

# KAM-RIGIDITY FOR PARABOLIC AFFINE ABELIAN ACTIONS

DANIJELA DAMJANOVIĆ, BASSAM FAYAD, AND MARIA SAPRYKINA

ABSTRACT. We show the following dichotomy for a linear parabolic  $\mathbb{Z}^2$ -action  $\rho_L$  on the torus with at least one step-2 generator:

- (i) Any affine  $\mathbb{Z}^2$ -action with linear part  $\rho_L$  has a  $\mathbb{Z}$ -factor that is either identity or genuinely parabolic, and is thus not KAM-rigid, or
- (ii) Almost every affine  $\mathbb{Z}^2$ -action with linear part  $\rho_L$  is KAM-rigid under volume preserving perturbations.

## CONTENTS

1. Introduction, statements and overview of the main proofs	2
1.1. Background and context	2
1.2. Main results and representative examples	4
1.3. Diophantine affine parabolic actions	9
1.4. KAM-rigidity	12
1.5. Examples of KAM-rigid actions	13
1.6. Overview of the proof of Theorem C	14
1.7. Comments on extensions and applications	17
1.8. Plan of the paper	18
2. Proof of Theorem C- the iteration part	18
2.1. Linearisation of the problem and the main iterative step: Proposition A	19
2.2. Iteration set-up	20
2.3. Convergence of the iterative scheme	24
2.4. Volume preservation of the conjugacy	25
3. Estimates of sums and double sums along the dual orbits	25
3.1. Notations	26
3.2. Estimates of the sums along the dual orbits of a step-2 matrix	28
3.3. Estimates of the double sums. The parabolic higher rank trick	29
4. Solution of the linearized problem. Proof of Proposition A	36
4.1. Approximating $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ by a coboundary	36
4.2. Application of Proposition D to truncated functions and vector fields	44
4.3. Proof of Proposition A	49
References	52

---

2010 *Mathematics Subject Classification.* 37C15, 37C85, 37D30.

*Key words and phrases.* Local rigidity, group actions, parabolic maps, KAM method.

## 1. INTRODUCTION, STATEMENTS AND OVERVIEW OF THE MAIN PROOFS

**1.1. Background and context.** A smooth (by which we mean  $C^\infty$ )  $\mathbb{Z}^k$ -action  $\rho$  on a smooth manifold  $M$  is said to be *locally rigid* if there exists a neighborhood  $\mathcal{U}$  of  $\rho$  in the space of smooth  $\mathbb{Z}^k$ -actions on  $M$ , such that for every  $\eta \in \mathcal{U}$  there is a smooth diffeomorphism  $h$  of  $M$  such that  $h \circ \rho(g) \circ h^{-1} = \eta(g)$ , for all  $g \in \mathbb{Z}^k$ .

When the rank of the acting group is  $k = 1$ , we are in the realm of classical dynamics ( $\mathbb{Z}$ -actions), where local rigidity in this strong form is not known to occur. Moreover, it is known that for affine maps on the torus local rigidity does not occur.

The only known situation in classical dynamics where a weaker form of local rigidity is proved, is the case of toral translations  $T_\alpha$  on  $\mathbb{T}^d$  with Diophantine frequency vectors  $\alpha$  (we exclude rigidity modulo infinite moduli from this discussion). Indeed, it follows from Arnold's normal form for perturbations of toral translations, that a volume preserving perturbation of  $T_\alpha$  with a Diophantine average translation vector  $\alpha$  is smoothly conjugated to  $T_\alpha$  [A]. We will call this phenomenon *KAM-rigidity*.

The situation is dramatically different for  $\mathbb{Z}^k$ -actions with  $k \geq 2$ , where local rigidity is more common.

For Anosov (hyperbolic) actions, an important breakthrough was the proof of local rigidity by Katok and Spatzier [KS]. The main tool in this context is the use of the action's invariant geometric structures [GK, KS], which after that proved useful in obtaining local rigidity for more general classes of partially hyperbolic actions with such geometric structures [NT, DK3, W1, W2, VW]. We note that there are many other local and global rigidity results for abelian partially hyperbolic actions than the ones mentioned above; we refrain from citing them all as our focus in this paper will be on local rigidity in the absence of any form of hyperbolicity and where there are no robust invariant geometric structures.

In this context two famous manifestations of rigidity for  $\mathbb{Z}^k$ -actions, are: KAM-rigidity of *simultaneously* Diophantine torus translations (see Definition 6) [M, DF, WX, P], and local rigidity for higher rank linear (and affine) partially hyperbolic actions on the torus [DK].

Simultaneously Diophantine torus translations generate  $\mathbb{Z}^k$ -actions which may have no Diophantine elements at all, so the result for single Diophantine translations does not apply, and one is forced to use commutativity of different action generators in a crucial way in order to obtain KAM rigidity.

For linear partially hyperbolic actions considered in [DK], the crucial assumption which leads to smooth rigidity is that such  $\mathbb{Z}^k$ -actions are of *higher rank*. A  $\mathbb{Z}^k$ -action by toral automorphisms is *higher rank* if there is a  $\mathbb{Z}^2$  subgroup such that all of its non-zero elements act by ergodic automorphisms. This condition is equivalent to the absence of  $\mathbb{Z}$ -factors of the action, namely: a higher rank  $\mathbb{Z}^k$ -action does not factor (possibly up to a finite index subgroup) to an action generated by a single automorphism of a (possibly different) torus. The condition is equivalent also to the exponential mixing for the action, which plays a crucial role in the proof of local rigidity in [DK]. Recently, it was announced in [W3] that exponential mixing leads to local rigidity for large classes of partially hyperbolic affine actions.

The above mentioned two classes of actions on the torus, elliptic ones generated by translations on one hand, and partially hyperbolic ones on the other hand, lie in the general class of *affine*  $\mathbb{Z}^k$ -actions on the torus. Affine actions are actions generated

by affine maps, and an affine map is a composition of a linear map and a translation. Such actions can be dynamically very different: they can be elliptic (the linear part of the action is the identity), or partially hyperbolic (the linear part contains a partially hyperbolic map, i.e., a map with some eigenvalues outside unit circle), or parabolic (the linear part acts by parabolic maps i.e. maps which have all eigenvalues 1), or can combine all these features. If the linear part contains a root of the identity, we take a finite index subgroup in the acting group which brings us to the general description above. In what follows we always assume that roots of the identity have been eliminated by passing to a finite index subgroup.

In this paper we focus on the most intricate and most surprising case of *parabolic actions*. In this case, single elements of the action (even under Diophantine conditions) are *not* KAM rigid, nor is there any mixing for the linear part of the action.

First we define a property of linear actions which will distinguish between actions that can be linear parts of KAM rigid affine actions, and those which cannot.

A linear  $\mathbb{Z}^2$ -action  $\rho_L$  on  $\mathbb{T}^d$  is *unlocked* if there is an affine  $\mathbb{Z}^2$ -action  $\rho$  with linear part  $\rho_L$  such that every  $\mathbb{Z}$ -factor of  $\rho$ , if it exists, is generated by a non-trivial translation. Thereby, we say a linear  $\mathbb{Z}^2$ -action  $\rho_L$  on  $\mathbb{T}^d$  is *locked* if for any affine action  $\rho$  with linear part  $\rho_L$ , there exists a  $\mathbb{Z}$ -factor which is *not* a translation  $T_\alpha$ ,  $\alpha \neq 0$ , see Definition 4.

Our analysis leads us to ask the following classification question for general affine  $\mathbb{Z}^2$ -actions on the torus:

**Question 1.** *Is it true that for any linear  $\mathbb{Z}^2$ -action  $\rho_L$  on  $\mathbb{T}^d$  there is the following dichotomy:*

- (i) *The action  $\rho_L$  is locked, or*
- (ii) *The action  $\rho_L$  is unlocked, and almost every affine action  $\rho$  with the linear part  $\rho_L$  is KAM-rigid.*

The previously mentioned classification results provide the positive answer to the above question for large classes of actions. For simultaneously Diophantine translation actions the linear part is the identity and thus is unlocked, so it fits into the case (ii) of the question. Since simultaneously Diophantine condition is a full measure condition, the works [M, DF, WX, P] give the positive answer to the question in this case. For affine actions with higher rank partially hyperbolic linear part the linear part is unlocked, since it has no  $\mathbb{Z}$ -factors at all. Hence, these actions also fall in category (ii). By [DK], such actions have the local rigidity property (and, therefore, KAM rigidity property as well; in fact, KAM rigidity holds for all affine actions of this kind, not just for almost all). Moreover, these two classes combined also give the positive answer to the question: [DF] proves KAM rigidity for a full measure set of affine  $\mathbb{Z}^2$ -actions on the torus whose linear part is a direct product of a higher rank action and the identity.

The main goal of this paper is to develop new tools to study this question for *parabolic* actions. Our main result states that the answer to the question is affirmative for parabolic actions containing a step-2 element (see Definition 1). We hope that the tools developed in this paper will allow to address the classification question 1 in full generality.

In this work the step-2 assumption on at least one element of the action is important in the proof, as we will discuss in detail in §1.6. In a nutshell, the reason is that if the cohomological equation above a step-2 affine map with a Diophantine translation part

has a solution, then the solution is tame. This is not the case for higher step affine maps [DFS].

**1.2. Main results and representative examples.** We will be interested in the problem of local rigidity of volume preserving perturbations of affine parabolic  $\mathbb{Z}^2$ -actions on the torus  $\mathbb{T}^d$ , where  $d \in \mathbb{N}$ . By  $\lambda$  we will denote the Haar measure on the torus  $\mathbb{T}^d$ .

We begin by defining some basic notions.

**Definition 1.** We say that  $A \in SL(d, \mathbb{Z})$  is *step- $S$  parabolic* if  $(A - \text{Id})^S = 0$ , and  $(A - \text{Id})^{S-1} \neq 0$ . An affine map of  $\mathbb{T}^d$  with the linear part  $A$  is defined via

$$a(x) = Ax + \alpha \pmod{1}, \quad x \in \mathbb{T}^d,$$

where  $\alpha \in \mathbb{R}^d$ . An affine map  $a$  is said to be *step- $S$*  if its linear part  $A$  is step- $S$ . We denote by  $\text{Aff}_S(\mathbb{T}^d)$  the space of all parabolic affine maps of step at most  $S$ .

A step- $S$  affine parabolic  $\mathbb{Z}^2$ -action on  $\mathbb{T}^d$  is a homomorphism  $\rho : \mathbb{Z}^2 \rightarrow \text{Aff}_S(\mathbb{T}^d)$ . We denote by  $\rho_L$  the action generated by the linear part of  $\rho$ .

Note that every affine map whose linear part is a unipotent matrix  $A \in SL(d, \mathbb{Z})$  (i.e., a matrix with all eigenvalues 1) is step- $S$  parabolic for some  $S \leq d$ .

Let  $\rho$  be an affine parabolic  $\mathbb{Z}^2$ -action on the torus  $\mathbb{T}^d$ . Then  $\rho$  is generated by two commuting affine maps  $a$  and  $b$ , and we will denote  $\rho$  simply by its generators as  $\langle a, b \rangle$ . The linear parts  $A$  and  $B$  of  $a$  and  $b$ , respectively, also commute. But the commutativity of the linear parts is not enough to guarantee the commutativity of  $a$  and  $b$ . If  $a(x) = Ax + \alpha \pmod{1}$  and  $b(x) = Bx + \beta \pmod{1}$ , then the commutator  $[a, b]$  is a translation on  $\mathbb{T}^d$  by the vector  $(A - \text{Id})\beta - (B - \text{Id})\alpha$ , which needs to be an integer in order to have  $[a, b] = \text{Id}$ . We may choose the lifts of  $a$  and  $b$  to  $\mathbb{R}^d$  so that the integer vector  $(A - \text{Id})\beta - (B - \text{Id})\alpha$  is trivial, and so without loss of generality we will always consider the lifts  $Ax + \alpha$  and  $Bx + \beta$  of  $a$  and  $b$ , respectively, such that  $(A - \text{Id})\beta = (B - \text{Id})\alpha$ . We will work with these lifts the whole time, and we denote them by the same letters  $a$  and  $b$ , respectively.

We will also use the shorthand notation  $A + \alpha$  to denote the affine map on  $\mathbb{T}^d$  with the linear part  $A$  and the translation part  $\alpha$ .

**Definition 2.** We denote by  $\mathcal{T}(A, B)$  the set of possible translation parts  $(\alpha, \beta)$  in the affine actions with linear part  $\langle A, B \rangle$ , that is

$$\mathcal{T}(A, B) := \{\alpha, \beta \in \mathbb{R}^d \mid (A - \text{Id})\beta = (B - \text{Id})\alpha\}.$$

Given an affine map  $a(x) = Ax + \alpha \pmod{1}$ , a perturbation of  $a$  is a diffeomorphism of the torus which can be lifted to  $\mathbb{R}^d$ , where it has the form

$$F(x) = A(x) + \alpha + f(x)$$

for some small  $\mathbb{Z}^d$ -periodic vector-valued function  $f$ . Therefore we will simply write  $a + f$  for a small perturbation of  $a$ . We will be interested in the set of smooth volume preserving perturbations, which we denote by  $\text{Diff}_\lambda^\infty(\mathbb{T}^d)$ .

We define now the notion of *KAM rigidity* which is central in this work. Note that an affine parabolic action always has a translation factor. Therefore, one cannot hope that any form of local rigidity, stronger than the one available for translations, can hold. Only local rigidity of KAM type can be expected for these actions.

**Definition 3.** We say that an affine  $\mathbb{Z}^2$ -action  $\langle a, b \rangle$  is KAM-rigid under  $\lambda$ -preserving perturbations, if there exists  $\sigma \in \mathbb{N}$ ,  $r_0 \in \mathbb{N}$ ,  $r_0 \geq \sigma$ , and  $\varepsilon > 0$  satisfying the following:

If  $r \geq r_0$  and  $\langle F, G \rangle = \langle a + f, b + g \rangle$  is a smooth  $\lambda$ -preserving  $\mathbb{Z}^2$ -action such that

$$\|f\|_r \leq \varepsilon, \quad \|g\|_r \leq \varepsilon, \quad \hat{f} := \int_{\mathbb{T}^d} f d\lambda = 0, \quad \hat{g} := \int_{\mathbb{T}^d} g d\lambda = 0, \quad (1)$$

then there exists  $H = Id + h \in \text{Diff}_\lambda^\infty(\mathbb{T}^d)$  such that  $\|h\|_{r-\sigma} \leq C(a, b) \varepsilon$  and

$$H \circ (a + f) \circ H^{-1} = a, \quad H \circ (b + g) \circ H^{-1} = b,$$

where  $C(a, b)$  is a constant depending only on the action  $\langle a, b \rangle$ .

In this paper, we will use the term *KAM-rigid* for short reference to *KAM-rigid under  $\lambda$ -preserving perturbations* since this will be the only context in which we place ourselves.

It is not difficult to find examples of parabolic commuting actions  $\langle A, B \rangle$  such that all the affine actions with this linear part are not KAM-rigid. Indeed, the commutation condition may force the affine action to have, for any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , a rank-one factor, to which Arnold's KAM-rigidity cannot be applied. As we will see in the following examples, this happens if the rank-one factor of the affine action is either identity or genuinely parabolic (i.e., has a non-trivial linear part). We call such pairs  $\langle A, B \rangle$  *locked* (Definition 4 below).

Let  $E_{ij}$  denote the integer matrix which has 1 in the position  $(i, j)$ , the rest of the elements being 0.

**Example 1** (Affine actions with identity as a rank-one factor). Consider the linear action on  $\mathbb{T}^3$  generated by  $A = Id + E_{21}$ ,  $B = Id + E_{31}$ . The commutation condition implies that for all  $(\alpha, \beta) \in \mathcal{T}(A, B)$  we have  $\alpha_1 = \beta_1 = 0$ . Hence, any affine action  $\langle A + \alpha, B + \beta \rangle$ , restricted to a sub-torus corresponding to the variable  $x_1$ , equals identity (we say that the action projects to identity on the torus  $\mathbb{T}_1$  spanned by variable  $x_1$ , i.e., the action has the identity factor). Consequently, such affine action is not KAM-rigid, as explained below.

To see why the action in Example 1 is not KAM-rigid, we can use the following general fact.

**Proposition 1.** If a parabolic commuting action  $\langle A, B \rangle$  is lower triangular, and  $\langle A + \alpha, B + \beta \rangle$  is such that  $\alpha_1 = \beta_1 = 0$  then  $\langle A + \alpha, B + \beta \rangle$  is not KAM-rigid.

*Proof.* Keep  $b = B + \beta$  unchanged and perturb  $a = A + \alpha$  to  $F(x_1, \dots, x_d) = Ax + \alpha + (0, \dots, 0, \varepsilon \sin(2\pi x_1))$ . The maps  $b$  and  $F$  commute, and  $F$  satisfies (1). To see that  $A + \alpha$  is not conjugated to  $F$  by a volume preserving conjugacy, consider the special two-dimensional case:  $a(x_1, x_2) = A(x_1, x_2) = (x_1, x_2 + x_1)$  and  $F(x_1, x_2) = (x_1, x_2 + x_1 + \varepsilon \sin(2\pi x_1))$  (the general case is not different from it). Define the circle diffeomorphism  $g : x_1 \mapsto x_1 + \varepsilon \sin(2\pi x_1)$  and the conjugacy  $H(x_1, x_2) = (g(x_1), x_2)$ . Clearly,  $H$  does not preserve area, and it is easy to see that, up to translation,  $H$  is the only conjugacy between  $a$  and  $F$ .  $\square$

**Remark 1.** Note that if  $\langle a, b \rangle$  has identity as a rank-one factor of dimension at least 2, then it is straightforward to perturb the action as in (1) so that there is no conjugacy at all with the affine action. To do this, it is enough to perturb the identity factor itself in a volume preserving way such that (1) holds.

**Example 2** (Affine actions with a genuinely parabolic rank-one factor). *Let us assume that*

$$B = Id + E_{21} + E_{32}.$$

*If  $A$  is lower triangular and  $(A - Id)$  does not contain neither  $E_{2j}$  nor  $E_{3j}$  for any  $j$ , then for any  $(\alpha, \beta) \in \mathcal{T}(A, B)$  we have that  $\alpha_1 = \alpha_2 = 0$ , and  $A + \alpha$  acts as identity on the two-torus obtained by the projection on  $(x_1, x_2)$ . Then the affine action  $\langle A + \alpha, B + \beta \rangle$  has the skew shift  $(x_1, x_2) \mapsto (x_1 + \beta_1, x_2 + x_1 + \beta_2)$  of the two-torus as a rank-one factor, and is thus not KAM-rigid, as explained below.*

To see why the action in Example 2 is not KAM-rigid, we can use the following general fact.

**Proposition 2.** *For any  $r \in \mathbb{N}$ , for any parabolic affine map  $A + \alpha$  on  $\mathbb{T}^d$  with  $A \neq Id$ , for any  $\varepsilon > 0$  there exists  $f$  such that*

$$\|f\|_r \leq \varepsilon, \quad \widehat{f} := \int_{\mathbb{T}^d} f d\lambda = 0,$$

*and  $A + \alpha + f \in \text{Diff}_\lambda^\infty(\mathbb{T}^d)$  is not conjugated to  $A + \alpha$ .*

*Proof.* Without loss of generality, we can consider the case of the two-dimensional skew shift  $(x_1, x_2) \mapsto (x_1 + \beta_1, x_2 + x_1 + \beta_2)$ . This map can be perturbed into  $(x_1, x_2) \mapsto (x_1 + \beta_1 + \varepsilon \sin(2\pi(x_1 + x_2)), x_2 + x_1 + \beta_2)$ , which is a shifted classical standard map that is not conjugate to a skew shift.  $\square$

The phenomena in Examples 1 and 2 can be subsumed under the existence, for any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , of a rank-one factor for the affine action  $\langle A + \alpha, B + \beta \rangle$  that is either identity or a genuinely parabolic action, which overrules KAM-rigidity. This motivates the following definition.

**Definition 4** (Locked actions). *When the commuting linear action  $\langle A, B \rangle$  is such that for any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the affine action  $\langle A + \alpha, B + \beta \rangle$  has a rank-one factor that is either identity or a genuinely parabolic action (i.e., has a non-trivial linear part), we say that  $\langle A, B \rangle$  is locked. We call the action unlocked if it is not locked.*

An immediate corollary of Propositions 1 and 2 is the following.

**Corollary 3.** *If  $\langle A, B \rangle$  is locked, then for any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the action of  $\langle a, b \rangle$  is not KAM-rigid.*

The main result of this paper is to show that besides the locked actions, for actions having (at least) one step-2 generator, KAM-rigidity under  $\lambda$ -preserving perturbations holds almost surely in the choice of the translation part. We formulate this dichotomy as follows.

**Theorem A.** *Given a commuting action  $\langle A, B \rangle$  of parabolic matrices, where  $A$  is step-2, we have the following dichotomy.*

- (i) *Action  $\langle A, B \rangle$  is locked, thus for any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the action of  $\langle a, b \rangle$  is not KAM-rigid.*
- (ii) *Action  $\langle A, B \rangle$  is unlocked, and for almost every choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the action of  $\langle a, b \rangle$  is ergodic and KAM-rigid under volume preserving perturbations.*

In the case of step-2 actions, we have a more stringent alternative.

**Theorem B.** *Given a commuting pair  $\langle A, B \rangle$  of step-2 parabolic matrices, we have the following dichotomy.*

- (i) *For any choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the action of  $\langle a, b \rangle$  has a rank-one factor that is identity, and is therefore not ergodic and not locally rigid.*
- (ii) *For almost every choice of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the action of  $\langle a, b \rangle$  is ergodic and KAM-rigid under volume preserving perturbations.*

Corollary 3 states that (i) impedes KAM-rigidity. The proof of the dichotomy between (i) and (ii) is the main result of this paper that we formulate more precisely *via* Proposition 5 and Theorem C of the next section.

We note that the statements remain true if we replace the preservation of the volume  $\lambda$  (and also averages with respect to the volume) by preservation of any common invariant measure.

In the following observations we discuss the relevance of the assumptions of the main theorems.

**Observation 1** (KAM-rigidity *vs.* local rigidity: why do we need Diophantine conditions?). When a linear  $\mathbb{Z}^2$ -action is not higher rank, it has a rank-one factor that we can represent by a pair  $\langle \text{Id}, C \rangle$ . The absence of rigidity of a single linear map  $C$  implies that the local rigidity in this case can only be considered for affine actions. This work treats rigidity of unlocked affine parabolic actions under Diophantine conditions on the translation vectors of the action. Roughly speaking, we put the Diophantine conditions on all the frequency vectors associated to sub-tori on which some element of the action acts as a translation. As we will see in the next section, there may be a finite or infinite number of such conditions. However, there will unavoidably be some Diophantine conditions that must be satisfied, thus only KAM type rigidity can be considered. The necessary set of Diophantine conditions comes from the fact that the affine action always has a translation part (due to the existence of a common eigenspace of eigenvalue 1 for the commuting linear pair, see Definition 5 and §4.1.2). The other Diophantine conditions (cf. Definition 8) that are used are natural conditions that play a crucial role in the proof (cf. §4.1.3), although we do not see for the moment how to show that they are necessary for the result to hold (see Question 2).

**Observation 2** (On higher step parabolic actions: why do we assume that one generator is step-2?). The assumption that one element of the action is step-2 plays a heavy role in the proof. In particular, it is crucial in defining a tame candidate for a conjugacy at each step of the inductive KAM conjugacy scheme. More details about the use of this assumption will be given in §1.6. We note that working under this, probably restrictive, assumption requires introducing some new ideas and techniques. Moreover, some phenomena, like resonances, appear only along sub-tori where both generators are step-2 (see §1.3). This is why we prefer to focus on the special case described by Theorem A and keep the study of the general case for a future work.

**Observation 3** (From parabolic actions to general affine actions: why do we focus on parabolic actions?). If a non higher rank linear action  $\langle A, B \rangle$  has a rank-one factor  $\langle \text{Id}, C \rangle$  where  $C$  does not have 1 as an eigenvalue, then, due to the commutation constraint,  $\text{Id}$  is still a factor for any affine action with the linear part  $\langle A, B \rangle$ . As a consequence, local rigidity would fail for the affine actions the same way it fails for the linear one. Based on this argument and on the fact that higher rank actions (i.e., those



for which all the elements of the linear part are ergodic, thus partially hyperbolic) are locally rigid [DK], the case of parabolic actions naturally appears as the main problem to settle in order to give a general classification of affine abelian actions on the torus in terms of KAM local rigidity.

**Observation 4** (The case of  $\mathbb{Z}^k$  actions for  $k \geq 3$ ). The same methods we use here provide the KAM-rigidity result for certain classes of affine  $\mathbb{Z}^k$ -actions with  $k \geq 3$ , as well, see Remark 4 for more details. It is useful to note that the higher the rank  $k$  of the acting group is, the "more locked" the action can become. An example of maximal rank parabolic abelian linear action on  $\mathbb{T}^4$  is the  $\mathbb{Z}^4$ -action generated by  $\text{Id} + E_{12}, \text{Id} + E_{14}, \text{Id} + E_{32}$  and  $\text{Id} + E_{34}$ . This action is completely locked in the strongest possible sense: there are no affine non-linear  $\mathbb{Z}^4$ -actions which have this action as a linear part, at all. We expect that the same holds for any maximal rank parabolic abelian linear action on any  $\mathbb{T}^d$ , they are locked. In these cases there is no KAM-rigidity. This points to the fact that parabolic  $\mathbb{Z}^2$ -actions are the most common situation in which we could expect to have KAM-rigidity.

**Observation 5** (The role of commutativity in the KAM-rigidity. The parabolic higher rank trick). As in [M] and [DK], our proof of local rigidity relies on a KAM inductive conjugacy scheme (see §1.6 for an outline of the scheme). At each step of the scheme, a system of cohomological equations must be solved up to a quadratic error.

In [M], each equation of the system is a cohomological equation above a circle rotation. Hence each individual equation has a formal solution provided vanishing of averages, but this solution may not be tame because each individual angle is not necessarily Diophantine. The main observation by Moser is that a cocycle relation forces the formal solutions to coincide and to be tame. This implies that the commutation relation allows one to find a tame solution to the system up to a quadratic error.

In [DK], the system of cohomological equations consists of individual equations that have tame solutions modulo a countable set of obstructions. The commutation, or the higher rank trick, is used to show that these obstructions can be removed up to a quadratically small error.

As it will be explained in detail in §1.6, our approach to proving KAM-rigidity for parabolic actions combines these two mechanisms of local rigidity.

The main challenge in our work is to replace the partially hyperbolic higher rank trick by a parabolic one. To explain this a little better, we risk a technical description that may look obscure now but that will become much clearer from the detailed overview of the proof in §1.6 as well as from the introduction of Section 4.

The higher rank trick usually relies on the exponential growth of integer vectors (Fourier frequencies) under the dual action of partially hyperbolic matrices. Also, the mechanism that lies at the heart of the higher rank trick of [DK] is the following consequence of the partial hyperbolicity of the action: for an integer vector  $\bar{m}$  that is lowest (with smallest norm) on its orbit under the dual action  $\bar{A}$  of the first generator of the action, it is possible to iterate by  $\bar{B}$  in one of the two directions (future or past) so that  $\|\bar{A}^k \bar{B}^l \bar{m}\|$  be always larger than  $c\|\bar{m}\|$ .

A main difficulty in our work is to replace the above argument by the fact that parabolic actions only grow at a polynomial rate. Much more annoying is the fact that for some integer vectors  $\bar{m}$  that are lowest on their  $\bar{A}$  orbit, it is possible that the



iterates by  $\bar{B}^l$  be decreasing in norm during a long time in both directions of  $l$ , before starting to increase. The challenge is to make sure that by choosing one direction of iteration for  $\bar{B}$ , the double iterates  $\|\bar{A}^k \bar{B}^l m\|$  remain always larger than  $\|m\|^\delta$  for some  $\delta > 0$  independent of  $m$  ( $\delta$  is comparable to  $1/S$  where  $S$  is the step of the action). This parabolic version of the higher rank trick is done in §3.3, where "being unlocked" property of the action again plays a major role.

There has been very few local rigidity results for parabolic actions. One example are actions by left multiplication on nilmanifolds. These are parabolic, and a form of local rigidity for such  $\mathbb{R}^2$ -actions on 2-step nilmanifolds was obtained in [D], under Diophantine conditions. Results of similar type were obtained for  $\mathbb{Z}^2$ -actions on Heisenberg nilmanifolds in [DT]. More recently, in [ZW], it is proved that certain large abelian parabolic actions on homogeneous spaces of semisimple Lie groups have strong local rigidity properties.

In the remainder of the introduction we give precise definition and precise formulation of the main rigidity result, as well as the overview of the proofs, examples and comments on possible applications.

**1.3. Diophantine affine parabolic actions.** In this section we define the full measure Diophantine conditions required on the pair  $(\alpha, \beta) \in \mathcal{T}(A, B)$  in order to guarantee KAM-rigidity. It will be a combination of two types of conditions: simultaneously Diophantine condition for the maximal translation factor of the action, and Diophantine conditions for the translation parts of special elements of the action that we refer to as resonances.

#### 1.3.1. The maximal translation factor.

**Definition 5 (Maximal translation factor).** *We say that the action  $\langle A, B \rangle$  has a maximal identity factor if there is a torus  $\mathbb{T}_1$  of dimension  $d_1$  such that  $\langle A, B \rangle$ , restricted to this torus, equals identity. The action of  $\langle a, b \rangle$  restricted to this factor is called the maximal translation factor of  $\langle a, b \rangle$ .*

**Definition 6 (Simultaneously Diophantine vectors).** *We say that a pair of vectors  $(\alpha, \beta) \in \mathbb{T}^d \times \mathbb{T}^d$  is simultaneously Diophantine if there exists  $\gamma, \tau > 0$  such that*

$$\max\{|1 - e(k, \alpha)|, |1 - e(k, \beta)|\} > \frac{\gamma}{|k|^\tau},$$

where  $e(m, x) = e^{2\pi i(m, x)}$ . We denote this property by  $(\alpha, \beta) \in \text{SDC}(\gamma, \tau)$ .

Observe that SDC-pairs of vectors form a set of full Haar measure in  $\mathbb{T}^d \times \mathbb{T}^d$ .

**1.3.2. Resonant vectors.** In what follows we will use the dual action corresponding to the linear part  $\langle A, B \rangle$ , induced on  $\mathbb{Z}^d$ . For a matrix  $A$ , the dual action on  $\mathbb{Z}^d$  is denoted by

$$\bar{A} := (A^{tr})^{-1}, \quad \bar{A} = \text{Id} + \hat{A}.$$

For a general  $m$  and  $(k, l) \in \mathbb{Z} \times \mathbb{Z}$ ,  $\bar{A}^k \bar{B}^l m$  has a polynomial expression (see Lemma 13). However, if  $m$  is such that there exists a pair  $(k, l) \in \mathbb{Z} \times \mathbb{Z} \setminus \{(0, 0)\}$  satisfying  $\bar{A}^k \bar{B}^l m = m$ , then, since this implies that  $\bar{A}^{ik} \bar{B}^{il} m = m$  for all  $i \in \mathbb{Z}$ , we necessarily have (even if  $A$  and  $B$  are higher step):

$$\bar{A}^k \bar{B}^l m - m = k \hat{A} m + l \hat{B} m = 0.$$

Hence, if  $\bar{A}m \neq m$  or  $\bar{B}m \neq m$ , we can associate to such an  $m$  a unique pair  $(k, l) \in \mathbb{N} \times \mathbb{Z}$  such that either  $(k, l) = (1, 0)$ , or  $(k, l) = (0, 1)$ , or  $k$  and  $l$  are mutually prime and  $k > 0$ . For all these cases, we use the same notation  $k \wedge l = 1$  and say that  $m$  is resonant and that  $(k, l)$  is its associated resonance pair. Notice that, due to commutativity, if  $m$  is resonant, then any other integer vector on the  $\langle \bar{A}, \bar{B} \rangle$ -orbit of  $m$  is also resonant with the same resonance pair. So resonance pairs are attached to orbits, rather than individual vectors. We summarise the above discussion in the following

**Definition 7 (Resonant vectors and resonance pairs).** Any vector  $m \in \mathbb{Z}^d \setminus \{0\}$  such that  $\bar{A}^k \bar{B}^l m = m$  for some  $(k, l)$ , while either  $\bar{A}m \neq m$  or  $\bar{B}m \neq m$ , is called a resonant vector. We will use the following notations:

$\mathcal{C}_2(k, l)$  denotes the set of all resonant  $m$  associated to the resonance pair  $(k, l)$ ,

$\mathcal{C}_2 = \mathcal{C}_2(A, B) = \bigcup_{k \wedge l = 1} \mathcal{C}_2(k, l)$  denotes the set of all resonant vectors,

$\mathcal{Q}(A, B)$  denotes the set of all resonance pairs  $(k, l) \in \mathbb{Z}^2$ .

The following lemma shows that the norm of the resonant pair is bounded by the norm of any of the corresponding resonances. It is therefore bounded by the smallest one of them on the  $\langle \bar{A}, \bar{B} \rangle$ -orbit.

**Lemma 4.** Let  $a = A + \alpha$  and  $b = B + \beta$  be commuting affine parabolic maps. If  $(k, l) \in \mathbb{Z}^2$  is the (unique) pair associated to the resonance  $m$  as in Definition 8, then there exists  $C = C(A, B) > 0$  such that

$$C(|k| + |l|) \leq |m|.$$

*Proof.* Let  $m$  be a resonant vector, i.e., let  $\bar{A}^k \bar{B}^l m = m$ . As explained earlier, this implies that  $k\hat{A}m = -l\hat{B}m$ . Consider now the two integer vectors:  $x = \hat{A}m$  and  $y = \hat{B}m$ . By assumption,  $k$  and  $l$  are mutually prime. This implies, in particular, that each component  $x_j$  of the vector  $x$  is divisible by  $l$ . Hence,  $x_j \geq |l|$ . Therefore, there exists a constant  $C(A)$  (depending on  $\hat{A}$ ) such that  $|m| \geq C(A)|l|$ . In the same way,  $|m| \geq C(B)|k|$ , which implies the statement.  $\square$

**Definition 8 (Diophantine resonances).** Let  $\langle a, b \rangle$  be an affine parabolic  $\mathbb{Z}^2$ -action. The number  $\alpha_{k,l} = a^k b^l - A^k B^l$  will be called the translation part of the element  $(k, l)$  of the action.

We say that a resonance  $m \in \mathcal{C}_2(k, l)$  is  $(\gamma, \tau)$ -Diophantine, if

$$|1 - e(m, \alpha_{k,l})| > \frac{\gamma}{|m|^\tau}. \quad (2)$$

**Remark 2.** The set of resonant vectors and resonance pairs for a given action may be empty, finite non empty, or infinite, as we will see in the examples at the end of this section.

**1.3.3. Diophantine property for actions.** We are ready to define the Diophantine parabolic affine actions, for which the main local rigidity result holds.

**Definition 9 (Diophantine actions).** Given  $\gamma, \tau > 0$  and a parabolic affine  $\mathbb{Z}^2$ -action  $\langle a, b \rangle$ , where  $a$  is step-2, we say that  $\langle a, b \rangle$  is  $(\gamma, \tau)$ -Diophantine if:

- (1) the maximal translation factor of  $\langle a, b \rangle$  is  $(\gamma, \tau)$ -simultaneously Diophantine (as in Definition 6), and
- (2) every resonance  $m \in \mathcal{C}_2(A, B)$  is  $(\gamma, \tau)$ -Diophantine (as in Definition 8).

**Example 3.** Let  $\langle A, B \rangle$  be the action on  $\mathbb{T}^2$  generated by  $A = Id + E_{21}$  and  $B = Id$ . The affine action  $\langle A + (\alpha, 0), B + (0, \beta) \rangle$  is  $(\gamma, \tau)$ -Diophantine if and only if both  $\alpha$  and  $\beta$  are  $(\gamma, \tau)$ -Diophantine.

*Proof.* Indeed, in this case  $\alpha_1 = \alpha, \beta_1 = 0$  is the translation part of the affine action, and the SDC condition reduces to the Diophantine condition on  $\alpha$ .

In this example all the vectors  $(m_1, m_2)$  with  $m_2 \neq 0$  are resonant with the same resonance pair  $(0, 1)$ , and the Diophantine condition on the resonance reduces to the Diophantine condition on  $\beta$ .  $\square$

The following simple observation is an important step in establishing the dichotomies in Theorems A and B.

**Proposition 5.** Fix  $\tau > d$ . Let  $\langle A, B \rangle$  be a linear parabolic  $\mathbb{Z}^2$ -action. We have the following alternative:

- (i)  $\langle A, B \rangle$  is locked as in Definition 4.
- (ii)  $\langle A, B \rangle$  is unlocked, and for almost every  $(\alpha, \beta) \in \mathcal{T}(A, B)$ ,  $\langle A + \alpha, B + \beta \rangle$  is  $(\gamma, \tau)$ -Diophantine for some  $\gamma > 0$ .

In case  $A$  and  $B$  are step-2, alternative (i) can be reduced to the existence of a rank-one factor that is identity.

**Remark 3.** Notice that no step-2 assumption is made on any generator in the first part of the Proposition.

*Proof of Proposition 5.* Consider the maximal translation factor of  $\langle a, b \rangle$  generated by the pair of translation vectors denoted by  $(\alpha^{(1)}, \beta^{(1)})$ . The condition  $(\alpha, \beta) \in \mathcal{T}(A, B)$  imposes some relations over  $\mathbb{Z}$  between the coordinates of the vectors  $\alpha^{(1)}$  and  $\beta^{(1)}$ . Then we have two possible scenarios.

The first one is that there exists a vector  $\bar{k} \in \mathbb{Z}^{d_1} \setminus \{0\}$  such that  $(\bar{k}, \alpha^{(1)}) = (\bar{k}, \beta^{(1)}) = 0$  for all  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , in which case a change of coordinates with  $X_1 := (\bar{k}, x^{(1)})$  will exhibit a one-dimensional rank-one factor on which the action is identity.

If the first scenario does not hold, then for all  $k \in \mathbb{Z}^{d_1} \setminus \{0\}$  either  $(k, \alpha^{(1)})$  or  $(k, \beta^{(1)})$  is not identically zero on  $\mathcal{T}(A, B)$ . In this case we can split the set of integers,  $\mathbb{Z}^{d_1} \setminus \{0\} = \mathcal{Z}_1 + \mathcal{Z}_2$ , in such a way that for  $k \in \mathcal{Z}_1$ ,  $(k, \alpha^{(1)})$  is not identically zero on  $\mathcal{T}(A, B)$ , and for  $k \in \mathcal{Z}_2$ ,  $(k, \beta^{(1)})$  is not identically zero on  $\mathcal{T}(A, B)$ . Now, for any  $k \in \mathcal{Z}_1$ , for any  $\delta > 0$ , we have that  $\lambda \{ \alpha^{(1)} \in \mathbb{T}^{d_1} : \|(k, \alpha^{(1)})\| \leq \delta |k|^{-d_1-1} \} \leq c \delta |k|^{-d_1-1}$  for a constant  $c = c(d)$ . Summing over all  $k \in \mathcal{Z}_1$  and then over all  $k \in \mathcal{Z}_2$ , and using Arcela-Ascoli theorem, we get that for almost every  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , there exists  $\gamma > 0$  such that for each  $k \in \mathcal{Z}_1$ , it holds that  $\|(k, \alpha^{(1)})\| \geq \gamma |k|^{-d_1-1}$ , and for each  $k \in \mathcal{Z}_2$  it holds that  $\|(k, \beta^{(1)})\| \geq \gamma |k|^{-d_1-1}$ . This implies that for almost every  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , the maximal translation factor of  $\langle a, b \rangle$  is  $(\gamma, \tau)$ -simultaneously Diophantine.

Next, we consider a resonance  $m \in \mathcal{C}_2(A, B)$  and let  $k \wedge l = 1$  be its unique corresponding vector such that  $k\hat{A}m + l\hat{B}m = 0$ . We then have two possible cases.

CASE 1. There exists a resonance  $m$  such that for every  $(\alpha, \beta) \in \mathcal{T}(A, B)$  it holds that  $e(m, \alpha_{k,l}) = 1$ . Then we prove the following.

**Lemma 6.** In the assumptions of Case 1, the action  $\langle a, b \rangle$  has a rank-one factor that is genuinely parabolic.

*Proof.* Let  $m \in \mathcal{C}_2(k, l)$  be a resonance, in which case for  $A' := A^k B^l$  we have  $\bar{A}'m = m$ , while either  $\bar{A}m \neq m$  or  $\bar{B}m \neq m$ . For definiteness, assume that  $\bar{B}m \neq m$ . We also have that for  $\alpha' := \alpha_{k,l}$ ,  $e(m, \alpha') = 1$ .

After a change of variables we can assume that  $m$  is one of the basis vectors, that is,  $m_i = 0$  for  $i \neq i_1$  and  $m_{i_1} = 1$ . We also assume that both matrices  $B$  and  $A'$  have 1-s on the main diagonal.

Since  $\bar{B}m \neq m$  and  $\bar{A}'m = m$ , we have that  $B$  contains some  $E_{i_1 i_2}$  while  $A'$  does not contain any  $E_{i_1 *}$  (where  $*$  ranges through possible indices). By another change of coordinates, we can assume that  $B$  does not contain any other  $E_{i_1 *}$  besides  $E_{i_1 i_2}$ . (Indeed, if  $B$  contains  $\sum_{j=2}^s k_j E_{i_1 i_j}$ , we can use the coordinate change  $x_{i_2} \mapsto \sum_{j=2}^s k_j x_{i_j}$ ,  $x_k \mapsto x_k$  for  $k \neq i_2$ ).

By the commutativity of  $a'$  and  $b$  we get that  $\alpha'_{i_2} = 0$ , and that  $A'$  contains no  $E_{i_2 *}$  (otherwise  $\hat{A}'\hat{B}$  would contain no  $E_{i_1 *}$  while  $\hat{B}\hat{A}'$  would contain some). Also, the hypothesis  $e(m, \alpha_{k,l}) = 1$  translates into  $\alpha'_{i_1} = 0$ .

If  $B$  has no element of the type  $E_{i_2 *}$ , we conclude that the action  $\langle a', b \rangle$  where  $a' = a^k b^l$  factors on the torus  $\mathbb{T}_{i_2, i_1}$  on which  $a'$  acts as identity, while  $b$  is genuinely parabolic.

If  $B$  had an element  $E_{i_2 i_3}$ , then again after a change of coordinates, we can assume that  $B$  does not contain any other  $E_{i_2 *}$  besides  $E_{i_2 i_3}$ .

As before, we have two consequences: 1)  $\alpha'_{i_3} = 0$ , and 2)  $A'$  contains no  $E_{i_3 *}$  (otherwise  $\hat{A}'\hat{B}$  would contain no  $E_{i_2 *}$  while  $\hat{B}\hat{A}'$  would contain some).

If  $B$  has no element of the type  $E_{i_3 *}$  we conclude that the torus  $\mathbb{T}_{i_3, i_2, i_1}$  is a factor of the action  $\langle a', b \rangle$  on which  $a' = a^k b^l$  acts as identity, while  $b$  is genuinely parabolic.

Arguing inductively, we obtain the proof of the lemma.  $\square$

If Case 1 does not hold, then we must be in the following case:

CASE 2. For every resonance  $m$ , there exists  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , such that  $e(m, \alpha_{k,l}) \neq 1$ , then by linearity of  $\alpha_{k,l}$  in the variables of  $(\alpha, \beta)$ , we see that the measure of  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , such that

$$|1 - e(m, \alpha_{k,l})| \leq \frac{\gamma}{|m|^{d+1}}$$

is less than  $c \frac{\gamma}{|m|^{d+1}}$  for a constant  $c = c(d)$ . Summing up over all possible resonances and using Arcela-Ascoli theorem, we get that for almost every  $(\alpha, \beta) \in \mathcal{T}(A, B)$ , there exists  $\gamma > 0$  such that every resonance is  $(\gamma, d+1)$  Diophantine.

From the proof of Lemma 6, we see that Case 1 cannot happen if the action is step-2.  $\square$

**1.4. KAM-rigidity.** Now we are ready to formulate precisely part (ii) of Theorems A and B. The following is our main rigidity result.

**Theorem C.** *Let  $\langle A, B \rangle$  be an unlocked linear parabolic  $\mathbb{Z}^2$ -action with (at least) one step-2 generator. If  $(\alpha, \beta) \in \mathcal{T}(A, B)$  are such that  $\langle a, b \rangle$  is  $(\gamma, \tau)$ -Diophantine for some  $\gamma > 0$  and  $\tau > 0$ , then  $\langle a, b \rangle$  is KAM-rigid.*

Theorems A and B follow directly from Theorem C and Proposition 5.

**Remark 4.** *If an affine  $\mathbb{Z}^k$ -action,  $k \geq 3$ , contains a Diophantine affine  $\mathbb{Z}^2$ -action with at least one step-2 generator, then Theorem C directly implies KAM-rigidity for the  $\mathbb{Z}^k$ -action. This is because the smooth conjugacy provided by Theorem C for the  $\mathbb{Z}^2$ -action*

would then conjugate the whole  $\mathbb{Z}^k$ -action perturbation. This simple observation is a consequence of the commutation and the ergodicity of the Diophantine  $\mathbb{Z}^2$ -action. This argument has been already used in [DK] (see Lemma 3.2 in [DK]) to draw the conclusion about local rigidity for a  $\mathbb{Z}^k$ -action from that of its  $\mathbb{Z}^2$ -subaction.

### 1.5. Examples of KAM-rigid actions.

To begin with, let us return to the simple Example 3.

**Proposition 7.** *Let  $\langle A, B \rangle$  be the action on  $\mathbb{T}^2$  given by  $A = Id + E_{21}$  and  $B = Id$ . If  $\alpha$  and  $\beta$  are Diophantine numbers, then the affine action  $\langle A + (\alpha, 0), B + (0, \beta) \rangle$  is KAM-rigid.*

*Proof.* As explained earlier, when  $\alpha_1$  is Diophantine, the translation factor of the action is SDC. On the other hand, all the resonances are of the form  $(m_1, m_2) = (0, m_2)$ ,  $m_2 \neq 0$ , with the corresponding resonance pair  $(0, 1)$ . Since  $\alpha_{0,1} = (0, \beta)$ , we have the following. When  $\beta$  is  $(\gamma, \tau)$ -Diophantine, condition (2) holds with the constants  $(\gamma, \tau)$ . Hence, Theorem C implies the KAM-rigidity of  $\langle A + (\alpha, 0), B + (0, \beta) \rangle$ .  $\square$

It is clear that if  $\alpha$  is Liouville, the corresponding action will not be KAM-rigid. To see this, just perturb  $A + (\alpha, 0) = (x_1, x_2) \mapsto (x_1 + \alpha, x_2 + x_1)$  to  $(x_1, x_2) \mapsto (x_1 + \alpha, x_2 + x_1 + \varepsilon\varphi(x))$ , where  $\varphi(x)$  is a smooth function with the zero mean that is not a coboundary above the rotation of angle  $\alpha$ . However, although our proof of KAM-rigidity heavily uses the Diophantine property of resonances, we are not able to settle whether  $\beta$  Diophantine is a necessary condition for KAM-rigidity.

**Question 2.** *Is the action  $\langle A + (\alpha, 0), B + (0, \beta) \rangle$  KAM-rigid when  $\alpha$  is Diophantine and  $\beta$  is Liouville?*

The following example provides a KAM-rigid action having infinitely many resonances with infinitely many resonance pairs.

*Proof.* The commutation condition  $(\alpha, \beta) \in \mathcal{T}(A, B)$  is satisfied if and only if  $\beta_1 = \alpha_4$ ,  $\beta_2 = 0$ , and  $\beta_3 = \alpha_3 = -\alpha_2$ . The translation factor of the action is the three-torus corresponding to the first three coordinates, and the translations are  $\alpha^{(1)} = (\alpha_1, \alpha_2, -\alpha_2)$  and  $\beta^{(1)} = (\alpha_4, 0, -\alpha_2)$ . It is easy to see that if the vectors  $(\alpha_1, \alpha_2)$  and  $(\alpha_4, -\alpha_2)$  are Diophantine, then the pair  $(\alpha^{(1)}, \beta^{(1)})$  is SDC. Indeed, denoting by  $\|x\|$  the closest distance from  $x \in \mathbb{R}$  to the integers, for any  $m = (m_1, m_2, m_3)$  we have:

$$\begin{aligned} \max(\|(m, \alpha^{(1)})\|, \|(m, \beta^{(1)})\|) &= \max(\|m_1\alpha_1 + (m_2 - m_3)\alpha_2\|, \|m_1\alpha_4 - m_3\alpha_2\|) \\ &\geq \gamma(|m_1| + |m_2 - m_3|)^{-\tau} + \gamma(|m_1| + |m_3|)^{-\tau} \\ &\geq \gamma'|m|^{-\tau'}. \end{aligned}$$

Let us turn to the resonances. They are the set of  $m$  such that  $m_6 = 0$ , and the two vectors,  $v_m = (m_5, m_7)$  and  $w_m = (m_4, m_4 + m_7)$ , are collinear and not both zero at the same time. Hence, at least one of  $m_5$  or  $m_7$  does not vanish. The resonance pairs are the pairs  $k_m \wedge l_m = 1$  such that  $k_m v_m + l_m w_m = 0$ . For example,

$$m = (m_1, m_2, m_3, n + 1, n, 0, n(n + 1)), \quad (m_1, m_2, m_3, n) \in \mathbb{Z}^3 \times \mathbb{Z}^*$$

is a resonant vector with the resonance pair  $(n + 1, -n)$ . Finally, fix any resonant  $m$  and observe that, since there are no constraints on  $\alpha_5, \alpha_7$  and since at least one of  $m_5$

or  $m_7$  does not vanish, the Diophantine condition on  $\alpha_{k_m, l_m}$  is satisfied for almost every  $(\alpha, \beta) \in \mathcal{T}(A, B)$ .  $\square$

**1.6. Overview of the proof of Theorem C.** The proof is based on an inductive scheme of successive conjugations of the perturbed action  $\langle F, G \rangle$ , where  $F = a + f$  and  $G = b + g$  to the affine action  $\langle a, b \rangle$ . As usually in the KAM approach, the linearized conjugacy equations are solved at each step of the induction with a loss of derivatives, which can be caused, for example, by small divisors or by other reasons. The *a priori* damaging effect of this loss is tamed out by the quadratic speed of convergence of the scheme.

In our context, the linearized conjugacy equations, often called the cohomological equations, are essentially of the following form:

$$\begin{aligned} h \circ a - Ah &= f, \\ h \circ b - Bh &= g. \end{aligned} \tag{3}$$

Two main differences with the classical KAM schemes that appear in our context are the following:

- (i) Cohomological equations (3) above each individual generator of the action are, in general, not solvable because of the existence of an infinite countable set of obstructions. These were first evidenced in the step-2 example in the work of Katok and Robinson [KR]. (Large set of distributional obstructions was likewise found for any step nilflows in [FF]. Also see [CF] for a study of certain cases of cohomological equations above abelian actions).
- (ii) In the case of parabolic affine maps of step 3 and higher, and with Diophantine translation part, if the solution to one of the equations of (3) exists, then it is smooth if the right-hand side is smooth. However, the loss of the number of derivatives is not fixed (in other words, the linearized cohomological equation is stable, but the solutions are not tame).

In a separate work we show that, for the simplest  $C^r$  step-3 map  $(x_1, x_2, x_3) \mapsto (x_1 + \alpha_1, x_2 + x_1, x_3 + x_2)$ , the loss of derivatives is roughly  $r/2$ , even for the nicest Diophantine angles  $\alpha_1$ . This constitutes a notable difference with the step-2 case, for which [KR] showed tameness of the solutions when they did exist.

To address (i), the usual path is to exploit the commutation relation to find approximate solutions to the cohomological equations. This was done in two related problems in the past. First, by Moser [M], who showed that SDC commuting circle rotations are locally rigid under the condition of preserving the rotation number. This was extended to higher dimension in [DF, WX, P]. Second, by Damjanović and Katok who proved in [DK] the local rigidity of higher rank partially hyperbolic affine abelian actions on the torus (i.e., actions, all of whose elements are ergodic automorphisms or affine maps with such linear parts).

In Moser's case the objective is to linearize a commuting pair  $R_{\alpha_i} + f_i$ ,  $i = 1, 2$ . The cohomological equations take the form  $h(x + \alpha_i) - h(x) = f_i(x) - \int f_i$ . They have formal solutions above the generators  $R_{\alpha_i}$ , and the commutation relation allows to upgrade the formal solutions into the approximate tame solutions. In fact, Moser's trick is to define, for each Fourier mode  $n$ , the corresponding coefficient  $h_n$  of the conjugacy map, using either one or the other of the linearized conjugacy equations, according to which  $\|n\alpha_i\|$ , for  $i = 1$  or for  $i = 2$ , is "not too small" as granted by the SDC-condition.



As a result, one gets a candidate conjugacy  $h$  that is *tame*, i.e., of the same order as the nonlinearities  $f$  and  $g$  with a fixed loss of the number of derivatives. Moreover, the commutation relation plus the SDC-condition insure that the constructed  $h$  solves the cohomological equations with a *quadratic* error (with a small abuse of notations, by *quadratic* we will mean that the error is of order of a power  $k > 1$  in the nonlinearities  $f$  and  $g$ , with a fixed loss of the number of derivatives).

The above procedure allows to implement the classical KAM quadratic scheme, with the issue of the constant terms  $\int f_i$  being resolved due to the condition of the preservation of the rotation numbers.

In [DK] the individual equations as in (3) have a tame solution provided a countable set of obstructions vanish, each one being formally computed as weighted sums along the dual orbit of Fourier coefficients of the nonlinearities  $f$  and  $g$ . The commutation relation in this case allows to get quadratic approximations of the nonlinearities by functions whose obstructions vanish. This was labelled "highr-rank trick". Here again, the approximation is quadratic with a finite loss of the number of derivatives.

Our proof of KAM-rigidity for parabolic actions combines two mechanisms of local rigidity: "Moser's trick" and "higher rank trick". The translation part of the action and the resonances (Fourier modes that are invariant along some element of the dual action) are treated using a mechanism, similar to Moser's trick. This is where the Diophantine conditions of Definition 9 on the action play a crucial role. It has to be noted that the use of Moser's trick for the resonances brings some technical challenges that affect the whole proof. Indeed, for a resonant Fourier mode  $m$ , we need to use the element  $F^k G^l$  of the action, where  $(k, l)$  is the resonance pair associated to  $m$ . This forces us to work out the linearization KAM scheme at each step for *a large number of elements* of the action, and not only for the two generators. Of course, we cannot control *all* of the nonlinearities in  $F^k G^l$  for all resonance pairs  $(k, l)$  at each step, because  $k$  and  $l$  can be arbitrarily large. Fortunately, the resonance pairs associated to a resonant mode  $m$  are of the order of  $m$  (see Lemma 4). This means that if, at a given step of the KAM scheme, we truncate the nonlinearities up to order  $N$  before finding an approximative solution of the linearized conjugacy equation, we will only need to control  $F^k G^l$  for  $k$  and  $l$  of order  $N$ . This can easily be included in the induction due to the parabolic nature of  $a$  and  $b$ .

For "non-resonant" Fourier modes, it is a higher rank trick approach similar to [DK] that is invoked. Indeed, for a non-resonant mode  $m$  we can define  $h_m$  via the sum of the Fourier coefficients of the nonlinearity  $f$  along the dual orbit of the step-2 generator of the affine action, taken in the "good direction": either in the future or in the past (in a similar way to what is done in Livschits theory). The fact that the generator is step-2 implies that the Fourier modes, involved in these partial sums, grow either for the past or the future sum, which allows us to define a tame candidate conjugacy  $h$ , as observed in [KR]. Observe that difficulty (ii) mentioned above shows that the mere definition of a candidate tame conjugacy when no element of the action is step-2 is already a challenge for the general higher step case. Other difficulties appear in relation with the applicability of the parabolic higher rank trick that will be explained in the next paragraph.

Once  $h$  is constructed, we see that it is only at special modes  $\bar{m}$  that are lowest (in norm) on their dual orbit along  $\bar{A}$  that the constructed  $h$  does not solve the cohomological



equation above  $a$  (at  $\bar{m}$ , the good direction switches from past to future). The error in solving the equation at  $\bar{m}$  is indeed the full sum along  $\bar{m}$  of the Fourier coefficients of the nonlinearity along  $\bar{A}$ . These sums, having the form  $\Sigma_m^A(f) = \sum_{k \in \mathbb{Z}} f_{\bar{A}^k m} \lambda_m^{(k)}$  (where  $\lambda_m^{(k)}$  are "innocuous" multipliers of modulus one related to the translation part of the action, see §3.1 for the exact definitions), are the obstructions to solving the cohomological equations above  $\bar{A}$ .

The higher rank trick uses commutativity to show that this full sum is equal to a double sum of a quadratic function  $\phi$  measuring the error of the pair  $(f, g)$  in (3) from forming a cocycle above the action  $\langle a, b \rangle$  (see Section 2.1 and Section 4.1.4 for more explanations). It appears to be fruitful to express the obstructions as the following double sums:

$$\Sigma_m^A(f) = \sum_{l \geq 0} \sum_{k \in \mathbb{Z}} \phi_{\bar{A}^k \bar{B}^l m} \lambda_m^{(k)} \mu_m^{(l)} = - \sum_{l \leq -1} \sum_{k \in \mathbb{Z}} \phi_{\bar{A}^k \bar{B}^l m} \lambda_m^{(k)} \mu_m^{(l)}. \quad (4)$$

There is an important difference between the phenomenon that lies behind the control of the double sums in our case, compared to the partially hyperbolic case. In the *partially hyperbolic* higher rank case treated in [DK], the Fourier modes that appear in the double sums in one of the two directions (future or past for  $\bar{B}$ ) are essentially increasing due to the partial hyperbolicity of the action, and this immediately leads to approximate solutions of (3) with quadratic errors with finite loss of the number of derivatives.

In our case, due to the presence of a higher step generator in the action, there may be no growth in either direction along the dual orbits that appear in the double sums. In fact, it always happens for some modes  $m$  that the double orbits appearing in (4) decay in both directions from  $|m|$  to  $|m|^{1/(S-1)}$ , where  $S$  is the step of the action. This is the difficulty (ii) mentioned above.

One of the key ingredients of our argument is the proof of the fact that for a *unlocked* parabolic linear action with at least one step-2 generator, the fall from  $|m|$  to  $|m|^{1/(S-1)}$  is the worst that can happen. Our proof uses the presence of a step-2 element, and its extension to higher step actions is another challenge in the study of the general case.

This means that the error in solving the first equation in (3) with the conjugating transformation  $h$  we constructed is quadratic, but with a loss of a certain *proportion* of the number of the derivatives that are considered (a proportion  $(S-2)/(S-1)$  for step- $S$  maps), even under the nicest Diophantine conditions.

The good news is that this loss of derivatives appears only in the quadratic error and not in the estimate of the conjugating map (for this, the step-2 assumption on one generator is crucial). As a consequence, this important loss of derivatives does not affect the convergence of the KAM scheme, for which it suffices to have a quadratic control of  $C^0$  norms of the error (in fact  $L^2$  would be sufficient).

Once it is shown that  $h$  solves the first equation of (3) up to a quadratic error (in  $C^0$  norm), the commutation relation and the fact that  $A$  is step-2 can be used again to show that  $h$  also solves the second equation with a quadratic error (see Section 4.1.1).

Finally, we point out to the fact that equations (3) can be solved as usual up to a set of  $2d$  constant terms that account for the averages of  $f$  and  $g$ . Unlike in Moser's case of commuting circle diffeomorphisms, these constants are not all related to some dynamical invariants. However, we can use the volume preservation of the perturbed action and the zero average of the nonlinearities to fix the averages of the conjugating diffeomorphisms at each step of the KAM scheme, so that the constant terms become

absorbed in the quadratic error. This third difference with the usual KAM scheme is explained in detail at the end of §2.2.

As remarked before, our arguments remain true if we replace the preservation of the volume  $\lambda$  by that of any common invariant measure for  $F$  and  $G$ . It suffices to replace  $\lambda$  by an arbitrary common invariant measure in all the text. Indeed, we do not use that  $\lambda$  is invariant by  $a$  and  $b$  in the proof of the linearization. Moreover, since  $a^k b^l$  is uniquely ergodic for some  $k$  and  $l$ , the linearisation implies, in fact, that there is a unique invariant measure for the action  $\langle F, G \rangle$ , and that this measure is the pullback of the Haar measure by the conjugacy.

**1.7. Comments on extensions and applications.** There are natural questions raised by our result as to what extent the method developed here is applicable to more general situations. We comment on this below.

◊ **On applications to non-abelian actions.** We note that there are classes of solvable affine actions to which our result in Theorem C can be directly applied. An *abelian-by-cyclic* group  $G$  is a finitely presented torsion free group admitting a short exact sequence  $0 \rightarrow \mathbb{Z}^k \rightarrow G \rightarrow \mathbb{Z} \rightarrow 0$  (see [WX] for detailed discussion on ABC groups). In this context, we call the subgroup  $\mathbb{Z}^k$  the abelian part of  $G$ . Let  $\rho : G \rightarrow \text{Aff}(\mathbb{T}^d)$  be an affine action of  $G$  such that  $\rho(\mathbb{Z}^k)$  is parabolic. Then from the KAM rigidity result for the action  $\rho(\mathbb{Z}^k)$  one may derive KAM rigidity for the  $G$  action. This way of obtaining KAM rigidity for an ABC action from KAM rigidity of its abelian part has been used before in [WX] but in the special case where the abelian part  $\rho(\mathbb{Z}^k)$  is generated by translations. More recently, in [P2], actions on  $\mathbb{T}^3$  of the following particular ABC group:  $\Gamma = \langle U, V, F : UV = VU, FU = U^2VF, FV = UVF \rangle$  have been studied. An example of a  $\Gamma$  action on  $\mathbb{T}^3$  is when  $U = \text{Id} + E_{12}$  and  $V = \text{Id} + E_{13}$  and

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

The abelian action  $\langle U, V \rangle$  is unlocked and thus Diophantine affine actions with such linear part are KAM rigid by Theorem C, which in turn implies KAM rigidity for an affine  $\Gamma$  action with such abelian part.

Similar to our dichotomy result, we expect to use the method we developed in this paper to obtain a classification result for linear ABC actions  $\rho_L : G \rightarrow \text{Aut}(\mathbb{T}^d)$  having a parabolic abelian part. There are roughly 3 main cases:

(i)  $\rho_L$  is locked: for every affine  $G$  action  $\rho$  with linear part  $\rho_L$ ,  $\rho$  has a rank-one factor that is either identity or a genuinely parabolic action.

(ii)  $\rho_L$  is unlocked but *has a locked abelian part*: for every affine  $G$  action  $\rho$  with linear part  $\rho_L$ ,  $\rho(\mathbb{Z}^k)$  has a rank-one factor that is either identity or a genuinely parabolic action, but  $\rho_L$  is unlocked.

(iii)  $\rho_L$  *has an unlocked abelian part*.

It is in the case (iii) where Theorem C applies. In the case (ii), even though Theorem C does not apply directly, we expect our method and even the constructions of solutions from our proofs, to apply.

It is a curious algebraic question to determine which solvable groups acting on the torus by automorphisms can have unlocked abelian part. The group  $\Gamma$  described above allows on  $\mathbb{T}^3$  both a locked action (case (i)) and an action with unlocked abelian part

(case (iii)), but does not allow case (ii) [P2]. We remark that the 3 dimensional discrete Heisenberg group  $H_3$  generated by three elementary matrices (i.e. matrices of the form  $\text{Id} + E_{ij}$ ) on any  $\mathbb{T}^d$  is locked (in particular it has a locked abelian part) and it is not clear if  $H_3$  linear actions by toral automorphisms are always locked.

◊ **On connection to nilflows.** Given a nilpotent Lie group  $N$  of step  $k$ , and a lattice  $\Gamma$  in  $N$ , the quotient  $N/\Gamma$  is a nilmanifold of step  $k$ . Any one-parameter subgroup of  $N$  defines, via left-multiplication on  $N/\Gamma$ , a smooth nilflow. Similarly, a subgroup  $A$  of  $N$  isomorphic to  $\mathbb{R}^k$  defines an  $\mathbb{R}^k$  nilaction on  $N/\Gamma$ . While it was proved by Flaminio and Forni [FF] that nilflows have infinite dimensional cohomology, nilactions can have finite dimensional cohomology as in [CF] or in [D]. In [D] this was used for proving a KAM type of local rigidity result for a class of nilactions with strong Diophantine properties, on 2-step nilmanifolds. There is a close connection between the actions which we consider in this paper and nilactions. Namely, one gets a parabolic affine  $\mathbb{Z}^k$  action if one considers return maps of an  $\mathbb{R}^k$  nilaction to a certain section, and  $\mathbb{R}^k$  nilactions can be viewed as suspensions over such  $\mathbb{Z}^k$  actions on the torus. We hope that some of the ideas developed here to study the KAM-rigidity of parabolic actions on the torus could be useful in the local rigidity study of nilactions, that is a more general problem where, besides the works cited above, there has been yet no progress.

**1.8. Plan of the paper.** The rest of the paper is devoted to the proof of Theorem C. The proof is divided into three parts. In §2 we state the main inductive KAM conjugacy step, Proposition A, and then show how to deduce Theorem C from it. In §3 we give some necessary estimates on sums and double sums along the dual orbits of  $A$  and  $B$  that serve for constructing the approximate solutions to the cohomological equations that appear in the linearized conjugacy equations. In §4 we use the latter estimates to prove Proposition A. Each part will start with a detailed introduction of its content and of the ideas that are involved in the proofs.

## 2. PROOF OF THEOREM C- THE ITERATION PART

The proof is based on a KAM scheme, with three peculiarities which distinguish it from the usual way KAM schemes are applied to proving local rigidity.

The usual KAM iteration goes as follows: we start with an  $\varepsilon$ -perturbation  $\langle F, G \rangle$  of  $\langle a, b \rangle$ . By linearising the conjugacy problem and by solving the linear equation approximately, we produce a conjugacy  $\mathcal{H}_1 = (\text{Id} + \mathbf{h}_1)$  which conjugates  $\langle F, G \rangle$  to an action  $\langle F_1, G_1 \rangle$  which is an  $\varepsilon^k$ -perturbation of  $\langle a, b \rangle$ , where  $k > 1$ . Then we say that  $\langle F_1, G_1 \rangle$  is a *quadratically small* perturbation of  $\langle a, b \rangle$ , with respect to how far  $\langle F, G \rangle$  was from  $\langle a, b \rangle$ . This process is repeated, and at the  $n$ -th step of iteration we build conjugacies  $\mathcal{H}_n = (\text{Id} + \mathbf{h}_1) \circ \dots \circ (\text{Id} + \mathbf{h}_n)$  that satisfy

$$\begin{cases} \mathcal{H}_n^{-1} \circ F \circ \mathcal{H}_n &= a + \mathbf{f}_{n+1}, \\ \mathcal{H}_n^{-1} \circ G \circ \mathcal{H}_n^{-1} &= b + \mathbf{g}_{n+1}, \end{cases} \quad (5)$$

where  $\mathbf{f}_{n+1}$  and  $\mathbf{g}_{n+1}$  are of order  $\varepsilon_n^k = \varepsilon_{n+1}$ , while  $\mathbf{h}_n$  is of order  $\varepsilon_n$ .

Truncation (or more generally, applying smoothing operators) is typically used only to remedy a fixed loss of regularity at each step of iteration while solving the linearized problem. In our case here, due to the (possible) presence of infinitely many resonances,

without truncation we might not have any quadratic estimates for the error. This is the first peculiarity of the proof, the corresponding details are contained in §4.1.1.

The other one is that at every step of the KAM procedure we solve the linearised equations approximately *only* up to a constant term. This constant term can be large, it makes the error at the  $n$ -th step of order  $\varepsilon_n$  instead of  $\varepsilon_{n+1}$  (as we would like), so *a priori* there need not be any convergence of the sequence  $\mathcal{H}_n$ . This is where we use the volume preservation assumption. Namely, the volume preservation assumption allows us to adjust the average of  $\mathbf{h}_n$  at step  $n$ , so that the total new error  $(\mathbf{f}_{n+1}, \mathbf{g}_{n+1})$  becomes of order  $\varepsilon_{n+1}$ . The same approach was used by Herman for Diophantine torus translations [H]. Application of this approach in the context of group actions meets certain difficulties. This is explained in §2.2.

The third feature of the proof, which has not appeared much in similar problems, is that, even though the estimates for  $\mathbf{h}_n$  at each step are tame, the estimates for the error at each step are not tame. Namely, the loss of the number of derivatives is not a fixed constant as usually, but a proportion (that goes to 1 when the second generator's step goes to infinity) of the number of derivatives. However, this does not affect the convergence of the scheme. Similar observation was used recently in [ZW].

### 2.1. Linearisation of the problem and the main iterative step: Proposition A.

Given small perturbations  $a + \mathbf{f}$  and  $b + \mathbf{g}$  of the two action generators  $a$  and  $b$ , and the commutativity condition among them:

$$(a + \mathbf{f}) \circ (b + \mathbf{g}) = (b + \mathbf{g}) \circ (a + \mathbf{f}), \quad (6)$$

we wish to solve for  $H = id + \mathbf{h}$  the conjugacy problem

$$H \circ (a + \mathbf{f}) = a \circ H, \quad H \circ (b + \mathbf{g}) = b \circ H. \quad (7)$$

The commutativity condition (6) can be rewritten as

$$\mathbf{f}(b + \mathbf{g}) - B\mathbf{f} - (\mathbf{g}(a + \mathbf{f}) - A\mathbf{g}) = 0,$$

which permits to see condition (6) as a sum of a linear operator applied to  $\mathbf{f}, \mathbf{g}$  plus a non-linear part which is quadratic in  $\mathbf{f}, \mathbf{g}$ :

$$[\mathbf{f} \circ b - B\mathbf{f} - (\mathbf{g} \circ a - A\mathbf{g})] + [\mathbf{f}(b + \mathbf{g}) - \mathbf{f} \circ b - (\mathbf{g}(a + \mathbf{f}) - \mathbf{g} \circ a)] = 0. \quad (8)$$

Now we introduce some notations. For any given  $\mathbf{h}$ , let

$$\begin{aligned} D_{1,0}\mathbf{h} &:= \mathbf{h} \circ a - A\mathbf{h}, \\ D_{0,1}\mathbf{h} &:= \mathbf{h} \circ b - B\mathbf{h}. \end{aligned}$$

With this notations, equation (8) gets the form

$$D_{0,1}\mathbf{f} - D_{1,0}\mathbf{g} = -\mathbf{f}(b + \mathbf{g}) + \mathbf{f} \circ b + \mathbf{g}(a + \mathbf{f}) - \mathbf{g} \circ a.$$

Similarly, since we are looking for the conjugating map  $H$  in a neighborhood of the identity, i.e., in the form  $H = Id + \mathbf{h}$  with  $\mathbf{h}$  small, our conjugacy problem (7) is linearised as

$$\begin{aligned} D_{1,0}\mathbf{h} &= \mathbf{f} + [\mathbf{h}(a + \mathbf{f}) - \mathbf{h} \circ a], \\ D_{0,1}\mathbf{h} &= \mathbf{g} + [\mathbf{h}(b + \mathbf{g}) - \mathbf{h} \circ b]. \end{aligned}$$

At the  $n$ -th step of the iteration process, for given  $\mathbf{f}_n, \mathbf{g}_n$  we show that we can find  $\tilde{\mathbf{f}}_n, \tilde{\mathbf{g}}_n, \mathbf{h}_n$  and vectors  $\mathbf{V}_n$  and  $\mathbf{W}_n$  such that

$$\begin{aligned} D_{1,0}\mathbf{h}_n &= \mathbf{f}_n + \tilde{\mathbf{f}}_n + \mathbf{V}_n, \\ D_{0,1}\mathbf{h}_n &= \mathbf{g}_n + \tilde{\mathbf{g}}_n + \mathbf{W}_n, \end{aligned}$$

where  $\mathbf{V}_n$  and  $\mathbf{W}_n$  are of the same order as  $\mathbf{f}_n, \mathbf{g}_n$ , and the new functions  $\tilde{\mathbf{f}}_n, \tilde{\mathbf{g}}_n$  are quadratic. In later sections, when we set up the iteration process, we will see that the volume preservation assumption will force the constant terms  $\mathbf{V}_n$  and  $\mathbf{W}_n$  to be of quadratic order as well.

Let us formulate the main iterative step as a proposition. In later sections Proposition A will be used to perform iterations, show their convergence and prove the main result Theorem C.

In what follows, we say that an affine action  $\langle a, b \rangle$  is *unlocked* if its linear part  $\langle A, B \rangle$  is unlocked. If  $\langle a, b \rangle$  is a Diophantine affine action, then its linear part is automatically unlocked, but we stress this in the statements since the property of  $\langle A, B \rangle$  being unlocked will play a crucial role in the proofs.

**Proposition A.** *Let  $\langle a, b \rangle$  be an unlocked  $(\gamma, \tau)$ -Diophantine parabolic affine  $\mathbb{Z}^2$  action, where  $a$  is step-2. Let  $F = a + \mathbf{f}$  and  $G = b + \mathbf{g}$  be  $C^\infty$  commuting diffeomorphisms generating a perturbation  $\langle F, G \rangle$  of  $\langle a, b \rangle$ . For  $r \geq 0$ , let  $\Delta_r = \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\}$ .*

*There exist constants  $C, C_r, C_{r'}$  and  $D = D(a, b, \gamma, \tau, d)$  such that for any  $N \in \mathbb{N}$  there exist vector fields  $\tilde{\mathbf{f}}_N, \tilde{\mathbf{g}}_N, \mathbf{h}_N$ , and vectors  $\mathbf{V}$  and  $\mathbf{W}$  such that*

$$\begin{aligned} D_{1,0}\mathbf{h}_N + \tilde{\mathbf{f}}_N &= \mathbf{f} + \mathbf{V}, \\ D_{0,1}\mathbf{h}_N + \tilde{\mathbf{g}}_N &= \mathbf{g} + \mathbf{W}, \end{aligned}$$

and the following estimates hold whenever  $0 \leq r, D < r'$ :

$$\begin{aligned} \|\mathbf{h}_N\|_r &\leq C_r N^D \Delta_r, \\ \|\tilde{\mathbf{f}}_N\|_0, \|\tilde{\mathbf{g}}_N\|_0 &\leq C N^D \Delta_0 \Delta_1 + C_{r'} N^{-r'+D} \Delta_{r'}, \\ \|\tilde{\mathbf{f}}_N\|_r, \|\tilde{\mathbf{g}}_N\|_r &\leq C_r N^D \Delta_r, \\ |\mathbf{V}|, |\mathbf{W}| &\leq C \Delta_0. \end{aligned} \tag{9}$$

The proof of Proposition A is postponed to §3.

**2.2. Iteration set-up.** In this section we set up the iteration which we use to prove Theorem C. The iterative step consists of three sub-steps: linearization, application of Proposition A and adjusting the average of the conjugating diffeomorphism by using the volume preservation of the perturbation.

**Proposition 8.** *Let  $\langle a, b \rangle$  be an unlocked  $(\gamma, \tau)$ -Diophantine parabolic affine  $\mathbb{Z}^2$  action, where  $a$  is step-2. There exists a constant  $D > 0$  only depending on the action  $\langle a, b \rangle$ , for which the following holds.*

*Let  $\langle a + \mathbf{f}, b + \mathbf{g} \rangle$  be a  $C^\infty$  volume preserving perturbation such that*

$$\text{ave}(\mathbf{f}) = \text{ave}(\mathbf{g}) = 0.$$

Assume that we have constructed a conjugation up to the  $n$ -th step,  $\mathcal{H}_{n-1} = (\text{Id} + \mathbf{h}_{n-1}) \circ \dots \circ (\text{Id} + \mathbf{h}_1)$ , such that  $\text{ave}((\mathcal{H}_{n-1}) - \text{Id}) = 0$  and

$$\mathcal{H}_{n-1} \circ \langle a + \mathbf{f}, b + \mathbf{g} \rangle \circ \mathcal{H}_{n-1}^{-1} = \langle a + \mathbf{f}_n, b + \mathbf{g}_n \rangle.$$

Denote  $\Delta_{r,n} := \max\{\|\mathbf{f}_n\|_r, \|\mathbf{g}_n\|_r\}$ .

Then for any  $N \in \mathbb{N}$  there exists  $\mathbf{h}_n$  (which depends on  $N$ ) such that for  $0 \leq r, D \leq r'$  and certain constants  $C, C_r, C_{r'}$ , if  $N^D \Delta_{1,n} < 1$ , then we have:

- (i)  $\|\mathbf{h}_n\|_r \leq C_r N^D \Delta_{r,n}$ ;
- (ii) For

$$\mathbf{f}_{n+1} := (\text{Id} + \mathbf{h}_n) \circ (a + \mathbf{f}_n) \circ (\text{Id} + \mathbf{h}_n)^{-1} - a,$$

$$\mathbf{g}_{n+1} := (\text{Id} + \mathbf{h}_n) \circ (b + \mathbf{g}_n) \circ (\text{Id} + \mathbf{h}_n)^{-1} - b,$$

$$\Delta_{r,n+1} := \max\{\|\mathbf{f}_{n+1}\|_r, \|\mathbf{g}_{n+1}\|_r\},$$

the following estimates hold:

$$\begin{aligned} \Delta_{0,n+1} &\leq C_r N^D \Delta_{0,n} \Delta_{1,n} + C_{r'} N^{-r'+D} \Delta_{r',n}, \\ \Delta_{r,n+1} &\leq C_r N^D \Delta_{r,n}; \end{aligned} \tag{10}$$

- (iii) For  $\mathcal{H}_n := (\text{Id} + \mathbf{h}_n) \circ \mathcal{H}_{n-1}$  we have:  $\text{ave}(\mathcal{H}_n - \text{Id}) = 0$ .

*Proof.* The non-linear problem is to find  $\mathbf{h}_n$  such that

$$(\text{Id} + \mathbf{h}_n) \circ \langle a + \mathbf{f}_n, b + \mathbf{g}_n \rangle = \langle a + \mathbf{f}_{n+1}, b + \mathbf{g}_{n+1} \rangle \circ (\text{Id} + \mathbf{h}_n) \tag{11}$$

with  $\mathbf{h}_n$  and  $\mathbf{f}_{n+1}, \mathbf{g}_{n+1}$  satisfying the estimates of the proposition.

Here is a brief outline of the proof. After linearizing the above non-linear problem, we will first apply Proposition A to determine  $\mathbf{h}_n$  and vectors  $\mathbf{V}_n, \mathbf{W}_n$  such that  $|\mathbf{V}_n| + |\mathbf{W}_n| \leq C \Delta_{0,n}$  and

$$(\text{Id} + \mathbf{h}_n) \circ \langle a + \mathbf{f}_n, b + \mathbf{g}_n \rangle = \langle a + \mathbf{f}_{n+1} + \mathbf{V}_n, b + \mathbf{g}_{n+1} + \mathbf{W}_n \rangle \circ (\text{Id} + \mathbf{h}_n), \tag{12}$$

where  $\mathbf{h}_n$  and  $\mathbf{f}_{n+1}, \mathbf{g}_{n+1}$  satisfy the estimates in (i) and (ii). We observe that if we change  $\mathbf{h}_n$  by adding to it a translation vector of order  $\Delta_{0,n}$ , then equation (12) will still hold with some new  $\mathbf{f}_{n+1}, \mathbf{g}_{n+1}, \mathbf{V}_n, \mathbf{W}_n$  that satisfy the same estimates. By adequately choosing the translation vector, based on the volume preservation condition and the zero average condition on the initial perturbation (and the inductive condition  $\text{ave}((\mathcal{H}_{n-1}) - \text{Id}) = 0$ ), we will be able to absorb the constants  $\mathbf{V}_n$  and  $\mathbf{W}_n$  into  $\mathbf{f}_{n+1}$  and  $\mathbf{g}_{n+1}$ .

**Linearization.** We begin by linearizing the non-linear conjugation problem. Equation (12) is rewritten in a way that expresses the error in the new perturbation  $(\mathbf{f}_{n+1}, \mathbf{g}_{n+1})$  in terms of the linearization of the non-linear conjugation problem above and additional errors:

$$\begin{aligned} \mathbf{f}_{n+1}(\text{Id} + \mathbf{h}_n) &= (\mathbf{h}_n \circ a - A\mathbf{h}_n) + \mathbf{f}_n - \mathbf{V}_n + (\mathbf{h}_n \circ (a + \mathbf{f}_n) - \mathbf{h}_n \circ a), \\ \mathbf{g}_{n+1}(\text{Id} + \mathbf{h}_n) &= (\mathbf{h}_n \circ b - B\mathbf{h}_n) + \mathbf{g}_n - \mathbf{W}_n + (\mathbf{h}_n \circ (b + \mathbf{g}_n) - \mathbf{h}_n \circ b). \end{aligned} \tag{13}$$

To estimate the left-hand side in the equations above, we need to estimate the following two terms:

$$\begin{aligned} E_1 &:= \mathbf{h}_n \circ a - A\mathbf{h}_n + \mathbf{f}_n - \mathbf{V}_n, \quad \mathbf{h}_n \circ b - B\mathbf{h}_n + \mathbf{g}_n - \mathbf{W}_n, \\ E_2 &:= \mathbf{h}_n \circ (a + \mathbf{f}_n) - \mathbf{h}_n \circ a, \quad \mathbf{h}_n \circ (b + \mathbf{g}_n) - \mathbf{h}_n \circ b. \end{aligned} \tag{14}$$

Bellow we estimate both terms,  $E_1$  and  $E_2$ , in  $C^0$  norm for the transformation  $\mathbf{h}_n$  provided by in Proposition A. This will imply the estimate for the  $C^0$  norm of  $\mathbf{f}_{n+1}$  and  $\mathbf{g}_{n+1}$ .

### Applying Proposition A.

Apply now Proposition A to  $\mathbf{f}_n, \mathbf{g}_n$ . Fix  $N \in \mathbb{N}$ . For the fixed  $N$ , from Proposition A we obtain  $\mathbf{h}_n := (\mathbf{h}_n)_N, (\mathbf{f}_n)_N, (\mathbf{g}_n)_N$  and vectors  $\mathbf{V}_n$  and  $\mathbf{W}_n$ . The first estimate in Proposition A gives directly:

$$\|\mathbf{h}_n\|_r \leq C_r N^D \Delta_{r,n}. \quad (15)$$

Observe that the assumption that  $N^D \Delta_{1,n}$  is bounded by a constant implies that  $\|\mathbf{h}_n\|_1$  is bounded by a constant. It is a common fact (see for example [L, Lemma AII.26]) that the inverse map  $(\text{Id} + \mathbf{h}_n)^{-1} = \text{Id} + \mathbf{h}'_n$  is such that  $\mathbf{h}'_n$  also satisfies the estimate:

$$\|\mathbf{h}'_n\|_r \leq C_r \|\mathbf{h}_n\|_r. \quad (16)$$

The error  $E_1$  is precisely  $((\mathbf{f}_n)_N, (\mathbf{g}_n)_N)$ , so from the second estimate in Proposition A we get for any  $r' > 0$ :

$$\|E_1\|_0 \leq C N^D \Delta_{0,n} \Delta_{1,n} + C_{r'} N^{-r'+D} \Delta_{r',n}. \quad (17)$$

The estimate for  $E_2$  follows by using the standard estimates (see for example Appendix in [DF]) and estimate (15):

$$\|E_2\|_0 \leq C \|\mathbf{h}_n\|_1 \Delta_{0,n} \leq C N^D \Delta_{1,n} \Delta_{0,n}. \quad (18)$$

Putting the two errors together, we have that the new error satisfies:

$$\Delta_{0,n+1} \leq C N^D \Delta_{0,n} \Delta_{1,n} + C_{r'} N^{-r'+D} \Delta_{r',n}. \quad (19)$$

The estimate for  $C^r$  norms of  $\mathbf{f}_{n+1}$  and  $\mathbf{g}_{n+1}$  for any  $r$  (the second estimate in (10)) follows from the definition (13) of these maps. We show how the estimate follows for  $\mathbf{f}_{n+1}$ . For  $\mathbf{g}_{n+1}$  the proof is the same.

From (13) we can write:

$$\mathbf{f}_{n+1} = \tilde{\mathbf{f}}_n (\text{Id} + \mathbf{h}'_n) + (\mathbf{h}_n \circ (a + \mathbf{f}_n) - \mathbf{h}_n \circ a) \circ (\text{Id} + \mathbf{h}'_n),$$

where  $\text{Id} + \mathbf{h}'_n = (\text{Id} + \mathbf{h}_n)^{-1}$ , and  $\mathbf{h}'_n$  satisfies estimate (16). Then by applying standard estimate for the composition of maps (see for example [Hör, Theorem A.8]) and the bound for  $\|\tilde{\mathbf{f}}_n\|_r$  which we have from Proposition A, we get:

$$\begin{aligned} \|\mathbf{f}_{n+1}\|_r &\leq \|\tilde{\mathbf{f}}_n (\text{Id} + \mathbf{h}'_n)\|_r + \|\mathbf{h}_n \circ (a + \mathbf{f}_n) - \mathbf{h}_n \circ a\|_r \circ (\text{Id} + \mathbf{h}'_n) \\ &\leq C_r (\|\tilde{\mathbf{f}}_n\|_r + \|\mathbf{h}'_n\|_r) \leq C_r N^D \Delta_{r,n}. \end{aligned}$$

**Adjusting the average of the conjugating diffeomorphism.** Now we will adjust the average of  $\mathbf{h}_n$  in such a way that the constant terms  $\mathbf{V}_n$  and  $\mathbf{W}_n$  in (13) are forced to be as small as  $\|\mathbf{f}_{n+1}\|_0$  and  $\|\mathbf{g}_{n+1}\|_0$ , respectively. The crucial role here is played by the assumptions on the volume preservation and zero averages of the initial errors. The adjustment of the average of  $\mathbf{h}_n$  will not depend on the action elements, as will be seen in Lemma 10. We will check that it works on one action generator, the other generator can be treated in the same way.



**Lemma 9.** Suppose that  $\mathbf{h}$  is such that

$$a + \mathbf{f}_{n+1} + \mathbf{V}_n = (\text{Id} + \mathbf{h}) \circ (a + \mathbf{f}_n) \circ (\text{Id} + \mathbf{h})^{-1}, \quad (20)$$

where  $|\mathbf{V}_n| = O(\Delta_{0,n})$  and  $\mathbf{f}_{n+1}$  satisfies estimate (10).

For any vector  $C$  such that  $|C| = O(\Delta_{0,n})$ , the function  $\hat{\mathbf{h}} = \mathbf{h} + C$  satisfies

$$a + \hat{\mathbf{f}}_{n+1} + \hat{\mathbf{V}}_n = (\text{Id} + \hat{\mathbf{h}}) \circ (a + \mathbf{f}_n) \circ (\text{Id} + \hat{\mathbf{h}})^{-1}, \quad (21)$$

where  $|\hat{\mathbf{V}}_n| = O(\Delta_{0,n})$  and  $\hat{\mathbf{f}}_{n+1}$  satisfies (10).

*Proof.* Using (20), we can write:

$$\begin{aligned} (\text{Id} + \mathbf{h} + C) \circ (a + \mathbf{f}_n) &= (\text{Id} + \mathbf{h}) \circ (a + \mathbf{f}_n) + C = (a + \mathbf{f}_{n+1} + \mathbf{V}_n + C) \circ (\text{Id} + \mathbf{h}) \\ &= (a + \mathbf{f}_{n+1} + \mathbf{V}_n + C) \circ (\text{Id} - C) \circ (\text{Id} + \mathbf{h} + C) \\ &= (a + \mathbf{f}_{n+1} \circ (\text{Id} - C) + (\mathbf{V}_n - (A - \text{Id})C)) \circ (\text{Id} + \mathbf{h} + C) \\ &= (a + \hat{\mathbf{f}}_{n+1} + \hat{\mathbf{V}}_n) \circ (\text{Id} + \mathbf{h} + C), \end{aligned}$$

where  $\hat{\mathbf{f}}_{n+1} := \mathbf{f}_{n+1} \circ (\text{Id} - C)$  and  $\hat{\mathbf{V}}_n := \mathbf{V}_n - (A - \text{Id})C$ . Estimate (10) holds then for  $\hat{\mathbf{f}}_{n+1}$  since it holds for  $\mathbf{f}_{n+1}$  and  $|C| = O(\Delta_{0,n})$ . Obviously,  $|\hat{\mathbf{V}}_n|$  is of the same order of magnitude as  $|C| = O(\Delta_{0,n})$ .  $\square$

In the previous part of the proof we constructed  $\mathbf{h}_n$  such that

$$a + \mathbf{f}_{n+1} - \mathbf{V}_n = (\text{Id} + \mathbf{h}_n)^{-1} \circ (a + \mathbf{f}_n) \circ (\text{Id} + \mathbf{h}_n).$$

If  $\mathbf{h}_n$  satisfies equation (20), we can apply Lemma 9 to adjust the average of  $\mathbf{h}_n$ . The following Lemma explains how the constant vector  $C$  is chosen at the  $n$ -th step of the iteration. Let  $H_{n-1} = \mathcal{H}_{n-1} - \text{Id}$ , and recall that, by assumption,  $\text{ave}(H_{n-1}) = 0$ .

**Lemma 10.** Let  $C = -\int_{\mathbb{T}^d} \mathbf{h}_n \circ \mathcal{H}_{n-1}$ , and let  $\hat{\mathbf{h}}_n = \mathbf{h}_n + C$ . Let  $\hat{\mathcal{H}}_n = (\text{id} + \hat{\mathbf{h}}_n) \circ \mathcal{H}_{n-1} = \text{Id} + \hat{H}_n$ . Then  $\text{ave}(\hat{H}_n) = 0$ .

*Proof.*  $\hat{\mathcal{H}}_n = (\text{Id} + \hat{\mathbf{h}}_n) \circ \mathcal{H}_{n-1}$  implies

$$\hat{H}_n = H_{n-1} + \hat{\mathbf{h}}_n \circ \mathcal{H}_{n-1}.$$

By the inductive assumption,  $\text{ave}(H_{n-1}) = 0$ . Then, by taking averages of both sides of the equation above, we get that  $\text{ave}(\hat{H}_n) = 0$ .  $\square$

After choosing  $C$  as in Lemma 10, by applying Lemma 9, we get the equation:

$$a + \hat{\mathbf{f}}_{n+1} + \hat{\mathbf{V}}_n = (\text{Id} + \hat{\mathbf{h}}_n) \circ (a + \mathbf{f}_n) \circ (\text{Id} + \hat{\mathbf{h}}_n)^{-1},$$

which implies

$$a + \hat{\mathbf{f}}_{n+1} + \hat{\mathbf{V}}_n = (\text{Id} + \hat{H}_n) \circ (a + \mathbf{f}) \circ (\text{Id} + \hat{H}_n)^{-1}.$$

From this, by composing on the right with  $\text{Id} + \hat{H}_n$  we get

$$A\hat{H}_n + \hat{\mathbf{f}}_{n+1}(\text{Id} + \hat{H}_n) + \hat{\mathbf{V}}_n = \mathbf{f} + \hat{H}_n \circ (a + \mathbf{f}).$$

By taking averages with respect to the volume of both sides of the equation above and using the assumptions that  $\mathbf{f}$  has zero average and that  $a + f$  is volume preserving, it follows that  $\hat{\mathbf{V}}_n = -\hat{\mathbf{f}}_{n+1}(\text{Id} + \hat{H}_n)$ . This implies that  $\hat{\mathbf{V}}_n = O(\|\hat{\mathbf{f}}_{n+1}\|_0)$ , which means that the constant  $\hat{\mathbf{V}}_n$  can be absorbed by  $\hat{\mathbf{f}}_{n+1}$ .

From Lemma 9 we have that estimates (10) hold for  $\hat{\mathbf{f}}_{n+1}$ . Finally, we proclaim the new  $\mathbf{f}_{n+1}$  to be  $\hat{\mathbf{f}}_{n+1}$ .  $\square$

**2.3. Convergence of the iterative scheme.** Once we have the result of Proposition 8, the set-up of the KAM scheme and its convergence is essentially the same as in usual applications of KAM method (see for example Section 5.4 in [DK]).

Assume that  $\langle a, b \rangle$  is a  $(\gamma, \tau)$ -Diophantine action. Let  $\langle a + \mathbf{f}, b + \mathbf{g} \rangle$  be a small smooth perturbation of  $\langle a, b \rangle$ . Since we are proving KAM rigidity, we also assume that  $\langle a + \mathbf{f}, b + \mathbf{g} \rangle$  is volume preserving and that  $\mathbf{f}$  and  $\mathbf{g}$  have zero average.

Given the initial perturbation above, we let:

$$\mathbf{f}_1 = \mathbf{f}; \quad \mathbf{g}_1 = \mathbf{g}.$$

Recall that we use the notation  $\Delta_{r,1} := \max\{\|\mathbf{f}_1\|_r, \|\mathbf{g}_1\|_r\}$ .

Let  $D$  be the constant from Proposition 8 which depends only on  $\langle a, b \rangle$ .

Fix  $k = \frac{4}{3}$ , and let  $l = 8D + 16$ .

At the first step we assume that:

$$\Delta_{0,1} < \varepsilon, \quad \Delta_{l,1} < \varepsilon^{-1}$$

for a small  $\varepsilon > 0$ . We will show that  $\varepsilon$  can be chosen so small that the iterative process converges.

We describe now the iterative process. By Proposition 8, there exists  $\mathbf{h}_1$  with the estimates claimed in the proposition. Then the transformation  $\text{Id} + \mathbf{h}_1$  conjugates  $\langle a + \mathbf{f}_1, b + \mathbf{g}_1 \rangle$  to a new perturbation, which we call  $\langle a + \mathbf{f}_2, b + \mathbf{g}_2 \rangle$ . This procedure is iterated.

At this point we still have the freedom to choose the truncation at level  $N$  when applying Proposition 8 at the  $n$ -th step of the iteration. If at step  $n$  we choose the truncation to be

$$N_n = \varepsilon_n^{-\frac{1}{3(D+2)}},$$

where  $\varepsilon_n = \varepsilon^{(k^n)}$ , then for sufficiently small  $\varepsilon$  we can show inductively that the following estimates hold for all  $n$ :

$$\begin{aligned} \Delta_{0,n} &< \varepsilon_n = \varepsilon^{(k^n)}, \\ \Delta_{l,n} &< \varepsilon_n^{-1}, \end{aligned} \tag{22}$$

and

$$\|\mathbf{h}_n\|_1 < \varepsilon_n^{\frac{1}{2}}. \tag{23}$$

Here we use the letter  $C$  to denote any constant which depends only on the fixed  $l, D$  and the unperturbed action  $\langle a, b \rangle$ .

Suppose that we are at the  $n$ -th step of iteration and that estimates (22) hold for  $n$ . First, we check that the condition  $N^D \Delta_{1,n} < 1$  holds by using the standard interpolation inequality

$$\Delta_{1,n} \leq C \Delta_{0,n}^{1-\frac{1}{l}} \Delta_{l,n}^{\frac{1}{l}}, \tag{24}$$

and assumptions (22):

$$N_n^D \Delta_{1,n} \leq C \varepsilon_n^{\frac{-D}{3(D+2)}} \varepsilon_n^{1-\frac{1}{l}} \varepsilon_n^{-\frac{1}{l}} = C \varepsilon^{\frac{-D}{3(D+2)} + 1 - \frac{2}{l}}.$$

Since for the chosen value of  $l$  the term  $\frac{-D}{3(D+2)} + 1 - \frac{2}{l}$  is positive, by choosing initial  $\varepsilon$  sufficiently small, we get that the above expression is smaller than 1.

Then the maps  $\mathbf{h}_n$ ,  $\mathbf{f}_{n+1}$  and  $\mathbf{g}_{n+1}$  are constructed by applying Proposition 8. The same proposition, combined with (22) and the interpolation inequalities (24), together with our choice of  $l$ , imply that (23) holds for a sufficiently small  $\varepsilon$ :

$$\|h_n\|_1 \leq CN_n^D \Delta_{1,n} \leq CN_n^D \Delta_{0,n}^{1-\frac{1}{l}} \Delta_{l,n}^{\frac{1}{l}} \leq C\varepsilon_n^{-\frac{D}{3(D+2)}+1-\frac{2}{l}} \leq \varepsilon_n^{\frac{1}{2}}.$$

Now we check that (22) holds for  $n$  replaced by  $n+1$ .

First we compute the bounds for the  $l$  norms by using the estimates of Proposition 8:

$$\Delta_{l,n+1} \leq CN_n^D \Delta_{l,n} \leq C\varepsilon_n^{\frac{-D}{3(D+2)}} (1 + \varepsilon_n^{-1}) \leq 2C\varepsilon_n^{\frac{-D}{3(D+2)}-1} < \varepsilon_n^{-\frac{1}{3}-1} = \varepsilon_n^{-\frac{4}{3}} = \varepsilon_{n+1}^{-1}.$$

Finally, we estimate the 0-norms (by using again estimates in Proposition 8 and the interpolation inequality):

$$\begin{aligned} \Delta_{0,n+1} &\leq CN_n^D \Delta_{0,n}^{1-\frac{1}{l}} \Delta_{l,n}^{\frac{1}{l}} \Delta_{0,n} + CN_n^{-l+D} \Delta_{l,n} \\ &\leq C(\varepsilon_n^{-\frac{D}{3(D+2)}+2-\frac{2}{l}} + \varepsilon_n^{\frac{l-D}{3(D+2)}-1}) \leq \varepsilon_n^{\frac{4}{3}} = \varepsilon_{n+1}, \end{aligned}$$

since, given our choice of  $l$ , both expressions  $-\frac{D}{3(D+2)}+2-\frac{2}{l}$  and  $\frac{l-D}{3(D+2)}-1$  are strictly larger than  $\frac{4}{3}$ .

Therefore the estimates in (22) hold for all  $n$ . This implies the convergence of  $\mathcal{H}_n$  in the  $C^1$  norm to some  $\mathcal{H}_\infty$ , which conjugates the initial perturbation to  $\langle a, b \rangle$ . The fact that the conjugation  $\mathcal{H}_\infty$  is  $C^m$  for every  $m > 0$  (i.e., that the process converges in any norm) is proved in a standard way by using interpolation estimates (see for example the end of Section 5.4 in [DK]).

**2.4. Volume preservation of the conjugacy.** Now we have that  $\mathcal{H}_\infty$  conjugates the perturbation  $\langle F, G \rangle$  to  $\langle a, b \rangle$ . Since  $\langle F, G \rangle$  is assumed to be volume preserving, the conjugation relation implies that the pushforward of the volume by  $\mathcal{H}_\infty$  is invariant under  $\langle a, b \rangle$ . Since  $a^k b^l$  is uniquely ergodic for some  $k$  and  $l$ , the map  $\mathcal{H}_\infty$  is volume preserving.

The proof of Theorem C is now completed *modulo* the proof of Proposition A. The rest of the paper is dedicated to the proof of Proposition A.  $\square$

### 3. ESTIMATES OF SUMS AND DOUBLE SUMS ALONG THE DUAL ORBITS

In this subsection we give the necessary estimates on sums and double sums along the dual orbits of  $A$  and  $B$  that will be crucial in solving the cohomological equations and proving Proposition A. The main results of this section are Propositions B and C.

In Proposition B we deal with partial sums along the step-2 dual orbits. These sums will be used in the proof of Proposition A for estimating the norms of the conjugacies. We also give a first estimation of full sums along the step-2 dual orbits that will be the key for estimating the error in solving the cohomological equations at the resonant Fourier modes.

As explained in Section 4.1.4, a full sum along the step-2 dual orbit can be reinterpreted, *via* the higher rank trick, as a double sum of a quadratically small function  $\phi$  measuring the error of the pair  $(f, g)$  in (3) from forming a cocycle above the action  $\langle a, b \rangle$ . In Proposition C, we deal with these double sums. The estimates we obtain in this proposition will serve for estimating the error in solving the cohomological equations at non-resonant Fourier modes.

**3.1. Notations.** In this subsection we summarise the notations used in the rest of the paper.

- Assume that  $\langle A, B \rangle$  is unlocked commuting linear parabolic action. We consider two commuting affine maps  $a(x) = Ax + \alpha$  and  $b(x) = Bx + \alpha$  on  $\mathbb{T}^d$ , where  $a$  is step-2 and  $b$  is step- $S$  (see Definition 1). Elements of the step- $S$  action  $\langle a, b \rangle : \mathbb{Z}^2 \rightarrow \text{Diff}_\lambda^\infty(\mathbb{T}^d)$  are denoted by  $a^k b^l$ ,  $(k, l) \in \mathbb{Z}^2$ .

- Let

$$A = \text{Id} + \tilde{A}; \quad B = \text{Id} + \tilde{B}.$$

In these notations,  $a$  being step-2 and  $b$  being step- $S$  implies:  $\tilde{A}^2 = \tilde{B}^S = 0$ .

- Let  $\bar{A} = (A^{tr})^{-1}$ ,  $\bar{B} = (B^{tr})^{-1}$ . The linear action of  $\langle \bar{A}, \bar{B} \rangle$  on  $\mathbb{Z}^d$  is called the dual action of  $\langle A, B \rangle$ . Let

$$\bar{A} = \text{Id} + \hat{A}; \quad \bar{B} = \text{Id} + \hat{B}.$$

Clearly,  $\bar{A}$  and  $\bar{B}$  are also step-2 and step- $S$ , respectively, which implies  $\hat{A}^2 = \hat{B}^S = 0$ .

- For  $S$  being the step of the action, let

$$\eta = 0.99 \frac{1}{S};$$

- To each  $m \in \mathbb{Z}^d$  we associate  $s = s(m)$ , called the step of  $m$ , such that

$$\hat{B}^s m = 0, \quad \hat{B}^{s-1} m \neq 0.$$

Denote

$$\delta = \delta(m) = 0.99 \frac{1}{s}. \tag{25}$$

Clearly, we have  $s(m) \leq S$  for any  $m \in \mathbb{Z}^d$ , and hence,  $\delta(m) \geq \eta$ .

- For each  $(k, l) \in \mathbb{Z}^2$ , let  $\alpha_{k,l}$  stand for the translation part of  $a^k b^l$ :

$$\alpha_{k,l} := a^k b^l - A^k B^l.$$

- Given a continuous function  $h : \mathbb{T}^d \rightarrow \mathbb{R}$ , denote its Fourier coefficients by  $h_m$ :

$$h(x) = \sum_{m \in \mathbb{Z}^d} h_m e(m, x), \quad e(m, x) := e^{2\pi i(m, x)}.$$

In these notations, for  $a(x) = Ax + \alpha$  we have:

$$h \circ a = \sum_{m \in \mathbb{Z}^d} h_{\bar{A}m} e(\bar{A}m, \alpha) e(m, x), \quad (h \circ a)_m = h_{\bar{A}m} e(\bar{A}m, \alpha).$$

- For  $(k, l) \in \mathbb{Z}^2$  denote by  $\partial_{k,l}$  the coboundary operator: for  $h \in C^\infty(\mathbb{T}^d)$  let

$$\partial_{k,l}(h) := h(a^k b^l) - h.$$

In particular, the expression for the  $m$ -th Fourier coefficient of a coboundary is

$$(\partial_{k,l}(h))_m = h_{\bar{A}^k \bar{B}^l m} e(\bar{A}^k \bar{B}^l m, \alpha_{k,l}) - h_m.$$

- We will work with the maps of the type  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ , the usual notation being:  $p(k, l) \in C^\infty$  for  $(k, l) \in \mathbb{Z}^2$ . For such maps we define the operator  $Lp : \mathbb{Z}^2 \times \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ , by the following. For any  $(k, l), (s, t) \in \mathbb{Z}^2 \times \mathbb{Z}^2$ , denote

$$Lp((k, l), (s, t)) := \partial_{k,l} p(s, t) - \partial_{s,t} p(k, l).$$

- Let  $m$  be such that  $\bar{A}m \neq m$  (hence,  $\hat{A}m \neq 0$ ). Define

$$\mathcal{M}(A) = \{m \in \mathbb{Z}^d \mid \langle m, \hat{A}m \rangle > 0\},$$

$$\mathcal{N}(A) = \{m \in \mathbb{Z}^d \mid \langle m, \hat{A}m \rangle < 0\}.$$

- Suppose that  $A$  is step-2, and  $m$  is such that  $\bar{A}m \neq m$ . Then we have  $\bar{A}^k m = m + k\hat{A}m$ . We say that  $\bar{m}$  is the lowest point on the  $\bar{A}$ -orbit of  $m$  if

$$|\bar{m}| \leq |\bar{m} + k\hat{A}\bar{m}| \quad \text{for all } k \in \mathbb{Z}.$$

Then we have a "switch":  $\bar{m} \in \mathcal{M}(A)$  but  $\bar{A}^{-1}\bar{m} \in \mathcal{N}(A)$ , or vice versa. Note that  $\bar{m}$  is the only point on the corresponding  $\bar{A}$ -orbit in which the "switch" between  $\mathcal{N}(A)$  and  $\mathcal{M}(A)$  happens.

We write  $m = \bar{m}$  to say that  $m$  is the lowest point on its own  $\bar{A}$ -orbit, and  $m \neq \bar{m}$  otherwise.

- Let

$$\lambda_m^{(-1)} := \lambda_m := e(m, \alpha), \quad \mu_m^{(-1)} := \mu_m := e(m, \beta),$$

$$\lambda_m^{(k)} = \lambda_{\bar{A}m} \lambda_{\bar{A}^2 m} \cdots \lambda_{\bar{A}^k m}, \quad k = 1, 2, \dots, \quad \lambda_m^{(0)} = 1,$$

$$\lambda_m^{(k)} = (\lambda_m \lambda_{\bar{A}^{-1}m} \cdots \lambda_{\bar{A}^{k+1}m})^{-1}, \quad k = -2, -3, \dots,$$

$$\mu_m^{(k)} = \mu_{\bar{B}m} \mu_{\bar{B}^2 m} \cdots \mu_{\bar{B}^k m}, \quad k = 1, 2, \dots, \quad \mu_m^{(0)} = 1,$$

$$\mu_m^{(k)} = (\mu_m \mu_{\bar{B}^{-1}m} \cdots \mu_{\bar{B}^{k+1}m})^{-1}, \quad k = -2, -3, \dots$$

- For  $A$  of step-2 and  $m \in \mathbb{Z}$ , consider the following partial sums over the dual orbit of  $A$ :

$$\Sigma_m^{+,A}(f) := \sum_{k=0}^{\infty} f_{\bar{A}^k m} \lambda_m^{(k)}, \quad \Sigma_m^{-,A}(f) := \sum_{k=-\infty}^{-1} f_{\bar{A}^k m} \lambda_m^{(k)},$$

$$\Sigma_m^A(f) := \Sigma_m^{+,A}(f) + \Sigma_m^{-,A}(f).$$

The last two-sided sum defines the so-called "obstruction operator".

- Let  $U \subset \mathbb{Z}^d$  be a set that is invariant under the action of  $\langle \bar{A}, \bar{B} \rangle$ , i.e., for any  $m \in U$  we have  $\bar{A}^s \bar{B}^t m \in U$ . Consider a set of real numbers, indexed by  $U$ :  $\xi = (\xi_m) = \{\xi_m \mid m \in U\}$ . With a little abuse of notation, we let the operator  $\partial_{s,t}$  act on  $\xi$ . Namely,

$$(\partial_{s,t}\xi)_m = \xi_{\bar{A}^s \bar{B}^t m} e(\bar{A}^s \bar{B}^t m, \alpha_{s,t}) - h_m.$$

Since the set of indices is invariant under the action, this expression is well-defined.

Comment: This notation is needed because we will define the conjugating functions  $h$  by Fourier coefficients in different ways for different (invariant) sets of indices:  $\mathcal{C}_1$ ,  $\mathcal{C}_2$  or  $\mathcal{C}_3$  (see below), and will need to solve equations in terms of Fourier coefficients before we have defined  $h$  as a function.

- The following splitting of  $\mathbb{Z}^d \setminus \{0\}$  will be used in our analysis.

$$\mathbb{Z}^d \setminus \{0\} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3,$$

where the sets  $\mathcal{C}_j$  are defined as follows.

$\mathcal{C}_1$ . (**Degenerate case**).  $\mathcal{C}_1$  is the set of  $m$  for which  $\bar{A}m = \bar{B}m = m$ .

$\mathcal{C}_2$ . (**Resonant non-degenerate case**). For  $(k, l) \in \mathbb{Z}^2 \setminus \{0\}$  we say that  $m \in \mathcal{C}_2(k, l)$  if the following holds:

- $m \notin \mathcal{C}_1$ ;
- $\bar{A}^k \bar{B}^l m = m$ .

We define  $\mathcal{C}_2 = \bigcup_{(k,l) \in \mathbb{Z}^2 \setminus \{0\}} \mathcal{C}_2(k, l)$ .

$\mathcal{C}_3$ . (**Non-resonant case**).  $\mathcal{C}_3 = \mathbb{Z}^2 \setminus (\{0\} \cup \mathcal{C}_1 \cup \mathcal{C}_2)$ .

**3.2. Estimates of the sums along the dual orbits of a step-2 matrix.** In the constructions that follow we will work with vectors  $m$  lying in certain subsets of  $\mathbb{Z}^d$  that are invariant under the action of  $\bar{A}$  and  $\bar{B}$ . Recall the notations  $\mathcal{M}(A)$  and  $\mathcal{N}(A)$  from Sec. 3.1.

**Proposition B.** *Let  $A$  be step-2 and suppose that  $\bar{A}m \neq m$ .*

(1) *Consider a set  $U \subset \mathbb{Z}^d$  that is invariant under the action of  $\bar{A}$ . Let  $(\xi_m) = \{\xi_m \in \mathbb{R} \mid m \in U\}$ , and suppose that for all  $m \in U$  we have  $|\xi_m| \leq c_0 |m|^{-r}$ . Then there exists  $c > 0$  such that*

- *If  $m \in \mathcal{M}(A) \cap U$ , then  $\sum_{k=0}^{\infty} |\xi_{\bar{A}^k m}| \leq c |m|^{-r+1}$ ,*
- *If  $m \in \mathcal{N}(A) \cap U$ , then  $\sum_{k=-\infty}^{-1} |\xi_{\bar{A}^k m}| \leq c |m|^{-r+1}$ ,*
- *If, moreover,  $m = \bar{m}$  (i.e.,  $m$  is the lowest point on its  $\bar{A}$ -orbit), then*

$$\sum_{k=-\infty}^{\infty} |\xi_{\bar{A}^k m}| \leq c |m|^{-r+1}.$$

(2) *Consider the sets  $U$  and  $(\xi_m)$  as in (1) and  $\lambda_m^{(1)}$  as in Sec. 3.1. Suppose that for each  $m \in U$ , the set of numbers  $(\zeta_m)$  satisfies*

$$\zeta_{\bar{A}m} \lambda_m^{(1)} - \zeta_m = \xi_m.$$

*Then there exists  $c > 0$  such that for each  $m \in U$  we have:*

$$|\zeta_m| \leq c |m|^{-r+1}.$$

(3) *There exists  $c > 0$  such that for any function  $\xi \in C^r$ , we have:*

- *If  $m \in \mathcal{M}(A)$ , then  $|\Sigma_m^{+,A}(\xi)| \leq c \|\xi\|_r |m|^{-r+1}$ ,*
- *If  $m \in \mathcal{N}(A)$ , then  $|\Sigma_m^{-,A}(\xi)| \leq c \|\xi\|_r |m|^{-r+1}$ .*
- *If, moreover,  $m = \bar{m}$  is the lowest point on its  $\bar{A}$ -orbit, then*

$$|\Sigma_m^A(\xi)| \leq c |m|^{-r+1}.$$

**Proof.** Item (1) follows directly from the estimate below, that will be used several times in the paper.

**Sublemma 1.** *Let  $A$  be step-2 and  $\bar{A}m \neq m$ . Then there exists  $c = c(r, A) > 0$  such that we have:*

- *If  $m \in \mathcal{M}(A)$ , then  $\sum_{k=0}^{\infty} |\bar{A}^k m|^{-r} \leq c |m|^{-r+1}$ ;*
- *If  $m \in \mathcal{N}(A)$ , then  $\sum_{k=-\infty}^{-1} |\bar{A}^k m|^{-r} \leq c |m|^{-r+1}$ .*
- *If, moreover,  $m = \bar{m}$  (i.e.,  $m$  is the lowest point on its  $\bar{A}$ -orbit), then*

$$\sum_{k=-\infty}^{\infty} |\bar{A}^k m|^{-r} \leq c |m|^{-r+1}.$$

*Proof.* Since  $\bar{A}$  is step-2 and  $\bar{A}m \neq m$ , for any  $m \in \mathbb{Z}^d$  and  $k \in \mathbb{Z}$ , we have  $\bar{A}^k m = m + k\hat{A}m$ .

Consider the case  $m \in \mathcal{M}(A)$ , i.e.,  $\langle m, \hat{A}m \rangle \geq 0$ . Then we have:  $|m + \hat{A}m| > |m|$  and  $\langle m + \hat{A}m, \hat{A}m \rangle > 0$  (notice the strict inequality). Let  $p$  denote the projection of the vector  $\hat{A}m$  onto the vector  $m + \hat{A}m$ . Clearly,  $p$  is non-zero and has the same direction as  $m + \hat{A}m$ . Therefore,  $|m + k\hat{A}m| = |(m + \hat{A}m) + (k-1)\hat{A}m| \geq |m + \hat{A}m| + (k-1)|p| > |m| + (k-1)|p|$ . Hence, for an appropriate constant  $c = c(r, A)$  we have:

$$\sum_{k=0}^{\infty} |\bar{A}^k m|^{-r} \leq \sum_{k=0}^{\infty} |(m + k\hat{A}m)|^{-r} \leq c|m|^{-r+1}.$$

To justify the last inequality, note that for any  $x, y > 0$ , and  $r > 1$  we have:

$$\sum_{l=1}^{\infty} (x + ly)^{-r} \leq \frac{c_1}{y(r-1)} (x + y)^{-r+1},$$

which can be proved by comparison with the integral  $y^{-r} \int_{t=1}^{\infty} (\frac{x}{y} + t)^{-r} dt$ .

The case of  $m \in \mathcal{N}(A)$ , i.e.,  $\langle m, \hat{A}m \rangle < 0$ , is similar. Indeed, the projection of the vector  $(-\hat{A}m)$  onto  $m$  has the same direction as  $m$ , and the above calculation holds.

Now let  $m = \bar{m}$  be the lowest point on its  $\bar{A}$ -orbit. Then, in particular,  $|m + \hat{A}m| > |m|$  and  $|m - \hat{A}m| > |m|$ . This implies, for example by studying the triangle with two sides formed by vectors  $m + \hat{A}m$  and  $m - \hat{A}m$ , that we have both  $\langle m + \hat{A}m, \hat{A}m \rangle > 0$  and  $\langle m - \hat{A}m, \hat{A}m \rangle > 0$ . Then we can use the two estimates above to conclude that

$$\sum_{k=-\infty}^{\infty} |\bar{A}^k m|^{-r} = \sum_{k=-\infty}^{-1} |\bar{A}^k m|^{-r} + \sum_{k=0}^{\infty} |\bar{A}^k m|^{-r} \leq c'|m|^{-r+1}.$$

□

To prove (2), for a fixed  $m$  write the given equation at the points  $\bar{A}^k m$  either for  $k \geq 0$  or for  $k \leq -1$ , multiply by appropriate constants and add up, obtaining a telescopic sum on the left-hand side. Then we get that  $|\zeta_m| \leq \sum_{k=0}^{\infty} |\xi_{\bar{A}^k m}|^{-r}$  or  $|\zeta_m| \leq \sum_{k=-\infty}^{-1} |\xi_{\bar{A}^k m}|^{-r}$ . The estimate follows from (1).

To prove (3), recall that the Fourier coefficients of any  $\xi \in C^r$  satisfy for all  $m \in \mathbb{Z}^d$ :

$$|\xi_m| \leq \|\xi\|_r |m|^{-r}.$$

Then for each  $k \in \mathbb{Z}$  we have:  $|\xi_{\bar{A}^k m}| \leq \|\xi\|_r |\bar{A}^k m|^{-r}$ . The result reduces to that of item (1). □

**3.3. Estimates of the double sums. The parabolic higher rank trick.** The double sums will be used for the case  $m \in \mathcal{C}_3$  (non-resonant case). For each  $m$ , one of the double sums is easier to estimate than the other. The corresponding sign of  $l$  will be called the "good sign" of  $l$  for the given  $m$ .

**Proposition C (Estimate of the double sums).** *Assume that  $\langle a, b \rangle$  is unlocked parabolic affine step-S action, where  $a$  is step-2. Suppose that  $m \in \mathcal{C}_3$  is the lowest point on its  $\bar{A}$ -orbit.*



For  $r$  sufficiently large there exists a constant  $c = c(r, A, B) > 0$  such that for  $\eta = 0.99/S$ , at least one of the following holds:

$$\sum_{k \in \mathbb{Z}} \sum_{l \geq 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq c|m|^{-\eta r+8}, \quad \sum_{k \in \mathbb{Z}} \sum_{l < 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq c|m|^{-\eta r+8}.$$

The proof of Proposition C is crucial for our analysis, it is rather technical and takes up the rest of this section.

**3.3.1. Implications of being unlocked.** Recall the notion of being unlocked from Definition 4. Let us make two observations.

**Lemma 11.** *Suppose that the action  $\langle a, b \rangle$  is unlocked. If  $\hat{B}^2 m \neq 0$ , then  $\hat{A}m \neq 0$ .*

*Proof.* Let  $\hat{B}^2 m \neq 0$ , and suppose by contradiction that  $\hat{A}m = 0$ . Consider the function  $g(x) = e(m, \tilde{B}x)$ . Observe that

$$g(ax) = g(Ax + \alpha) = e(m, \tilde{B}(x + \tilde{A}x + \alpha)) = e(m, \tilde{B}x)e(m, \tilde{B}\tilde{A}x)e(m, \tilde{B}\alpha).$$

Since  $m\tilde{A} = (\hat{A}m)^t = 0$ , and since (by commutativity  $ab = ba$ ) we have  $\tilde{B}\alpha = \tilde{A}\beta = 0$ , we conclude that  $g(ax) = g(x)$ .

On the other hand,

$$g(bx) = e(m, \tilde{B}x)e(m, \tilde{B}^2x)e(m, \tilde{B}\beta) = g(x)e(m, \tilde{B}^2x)e(m, \tilde{B}\beta).$$

Since  $m\tilde{B}^2 = \hat{B}^2 m \neq 0$ , we conclude that the action  $(a, b)$  has a rank-one factor that is not a translation.  $\square$

The second observation is

**Lemma 12.** *Suppose that the action  $\langle a, b \rangle$  is unlocked and  $\hat{A}^2 m = 0$ . If  $\hat{B}^s m = 0$  for some  $s \geq 2$ , then  $\hat{A}\hat{B}^{s-1}m = 0$ .*

*Proof.* Let  $\hat{B}^s m = 0$ , and suppose by contradiction that  $\hat{A}\hat{B}^{s-1}m \neq 0$ . Define  $f(x) = e(m, \tilde{A}\tilde{B}^{s-1}x)$ . One easily verifies that  $f(ax) = f(x)$  and  $f(bx) = f(x)$  (using relations  $\tilde{A}\tilde{B}^{s-1}\alpha = \tilde{A}^2\tilde{B}^{s-2}\beta$ ,  $\tilde{A}\tilde{B}^{s-1}\beta = \tilde{B}^s\alpha$ ). Hence the action  $\langle A + \alpha, B + \beta \rangle$  has a rank-one factor equal to identity, contradicting the assumption.  $\square$

**3.3.2. Polynomial expansion of  $\bar{A}^k \bar{B}^l m$ .** Recall the notations from §3.1.

**Lemma 13.** *Assume that  $\langle a, b \rangle$  is unlocked parabolic affine action, and  $a$  is step-2. For any  $m \in \mathcal{C}_3$ ,  $k, l \in \mathbb{Z}$ ,  $s = s(m)$ , there exists  $t = t(m, A, B)$ ,  $1 \leq t \leq s-1$ , such that*

$$\bar{A}^k \bar{B}^l m = m + k\hat{A}m + \sum_{j=1}^{t-1} c_j l^j \hat{B}^j (m + k\hat{A}m) + \sum_{j=t}^{s-1} c_j l^j \hat{B}^j m, \quad (26)$$

where  $c_1 = 1$ , and all  $c_j$  are positive constants that can be computed explicitly.

*Proof.* **Case**  $s = s(m) = 2$ . Here we have  $\hat{A}^2 m = \hat{B}^2 m = 0$  (since  $s(m) = 2$ ), and  $\hat{A}m \neq 0$ ,  $\hat{B}m \neq 0$  (since  $m \in \mathcal{C}_3$ ). Therefore,  $\bar{A}^k m = m + k\hat{A}m$  and  $\bar{B}^l m = m + l\hat{B}m$ . By Lemma 12,  $\hat{B}^2 m = 0$  implies that  $\hat{A}\hat{B}m = 0$ , which gives the result.

**Case**  $s = s(m) \geq 3$ . Here we have  $\hat{B}^2 m \neq 0$ , so, by Lemma 11,  $\hat{A}m \neq 0$ . Since  $\hat{A}^2 m = \hat{B}^s m = 0$ , we have  $\bar{A}^k m = m + k\hat{A}m$  and  $\bar{B}^l m = m + \sum_{j=1}^{s-1} c_j l^j \hat{B}^j m$ . Note that, by Lemma 12,  $\hat{A}\hat{B}^{s-1}m = 0$ . Composing the two expressions above and using the commutativity gives the result.  $\square$

Recall that for  $m \in \mathcal{C}_3$  we have the following two possibilities:

- $s(m) = 2$ , in which case  $\widehat{B}^2 m = 0$ , and for all  $(k, l) \in \mathbb{Z}^2 \setminus \{0\}$  we have  $k\widehat{A}m + l\widehat{B}m \neq 0$ ;
- $s(m) \geq 3$ , in which case  $\widehat{B}^2 m \neq 0$ .

We will use different ways of controlling the double sums for the two cases above.

**3.3.3. Proof of Proposition C, case  $s(m) = 2$ .** In this section we fix  $m \in \mathcal{C}_3$ ,  $s(m) = 2$ , and study the growth properties of

$$v_{k,l} := \bar{A}^k \bar{B}^l m = m + k\widehat{A}m + l\widehat{B}m. \quad (27)$$

The following two lemmas prove that there exists a constant  $c = c(A) > 0$  such that if  $m$  is the lowest point on its  $\bar{A}$ -orbit, then for all  $k \in \mathbb{Z}$  and either for all  $l \in \mathbb{N}$  or for all  $l \in (-\mathbb{N})$  we have:

$$|v_{k,l}| > c|m|.$$

**Lemma 14.** *Let  $m \in \mathcal{C}_3$ ,  $s(m) = 2$ . If  $|\widehat{A}m| > |m|$ , then for all  $k, l \in \mathbb{Z}$  we have:*

$$|v_{k,l}| > \|\widehat{A}\|^{-1}|m|.$$

*Proof.* Assume the contrary:  $|v_{k,l}| \leq \|\widehat{A}\|^{-1}|m|$ . Apply  $\widehat{A}$  to equality (27). Since  $\widehat{A}^2 m = \widehat{A}\widehat{B}m = 0$ , we have:

$$|\widehat{A}m| = |\widehat{A}v_{k,l}| \leq \|\widehat{A}\| |v_{k,l}| \leq |m|,$$

contradicting the assumption of the lemma.  $\square$

**Lemma 15.** *Let  $m \in \mathcal{C}_3$  be the lowest point on its  $\bar{A}$ -orbit, and  $s(m) = 2$ .*

*If  $|\widehat{A}m| \leq |m|$ , then for all  $k \in \mathbb{Z}$  and either for all  $l \in \mathbb{N}$  or for all  $l \in (-\mathbb{N})$  we have:*

$$|v_{k,l}| \geq |m|/2.$$

*Proof.* Since  $m \in \mathcal{C}_3$ , we have  $\widehat{A}m \neq 0$ . Denote  $\mathcal{V}_m = \text{span}\{\widehat{A}m\}$ , and let  $m^\perp$  and  $(\widehat{B}m)^\perp$  stand for the projections of  $m$  and  $\widehat{B}m$ , respectively, onto the orthogonal complement of  $\mathcal{V}_m$ . Then

$$|\bar{A}^k \bar{B}^l m| = |m + k\widehat{A}m + l\widehat{B}m| \geq |m^\perp + l(\widehat{B}m)^\perp|.$$

To prove the lemma, it is enough to show that  $|m^\perp| \geq |m|/2$ . When this is done, we choose the "good sign" of  $l$  to be positive if the angle between the vectors  $m^\perp$  and  $(\widehat{B}m)^\perp$  is acute, and negative otherwise. If we choose  $l$  of good sign, then  $|\bar{A}^k \bar{B}^l m| \geq |m^\perp| \geq |m|/2$ .

Let us estimate  $|m^\perp|$ . Suppose that the angle  $\theta$  between the vectors  $m$  and  $\widehat{A}m$  satisfies  $0 < \theta \leq \pi/2$  (otherwise, use  $(-\widehat{A}m)$  instead of  $\widehat{A}m$ ). Since  $m$  is the lowest point on its  $\widehat{A}$ -orbit, we have:  $|m - \widehat{A}m| \geq |m|$ . We see in this case that the projection of  $m$  onto  $\mathcal{V}_m$  satisfies

$$\|\text{proj}_{\mathcal{V}_m} m\| \leq \|\widehat{A}m\|/2.$$

Hence,

$$|m^\perp| \geq |m| - \|\text{proj}_{\mathcal{V}_m} m\| \geq |m| - \|\widehat{A}m\|/2 \geq |m| - |m|/2 = |m|/2.$$

$\square$

**Lemma 16** (Estimate of the double sums,  $s(m) = 2$ ). *Assume that  $\langle a, b \rangle$  is a unlocked parabolic affine action, and  $a$  is step-2. Suppose that  $m \in \mathcal{C}_3$ ,  $s(m) = 2$ , and  $m$  is the lowest point on its  $\bar{A}$ -orbit.*

*For  $r$  sufficiently large there exists a constant  $c = c(r, A, B) > 0$  such that at least one of the following holds:*

$$\sum_{k \in \mathbb{Z}} \sum_{l \geq 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq C|m|^{-r+8}, \quad \sum_{k \in \mathbb{Z}} \sum_{l < 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq C|m|^{-r+8}.$$

*Proof.* Assume without loss of generality that the good sign of  $l$  is positive.

Denote  $u := \hat{A}m$ ,  $v := \hat{B}m$  and let  $C_{k,l} = k\hat{A}m + l\hat{B}m$ . Since  $u$  and  $v$  are non-parallel integer vectors whose sizes satisfy, for some  $c_0 > 0$ ,

$$1 \leq |u|, |v| \leq c_0|m|,$$

one can show that the angle  $\theta$  between  $u$  and  $v$  satisfies  $\sin \theta \geq c_1|m|^{-2}$ , and hence  $\gamma := |\cos \theta| \leq 1 - c_2|m|^{-4}$ . Note that for all  $k, l \in \mathbb{Z}$  we have the following two inequalities:

$$\begin{aligned} |C_{k,l}|^2 &= |ku + lv|^2 = k^2|u|^2 + l^2|v|^2 + 2kl|u||v|\gamma, \\ (k|u| + l|v|)^2 &= k^2|u|^2 + l^2|v|^2 + 2kl|u||v| \geq 0. \end{aligned}$$

Hence,

$$\begin{aligned} |C_{k,l}|^2 &\geq |C_{k,l}|^2 - \gamma(k|u| + l|v|)^2 = (1 - \gamma)(k^2|u|^2 + l^2|v|^2) \\ &\geq c_2|m|^{-4}(k^2|u|^2 + l^2|v|^2) \geq c_2|m|^{-4}(k^2 + l^2). \end{aligned}$$

Now, if  $k^2 + l^2 \geq c_3|m|^8$ , then for  $|m| > 1$  we have  $|m| \leq |C_{k,l}|/2$ , and

$$|\bar{A}^k \bar{B}^l m| \geq |C_{k,l}| - |m| \geq |C_{k,l}|/2 \geq |m|.$$

Finally, we split the desired sum:

$$\sum_{k \in \mathbb{Z}} \sum_{l \geq 0} |\bar{A}^k \bar{B}^l m|^{-r} = \Sigma_1 + \Sigma_2,$$

where  $\Sigma_1$  contains the terms  $|\bar{A}^k \bar{B}^l m|^{-r}$  corresponding to  $k^2 + l^2 \leq c_3|m|^8$ , and  $\Sigma_2$  contains those with  $k^2 + l^2 > c_3|m|^8$ .

The sum  $\Sigma_1$  contains  $\leq 4c_3|m|^8$  terms. By Lemmas 14 and 15, for a certain  $c = c(r, A, B)$  we have for all  $m$ :

$$|v_{k,l}|^{-r} \leq c|m|^{-r},$$

so

$$\Sigma_1 \leq c_4|m|^{-r+8}.$$

We estimate  $\Sigma_2$  by comparison with an integral:

$$\begin{aligned} \Sigma_2 &= \sum_{k^2 + l^2 \geq |m|^8} |v_{k,l}|^{-r} \leq c_5|m|^{2r} \sum_{k^2 + l^2 \geq |m|^8} (k^2 + l^2)^{-r/2} \\ &\leq c_5|m|^{2r} \int_{x^2 + y^2 \geq |m|^8} (x^2 + y^2)^{-r/2} dx dy \leq c_6|m|^{-2r+8} \leq c_6|m|^{-r+8}. \end{aligned}$$

The combination of the estimates for  $\Sigma_1$  and  $\Sigma_2$  provides the desired result.  $\square$

The following subsections contain the proof of Proposition C for the case  $s(m) \geq 3$ . We assume that  $m \in \mathcal{C}_3$ ,  $s = s(m) \geq 3$ , and study the growth properties of  $|\bar{A}^k \bar{B}^l m|$ , given by formula (26):

$$v_{k,l} := \bar{A}^k \bar{B}^l m = m + k\hat{A}m + \sum_{j=1}^t c_j l^j \hat{B}^j(m + k\hat{A}m) + \sum_{j=t+1}^{s-1} c_j l^j \hat{B}^j m.$$

**3.3.4. The case  $s(m) \geq 3$ : Estimate for small  $l$ .** Recall the notation  $\delta = \delta(m) = 0.99\frac{1}{s}$  from Sec. 3.1.

**Lemma 17.** *Assume that  $\langle a, b \rangle$  is unlocked parabolic affine action, and  $a$  is step-2. For any  $\xi > 0$ , there exists  $c = c(A, B, \xi) > 0$  such that for any  $m \in \mathcal{C}_3$  with  $s(m) \geq 3$ , being the lowest point on its  $\bar{A}$ -orbit, we have for any  $k \in \mathbb{Z}$  and for any  $|l| < |\xi m|^\delta$ :*

$$|\bar{A}^k \bar{B}^l m| \geq c|m|^\delta. \quad (28)$$

*Proof.* Assume that  $\xi = 1$ ; the same proof holds for any  $\xi > 0$ . Denote  $v_{k,l} := \bar{A}^k \bar{B}^l m$  for brevity. First consider "large"  $m$ , such that  $|m| > C_0$  for an appropriate constant  $C_0$ . For this  $m$ , suppose by contradiction that  $|v_{k,l}| < |m|^\delta$  for some  $|l| \leq |m|^\delta$ . By assumption,  $m$  is the lowest point on its  $\bar{A}$ -orbit, so  $|m + k\hat{A}m| \geq |m|$  for any  $k \in \mathbb{Z}$ . Applying inductively  $\hat{B}^{s-j}$ ,  $j = 1, \dots, s-1$  to equation (26), we get for a certain constant  $C = C(A, B)$ :

$$|v_{k,l} - (m + k\hat{A}m)| \leq C|m|^{s\delta} \leq C|m|^{0.99}.$$

If  $m$  satisfies  $|m| \geq (2C)^{100} := C_0$ , then the latter implies

$$|v_{k,l} - (m + k\hat{A}m)| \leq |m|/2.$$

Since  $|m + k\hat{A}m| \geq |m|$ , we conclude that  $|v_{k,l}| \geq |m|/2$  which is in contradiction with our assumption that  $|v_{k,l}| \leq |m|^\delta$ . Thus, we have proved the desired estimate for all  $|m| \geq (2C)^{100}$ . If  $|m| < (2C)^{100}$ , the estimate is achieved by the choice of a sufficiently small constant  $c(A, B)$ .  $\square$

**3.3.5. The case  $s(m) \geq 3$ : Linear Drift in  $l$ .** This is the section where the "good sign of  $l$ " for the given  $m \in \mathbb{Z}^d$  plays the crucial role.

**Lemma 18 (Linear Drift in  $l$ ,  $s(m) \geq 3$ ).** *Assume that  $\langle a, b \rangle$  is unlocked parabolic affine action, and  $a$  is step-2. There exists a constant  $c = c(A, B) > 0$  such that for any  $m \in \mathcal{C}_3$ ,  $s(m) \geq 3$ , the following holds: for all  $k \in \mathbb{Z}$  and either for all  $l \geq 0$  or for all  $l < 0$  we have:*

$$|\bar{A}^k \bar{B}^l m| \geq c|l|.$$

*Proof.* Recall that  $s = s(m) \geq 3$ , which means that  $\hat{B}^s m = 0$ ,  $\hat{B}^{s-1} m \neq 0$ . In particular,  $\hat{B}^2 m \neq 0$ . By Lemma 12, the assumption on being unlocked implies  $\hat{A}\hat{B}^{s-1} m = 0$ . Denote

$$\mathcal{V}_m = \text{span} \{ \hat{B}^l m, \hat{A}\hat{B}^{l'} m \mid l \in [2, s-1], l' \in [0, s-2] \},$$

where the terms  $\widehat{A}\widehat{B}^{l'}m$  may vanish starting from some  $l' = t$ ,  $t \geq 1$ . Let  $m^\perp$  and  $(\widehat{B}m)^\perp$  denote the orthogonal projections of  $m$  and  $\widehat{B}m$ , respectively, onto the orthogonal complement of  $\mathcal{V}_m$ . Let us show that for some constant  $c = c(A, B) > 0$  we have

$$|(\widehat{B}m)^\perp| \geq c > 0.$$

We consider two subcases.

**Case**  $\widehat{A}\widehat{B}^{s-2}m = 0$ . Here we have:

$$0 \neq \widehat{B}^{s-1}m = \widehat{B}^{s-2}(\widehat{B}m)^\perp.$$

Since  $\widehat{B}^{s-1}m \neq 0$  is an integer, we have  $|\widehat{B}^{s-1}m| \geq 1$ . Since the norm of  $\widehat{B}^{s-2}$  is bounded away from zero, we have  $|(\widehat{B}m)^\perp| \geq c_0(A, B) > 0$ .

**Case**  $\widehat{A}\widehat{B}^{s-2}m \neq 0$ . In this case we have

$$0 \neq \widehat{A}\widehat{B}^{s-2}m = \widehat{A}\widehat{B}^{s-3}(\widehat{B}m) = \widehat{A}\widehat{B}^{s-3}(\widehat{B}m)^\perp.$$

Since  $\widehat{A}\widehat{B}^{s-2}m \neq 0$  is an integer, we have  $|\widehat{A}\widehat{B}^{s-2}m| \geq 1$ . Since the norm of  $\widehat{A}\widehat{B}^{s-3}$  is bounded away from zero, this implies  $|(\widehat{B}m)^\perp| \geq c_1(A, B) > 0$ .

To complete the proof, recall that, by (26),

$$\begin{aligned} |v_{k,l}| &= |\bar{A}^k \bar{B}^l m| = |m + k\widehat{A}m + \sum_{j=1}^{t-1} c_j l^j \widehat{B}^j(m + k\widehat{A}m) + \sum_{j=t}^{s-1} c_j l^j \widehat{B}^j m| \\ &\geq |m^\perp + l(\widehat{B}m)^\perp|. \end{aligned}$$

Choose the "good sign of  $l$ " to be positive if the vectors  $m^\perp$  and  $l(\widehat{B}m)^\perp$  form an acute angle, and negative otherwise. For this sign of  $l$  we get the desired result.  $\square$

**3.3.6. The case  $s(m) \geq 3$ : Drift in  $k$ .** Recall the notations  $s = s(m)$  and  $\delta = \delta(m)$  from Section 3.1.

**Lemma 19 (Drift in  $k$ ).** *Assume that  $\langle a, b \rangle$  is unlocked parabolic affine action, and  $a$  is step-2. There exist positive constants  $\xi = \xi(A, B)$  and  $C = C(A, B)$  such that for any  $m \in \mathcal{C}_3$ , for any  $k, l$  satisfying  $|k| \geq \xi|m|$ ,  $|l| \leq |k|^\delta$  with  $\delta = \delta(m)$  defined in Section 3.1, we have:*

$$|\bar{A}^k \bar{B}^l m| \geq C|k|^\delta.$$

*Proof.* Let  $s = s(m)$  be the step of  $m$ , defined in Section 3.1. Since  $\delta = \delta(m) = 0.99/s$ , condition  $|l| \leq |k|^\delta$  implies  $|k| \geq |l|^s$ . Let  $p \in [0, s-1]$  be the largest integer such that  $\widehat{A}\widehat{B}^p m \neq 0$ , and observe that from (26) and

$$\widehat{B}^p \bar{A}^k \bar{B}^l m = \widehat{B}^p m + k\widehat{A}\widehat{B}^p m + \mathcal{O}(l^{s-1}), \quad (29)$$

where  $\mathcal{O}(l^{s-1})$  denotes the terms free from  $k$  with the maximal power of  $l$  being  $s-1$ . If we assume that  $|k| \geq \xi(A, B)|m|$  with  $\xi(A, B)$  sufficiently large, then the linear term in  $k$  is dominant in (29) so that  $|\widehat{B}^p \bar{A}^k \bar{B}^l m| \geq |k|/2$ , thus  $|\bar{A}^k \bar{B}^l m| \geq C|k|$  for a certain positive constant  $C$ .  $\square$

**3.3.7. Proof of Proposition C, case  $s(m) \geq 3$ .** We now turn to the effective control of the double sums.

*Proof.* Assume without loss of generality that the good sign of  $l$  is positive. Let  $\xi = \xi(A, B) > 0$  be the constant from Lemma 19, and let  $\delta = \delta(m) = 0.99/s(m)$ , as before. We split the sum into the following five partial sums, each of which will be estimated separately:

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \sum_{l \geq 0} |\bar{A}^k \bar{B}^l m|^{-r} &= \left( \sum_{|k| \leq |\xi m|} \sum_{l \leq |\xi m|^\delta} + \sum_{|k| \leq |\xi m|} \sum_{l > |\xi m|^\delta} + \right. \\ &\quad \left. \sum_{|k| > |\xi m|} \sum_{l > |k|^\delta} + \sum_{|k| > |\xi m|} \sum_{l < |\xi m|^\delta} + \sum_{|k| > |\xi m|} \sum_{|\xi m|^\delta \leq l \leq |k|^\delta} \right) |\bar{A}^k \bar{B}^l m|^{-r} \\ &:= \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4 + \Sigma_5. \end{aligned}$$

The following elementary estimate is used several times below: for any  $p, r > 0$ , we have:

$$\sum_{j \geq p} j^{-r} \leq c_0(r) p^{-r+1}.$$

**Estimate of  $\Sigma_1$ .** By Lemma 17, there exists  $c = c(A, B)$  such that for all  $l \leq |\xi m|^\delta$  and for all  $k \in \mathbb{Z}$  we have  $|\bar{A}^k \bar{B}^l m| \geq c|m|^\delta$ . The sum  $\Sigma_1$  contains  $\leq 3|\xi m|^{\delta+1}$ . Hence,

$$\Sigma_1 \leq 3|\xi m|^{\delta+1} (c|m|^\delta)^{-r} < c_1 |m|^{-r\delta+2}.$$

**Estimate of  $\Sigma_2$  and  $\Sigma_3$ .** By Lemma 18, there exists  $c = c(A, B)$  such that for all  $k \in \mathbb{Z}$  and for all  $l \geq 0$  we have:  $|\bar{A}^k \bar{B}^l m| \geq cl$ . In the case of  $\Sigma_2$  we have  $|k| \leq \xi|m|$ , so

$$\Sigma_2 \leq 3\xi |m| c^{-r} \sum_{l > |\xi m|^\delta} l^{-r} \leq c_2 |m|^{-r\delta+2}.$$

In the case of  $\Sigma_3$  we have:

$$\Sigma_3 \leq 3c^{-r} \sum_{k > \xi|m|} \sum_{l > k^\delta} l^{-r} \leq \tilde{c}_3 \sum_{k > \xi|m|} (k^\delta)^{-r+1} \leq c_3 |m|^{-r\delta+2}.$$

**Estimate of  $\Sigma_4$  and  $\Sigma_5$ .** By Lemma 19, if  $|k| > \xi|m|$  and  $0 \leq l \leq |k|^\delta$ , then  $|\bar{A}^k \bar{B}^l m| \geq C|k|^\delta$ . Therefore,

$$\Sigma_4 \leq C|\xi m|^\delta \sum_{k > \xi|m|} k^{-\delta r} \leq c_4 |m|^{-\delta r+2},$$

and

$$\Sigma_5 \leq \sum_{l \geq |\xi m|^\delta} \sum_{k \geq l^{1/\delta}} k^{-r\delta} \leq c_5 \sum_{l \geq |\xi m|^\delta} (l^{1/\delta})^{-r\delta+1} \leq c_5 |m|^{-r\delta+2}.$$

Recall that for any  $m \in \mathbb{Z}^d$  we have  $\delta(m) = 0.99/s(m) \geq \eta = 0.99/S$ , where  $S$  is the step of  $B$ . Therefore,  $|m|^{-r\delta+2} \leq |m|^{-r\eta+2}$ . Summing up the above estimates, we obtain the desired result.  $\square$

#### 4. SOLUTION OF THE LINEARIZED PROBLEM. PROOF OF PROPOSITION A

A way of interpreting the statement of Proposition A is the following. A perturbation  $\langle F, G \rangle$  of the action  $\langle a, b \rangle$  defines a map  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$  by  $\mathbf{p}(k, l) := F^k G^l - a^k b^l$ ,  $(k, l) \in \mathbb{Z}^2$ . Proposition A (in fact) claims that there exists a tame map which projects  $\mathbf{p}$  to the space of (twisted) coboundaries over  $\langle a, b \rangle$  in such a way that the complement of this projection has quadratic estimates with respect to  $\mathbf{p}$ . How is commutativity going to give us that the error we make while projecting is quadratically small? Commutativity relations for all action elements tell us that certain linear operator  $\mathbf{Lp}$  (see §4.2.2) defined on  $\mathbf{p}$  is bounded (roughly) by the size of the square of  $\mathbf{p}$ . So the core of the problem is to produce a projection of  $\mathbf{p}$  to the space of (twisted) coboundaries over  $\langle a, b \rangle$  so that the complement of this projection (the error we are making) can be bounded by the size of  $\mathbf{Lp}$ . This is done in Sections 4.1.1 and 4.2. The final §4.3 contains the proof of Proposition A. It is in this proof that we use the fact that the commutativity assumption implies that  $\mathbf{Lp}$  is quadratically small with respect to  $\mathbf{p}$ .

This interpretation of the statement of Proposition A is useful for understanding its proof. Namely, even though the statement of Proposition A contains only  $\mathbf{f}$  and  $\mathbf{g}$  (in the notations of this section it means that  $\mathbf{f} = \mathbf{p}(1, 0)$  and  $\mathbf{g} = \mathbf{p}(0, 1)$ ), in order to produce the estimates, *we need to use the whole map*  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$ , not just  $\mathbf{f}$  and  $\mathbf{g}$ . We explain this point more in §4.1 after the statement of Proposition D that contains the main estimates on the conjugacy and the error.

The plan of the proof of Proposition A is the following. We start by constructing projections to coboundaries for *function-valued* maps  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ . The main result that leads to Proposition A is Proposition D which we state in §4.1 and prove in §4.1.1. Proposition D contains the crucial estimates for the convergence of the iteration process. In §4.2 we use Proposition D to deduce the corresponding statement, Proposition E, for truncations of  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ , which we then inductively apply to obtain Proposition F for the truncated *vector field-valued* map  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$ . This passage from a function-valued map  $p$  to a vector field-valued map  $\mathbf{p}$  is quite direct due to the fact that our action  $\langle a, b \rangle$  has a *parabolic* linear part. Similar inductive argument has been used in all the other works which use KAM method for parabolic actions ([D], [DK2], [ZW], [DT]). Finally, the main result for vector field-valued maps  $\mathbf{p}$  (Proposition F) is used in §4.3 to prove the main iterative step, Proposition A.

**4.1. Approximating  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$  by a coboundary.** We start with a set of functions,  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ . Recall the definitions of  $\partial_{k,l}(h)$  and  $Lp((k, l), (s, t))$  from §3.1. Here we introduce some extra notations. For a fixed natural number  $N$  we define  $\mathcal{Q}_N$  to be the set consisting of all the resonant pairs corresponding to the resonant vectors of norm less than  $N$  (see Lemma 4 for the bound on the norm of resonant pairs with respect to the resonant vector, and for the definition of constant  $C$  which appears in the definition below). In other words,

$$\mathcal{Q}_N := \{(k, l) \in \mathcal{R}(A, B) : C(|k| + |l|) < N\} \cup \{(1, 0), (0, 1)\}. \quad (30)$$

For the simplicity of notations, we introduce the following norms, for any  $r \geq 0$ :

$$\begin{aligned} \|p\|_r &:= \max\{\|p(1, 0)\|_r, \|p(0, 1)\|_r\}, \\ \|Lp\|_{r,N} &:= \max\{\|Lp((1, 0), (k, l))\|_r, \|Lp((0, 1), (k, l))\|_r : (k, l) \in \mathcal{Q}_N\}. \end{aligned} \quad (31)$$



From this point on,  $\kappa$  will denote a constant which depends only on the action  $\langle a, b \rangle$  and the regularity  $r$ , but along the way it will absorb other constants which appear in the estimates.

The main result we prove here is:

**Proposition D.** *Let  $\langle a, b \rangle$  be an unlocked  $(\gamma, \tau)$ -Diophantine step- $S$  parabolic affine action, where  $a$  is step-2, and let  $r > 0$ . There exist constants  $\mu = \mu(S) > 0$ ,  $\sigma = \sigma(\langle a, b \rangle)$  and  $\kappa = \kappa(r, \langle a, b \rangle)$  such that for any map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$  there exists a  $C^\infty$  function  $h$  such that*

$$\|h\|_r \leq \kappa \|p\|_{r+\sigma}, \quad (32)$$

and the map  $\tilde{p} : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$  defined by

$$\tilde{p}(s, t) := \partial_{s,t} h - p(s, t) + \text{ave}(p(s, t)), \quad (s, t) \in \mathbb{Z}^2, \quad (33)$$

satisfies the following estimate:

$$\|\tilde{p}(s, t)\|_r \leq \|p(s, t)\|_r + \kappa(|s| + |t|)^{dr} \|p\|_{r+\sigma}. \quad (34)$$

Moreover, if  $p$  is truncated up to  $N$ , then  $\tilde{p}$  satisfies even the following estimate for any  $(s, t) \in \mathcal{Q}_N$ :

$$\|\tilde{p}(s, t)\|_r \leq \kappa(|s| + |t|)^{d(\mu r + \sigma)} \|Lp\|_{\mu r + \sigma, N}. \quad (35)$$

The proof of the proposition is lengthy and takes up all of the next section. Here is a short overview of the proof.

We will define  $h$  via its Fourier coefficients  $h_m$ , in different ways depending on  $m$ . We need to apply different arguments in the following three cases: when the orbit of  $m$  under the dual linear action  $\langle \bar{A}, \bar{B} \rangle$  is a single point, when it is finite under one element of the action (but not under all elements), or when it is infinite. If the orbit is a single point we are in the degenerate case. The second case is when  $m$  is resonant, otherwise  $m$  is non-resonant. As explained in §1.3, to each resonance  $m$  we can attach a unique resonance pair  $(k, l)$  for which  $\bar{A}^k \bar{B}^l m = m$ .

The special (degenerate) case when the  $\langle \bar{A}, \bar{B} \rangle$ -orbit of  $m$  is a single point, that is when  $\bar{A}m = \bar{B}m = m$ , is dealt with in the same way as in the original proof of Moser in [M] (see §4.1.2).

Next, we have the situation when  $\bar{A}m = m$  and  $\bar{B}m \neq m$ . This is a  $(1, 0)$  resonance. In this case the fact that  $\langle A, B \rangle$  is unlocked implies that  $\bar{B}^2 m = 0$  (so  $m$  is step 2 for  $\bar{B}$ ). Then we use the generator  $b$  to construct  $h_m$ . To show that  $\tilde{p}$  satisfies the needed estimate in this case, we will need the Diophantine condition on the translation vector  $\alpha_{1,0}$ .

If  $\bar{A}m \neq m$ , we use the generator  $a$  to construct  $h_m$ , and use the fact that  $\bar{A}$  is step 2 to estimate  $h_m$ . To obtain the estimate for  $\tilde{p}$  we use different strategies for resonant and non-resonant  $m$ .

When  $m$  is resonant, we will use the corresponding resonant pair  $(k, l)$  and the action element  $a^k b^l$  to estimate the error. This is exactly where we need to use all the elements of the action, i.e., the map  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$ , and not just two generators  $\mathbf{f}$  and  $\mathbf{g}$ . Moreover, it is here that we will use the Diophantine assumptions on the translation parts  $\alpha_{k,l} = a^k b^l - A^k B^l$ . To control the number of action elements we use, we need to truncate the given data first. This is why the crucial error estimate in Proposition D is stated only for truncated maps. In our arguments, as explained in Lemma 4, the norm of the resonant pair will be bounded by the norm of the resonance. Therefore, for the

estimate of the  $N$ -truncated maps, we only need to consider the resonant pairs for the resonances bounded by  $N$ . The treatment of all the resonant cases is done in §4.1.3.

Finally, if the  $\langle \bar{A}, \bar{B} \rangle$ -orbit of  $m$  is infinite, we use the double sums estimates. This part of the argument uses §3.3 and is contained in §4.1.4.

**4.1.1. Proof of Proposition D.** Let us pass to defining and estimating the numbers  $h_m$  and  $(\tilde{p}(s, t))_m$ . The arguments will strongly depend on  $m$ . Namely, we always set  $h_0 = 0$ , and for  $m$  lying in each of the three subsets,  $\mathcal{C}_1$ ,  $\mathcal{C}_2$  and  $\mathcal{C}_3$ , defined in §3.1, we have to develop a special approach. A more precise statement of Proposition D is the following.

**Proposition 20.** *Let  $\langle a, b \rangle$  be an unlocked  $(\gamma, \tau)$ -Diophantine parabolic step- $S$  affine action, where  $a$  is step-2, and let  $r > 0$ . There exist constants  $\eta > 0$  ( $\eta = 0.99/S$ ),  $\sigma = \sigma(\langle a, b \rangle)$  and  $\kappa = \kappa(r, \langle a, b \rangle)$  such that, for any map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$  there exists a set of numbers  $(h_m)$ ,  $m \in \mathbb{Z}^d \setminus \{0\}$  such that*

$$|h_m| \leq \kappa \max\{\|p(1, 0)\|_r, \|p(0, 1)\|_r\} |m|^{-r+1+\tau} \quad (36)$$

with the following property. Define a new map  $\tilde{p}$  from  $\mathbb{Z}^2$  into the space of formal Fourier series as follows:

$$(\tilde{p}(s, t))_m := h_{\bar{A}^s \bar{B}^t m} e(\bar{A}^s \bar{B}^t m, \alpha_{s,t}) - h_m - (p(s, t))_m, \quad (s, t) \in \mathbb{Z}^2. \quad (37)$$

Then it satisfies, for any  $r > 8/\eta$ :

$$|\tilde{p}(s, t)_m| \leq \begin{cases} \kappa(|s| + |t|)^{dr} \max\{\|Lp((s, t), (1, 0))\|_r, \|Lp((s, t), (0, 1))\|_r\} |m|^{-\eta r + 9 + \tau}, & m \in \mathcal{C}_1 \cup \mathcal{C}_3 \\ \kappa(|s| + |t|)^{dr} \max\{\|Lp((1, 0), (k, l))\|_r, \|Lp((0, 1), (k, l))\|_r, \\ \quad \|Lp((1, 0), (s, t))\|_r, \|Lp((0, 1), (s, t))\|_r\} |m|^{-r + \tau + 2}, & m \in \mathcal{C}_2(k, l). \end{cases} \quad (38)$$

Moreover, for  $m \in \mathcal{C}_2(k, l)$  we have  $|k| + |l| \leq C(A, B)|m|$ .

**Proof of Proposition D from Proposition 20.** Recall that in (31) we defined:

$$\|p\|_r := \max\{\|p(1, 0)\|_r, \|p(0, 1)\|_r\}.$$

Then estimate (36) directly implies estimate (32) for  $h$  with a loss of  $\sigma := \tau + d + 2$  derivatives.

The map  $\tilde{p}$ , defined in (33), satisfies the linear estimate (34), which follows from its definition and estimate (32) for  $h$ :

$$\|\tilde{p}(s, t)\|_r \leq \|\partial_{s,t} h\|_r + \|p(s, t)\|_r \leq \|p(s, t)\|_r + \kappa(|s| + |t|)^{rd} \|p\|_{r+\sigma}.$$

If  $p$  is truncated up to  $N$ , then by taking the maximum on the right hand side of (38) over all resonant pairs  $(k, l) \in \mathcal{Q}_N$  (which is a finite set), we get for any  $(s, t) \in \mathcal{Q}_N$  (see definition (31)):

$$\begin{aligned} |\tilde{p}(s, t)_m| &\leq \kappa(|s| + |t|)^{dr} \max_{(k,l) \in \mathcal{Q}_N} \{\|Lp((1, 0), (k, l))\|_r, \|Lp((0, 1), (k, l))\|_r, \\ &\quad \|Lp((1, 0), (s, t))\|_r, \|Lp((0, 1), (s, t))\|_r\} |m|^{-\eta r + \tau + 9} \\ &\leq \kappa(|s| + |t|)^{dr} \|Lp\|_{r,N} |m|^{-\eta r + \tau + 9}. \end{aligned}$$

This implies that

$$\sup |\tilde{p}(s, t)_m| |m|^{\eta r - \tau - 9} \leq \kappa(|s| + |t|)^{dr} \|Lp\|_{r,N},$$

which (by making a substitution  $r := \eta r - \tau - 9$ ) gives:

$$\sup_m \{ |\tilde{p}(s, t)_m| |m|^r \} \leq \kappa(|s| + |t|)^{d(\eta^{-1}r + \eta^{-1}(\tau+9))} \|Lp\|_{\eta^{-1}r + \eta^{-1}(\tau+9), N}.$$

Because of the well known norm comparison:  $\|\tilde{p}(s, t)\|_r \leq C \sup_m \{ |\tilde{p}(s, t)_m| |m|^{r+d+2} \}$ , we have:

$$\|\tilde{p}(s, t)\|_r \leq \kappa(|s| + |t|)^{d(\eta^{-1}(r+d+2) + \eta^{-1}(\tau+9))} \|Lp\|_{\eta^{-1}(r+d+2) + \eta^{-1}(\tau+9), N}.$$

Now let  $\mu := \eta^{-1}$  (recall that  $\eta < 1$ ), define the new  $\sigma := \eta^{-1}(\tau + 9 + d + 2)$  to obtain the final estimate  $\|\tilde{p}(s, t)\|_r \leq \kappa(|s| + |t|)^{d(\mu r + \sigma)} \|Lp\|_{\mu r + \sigma, N}$ .  $\square$

In the rest of this section we prove Proposition 20. We will split the proof into three subsections according to  $m \in \mathcal{C}_1$ ,  $m \in \mathcal{C}_2$  or  $m \in \mathcal{C}_3$ .

**4.1.2. Proof of Proposition 20 in the case  $m \in \mathcal{C}_1$ .** Let  $m \in \mathcal{C}_1$ , i.e., we have  $\bar{A}m = \bar{B}m = m$ . Since the action is assumed to be  $(\gamma, \tau)$ -Diophantine, we have either  $|e(m, \alpha) - 1| \geq \gamma \|m\|^{-\tau}$ , or  $|e(m, \beta) - 1| \geq \gamma \|m\|^{-\tau}$ . Let  $f = p(1, 0)$  and  $g = p(0, 1)$ .

Define  $h_m$  as follows:

$$h_m := \begin{cases} (e(\alpha, m) - 1)^{-1} f_m, & \text{if } |1 - e(\alpha, m)| \geq \gamma \|m\|^{-\tau}, \\ (e(\beta, m) - 1)^{-1} g_m, & \text{if } |1 - e(\alpha, m)| < \gamma \|m\|^{-\tau}. \end{cases} \quad (39)$$

Then we have the following

**Proposition 21.** *Let  $\langle a, b \rangle$  be a  $(\gamma, \tau)$ -Diophantine parabolic affine action, and let a map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$ , be given. For  $m \in \mathcal{C}_1$ , let  $h_m$  be defined as in (39). Then*

$$|h_m| \leq \gamma \max\{\|f\|_r, \|g\|_r\} |m|^{-r+\tau},$$

and for any  $(s, t) \in \mathbb{Z}^2$ , the number  $(\tilde{p}(s, t))_m$  defined by formula (37), which in this case has the form:

$$(\tilde{p}(s, t))_m = h_m(e(m, \alpha_{s,t}) - 1) - (p(s, t))_m,$$

satisfies

$$|(\tilde{p}(s, t))_m| \leq c \max\{\|Lp((s, t), (1, 0))\|_r, \|Lp((s, t), (0, 1))\|_r\} |m|^{-r+\tau}.$$

**Proof.** Suppose first that  $m$  is such that  $|1 - e(\alpha, m)| \geq \gamma \|m\|^{-\tau}$ , in which case  $h_m = (e(\alpha, m) - 1)^{-1} f_m$ . Then

$$\begin{aligned} (Lp((s, t), (1, 0)))_m &= (\partial_{s,t} f)_m - (\partial_{1,0} p(s, t))_m \\ &= (e(\alpha_{s,t}, m) - 1) f_m - (e(\alpha, m) - 1) (p(s, t))_m = \\ &= (e(\alpha_{s,t}, m) - 1) h_m (e(\alpha, m) - 1) - (e(\alpha, m) - 1) (p(s, t))_m = \\ &= (e(\alpha, m) - 1) (e(\alpha_{s,t}, m) - 1) h_m - (p(s, t))_m \\ &= (e(\alpha, m) - 1) (\tilde{p}(s, t))_m. \end{aligned}$$

Estimate  $|1 - e(\alpha, m)| \geq \gamma \|m\|^{-\tau}$  implies the result. The case when  $h_m = (e(\beta, m) - 1)^{-1} g_m$  is treated in the same way.

Directly from the definition of  $h_m$  and from the SDC-condition on  $\alpha$  and  $\beta$  we obtain the bound for  $|h_m|$ :

$$\begin{aligned} |h_m| &\leq \max\{|e(\beta, m) - 1|^{-1}, |e(\alpha, m) - 1|^{-1}\} \max\{|f_m|, |g_m|\} \\ &\leq \gamma |m|^\tau |m|^{-r} \max\{\|f\|_r, \|g\|_r\}. \end{aligned}$$

□

It is straightforward that Proposition 21 implies Proposition 20 in case  $m \in \mathcal{C}_1$ .

**4.1.3. Proof of Proposition 20 in the case  $m \in \mathcal{C}_2$ .** Let  $\langle a, b \rangle$  be an unlocked parabolic affine action, where  $a$  is step-2. Let  $m \in \mathcal{C}_2(k, l)$ , i.e., at least one of  $\bar{A}m$  and  $\bar{B}m$  is different from  $m$ , and there exists  $(k, l) \in \mathbb{Z}^2 \setminus \{0\}$  such that  $\bar{A}^k \bar{B}^l m = m$  (thus  $k\hat{A}m + l\hat{B}m = 0$ ). By Lemma 11,  $\mathcal{C}_2(k, l)$  can be divided into two sub-cases:

$\mathcal{C}'_2(k, l)$ :  $\bar{A}m \neq m$ ;  
 $\mathcal{C}''_2(k, l)$ :  $\bar{A}m = m$ , while  $\bar{B}m \neq m$ . Note that in this case, by Lemma 11, we have  $\bar{B}^2m = m$ .

The sets  $\mathcal{C}'_2(k, l)$  and  $\mathcal{C}''_2(k, l)$  are invariant under the action of  $\langle A, B \rangle$  due to the commutativity of the action. Consider a map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$ . Denote  $f = p(1, 0)$  and  $g = p(0, 1)$ .

For  $m \in \mathcal{C}'_2(k, l)$  we define:

$$h_m = \begin{cases} \Sigma_m^{+,A}(f), & m \in \mathcal{M}(A), \\ \Sigma_m^{-,A}(f), & m \in \mathcal{N}(A). \end{cases} \quad (40)$$

For  $m \in \mathcal{C}''_2(k, l)$  we define:

$$h_m = \begin{cases} \Sigma_m^{+,B}(g), & m \in \mathcal{M}(B), \\ \Sigma_m^{-,B}(g), & m \in \mathcal{N}(B). \end{cases} \quad (41)$$

To understand our choice for  $h_m$ , think of  $C^\infty$ -functions  $h$  and  $f$  satisfying  $h \circ a - h = f$ . Then the Fourier coefficients are related by

$$(h \circ a - h)_m = h_{\bar{A}m} \lambda_m^{(1)} - h_m = f_m.$$

Iterating this equality by  $\bar{A}$  either in the positive or in the negative direction while multiplying by appropriate constants, one obtains a telescopic sum equal to  $h_m$  as given in formula (40). The following proposition is the main statement of this section. It is straightforward that it implies Proposition 20 in case  $m \in \mathcal{C}_2$ .

**Proposition 22.** *Assume that  $\langle a, b \rangle$  is an unlocked  $(\gamma, \tau)$ -Diophantine parabolic affine action, where  $a$  is step-2. Consider a map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$ , denote  $f = p(1, 0)$  and  $g = p(0, 1)$ . Given  $m \in \mathcal{C}_2(k, l)$  for some  $(k, l) \in \mathbb{Z}^2 \setminus \{0\}$ , define  $h_m$  as above.*

*Then there exists  $\kappa = \kappa(\gamma, \tau, A, B) > 0$  such that*

$$|h_m| \leq \kappa \max\{\|f\|_r, \|g\|_r\} |m|^{-r+1},$$

*and for any  $(s, t) \in \mathbb{Z}^2$ , the number  $(\tilde{p}(s, t))_m$  defined by formula (37), i.e.,*

$$\tilde{p}(s, t)_m = h_{\bar{A}^s \bar{B}^t m} e(\bar{A}^s \bar{B}^t m, \alpha_{s,t}) - h_m - (p(s, t))_m,$$

*satisfies:*

$$|(\tilde{p}(s, t))_m| \leq \kappa(|s| + |t|)^{dr} \max\{\|Lp((1, 0), (k, l))\|_r, \|Lp((0, 1), (k, l))\|_r, \\ \|Lp((1, 0), (s, t))\|_r, \|Lp((0, 1), (s, t))\|_r\} |m|^{-r+\tau+2}. \quad (42)$$

*In addition,  $|k| + |l| \leq c|m|$  for a certain  $c = c(A, B) > 0$ .*

*Proof.* Let us present the proof of Proposition 22 modulo certain lemmas, that are proved below. First consider  $m \in \mathcal{C}'_2(k, l)$ ; the arguments for  $m \in \mathcal{C}''_2(k, l)$  are similar.

(i) Define  $h_m$  by (40). The estimate for  $|h_m|$  follows from Proposition B part (3).

(ii) We start by proving estimate (42) for  $(s, t) = (1, 0)$ . This is done in Lemma 23. Namely, we observe in that Lemma that the error term  $\tilde{f}_m = \tilde{p}(1, 0)_m = h_{\bar{A}m} \lambda_m^{(1)} - h_m - f_m$ , vanishes for all  $m$  except for those  $m$  that are lowest in norm on their  $\bar{A}$  orbit. In the latter case, we will derive from the commutation relation formula (43) that we repeat here

$$(e(m, \alpha_{k,l}) - 1) \tilde{f}_m = \Sigma_m^A(Lp((1, 0), (k, l))).$$

After that, the right-hand side is bounded above by the norm of  $Lp((1, 0), (k, l))$  (because  $m$  is lowest on its orbit), while the term  $|e(m, \alpha_{k,l}) - 1|$  is bounded below by the Diophantine condition for the resonances. It is here that the Diophantine conditions on the resonances play a crucial role: this condition, combined with formula (43), implies:

$$|\tilde{f}_m| \leq c \|Lp((1, 0), (k, l))\|_r |m|^{-r+1+\tau}.$$

(iii) Use step (ii) above to prove estimate (42) for all  $(s, t)$ . Lemma 24 derives the estimates on  $(\tilde{p}(s, t))_m$  for any  $(s, t)$  from those on  $\tilde{f}_m = (\tilde{p}(1, 0))_m$  (or on  $\tilde{g}_m = (\tilde{p}(0, 1))_m$ , which will be relevant for  $m \in \mathcal{C}''_2(k, l)$ ). We use Lemma 24 with  $\mathcal{K} = c \|Lp((1, 0), (k, l))\|_r$  and  $\rho = -r + 1 + \tau$  to get (42) for all  $(s, t)$ .

The arguments for  $m \in \mathcal{C}''_2(k, l)$  are similar: define  $h_m$  by (41) and estimate  $|h_m|$  with the help of part (3) of Proposition B; estimate  $|\tilde{g}_m| := |(\tilde{p}(0, 1))_m|$  via  $\|Lp((0, 1), (k, l))\|_r$  with the help of Lemma 23, and use it instead of  $|\tilde{f}_m|$  to get formula (42) for all  $(s, t)$ .  $\square$

The following lemma provides the proof for item (ii) above. Below we write  $m = \bar{m}$  to say that  $m$  is the lowest (in norm) point on its  $\bar{A}$  orbit, and  $m \neq \bar{m}$  otherwise.

**Lemma 23.** *Assume that  $\langle a, b \rangle$  is an unlocked  $(\gamma, \tau)$ -Diophantine parabolic affine action, where  $a$  is step-2. Let  $m \in \mathcal{C}_2(k, l)$ . Denote  $f = p(1, 0)$ ,  $\tilde{f}_m = (\tilde{p}(1, 0))_m$ ,  $g = p(0, 1)$ ,  $\tilde{g}_m = (\tilde{p}(0, 1))_m$ , and let  $h_m$  be as in (40), (41). Then there exists a constant  $c = c(r, A, B) > 0$  such that for  $m = \bar{m}$  we have:*

$$\begin{cases} |\tilde{f}_m| \leq c \|Lp((1, 0), (k, l))\|_r |m|^{-r+1+\tau} & \text{if } m \in \mathcal{C}'_2(k, l), \\ |\tilde{g}_m| \leq c \|Lp((0, 1), (k, l))\|_r |m|^{-r+1+\tau} & \text{if } m \in \mathcal{C}''_2(k, l). \end{cases}$$

For  $m \neq \bar{m}$  we have  $\tilde{f}_m = \tilde{g}_m = 0$ .

*Proof.* Assume that  $m \in \mathcal{C}'_2(k, l)$ , the case  $m \in \mathcal{C}''_2(k, l)$  being similar. By the definition of  $h_m$ , we have:  $\tilde{f}_m = h_{\bar{A}m} \lambda_m^{(1)} - h_m - f_m$ , and formally we can express:

$$\tilde{f}_m = \begin{cases} \Sigma_m^A(f), & m = \bar{m}, \\ 0, & \text{otherwise.} \end{cases}$$

Assume that  $m = \bar{m}$ . Using the definition of  $h_m$ , we get:

$$\begin{aligned} (Lp((1, 0), (k, l)))_m &= (\partial_{1,0} p(k, l))_m - (\partial_{k,l} f)_m \\ &= (\partial_{1,0} p(k, l))_m - (e(m, \alpha_{k,l}) - 1) f_m. \end{aligned}$$

Note that, by the commutativity of  $a^k b^l$  and  $a$ , we get the relation:  $A\alpha_{k,l} = \alpha_{k,l}$ . Therefore, the term  $(e(m, \alpha_{k,l}) - 1)$  is not changed when  $m$  moves along the  $\bar{A}$ -orbit.

Take the weighted sum  $\Sigma_m^A$  on both sides and recall that for  $m = \bar{m}$  we have  $\Sigma_m^A(f) = \tilde{f}_m$ :

$$\Sigma_m^A(Lp((1, 0), (k, l))) = (e(m, \alpha_{k,l}) - 1) \Sigma_m^A(f) = (e(m, \alpha_{k,l}) - 1) \tilde{f}_m. \quad (43)$$

Since  $Lp((1, 0), (s, t)) \in C^r$ , we can use Proposition B, which gives us:

$$|\Sigma_m^A(Lp((1, 0), (k, l)))| \leq \|Lp((1, 0), (k, l))\|_r |m|^{-r+1}.$$

Using the Diophantine assumption on resonances, we conclude:

$$|\tilde{f}_m| \leq \|Lp((1, 0), (k, l))\|_r |m|^{-r+1} |e(m, \alpha) - 1|^{-1} \leq \|Lp((1, 0), (k, l))\|_r |m|^{-r+1+\tau}.$$

□

The following lemma derives the estimates on  $(\tilde{p}(s, t))_m$  for any  $(s, t)$  from those on  $\tilde{f}_m = (\tilde{p}(1, 0))_m$  or  $\tilde{g}_m = (\tilde{p}(0, 1))_m$ , providing the details for item (iii).

**Lemma 24.** *Assume that  $\langle a, b \rangle$  is an unlocked parabolic affine action, and  $a$  is step-2. Consider a map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$ , and let  $\{h_m\}_{m \in \mathcal{U}}$  be given, where  $\mathcal{U}$  is some  $\langle \bar{A}, \bar{B} \rangle$ -invariant set. Define for every  $(s, t) \in \mathbb{Z}^2$ ,  $\{(\tilde{p}(s, t))_m\}_{m \in \mathbb{Z}}$  via*

$$(\tilde{p}(s, t))_m = h_{\bar{A}^s \bar{B}^t m} e(\bar{A}^s \bar{B}^t m, \alpha_{s,t}) - h_m - (p(s, t))_m.$$

Suppose that there exists  $\mathcal{K} > 0$  and  $0 < \rho \leq r$ , such that we have

$$\text{either } |(\tilde{p}(1, 0))_m| \leq \mathcal{K} |m|^{-\rho} \text{ for all } m \in \mathcal{U}, \quad \text{or } |(\tilde{p}(0, 1))_m| \leq \mathcal{K} |m|^{-\rho} \text{ for all } m \in \mathcal{U}.$$

Then there exists a constant  $c = c(A, B) > 0$  such that for any  $(s, t) \in \mathbb{Z}^2$  we have:

$$|(\tilde{p}(s, t))_m| \leq c(|s| + |t|)^{dr} (\max\{\|Lp((s, t), (1, 0))\|_r, \|Lp((s, t), (0, 1))\|_r\} + \mathcal{K}) |m|^{-\rho+1}. \quad (44)$$

*Proof.* Suppose first that for all  $m$  in an  $\bar{A}$ -invariant set  $\mathcal{U}$  we have  $|(\tilde{p}(1, 0))_m| \leq \mathcal{K} |m|^{-\rho}$ . Denote  $f_m := (p(1, 0))_m$  and  $\tilde{f}_m := (\tilde{p}(1, 0))_m$  for brevity.

For the sequence  $\{h_m\}$  we formally have:

$$\partial_{s,t}(\partial_{k,l} h)_m = \partial_{k,l}(\partial_{s,t} h)_m.$$

To see this notice that any smooth function  $H$  one can verify, using only the commutativity relation  $ab = ba$ , that  $\partial_{s,t} \partial_{k,l} H = \partial_{k,l} \partial_{s,t} H$ . Hence, under the commutativity condition, the Fourier coefficients of a smooth function  $H$  satisfy for each  $m$ :  $\partial_{s,t}(\partial_{k,l} H)_m = \partial_{k,l}(\partial_{s,t} H)_m$ . This implies the desired relation for the sequence of numbers  $\{h_m\}$  (this relation can be also verified directly).

By the definition of the set of numbers  $(h_m)$ ,

$$\begin{aligned} (Lp((1, 0), (s, t)))_m &= (\partial_{1,0} p(s, t))_m - (\partial_{s,t} f)_m \\ &= (\partial_{1,0} p(s, t))_m - (\partial_{s,t} \partial_{1,0} h - \partial_{s,t} \tilde{f})_m \\ &= (\partial_{1,0} (p(s, t) - \partial_{s,t} h))_m + (\partial_{s,t} \tilde{f})_m \\ &= (\partial_{1,0} \tilde{p}(s, t))_m + (\partial_{s,t} \tilde{f})_m. \end{aligned}$$

Hence,

$$(\partial_{1,0} \tilde{p}(s, t))_m = (Lp((1, 0), (s, t)))_m - (\partial_{s,t} \tilde{f})_m.$$

Since  $Lp((1, 0), (s, t)) \in C^r$  and  $\rho \leq r$ , we have

$$|(Lp((1, 0), (s, t)))_m| \leq \|Lp((1, 0), (s, t))\|_r |m|^{-\rho}.$$

Let us estimate  $(\partial_{s,t}\tilde{f})_m = \tilde{f}_{\bar{A}^s\bar{B}^t m} - \tilde{f}_m$ . To bound  $|\tilde{f}_{\bar{A}^s\bar{B}^t m}|$  notice that, since the linear part of the action  $\langle a, b \rangle$  is parabolic, for some constant  $c_0 = c_0(A, B)$  we have:

$$\|\bar{A}^{-s}\bar{B}^{-t}\| \leq c_0(|s| + |t|)^d.$$

Hence, for any  $(s, t)$  we have:

$$|m| \leq \|\bar{A}^{-s}\bar{B}^{-t}\| |\bar{A}^s\bar{B}^t m| \leq c_0(|s| + |t|)^d |\bar{A}^s\bar{B}^t m|,$$

and thus  $|\bar{A}^s\bar{B}^t m|^{-\rho} \leq c_1(|s| + |t|)^{dr} |m|^{-\rho}$ , and therefore

$$\begin{aligned} |(\partial_{s,t}\tilde{f})_m| &\leq |\tilde{f}_{\bar{A}^s\bar{B}^t m}| + |\tilde{f}_m| \leq \mathcal{K}(|\bar{A}^s\bar{B}^t m|^{-\rho} + |m|^{-\rho}) \\ &\leq c_2 \mathcal{K}(|s| + |t|)^{dr} |m|^{-\rho} \end{aligned}$$

for some  $c_1, c_2 > 0$  only depending on  $(A, B)$ . Finally, we obtain:

$$\begin{aligned} |(\partial_{1,0}\tilde{p}(s, t))_m| &\leq |(Lp((1, 0), (s, t)))_m + (\partial_{s,t}\tilde{f})_m| \\ &\leq \left( \|Lp((1, 0), (s, t))\|_r + c_2 \mathcal{K}(|s| + |t|)^{dr} \right) |m|^{-\rho}. \end{aligned}$$

By Proposition B (2), this implies that

$$\begin{aligned} |(\tilde{p}(s, t))_m| &\leq c_3 \left( \|Lp((1, 0), (s, t))\|_r + c_1 \mathcal{K}(|s| + |t|)^{dr} \right) |m|^{-\rho+1} \\ &\leq c(|s| + |t|)^{dr} (\|Lp((1, 0), (s, t))\|_r + \mathcal{K}) |m|^{-\rho+1} \end{aligned}$$

for some  $c = c(A, B) > 0$ . The case  $|(\tilde{p}(0, 1))_m| \leq \mathcal{K}|m|^{-\rho}$  is similar, we just have to use  $\partial_{0,1}$ ,  $g_m := (p(0, 1))_m$  and  $\tilde{g}_m := (\tilde{p}(0, 1))_m$  instead of  $\partial_{1,0}$ ,  $f_m$  and  $\tilde{f}_m$ , respectively.  $\square$

**4.1.4. Proof of Proposition 20 in the case  $m \in \mathcal{C}_3$ .** Let  $m \in \mathcal{C}_3$ ; denote  $f = p(0, 1)$ . We define  $h_m$  by

$$h_m = \begin{cases} \Sigma_m^{+,A}(f), & m \in \mathcal{M}(A), \\ \Sigma_m^{-,A}(f), & m \in \mathcal{N}(A). \end{cases} \quad (45)$$

**Proposition 25.** Assume that  $\langle a, b \rangle$  is unlocked step- $S$  parabolic affine action, where  $a$  is step-2. Let a map  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$  be given. For  $m \in \mathcal{C}_3$ , define  $h_m$  by (45).

There exists a constant  $\kappa = \kappa(\gamma, r, A, B)$  such that

$$|h_m| \leq \kappa \|f\|_r |m|^{-r+1}, \quad (46)$$

and, defining for each  $(s, t)$  the number  $(\tilde{p}(s, t))_m$  as in formula (37), i.e.,

$$(\tilde{p}(s, t))_m = h_{\bar{A}^s\bar{B}^t m} e(\bar{A}^s\bar{B}^t m, \alpha_{s,t}) - h_m - (p(s, t))_m,$$

for  $\eta = 0.99/S$ , for any  $r > 8/\eta$ , we have the estimate:

$$|\tilde{p}(s, t)_m| \leq \kappa(|s| + |t|)^{dr} \max\{\|Lp((1, 0), (0, 1))\|_r, \|Lp((s, t), (1, 0))\|_r\} |m|^{-\eta r+9}. \quad (47)$$

*Proof.* The proof of this proposition is done in 3 steps similarly to that of Proposition 22. Steps (i) and (iii) of the proof rely on the same lemmas (applied with slightly different constants). The important difference lies in the proof of step (ii) that relies on Proposition C on the control of the double sums along the dual orbit of a lowest point on an  $\bar{A}$ -orbit. Here are the steps:

- (i) Estimate (46) follows from part (3) of Proposition B.
- (ii) Based on Proposition C, we will show the following:



**Lemma 26.** *Assume that  $\langle a, b \rangle$  is unlocked step- $S$  parabolic affine action, where  $a$  is step-2, and let  $m \in \mathcal{C}_3$ . As before, denote  $f = p(1, 0)$ ,  $\tilde{f}_m = (\tilde{p}(1, 0))_m$ , and let  $h_m$  be as in (45). Then there exists a constant  $c = c(r, A, B) > 0$  such that for  $\eta = 0.99/S$ , if  $m = \bar{m}$  (i.e.,  $m$  is the lowest point on its  $\bar{A}$  orbit), we have:*

$$|\tilde{f}_m| \leq c \|Lp((1, 0), (0, 1))\|_r |m|^{-\eta r + 8}; \quad (48)$$

if  $m \neq \bar{m}$ , then  $\tilde{f}_m = 0$ .

(iii) Lemma 26, followed by Lemma 24 with  $\mathcal{K} = c \|Lp((1, 0), (0, 1))\|_r$  and  $\rho = -\eta r + 8$ , implies (47) for arbitrary  $(s, t)$ .  $\square$

Thus it only remains now to show Lemma 26.

*Proof of Lemma 26.* By the definition of  $h_m$  we have:

$$\tilde{f}_m = h_{\bar{A}m} e(\bar{A}m, \alpha) - h_m - f_m = \begin{cases} -\Sigma_m^A(f), & m = \bar{m}, \\ 0, & m \neq \bar{m}. \end{cases} \quad (49)$$

Assume that  $m = \bar{m}$ . Denote  $\phi := Lp((1, 0), (0, 1))$  for brevity. Note that for any  $m \in \mathcal{C}_3$  the definition of  $\phi$  implies directly that the following holds in terms of the formal power series:

$$\Sigma_m^A(f) = \sum_{l \geq 0} \sum_{k \in \mathbb{Z}} \phi_{\bar{A}^k \bar{B}^l m} \lambda_m^{(k)} \mu_m^{(l)} = - \sum_{l \leq -1} \sum_{k \in \mathbb{Z}} \phi_{\bar{A}^k \bar{B}^l m} \lambda_m^{(k)} \mu_m^{(l)}. \quad (50)$$

Since  $\phi \in C^r$ , for each  $m$ , its  $m$ -th Fourier coefficient satisfies  $|\phi_m| \leq \|\phi\|_r |m|^{-r}$ . Hence,

$$|\Sigma_m^A(f)| \leq \|\phi\|_r \sum_{l \geq 0} \sum_{k \in \mathbb{Z}} |\bar{A}^k \bar{B}^l m|^{-r} = \|\phi\|_r \sum_{l \leq -1} \sum_{k \in \mathbb{Z}} |\bar{A}^k \bar{B}^l m|^{-r}.$$

The desired estimate (48) in this case follows directly from the estimate of the above double sums, namely it is proved in Proposition C that for  $r$  sufficiently large there exists a constant  $c = c(r, A, B) > 0$  such that at least one of the following holds:

$$\sum_{k \in \mathbb{Z}} \sum_{l \geq 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq c |m|^{-\eta r + 8}, \quad \sum_{k \in \mathbb{Z}} \sum_{l < 0} |\bar{A}^k \bar{B}^l m|^{-r} \leq c |m|^{-\eta r + 8}.$$

$\square$

## 4.2. Application of Proposition D to truncated functions and vector fields.

The application of Proposition D to truncated functions is direct. Application to truncated vector fields requires an iteration process as a consequence of the fact that the linear part of the unperturbed action is parabolic. Similar iterative procedure has been used before in [D], [DK2].

**4.2.1. Truncated functions.** For a general smooth function  $v$  on  $\mathbb{T}^d$  (or a vector field) and for  $N \in \mathbb{N}$  the truncation  $T_N v$  is obtained by cutting off the Fourier series of  $v$  with index  $\|n\| \geq N$ . The residue operator is defined as  $R_N := Id - T_N$ .

For a map  $p : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$ , and  $N \in \mathbb{N}$ , define the truncation  $T_N p$  by  $(T_N p)(s, t) = T_N p(s, t)$ , and  $R_N p := p - T_N p$ .

Then the operators  $T_N$  and the residue operators  $R_N := Id - T_N$  satisfy the following estimates for all  $N \in \mathbb{N}$  and all  $0 < r \leq r'$ :

$$\begin{aligned} \|T_N v\|_{r'} &\leq C_{r,r'} N^{r'-r+d} \|v\|_r, \\ \|R_N v\|_r &\leq C_{r,r'} N^{r-r'+d} \|v\|_{r'}. \end{aligned} \quad (51)$$

Observe that  $\text{ave}(T_N v) = \text{ave}(v)$ .

The following statement is a direct application of our main technical result, Proposition [D](#), to the truncations.

**Proposition E.** *Let  $\langle a, b \rangle$  be an unlocked  $(\gamma, \tau)$ -Diophantine parabolic step- $S$  affine action, where  $a$  is step-2, and let  $r > 0$ . There exist constants  $\mu = \mu(S) > 0$ ,  $\sigma = \sigma(\langle a, b \rangle)$  and  $\kappa = \kappa(r, \langle a, b \rangle)$  such that for any  $p : \mathbb{Z}^2 \mapsto C^\infty(\mathbb{T}^d)$  there exists  $V : \mathbb{Z}^2 \rightarrow \mathbb{R}$  such that for every fixed  $N \in \mathbb{N}$  and the truncation  $q = T_N p$ , there exist  $h \in C^\infty(\mathbb{T}^d)$  and  $\tilde{q} : \mathbb{Z}^2 \rightarrow C^\infty(\mathbb{T}^d)$  satisfying*

$$q(s, t) = \partial_{s,t} h + \tilde{q}(s, t) + V(s, t), \quad (s, t) \in \mathbb{Z}^2,$$

and the following estimates hold:

$$\begin{aligned} \|h\|_r &\leq \kappa \|q\|_{r+\sigma}, \\ \|\tilde{q}(s, t)\|_r &\leq \|q(s, t)\|_r + \kappa(|s| + |t|)^{rd} \|q\|_{r+\sigma}, \\ \|\tilde{q}(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} \|Lq\|_{\mu r + \sigma, N}, \\ V(s, t) &= \text{ave}(p(s, t)) = \text{ave}(q(s, t)), \end{aligned} \quad (52)$$

where  $\bar{r} := \mu r + \sigma$ .

Note that in the above proposition,  $h$  and  $\tilde{q}$  depend on  $N$ .

**4.2.2. Truncated vector fields.** For a map  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$  and  $N \in \mathbb{N}$ , define the truncation  $T_N \mathbf{p}$  by  $(T_N \mathbf{p})(s, t) = T_N \mathbf{p}(s, t)$ , and  $R_N \mathbf{p} := \mathbf{p} - T_N \mathbf{p}$ . We have the same estimates for the operators  $T_N$  and  $R_N$  on vector fields as in [\(51\)](#).

For  $\mathbf{h} \in \text{Vect}^\infty(\mathbb{M})$  we define

$$D_{s,t} : \mathbf{h} \mapsto \mathbf{h} \circ \rho(s, t) - \rho_0(s, t) \mathbf{h},$$

where  $(s, t) \in \mathbb{Z}^2$ . The operator  $\mathbf{L}$  is then defined by

$$(\mathbf{L}\mathbf{p})((k, l), (s, t)) = D_{s,t} \mathbf{p}(k, l) - D_{k,l} \mathbf{p}(s, t) \quad (53)$$

for any  $(k, l), (s, t) \in \mathbb{Z}^2$ .

Recall that the set  $\mathcal{Q}_N$  is defined in [\(30\)](#). For  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$  and  $r \geq 0$  let

$$\begin{aligned} \|\mathbf{p}\|_r &:= \max\{\|\mathbf{p}(1, 0)\|_r, \|\mathbf{p}(0, 1)\|_r\}, \\ \|\mathbf{L}\mathbf{p}\|_{r,N} &:= \max\{\|\mathbf{L}\mathbf{p}((1, 0), (k, l))\|_r, \|\mathbf{L}\mathbf{p}((0, 1), (k, l))\|_r : (k, l) \in \mathcal{Q}_N\}. \end{aligned}$$

**Proposition F.** *Let  $\langle a, b \rangle$  be an unlocked step- $S$  parabolic  $(\gamma, \tau)$ -Diophantine affine action, where  $a$  is step-2, and let  $r > 0$ . There exist constants  $\mu = \mu(S) > 0$ ,  $\sigma = \sigma(\langle a, b \rangle)$  and  $\kappa = \kappa(r, \langle a, b \rangle)$  such that for any  $\mathbf{p} : \mathbb{Z}^2 \mapsto \text{Vect}^\infty(\mathbb{T}^d)$  there exists  $\mathbf{V} : \mathbb{Z}^2 \rightarrow \mathbb{R}^d$  such that for every fixed  $N \in \mathbb{N}$  and the truncation  $q = T_N p$ , there exist  $\mathbf{h} \in \text{Vect}^\infty(\mathbb{T}^d)$  and  $\tilde{\mathbf{q}} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$  satisfying*

$$\mathbf{q}(s, t) = D_{s,t} \mathbf{h} + \tilde{\mathbf{q}}(s, t) + \mathbf{V}(s, t), \quad (s, t) \in \mathbb{Z}^2,$$

and the following estimates hold

$$\begin{aligned}
\|\mathbf{h}\|_r &\leq \kappa \|\mathbf{q}\|_{r+\sigma}, \\
\|\tilde{\mathbf{q}}(s, t)\|_r &\leq \|\mathbf{q}(s, t)\|_r + \kappa(|s| + |t|)^{rd} \|\mathbf{q}\|_{r+\sigma}, \\
\|\tilde{\mathbf{q}}(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} \|\mathbf{L}\mathbf{q}\|_{\mu r + \sigma, N}, \\
\|\mathbf{V}(1, 0)\| &\leq \|\mathbf{p}(1, 0)\|_0; \quad \|\mathbf{V}(0, 1)\| \leq \|\mathbf{p}(0, 1)\|_0,
\end{aligned} \tag{54}$$

where  $\bar{r} := \mu r + \sigma$ .

Note that in the above Proposition,  $\mathbf{h}$  and  $\tilde{\mathbf{q}}$  depend on  $N$ .

*Proof.* Since  $A$  and  $B$  are commuting unipotent matrices, there exists a basis in  $\mathbb{R}^d$  in which both of them are upper triangular. We choose this basis to represent  $\mathbf{f}$ ,  $\mathbf{g}$  and  $\mathbf{h}$  in coordinate form. Also, we use this basis to define the norm of an arbitrary  $\mathbf{h}$ : if  $\mathbf{h} = (h_1, \dots, h_d)$  in the chosen coordinates, then we define  $\|\mathbf{h}\|_r = \max\{\|h_i\|_r : i = 1, \dots, d\}$ . For an element  $(s, t) \in \mathbb{Z}^2$ , let  $\mathcal{A}^{(s, t)}$  denote the matrix  $A^s B^t$  in the chosen basis. Notice that the matrix  $\mathcal{A}^{(s, t)}$  is upper triangular. Moreover, it has a polynomial growth of coefficients with respect to  $(s, t)$ , so we can assume that any element of the matrix  $\mathcal{A}^{(s, t)}$  has size at most  $C(|s| + |t|)^S$ , where  $C$  is a constant depending on  $A$  and  $B$ , and  $S$  is the step of the action.

Recall that in the beginning of this section, for  $(s, t) \in \mathbb{Z}^2$ , we defined the operator

$$D_{s, t} \mathbf{h} := \mathbf{h} \circ (a^s b^t) - A^s B^t \mathbf{h}.$$

In the chosen basis, this operator will have the form

$$D_{s, t} \mathbf{h} = \bar{\partial}_{s, t} \mathbf{h} + \mathcal{A}^{(s, t)} \mathbf{h}, \tag{55}$$

where  $\bar{\partial}_{s, t}$  denotes the "diagonal" operator acting on  $\mathbf{h}$  coordinatewise by the operator  $\partial_{s, t}$ :

$$D_{s, t} = \begin{pmatrix} \partial_{s, t} & a_{12}^{s, t} & a_{13}^{s, t} & \cdots & a_{1d}^{s, t} \\ 0 & \partial_{s, t} & a_{23}^{s, t} & \cdots & a_{2d}^{s, t} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \partial_{s, t} \end{pmatrix}.$$

This upper triangular form allows us to construct  $\mathbf{h}$  inductively, starting from its last coordinate, and then continuing upwards to the first coordinate. At each step of this inductive procedure, we use all the previously constructed coordinates of  $\mathbf{h}$ .

To keep track of the estimates during the induction, it is convenient to consider the operator  $L$  in the upper triangular form as well. Thus, for any  $\mathbf{q}$  we let  $\mathbf{q}(s, t) = (q_1(s, t), \dots, q_d(s, t))$  be the coordinates of the vector field  $\mathbf{q}(s, t)$  in the chosen basis, and introduce the following operator:

$$\mathbf{L}\mathbf{q}((k, l), (s, t)) = \begin{pmatrix} \partial_{k, l} & a_{12}^{k, l} & \cdots & a_{1d}^{k, l} \\ 0 & \partial_{k, l} & \cdots & a_{2d}^{k, l} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \partial_{k, l} \end{pmatrix} \begin{pmatrix} q_1(s, t) \\ q_2(s, t) \\ \cdots \\ q_d(s, t) \end{pmatrix} - \begin{pmatrix} \partial_{s, t} & a_{12}^{s, t} & \cdots & a_{1d}^{s, t} \\ 0 & \partial_{s, t} & \cdots & a_{2d}^{s, t} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \partial_{s, t} \end{pmatrix} \begin{pmatrix} q_1(k, l) \\ q_2(k, l) \\ \cdots \\ q_d(k, l) \end{pmatrix}. \tag{56}$$

Given  $\mathbf{q}(s, t)$ , we first take out all of its averages and call the vector of averages by

$$\mathbf{V}(s, t) = (\text{ave}(q_1(s, t)), \dots, \text{ave}(q_d(s, t))).$$

Now we can work with  $\mathbf{q}(s, t)$  assuming that its averages are 0. Then the last estimate of Proposition F follows directly from the definition of  $\mathbf{V}(1, 0)$  and  $\mathbf{V}(0, 1)$ .

Let us proceed with the inductive construction of  $\mathbf{h}$ . We start with the last coordinate  $q_d$ . By applying Proposition E to  $q_d$  we obtain  $h_d$  and  $\tilde{q}_d$  such that:

$$\begin{aligned} q_d(s, t) &= \partial_{s,t} h_d + \tilde{q}_d(s, t), \\ \|h_d\|_r &\leq \kappa \|q\|_{r+\sigma} \leq \|\mathbf{q}\|_{r+\sigma}, \\ \|\tilde{q}_d(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} \|Lq_d\|_{\mu r+\sigma, N} \leq \kappa(|s| + |t|)^{d\bar{r}} \|\mathbf{Lq}\|_{\mu r+\sigma, N}. \end{aligned} \quad (57)$$

Now we turn to constructing  $h_{d-1}$ . First, define  $h'_d(s, t) := a_{d-1,d}^{s,t} h_d$  for any  $(s, t) \in \mathbb{Z}^2$  and observe that

$$\|h'_d(s, t)\|_r \leq |a_{d-1,d}^{s,t}| \|h_d\|_r \leq \kappa(|s| + |t|)^d \|q_d\|_{r+\sigma}. \quad (58)$$

Since we have put (55) in the upper triangular form, we have to show the following.

**Lemma 27.** *There exist  $h_{d-1}$  and  $\tilde{q}_{d-1}$  such that*

$$\begin{aligned} (q_{d-1} - h'_d)(s, t) &= \partial_{s,t} h_{d-1} + \tilde{q}_{d-1}(s, t), \\ \|h_{d-1}\|_r &\leq \kappa \|\mathbf{q}\|_{r+2\sigma}, \\ \|\tilde{q}_{d-1}(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} N^{d_2} \|\mathbf{Lq}\|_{\mu' r+\sigma', N}, \end{aligned}$$

where  $d_2 := \max\{d, d\bar{r}\}$ ,  $\mu' := \mu^2$  and  $\sigma' := \mu\sigma + \sigma$ .

*Proof.* By substituting the approximation of  $q_d$  of (57) into the line  $d-1$  of the operator of (56), we get

$$\begin{aligned} \mathbf{Lq}((k, l), (s, t))_{d-1} &= \\ &= Lq_{d-1}((k, l), (s, t)) + a_{d-1,d}^{s,t} (\partial_{k,l} h_d + \tilde{q}_d(k, l)) - a_{d-1,d}^{k,l} (\partial_{s,t} h_d + \tilde{q}_d(s, t)) \\ &= L(q_{d-1} - h'_d)((k, l), (s, t)) + a_{d-1,d}^{s,t} \tilde{q}_d(k, l) - a_{d-1,d}^{k,l} \tilde{q}_d(s, t). \end{aligned}$$

From this equality and the estimates for  $\tilde{q}_d$  in (57), we obtain the following estimate for all  $(k, l), (s, t) \in \mathcal{Q}_N$ :

$$\begin{aligned} \|L(q_{d-1} - h'_d)((k, l), (s, t))\|_r &\leq \\ &\leq \|\mathbf{Lq}((k, l), (s, t))_{d-1}\|_r + |a_{d-1,d}^{s,t}| \|\tilde{q}_d(k, l)\|_r + |a_{d-1,d}^{k,l}| \|\tilde{q}_d(s, t)\|_r \\ &\leq \|\mathbf{Lq}((k, l), (s, t))_{d-1}\|_r + \kappa(|k| + |l|)^{d_2} (|s| + |t|)^{d_2} \|Lq_d\|_{\mu r+\sigma, N}, \end{aligned} \quad (59)$$

where  $d_2 = \max\{d, d\bar{r}\}$ . By taking the maximum of  $\|L(q_{d-1} - h'_d)((k, l), (s, t))\|_r$  for  $(k, l) \in \mathcal{Q}_N$  and  $(s, t) \in \{(1, 0), (0, 1)\}$ , we get from inequality (59):

$$\|L(q_{d-1} - h'_d)\|_{r, N} \leq \|\mathbf{Lq}_{d-1}\|_{r, N} + \kappa N^{d_2} \|Lq_d\|_{\mu r+\sigma, N}. \quad (60)$$

Now we apply Proposition E to  $q_{d-1} - h'_d$ , and find  $h_{d-1}$  such that

$$\begin{aligned} \|h_{d-1}\|_r &\leq \\ &\leq \kappa \|q_{d-1} - h'_d\|_{r+\sigma} = \kappa \max\{\|q_{d-1}(1, 0) - h'_d(1, 0)\|_{r+\sigma}, \|q_{d-1}(0, 1) - h'_d(0, 1)\|_{r+\sigma}\} \\ &\leq \kappa (\|q_{d-1}\|_{r+\sigma} + \|h'_d\|_{r+\sigma}) \leq \kappa (\|q_{d-1}\|_{r+\sigma} + \|q_d\|_{r+2\sigma}) \\ &\leq \kappa \|\mathbf{q}\|_{r+2\sigma}, \end{aligned}$$

where we used the estimate for  $\|h'_d\|_{r+\sigma}$  from (58).

For the new error in coordinate  $d-1$  we use the estimate obtained in Proposition E and (60) to obtain

$$\begin{aligned} \|\tilde{q}_{d-1}(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} \|L(q_{d-1} - h'_d)\|_{\mu r + \sigma, N} \\ &\leq \kappa(|s| + |t|)^{d\bar{r}} (\|\mathbf{L}\mathbf{q}_{d-1}\|_{\mu r + \sigma, N} + \kappa N^{d_2} \|Lq_d\|_{\mu(\mu r + \sigma) + \sigma, N}) \\ &\leq \kappa(|s| + |t|)^{d\bar{r}} N^{d_2} \|\mathbf{L}\mathbf{q}\|_{\mu' r + \sigma', N}, \end{aligned}$$

where  $\mu' := \mu^2$  and  $\sigma' := \mu\sigma + \sigma$ .  $\square$

The rest of the inductive construction goes along the same lines. Namely, assume that  $h_d, \tilde{q}_d, h_{d-1}, \tilde{q}_{d-1}, \dots, h_i, \tilde{q}_i$  have all been constructed, and assume that they satisfy the following estimates for every  $j = i, \dots, d$  and  $(s, t) \in \mathbb{Z}^2$ :

$$\begin{aligned} \|h_j\|_r &\leq \kappa \|\mathbf{q}\|_{r + (d-j+2)\sigma}, \\ \|\tilde{q}_j(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} N^{(j+1)d_2} \|\mathbf{L}\mathbf{q}\|_{\eta_{j+1}r + \sigma_{j+1}, N}, \end{aligned}$$

where  $\eta'_{j+1} := \eta \cdot \eta_j$ ,  $\sigma_{j+1} := \eta\sigma_j + \sigma$  and  $d^{(j+1)} := \max\{d, d^{(j)}\}$ .

In  $(i-1)$ -st equation in (56) we substitute all the  $q_j$  coordinates with  $q_j(s, t) = \partial_{s,t}h_j + \tilde{q}_j(s, t)$ . Then the  $(i-1)$ -st equation in (56) becomes:

$$\begin{aligned} \mathbf{L}\mathbf{q}((k, l), (s, t))_{i-1} &= Lq_{i-1}((k, l), (s, t)) + \sum_{j=i}^d a_{1,j}^{s,t} q_j(k, l) - \sum_{j=i}^d a_{j,d}^{k,l} q_j(s, t) \\ &= Lq_{i-1}((k, l), (s, t)) + \sum_{j=i}^d a_{1,j}^{s,t} (\partial_{k,l}h_j + \tilde{q}_j(k, l)) - \sum_{j=i}^d a_{j,d}^{k,l} (\partial_{s,t}h_j + \tilde{q}_j(s, t)) \\ &= L(q_{i-1} - h'_i)((k, l), (s, t)) + \sum_{j=i}^d a_{1,j}^{s,t} \tilde{q}_j(k, l) - \sum_{j=i}^d a_{j,d}^{k,l} \tilde{q}_j(s, t), \end{aligned}$$

where we defined  $h'_i(s, t) := \sum_{j=i}^d a_{j,d}^{s,t} h_j$ .

Now we do the same as in Lemma 27: we apply Proposition E to  $q_{i-1} - h'_i$  to obtain  $h_{i-1}$  and  $\tilde{q}_{i-1}$ , such that

$$q_{i-1} - h'_i = \partial_{s,t}h_{i-1} + \tilde{q}_{i-1}.$$

The same procedure as in Lemma 27 gives:

$$\begin{aligned} \|h_j\|_r &\leq \kappa \|\mathbf{q}\|_{r + (d-i+2)\sigma}, \\ \|\tilde{q}_{i-1}(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} N^{(i+1)d_2} \|\mathbf{L}\mathbf{q}\|_{\mu_{i+1}r + \sigma_{i+1}, N}, \end{aligned}$$

where  $\eta'_{i+1} := \mu \cdot \mu_i$ ,  $\sigma_{i+1} := \mu\sigma_i + \sigma$ .

After passing through all the  $d$  steps we redefine the constants  $\sigma$  and  $\mu$ . Namely, will have a multiplicative loss  $\mu := \mu^d$  and an additive loss  $\sigma := \max\{\mu^d\sigma + d \cdot \sigma, (d+2)\sigma\}$  of the number of the derivatives for the newly constructed  $\tilde{\mathbf{q}}$  with respect to  $\mathbf{L}\mathbf{q}$ . In other words, for every  $j = 1, \dots, d$ , we have with the newly constructed  $\sigma$  and  $\mu$  the following estimates:

$$\begin{aligned} \|h_j\|_r &\leq \kappa \|\mathbf{q}\|_{r + \sigma}, \\ \|\tilde{q}_j(s, t)\|_r &\leq \kappa(|s| + |t|)^{d\bar{r}} N^{D_1} \|\mathbf{L}\mathbf{q}\|_{\mu r + \sigma, N}, \end{aligned}$$

where  $D_1 = D_1(d, r)$ . By defining  $\mathbf{h} = (h_1, \dots, h_d)$  and  $\tilde{\mathbf{q}} = (\tilde{q}_1, \dots, \tilde{q}_d)$  we obtain the estimates claimed in the Proposition F for  $\mathbf{h}$  and  $\tilde{\mathbf{q}}(s, t)$ .

The "linear" estimate for  $\|\tilde{\mathbf{q}}(s, t)\|_r$  follows directly from the fact that by construction  $\mathbf{q}(s, t) = D_{s,t}\mathbf{h} + \tilde{\mathbf{q}}(s, t)$ , so the estimate for  $\mathbf{h}$  implies:

$$\begin{aligned} \|\tilde{\mathbf{q}}(s, t)\|_r &\leq \|D_{s,t}\mathbf{h}\|_r + \|\mathbf{q}(s, t)\|_r \leq (|s| + |t|)^{dr} \|\mathbf{h}\|_r + \|\mathbf{q}(s, t)\|_r \\ &\leq \kappa(|s| + |t|)^{dr} \|\mathbf{q}\|_{r+\sigma} + \|\mathbf{q}(s, t)\|_r. \end{aligned}$$

By denoting  $D = d(d+1)$ , we obtain the estimates claimed in the Proposition F.  $\square$

**4.3. Proof of Proposition A.** Recall that in our setup for a perturbation  $\langle F, G \rangle$  of the action  $\langle a, b \rangle$ , we define the map  $\mathbf{p} : \mathbb{Z}^2 \rightarrow \text{Vect}^\infty(\mathbb{T}^d)$  by  $\mathbf{p}(k, l) := F^k G^l - a^k b^l$ ,  $(k, l) \in \mathbb{Z}^2$ . With the notations  $\mathbf{p}(1, 0) = \mathbf{f}$  and  $\mathbf{p}(0, 1) = \mathbf{g}$ , the action  $\tilde{\rho}$  is generated by the maps  $a + \mathbf{f}$  and  $b + \mathbf{g}$ .

Let  $N$  be fixed, and let  $\mathbf{q}_N = T_N \mathbf{p}$  be the truncation of  $\mathbf{p}$ . First we define  $\mathbf{V}(k, l) = -\text{ave}(\mathbf{p}(k, l))$ . (With a little abuse of notation) denote  $\mathbf{V} = \mathbf{V}(1, 0) := -\text{ave}(\mathbf{f})$  and  $\mathbf{W} = \mathbf{V}(0, 1) = -\text{ave}(\mathbf{g})$ . It is then clear that the last estimate in (9) holds.

Now we apply Proposition F to  $\mathbf{p}$  and its truncation  $\mathbf{q}_N = T_N \mathbf{p}$  in order to obtain  $\mathbf{h}_N$ . Then as in Proposition F we define  $\tilde{\mathbf{q}}_N(k, l) := \mathbf{q}_N(k, l) - D_{k,l}\mathbf{h}_N + \mathbf{V}(k, l)$ , and so

$$\tilde{\mathbf{p}}_N(k, l) := \mathbf{p}(k, l) - D_{k,l}\mathbf{h}_N + \mathbf{V}(k, l) = \tilde{\mathbf{q}}_N + R_N \mathbf{p}. \quad (61)$$

Finally, we let

$$\tilde{\mathbf{f}}_N = \tilde{\mathbf{p}}_N(1, 0), \quad \tilde{\mathbf{g}}_N = \tilde{\mathbf{p}}_N(0, 1).$$

We begin by estimating  $\mathbf{p}(k, l) = (a + \mathbf{f})^k \circ (b + \mathbf{g})^l - a^k b^l$ .

**Lemma 28.** For  $(k, l) \in \mathbb{Z}^2$ ,

$$\|\mathbf{p}(k, l)\|_r \leq C_r(|k| + |l|)^{3d} \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\}. \quad (62)$$

*Proof.* Developing the expression for  $\mathbf{p}(k, l)$  as  $\mathbf{p}(k, l) = (a + \mathbf{f})^k \circ (b + \mathbf{g})^l - a^k b^l = a(a + \mathbf{f})^{k-1} \circ (b + \mathbf{g})^l + \mathbf{f} \circ (a + \mathbf{f})^{k-1} \circ (b + \mathbf{g})^l - a^k b^l$  and continuing inductively, one gets

$$\begin{aligned} \mathbf{p}(k, l) &= \sum_{j=0}^{k-1} s_j \mathbf{f} \circ (a + \mathbf{f})^j + \sum_{j=0}^{l-1} t_j \mathbf{g} \circ (a + \mathbf{f})^k \circ (b + \mathbf{g})^j \\ &= \sum_{j=0}^{k-1} s_j f \circ (a^j + \xi_j) + \sum_{j=0}^{l-1} t_j g \circ (a^k b^j + \eta_j). \end{aligned}$$

Here  $s_j, t_j$  are appropriate compositions of type  $A^i B^j$  with  $i \leq k, j \leq l$ . The sums contain  $|k| + |l|$  terms, each coefficient can be estimated by  $\|A^k B^l\| \leq C(|k| + |l|)^S$ . The  $r$ -norm of each of these terms can be estimated by  $c\|A^k B^l\|^r \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\} \leq cr(|k| + |l|)^S \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\}$ . Hence,  $\|\mathbf{p}(k, l)\|_r \leq C_r(|k| + |l|)(|k| + |l|)^{2S} \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\} \leq C_r(|k| + |l|)^{3S} \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\}$ . Since  $S \leq d$ , we can then bound this from above by  $C_r(|k| + |l|)^{3d} \max\{\|\mathbf{f}\|_r, \|\mathbf{g}\|_r\}$ .  $\square$

Let us now move to proving that the main estimates (9) in of Proposition A hold for  $\mathbf{h}_N, \tilde{\mathbf{f}}_N$  and  $\tilde{\mathbf{g}}_N$ . We start with  $\mathbf{h}_N$ . Directly from the first estimate in (54) of Proposition F, and truncation estimate (51) it follows that:

$$\|\mathbf{h}_N\|_r \leq \kappa \|\mathbf{q}_N\|_{r+\sigma} \leq \kappa N^\sigma \|\mathbf{p}\|_r.$$

The "linear" estimate for  $\tilde{\mathbf{f}}_N$  and  $\tilde{\mathbf{g}}_N$  (the third estimate in Proposition A) follows directly from the estimate for  $\mathbf{h}_N$ :

$$\|\tilde{\mathbf{f}}_N\|_r = \|\tilde{\mathbf{p}}_N(1, 0)\|_r \leq \|\mathbf{p}_N(1, 0) - D_{1,0}\mathbf{h}_N + \mathbf{V}\|_r \leq C_r N^{D'} \Delta_r.$$

Estimate for  $\tilde{\mathbf{g}}_N = \tilde{\mathbf{p}}_N(0, 1)$  follows exactly in the same way.

Now we will use the fact that  $a^k b^l + \mathbf{p}(k, l)$  is a commutative action, in order to obtain the quadratic estimate for the error  $\tilde{\mathbf{q}}$ . This is done in two steps: the first one is to show that  $\|\mathbf{Lp}\|_{r,N}$  is quadratic by using the fact that  $a^k b^l + \mathbf{p}(k, l)$  is a commutative action. The second one is to compare  $\|\mathbf{Lq}\|_{r,N}$  to  $\|\mathbf{Lp}\|_{r,N}$ .

Recall that the operator  $\mathbf{L}$  acts on  $\mathbf{p}$  by the formula  $\mathbf{Lp}((k, l), (s, t)) = D_{s,t}\mathbf{p}(k, l) - D_{k,l}\mathbf{p}(s, t)$  for any  $(k, l), (s, t) \in \mathbb{Z}^2$ .

**Lemma 29.** *For any  $N \in \mathbb{N}$  and  $r \geq 0$  the following holds:*

$$\|\mathbf{Lp}\|_{r,N} \leq C_r N^{4d} \Delta_{r+1} \Delta_0.$$

*Proof.* First, we notice that the commutativity of the action  $a^k b^l + \mathbf{p}(k, l)$  implies that

$$(a^k b^l + \mathbf{p}(k, l)) \circ (a^s b^t + \mathbf{p}(s, t)) = (a^s b^t + \mathbf{p}(s, t)) \circ (a^k b^l + \mathbf{p}(k, l)).$$

Therefore, the operator  $\mathbf{L}$ , besides its linear form, has also a non-linear expression (on the right below), in particular:

$$\begin{aligned} D_{(s,t)}\mathbf{p}(k, l) - D_{k,l}\mathbf{p}(s, t) &= \mathbf{Lp}((k, l), (s, t)) \\ &= \mathbf{p}(k, l)(a^s b^t + \mathbf{p}(s, t)) - \mathbf{p}(k, l)(s, t) + \mathbf{p}(s, t)(a^k b^l + \mathbf{p}(k, l)) - \mathbf{p}(s, t)(a^k b^l). \end{aligned} \quad (63)$$

From the non-linear expression for  $\mathbf{Lp}((k, l), (s, t))$  and the classical estimates for compositions ([Hör] Theorem A.8) we obtain the "quadratic" estimate:

$$\begin{aligned} \|\mathbf{Lp}((k, l), (s, t))\|_r &= \|\mathbf{p}(k, l)(a^s b^t + \mathbf{p}(s, t)) - \mathbf{p}(k, l)(a^s b^t)\|_r \\ &\quad + \|\mathbf{p}(s, t)(a^k b^l + \mathbf{p}(k, l)) - \mathbf{p}(s, t)(a^k b^l)\|_r \\ &\leq C_r N^d (\|\mathbf{p}(k, l)\|_{r+1} \|\mathbf{p}(s, t)\|_0 + \|\mathbf{p}(k, l)\|_0 \|\mathbf{p}(s, t)\|_{r+1}) \\ &\leq C_r N^d (|k| + |l|)^{3d} (|s| + |t|)^{3d} \Delta_{r+1} \Delta_0. \end{aligned}$$

Recall that  $\|\mathbf{Lp}\|_{r,N} := \max\{\|\mathbf{Lp}((1, 0), (k, l))\|_r, \|\mathbf{Lp}((0, 1), (k, l))\|_r : (k, l) \in \mathcal{Q}_N\}$  and that for  $(k, l) \in \mathcal{Q}_N$  we have that  $C(|k| + |l|) \leq N$ . From the above inequality, by taking maximum over  $(k, l) \in \mathcal{Q}_N$ , we get the required estimate for  $\|\mathbf{Lp}\|_{r,N}$ .  $\square$

Next we will use the following basic estimate of the norm of the operator  $D_{k,l}$ :

**Lemma 30.** *For  $(k, l) \in \mathcal{Q}_N$  and  $r \geq 0$*

$$\|D_{k,l}\mathbf{p}(s, t)\|_r \leq C_r N^{dr} \|\mathbf{p}(s, t)\|_r.$$

*Proof.* To show this bound we only need to observe that for any affine map  $a : x \rightarrow Ax + \alpha$  and any function  $\phi$  the norm  $\|\phi \circ a\|_r$  is bounded above by  $\|A\|^r \|\phi\|_r$ . Since in the operator  $D_{k,l}$  we compose with  $a^k b^l$ , which has linear part  $A^k B^l$ , and the norms of matrices  $A^k B^l$  are (up to a positive constant) bounded by  $(|k| + |l|)^d$ , then  $(k, l) \in \mathcal{Q}_N$  directly implies the required estimate, because for  $(k, l) \in \mathcal{Q}_N$ ,  $C(|k| + |l|) \leq N$  by Lemma 4.  $\square$



**Lemma 31.** For any  $N \in \mathbb{N}$  and  $r' \geq r \geq 0$  the following holds:

$$\|\mathbf{Lq}\|_{r,N} \leq C_r N^{r+5d} \Delta_1 \Delta_0 + C_{r,r'} N^{r-r'+2d_2} \Delta_{r'},$$

where  $d_2 = d_2(r) := \max\{d, dr\}$ .

*Proof.* Because  $\mathbf{L}$  is a linear operator we have the following:

$$\mathbf{Lp}((k, l), (s, t)) = \mathbf{L}(T_N \mathbf{p})((k, l), (s, t)) + \mathbf{L}(R_N \mathbf{p})((k, l), (s, t)). \quad (64)$$

The term  $\mathbf{L}(R_N \mathbf{p})((k, l), (s, t))$  for  $(k, l) \in \mathcal{Q}_N$  and  $(s, t) \in \{(1, 0), (0, 1)\}$  is estimated by using the estimates on the truncation operators, (51), the linear form (64) of  $\mathbf{L}$  and Lemmas 30 and 28 (again, recall that for  $(k, l) \in \mathcal{Q}_N$  we have that  $C(|k| + |l|) \leq N$ ):

$$\begin{aligned} \|\mathbf{L}(R_N \mathbf{p})((k, l), (s, t))\|_r &\leq \|D_{s,t} R_N \mathbf{p}(k, l) - D_{k,l} R_N \mathbf{p}(s, t)\|_r \\ &\leq C_r (|s| + |t|)^{dr} \|R_N \mathbf{p}(k, l)\|_r + C_r (|k| + |l|)^{dr} \|R_N \mathbf{p}(s, t)\|_r \\ &\leq C_{r,r'} (N^{r-r'} \|\mathbf{p}(k, l)\|_{r'} + N^{dr} N^{r-r'} \|\mathbf{p}(s, t)\|_r) \\ &\leq C_{r,r'} (N^{r-r'+3d} \|\mathbf{p}\|_{r'} + N^{r-r'+dr} \|\mathbf{p}\|_{r'}) \\ &\leq C_{r,r'} N^{r-r'+d_3} \|\mathbf{p}\|_{r'}, \end{aligned}$$

where  $d_3 = d_3(r) := \max\{3d, dr\}$ .

Using this and (64), as well as (51) and Lemma 29, we obtain for  $0 \leq r'' \leq r \leq r'$  we have:

$$\begin{aligned} \|\mathbf{Lq}((k, l), (s, t))\|_r &\leq \|\mathbf{Lp}((k, l), (s, t))\|_r + \|\mathbf{L}(R_N \mathbf{p})((k, l), (s, t))\|_r \\ &\leq \|T_N \mathbf{Lp}((k, l), (s, t))\|_r + \|R_N \mathbf{Lp}((k, l), (s, t))\|_r \\ &\quad + \|\mathbf{L}(R_N \mathbf{p})((k, l), (s, t))\|_r \\ &\leq C_{r,r''} N^{r-r''+d} \|\mathbf{Lp}((k, l), (s, t))\|_{r''} + C_{r,r'} N^{r-r'+d} \|\mathbf{Lp}((k, l), (s, t))\|_{r'} \\ &\quad + C_{r,r'} N^{r-r'+d_3} \Delta_{r'} \\ &\leq C_{r,r''} N^{r-r''+5d} \Delta_{r''+1} \Delta_0 + C_{r,r'} N^{r-r'+d_3} \Delta_{r'}, \end{aligned}$$

where we applied the estimate  $C(|k| + |l|) \leq N$  that is valid for all  $(k, l) \in \mathcal{Q}_N$ . Now by setting  $r'' = 0$ , and by taking the maximum over  $(k, l) \in \mathcal{Q}_N$  and  $(s, t) \in \{(1, 0), (0, 1)\}$ , we obtain the required estimate for  $\|\mathbf{Lq}\|_{r,N}$ .  $\square$

From the third estimate in Proposition F and Lemma 31, for some fixed constant  $D > 0$ , for any  $r' > 0$  and for a fixed  $d_2 = d_2(\sigma)$ , we have:

$$\begin{aligned} \|\tilde{\mathbf{q}}_N(1, 0)\|_0 &\leq \kappa N^D \|\mathbf{Lq}_N\|_{\sigma,N} \\ &\leq \kappa N^{D+5d+\sigma} \Delta_1 \Delta_0 + C_{r'} N^{D-r'+\sigma+d_3} \Delta_{r'} \\ &\leq \kappa N^{D'} \Delta_1 \Delta_0 + C_{r'} N^{-r'+D'} \Delta_{r'}, \end{aligned}$$

where we define  $D' = D + 5d + d_3 + \sigma$ . Notice that  $d_3 = d_3(\sigma)$ , and consequently  $D'$ , depends only on the action  $\langle a, b \rangle$  and  $d$ . Recall that in (61) we defined  $\tilde{\mathbf{p}}_N = \tilde{\mathbf{q}}_N + R_N \mathbf{p}$  for every  $N$ . By combining the above estimate and the truncation estimates (51) for the operator  $R_N$ , we get a similar bound, with possibly new constants  $\kappa$  and  $C_{r'}$ :

$$\|\tilde{\mathbf{f}}_N\|_0 = \|\tilde{\mathbf{p}}_N(1, 0)\|_0 \leq \kappa N^{D'} \Delta_1 \Delta_0 + C_{r'} N^{-r'+D'} \Delta_{r'}.$$

Finally, we declare the new  $D$  to be the  $D'$ . Estimates for  $\tilde{\mathbf{g}}_N = \tilde{\mathbf{p}}_N(0, 1)$  are proved in exactly the same way. This completes the proof of all the estimates claimed in Proposition A.  $\square$

**ACKNOWLEDGMENT.** The authors are grateful to Livio Flaminio and Giovanni Forni for stimulating discussions on the subject. The first author was supported by the Swedish Research Council grant VR 2019-04641.

## REFERENCES

- [A] V. I. Arnol'd. Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations. Russian Math. Surveys, 18(5):9–36, 1963.
- [CG] J-P Conze, Y. Guivarc'h, Ergodicity of group actions and spectral gap, applications to random walks and Markov shifts, DCDS, 2013, Volume 33, Issue 9: 4239–4269.
- [CF] S. Cosentino, L. Flaminio, Equidistribution for higher-rank Abelian actions on Heisenberg nilmanifolds, JMD 2015, Volume 9: 305–353.
- [D] D. Damjanović, Abelian actions with globally hypoelliptic leafwise Laplacian and rigidity, Journal d'Analyse Mathématique, Volume 129, Issue 1, 139–163 (2016).
- [DF] D. Damjanović, B. Fayad, On Local Rigidity of Partially Hyperbolic Affine  $\mathbb{Z}^k$  Actions, J. Reine Angew. Math., 751 (2019), 1–26
- [DFS] D. Damjanović, B. Fayad, M. Saprykina, On the absence of tame solutions of the cohomological equation above parabolic transformations. *In preparation*.
- [DK] D. Damjanović, A. Katok, Local Rigidity of Partially Hyperbolic Actions of  $\mathbb{Z}^k$  and  $\mathbb{R}^k$ ,  $k \geq 2$ . I. KAM method and actions on the Torus, Annals of Mathematics, Vol. 172, No. 3, 1805–1858 (2010).
- [DK2] D. Damjanović, A. Katok, Local Rigidity of Parabolic Homogeneous Actions: I. A Model Case, Journal of Modern Dynamics, Vol. 5, no. 2, 203–235 (2011).
- [DK3] D. Damjanović, A. Katok, Local Rigidity of Partially Hyperbolic Actions. II: The Geometric Method and Restrictions of Weyl Chamber Flows on  $SL(n, \mathbb{R})/\Gamma$ , International Mathematics Research Notices, Volume 2011, Issue 19, 2011, Pages 4405–4430,
- [DT] D. Damjanović, J. Tanis, Transversal local rigidity of discrete abelian actions on Heisenberg nilmanifolds, Ergodic Theory and Dynamical Systems, (2021) 1–41.
- [FF] L. Flaminio, G. Forni, On the cohomological equation for nilflows, 2007, Volume 1, Issue 1: 37–60.
- [GK] M. Guysinsky and A. Katok, Normal forms and invariant geometric structures for dynamical systems with invariant contracting foliations, Math. Res. Letters, 5, (1998), 149–163.
- [H] M. R. Herman, Sur la conjugaison différentiable des difféomorphismes du cercle a des rotations, Inst. Hautes Études Sci. Publ. Math., 49 (1979), 5–233.
- [Hör] L. Hörmander, The boundary problems of physical geodesy, Archive for Rational Mechanics and Analysis 62 (1976), no. 1, 1–52.
- [KR] A. Katok, E. A. Robinson Jr, Cocycles, cohomology and combinatorial constructions in ergodic theory - 2001
- [KS] A. Katok, R. Spatzier, Trudy Matematicheskogo Instituta imeni V.A. Steklova, 1997, Volume 216, Pages 292–319.
- [L] V. F. Lazutkin, KAM Theory and Semiclassical Approximations to Eigenfunctions, Ergeb. Math. Grenzgeb. 24, Springer-Verlag, New York, 1993.
- [M] J. Moser, On Commuting Circle Mappings and Simultaneous Diophantine Approximations, Math. Z., 205:1 (1990), 105–121.
- [NT] V. Nitya, A. Török, Local rigidity of certain partially hyperbolic actions of product type, Ergodic Theory and Dynamical Systems 21 (2001), 1213–1237.
- [P] B. Petković, Classification of Perturbations of Diophantine  $\mathbb{Z}^m$  Actions on Tori of Arbitrary Dimension, Regul. Chaotic Dyn., 2021, Volume 26, Issue 6, Pages 700–716
- [P2] B. Petković, PhD Thesis, KTH 2022.

- [VW] K. Vinhage Z.J. Wang, Local rigidity of higher rank homogeneous abelian actions: a complete solution via the geometric method, *Geometriae Dedicata* volume 200, pages 385–439 (2019).
- [W1] Z. J. Wang. Local rigidity of partially hyperbolic actions. *J. Mod. Dyn.*, 4(2):271– 327, 2010.
- [W2] Z. J. Wang. New cases of differentiable rigidity for partially hyperbolic actions: symplectic groups and resonance directions. *J. Mod. Dyn.*, 4(4):585–608, 2010.
- [ZW] Z. J. Wang, Local rigidity of parabolic algebraic actions, arXiv:1908.09181.
- [W3] Z. J. Wang, Local rigidity of higher rank partially hyperbolic algebraic actions, arXiv:1103.3077
- [WX] A. Wilkinson and J. Xue, Rigidity of some abelian-by-cyclic solvable group actions on  $\mathbb{T}^N$ . *Commun. Math. Phys.* 376 (2020), 1223–1259.

(Danijela Damjanović) DEPARTMENT OF MATHEMATICS, KUNGLIGA TEKNISKA HÖGSKOLAN, LINDSTEDTSVÄGEN 25, SE-100 44 STOCKHOLM, SWEDEN

*Email address:* `ddam@kth.se`

(Bassam Fayad) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, 2307 KIRWAN HALL, COLLEGE PARK, USA

*Email address:* `bassam@umd.edu`

(Maria Saprykina) DEPARTMENT OF MATHEMATICS, KUNGLIGA TEKNISKA HÖGSKOLAN, LINDSTEDTSVÄGEN 25, SE-100 44 STOCKHOLM, SWEDEN

*Email address:* `masha@kth.se`