Properties of mixing BV vector fields

Stefano Bianchini, Martina Zizza

June 2, 2023

Abstract

We consider the density properties of divergence-free vector fields $b \in L^1([0,1], \mathrm{BV}([0,1]^2))$ which are ergodic/weakly mixing/strongly mixing: this means that their Regular Lagrangian Flow X_t is an ergodic/weakly mixing/strongly mixing measure preserving map when evaluated at $t = 1$. Our main result is that there exists a G_δ-set $\mathcal{U} \subset L^1_{t,x}([0,1]^3)$ containing all divergence-free vector fields such that

1. the map Φ associating b with its RLF X_t can be extended as a continuous function to the G_δ-set $\mathcal{U}$;
2. ergodic vector fields b are a residual G_δ-set in $\mathcal{U}$;
3. weakly mixing vector fields b are a residual G_δ-set in $\mathcal{U}$;
4. strongly mixing vector fields b are a first category set in $\mathcal{U}$;
5. exponentially (fast) mixing vector fields are a dense subset of $\mathcal{U}$.

The proof of these results is based on the density of BV vector fields such that $X_{t=1}$ is a permutation of subsquares, and suitable perturbations of this flow to achieve the desired ergodic/mixing behavior. These approximation results have an interest of their own.
A discussion on the extension of these results to $d \geq 3$ is also presented.

Preprint SISSA 19/2021/MATE[1]

Key words: ergodicity, mixing, Baire Category Theorem, divergence-free vector fields, Regular Lagrangian Flows, rates of mixing.

MSC2020: 26A21, 35Q35, 37A25.

Contents

[1]Data sharing not applicable to this article as no datasets were generated or analysed during the current study

1 Introduction

Consider a divergence-free vector field $b : \mathbb{R}^+ \times \mathbb{T}^d \to \mathbb{R}^d$ and the continuity equation

$$\partial_t \rho_t + D \cdot (\rho_t b_t) = 0, \quad \rho_{t=0} = \rho_0. \tag{1.1}$$

In recent year the following question has been addressed: is the solution ρ_t approaching weakly a constant as $t \to \infty$? The meaning of "approaching a constant" is usually formalized as

$$\rho_t \underset{t\to\infty}{\to} \fint \rho_0 \mathcal{L}^d \quad \text{weakly in } L^2, \tag{1.2}$$

($\mathcal{L}^d$ is the d-dimensional) since $\|\rho_t\|_{L^p}$ is constant (at least for positive solutions and sufficiently regular vector fields) and this is referred to as *functional mixing* (another notion of mixing is the *geometric mixing* introduced in [**Bressan˙conj**], but for our purposes the functional mixing above is the most suitable, since it is related to Ergodic Theory).
Without any functional constraint on the space derivative Db_t, it is quite easy to obtain mixing in finite time: a well known example is [**Depauw**]. A similar idea, used in a nonlinear setting, can be found in [**Bressan˙illposed**]. See also [**finite**] for completeness. The problem is usually formulated as follows: assume that $b \in L^\infty_{t,x} \cap L^\infty_t W^{s,p}_x$, what is the maximal speed of convergence in (1.2)?

This question has been addressed in several papers. In [**Alberti˙mix**] the 2d-case has been throughly analyzed, and the main results are the explicit construction of mixing vector fields when the initial data is fixed: the authors are able to achieve the optimal exponential mixing rate for the case $W^{1,p}$, $p > 1$, and study also the case $s < 1$ (mixing in finite time) and $s > 1$ (mixing at a polynomial rate). Recall that for $s = 1, p > 1$ the mixing is at most exponential [**Crippa˙DeLellis**], while the same estimate in $W^{1,1}$ (or equivalently BV) is still open [**Bressan˙conj**]. In [**YaoZlatos**, **univ:mixer**] it is discussed the existence of universal mixers: that is divergence-free vector fields that mix *any* initial data. In particular, in [**univ:mixer**] the authors construct a vector field which mixes at an exponential rate every initial data, and it belongs to $L^\infty_t W^{s,p}_x$ for $s < \frac{1+\sqrt{5}}{2}$, $p \in [1, \frac{2}{2s+1-\sqrt{5}})$. The autonomous 2d vector field is special, having an Hamiltonian structure: indeed in [**Marconi˙Bonicatto˙poly**] the authors show that the mixing is polynomial with rate t^{-1} when $b \in \mathrm{BV}$.

In this paper we consider the different problem: how many vector fields are mixing? More precisely, we study the mixing properties of flows generated in the unit square $K = [0,1]^2$ by divergence-free vector fields $b : [0,1] \times K \to \mathbb{R}^2$ belonging to the space $L^\infty([0,1], \mathrm{BV}(K))$: to avoid problems at the boundary, we assume that the vector field b is divergence-free and BV when extended to whole $\mathbb{R}^2$. In order to shorten the notation, we will sometimes write $\mathrm{BV}(K)$, $K = [0,1]^2$ as the space $\mathrm{BV}(\mathbb{R}^2) \cap \{\operatorname{supp} b \subset K\}$.
All the results stated here can be extended to the case $x \in \mathbb{T}^2$ with minor modifications; our choice is in the spirit of [**Shnirelman**].
In this setting, there exists a unique flow $t \to X_t \in C([0,1], L^1(K;K))$ (called Regular Lagrangian Flow (RLF)) of the ODE

$$\begin{cases} \frac{d}{dt} X_t(y) = b(t, X_t(y)), \\ X_{t=0}(y) = y, \end{cases}$$

which is measure-preserving and stable, see [**Ambrosio:BV**, **Ambrosio:Luminy**] and Section 2.2. Our idea is to consider the $\mathcal{L}^2$-a.e. invertible measure preserving map $X_{t=1} : K \to K$ as an *automorphism* of the measure space $(K, \mathcal{B}(K), \mathcal{L}^2 \llcorner_K)$ and apply the tools of Ergodic Theory. Here and in the following $\mathcal{L}^2 \llcorner_K$ is the Lebesgue measure on K and $\mathcal{B}(K)$ are the Borel subsets of K. We call $G(K)$ the group of automorphisms of K. We underline that the additional difficulty is to retain that the maps under consideration are generated by a divergence-free vector field in $L^\infty_t \mathrm{BV}_x$.

There is a rich literature in Ergodic Theory that has deeply investigated the genericity properties of mixing for invertible and measure-preserving maps. These results are due mostly to Oxtoby and Ulam [**Oxtoby**], Halmos [**Halmos:ergodic**, **Halmos:weak:mix**], Katok and Stepin [**Katok**] and Alpern [**Alpern**]. They proved that the set of ergodic transformation is a residual (or comeagre) G_δ-set (i) in the set $G(K)$ with the *neighbourhood topology*[2] [**Halmos:weak:mix**], (ii) in the set of measure-preserving homeomorphisms of a connected manifold with the strong topology[3] [**Oxtoby**, **Katok**]. Moreover, the transformations satisfying a stronger condition known as *weak mixing*, that is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\big]^2 = 0,$$

for every A, B measurable sets, are still a residual G_δ set [**Halmos:weak:mix**, **Katok**]. In 1976 Alpern showed that these problems are indeed connected by using the Annulus Theorem [**Alpern**]. A different result holds for strongly mixing maps, i.e., such that

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

It was shown firstly by Rokhlin in [**Rokhlin**] (see [**Weiss**] for an exposition of Rokhlin's work) and then by D. Ornstein [**Halmos:lectures**] that (strongly) mixing maps are a first category set in the neighborhood topology.

In these settings, the genericity properties of measure-preserving (weakly) mixing or ergodic maps are fairly understood; to our knowledge a similar analysis has not been done for flows generated by vector fields with additional regularity requirements (e.g. $b \in L^\infty_t \mathrm{BV}_x$). The aim of our work is to extend the above genericity results to divergence-free vector fields whose Regular Lagrangian Flow is ergodic and weakly mixing (in dimension $d = 2$, but see the discussion below on the extension to every dimension $d \geq 3$).
We remark that here we are looking to genericity properties of mixing in the topological sense, and not a.e. mixing w.r.t. some probability measure in the space of vector fields (e.g. [**Bedrossian**]): while there is some relation between the two notions, in general one result does not imply the other.

Let $b \in L^\infty([0,1], \mathrm{BV}(K))$ be a divergence-free vector field.

Definition 1.1. We say that b is *ergodic (weakly mixing, strongly mixing)* if its unique Regular Lagrangian Flow evaluated at time $t = 1$ is ergodic (respectively weakly mixing, strongly mixing).

In the original setting (1.1), if $b_{t+1} = b_t$ (i.e., it is time periodic of period 1), then it is fairly easy to see that if b is strongly mixing as in the above definition then (1.2) holds, while for weakly mixing vector fields b it holds the weaker limit

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \bigg(\int_K \Big(\rho_t - \fint \rho_0 \mathcal{L}^2\Big) \phi \mathcal{L}^2 \bigg)^2 dt = 0, \quad \forall \phi \in L^2(K).$$

Our main result is the following.

Theorem 1.2. *There exists a G_δ-subset $\mathcal{U} \subset L^1([0,1]; L^1(K)) \cap \{b : D \cdot b = 0\}$ containing all divergence-free vector fields in $L^\infty([0,1]; \mathrm{BV}(K))$ with the following properties:*

1. *the map Φ associating b with its RLF X_t,*

$$\Phi : \{b \in L^\infty_t \mathrm{BV}_x : D \cdot b_t = 0\} \to C([0,1], L^1(K)),$$

can be extended as a continuous function to the G_δ-set $\mathcal{U}$;

2. *ergodic vector fields b are a residual G_δ-set in $\mathcal{U}$;*

3. *weakly mixing vector fields b are a residual G_δ-set in $\mathcal{U}$;*

4. *strongly mixing vector fields b are a first category set in $\mathcal{U}$;*

5. *exponentially (fast) mixing vector fields are a dense subset of $\mathcal{U}$.*

[2]The neighbourhood topology is indeed the convergence in measure, see Subsection 3.1.
[3]A sequence of maps $T_n \to T$ in the strong topology if $T_n \to T$ and $T_n^{-1} \to T^{-1}$ uniformly on K.

We will reasonably call the flow $X_t = \Phi(b)$, $b \in \mathcal{U}$, as the Regular Lagrangian Flow of b, even if we are outside the setting where RLF are known to be unique: however X_t is the unique flow which can be approximated by RLF of smoother vector fields b^n as $b^n \to b$ in L^1. The existence of such a set $\mathcal{U}$ is due to purely topological properties of metric spaces (Proposition 2.3).

Our proof adapts some ideas from [**Halmos:weak:mix**] to our setting: we give an outline of Halmos' analysis. First of all, both ergodic automorphisms and weakly mixing automorphisms are a G_δ-set [**Halmos:ergodic**, **Halmos:weak:mix**]. Next, it is shown that the mixing properties are invariant under conjugation, i.e., if $T : [0,1] \to [0,1]$ is weakly/strongly mixing and $R : [0,1] \to [0,1]$ is an automorphism, then $R \circ T \circ R^{-1}$ is weakly/strongly mixing too. It remains to be proved that weakly mixing maps are dense: define a *permutation* as an automorphism of $[0,1]$ sending dyadic intervals (subintervals of $[0,1]$ with dyadic endpoints) into dyadic intervals by translation (in dimension greater than 1 the map translates dyadic subcubes). Cyclic permutations (i.e., permutations made by a unique cycle) of the same intervals are clearly conjugate. One of the key ingredients of Halmos' proof is that, for every non periodic automorphism (i.e., $T^n x \neq x$ for all n in a conegligible set of points x), there exists a cyclic permutation close to it in the neighbourhood topology, and by the previous observation about conjugation of permutations one deduces that if T is non-periodic then the maps of the form $R \circ T \circ R^{-1}$ form a dense set. In particular, the weakly mixing maps are a G_δ-set containing a non periodic map, hence this set is residual.

In our setting, the fact that ergodic/weakly mixing vector fields form a G_δ-subset of $\mathcal{U}$ is an easy consequence of the Stability Theorem for Regular Lagrangian Flows and the definition of the map Φ (see Point (1) and Proposition 3.9). Indeed, since both ergodic automorphisms and weakly mixing automorphisms are a G_δ-set [**Halmos:ergodic**, **Halmos:weak:mix**], then by the continuity of the map Φ associating b with the RLF X_t, ergodic and weakly mixing vector fields are a G_δ-set also. Unfortunately, we cannot use conjugation of a RLF X_t with an automorphism R of K, since in general $R \circ X_{t=1} \circ R^{-1}$ is not a RLF generated by $b \in L^\infty_t BV_x$ (or even $b \in \mathcal{U}$). However, we are able to prove the density in $\mathcal{U}$ of vector fields $b \in L^\infty_t \mathrm{BV}_x$ whose RLF is a cyclic permutation of subsquares of K, which is the natural extension of the permutation of intervals used in [**Halmos:weak:mix**]. More precisely, the map $T = X_{t=1}$ sends by a rigid translation subsquares of some rational grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$, where $D \in \mathbb{N}$, into subsquares of the same grid (it will be clear later that being dyadic as in [**Halmos:weak:mix**] is not relevant, see Lemma 5.1 and Remark 5.3), and as a permutation of subsquares it is made by a single cycle. The precise statement is the following.

Theorem 1.3. *Let $b \in L^\infty([0,1], \mathrm{BV}(K))$ be a divergence-free vector field. Then for every $\epsilon > 0$ there exist $1 \ll D \in \mathbb{N}$, two positive constants C_1, C_2 and a divergence-free vector field $b^c \in L^\infty([0,1], \mathrm{BV}(K))$ such that*

$$||b - b^c||_{L^1(L^1)} \leq \epsilon, \quad ||\mathrm{Tot.Var.}(b^c)(K)||_\infty \leq C_1 ||\mathrm{Tot.Var.}(b)(K)||_\infty + C_2 \tag{1.3}$$

and the map $X^c_{t=1} : K \to K$, where $X^c_t : [0,1] \times K \to K$ is the flow associated with b^c, is a D^2-cycle of subsquares of size $\frac{1}{D}$.

The above approximation is the most technical part of the paper, and it is the point which forces to state the theorem in $\mathcal{U}$ and not in the original space $b \in L^\infty_t \mathrm{BV}_x$: indeed, while achieving the density in the $L^1_{t,x}$-topology, the total variation increases because of the constants C_1, C_2 in (1.3). (It is possible to improve the first estimate of (1.3) to $\|b - b^c\|_{L^\infty L^1} \leq \epsilon$, see Remark 4.7, but to avoid additional technicalities we concentrate on the simplest results leading to Theorem 1.2.)

We remark that the above approximation result is sufficient to prove that strongly mixing vector fields are a set of first category (Proposition 3.9): indeed, Theorem 1.3 shows the density in $\mathcal{U}$ of divergence-free vector fields whose flow is made of periodic trajectories with the same period D^2. This observation is the key to obtain Point (4) of Theorem 1.2.

Looking at cyclic permutations of subsquares is an important step to obtain ergodic (and then weakly mixing and strongly mixing) vector fields: indeed, instead of studying the map $X^c_{t=1}$ (the RLF generated by b^c of Theorem 1.3 above) in the unit square K with the Lebesgue measure $\mathcal{L}^2$, it is sufficient to work in the finite space made of the centers of the subsquares

$$\Omega = \left\{ x = \left(\frac{k_1 - 1/2}{D}, \frac{k_2 - 1/2}{D} \right), k_1, k_2 = 1, \dots, D \right\}, \tag{1.4}$$

where the measure-preserving transformation $X^c_{t=1}$ reduces to a cyclic permutations. In particular in Ω it is already ergodic.

Since we cannot use the conjugation argument as we observed above, the final steps of the proof of Theorem 1.2 differ from the ones of [**Halmos:weak:mix**]. Indeed we give a general procedure to perturb vector fields $b^c \in L^\infty_t \mathrm{BV}_x$ (whose RLF X^c_t at $t = 1$ is a cyclic permutation of subsquares) into ergodic vector fields b^e (strongly mixing vector fields b^s) still belonging to $L^\infty_t \mathrm{BV}_x$: here the explicit form of X^c plays a major role, allowing us to construct *explicitly* the perturbations to b^c (Subsections 4.2,4.3).
The key idea is to apply the (rescaled) *universal mixer* vector field (introduced in [**univ:mixer**]) whose RLF at time $t = 1$ is the *Folded Baker's map*

$$U = U \llcorner_{t=1} = \begin{cases} \left(-2x + 1, -\frac{y}{2} + \frac{1}{2}\right) & x \in \left[0, \frac{1}{2}\right), \\ \left(2x - 1, \frac{y}{2} + \frac{1}{2}\right) & x \in \left(\frac{1}{2}, 1\right], \end{cases} \qquad y \in [0, 1], \tag{1.5}$$

to the subsquares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ given by Theorem 1.3.
In order to achieve ergodicity, it is sufficient to apply the universal mixer U inside a single subsquare, because the action of $X^c_{t=1}$ is already ergodic in Ω being a cyclic permutation. Together with the fact that ergodic vector fields are a G_δ-set, this gives the proof of Point (2) of Theorem 1.2. The perturbation to achieve exponential mixing is more complicated, since we need to transfer mass across different subsquares. The idea is to apply the universal mixer U to adjacent couples of subsquares, a procedure which assures that the mass of ρ is eventually equidistributed among all subsquares. The exponential mixing is a consequence of the finiteness of Ω and the properties of U (see Proposition 4.10). This concludes the proof of Point (5) of Theorem 1.2, and since strongly mixing vector fields are a subset of weakly mixing vector fields we obtain also Point (3), concluding the proof of the theorem.

It should not be surprising that exponentially mixing vector fields are a dense subset of $\mathcal{U}$. Even if this G_δ contains vector fields whose behaviour is far from mixing (as for example horizontal shears) the key point is that any vector field can be approximated by *permutation vector fields*, which are the building blocks for any mixing behaviour. We point out that our construction does not provide any example of a smooth mixing vector field: an interesting open question is the construction of a time-periodic vector field with smooth regularity in space, since the one constructed in [**Bedrossian**] does not satisfy the periodicity in time.

A completely analogous result can be obtained in any dimension by adapting the above steps, at the cost of additional heavy technicalities. In this work we decided to sketch the proof of the key estimates (i.e., the ones requiring new ideas) in the general case (see Section 5.1.1 and the comments in next section).

1.1 A discussion of the key points and results leading to the main theorem

The sketch of the proof given above includes steps which are somehow standard either in the theory of linear transport or in ergodic analysis. Other points of the proof instead require to introduce new ideas or at least to significantly develop tools present in the literature. This section is devoted to expand these critical parts in order to help the reader to understand the novelties contained in this work. In most proofs we tried to get the best possible result in term of the L^1 and BV norms, with the hope to prove a similar statement inside the closed subset $\{b : \|b_t\|_{L^1}, \mathrm{Tot.Var.}(b_t) \leq C\} \subset L^1_{t,x}$. However there are some delicate arguments where we have to increase the total variation of b of a fixed amount: we will point out these points here below.

The most technical part of this paper is the proof of the approximation theorem through cyclic permutations, Theorem 1.3. It is based on two results: the first one, whose proof is the content of Section 5, is an approximation through vector fields whose RLF X_t at time $t = 1$ is a permutation of subsquares. The second one exploits the classical result that every permutation is a product of disjoint cycles (Proposition 4.5), in order to merge these cycles into a single one.

1.1.1 Approximations of flows by permutations

The approximation through divergence-free vector fields b whose flow at $t = 1$ is a permutation of squares has been already studied in [**Shnirelman**] in the context of generalized flows for incompressible fluids. Indeed the starting point of Section 5 is Lemma 5.1, whose statement is almost identical to Lemma 4.3 of [**Shnirelman**]: it says that if T is a smooth map sufficiently close to identity, there exists an arbitrarily close flow σ_t, $t \in [0,1]$, such that $\sigma_{t=0} = T$ and $\sigma_{t=1}$ maps affinely rectangles whose edges are on a dyadic grid into rectangles belonging to the same grid. Even if the ideas of the proof are completely similar to the original ones, we choose to make them more explicit (see also Remark 5.3 for some comments on the original proof).

At this point the proof diverges from [**Shnirelman**], due to the fact that in his case one has to control the L^2-norm of the vector field while here we need to build a perturbation of a vector field (not of a map) and to estimate its BV_x-norm. In Lemma 5.9 we prove that the perturbation σ_t constructed in the above paragraph (i.e., in Lemma 5.1) can be encapsulated inside the flow X_t so that the resulting vector field is close in $L^\infty_t L^1_x$ and remains in $L^\infty_t \mathrm{BV}_x$ if the grid is sufficiently small, always under the assumption that X_t is close to identity.

We finally arrive to the approximation theorem through permutations (Theorem 5.14), which we think that can have an independent interest:

Theorem 1.4. *Let $b \in L^\infty([0,1]; BV(K))$ be a divergence-free vector field. Then for every $\epsilon > 0$ there exist $\delta', C_1, C_2 > 0$ positive constants, $D \in \mathbb{N}$ arbitrarily large and a divergence-free vector field $b^\epsilon \in L^\infty([0,1]; BV(K))$ such that*

1. $\operatorname{supp} b^\epsilon_t \subset\subset K^{\delta'} = [\delta', 1-\delta']^2$,
2. *it holds*
$$\|b - b^\epsilon\|_{L^\infty(L^1)} \leq \epsilon, \quad ||\mathrm{Tot.Var.}(b^\epsilon)(K)||_\infty \leq C_1 ||\mathrm{Tot.Var.}(b)(K)||_\infty + C_2, \tag{1.6}$$
3. *the map $X^\epsilon\llcorner_{t=1}$ generated by b^ϵ at time $t=1$ translates each subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ into a subsquare of the same grid, i.e., it is a permutation of squares.*

We remark that in the statement of Theorem 5.14 it is also assumed that there exists $\delta > 0$ such that for $\mathcal{L}^1$-a.e. $t \in [0,1]$, $\operatorname{supp} b_t \subset\subset K^\delta$. This is for technical reason, as standard approximation methods allows to deduce the Theorem 1.4 above from Theorem 5.14.

The starting point of its proof is to divide the time interval $[0,1]$ into subintervals $[t_i, t_{i+1}]$ and apply the previous perturbations (Lemma 5.1) to b_t, $t \in [t_i, t_{i-1}]$. We however need an additional mechanism in order to obtain a permutation of subsquares and not a piecewise affine map at $t=1$, as it would be the case if we only use the perturbations above.

The introduction of this new perturbation is done in Section 5.3: the idea is that if a measure preserving map T is diagonal with rational eigenvalues, then there exists a subgrid and a map R made by two rotations such that $T \circ R$ maps subsquares of the new grid into subsquares instead of rectangles (Lemma 5.12). The key point is that the total variation of the new map is bounded independently on the grid size, while the L_1-norm converges to 0 as the grid becomes smaller and smaller. This gives better BV estimates than the construction of [**Shnirelman**]. In the proof of the theorem, this rotation mechanism has to act differently in each subrectangle. The procedure illustrated in Figure 1) has to be done during the time evolution. The interesting part of the above theorem is the form of the estimate for the Total Variation in (1.6). The constant C_1 comes out from the approximation argument of Shnirelman: it means that the total variation of the piecewise affine approximation is, as expected, of the order of $\mathrm{Tot.Var.}(b_t)$: we believe that this constant C_1 can be optimized, but it is not necessary here, because the hard term is the one leading to C_2. Indeed, the second constant comes from the rotation mechanism: performing a rotation inside a rectangle costs, in terms of the total variation, as the area of the rectangle see Lemma 2.6).

1.1.2 From permutations of subsquares to ergodic/exponential mixing

The advantage of having a flow X_t such that $X_{t=1}$ is a permutations of subsquares is that its action is sufficiently simple to perturb in order to achieve a desired property. Nevertheless it requires some smart constructions, since in any case we need to control the L^1-distance and the BV norm.

The first step is to perturb a permutation of subsquares into a cyclic permutation of subsquares, i.e., a permutation made of a single cycle: this is clearly a necessary condition for ergodicity. Roughly speaking, the idea is to exchange two adjacent subsquares belonging to different cycles in order to merge them. We do this operation in two steps. In Lemma 4.4 it is shown that one can arbitrarily refine the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ into $\mathbb{N} \times \mathbb{N}\frac{1}{DM}$ so that each cycle of length k in the original grid becomes a

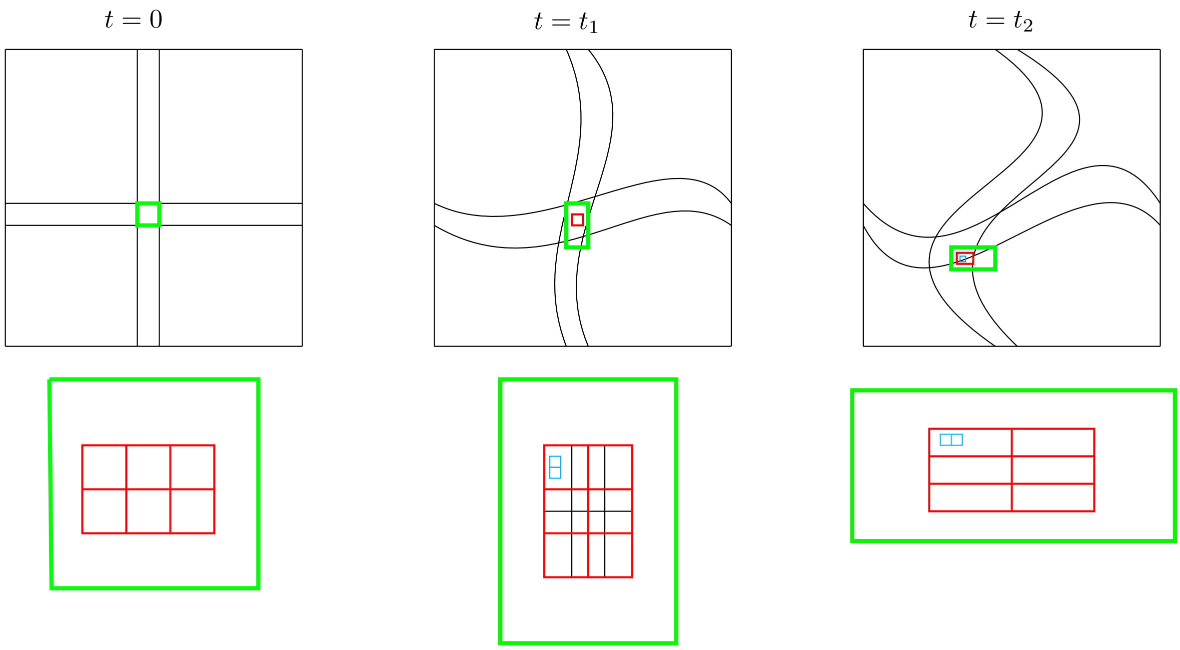


Figure 1: A graphic explanation of the action of rotations: the three top frames above shows the evolution of curves at different times t_i, where the perturbation of Lemma 5.1 takes care of transforming the green square affinely into the green rectangles. The three bottom frames is the action of the rotation on a finer grid: the red rectangle is chosen so that the action of the affine map $X_{t=t_1}$ coincides with a rotation by $\pi/2$ (as a set, but the black grid (image through an affine map) is not the image of the red grid after a rotation), and then the red grid is mapped into itself when composed with a rotation of $\pi/2$ inside the red rectangle. At the next step, one chooses again a finer grid (the light blue one) to perform the same transformation, so that the blue grid is mapped into itself.

cycle of length kM^2 in the new one. Moreover the perturbation is going to 0 in $L^\infty_t L^1_x$ as $M \to \infty$ and its $L^\infty \mathrm{BV}_x$ is arbitrarily small when D is large.
The above result allows now to exchange sets of size $(DM)^{-1}$ when merging cycles: this is done in Proposition 4.5. This proposition faces a new problem: in the previous case the exchange of subsquares of size $(DM)^{-1}$ occurs within the same subsquare of size D^{-1}: the latter is only deformed during the evolution and hence the merging can be done in the whole time interval $[0, 1]$. In the case of Proposition 4.5, instead, we are exchanging subsquares of size $(DM)^{-1}$ which are then shifted away during the flow, since they belong to different subquares of the grid D^{-1}. This requires to do the exchange sufficiently fast (i.e., during the time where they share a common boundary, Remark 4.7), or to freeze the evolution for an interval of time $[0, \delta]$ and perform the exchanges here and then let the flow permuting the subsquares to evolve during the remaining time interval $[\delta, 1]$. We choose for simplicity this second line, being easier and not changing the final result: notice however that now the constant M plays the role of controlling the constant δ^{-1}, appearing because the exchange action occurs in the time interval $[0, \delta]$.

Once we have a cyclic permutation of subsquares, the perturbation to get an ergodic vector field is straighforward.

To achieve the exponential mixing, instead, we need to transfer mass across different subsquares, and hence we face again the problem of Proposition 4.5 above: we let the mixing action occurs in an interval of time where the evolution is frozen, and then let the cyclic permutation to act in the time interval $[\delta, 1]$ (see also Remark 4.11). The idea is again to use the universal mixer (1.5) to exchange mass across to nearby subsquares. The additional difficulty here is that in order to avoid resonant phenomena we mix all squares with 2 neighboring ones, so that by simple computations the Markov Shift obtained through this map is exponentially mixing, Proposition 4.10.

To collect all the above results into a proof of Point (5) of Theorem 1.2 is not difficult at this point, and we devote a section (Section 4.4 and Corollary 3.10) to shows how to merge these result and get the desired statement.

1.2 Plan of the paper

The paper is organized as follows.

In Section 2, after listing some of the notation used in the paper, we give a short overview on BV functions (Section 2.1) and Regular Lagrangian Flows (Section 2.2), proving the extension of the continuous dependence to a complete set $\mathcal{U}$ in Proposition 2.3 (providing the proof of Point (1) of Theorem 1.2) and stating some technical estimates on composition of maps (Theorem 2.4 and (2.2), (2.3)) and on the vector field (2.5) generating a rotation (Lemma 2.6).

In Section 3 we collect some classical results in Ergodic Theory which are needed for Theorem 1.2, and also give the proof of the G_δ-properties of the set of ergodic/weakly mixing vector fields of Theorem 1.2. First we introduce the basic definitions, then in Section 3.1 we clarify the relation with the neighborhood topology and the L^1-topology used in Theorem 1.2.
In Section 3.2 we restate in our setting the well known fact that weakly mixing are a G_δ-set, as well as the first category property of strongly mixing vector fields (Proposition 3.9). The proof of the remaining parts of Theorem 1.2 is a corollary of the previous statement (Corollary 3.10), if we know that the strongly mixing vector fields are dense.
The construction of exponential mixing vector fields is based on the analysis of Markov Shift: in Section 3.3 we give the results which are linked to our construction.

In Section 4 we present the proof of the the density of exponentially mixing vector fields, under the assumption that permutation flows are dense in $L^1_t \, \mathrm{BV}_x$ w.r.t. the $L^1_{t,x}$-norm. We decide to put first this construction because it is in some sense independent on the proof of the density of permutation flows: the idea is that different functional settings can be studied by changing this last part (i.e., the density of permutation flow), while keeping the construction of approximation by permutations more or less the same. The first statement is Lemma 4.4 which allows to partition the subsquares of a given cycle into smaller subsquares still belonging to the same cycle. The usefullness of this estimate is shown in Proposition 5.4, where we need to exchange mass only on an area which is of order M^{-2}, and hence obtaining that the perturbation is small in $L^1_{t,x}$ and $L^\infty_t BV_x$ (Proposition 4.8 of Remark 4.7 addresses the problem of exchanging two subsquares during the evolution, a refinement not needed for the proof of Theorem 1.2). The last two subsection address the density of ergodic vector fields (Proposition 4.9) and of exponentially mixing vector fields (Proposition 4.10): the basic idea is the same (i.e., perturb the cyclic permutation). Section 4.4 shows at this point how the assumptions of Corollary 3.10 are verified, concluding the proof of Theorem 1.2 under the assumption of the density of vector fields whose flow is a permutation of subsquares.

The last section, Section 5, proves the cornerstone approximation result, i.e., the density of vector fields whose flow at $t = 1$ is a permutation of subsquares, Theorem 5.14 (whose statement is the same of Theorem 1.6).
In Section 5.1 we approximate a smooth flow close to identity with a BV flow which is locally affine in subrectangles: Lemma 5.1 considers the 2d-case as in [**Shnirelman**], while the needed variations for the d-dimensional case are in Section 5.1.1.
The BV estimates for such perturbed flow are studied in Section 5.2. A preliminary result (Lemma 5.7) takes care of the conditions that the area of the subsquares has to be a dyadic rational, while the key estimates are in Lemma 5.9: an important fact is that as the grid becomes finer the perturbation becomes smaller.
An ingredient for obtaining a flow which is a permutation of subsquares is the use of rotations: in Section 5.3 we study these elementary transformations.
The main approximation theorem, Theorem 5.14, is stated and proved in Section 5.4. Its proof uses all the ingredients of the previous sections, and an additional argument on how to encapsulate rotations in order to control the total variation.

2 Preliminaries and notation

First, a list of standard notations used throughout this paper.

- $\Omega \subset \mathbb{R}^n$ denotes in general an open set; $\mathcal{B}(\Omega)$ denotes the σ-algebra of Borel sets of Ω;
- $\mathrm{dist}(x, A)$ is the distance of x from the set $A \subset \Omega$, defined as the infimum of $|x - y|$ as y varies in A;

- $\forall A \subset \Omega$, $\mathring{A}$ denotes the interior of A and ∂A its boundary, moreover, if $\epsilon > 0$, then A^ϵ is the ϵ-neighbourhood of A, that is
$$A^\epsilon = \{x \in \Omega : \mathrm{dist}(x, \partial A) \le \epsilon\};$$
- $\mathcal{M}_b(\Omega)$ bounded Radon measures;
- if $\nu \in \mathcal{M}_b(\Omega)$ then $\|\nu\|$ denotes its total variation;
- $\mathrm{BV}(\Omega)$ is the set of functions with bounded variation, and if $u \in \mathrm{BV}(\Omega)$ we will use instead Tot.Var.(u) to denote $\|Du\|$;
- $\mathcal{L}^d$ denotes the d-dimensional Lebesgue measure on $\mathbb{R}^d$, and $\mathcal{H}^k$ the k-dimensional Hausdorff measure;
- $K = [0,1]^2$ is the unit square;
- $\mathcal{L}^2 \llcorner_K$ denotes the normalized Lebesgue measure on K;
- let $b : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2$, and let $t, s \in [0,1]$ then we denote by $X(t,s,x)$ a solution of
$$\begin{cases} \dot{x}(t) = b(t, x(t)) \\ x(s) = x, \end{cases}$$
moreover we will use $X(t)(x)$ or (alternatively $X_t(x)$) for $X(t,0,x)$ (in our setting as a flow the function $X(t,s,x)$ is unique a.e.);
- (S, Σ, μ) denotes a locally compact separable metric space where μ is a normalized complete measure;
- $G(S)$ denotes the space of automorphisms of S.

2.1 BV functions

In this subsection we recall some results concerning functions of bounded variation. For a complete presentation of the topic, see [**AFP**]. Let $u \in \mathrm{BV}(\Omega; \mathbb{R}^m)$ and $Du \in \mathcal{M}_b(\Omega)^{n \times m}$ the $n \times m$-valued measure representing its distributional derivative. We recall the decomposition of the measure Du
$$Du = D^{\mathrm{cont}}u + D^{\mathrm{jump}}u = D^{\mathrm{a.c.}}u + D^{\mathrm{cantor}}u + D^{\mathrm{jump}}u,$$
where $D^{\mathrm{cont}}u, D^{\mathrm{a.c.}}u, D^{\mathrm{cantor}}u, D^{\mathrm{jump}}u$ are respectively the continuous part, the absolutely continuous part, the Cantor part and the jump part of the measure. We also recall that for $u \in BV(\Omega)$ the following estimate on the translation holds: for every $C \subset \Omega$ compact and $z \in \mathbb{R}^n$ such that $|z| \le \mathrm{dist}(C, \partial\Omega)$
$$\int_C |u(x+z) - u(x)| dx \le \left| \sum_{i=1}^n z_i D_i u \right| (C^{|z|}). \tag{2.1}$$

2.2 Regular Lagrangian Flows

Throughout the paper we will consider divergence-free vector fields $b : [0,1] \times K \to \mathbb{R}^2$ in the space $L^\infty([0,1]; \mathrm{BV}(K))$ (in short $b \in L^\infty_t \mathrm{BV}_x$) such that $\mathrm{supp}(b_t) \subset\subset \mathring{K}$ for $\mathcal{L}^1$-a.e. $t \in [0,1]$: it is standard to extend the analysis to divergence-free BV-vector fields in $\mathbb{R}^2$ with support in K. When the velocity field b is Lipschitz, then its *flow* is well-defined in the classical sense, indeed it is the map $X : [0,1] \times K \to K$ satisfying
$$\begin{cases} \frac{d}{dt} X_t(x) = b(t, X_t(x)); \\ X_0(x) = x. \end{cases}$$
But when we allow the velocity fields to be discontinuous (as in our case BV regular in space) we can still give a notion of a flow (namely the *Regular Lagrangian Flow*). These flows have the advantage to allow rigid *cut and paste* motions, since they do not preserve the property of a set to be connected. More in detail, we give the following

Definition 2.1. Let $b \in L^1([0,1] \times \mathbb{R}^2; \mathbb{R}^2)$. A map $X : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2$ is a *Regular Lagrangian Flow* (RLF) for the vector field b if

1. for a.e. $x \in \mathbb{R}^2$ the map $t \to X_t(x)$ is an absolutely continuous integral solution of

$$\begin{cases} \frac{d}{dt}x(t) = b(t, x(t)); \\ x(0) = x. \end{cases}$$

2. there exists a positive constant C independent of t such that

$$\mathcal{L}^2(X_t^{-1}(A)) \leq C\mathcal{L}^2(A), \quad \forall A \in \mathcal{B}(\mathbb{R}^2).$$

DiPerna and Lions proved existence, uniqueness and stability for Sobolev vector fields with bounded divergence [**DiPerna:Lions**], while the extension to the case of BV vector fields with divergence in L^1 has been done by Ambrosio in [**Ambrosio:BV**]. When dealing with divergence-free vector fields b the unique Regular Lagrangian Flow $t \to X_t$ associated with b is a flow of measure-preserving maps, main objects of investigations in Ergodic Theory. In the sequel we will build flows of measure-preserving maps originating from divergence-free vector fields; more precisely, if a flow $X : [0,1] \times K \to K$ is invertible, measure-preserving for $\mathcal{L}^1$-a.e. t and the map $t \to X_t$ is differentiable for $\mathcal{L}^1$-a.e. t and $\dot{X}_t \in L^1(K)$, then the *vector field associated with* X_t is the divergence-free vector field defined by

$$b_t(x) = b(t,x) = \dot{X}_t(X_t^{-1}(x)).$$

Theorem 2.2 (Stability, Theorem 6.3 [**Ambrosio:Luminy**])**.** *Let $b_n, b \in L^\infty([0,1], \mathrm{BV}(K))$ be divergence-free vector fields and let X^n, X be the corresponding Regular Lagrangian Flows. Assume that*

$$||b_n - b||_{L^1_{t,x}} \to 0 \quad \text{as } n \to \infty,$$

then

$$\lim_{n\to\infty} \int_K \sup_{t\in[0,1]} |X_t^n(x) - X_t(x)| dx = 0.$$

In this setting we can extend the family of vector fields we consider to a Polish subspace of $L^1_{t,x}$ in which we still have a notion of uniqueness. This extension allows us to apply Baire Category Theorem for the results of genericity that we will give for weakly mixing vector fields.

Proposition 2.3 (Extension)**.** *Let*

$$\Phi : \{b \in L^\infty_t \mathrm{BV}_x : D \cdot b_t = 0\} \subset \{b \in L^1([0,1], L^1(K)), D \cdot b_t = 0\} \to C([0,1], L^1(K))$$

the map that associates b with its unique Regular Lagrangian Flow X_t. Then Φ can be extended as a continuous function to a G_δ-set $\mathcal{U}$ containing $\{b \in L^\infty_t \mathrm{BV}_x : D \cdot b_t = 0\}$.

This proposition proves Point (1) of Theorem 1.2.

Proof. We recall that for every $f : A \to Z$ continuous where $A \subset W$ is metrizable and Z is a complete metric space, there exists a G_δ-set $A \subset G$ and a continuous extension $\tilde{f} : G \to Z$ (Proposition 2.2.3, [**Srivastava**]). Thus we have to prove the continuity of the map Φ which follows by

$$\begin{aligned} \|\Phi(b^n) - \Phi(b)\|_{C_t L^1_x} &= \sup_{t\in[0,1]} \int_K \big|X_t^n(x) - X_t(x)\big| dx \\ &\leq \int_K \sup_{t\in[0,1]} \big|X_t^n(x) - X_t(x)\big| dx. \end{aligned}$$

This concludes the proof. □

We will also use the following tools to prove the main approximation theorems of the paper. The first one gives a rule to compute the total variation of the composition of vector fields, while the second one is a direct computation of the cost, in terms of the total variation of the vector field whose flow rotates rectangles.

Theorem 2.4 (Change of variables, Theorem 3.16, [**AFP**]). *Let Ω, Ω' two open subsets of $\mathbb{R}^n$ and let $\phi : \Omega \to \Omega'$ invertible with Lipschitz inverse, then $\forall u \in BV(\Omega')$ the function $v = u \circ \phi$ belongs to $BV(\Omega)$ and*

$$\mathrm{Tot.Var.}(v) \leq \mathrm{Lip}(\phi^{-1})^{n-1} \mathrm{Tot.Var.}(u).$$

Corollary 2.5. *Let $\Omega, \Omega' \subset \mathbb{R}^n$ be two open sets where $\partial\Omega'$ is Lipschitz and let $\phi : \bar{\Omega} \to \bar{\Omega}'$ invertible with Lipschitz inverse, then $\forall u \in BV(\mathbb{R}^n)$ the function*

$$v = \begin{cases} u \circ \phi & x \in \Omega, \\ 0 & \textit{otherwise}, \end{cases}$$

belongs to $BV(\mathbb{R}^n)$ and

$$\begin{aligned} \mathrm{Tot.Var.}(v)(\mathbb{R}^n) &\leq \mathrm{Lip}(\phi^{-1})^{n-1} \big(\mathrm{Tot.Var.}(u)(\Omega') + \|\mathrm{Tr}(u, \partial\Omega')\|_{L^1(\mathcal{H}^{n-1} \llcorner_{\partial\Omega'})} \big) \\ &= \mathrm{Lip}(\phi^{-1})^{n-1} \mathrm{Tot.Var.}(u \llcorner_{\Omega'})(\mathbb{R}^n). \end{aligned}$$

In the following, we have often to study the properties of the vector field b_3 associated with the composition $Y_3(t)$ of two smooth measure preserving flows $t \mapsto Y_i(t)$, $i = 1, 2$, with associated vector fields b_1, b_2. By direct computation

$$\begin{aligned} b_3(t, Y_3(t, y)) = \partial_t Y_1(t, Y_2(t, y)) &= b_1(t, Y_3(t, y)) + \nabla Y_1(t, Y_2(t, y)) b_2(t, Y_2(t, y)), \\ b_3(t, x) &= b_1(t, x) + \nabla Y_1(t, Y_2(t, Y_3^{-1}(t, x))) b_2(t, Y_2(t, Y_3^{-1}(t, x))) \\ &= b_1(t, x) + \nabla Y_1(t, Y_1^{-1}(t, x)) b_2(t, Y_1^{-1}(t, x)). \end{aligned} \tag{2.2}$$

Hence using Theorem 2.4 we conclude that (being $Y_1 \circ Y_2$ measure preserving too)

$$\begin{aligned} \mathrm{Tot.Var.}(b_3) &\leq \mathrm{Tot.Var.}(b_1) + \mathrm{Lip}(Y_1)^{n-1} \mathrm{Tot.Var.}(DY_1(t) b_2) \\ &\leq \mathrm{Tot.Var.}(b_1) + \|\nabla Y_1\|_\infty^n \mathrm{Tot.Var.}(b_2) + \|\nabla Y_1\|_\infty^{n-1} \|b_2\|_\infty \mathrm{Tot.Var.}(DY_1(t)). \end{aligned} \tag{2.3}$$

Throughout the paper we will extensively use a flow rotating rectangles and the vector field associated with it. More precisely we define the *rotation flow* $r_t : K \to K$ for $t \in [0, 1]$ in the following way: call

$$V(x) = \max \left\{ \left| x_1 - \frac{1}{2} \right|, \left| x_2 - \frac{1}{2} \right| \right\}^2, \quad (x_1, x_2) \in K.$$

Then the *rotation field* is $r : K \to \mathbb{R}^2$

$$r(x) = \nabla V^\perp(x), \tag{2.4}$$

where $\nabla^\perp = (-\partial_{x_2}, \partial_{x_1})$ is the orthogonal gradient. Finally the rotation flow r_t is the flow of the vector field r, i.e., the unique solution to the following ODE system:

$$\begin{cases} \dot{r}_t(x) = r(r_t(x)), \\ r_0(x) = x. \end{cases} \tag{2.5}$$

This flow rotates the cube counterclockwise of an angle $\frac{\pi}{2}$ in a unit interval of time.

Lemma 2.6. *Let $R \subset \mathbb{R}^2$ a rectangle of sides $a, b > 0$. Consider the rotating flow*

$$R_t = \chi^{-1} \circ r_t \circ \chi,$$

where $\chi : R \to K$ is the affine map sending R into the unit cube and r_t is the rotation flow defined in (2.5)*. Let b_t^R the divergence-free vector field associated with R_t. Then*

$$\mathrm{Tot.Var.}(b_t^R)(\mathbb{R}^2) = 4a^2 + 4b^2, \quad \forall t \in [0, 1].$$

Proof. The potential V generating the rotation of $\pi/2$ in this case is the function

$$V(x) = \max \left\{ \frac{b}{a} \left(x_1 - \frac{a}{2} \right)^2, \frac{a}{b} \left(x_2 - \frac{b}{2} \right)^2 \right\},$$

where we assume that $R = [0, a] \times [0, b]$, so that the vector field is given by

$$r(x) = \nabla^\perp V = \begin{cases} \big(0, \frac{2b}{a} (x_1 - \frac{a}{2})\big) & |x_1| \geq \frac{b}{a} |x_2|, 0 \leq x_1 \leq a, \\ \big(-\frac{2a}{b} (x_2 - \frac{b}{2}), 0\big) & |x_1| < \frac{b}{a} |x_2|, 0 \leq x_2 \leq b. \end{cases}$$

Hence by elementary computations

$$\|D^{\mathrm{cont}} r\| = a^2 + b^2, \quad \|D^{\mathrm{jump}} r\| = 3a^2 + 3b^2,$$

and then we conclude. ☐

3 Ergodic Theory

We will consider flows of divergence-free vector fields from the point of view of Ergodic Theory. Even if we apply the results to the case $(K, \mathcal{B}(K), \mathcal{L}^2 \llcorner_K)$ in this section we will give the notions of ergodicity and mixing in more general spaces [**Ergodic:theory**]. More precisely, let (Ω, Σ, μ) be a locally compact separable metric space where μ is complete and normalized, that is $\mu(\Omega) = 1$.

Definition 3.1. An *automorphism* of the measure space (Ω, Σ, μ) is a one-to-one map $T : \Omega \to \Omega$ bi-measurable and measure-preserving, that is

$$\mu(A) = \mu(T(A)) = \mu(T^{-1}(A)), \quad \forall A \in \Sigma.$$

We call $G(\Omega)$ the group of automorphisms of the measure space (Ω, Σ, μ).

Definition 3.2. A *flow* $\{X_t\}$, $t \in \mathbb{R}$, is a one-parameter group of automorphisms of (Ω, Σ, μ) such that for every $f : \Omega \to \mathbb{R}$ measurable, the function $f(X_t(x))$ is measurable on $\Omega \times \mathbb{R}$.

Definition 3.3. Let $T : \Omega \to \Omega$ an automorphism. Then

- T is ***ergodic*** if for every $A \in \Sigma$

$$T(A) = A \quad \Rightarrow \quad \mu(A) = 0 \text{ or } \mu(A) = 1; \tag{3.1}$$

- T is ***weakly mixing*** if $\forall A, B \in \Sigma$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\big]^2 = 0; \tag{3.2}$$

- T is ***(strongly) mixing*** if $\forall A, B \in \Sigma$

$$\lim_{n \to \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B). \tag{3.3}$$

Remark 3.4. It is a well-known and quite elementary fact that strongly mixing $\Rightarrow$ weakly mixing $\Rightarrow$ ergodic.

We can give the analogous definitions for the flow:

Definition 3.5. Let $\{X_t\}$ a flow of automorphisms. Then

- $\{X_t\}$ is ***ergodic*** if for every $A \in \Sigma$

$$X_t(A) = A \quad \Rightarrow \quad \mu(A) = 0 \text{ or } \mu(A) = 1; \tag{3.4}$$

- $\{X_t\}$ is ***weakly mixing*** if $\forall A, B \in \Sigma$

$$\lim_{t \to \infty} \fint_0^t \left[\int_\Omega \chi_A(X_{-s}(x))\chi_B(x) d\mu - \mu(A)\mu(B)\right]^2 ds = 0; \tag{3.5}$$

- $\{X_t\}$ is ***(strongly) mixing*** if $\forall A, B \in \Sigma$

$$\lim_{t \to \infty} \int_\Omega \chi_A(X_{-t}(x))\chi_B(x) d\mu = \mu(A)\mu(B). \tag{3.6}$$

3.1 The neighbourhood topology as a convergence in measure.

To get a genericity result it is necessary to identify the correct topology on $G(\Omega)$. Following the work of Halmos [**Halmos:weak:mix**] we define the *neighbourhood topology* as the topology generated by the following base of open sets: let $T \in G(\Omega)$ then

$$N(T) = \{S \in G(\Omega) : |T(A_i) \triangle S(A_i)| < \epsilon, \quad i = 1, \dots, n\},$$

where $\epsilon > 0$ and $A_i \in \Sigma$ are measurable sets.

Since for our purposes we will consider the L^1 topology on $G(\Omega)$, we recall the following

Proposition 3.6. *Let $\{T_n\}, T \subset G(\Omega)$ and assume that $T_n \to T$ in measure. Then $T_n \to T$ in the neighbourhood topology. Conversely, if $T_n \to T$ in the neighbourhood topology, then T_n converges to T in measure.*

Since in our case Ω is a compact set, then the convergence in measure is equivalent to the convergence in L^1: hence we will use the L^1 topology for maps as in Proposition 2.3.

We will be concerned with flows of vector fields extended periodically to the real line, that is $b(t+1) = b(t)$. Even if X_t is not a flow of automorphisms, the quantities in the r.h.s. of (3.5),(3.6) can be computed and are related to the mixing properties of $T = X_{t=1}$. Also the ergodic properties of $T = X_1$ are equivalent to an ergodic property of X_t.

Let $\{X_s\}_{s\in[0,1]}$ be a family of automorphisms of Ω such that $s \to X_s$ is continuous (hence uniformly continuous) with respect to the neighborhood topology of $G(\Omega)$. Let $T = X_{t=1}$ and define

$$X_t = X_s \circ T^n = T^n \circ X_s, \quad t = n + s, s \in [0,1).$$

Lemma 3.7. *The following hold*

1. *if T is ergodic then for every set $A \in \Sigma$*
$$\fint_0^t \chi_{X_s(A)} ds \to_{L^1} |A|;$$
2. *T is weakly mixing iff for every $A, B \in \Sigma$*
$$\lim_{t\to\infty} \fint_0^t \big[|X_s(A) \cap B| - |A||B|\big]^2 ds = 0;$$
3. *T is mixing iff $\forall A, B \in \Sigma$*
$$\lim_{t\to\infty} \big|X_t(A) \cap B\big| = |A||B|.$$

The proof of this lemma is given in Appendix 6, since we believe it is standard and not strictly related to our results.

Definition 3.8. Let $b \in L^\infty([0,1], \mathrm{BV}(\mathbb{R}^2))$, $\operatorname{supp} b_t \subset K$, be a divergence-free vector field. We will say that b is ergodic (weakly mixing, strongly mixing) if its unique RLF X_t evaluated at $t = 1$ is ergodic (respectively weakly mixing, strongly mixing).

3.2 Genericity of weakly mixing

Let $\mathcal{U}$ be the G_δ-subset of $L^1_{t,x}$ where the Regular Lagrangian Flow can be uniquely extended by continuity (Proposition 2.3). The first statement has the same proof of [Theorem 2, [**Halmos:weak:mix**]] and [Page 77,[**Halmos:lectures**]]:

Proposition 3.9. *The set of ergodic/weakly-mixing vector fields is a G_δ-set in $\mathcal{U}$, the set of strongly mixing is a first category set.*

We repeat the proof for convenience only for weakly/strongly mixing, the case for ergodic vector fields is completely analogous [**Halmos:ergodic**].

Proof. Since the map $\tilde{\Phi}(b)(t = 1) = T(b)$ defined in Proposition 2.3 is continuous from $\mathcal{U}$ into $L^1(K, K)$, it is enough to prove that the set of weakly mixing maps is a G_δ. For simplicity we define a new topology on $G(K)$ that coincides with the neighbourhood topology known as *Von Neumann strong neighbourhood topology.* Given $T \in G(K)$, define a linear operator $T : L^2(K, \mathbb{C}) \to L^2(K, \mathbb{C})$ by

$$(Tf)(x) = f(Tx) \quad \forall f \in L^2(K, \mathbb{C})$$

such that $||Tf||_{L^2} = ||f||_{L^2}$ (being T measure-preserving). Consider f_i a countable dense subset in L^2: a base of open sets in the strong neighbourhood topology is given by

$$N(T) = \{S \in G(K) : ||Tf_i - Sf_i||_2 \le \epsilon, \quad i = 1, \dots, n\}.$$

Then we define

$$E(i, j, m, n) = \big\{T \in G(K) : \big|(T^n f_i, f_j) - (f_i, 1)(1, f_j)\big| < 2^{-m}\big\},$$

where $(\cdot,\cdot)$ denotes the scalar product in L^2. Simply observing that $T \to (Tf, g)$ is continuous in the strong neighbourhood topology then by Proposition 3.6 it follows that $E(i,j,n,m)$ is open in $L^1(K,K)$, and then

$$G = \bigcap_{i,j,m} \bigcup_n E(i,j,m,n)$$

is a G_δ-set. By the Mixing Theorem [Theorem 2, page 29, [**Ergodic:theory**]] G coincides with the set of weakly mixing maps in $L^1(K,K)$. Indeed if T is not mixing, then there exists $f \neq 0$ and a complex eigenvalue $\lambda \in \{|z| = 1, z \neq 0, 1\}$ such that $Tf = \lambda f$. We can assume that f is orthogonal to the eigenvector 1, that is $(f, 1) = 0$, and also that $||f||_2 = 1$. Now choose i such that $||f - f_i|| \leq \epsilon$ for some ϵ to be chosen later and take $f_j = f_i$. Then

$$\begin{aligned}
1 &= |(T^n f, f) - (f,1)(f,1)| \\
&\leq |(T^n f, f) - (T^n f, f_i)| + |(T^n f, f_i) - (T^n f_i, f_i)| + |(T^n f_i, f_i) - (f_i, 1)(1, f_i)| \\
&\quad + |(f_i,1)(1,f_i) - (f,1)(f_i,1)| + |(f,1)(1,f_i) - (f,1)(f,1)| \\
&\leq 2||f - f_i||_2 + 2||f_i||_2||f - f_i||_2 + |(T^n f_i, f_i) - (f_i,1)(1,f_i)|,
\end{aligned}$$

so since $||f_i||_2 \leq 1 + \epsilon$ we get that

$$1 \leq 2\epsilon + 2(1+\epsilon)\epsilon + |(T^n f_i, f_i) - (f_i,1)(1,f_i)|.$$

With the choice of $\epsilon > 0$ small enough we get that $\frac{1}{2} \leq |(T^n f_i, f_i) - (f_i,1)(1,f_i)|$, that is $T \notin G$. This concludes the proof of the first part of the statement.
We next prove that the set of strongly mixing vector fields is a first category set. Let $A \subset K$ be a measurable set such that $|A| = \frac{1}{2}$. Then define the F_σ-set

$$F = \bigcup_n \bigcap_{k>n} \left\{ T \in G(K) : \left| |(T^{-k}(A) \cap A)| - \frac{1}{4} \right| \leq \frac{1}{5} \right\}.$$

Clearly strongly mixing maps are contained in F by definition and therefore strongly mixing vector fields are contained in $\tilde{F} = \tilde{\Phi}^{-1}(t=1)(F)$. This $\tilde{F}$ is a set of first category: indeed consider the set

$$\bigcup_{k>n} \tilde{\Phi}^{-1}(t=1) \left(\left\{ T \in G(K) : \left| |(T^{-k}(A) \cap A)| - \frac{1}{4} \right| \leq \frac{1}{5} \right\} \right)^c. \tag{3.7}$$

By our main result (Theorem 5.14) $\forall b \in \mathcal{U}$ for all $n \in \mathbb{N}$ there exists $k > n$ and $b^p \in L^\infty_t(\mathrm{BV}_x)$ such that the RLF $X^p_{t=1}$ associated with b^p evaluated at $t = 1$ is a permutation of subsquares of period k. Hence

$$\bigcup_{k>n} \{ b \in L^\infty_t(\mathrm{BV}_x) \text{ permutation of period } k \}$$

is dense and contained in (3.7), so that we conclude that (3.7) is open and dense for all n, i.e., F is of first category. □

Corollary 3.10. *Assume that the set*

$$SM = \big\{ b \in \mathcal{U} : b \text{ is strongly mixing} \big\}$$

is dense in $\mathcal{U}$. *Then the set of weakly mixing vector fields is residual.*

Proof. Elementary. □

Our aim will be to prove the assumption of the above corollary, which together with Proposition 3.9 will conclude the proof of Theorem 1.2 once we show that the dense set of strongly mixing vector fields are actually exponentially mixing.

Remark 3.11. The above situation, namely

- b strongly mixing is dense in $\mathcal{U}$,
- b weakly mixing is second category in $\mathcal{U}$,

is in some sense the best situation we can hope in $\mathcal{U}$. Indeed, the strongly mixing vector fields are a set of first category and then it is not a "fat" set. On the other hand, the weakly mixing vector fields would be a "fat" set once we know their density, which one deduces from the density of the strongly mixing vector fields.

3.3 Markov Shifts

When dealing with finite spaces $X = \{1, \dots, n\}$ and processes whose outcome at time k depends only on their outcome at time $k-1$ it is easier to determine some statistical properties of the dynamical system, as ergodicity and mixing (see for a reference [**Mane**],[**Viana**]). More precisely let $B(n) = \{\theta : \mathbb{Z} \to X\}$ the space of sequences and define a cylinder

$$C(m, k_1, \dots, k_r) = \Big\{\theta \in B(n) : \theta(m+i) = k_{i+1},\ i = 0, \dots, r-1\Big\}$$

where $m \in \mathbb{Z}$ and $k_i \in X$. Therefore, since the Borel σ-algebra on $B(n)$ is generated by disjoint union of cylinders, we can define a probability measure μ on $B(n)$ simply determining its value on cylinders. A *Markov measure* μ is a probability measure on $B(n)$ for which there exist $p_i > 0, P_{ij} \geq 0$, $i, j = 1, \dots n$, with

$$\sum_i p_i = \sum_j P_{ij} = 1, \quad \sum_i p_i P_{ij} = p_j,$$

such that

$$\mu(C(m, k_1, \dots, k_r)) = p_{k_1} P_{k_1 k_2} \dots P_{k_{r-1} k_r}$$

for every cylinder $C(m, k_1, \dots, k_r)$. The P_{ij} are called *transition probabilities* and $P = (P_{ij})$ is the *transition matrix*. The transition matrix is a stochastic matrix, that is $\sum_j P_{ij} = 1$ for every i. Now define $P_{ij}^{(m)}$ the coefficients of the matrix P^m.

Definition 3.12. A matrix P with positive coefficients is *irreducible* if $\forall i, j$ there exists m such that $P_{ij}^{(m)} > 0$.

Definition 3.13. A matrix P with positive coefficients is *aperiodic* if there exists m such that $P_{ij}^{(m)} > 0$ $\forall i, j$.

A *Markov shift* is a map $\sigma : (B(n), \mu) \to (B(n), \mu)$ such that

$$\sigma(\theta)(i) = \theta(i+1), \quad \forall \theta \in B(n).$$

Then it can be proved that $\sigma_\sharp \mu = \mu$. We conclude this subsection with the following results on ergodicity and mixing properties of Markov shifts (see [**Viana**], Chapter 7).

Proposition 3.14 (Ergodicity)**.** *The following are equivalent:*

1. $\sigma : (B(n), \mu) \to (B(n), \mu)$ *is ergodic;*
2. P *is irreducible;*
3. $\displaystyle\lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} P_{ij}^{(k)} = p_j.$

Proposition 3.15 (Mixing)**.** *The following are equivalent:*

1. $\sigma : (B(n), \mu) \to (B(n), \mu)$ *is strongly mixing;*
2. P *is aperiodic;*
3. $\displaystyle\lim_{m \to \infty} P_{ij}^{(m)} = p_j.$

4 Density of Strongly Mixing vector fields

In Section 5 we prove that *permutation vector fields* (i.e., vector fields whose RLF X_t, when evaluated at $t = 1$, is a permutation of subsquares, Point (3) of Theorem 5.14) are dense in $\mathcal{U}$. In this section, we show that each permutation vector field can be approximated by a vector field whose RLF evaluated at time $t = 1$ is a unique cycle. This approximation result will be used to get first the density of ergodic vector fields, then the density of strongly mixing vector fields.

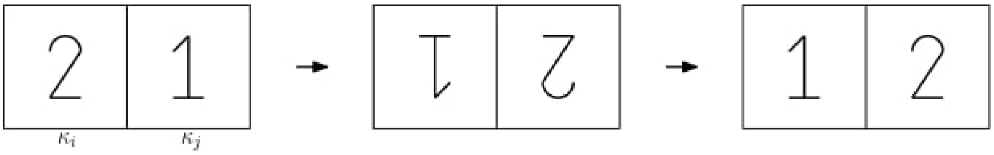


Figure 2: The action of the *transposition flow* T_t.

4.1 Cyclic permutations of squares

We start by recalling some basic facts about permutations. Denote by S_n the set of permutations of the elements $\{1, \dots, n\}$.

Definition 4.1. Let $\sigma \in S_n$ be a permutation and $k \le n \in \mathbb{N}$. We say that σ is a *k-cycle* c (or simply a *cycle*) if there exist k distinct elements $a_1, \dots, a_k \in \{1, \dots, n\}$ such that

$$\sigma(a_i) = a_{i+1}, \quad \sigma(a_k) = a_1, \quad \sigma(x) = x \quad \forall x \neq a_1, \dots, a_k.$$

We identify the permutation with the ordered set $c = (a_1 a_2 \dots a_k)$. The number k is the *length* of the cycle. We say that c is *cyclic* if $k = n$. We call *transpositions* the 2-cycles.

Definition 4.2. Let c_1, c_2 be the cycles $c_1 = (a_1 \dots a_t)$ and $c_2 = (b_1 \dots b_s)$. We say that c_1, c_2 are *disjoint* cycles if $a_i \neq b_j$ for every $i = 1, \dots, t$, $j = 1, \dots, s$.

Recall the following result.

Theorem 4.3. *Every permutation $\sigma \in S_n$ is the product of disjoint cycles.*

From now on we will address flows X_t of divergence-free vector fields such that $X_{t=1}$ is a permutation of squares of size $\frac{1}{D}$.

Let us fix the size $D \in \mathbb{N}$ of the grid in the unit square K. We enumerate the D^2 subsquares of the grid and we consider S_{D^2} the set of the permutations of $\{\kappa_1, \dots, \kappa_{D^2}\}$. We say that two squares (ore more in general two rectangles) are *adjacent* if they have a common side. We will use also the word *adjacent* for cycles: two disjoint cycles of squares c_1, c_2 are adjacent if there exist $\kappa_1 \in c_1$, $\kappa_2 \in c_2$ adjacent subsquares. Two *adjacent* squares can be *connected* by a transposition, which can be defined simply as an exchange between the two squares: let κ_i, κ_j two adjacent squares of size $\frac{1}{D}$ and let $R = \kappa_i \cup \kappa_j$, then the *transposition flow* between κ_i, κ_j is $T_t(\kappa_i, \kappa_j) : [0,1] \times K \to K$ defined as

$$T_t(\kappa_i, \kappa_j) = \begin{cases} \chi^{-1} \circ r_{4t} \circ \chi & x \in \mathring{R},\ t \in \left[0, \frac{1}{2}\right], \\ \chi_i^{-1} \circ r_{4t} \circ \chi_i & x \in \mathring{\kappa}_i,\ t \in \left[\frac{1}{2}, 1\right], \\ \chi_j^{-1} \circ r_{4t} \circ \chi_j & x \in \mathring{\kappa}_j,\ t \in \left[\frac{1}{2}, 1\right], \\ x & \text{otherwise}, \end{cases} \tag{4.1}$$

where the map $\chi : R \to K$ is the affine map sending the rectangle R into the unit square K, χ_i, χ_j are the affine maps sending κ_i, κ_j into the unit square K and r is the rotation flow (2.5). This invertible measure-preserving flow has the property to exchange the two subsquares in the unit time interval (Figure 2). Moreover, by the computations done in Lemma 2.6, we can say that

$$\text{Tot.Var.}(\dot{T}_t(\kappa_i, \kappa_j))(\bar{R}) \le \frac{20}{D^2}. \tag{4.2}$$

Lemma 4.4. *Let $b \in L^\infty_t(\mathrm{BV}_x)$ be a divergence-free vector field and assume that its flow at time $t = 1$, namely $X_{\llcorner t=1}$, is a k-cycle of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ where $k, D \in \mathbb{N}$. Then for every $M = 2^p$, there exists $b^c \in L^\infty_t(\mathrm{BV}_x)$ divergence-free vector field such that*

$$||b - b^c||_{L^\infty(L^1)} \le \mathcal{O}\left(\frac{1}{D^3 M}\right),$$

$$||\text{Tot.Var.}(b^c - b)(K)||_\infty \le \mathcal{O}\left(\frac{1}{D^2}\right),$$

and the map $X^c_{t=1} : K \to K$ is a kM^2-cycle of squares of size $\frac{1}{DM}$, where $X^c_t : [0,1] \times K \to K$ is the flow associated with b^c.

Here and in the following we will write $T(\kappa_i) = \kappa_j$ meaning that T is a rigid translation of κ_i to κ_j. This to avoid cumbersome notation.

Proof. Let us call $T \doteq X_{t=1}$: being a cycle, there exist $\{\kappa_1, \dots, \kappa_k\} \subset \{1, \dots, D^2\}$ such that

$$T(\kappa_i) = \kappa_{i+1}, \quad T(\kappa_n) = \kappa_1, \quad T(x) = x \quad \text{otherwise}.$$

Now fix some $M = 2^p$ and divide each subsquare κ_i into M^2 subsquares κ_i^j with $j = 1, \dots, M^2$. Since T is a translation of subsquares and choosing cleverly the labelling $j \to \kappa_i^j$, then we have also $T(\kappa_i^j) = \kappa_{i+1}^j$ so that T is a permutation of subsquares κ_i^j. More precisely, it is the product of M^2 disjoint cycles of length k. The idea is to connect these cycles with transpositions in order to have a unique cycle of length kM^2: we will need a parturbation inside κ_1.

Divide the M^2 subsquares of κ_1 into $\frac{M^2}{2}$ couples $R_h^2 = \kappa_1^j \cup \kappa_1^{j'}$ with $h = 1, \dots, \frac{M^2}{2}$ and $\kappa_1^j, \kappa_1^{j'}$ are adjacent squares. In the time interval $\left[0, \frac{1}{2}\right]$ perform $\frac{M^2}{2}$ transpositions, one in each R_h^2, that is

$$X_t^c(x) = X_t \circ T_t^2(x), \quad t \in \left[0, \frac{1}{2}\right],$$

where the flow $T^2 : \left[0, \frac{1}{2}\right] \times K \to K$

$$T_t^2 \llcorner_{R_h^2} \doteq T_{2t}(\kappa_1^j, \kappa_1^{j'}) \quad \text{and} \quad T_t^2(x) = x \text{ otherwise},$$

is the transposition flow (4.1) between κ_1^j and $\kappa_1^{j'}$ as defined in (4.1) above. Then for $t \in \left[0, \frac{1}{2}\right]$ fixed,

$$\text{Tot.Var.}(b_t^c - b_t)(\kappa_1) \leq \mathcal{O}(1) 2 \frac{M^2}{2} \frac{20}{M^2 D^2},$$

where we have used (4.2). We observe that at this time step we have obtained $\frac{M^2}{2}$ disjoint $2k$-cycles.

In the time interval $\left[\frac{1}{2}, \frac{3}{4}\right]$ we divide the unit square into squares $R_h^4 = R_j^2 \cup R_{j'}^2$ with $h = 1, \dots, \frac{M^2}{4}$ where $R_j^2, R_{j'}^2$ are adjacent (in particular there exist $\kappa_1^j \subset R_j^2, \kappa_1^{j'} \subset R_{j'}^2$ adjacent squares). Now we perform $\frac{M^2}{4}$ transpositions of squares connecting the two rectangles $R_j^2, R_{j'}^2$ as in Figure 3. More precisely we define for $t \in \left[\frac{1}{2}, \frac{3}{4}\right]$

$$X_t^c(x) = X_t \circ T_t^4(x), \quad t \in \left[\frac{1}{2}, \frac{3}{4}\right],$$

where the flow $T^4 : \left[\frac{1}{2}, \frac{3}{4}\right] \times K \to K$

$$T_t^4 \llcorner_{R_h^4} \doteq T_{4t-2}(\kappa_1^j, \kappa_1^{j'}) \quad \text{and} \quad T_t^4(x) = x \text{ otherwise},$$

is the transposition flow (4.1) between κ_1^j and $\kappa_1^{j'}$. Again,

$$\text{Tot.Var.}(b_t^c - b_t)(\kappa_1) \leq \mathcal{O}(1) 4 \frac{M^2}{4} \frac{20}{M^2 D^2}.$$

Repeating the procedure (see Figure 3),

1. at the $2i-1$-th step we divide our initial square κ_1 into $\frac{M^2}{2^{2i-1}}$ rectangles (made of two squares of obtained at the step $2(i-1)$) so that we perform 2^{2p-i} transpositions of subsquares κ_1^j in the time interval $\left[\sum_{j=1}^{2i-2} \frac{1}{2^j}, \sum_{j=1}^{2i-1} \frac{1}{2^j}\right]$;

2. at the $2i$-th step, we divide our initial square κ_1 into $\frac{M^2}{2^{2i}}$ squares (made of 2 rectangles of the previous step) so that we perform 2^{2p-i} transpositions of subsquares κ_1^j in the time interval $\left[\sum_{j=1}^{2i-1} \frac{1}{2^j}, \sum_{j=1}^{2i} \frac{1}{2^j}\right]$.

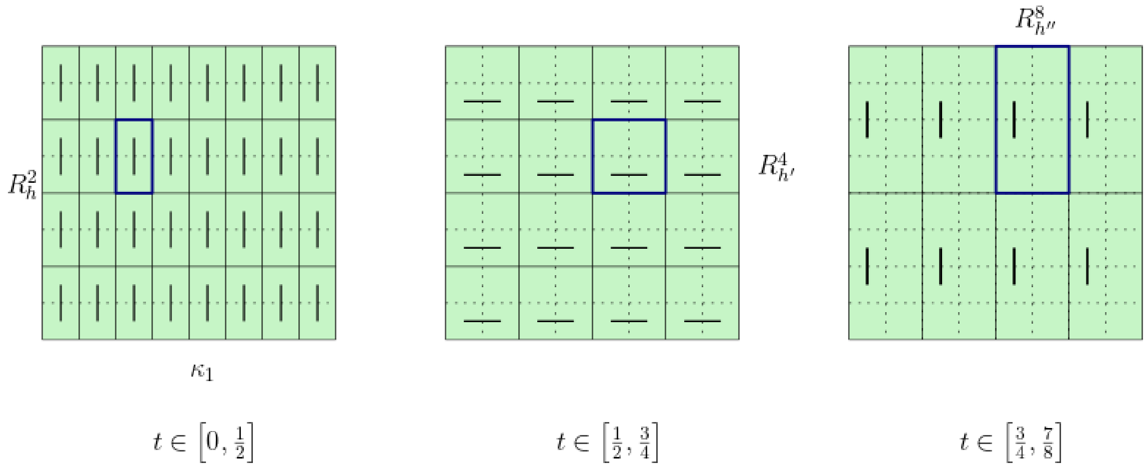


Figure 3: Subdivision of the initial square κ_1 into subrectangles/subsquares where transpositions (the bars) occurr between subsquares $\kappa_i^j \subset R_j^{2i}, \kappa_i^{j'} \subset R_{j'}^{2i}$ of side $\frac{1}{DM}$ (dotted lines). Notice that at the first/third step the initial square κ_1 is divided into rectangles (see Point (1) of the procedure), while at the second step it is divided into squares (see Point (2)).

In both cases we find in the interval $\left[\sum_{j=1}^{i-1} \frac{1}{2^j}, \sum_{j=1}^{i} \frac{1}{2^j}\right]$ that

$$\text{Tot.Var.}(b_t^c - b_t)(\kappa_1) \leq \mathcal{O}(1) 2^i \frac{M^2}{2^i} \frac{20}{M^2 D^2}.$$

Call $t_i = \sum_{j=1}^{i} 2^{-j}$. We will prove that the map $X(1, t_i) \circ X^c \llcorner_{t=t_i}$ is a permutation given by the product of $\frac{M^2}{2^i}$ disjoint $2^i k$-cycles simply by induction on i.

The case $i = 1$ is immediate from the definition. So let us assume that the property is valid for i and call $c_1, c_2, \ldots, c_{\frac{M^2}{2^i}}$ the disjoint $2^i k$-cycles made of rectangles of subsquares as in Figure 3, where we have ordered them in such a way that c_{2h-1}, c_{2h} with $h = 1, \ldots, \frac{M^2}{2^{i+1}}$ are adjacent along the long side. Then fix a couple of adjacent cycles, for simplicity c_1, c_2. Then

$$\begin{aligned} c_1 &= (\kappa_1^1 \quad \ldots \quad \kappa_{2^i k}^1), \\ c_2 &= (\kappa_1^2 \quad \ldots \quad \kappa_{2^i k}^2), \end{aligned}$$

and assume that there exist j, j' such that $\kappa_j^1, \kappa_{j'}^2$ are the adjacent subsquares in which we perform the transposition. By simply observing that

$$X^c \llcorner_{t=t_{i+1}}(x) = X^c \llcorner_{t=t_i}(x) + \int_{t_i}^{t_{i+1}} b^c(s, X_s^c(x)) ds,$$

we deduce that, when restricted to $c_1 \cup c_2$, the map $X(1, t_{i+1}) \circ X^c \llcorner_{t_{i+1}}$ is the following permutation

$$\begin{pmatrix} \kappa_1^1 & \ldots & \kappa_{j-1}^1 & \kappa_j^1 & \ldots & \kappa_{2^i k}^1 & \kappa_1^2 & \ldots & \kappa_{j'-1}^2 & \kappa_{j'}^2 & \ldots & \kappa_{2^i k}^2 \\ \kappa_2^1 & \ldots & \kappa_{j'}^2 & \kappa_{j+1}^1 & \ldots & \kappa_1^1 & \kappa_2^2 & \ldots & \kappa_j^1 & \kappa_{j'+1}^2 & \ldots & \kappa_1^2 \end{pmatrix}.$$

Clearly this is a single cycle of length $2^{(i+1)k}$, and it is supported on a rectangle.

The procedure stops at $t = \sum_{j=1}^{2p} \frac{1}{2^j}$ when we have obtained a unique $M^2 k$-cycle. Summing up, for t fixed

$$\text{Tot.Var.}(b_t^c - b)(K) \leq \mathcal{O}(1) \frac{20}{D^2},$$

that is

$$||\text{Tot.Var.}(b_t^c - b_t)(K)||_\infty \leq \mathcal{O}\left(\frac{1}{D^2}\right).$$

We conclude with the $L_t^\infty L_x^1$ estimate of the vector field: to do this computation it is necessary to observe that b_t and b_t^c differ only in the couples of adjacent squares in which we perform the transpositions. Using (2.2) and simple estimates on the rotation (2.4) we obtain, for $t \in [t_{i-1}, t_i]$ fixed,

$$\|b^c_t - b_t\|_1 \le \mathcal{O}(1)\frac{M^2}{2^i}2^i \frac{2}{DM}\frac{2}{D^2M^2} \le \mathcal{O}\left(\frac{1}{D^3M}\right),$$

which concludes the proof. □

We state now the approximation result by vector fields whose flow at time $t = 1$ is a unique cycle.

Proposition 4.5. *Let $b \in L^\infty_t(\mathrm{BV}_x)$ be a divergence-free vector field and assume that $b_t = 0$ for $t \in [0, \delta]$, $\delta > 0$, and its flow at time $t = 1$, namely $X\llcorner_{t=1}$, is a permutation of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ where $D \in \mathbb{N}$. Then for every $M = 2^p \gg 1$ there exists a divergence-free vector field $b^c \in L^\infty_t(\mathrm{BV}_x)$ such that*

$$||b - b^c||_{L^1(L^1)} \le \mathcal{O}\left(\frac{1}{DM^3}\right), \quad ||\mathrm{Tot.Var.}(b^c_t - b_t)(K)||_\infty \le \mathcal{O}\left(\frac{1}{\delta M^2}\right)$$

and the map $X^c_1 : K \to K$, being $X^c_t : [0,1] \times K \to K$ is the flow associated with b^c, is a M^2D^2-cycle of subsquares of size $\frac{1}{DM}$.

Proof of Proposition 4.5. Let us fix $\epsilon > 0$ and consider $M = 2^p$ to be chosen later. Let $C \doteq X\llcorner_{t=1}$ be a permutation, which we write by Theorem (4.3)

$$C = (\kappa^1_1 \dots \kappa^1_{k_1})(\kappa^2_1 \dots \kappa^2_{k_2}) \dots (\kappa^n_1 \dots \kappa^n_{k_n}) = c_1 \dots c_n,$$

where $\sum_{i=1}^n k_i \le D^2$. Define $c_{n+1}, \dots c_N$, $N = D^2 - \sum_i k_i + n$, the 1-cycles representing the subsquares that are sent into themselves. By the previous lemma we can also assume that $C\llcorner_{c_i}$ $i = 1, \dots, N$ is a cyclic permutation of subsquares a^i_{jk}, $j = 1, \dots, M^2$, of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{MD}$. To find a D^2M^2 cycle we should consider all the couples of adjacent subsquares (of size $\frac{1}{MD}$), and then we should connect them by transpositions in a precise way.

Fix c_1 and consider

$$C^1 = \{c_h \neq c_1 \ \text{ s.t. } c_h \text{ adjacent to } c_1\} = \{c^1_1, \dots, c^1_{|C^1|}\}.$$

Now for every $c^1_j \in C^1$ define by induction the disjoint families of cycles

$$C^2_j = \{c_h \notin \{c_1\} \cup C^1 \cup C^2_1 \cup \dots \cup C^2_{j-1} \text{ s.t. } c_h \text{ is adjacent to } c^1_j\},$$

and call

$$C^2 = C^2_1 \cup \dots \cup C^2_{|C^1|} = \{c^2_1, \dots, c^2_{|C^2|}\}.$$

At the $i-1$-th step we have

$$C^{i-1} = \{c^{i-1}_1, \dots, c^{i-1}_{|C^i|}\}.$$

and, for every $c^{i-1}_j \in C^{i-1}$,

$$C^i_j = \{c_h \notin \{c_1\} \cup C^1 \cup C^2 \cup \dots \cup C^{i-1} \cup C^i_1 \cup \dots C^i_{j-1} \text{ s.t. } c_h \text{ is adjacent to } c^{i-1}_j\}.$$

The procedure ends when we have arranged all c_i into sets C^i, and hence for some $K \in \mathbb{N}$ we obtain $C^{K+1} = \emptyset$ (see Figure 4). Indeed, by contradiction assume that $|\{c_1\} \cup C^1 \cup C^2 \cup \dots \cup C^K| < N$. Then this set has a boundary, i.e., there exists a cycle $c \notin \{c_1\} \cup C^1 \cup C^2 \cup \dots \cup C^K$ adjacent to a cycle of $\{c_1\} \cup C^1 \cup C^2 \cup \dots \cup C^K$, which is a contradiction by definition.

The partition

$$C(c_1) = \{c_1\} \cup C^2 \cup \dots C^K$$

has the natural structure of a directed tree: indeed every two cycles $c_i \in C^i$, $c_j \in C^j$ are connected by a unique sequence of cycles: the direction of each edge is given by the construction $c^{i-1}_j \to c_h$ whenever $c_h \in C^i_j$. This tree-structure gives us a selection of the $N-1$ couples of subsquares κ^j_i of disjoint cycles c_i in which we can perform a transposition among the subsquares a^i_{jk} to connect all of them in a unique D^2M^2-cycle. More precisely, for every connected couple c^{i-1}_j, c_h such that $c_h \in C^i_j$, there exist cubes $\kappa \in c^{i-1}_j, \kappa' \in c_h$, and hence there are adjacent subsquares $a \subset \kappa, a' \subset \kappa'$ of size $1/(MD)$: assuming $M \ge 4$, we can take a, a' not being on the corners of κ, κ', respectively.

Let $T_t : [0, \delta] \times K \to K$ be the transposition flow (4.1) acting in the selected $N-1$ couples of subsquares a, a' reparametrized on the time interval $[0, \delta]$ and define (being $X_t = \mathrm{id}$ for $t \in [0, \delta]$)

$$X_t^c(x) = \begin{cases} T_t(x) & t \in [0, \delta], \\ X_t \circ T_\delta(x) & t \in [\delta, 1]. \end{cases} \tag{4.3}$$

The transposition is well defined: indeed it can happen that c_i, c_j, c_k are adjacent cycles and the couples of adjacent squares (of size $\frac{1}{D}$) are κ_i, κ_j and κ_i, κ_k (where $\kappa_i \in c_i, \kappa_j \in c_j$ and $\kappa_k \in c_k$), that is: κ_i is in common. But since the transposition occurs between subsquares of size $\frac{1}{DM}$ nor belonging to the corners, it is always guaranteed that the transpositions act on disjoint subsquares. By using the explicit formula (2.4) we get that for $t \in [0, \delta]$

$$||b_t^c - b_t||_1 \leq \frac{\mathcal{O}(1)}{\delta} \left(\frac{N-1}{D^3 M^3} \right) \leq \frac{\mathcal{O}(1)}{\delta} \left(\frac{1}{DM^3} \right),$$

while for $t \in [\delta, 1]$ it clearly holds $b_t^c = b_t$. Coupling these last two estimates we get the $L_t^1 L_x^1$ estimate:

$$||b - b^c||_{L^1(L^1)} \leq \mathcal{O}\left(\frac{1}{DM^3} \right),$$

for $\delta << 1$ and M sufficiently large.

Next we compute the total variation for $t \in [0, \delta]$: by using (4.2), we get

$$\mathrm{Tot.Var.}(b_t^c)(K) \leq \frac{N-1}{\delta} \frac{20}{M^2 D^2} \leq \frac{1}{\delta} \frac{20}{M^2},$$

while for $t \in [\delta, 1-\delta]$ we find

$$\mathrm{Tot.Var.}(b_t^c)(K) = \mathrm{Tot.Var.}(b_t)(K),$$

therefore

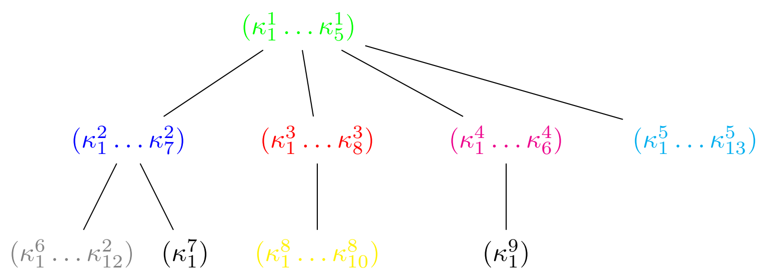
$$||\mathrm{Tot.Var.}(b_t^c - b_t)(K)||_\infty \leq \mathcal{O}\left(\frac{1}{\delta M^2} \right).$$

To conclude we have to prove that X_1^c is a unique cycle, which follows by the tree-structure of the selection of adjacent cycles. The end points of the tree are clearly cycles. By recurrence, assume that c_j^{i-1} is connected to cycles γ_h, each one made of all squares belonging to $c_h \in C_j^i$ and all subsequent cycles to c_h. It is fairly easy to see that the transposition merging c_j^{i-1} to each $c_h \in C_j^i$ generates a unique cycles γ_j^i, made of the cubes of c_j^{i-1} and all γ_h. We thus conclude that the map X_1^c is a cycle of size $M^2 D^2$. $\square$

Remark 4.6. An example of how the proof works is in Figure 4: the decomposition in cycles is

$$C = (\kappa_1^1 \dots \kappa_5^1)(\kappa_1^2 \dots \kappa_7^2)(\kappa_1^3 \dots \kappa_8^3)(\kappa_1^4 \dots \kappa_6^4)(\kappa_1^5 \dots \kappa_{13}^5)(\kappa_1^6 \dots \kappa_{12}^2)(\kappa_1^7)(\kappa_1^8 \dots \kappa_{10}^8)(\kappa_1^9)(\kappa_1^{10}).$$

The black arrow indicates the adjacent subsquares where the exchanges are performed: the tree of concatenation is then



Note that in the subsquares $(\kappa_2^1, \kappa_1^2), (\kappa_2^1, \kappa_1^3)$ and $(\kappa_5^1, \kappa_1^4), (\kappa_5^1, \kappa_1^5)$ the exchange occurs actually in the subsquares $(a_{2j}^1, a_{1\ell}^2), (a_{2j'}^1, a_{1\ell'}^3)$ and $(a_{5k}^1, a_{1\ell''}^4), (a_{5k'}^1, a_{1,\ell'''}^5)$, so that it is always acting on different couples of subsquares.

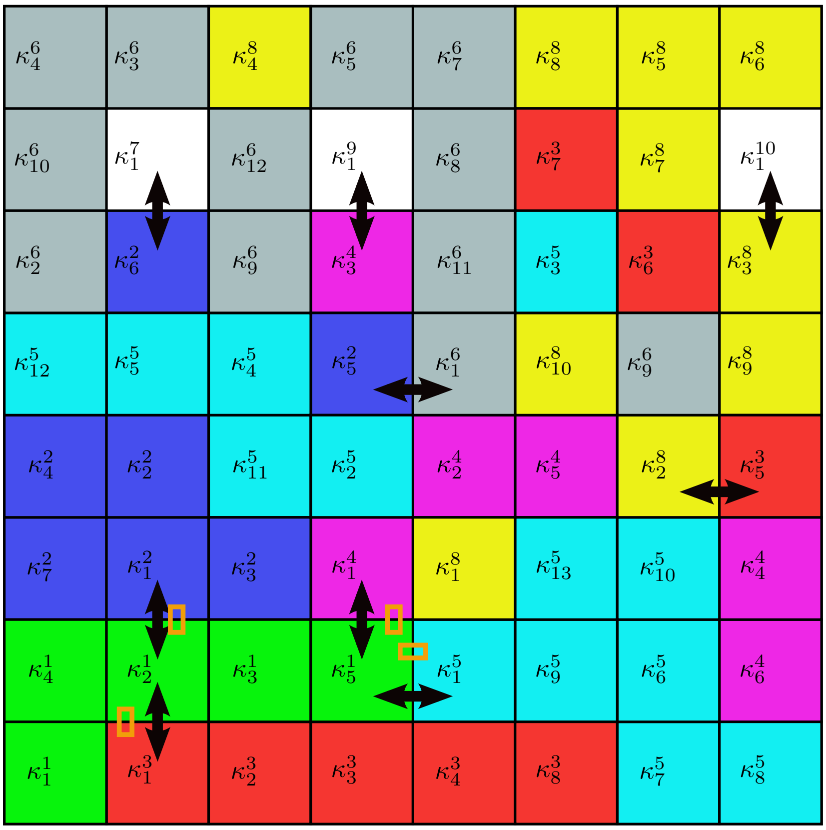


Figure 4: Concatenation of cycles in a specific example, Remark 4.6. The orange subrectangles are the couples of a^i_{jk} on which the transposition T_t of (4.3) acts.

Remark 4.7. The construction of the cyclic flow (4.3) gives us only the $L^1_t L^1_x$ estimate on the vector fields, which is what we need for our genericity result. We can get the more refined estimate in $L^\infty_t L^1_x$ allowing for mass flowing (when performing the transposition) during the time evolution of the flow X_t (see Figure 5). In this case, the time spent by the squares of size $(MD)^{-1}$ to transfer the mass is of order $(MD)^{-1}$, so that the vector field moving it should be of the order

$$\frac{\text{length}}{\text{time}} = \mathcal{O}(1), \quad \text{acting on a region of area } \frac{N-1}{(MD)^2} \leq M^{-2}. \tag{4.4}$$

Hence the $L^\infty_t L^1_x$ estimate can be obtained by (2.2) as

$$\|b^c_t - b_t\|_1 \leq \mathcal{O}(1) M^{-2},$$

while the total variation estimate becomes

$$\text{Tot.Var.}(b^c_t - b_t) = \mathcal{O}(1)\frac{D}{M}.$$

The statement one can prove is then the following.

Proposition 4.8. *Let $b \in L^\infty_t(\mathrm{BV}_x)$ be a divergence-free vector field and assume that its flow at time $t = 1$, namely $X_{\llcorner t=1}$, is a permutation of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ where $D \in \mathbb{N}$. Then for every $\epsilon > 0$ there exist $M = 2^p$ and $b^c \in L^\infty_t(\mathrm{BV}_x)$ a divergence-free vector field such that*

$$||b^c_t - b_t||_{L^1} \leq \mathcal{O}(M^{-2}) \leq \epsilon, \quad \text{Tot.Var.}(b^c_t - b_t)(K) \leq \mathcal{O}(D/M),$$

and the map $X^c_{t=1} : K \to K$, being $X^c_t : [0,1] \times K \to K$ the flow associated with b^c, is a M^2D^2-cycle of subsquares of size $\frac{1}{DM}$.

4.2 Density of ergodic vector fields

Starting from the cyclic permutation we have built in the previous section, we construct an ergodic vector field arbitrarily close to a given vector field in $L^\infty_t \mathrm{BV}_x$. The density of ergodic vector fields is

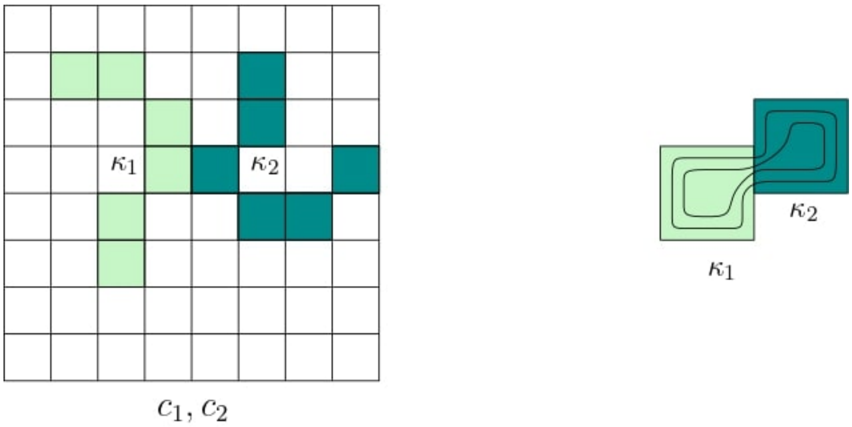


Figure 5: The two adjacent cycles c_1 (light green) and c_2 (blue) touch in κ_1 and κ_2, which exchange their mass during the time evolution.

not strictly relevant for the genericity result of weakly mixing vector fields, but it can be considered as a simple case study for the construction of strongly mixing vector fields. Moreover it will give a direct proof of Point (2) of Theorem 1.2.

We will use the *universal mixer* that has been constructed in [**univ:mixer**]: it is the time periodic divergence-free vector field $u \in L^\infty_t([0,1], BV_x(\mathbb{R}^2))$ whose flow $U_t : [0,1] \times K \to K$ of measure-preserving maps realizes at time $t = 1$ the *folded Baker's map*, that is

$$U = U\llcorner_{t=1} = \begin{cases} \left(-2x+1, -\frac{y}{2} + \frac{1}{2}\right) & x \in \left[0, \frac{1}{2}\right), \\ \left(2x-1, \frac{y}{2} + \frac{1}{2}\right) & x \in \left(\frac{1}{2}, 1\right], \end{cases} \qquad y \in [0,1], \tag{4.5}$$

(see Theorem 1, [**univ:mixer**]).

Proposition 4.9. *Let $b \in L^\infty_t(\mathrm{BV}_x)$ and let X_t be its RLF, and assume that $X_{t=1}$ is a cyclic permutation of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$. Then there exists $b^e \in L^\infty_t(\mathrm{BV}_x)$ divergence-free ergodic vector field such that*

$$\begin{aligned} &\|b - b^e\|_{L^\infty(L^1)} \le \mathcal{O}\left(\frac{1}{D^2}\right), \\ &\|\mathrm{Tot.Var.}(b^e)(K)\|_\infty \le \|\mathrm{Tot.Var.}(b)(K)\|_\infty + \mathcal{O}\left(\frac{1}{D^2}\right). \end{aligned} \tag{4.6}$$

Proof. Let us call $T = X\llcorner_{t=1}$ and $\kappa_1, \dots, \kappa_{D^2}$ the subsquares of the grid where the numbering is chosen such that

$$T(\kappa_i) = \kappa_{i+1}, \quad T(\kappa_n) = \kappa_1.$$

Let us define

$$X^e_t = \begin{cases} X_t \circ U^1_t & x \in \kappa_1, \\ X_t & \text{otherwise}, \end{cases}$$

where the flow $U^1_t = \theta^{-1} \circ U_t \circ \theta$ and θ is the affine map from κ_1 to K, i.e. ,$\theta(x,y) = (Dx, Dy)$.

We first prove the ergodicity of $T^e = X^e\llcorner_{t=1}$. Assume by contradiction that T^e is not ergodic, then there exists a measurable set B such that $T^e(B) = B$ and $0 < |B| < 1$. We claim that $|B \cap \kappa_1| > 0$. Indeed, since $|B| > 0$ there exists i such that $|B \cap \kappa_i| > 0$. If $i = 1$ we have nothing to prove, if not, since T^e is measure-preserving, then $|T^e(B \cap \kappa_i)| > 0$. But

$$0 < |T^e(B \cap \kappa_i)| = |T^e(B) \cap T^e(\kappa_i)| = |B \cap \kappa_{i+1}|$$

(we have used that the set B is invariant) and re-applying the map T^e sufficiently many times we have the claim. Moreover, $|B \cap \kappa_1| < \frac{1}{D^2}$. If not, that is $|B \cap \kappa_1| = \frac{1}{D^2}$, then $|B \cap \kappa_i| = \frac{1}{D^2}$ for every $i = 1, \dots, D^2$, again by using the fact that B is invariant and that $T^e(\kappa_i) = \kappa_{i+1}$ and $T^e(\kappa_{D^2}) = \kappa_1$. But now

$$|B| = \sum_{i=1}^{D^2} |B \cap \kappa_i| = \sum_{i=1}^{D^2} \frac{1}{D^2} = 1,$$

which is a contradiction, since $|B| < 1$. Now, the fact that $0 < |B \cap \kappa_1| < \frac{1}{D^2}$ implies that $U_1^1(B \cap \kappa_1) \neq B \cap \kappa_1$ because U_1^1 is mixing (and thus ergodic). But this is a contradiction because $T^e(B \cap \kappa_1) = B \cap \kappa_2$ and applying to both of them T^{D^2-1} we find that

$$U_1^1(B \cap \kappa_1) = T^{D^2-1}(T^e(B \cap \kappa_1)) = T^{D^2-1}(B \cap \kappa_2) = (T^e)^{D^2-1}(B \cap \kappa_2) = B \cap \kappa_1,$$

where we have used that $T^{D^2} = Id$. To prove the estimates (4.6) we have to observe first that U_t acts only on κ_1, then that it is the composition of two rotations (see [Figure 1, [**univ:mixer**]]), that is Tot.Var.$(\dot{U}_t(U_t^{-1}))(\bar{\kappa}_1) \leq \mathcal{O}\left(\frac{1}{D^2}\right)$ (see again Lemma 2.6). ☐

4.3 Density of strongly mixing vector fields

As in the previous section, we use the density of cyclic permutations to show that the vector fields whose flow is strongly mixing are dense in $\mathcal{U}$ with the $L^1_{t,x}$-topology. Again we use the universal mixer constructed in [**univ:mixer**]. The main result here is the following

Proposition 4.10. *Let $b \in L^\infty_t(\mathrm{BV}_x)$ and let X_t be its RLF, and assume that $b_t = 0$ for $t \in [0, 2\delta]$ and $X_{t=1}$ is a cyclic permutation of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$, $D = 2^p$. Then there exists $b^s \in L^\infty_t(\mathrm{BV}_x)$ divergence-free strongly mixing vector field such that*

$$\begin{aligned} &\|b - b^s\|_{L^1(L^1)} \leq \mathcal{O}\left(\frac{1}{\delta D}\right), \\ &\|\mathrm{Tot.Var.}(b^s)(K)\|_\infty \leq \|\mathrm{Tot.Var.}(b)(K)\|_\infty + \mathcal{O}\left(\delta^{-1}\right). \end{aligned} \tag{4.7}$$

In the proof it is shown that the mixing is actually exponential, in the sense that for every set in a countable family of sets $\{B_i\}_i$ generating the Borel σ-algebra it holds

$$\left|T^q(B_i) \cap B_j\right| - |B_i||B_j| = \mathcal{O}(1) c_{ij}^q, \quad c_{ij} < 1.$$

Proof. Let us call $T = X_{\llcorner t=1}$ and $\kappa_1, \dots, \kappa_{D^2}$ the subsquares of the grid where the numbering is chosen such that

$$T(\kappa_i) = \kappa_{i+1}, \quad T(\kappa_{D^2}) = \kappa_1.$$

If $\{1, \dots, D^2\} \ni \ell \mapsto j(\ell) \in \{1, \dots, D^2\}$ is an enumeration of κ_i such that $\kappa_{j(\ell)}, \kappa_{j(\ell+1)}$ are adjacent, consider the rescaled universal mixer $U_t^{\ell,\ell+1}$ acting on $\kappa_\ell, \kappa_{\ell+1}$ in the time interval $[0, \delta]$, whose generating vector field $b^{U^{\ell,\ell+1}}$ satisfies the estimates

$$\|b_t^{U^{\ell,\ell+1}}\|_{L^1} = \mathcal{O}(1)\frac{1}{\delta}\frac{1}{D^3}, \quad \mathrm{Tot.Var.}(b_t^{U^{\ell,\ell+1}}) = \mathcal{O}(1)\frac{1}{\delta}\frac{1}{D^2}.$$

The idea is to define the a new vector field as in (4.3)

$$X_t^s(x) = \begin{cases} M_t(x) & t \in [0, 2\delta], \\ X_t \circ M_{2\delta}(x) & t \in [2\delta, 1], \end{cases}$$

where the map M_t, $t \in [0, 2\delta]$, is defined as follows:

$$M_t(x) = \begin{cases} U_t^{\ell,\ell+1}(x) & t \in [0, \delta], \ell \text{ even}, \\ U_t^{\ell,\ell+1}(x) & t \in [\delta, 2\delta], \ell \text{ odd}. \end{cases}$$

The estimates (4.7) follows as in Proposition 4.9, so we are left with the proof that $T^s = X_1^s$ is strongly mixing.

The map T^s is the composition of 3 maps $T_3 \circ T_2 \circ T_1$ acting as follows (all indexes should be intended modulus D^2):

1. T_1 is the folded Baker's map U acting on the couples $\ell, \ell + 1$, $\ell = 0, 2, \dots$ even;

2. T_2 is the folded Baker's map U acting on the couples $\ell, \ell + 1$, $\ell = 1, 3, \dots$ odd;

3. T_3 is a cyclic permutation $\ell \to j^{-1}(j(\ell) + 1)$.

We first compute the evolution of a rectangle a of the form

$$a = 2^{-p}[k,k+1] \times 2^{-p'}[k',k'+1]\frac{1}{D}, \quad k = 0,\dots,2^p D - 1,\ k' = 0,\dots,2^{p'}D - 1,\ p,p' \in \mathbb{N}.$$

By definition of U (4.5) we obtain that if $p \geq 1$ then the map T_1 does not split a into disjoint rectangles, i.e.

$$T_1 a = 2^{1-p}[\tilde{k},\tilde{k}+1] \times 2^{-p'-1}[\tilde{k}',\tilde{k}'+1]\frac{1}{D}, \quad \tilde{k} = 0,\dots,2^{p-1}D - 1, \tilde{k}' = 0,\dots,2^{p'+1}D - 1,$$

and the same happens for T_2:

$$T_2 a = 2^{1-p}[\hat{k},\hat{k}+1] \times 2^{-p'-1}[\hat{k}',\hat{k}'+1]\frac{1}{D}, \quad \hat{k} = 0,\dots,2^{p-1}D - 1, \hat{k}' = 0,\dots,2^{p'+1}D - 1.$$

Hence if

$$a = 2^{-2p}[k,k+1] \times 2^{-2p'}[k',k'+1]\frac{1}{D}, \quad k = 0,\dots,2^{2p}D - 1,\ k' = 0,\dots,2^{2p'}D - 1,\ p,p' \in \mathbb{N}, \tag{4.8}$$

then

$$T_2 \circ T_1 a = 2^{2(1-p)}[\check{k},\check{k}+1] \times 2^{-2(p'+1)}[\check{k}',\check{k}'+1]\frac{1}{D}, \quad \check{k} = 0,\dots,2^{2(p-1)}D - 1, \check{k}' = 0,\dots,2^{2(p'+1)}D - 1,$$

and being the action of T_3 just a permutation, the final form $T^s a = T_3 \circ T_2 \circ T_1 a$ is again a rectangle. When $p = 0$, instead the rectangle a is mapped into two rectangles belonging to two different subsquares κ, κ'

$$T_1 a = [\tilde{k}_1,\tilde{k}_1+1] \times 2^{-p'-1}[\tilde{k}_1',\tilde{k}_1'+1]\frac{1}{D} \cup [\tilde{k}_2,\tilde{k}_2+1] \times 2^{-p'-1}[\tilde{k}_2',\tilde{k}_2'+1]\frac{1}{D},$$

and the action of T_2 divides $T_1 a$ into 4 rectangles of horizontal length $1/D$ belonging to 4 different subsquares. As before, T_3 just shuffles them into new locations.
The same happens when considering $(T^s)^{-1}$: if $p' \leq 1$ and a is given by (4.8) then $(T^s)^{-1}a$ is still a rectangle of side $2^{-2(p+1)} \times 2^{-2(p'-1)}\frac{1}{D}$, while for $p' = 0$ it is split into 4 rectangles with vertical size equal to $1/D$.

In particular, starting from two squares a, a' of side $(2^{-2p}D)^{-2}$, for $q \geq p$ the set $(T^s)^q a$ is made of disjoint rectangles whose horizontal side is D^{-1}, and $(T^s)^{-q}a'$ is made of disjoint rectangles whose vertical side is D^{-1}. Hence if the masses of $(T^s)^q a$, $(T^s)^{-q'}a'$ inside κ_i are $m_i(q)$, $m_i'(-q')$, then by Fubini

$$\mathcal{L}^2\big((T^s)^q a \cap (T^s)^{-q'}a'\big) = \sum_{i=1}^{D^2} D^2 m_i(q) m_i'(-q').$$

In order to prove the strong mixing it is enough to show that

$$m_i(q) \to \frac{\mathcal{L}^2(a)}{D^2},\ m_i'(-q') \to \frac{\mathcal{L}^2(a')}{D^2} \quad q,q' \to \infty.$$

Actually, we will show that the above convergence is exponential, which implies that the mixing is exponential. We prove the above exponential convergence for $m_i(q)$, the other being completely similar.

Once $(T^s)^q a$ has become a rectangle of horizontal side $1/D$, the distribution of mass by T^s is computed by the action of the following matrices on the vector $(m_i)_i$:

1. the matrix A_1 corresponding to the map T_1,

$$(A_1)_{\ell'\ell} = \frac{1}{2}\begin{cases} \delta_{\ell'\ell} + \delta_{\ell'(\ell-1)} & \ell' = 0,2,\dots, \\ \delta_{\ell'(\ell+1)} + \delta_{\ell'\ell} & \ell' = 1,3,\dots; \end{cases}$$

2. the matrix A_2 corresponding to the map T_2,

$$(A_2)_{\ell'\ell} = \frac{1}{2}\begin{cases} \delta_{\ell'\ell} + \delta_{\ell'(\ell+1)} & \ell' = 0,2,\dots, \\ \delta_{\ell'(\ell-1)} + \delta_{\ell'\ell} & \ell' = 1,3,\dots; \end{cases}$$

3. the permutation matrix A_3 corresponding to T_3.

Being the Markov process generated by the matrix $P = A_3A_2A_1$ finite dimensional, exponential mixing is equal to strong mixing, and we prove directly that P has a simple eigenvalue of modulus 1 whose eigenvector is necessarily the uniform distribution $(1/D^2, 1/D^2, \dots)$: in particular this gives that P is aperiodic (Definition 3.13 and Proposition 3.15). Indeed, for $v \in \mathbb{C}^D$ one considers the functional $|v|$, and by simple computations it holds $|A_3v| = |v|$ and

$$|A_1v| = |v| \quad \text{iff} \quad v_\ell = v_{\ell+1} \text{ for } \ell = 0, 2, \dots,$$

$$|A_2v| = |v| \quad \text{iff} \quad v_\ell = v_{\ell+1} \text{ for } \ell = 1, 3, \dots.$$

Hence the unique v such that $|Av| = |v|$ is $v = (1/D^2, 1/D^2, \dots)$, and 1 is a simple eigenvector. □

Remark 4.11. As in Remark 4.7, one could let the Bakers map to act during the time evolution of X_t, but in this case the distance in $L^\infty L^1$ would be of order 1. The problem is that the maps T_1, T_2 are acting on the whole set $K = [0,1]^2$, and the vector field $b^s_t - b_t$ is of order 1 as in (4.4).

4.4 Proof of the density of strongly mixing vector fields

We are now ready to prove the density of strongly mixing vector fields in $\mathcal{U}$, which implies the statement by Corollary 3.10. It will be obtained through the following steps.

1. Let $b \in \mathcal{U}$: by the very construction of the set $\mathcal{U}$ (Proposition 2.3), we can assume that $b \in L^\infty_t \mathrm{BV}_x$. Fix $\epsilon > 0$.

2. By the continuity of translation in L^1, we can take $0 < \delta \ll 1$ such that defining
$$b^\delta = \begin{cases} 0 & t \in [0, 3\delta), \\ \frac{1}{1-3\delta} b_{(t-3\delta)/(1-3\delta)} & t \in [3\delta, 1], \end{cases}$$
it holds
$$\|b^\delta - b\|_{L^1_{t,x}} < \frac{\epsilon}{4}.$$
Since
$$\|\mathrm{Tot.Var.}(b^\delta)\|_\infty = \frac{1}{1-3\delta} \|\mathrm{Tot.Var.}(b)\|_\infty$$
then $b^\delta \in \mathcal{U}$. Clearly we can also assume that b^δ is compactly supported in K.

3. Use Theorem 5.14 to approximate b^δ in $[3\delta, 1]$ with a vector field $b^{\epsilon\delta} \in L^\infty_t \mathrm{BV}_x \subset \mathcal{U}$ such that
$$\|b^\delta - b^{\epsilon\delta}\|_{L^1_{t,x}} < \frac{\epsilon}{4},$$
and such that its RLF is a permutation of squares of size D^{-1}. We can assume that
$$D \gg \frac{1}{\epsilon\delta}. \tag{4.9}$$

4. Apply Lemma 4.4 together with Proposition 4.5 to $b^{\epsilon\delta}$ for $t \in [2\delta, 1]$ obtaining a new vector field $b^{\epsilon\delta c} \in L^\infty_t \mathrm{BV}_x \subset \mathcal{U}$ such that
$$\|b^{\epsilon\delta} - b^{\epsilon\delta c}\|_{L^1_{t,x}} \leq \mathcal{O}\biggl(\frac{1}{DM}\biggr) < \frac{\epsilon}{4}$$
for $M = 2^{p'} \gg 1$, and such that its RLF is a single cycle of squares of size $(DM)^{-1}$.

5. Finally, apply Proposition 4.10 to $b^{\epsilon\delta c}$ in $t \in [0,1]$ obtaining a strongly (exponentially) mixing vector field $b^{\epsilon\delta cs} \in L^\infty_t \mathrm{BV}_x \subset \mathcal{U}$ such that
$$\|b^{\epsilon\delta c} - b^{\epsilon\delta cs}\|_{L^1_{t,x}} \leq \mathcal{O}\biggl(\frac{1}{\delta D}\biggr) < \frac{\epsilon}{4}$$
by using (4.9).

We thus conclude that for every $b \in L^\infty_t \mathrm{BV}_x$ and $\epsilon > 0$ there is a vector field $b^s \in L^\infty_t \mathrm{BV}_x$ exponentially mixing such that
$$\|b - b^s\|_{L^1_{t,x}} < \epsilon,$$
which is our aim.

5 Permutation Flow

In this section we prove the key tool of this paper, namely the approximation in L^1 of any BV vector field with another BV vector field such that its flow at $t = 1$ is a permutation of subsquares, i.e., it is a rigid translation of subsquares of a grid partition of $K = [0,1]^2$. The approach is inspired by [**Shnirelman**], with the additional difficulty that we need to control the BV norm of the approximating vector field. We will address also the d-dimensional case, explaining the additional technicalities needed to prove the same approximation result in the general case.
This section is divided into two parts: in the first one we collect some preliminary estimates which will be used as building blocks in the proof of the main theorem, while in the second part we state the main approximation theorem and give its proof.

5.1 Affine approximations of smooth flows

The next lemma is almost the same of [**Shnirelman**]. In order to follow the original Shnirelman's Lemma we require the subrectangles in the next lemma to be dyadic (i.e., their corners belong to a dyadic partition, see Remark 5.3 however), but we notice that the proof of the main theorem works in the same way just asking subrectangles with rational coordinates to be mapped affinely onto subrectangles with rational coordinates. At the end this section we will address the same lemma in the general case $d > 2$, which in the original paper is not proved.
Let T be a measure-preserving diffeomorphism $T : [0,1]^2 \to [0,1]^2$ of class C^3 and such that $T = \mathrm{id}$ in a neighborhood of $\partial[0,1]^2$. Assume that it is close to the identity, i.e., there exists $\delta > 0$ sufficiently small such that $||T - \mathrm{id}||_{C^1} \leq \delta$.

Lemma 5.1. *There exists $N \in \mathbb{N}$, $N = 2^p$, and a path of measure-preserving invertible maps $t \to \sigma_t$ piecewise smooth w.r.t. the time variable t such that $\sigma_0 = T$ and σ_1 maps arbitrarily small dyadic rectangles $P_{ij} \in \mathbb{N}\times\mathbb{N}\frac{1}{N} = K_N$ (meaning that their boundaries are in the net K_N) affinely onto dyadic rectangles $\tilde{P}_{ij} \in K_N$.*
Moreover, the map σ is of the form

$$\sigma_t = T \circ \xi_{3t}\mathbf{1}_{[0,1/3]}(t) + \zeta_{3t-1} \circ T \circ \xi_1\mathbf{1}_{[1/3,2/3]}(t) + \eta_{3t-2} \circ \zeta_1 \circ T \circ \xi_1\mathbf{1}_{[2/3,1]}(t). \tag{5.1}$$

where $\xi, \eta : [0,1] \times [0,1]^2 \to [0,1]^2$ are piecewise smooth and $\zeta : [0,1] \times [0,1]^2 \to [0,1]^2$ is smooth, so that for every $t \in [0,1]$, the map σ_t is piecewise smooth on each subrectangle κ and it extends continuously on $\bar{\kappa}$.
Finally, the space differential $D\sigma_1$ of $\sigma_{t=1}$ is a constant diagonal matrix in each subrectangle.

The number N is used in the next results in order to have that the perturbation is arbitrarily small in $L^1_{t,x}$.

Proof. The proof is given in 3 steps:

1. first by an arbitrarily small perturbation of the final configuration we make sure the area of the regions which will be mapped into rectangles is dyadic;
2. secondly we perturb along horizontal slabs in order to have that vertical sections of the slabs are mapped into vertical segments;
3. finally we perturb vertical slabs so that the image of particular rectangles are rectangles and vertical segments remains vertical segments.

The composition of all 3 maps with T as in (5.1) will be the movement σ_t. We will use the notation

$$[0,1]^2 \ni (x_1, x_2) \mapsto T(x_1, x_2) = (z_1, z_2) \in [0,1]^2$$

to avoid confusion between the final coordinates and the initial ones. When piecing together maps which are defined in closed sets with piecewise regular boundaries, we will neglect the negligible superposition of boundaries for simplicity: this slight inaccuracy should not generate confusion.
Step 0: initial grid and perturbation. For $N_0 = 2^{p_0} \gg 1$ define the horizontal and vertical slabs

$$H_j = [0,1] \times 2^{-p_0}[j-1, j], \quad V_i = 2^{-p_0}[i-1, i] \times [0,1], \quad i, j = 1, \dots, 2^{p_0}.$$

The image of the horizontal lines

$$x_1 \mapsto T(x_1, x_2)$$

can be written as graphs of functions

$$z_1 \mapsto g(z_1, x_2),$$

and divides every vertical slab V_i into $N_0 = 2^{p_0}$ parts

$$\tilde{\omega}_{ij} = \Big\{ (i-1)2^{-p_0} \leq z_1 \leq i2^{-p_0}, g(z_1, (j-1)2^{-p_0}) \leq z_2 \leq g(z_1, j2^{-p_0}) \Big\}.$$

Let $\zeta_t : [0,1]^2 \to [0,1]^2$ be a measure preserving flow, moving mass across the boundary of $\tilde{\omega}_{ij}$: we can assume w.l.o.g that the mass flow $\phi_{ij,i'j'}$ across the boundary from $\tilde{\omega}_{ij}$ to $\tilde{\omega}_{i'j'}$ occurs in the relative interior of $\partial\tilde{\omega}_{ij} \cap \partial\tilde{\omega}_{i'j'}$. The measure preserving condition requires that

$$\phi_{ij,(i-1)j} + \phi_{ij,(i+1)j} + \phi_{ij,i(j+1)} + \phi_{ij,i(j-1)} = 0.$$

Set $T' = \zeta_1 \circ T$ and consider the new curves

$$z_1 \mapsto g'(z_1, x_2), \quad \operatorname{Graph} g' = T'([0,1] \times \{x_2\}).$$

Let $\tilde{\omega}'_{ij}$ be the new regions

$$\tilde{\omega}'_{ij} = \Big\{ (i-1)2^{-p_0} \leq z_1 \leq i2^{-p_0}, g'(z_1, (j-1)2^{-p_0}) \leq z_2 \leq g'(z_1, j2^{-p_0}) \Big\},$$

whose new area is

$$\mathcal{L}^2(\tilde{\omega}'_{ij}) = \phi_{ij,i(j-1)} + \phi_{ij,i(j+1)}.$$

Starting with $\tilde{\omega}'_{11}$, we move a mass $\phi_{11,12} < 2^{-p_0-p_1} \ll 1$ so that

$$\mathcal{L}^2(\tilde{\omega}'_{11}) = 2^{-p_0-p_1} n_{11} \in 2^{-p_0-p_1}\mathbb{N}.$$

Hence a mass $-\phi_{11,21}$ is flowing to the region $\tilde{\omega}_{21}$. Assuming that we have

$$\mathcal{L}^2(\tilde{\omega}'_{i1}) = 2^{-p_0-p_1} n_{i1} \in 2^{-p_0-p_1}\mathbb{N},$$

and that the mass flowing by $\phi_{i1,i2}, \phi_{i1,(i+1)1}$ is $< 2^{-p_1}$, we consider two cases:

1. if $\phi_{i1,(i+1)1} \in 2^{-p_0-p_1}[0,1)$, then we flow a mass $\phi_{(i+1)1,(i+1)2} \in 2^{-p_0-p_1}[0,1)$ so that

   $$\mathcal{L}^2(\tilde{\omega}'_{(i+1)1}) = 2^{-p_0-p_1} n_{(i+1)1} \in 2^{-p_1}\mathbb{N},$$

   and the flow to the right is then

   $$\phi_{(i+1)1,(i+2)i} = \phi_{i1,(i+1)i} - \phi_{(i+1)1,(i+1)2} \in (-1,1)2^{-p_0-p_1},$$

   by the balance and because they have different sign;
2. if $\phi_{i1,(i+1)1} \in 2^{-p_0-p_1}(-1,0)$, then we flow a mass $\phi_{(i+1)1,(i+1)2} \in 2^{-p_0-p_1}(-1,0)$ and obtain the same estimate.

The last term $\tilde{\omega}'_{N_0 1}$ is computed by conservation: indeed

$$\sum_i \phi_{i1,i2} = 0,$$

and then

$$\mathcal{L}^2(T'([0,1] \times [0, 2^{-p_0}]) = 2^{-p_0} = \sum \mathcal{L}^2(\tilde{\omega}'_{i1}) = 2^{-p_0-p_1} \sum_{i=0}^{2^{p_0}-1} n_{i1} + \mathcal{L}^2(\tilde{\omega}'_{N_0 1}),$$

so that

$$\mathcal{L}^2(\tilde{\omega}'_{N_0 1}) = 2^{-p_0-p_1} \left(2^{p_1} - \sum_{i=0}^{2^{p_0}-1} n_{i1} \right) \in 2^{-p_0-p_1}\mathbb{N}.$$

The estimate of $\phi_{N_0 1, N_0 2}$ is automatic from the flow $\phi_{(N_0-1)1, N_0 1}$.

The above procedure is then repeated for each region

$$T([0,1] \times 2^{-p_0}[j-1,j]) = \bigcup_{i=1}^{N_0} \tilde{\omega}_{ij},$$

and the flow across each boundary is $\leq 2^{-p_0-p_1}$: the conservation of the measure of $T([0,1]\times[0,j]2^{-p_0})$ yields that the last element $\tilde{\omega}_{N_0 j}$ is again dyadic.

From now on we work with the map $T' = \zeta_1 \circ T$.

Step 1: perturbation along horizontal slabs. Consider the curves

$$z_2 \mapsto (T')^{-1}(z_1, z_2),$$

which can be parameterized as

$$x_2 \mapsto f'(z_1, x_2)$$

being T' close to the identity in C^1. In each H_j we can determine uniquely the value

$$x_{1,j}(z_1) = \fint_{(j-1)2^{-p_0}}^{j2^{-p_0}} f'(z_1, x_2) dx_2, \tag{5.2}$$

and since T' is close to identity, again every map $z_1 \mapsto x_{1,j}(z_1)$ is invertible: denote its inverse by $z_{1,j}(x_1)$.

In particular, we consider the values

$$x_{1,ij} = x_{1,j}(i2^{-p_0}). \tag{5.3}$$

By (5.2) it follows that

$$\big(x_{1,ij} - x_{1,(i-1)j}\big)2^{-p_0} = \mathcal{L}^2\big((T')^{-1}(\tilde{\omega}'_{ij})\big) \in 2^{-p_0-p_1}\mathbb{N}, \tag{5.4}$$

so that we deduce that the elements $x_{1,ij}$ are dyadic, i.e., $x_{1,ij} \in 2^{-p_1}\mathbb{N}$ (being $x_{1,0j} = 0$).

Consider the family of ordered curves parametrized by $x_1 \in [0,1]$

$$[0,1] \times [j-1,j]2^{-p_0} \ni t, x_2 \mapsto f'_{j,t}(x_1, x_2) = (1-t)x_1 + tf'(z_{1,j}(x_1), x_2),$$

and let $\xi_{j,t} : [0,1] \times [j-1,j]2^{-p_0} \to [0,1] \times [j-1,j]2^{-p_0}$ be the unique measure preserving map mapping each segment $\{x_1\} \times [j-1,j]2^{-p_0}$ into the image of $(f'_{j,t}(x_1,x_2), x_2)$, $x_2 \in [j-1,j]2^{-p_0}$. This map is uniquely defined by the balance of mass, which reads as

$$\int_{(j-1)2^{-p_0}}^{(\xi_{j,t})_2(x_1,x_2)} \partial_{x_1} f'_{j,t}(x_1, w) dw = x_2 - (j-1)2^{-p_0}. \tag{5.5}$$

Being $f'_{j,t}$ close to the identity, $\xi_{j,t}$ is smooth and close to the identity.

Let $\xi_t : [0,1]^2 \to [0,1]^2$ be the measure preserving map obtained by piecing together the maps $\xi_{j,t}$. By construction the map $T'' = T' \circ \xi_1$ maps each vertical segment $\{x_1\} \times [j-1,j]2^{-p_0}$ into the vertical segment

$$\{z_{1,j}(x_1)\} \times \big[g'(z_{1,j}(x_1), (j-1)2^{-p_0}), g'(z_{1,j}(x_1), j2^{-p_0})\big].$$

Step 2: construction of the affine maps. The next step is to rectify the pieces of curves

$$[i-1,i]2^{-p_0} \ni z_1 \mapsto g_j(z_1) = g'(z_1, j2^{-p_0}), \tag{5.6}$$

which are the horizontal slab of the sets $\tilde{\omega}'_{ij}$. Fixing a vertical slide v_i, one considers the unique measure preserving map $\eta_{i,t} : [i-1,i] \times [0,1] \to [i,i-1] \times [0,1]$ such that the segments (5.6) are mapped into vertical segments and such that maps the curve $g_j([i-1,i]2^{-p_0})$ into the curve

$$\begin{aligned} g'_{t,ij}(z_1) &= (1-t)g'(z_1, j2^{-p_0}) + t \fint_{(i-1)2^{-p_0}}^{i2^{-p_0}} g'(w, j2^{-p_0}) dw \\ &= (1-t)g'(z_1, j2^{-p_0}) + tz_{2,ij}. \end{aligned} \tag{5.7}$$

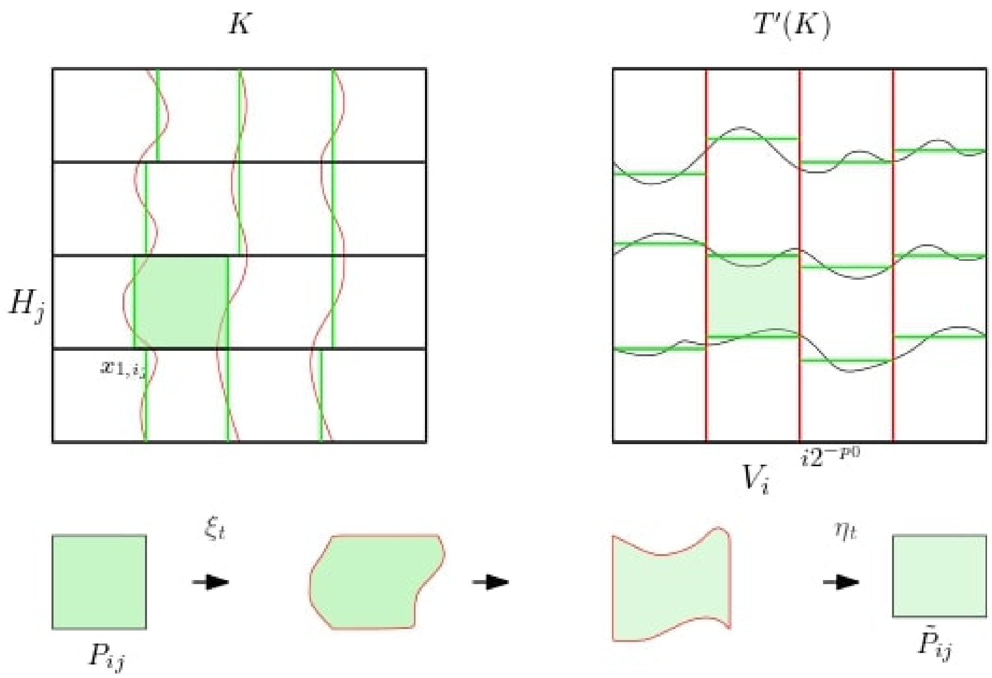


Figure 6: The action of σ_t in Lemma 5.1: first the map ξ_t moves the mass in H_j in order to map the vertical green segments into the counterimages of the vertical red segments; then the map T acts and the horizontal black boundaries of H_j becomes the black curves, but vertical segments remain vertical; finally the action of η_t rectifies the horizontal boundaries, while keeping vertical segments vertical.

In each $\tilde{\omega}'_{ij}$ this map is uniquely determined by the balance

$$\int_{(i-1)2^{-p_0}}^{(\eta_{i,t})_2(z_1,z_2)} \big(g'_{t,ij}(w) - g'_{t,i(j-1)}(w)\big)dw = \text{constant},$$

while the vertical coordinate is affine in each vertical segment.

Let $\eta_t : [0,1]^2 \to [0,1]^2$ be the measure preserving map obtained by piecing together the maps $\eta_{i,t}$.

Conclusion. Up to a time scaling, the map we are looking for is

$$\sigma_t = T \circ \xi_t \mathbf{1}_{[0,1]}(t) + \zeta_{t-1} \circ T \circ \xi_1 \mathbf{1}_{[1,2]}(t) + \eta_{t-2} \circ \zeta_1 \circ T \circ \xi_1 \mathbf{1}_{[2,3]}(t).$$

It is clearly measure preserving and at $t = 3$ it maps affinely the rectangles with dyadic coordinates

$$P_{ij} = \big[x_{1,(i-1)j}, x_{1,ij}\big] \times [j-1, j]2^{-p_0}$$

into the rectangles with dyadic coordinates

$$\tilde{P}_{ij} = [i-1, i]2^{-p_0} \times \big[z_{2,i(j-1)}, z_{2,ij}\big].$$

The values $x_{1,ij}$, $z_{2,ij}$ are given by (5.3), (5.7) and belong to $2^{-p_1}\mathbb{N}$. Thus $N_1 = N$ is the number of the statement.

The fact that σ_t is piecewise smooth and it extends continuously to the boundary of each P_{ij} are immediate from the construction, and its smallness follows by observing that as p_0, p_1 diverge the maps ξ, ζ, η converge to the identity. □

Being the rectangles dyadic in the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$, then we have the following

Corollary 5.2. *Every subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ is sent by σ_1 affinely by a diagonal matrix onto a subrectangle with rational coordinates.*

Remark 5.3. The previous lemma also tells us that the map σ_1 is piecewise affine, in particular there exists $N = 2^{p_1} \in \mathbb{N}$ refinement of the grid such that σ_1 maps each κ subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ affinely by a diagonal matrix onto a subrectangle q with rational coordinates. It's false, in general, that q has dyadic rational coordinates as stated in [**Shnirelman**]. More precisely, the previous lemma states that each

$$P_{ij} = [x_{1,(i-1)j}, x_{1,ij}] \times [j-1, j]2^{-p_0}$$

is sent into

$$\tilde{P}_{ij} = [i-1, i]2^{-p_0} \times [z_{2,i(j-1)}, z_{2,ij}],$$

where $x_{1,(i-1)j}, x_{1,ij}, z_{2,i(j-1)}, z_{2,ij}$ are dyadic. Call $\Delta x = x_{1,ij} - x_{1,(i-1)j}$ and $\Delta z = z_{2,ij} - z_{2,i(j-1)}$. Then by (5.4) $\Delta x = 2^{-p_1} n_{ij}$ with $n_{ij} \in \mathbb{N}$. Up to translation the perturbed map σ_1 (5.1) can be written as

$$\sigma_1 \llcorner_{P_{ij}} = \begin{pmatrix} \frac{2^{-p_0}}{\Delta x} & 0 \\ 0 & 2^{p_0}(\Delta z) \end{pmatrix}.$$

Take a subsquare $\kappa = [h-1, h]2^{-p_1} \times [k-1, k]2^{-p_1} \subset P_{ij}$, then $q = \sigma_1 \llcorner_{P_{ij}}(\kappa) = \left[\frac{2^{-p_0}}{n_{ij}}, 2^{p_0 - p_1}(\Delta z)\right]$, which is dyadic only with further requirements on n_{ij}. For a more detailed analysis consider H_j and call $\Delta x_i = x_{1,ij} - x_{1,(i-1)j} = 2^{-p_1} n_{ij}$, where $i = 1, \dots 2^{p_0}$. If we assume that every subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ is sent into a dyadic rectangle then we find the conditions

$$n_{ij} = 2^{m_{ij}}, \quad \forall i, j.$$

This condition tells us that, being measure-preserving,

$$\sigma_1 \llcorner_{P_{ij}} = \begin{pmatrix} 2^{-p_0 + p_1 - m_{ij}} & 0 \\ 0 & 2^{p_0 - p_1 + m_{ij}} \end{pmatrix},$$

that is, all possible matrices are of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 4 \end{pmatrix} \quad \dots \quad .$$

This condition is not compatible with the fact that σ_1 is an approximation of the original map T, which has been chosen to be close to the identity.

Remark 5.4. From the previous lemma it easily follows that, if N is the size of the grid, then every rectangle contained in the unit square K is sent by the perturbed flow into a union of rectangles.

Remark 5.5. To use Theorem 2.4 we observe that in our case the change of variables ϕ is given by the flow X_t. In particular, since X_t is close to the identity with all its derivatives, the costant C_{X_t} given by the previous theorem, is $C_{X_t} \leq (1+\delta)^{d-1}$ ($d = 2$ here).

5.1.1 The d-dimensional case

The analysis of the general case can be done as follows.

The starting point is the following approximation assumption in $d-1$-dimension.

Assumption 5.6. If the $\mathcal{L}^{d-1}$-measure-preserving diffeomorphism $T : [0,1]^{d-1} \to [0,1]^{d-1}$ is sufficiently close to the identity and equal to id in a neighborhood of $\partial[0,1]^{d-1}$, then there exists $N \in \mathbb{N}$, $N = 2^p$, and a measure-preserving piecewise smooth invertible map σ close to T such that $T \circ \sigma$ maps dyadic rectangles $P_{ij} \in \frac{\mathbb{N}^{d-1}}{N}$ onto dyadic rectangles $\tilde{P}_{ij} \in \frac{\mathbb{N}^{d-1}}{N}$ by a diagonal linear map (up to a translation).

The above assumption is true for $d = 3$: indeed if σ_t is the map of Lemma 5.1 and T is any $\mathcal{L}^2$-measure-preserving diffeomorphism $T : [0,1]^2 \to [0,1]^2$, as in the previous assumption, then $\sigma = T^{-1} \circ \sigma_{t=1}$ does the job.

Now let $T : [0,1]^d \to [0,1]^d$ be a diffeomorphism sufficiently close to the identity and equal to the identity near $\partial[0,1]^d$ (Figure 7). We will not address the perturbation ζ used to obtain dyadic parallelepipeds (Step 0 of the proof above), being the idea completely similar to the $2d$-case. We will also neglect the time dependence (i.e., how to split $t \in [0,1]$ into time intervals where the different maps are acting), because it is a fairly easy extension of the $2d$ case.

Step 1. Consider the curves

$$z_d \mapsto T^{-1}(z_1, \dots, z_d).$$

The first step is to perturb T to a map T' in order to have that the above curves are segments along the x_d-direction in each slab $x_d \in [k_d, k_d + 1]/N$ (Figure 8).

Being T close to the identity, the surface $T^{-1}(\{z_{d-1} = \text{const.}\})$ is parameterized by $x_1, \dots, x_{d-2}, x_d$, and then in each strip

$$(x_1, \dots, x_{d-2}) = \text{const.}, \quad x_{d-1} \in [0,1], \ x_d \in [k_d, k_d + 1] \frac{1}{N}$$

one can use the same measure preserving map ξ_1 defined in Step 1 of the proof of Lemma 5.1 above to obtain a perturbation $\hat{T} = T \circ \xi$ such that

$$\hat{T}^{-1}(\{z_{d-1} = \text{const.}\}) \cap \big\{x_d \in [k_d, k_d + 1]/N\big\}$$

is independent of x_d, in the sense that it is the graph of a function depending only on $x_1, \dots, x_{d-2}$ times the segment $x_d \in [k_d, k_d + 1]/N$.
Disintegrate the Lebesgue measure $\mathcal{L}^d$ as

$$\mathcal{L}^d \llcorner_{\{x_d \in [k_d, k_d+1]/N\}} = \int \Big[a(x_1, \dots, x_{d-2}, z_{d-1}) dx_1 dx_{d-2} dx_d \Big] dz_{d-1},$$

according to the partition $\hat{T}^{-1}(z_{d-1} = \text{const})$ (the density a does not depend on x_d because the surfaces contains the segments along x_d), and consider the 2-dimensional surfaces

$$\hat{T}^{-1}(z_{d-1} = \text{const}) \cap \{x_1, \dots, x_{d-3} = \text{const}\}.$$

We use the same map ξ_1 of Step 1 of the proof above to rectify the curves

$$\begin{aligned} z_d \mapsto f(x_1, \dots, x_{d-3}, z_{d-2}, z_{d-1}; z_d) = (\hat{T})^{-1}(z_{d-2}, z_{d-1} = \text{const}) \\ \cap \{x_1, \dots, x_{d-3} = \text{const}\} \cap \{x_d \in [k_d, k_d + 1]/N\}. \end{aligned}$$

The main difference w.r.t. the maps (5.2), (5.5) is that instead of the Lebesgue measure we use the density $a(x_1, \dots, x_{d-2}, z_{d-1})$. Eventually, the composition of the two maps above gives a new map $\check{T}$ such that $(\check{T})^{-1}(z_{d-2}, z_{d-1} = \text{const})$ is a $(d-2)$-dimensional surface made of the graph of a function depending on $x_1, \dots, x_{d-3}$ times the segment $x_d \in k_d, k_d + 1]/N$.
The argument is then repeated in the $d-2$-regions $(\check{T})^{-1}(z_{d-2}, z_{d-1} = \text{const})$ (i.e., disintegrate the Lebesgue measure and shift along the x_{d-3} direction to rectify $(\check{T})^{-1}(z_{d-3}, \dots, z_{d-1} = \text{const})$, and so on until we obtain that a new map T' such that

$$(T')^{-1}(\{z_1, \dots, z_{d-1} = \text{const.}\}) \cap \big\{x_d \in [k_d, k_d + 1]/N\big\}$$

is independent on x_d. This means that lines along the z_d are mapped back into N segments of length $1/N$ along x_d.

Step 2. Differently from the $2d$ case, it is not enough to perturb the vertical slab as in Step 2, since the sets $(T')^{-1}(\{z_k = \text{const.}\})$ are not of (piecewise) the form $\{x_k = \text{const.}\}$. Observe however that in each slab $\{x_d \in [k_d, k_d + 1]/N\}$ the map

$$(x_1, \dots, x_{d-1}) = (T'_{k_d})^{-1}(z_1, \dots, z_{d-1})$$

is well defined, where $(T')^{-1}_{k_d}$ denotes the first $(d-1)$-components of $(T')^{-1}$ restricted to $\{x_d \in [k_d, k_d + 1]/N\}$: we have used the property that segments along z_d are mapped back into segments along x_d.

We use Assumption 5.6 to get a map $\sigma_{k_d} : [0,1]^{d-1} \to [0,1]^{d-1}$ such that $T''_{k_d} = T'_{k_d} \circ \sigma_{k_d}$ maps affinely parallepipeds of a grid $\mathbb{N}^{d-1}/(NN_1)$, $N_1 = 2^{p_1}$, into cubes of the same grid: we can take $N_1 \gg 1$ in order to be independent of k (Figure 9).
Hence the map T''_{k_d} maps parallelepipeds of the form

$$\prod_{i=1}^{d-1} \frac{[k_i, k_i + 1]}{NN_1} \times \frac{[k_d, k_d + 1]}{N}$$

into regions for the form

$$\bigg\{ z_d \in \big[g(x_1, \dots, x_{d-1}, k_d/N), g(x_1, \dots, x_{d-1}, (k_d + 1)/N) \big], z_i \in \prod_{i=1}^{d-1} \frac{[k'_i, k'_i + 1]}{NN_1} \bigg\},$$

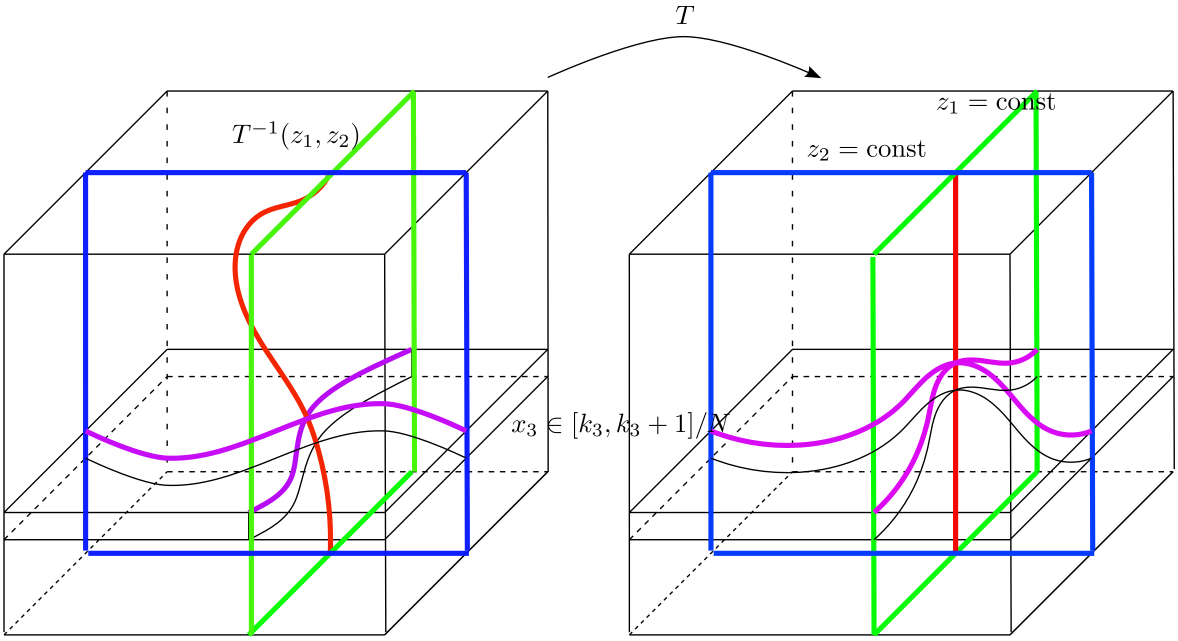


Figure 7: Starting point: the map $T : [0,1]^3 \to [0,1]^3$ maps the slap $x_3 \in [k_3, k_3 + 1]/N$ into a $3d$-set with purple intersections, and $T^{-1}(z_1, z_2)$ is the red curve at the left.

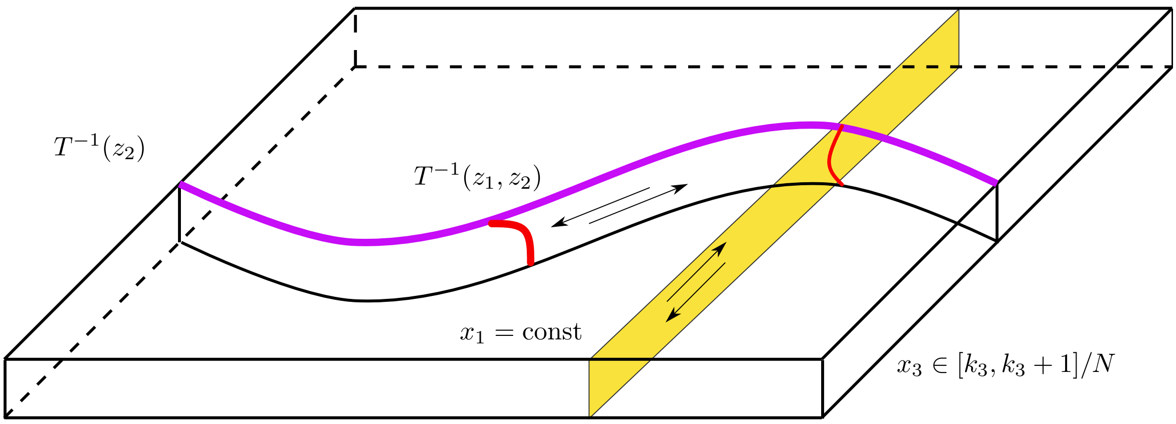


Figure 8: First move the mass in the yellow $2d$-rectangle $x_1 = \mathrm{const}$ so that its intersection with $T^{-1}(z_2)$ is vertical in $x_3 \in [k_3, k_3 + 1]/N$, next move the mass along $T^{-1}(z_2)$ so that $T^{-1}(z_1, z_2)$ is vertical in $x_3 \in [k_3, k_3 + 1]/N$.

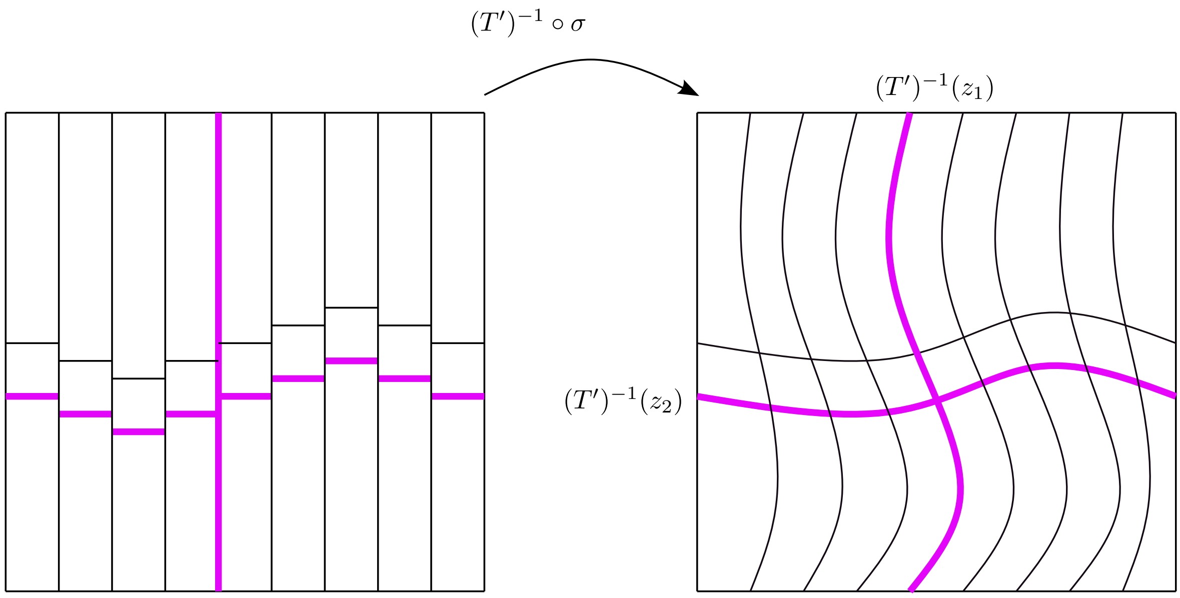


Figure 9: The recurrence assumption yields a map σ which maps affinely subsquares into rectangles: in the picture it is shows how it acts before the composition with T' (see also Figure 6).

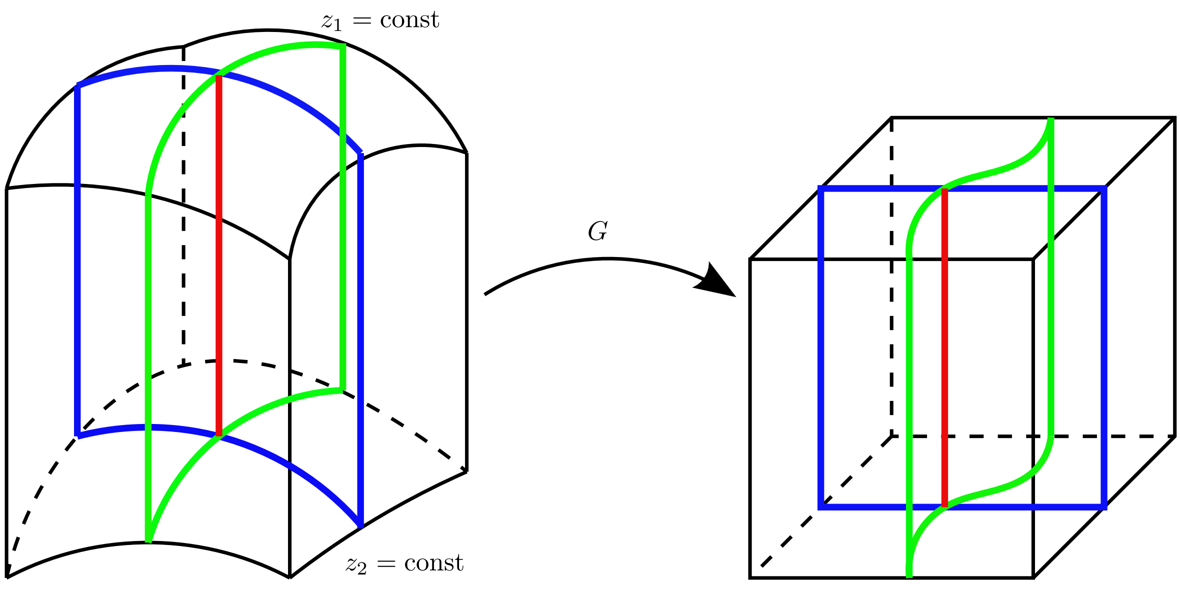


Figure 10: The last step is to map the subcubes deformed in the direction x_3 into parallelepipeds such that the Lebesgue measure is preserved and the map G is of triangular form: these conditions imply that $\tilde{T} = G \circ \bar{T}$ is affine.

and up to a translation it is a linear diagonal map in the first $d-1$ coordinates and segments along x_d remains along x_d.

Step 3. Piecing together the maps T''_{k_d}, we obtain a measure preserving map $\bar{T}$ close to T with the properties listed at the end of the previous step. We will use the fact that it is affine in the first $(d-1)$ coordinates to use a map similar η_1 of Step 2 of the proof of Lemma 5.1 to rectify the set

$$\bar{T}(\{x_d \in [k_d, k_d+1]/N\}) \cap \big\{z_j \in [k'_j, k'_j+1]/(NN_1), j = 1, \dots, d-1\big\}.$$

It is defined as the unique measure preserving map $G(z_1, \dots, z_{d-1})$ of the form

$$G(z_1, \dots, z_{d-1}) = \Big(G_1(z_1, \dots, z_{d-1}), G_2(z_2, \dots, z_{d-1}), \dots, G_{d-1}(z_{d-1}), G(z_1, \dots, z_d)\Big).$$

Note that since z_d enters in the last component, segments along z_d are mapped into segments along z_d, and the triangular form of the map assures its uniqueness (Figure 10).

The last part of the analysis is to deduce that if a measure preserving transformation $\tilde{T} = G \circ \bar{T} : [0,1]^d \to [0,1]^d$ is such that $\tilde{T}$ is of triangular form then it is the identity: we have rescaled every rectangle to a cube by linear scaling.

If the map has this triangular form, we conclude that the measure preserving condition reads as

$$\prod_{i=1}^{d} \partial_i \tilde{T}_i(x_i, \dots, x_d) = 1,$$

which together with

$$\int \partial_i \tilde{T}_i(x_i, \dots, x_d) dx_i = 1$$

gives $\partial_i T_i(x_i, \dots, x_d) = 1$, i.e., that the map is the identity.

5.2 BV estimates of perturbations

Let $X_t : K \to K$ be a smooth flow of measure-preserving diffeomorphisms and assume

$$\|X_t - \mathrm{id}\|_{C^3}, \|X_t^{-1} - \mathrm{id}\|_{C^3} \leq \wp, \tag{5.8}$$

with $\wp \ll 1$. Call $T(x) = X_{t=1}(x)$, and let $N = 2^{p_0} \in \mathbb{N}$ be the dimension of the grid given by Lemma 5.1. In this section we compute the BV norms of the perturbations of the form (5.1) constructed in Lemma 5.1.

We first address the action of the map ζ_t on $X(t)$. Define the perturbed flow

$$t \mapsto X'_t(x) = \zeta_t \circ X_t(x).$$

Lemma 5.7. *There exists a perturbation ζ_t as required by Step 0 of Lemma 5.1 such that if v is its associated vector field then*

$$\|\zeta_t - \mathrm{id}\|_{C_0} + 2^{-p_0}\|\nabla\zeta_t - \mathrm{id}\|_{C_0} + 2^{-2p_0}\|\nabla^2\zeta_t\|_{C_0} \leq \mathcal{O}(1)2^{-p_1},$$

$$\|v\|_{C_0} + 2^{-p_0}\|\nabla v\|_{C_0} + 2^{-2p_0}\|\nabla^2 v\|_{C^0} \leq \mathcal{O}(1)2^{-p_1}.$$

for p_1 sufficiently large.

Proof. The request of Step 0 of Lemma 5.1 is that the flow across the each region $\tilde{\omega}_{ij}$ is 0 (plus the dyadic condition on the new region $\tilde{\omega}_{ij}$). Hence the problem reduces in finding a suitable incompressible flow with a given boundary flux: we will construct a flow generated by a vector field constant in time.

Consider a function v_n on $\partial\omega_{ij}$ such that

- its support is at distance 2^{-p_0-2} from the corners of $\tilde{\omega}_{ij}$,
- the integral on each of the regular sides is the required flux $\phi_{ij,i'j'}$,
- $\|v_n\|_\infty, 2^{-p_0}\|v'_n\|, 2^{-2p_0}\|v''_n\| \leq \mathcal{O}(1)2^{-p_1}$, where v'_n is the derivative of v_n.

Its existence follows from the fact that ω_{ij} is close to a square of side 2^{-p_0}, being X close to the identity. The last point is a consequence of the fact that $|\phi_{ij,i'j'}| \leq 2^{-p_0-p_1}$.

The integral of v on $\partial\tilde{\omega}_{ij}$ is a potential function p, which is constant in the 2^{-p_0-2}-neighborhood of every corner and such that

$$\|p\|_{C_0}, 2^{-p_0}\|p'\|_{C_0}, 2^{-2p_0}\|p''\|_{C_0}, 2^{-3p_0}\|p'''\|_{C_0} \leq \mathcal{O}(1)2^{-p_0-p_1},$$

where p', p'' are its first and second derivative.

Extend p to a C^2-function inside $\tilde{\omega}_{ij}$: since this extension can be required to vary in a region of size 2^{-p_0-2}, we get

$$\|p\|_{C_0} \leq \mathcal{O}(1)2^{-p_0-p_1}, \quad \|\nabla p\|_{C_0} \leq \mathcal{O}(1)\frac{2^{-p_0-p_1}}{2^{-p_0-2}} = \mathcal{O}(1)2^{-p_1},$$

$$\|\nabla^2 p\|_{C_0} \leq \mathcal{O}(1)\frac{2^{-p_0-p_1}}{2^{-2p_0-4}} = \mathcal{O}(1)2^{p_0-p_1}.$$

$$\|\nabla^3 p\|_{C_0} \leq \mathcal{O}(1)\frac{2^{-p_0-p_1}}{2^{-3p_0-6}} = \mathcal{O}(1)2^{2p_0-p_1}.$$

In particular the vector field $v = \nabla^\perp p$ satisfies the statement, and if ζ_t is the flow generated by v then the same holds by the estimates

$$\|\zeta_t - \mathrm{id}\|_{C^0} \leq \|v\|_{C_0} t, \quad \|\nabla\zeta_t - \mathrm{id}\|_{C^0} \leq e^{\|\nabla v\|_{C_0} t} - 1,$$

$$\|\nabla^2\zeta(t)\|_{C^0} \leq e^{\|\nabla v\|_{C_0} t}\|\nabla^2 v\|_{C_0}\|\nabla\zeta\|^2_{C_0},$$

with $p_1 \gg 1$. □

Corollary 5.8. *If b' is the vector field associated with $X'_t = \zeta_t \circ X_t$, then*

$$\|b'_t - b_t\|_{C_1} \leq \mathcal{O}(1)2^{2p_0-p_1}.$$

Proof. From the formula (2.2)

$$b'(t,x) - b(t,x) = v(x) + \big(\nabla\zeta_t(t, \zeta_t^{-1}(x)) - \mathrm{id}\big)b(t,x),$$

where $v(x)$ is the time independent vector field associated with ζ_t. Hence from the previous lemma

$$\begin{aligned}\|b'(t) - b(t)\|_{C^1} &\leq \|v\|_{C^1} + \mathcal{O}(1)\|\zeta_t - \mathrm{id}\|_{C^2}\|b\|_{C^1} \\ &\leq \mathcal{O}(1)2^{2p_0-p_1}.\end{aligned}$$ □

Define the **perturbed flow** $t \to \tilde{X}_t$ as

$$\tilde{X}_t(x) = \begin{cases} X'(t, 0, \xi_t(x)) & t \in \left[0, \frac{1}{2}\right], \\ X'(t, 1, \eta(t, 1, w(x))) & t \in \left[\frac{1}{2}, 1\right], \end{cases}$$

where ξ_t and η_t are given by formula (5.1) of Lemma 5.1 (here since the map ζ is not needed we rescale ξ_t, η_t with $t \in [0, 1/2]$) and

$$w(x) = \eta_1 \circ T' \circ \xi_{\frac{1}{2}}(x), \quad T' = \zeta_1 \circ T.$$

Call $\tilde{b}_t$ the vector field associated with $\tilde{X}_t$.

Lemma 5.9 (BV estimates)**.** *There exists a positive constant $C = C(\wp)$ such that*

$$\|b - \tilde{b}\|_{L^\infty(L^1)} \leq \frac{C\wp}{N}, \quad \|\mathrm{Tot.Var.}(\tilde{b})\|_\infty \leq C\|\mathrm{Tot.Var.}(b)\|_\infty + \frac{C\wp}{N}. \tag{5.9}$$

Proof. From Corollary 5.8, we have that (for $p_1 \gg 1$)

$$\|b - b'\|_{C^1} \leq \mathcal{O}(1)2^{2p_0-p_1} \ll \frac{\wp}{N},$$

so that we are left to prove (5.9) with b' in place of b:

$$\|b' - \tilde{b}\|_{L^\infty(L^1)} \leq \frac{C\wp}{N}, \quad \|\mathrm{Tot.Var.}(\tilde{b}_t)\|_\infty \leq C\wp.$$

We will prove the above estimates for $t \in [0, 1/2]$, i.e., only for $X'_t \circ \xi_t$, being the analysis of $X'(t, 1, \eta(t, 1, w(x)))$ completely analogous.

We start by observing that there exists a constant $C > 0$ such that

$$\|\dot{\xi}_t\|_\infty \leq \frac{C\wp}{N}.$$

Indeed, the map $\xi^i_t = \xi_t \llcorner_{H_i}$ is given by the formulas (see Step 1 of the proof of Lemma 5.1)

$$\xi^i_{1,t}(x_1, x_2) = f'_{i,2t}(x_1, \xi^i_t(x_1, x_2)) = (1 - 2t)x_1 + 2tf'(z_{1,i}(x_1), \xi^i_{2,t}(x_1, x_2)),$$

$$\int_{(i-1)2^{-p_0}}^{\xi^i_{2,t}(x_1,x_2)} \partial_{x_1} f'_{i,2t}(x_1, w)dw = x_2 - (i - 1)2^{-p_0}.$$

Since it holds by (5.8)

$$\|f'\llcorner_{H_i} - \mathrm{id}\|_{C^3}, \|z_{1,i} - \mathrm{id}\|_{C^3} \leq \mathcal{O}(\wp), \tag{5.10}$$

then

$$\|f'_{i,2t}(x_1, w) - x_1\|_{C_3} \leq \mathcal{O}(\wp).$$

Hence the function

$$F(t, x_1, x_2, \xi) = \int_{(i-1)2^{-p_0}}^{\xi_2} \partial_{x_1} f'_{i,2t}(x_1, w)dw - x_2 + (i - 1)2^{-p_0}$$

satisfies

$$\|F(t, x_1, x_2, \xi) - (\xi_2 - x_2)\|_{C^2} = \left\| \int_{(i-1)2^{-p_0}}^{\xi_2} \big(\partial_{x_1} f'_{i,2t}(x_1, w) - 1\big)dw \right\|_{C_2} \leq \frac{\mathcal{O}(\wp)}{N}.$$

By the Implicit Function Theorem we deduce that

$$\|\xi^i_{2,t} - x_2\|_{C^2} \leq \frac{\mathcal{O}(\wp)}{N},$$

and in particular

$$\|\dot{\xi}^i_{2,t}\|_{C^0}, \|\nabla\dot{\xi}^i_{2,t}\|_{C^0} \leq \frac{\mathcal{O}(\wp)}{N}.$$

Similarly

$$\|\xi^i_{1,t} - x_1\|_{C^2} \leq \frac{\mathcal{O}(\wp)}{N},$$

and then

$$\|\dot{\xi}^i_{1,t}\|_{C^0}, \|\nabla\dot{\xi}^i_{1,t}\|_{C^0} \leq \frac{\mathcal{O}(\wp)}{N}.$$

We next estimate the total variation of the vector field

$$v_t = \dot{\xi}_t(\xi^{-1}_t(x)).$$

We will use the following elementary formulas:

$$\partial_{z_1} f'(z_1, x_2) = \frac{1}{\partial_{z_2} X^{-1}_2(z_1, z_2(z_1, x_2))} \tag{5.11}$$

$$\partial_{x_2} f'(z_1, x_2) = \frac{\partial_{z_2}(X_1)^{-1}(z_1, z_2(z_1, x_2))}{\partial_{z_2}(X_2)^{-1}(z_1, z_2(z_1, x_2))}, \quad X_2(z_1, z_2(z_1, x_2)) = x_2, \tag{5.12}$$

$$\partial_{x_1} z_1 = \frac{1}{\fint_{(i-1)2^{-p_0}}^{i2^{-p_0}} \partial_{z_1} f'(z_{1,i}(x_1), w)dw},$$

from which it follows

$$\sum_i ||\partial_{z_1} f'(z_{i,1}, x_2)\partial_{x_1} z_{i,1} - 1||_{L^1(H_i)} + \sum_i \|\partial_{x_2} f'(z_{i,1}, x_2)\|_{L^1(H_i)} \leq C||\text{Tot.Var.}(b)(K)||_\infty.$$

Indeed

$$\begin{aligned}
||\partial_{z_1} f'(z_{i,1}, x_2)\partial_{x_1} z_{i,1} - 1||_{L^1(H_i)} &\leq ||\partial_{x_1} z_{i,1}||_\infty ||\partial_{z_1} f'(z_{i,1}, x_2) - 1||_{L^1(H_i)} \\
&\quad + ||\partial_{x_1} z_{i,1} - 1||_{L^1(H_i)} \\
&\leq C||\partial_{z_1} f'(z_{i,1}, x_2) - 1||_{L^1(H_i)} \\
&\quad + C\left\| \fint_{(i-1)2^{-p_0}}^{i2^{-p_0}} \partial_{z_1} f'(z_{i,1}, x_2) dw - 1 \right\|_{L^1(H_i)} \\
&\leq C||\partial_{z_1} f'(z_{i,1}, x_2) - 1||_{L^1(H_i)},
\end{aligned}$$

therefore, by (5.11), we get

$$\begin{aligned}
&\sum_i ||\partial_{z_1} f'(z_{i,1}, x_2) - 1||_{L^1(H_i)} + \sum_i \|\partial_{x_2} f'(z_{i,1}, x_2)\|_{L^1(H_i)} \\
&\qquad \leq C||\nabla(X^{-1} - Id)||_1 \leq C\int_0^1 \text{Tot.Var.}(b_s)(K) ds \leq C||\text{Tot.Var.}(b)(K)||_\infty.
\end{aligned}$$

By the Implicit Function Theorem we recover the following estimate for $|\nabla\dot{\xi}|$:

$$\begin{aligned}
|\nabla\dot{\xi}| &\leq C\bigg(|\partial_{x_1} f'(z_{1,i}(x_1), \xi_2) - 1| + \int_{(i-1)2^{-p_0}}^{\xi_2} |\partial^2_{x_1} f'(z_{1,i}(x_1), w)| dw \\
&\qquad + \|\nabla f'(z_{1,i}(x_1), \xi_2)\|_{C^1} \big(|\dot{\xi}| + |\nabla\xi| + |\dot{\xi}||\nabla\xi|\big)\bigg) \\
&\leq C|\partial_{x_1} f'(z_{1,i}(x_1), \xi_2) - 1| + \frac{\mathcal{O}(\wp)}{N}.
\end{aligned}$$

Hence

$$\|\nabla\dot{\xi}\|_1 \leq C\|\text{Tot.Var.}(b)(K)||_\infty + \frac{\mathcal{O}(\wp)}{N}.$$

For the jump part, we estimate the vector v_t at the boundaries of H_i: from the definition

$$\begin{aligned}
\xi_t(x_1, (i-1)2^{-p_0}) &= (1-2t)x_1 + 2tf'\big(z_{1,i}(x_1), (i-1)2^{-p_0}\big), \\
\xi_t(x_1, i2^{-p_0}) &= (1-2t)x_1 + 2tf'\big(z_{1,i}(x_1), i2^{-p_0}\big),
\end{aligned}$$

We consider only the second one, being the analysis of the first completely similar. Differentiating $\xi_t(x_1, i2^{-p_0})$ w.r.t. t and using the definition of $z_{1,i}(x_1)$ we have

$$\begin{aligned}
\dot{\xi}_t(x_1, i2^{-p_0}) &= 2\big(f'\big(z_{1,i}(x_1), i2^{-p_0}\big) - x_1\big) \\
&= 2\bigg(f'\big(z_{1,i}(x_1), i2^{-p_0}\big) - \fint_{(i-1)/N}^{i/N} f'\big(z_{1,i}(x_1), w\big) dw\bigg).
\end{aligned}$$

and then

$$\begin{aligned}
|\dot{\xi}_t(x_1, i/N)| &\leq 2\int_{(i-1)/N}^{i/N} \big|\partial_{x_2} f'(z_{1,i}(x_1), w)\big| dw \\
&\leq C\int_{(i-1)/N}^{i/N} \big|\partial_{z_2}(X_1)^{-1}(z_{1,i}(x_1), z_2(z_{1,i}(x_1), w))\big| dw.
\end{aligned}$$

Hence, by (5.10) and from the definition of $z_2(z_1, x_2)$ in (5.12) we have that

$$\big\|(z_{1,i}(x_1), z_2(z_{1,i}(x_1), x_2)) - \text{id}\big\|_{C^1} \leq \frac{\mathcal{O}(\wp)}{N},$$

so that using

$$\big\{\big(z_{1,i}(x_1), z_2(z_{1,i}(x_1), x_2)\big), (x_1, x_2) \in H_i\big\} = X_1(H_i),$$

we have

$$
\begin{aligned}
&\int_0^1 |\dot{\xi}_t(x_1, i/N)| dx_1 \\
&\leq C \int_0^1 \int_{(i-1)/N}^{i/N} \big|\partial_{z_2}(X_1)^{-1}(z_{1,i}(x_1), z_2(z_{1,i}(x_1), x_2))\big| dx_1 dx_2 \\
&\leq C \int_{X_1(H_i)} \big|\partial_{z_2}(X_1)^{-1}(z_1, z_2)\big| dz.
\end{aligned}
\tag{5.13}
$$

Again, since the change of variable which associate t, x_1 with its position w_1 on the jump line $[0,1] \times \{i/N\}$ is given by

$$
t, x_1 \mapsto w_1 = (1-2t)x_1 + 2tf'(z_{1,i}(x_1), i/N),
$$

up to a constant $1 + \mathcal{O}(\wp)/N$ (again by (5.10)) the first integral in (5.13) corresponds to the jump part of $\dot{\xi}(x_1, i/N)$ on $[0,1] \times \{i/N\}$ when extended to 0 outside H_i.

The same estimate holds for the jump of $\dot{\xi}_t(x_1, (i-1)2^{-p_0})$ on $[0,1] \times \{(i-1)/N$.

We conclude that

$$
\|D^{\mathrm{jump}} v\| \leq C \sum_i \int_{X(H_i)} \big|\partial_{z_2}(X_1)^{-1}(z_1, z_2)\big| dz \leq C \|\mathrm{Tot.Var.}(b)\|_\infty.
$$

We thus deduce

$$
\mathrm{Tot.Var.}(v_t) \leq C \|\mathrm{Tot.Var.}(b)||_\infty + \frac{\mathcal{O}(\wp)}{N}.
$$

Collecting all estimates we have:

L^1 **estimate.** Fix $t \in \big[0, \frac{1}{2}\big]$. From (2.2) it follows that

$$
|\tilde{b}_t(x) - b_t(x)| \leq \|\dot{\xi}_t\|_\infty |\nabla X'_t((X')_t^{-1}(x)| \leq \frac{C\wp}{N}.
$$

BV estimate. Again from (2.2)

$$
\mathrm{Tot.Var.}(\tilde{b}_t - b_t) \leq \mathrm{Tot.Var.}\big(\nabla X_t(X_t^{-1}(x))\dot{\xi}_t(\tilde{X}_t^{-1}(x))\big),
$$

so that we have to compute the total variation of

$$
\nabla X_t(X_t^{-1}(x))\dot{\xi}_t(X_t^{-1}(x)).
$$

By using Theorem 2.4 we have

$$
\begin{aligned}
\mathrm{Tot.Var.}\big(\nabla X_t(X_t^{-1})\dot{\xi}_t(\tilde{X}_t^{-1})\big) &\leq \mathrm{Lip}(X_t)\mathrm{Tot.Var.}\big(\nabla X_t \dot{\xi}_t(\xi_t^{-1})\big) \\
&\leq \mathrm{Lip}(X_t)\mathrm{Tot.Var.}(\nabla X_t)\|\dot{\xi}_t\|_\infty \\
&\quad + \mathrm{Lip}(X_t)\|\nabla X_t\|_\infty \mathrm{Tot.Var.}(\dot{\xi}_t \circ \xi_t^{-1}) \\
&= \mathrm{Lip}(X_t)\mathrm{Tot.Var.}(\nabla X_t)\|\dot{\xi}_t\|_\infty \\
&\quad + \mathrm{Lip}(X_t)\|\nabla X_t\|_\infty \mathrm{Tot.Var.}(v_t).
\end{aligned}
$$

The first term can be estimated by

$$
\mathrm{Tot.Var.}(\nabla X_t)\|\dot{\xi}_t\|_\infty \leq \frac{\mathcal{O}(\wp)}{N},
$$

while the second term is controlled by

$$
\|\nabla X_t\|_\infty \mathrm{Tot.Var.}(v_t) \leq C\|\mathrm{Tot.Var.}(b)||_\infty + \frac{\mathcal{O}(\wp)}{N}.
$$

This is the statement. □

Remark 5.10. The above estimates can be obtained also for the d-dimensional case, since the maps used in that case is a composition of maps of the $2d$ case: the estimates are completely similar (but a lot more complicated).

Remark 5.11. We notice here that the constant C in front of the total variation is larger than 1: this fact is one of the reasons why we need to work in the G_δ-set $\mathcal{U}$.

5.3 BV estimates for rotations

In this part we address the analysis of rotations: these are needed because the map of Lemma 5.1 maps affinely subrectangles into subrectangles, while we need squares translated into squares.
The approach here differs from the one of [**Shnirelman**], because the rotations used in that paper have a BV norm which is not bounded by the area (Lemma 5.12) and instead it depends on the size of the squares (actually it blows up when the size of the squares goes to 0).

Let $N = 2^{p_0}, p_0 \in \mathbb{N}$, and $\sigma_1 : K \to K$ be respectively the dimension of the grid and the map given by Lemma 5.1.

Lemma 5.12. *There exists $M \in \mathbb{N}$ and a flow $\bar{R}_t : K \to K$ invertible, measure-preserving and piecewise smooth such that the map $\sigma_1 \circ \bar{R}_1$ translates each subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{M}$ into a subsquare of the same grid, i.e., it is a permutation of squares.*

In particular note that $\nabla(\sigma_1 \circ \bar{R}_1)$ is the identity inside each subsquare κ.

Proof. Fix κ a subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ and call $q = \sigma_1(\kappa)$ its image. Since $\sigma \llcorner \kappa$ is an affine measure-preserving map of diagonal form, then, up to translations,

$$\sigma_1(x) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} x, \qquad \forall x \in \kappa,$$

where $\lambda_1, \lambda_2 \in \mathbb{Q}_{>0}$ and $\lambda_1\lambda_2 = 1$. Therefore the rectangle q has horizontal side of length $\frac{\lambda_1}{N}$ and vertical side of length $\frac{1}{\lambda_1 N}$. Decompose now the square κ into rectangles R_{ij} with $i = 1, \dots, \frac{1}{l_1}$ and $j = 1, \dots, \frac{1}{l_2}$ with horizontal side of length $\frac{l_1}{N}$ and vertical side of length $\frac{l_2}{N}$. The numbers $\frac{1}{l_1}, \frac{1}{l_2} \in \mathbb{N}$ are chosen such that $\lambda_1 = \frac{l_2}{l_1}$, i.e., $\sigma_1(R_{ij}) = R_{ij}^{\perp}$, where $R_{ij}^{\perp}$ is the rotated rectangle counterclockwise of an angle $\frac{\pi}{2}$.
In each R_{ij} we perform a rotation given by the flow

$$R_t^{ij} = \chi^{-1} \circ r_t \circ \chi,$$

where $\chi : R_{ij} \to K$ is the affine map, up to translation, sending each R_{ij} into the unit square K, namely

$$\chi x = \begin{pmatrix} \frac{N}{l_1} & 0 \\ 0 & \frac{N}{l_2} \end{pmatrix} x, \quad \forall x \in R_{ij},$$

whereas $r_t : K \to K$ is the rotation flow (2.5). Finally define $R_t : K \to K$ such that $R_{t \llcorner R_{ij}} = R_t^{ij}$. This flow rotates the interior of each rectangle R_{ij} by $\pi/2$ during the time evolution.

Now we choose $M \in \mathbb{N}$ large enough to refine the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ in such a way that for every subsquare $\kappa \in \mathbb{N} \times \mathbb{N}\frac{1}{N}$, $\forall i, j$, each rectangle $R_{ij} \subset \kappa$ is the union of squares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{M}$ and each rectangle $q = \sigma_1(\kappa)$ is union of subsquares of $\mathbb{N} \times \mathbb{N}\frac{1}{M}$, which is possible since the vertices of the squares and rectangles we are considering are all rationals.
We claim that the map $\sigma_1 \circ R_t : K \to K$ is a flow of invertible, measure-preserving maps such that $\sigma_1 \circ R_1 : K \to K$ is a permutation of subsquares of size $\frac{1}{M}$ up to a rotation of $\pi/2$. Indeed, fix R_{ij} and assume that it contains a_{ij} subsquares κ_{ij}^h of size $\frac{1}{M} \times \frac{1}{M}$ along the horizontal side and b_{ij} subsquares along the vertical one. The rotation R_1 stretches each square κ_{ij}^h into a rectangle r_{ij}^h whose size is $\frac{l_1}{Ml_2} \times \frac{l_2}{Ml_1}$, i.e., $\frac{1}{\lambda_1 M} \times \frac{\lambda_1}{M}$. Now it is clear that $\sigma_1(r_{ij}^h)$ is a square of size $\frac{1}{M} \times \frac{1}{M}$.
The Jacobian of R_1 is

$$JR_1 = \begin{pmatrix} 0 & -\lambda_2 \\ \lambda_1 & 0 \end{pmatrix}.$$

Thus the composition of the two maps acts in each square r_{ij}^h as

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 0 & -\lambda_2 \\ \lambda_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

i.e., a rotation of $\pi/2$.

Define then the map $\bar{R}_t$ as

$$\bar{R}_{t \llcorner \kappa_{ij}^h} = R_t \circ r_t^{-\pi/2},$$

where $r_t^{-\pi/2}$ is the rotation of the square κ_{ij}^h of $-\pi/2$: now the map has Jacobian id in each square κ_{ij}^h. This is the map of the statement. $\square$

Remark 5.13. A completely similar construction can be done in dimension $d \geq 3$: in this case, in each cube $\kappa \in \mathbb{N}^d/N$ the piecewise affine map σ has the form (up to a translation)

$$\sigma = \mathrm{diag}(\lambda_1, \dots, \lambda_d), \quad \lambda_1 \lambda_2 \dots \lambda_d = 1.$$

Hence the subpartition of κ is done into parallelepipeds $\ell_1 \times \ell_2 \times \cdots \times \ell_d$ such that

$$\lambda_i = \frac{\ell_{i+1}}{\ell_i}.$$

The action of σ transform each of these parallelepipeds into the new ones $\ell_2 \times \ell_3 \times \cdots \times \ell_d \times \ell_1$, which is the range of the rotation

$$\begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \mathrm{id} & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \cdots\cdots\cdots \begin{pmatrix} 0 & -1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \mathrm{id} & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

(The above formula is the decomposition into $2d$ rotations.) Hence, as in Lemma 5.12 above, a rotation of the parallelpipeds $\ell_1 \times \cdots \times \ell_d$ and a counter-rotation of the subcubes of the parallelepipeds gives the transformation.

5.4 Main approximation theorem

We are ready to prove our main result.

Theorem 5.14. *Let $b \in L^\infty([0,1]; BV(K))$ be a divergence-free vector field and assume that there exists $\delta > 0$ such that for $\mathcal{L}^1$-a.e. $t \in [0,1]$, $\operatorname{supp} b_t \subset\subset K^\delta$. Then for every $\epsilon > 0$ there exist $\delta', C_1, C_2 > 0$ positive constants, $D \in \mathbb{N}$ arbitrarily large and a divergence-free vector field $b^\epsilon \in L^\infty([0,1]; BV(K))$ such that*

1. $\operatorname{supp} b^\epsilon_t \subset\subset K^{\delta'}$,
2. *it holds*

$$\|b - b^\epsilon\|_{L^\infty(L^1)} \leq \epsilon, \quad ||\mathrm{Tot.Var.}(b^\epsilon)(K)||_\infty \leq C_1 ||\mathrm{Tot.Var.}(b)(K)||_\infty + C_2, \tag{5.14}$$

3. *the map $X^\epsilon \llcorner_{t=1}$ generated by b^ϵ at time $t = 1$ translates each subsquare of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ into a subsquare of the same grid, i.e., it is a permutation of squares.*

Remark 5.15. Observe that the theorem can be easily extended to vector fields $b \in L^\infty([0,1], \mathrm{BV}(\mathbb{R}^2))$ such that $\operatorname{supp} b_t \subset K$. We keep here the original setting of [**Shnirelman**].

Remark 5.16. By inspection of the proof one can check that C_1, C_2 are independent of b. This is in any case not needed for the proof of the main theorem.

Remark 5.17. A possible approach would be to divide the time interval $[0,1]$ into sufficiently small time steps $\sim \tau$ in order to apply Lemmas 5.1 , 5.12 and hence to compose the resulting maps as done in [**Shnirelman**], however by Lemma 2.6,

$$||\mathrm{Tot.Var.}(\bar{R})(\kappa)||_\infty \sim \frac{\mathrm{Area}(K)}{\tau}$$

so that the total variation blows up as the time step goes to zero.

Proof. We divide the proof into several steps.

Step 1. Let $\rho \in C^\infty_c(\mathbb{R}^2)$ be a mollifier, and define

$$b^\alpha_t \doteq b_t * \rho_\alpha,$$

where $\rho_\alpha(x) \doteq \alpha^{-2} \rho(\frac{x}{\alpha})$ and $\alpha << 1$ is chosen such that $\operatorname{supp} b_{\alpha,t} \subset\subset K$.

By well known estimates (see (2.1)) we obtain

$$\|b_t^\alpha - b_t\|_{L^1} \leq \alpha \text{Tot.Var.}(b_t)(K), \quad \text{Tot.Var.}(b_t^\alpha)(K) \leq \text{Tot.Var.}(b_t)(K).$$

then if α is chosen such that $\alpha \leq \frac{\epsilon}{2||\text{Tot.Var.}(b)(K)||_\infty}$, we conclude that

$$||b_t^\alpha - b_t||_1 \leq \frac{\epsilon}{2}, \quad \text{Tot.Var.}(b_t^\alpha)(K) \leq \text{Tot.Var.}(b_t)(K)$$

and we have to prove the theorem for b^α. Moreover b^α satisfies the estimates

$$||b_t^\alpha||_\infty \leq \frac{1}{\alpha^2}||b_t||_1, \quad \|\nabla^n b_t^\alpha\|_\infty \leq \frac{C_n}{\alpha^{1+n}}\|\text{Tot.Var.}(b_t)(K)\|_\infty.$$

From now on we will call $b^\alpha = b$ to avoid cumbersome notations.

Step 2. Let us consider a partition of the time interval $0 = t_0 < t_1 < \cdots < t_n = 1$ where $n \in \mathbb{N}$ and $t_i = \frac{i}{n}$, where n will be chosen later on. Let us call $X_j \doteq X(t_j, t_{j-1}, x)$ and $X_j(t) \doteq X(t, t_{j-1}, x)$ defined for $t \in [t_{j-1}, t_j]$. Then each flow $X_j(t)$ is close to the identity with its derivatives, indeed

$$X_j(t, x) = x + \int_{t_{j-1}}^t b(s, X(s, t_{j-1}, x))ds, \tag{5.15}$$

so that

$$\|X_j(t) - \text{id}\|_{C^k} \leq C(k)(t - t_{j-1})\|b\|_{C^k}.$$

More precisely, if $\wp$ is the constant of (5.8), there exists $n \in \mathbb{N}$ such that

$$\|X_j(t) - \text{id}\|_{C^3}, \|X_j^{-1} - \text{id}\|_{C^3} \leq \wp, \quad \forall t \in [t_{j-1}, t_j], \quad \forall j = 1, \ldots, n.$$

Therefore we can apply Lemma 5.1 to each $X_j(t)$ finding $N_j = 2^{p_j}$ dyadic and $\tilde{X}_j : [t_{j-1}, t_j] \times K \to K$ with the property that, at time $t = t_j$, the map $\tilde{X}_j(t_j)$ sends subsquares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N_j}$ into rational rectangles with vertices in $\frac{\mathbb{N}}{N_j R_j}$. In particular, the eigenvalues of all affine maps σ for $\tilde{X}_j(t_j)$ belongs to $\frac{\mathbb{N}}{R_j}$. We can moreover assume that $N_j = N$ for all maps $\tilde{X}_j(t_j)$ by taking N sufficiently large. Finally from Lemma 5.9 we have that in each interval $[t_{j-1}, t_j]$ it holds

$$\|b - \tilde{b}_j\|_{L^\infty(L^1)} \leq \frac{C\wp}{N}, \quad \|\text{Tot.Var.}(\tilde{b}_j)\|_\infty \leq C\|\text{Tot.Var.}(b_t)\|_\infty + \frac{C\wp}{N}.$$

so that for $N \gg 1$ we have

$$\|b - \tilde{b}_j\|_{L^\infty(L^1)} \leq \epsilon, \quad ||\text{Tot.Var.}(\tilde{b}_j)(K)||_\infty \leq C||\text{Tot.Var.}(b)(K)||_\infty + \epsilon,$$

where $\tilde{b}_j$ is the vector field associated with $\tilde{X}_j$.

We define $t \to \tilde{X}_t$ the *perturbed flow*

$$\tilde{X}(t) = \begin{cases} \tilde{X}_1(t) & 0 \leq t \leq t_1, \\ \tilde{X}_2(t) \circ \tilde{X}_1(t_1) & t_1 \leq t \leq t_2, \\ \ldots & \\ \tilde{X}_{i+1}(t) \circ \tilde{X}_i(t_i) \circ \cdots \circ \tilde{X}_1(t_1) & t_i \leq t \leq t_{i+1}, \\ \ldots & \\ \tilde{X}_n(t) \circ \tilde{X}_{n-1}(t_{n-1}) \circ \cdots \circ \tilde{X}_1(t_1) & t_{n-1} \leq t \leq 1. \end{cases}$$

The map is clearly piecewise affine, and the eigenvalues of each affine piece σ belong to $\frac{\mathbb{N}}{\prod_j R_j}$.

Step 3. The map $\tilde{X}(1)$ has the property of sending subsquares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{N}$ into union of rational rectangles. Let $D = N(\prod_j R_j)^2$: we now show that $\tilde{X}(1)$ maps subsquares of the grid $\mathbb{N} \times \mathbb{N}\frac{1}{D}$ into rational rectangles.

Let $R = \prod_j R_j$ and assume that the map $\tilde{X}(1, t_{j+1})$ maps the subsquares of the grid

$$\mathbb{N} \times \mathbb{N}\frac{1}{NR\prod_{k=j+1}^n R_k}$$

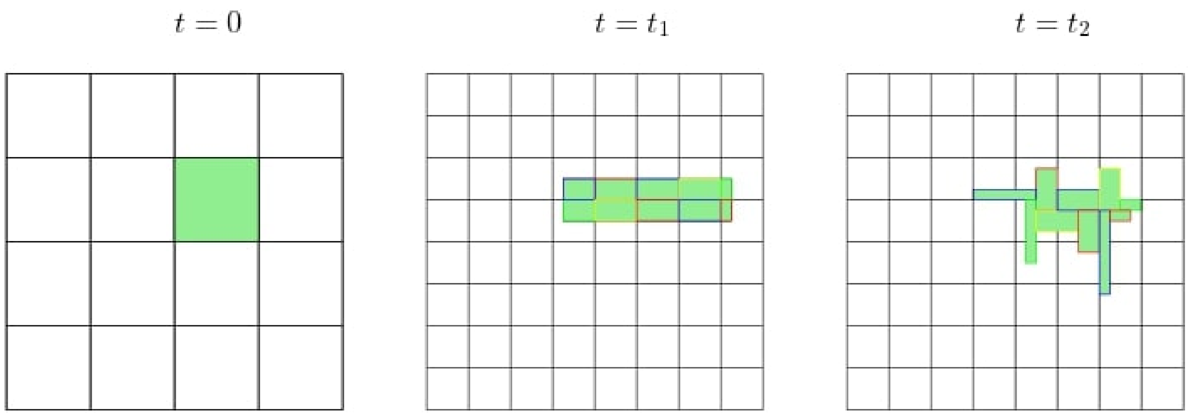


Figure 11: A square is sent into a union of rectangles by the map $\tilde{X}(1)$.

into rational rectangles. Since $\tilde{X}_j$ maps affinely the subsquares of $\mathbb{N}\times\mathbb{N}\frac{1}{N}$ into rectangles of the grid

$$\mathbb{N}\times\mathbb{N}\frac{1}{NR_j} \subset \mathbb{N}\times\mathbb{N}\frac{1}{NR\prod_{k=j+1}^{n} R_k},$$

then

$$\tilde{X}_j^{-1}\left(\mathbb{N}\times\mathbb{N}\frac{1}{NR\prod_{k=j+1}^{n} R_k}\right) \subset \mathbb{N}\times\mathbb{N}\frac{1}{NR\prod_{k=j}^{n} R_k}.$$

In particular we obtain that

$$\tilde{X}^{-1}\left(\mathbb{N}\times\mathbb{N}\frac{1}{NR}\right) \subset \mathbb{N}\times\mathbb{N}\frac{1}{NR^2}.$$

We rename the flow $\tilde{X}_t^D$ to indicate the size of the grid on which it acts as a piecewise affine map. Note that the above estimates (as well as the next ones) improve whenever N becomes larger, so that the size of the grid D can be taken arbitrarily large.

Step 4. To conclude the proof we want to modify the flow $\tilde{X}_t^D$ slightly in such a way that the new flow $\check{X}_t^D$ evaluated at $t=1$ sends subsquares into subsquares by translations. The key idea is to perform rotations as in Lemma 5.12 balancing two effects: one one hand the cost of a rotation is at least of the order of the area (Lemma 2.6), on the other hand if the squares are too much deformed the cost is exponentially large w.r.t. the total variation used to deform the square. The idea will be to use these rotations only when the deformation reaches a critical threshold.

First let us fix κ_0 a subsquare of the grid $\mathbb{N}\times\mathbb{N}\frac{1}{D}$ and call κ_i its images through the maps $\kappa_i = \tilde{X}_{t=t_i}^D(\kappa_0)$. Since each map $\tilde{X}_i(t_i)$ is affine and measure-preserving on κ_{i-1}, up to a translation it can be represented as

$$\begin{pmatrix} \sigma_i & 0 \\ 0 & \frac{1}{\sigma_i} \end{pmatrix}$$

where $\sigma_i \in \mathbb{Q}$ and $|\sigma_i - 1|, |\frac{1}{\sigma_i} - 1| \leq \wp \ll 1$, where $\wp$ is the one given by the partition. Moreover, being

$$\nabla^2 \tilde{X}_{t=t_j}^D(x) = 0 \tag{5.16}$$

whenever x belongs to the interior of the subsquares, we deduce that for the same x

$$|\nabla^2 \tilde{X}_t^D(x)| \leq \wp.$$

We can also observe that, by (5.15)

$$|\sigma_i - 1|, \left|\frac{1}{\sigma_i} - 1\right| \leq \frac{C}{\mathcal{L}^2(\kappa_0)} \int_{t_{i-1}}^{t_i} \mathrm{Tot.Var.}(b_s^D)(\kappa_0) ds.$$

By elementary computations one can prove that

$$\left|\prod_i \sigma_i - 1\right|, \left|\prod_i \frac{1}{\sigma_i} - 1\right| \leq \max\left\{\sigma, \frac{1}{\sigma}\right\} \sum_i |\sigma_i - 1|.$$

Hence if we have the bound

$$3 \leq \left(\sigma_{j+1} \cdot \ldots \cdot \sigma_i + \frac{1}{\sigma_{j+1} \cdot \ldots \cdot \sigma_i}\right) \leq 4,$$

then

$$\frac{3}{8}\mathcal{L}^2(\kappa_j) \leq \int_{t_j}^{t_i} \text{Tot.Var.}(b_s^D)(\kappa_0) ds. \tag{5.17}$$

The idea is to find now a new sequence of times $\{t_{i_j}\} \subset \{t_i\}$, $j = 1, \ldots, n'$ and $i = 1, \ldots, n$, in which we perform the rotations of Lemma 2.6 in order to have both the total variation is controlled be the total variation of $\tilde{b}^D$ and the property of sending subsquares into subsquares by translation.

Let us start defining $t_{i_0} = 0$ and

$$t_{i_1} = \min T_0,$$

where

$$T_0 = \left\{t_i > 0 : 3 \leq \left(\sigma_1 \cdot \cdots \cdot \sigma_i + \frac{1}{\sigma_1 \cdot \cdots \cdot \sigma_i}\right) \leq 4\right\}.$$

Then two situations may occur.

1. The set T_0 is empty, that is $\left(\sigma_1 \ldots \sigma_i + \frac{1}{\sigma_1 \ldots \sigma_i}\right) \leq 3$ for all i. In this case we perform a rotation in $[0, 1]$, that is $R^1 : [0, 1] \times K \to K$ (as in Lemma 5.12) where $R^1 \llcorner_{K \setminus \mathring{\kappa}_0}(x) = x$ and it is such that $X^D \circ R^1 \llcorner_{t=1}$ sends subsquares of κ_0 into subsquares of κ_n. In this case

$$\hat{X}_t^D = \tilde{X}_t^D \circ R_t^1$$

$$\hat{b}_t^D(x) = \tilde{b}_t^D(x) + \nabla \tilde{X}_t^D \circ (\tilde{X}_t^D)^{-1}(x) \dot{R}_t^1((\tilde{X}_t^D \circ R_t^1)^{-1}(x)) \quad x \in \kappa_0$$

(where we have recalled that all functions can be extended smoothly to κ_0) and

$$\begin{aligned}
\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) &\leq \|\nabla \tilde{X}_t^D\|_\infty^2 TV(\dot{R}_t(R_t^{-1}))(\kappa_0) \\
&\quad + \|\nabla \tilde{X}_t^D\|_\infty \|\dot{R}\|_\infty \text{Tot.Var.}(\nabla \tilde{X}_t^D)(\kappa_0) \\
&\leq \|\nabla \tilde{X}_t^D\|_\infty \frac{4}{D^2}\left(\sigma_1 \ldots \sigma_n + \frac{1}{\sigma_1 \ldots \sigma_n}\right) + \frac{\mathcal{O}(\wp)}{D^3} \\
&\leq \|\nabla \tilde{X}_t^D\|_\infty \frac{12}{D^2} + \frac{\mathcal{O}(\wp)}{D^3}.
\end{aligned} \tag{5.18}$$

We have observed that

$$\nabla^2 \tilde{X}_{t=t_j}^D(x) = 0 \quad \text{for } x \in \mathring{\kappa}_0 \text{ by (5.16)},$$

so that for $t \in [t_{j-1}, t_j]$

$$\|\nabla^2 \tilde{X}_t^D\|_\infty \leq \mathcal{O}(1) \int_{t_{j-1}}^t |\nabla^2 b(s, \tilde{X}_s^D)| ds = \frac{O(\wp)}{D^2}.$$

Since $\|\nabla \tilde{X}_t^D\|_\infty \leq C$ by the assumptions that these are sets with small deformations, we obtain

$$\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) \leq \frac{\mathcal{O}(1)}{D^2} = \mathcal{O}(1)\mathcal{L}^2(\kappa_0),$$

where we have used (5.17).

2. The set T_0 is non empty. Then if $t_{i_1} = 1$ the procedure stops and you perform a rotation as in Lemma 5.12 in $[0, 1]$ finding

$$\begin{aligned}
\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) &\leq \|\nabla \tilde{X}_t^D\| \frac{4}{D^2}\left(\sigma_1 \ldots \sigma_n + \frac{1}{\sigma_1 \ldots \sigma_n}\right) + \frac{\mathcal{O}(\wp)}{D^3} \\
&\leq \mathcal{O}(1)\|\nabla \tilde{X}_t^D\| \int_0^1 \text{Tot.Var.}(\tilde{b}_s^D)(\kappa_0) ds + \frac{\mathcal{O}(\wp)}{D^3},
\end{aligned}$$

where we have used (5.17) and we have estimated the higher order term as in (5.18).

If $t_{i_1} < 1$ we compute

$$t_{i_2} = \min T_1,$$

where

$$T_1 = \left\{ t_i > t_{i_1} : 3 \leq \left(\sigma_{i_1+1} \dots \sigma_i + \frac{1}{\sigma_{i_1+1} \dots \sigma_i} \right) \leq 4 \right\}.$$

If $T_1 = \emptyset$ we stop the procedure and we perform a rotation in $[t_{i_1-1}, 1] = [0, 1]$ finding, for $t \in [t_{i_1-1}, 1]$

$$\begin{aligned}
\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) &\leq \|\nabla \tilde{X}_t^D\| \frac{4}{D^2} \frac{1}{1 - t_{i_0}} \left(\sigma_{i_1-1} \dots \sigma_n + \frac{1}{\sigma_{i_0} \dots \sigma_n} \right) \\
&\quad + \frac{\mathcal{O}(\wp)}{D^3} \\
&\leq \mathcal{O}(1) \frac{1}{1 - t_{i_0}} \int_{t_{i_0}}^{t_{i_1}} \text{Tot.Var.}(\tilde{b}_s^D)(\kappa_0) ds + \frac{\mathcal{O}(\wp)}{D^3} \\
&\leq \mathcal{O}(1) \fint_{t_{i_0}}^{1} \text{Tot.Var.}(\tilde{b}_s^D)(\kappa_0) ds + \frac{\mathcal{O}(\wp)}{D^3}.
\end{aligned} \tag{5.19}$$

If T_1 is non empty we perform a rotation in $[0, t_{i_1}]$ finding for $t \in [0, t_{i_1}]$ the following estimate

$$\begin{aligned}
\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) &\leq \|\nabla \tilde{X}^D\| \frac{4}{D^2} \frac{1}{t_{i_1} - 1} \left(\sigma_1 \dots \sigma_{i_1} + \frac{1}{\sigma_1 \dots \sigma_{i_1}} \right) + \frac{\mathcal{O}(\wp)}{D^3 t_{i_1}} \\
&\leq \mathcal{O}(1) \fint_0^{t_{i_1}} \text{Tot.Var.}(\tilde{b}_s^D)(\kappa_0) ds + \frac{\mathcal{O}(n\wp)}{D^3}.
\end{aligned}$$

Again if $t_{i_2} = 1$ the procedure stops and we perform another rotation in $[t_{i_1}, t_{i_2}]$. If not we consider the set T_2 and we proceed.

The general step if the following: we consider the set

$$T_{j+1} = \left\{ t_i > t_{i_j} : 3 \leq \left(\sigma_1 \cdot \dots \cdot \sigma_i + \frac{1}{\sigma_1 \cdot \dots \cdot \sigma_i} \right) \leq 4 \right\}.$$

If T_{j+1} is empty, then we perform a rotation in $[t_{i_{j-1}}, 1]$ finding for $t \in [t_{i_{j-1}}, 1]$ the same estimate as (5.19). If T_{j+1} is non empty, we perform a rotation in $[t_{i_{j-1}}, t_{i_j}]$ and we consider

$$T_{j+2} = \left\{ t_i > t_{i_{j+1}} : 3 \leq \left(\sigma_1 \cdot \dots \cdot \sigma_i + \frac{1}{\sigma_1 \cdot \dots \cdot \sigma_i} \right) \leq 4 \right\}.$$

At the end of this procedure there are two possible scenarios: let $n' = \sup\{j : T_j \neq \emptyset\}$. If $t_{i_{n'}} = 1$, the procedure ends by performing a rotation in $[t_{i_{n'-1}}, 1]$. If we find $t_{i_{n'}} < 1$ with the property that

$$\left(\sigma_{i_{n'}+1} \dots \sigma_i + \frac{1}{\sigma_{i_{n'}+1} \dots \sigma_i} \right) \leq 3$$

for all $i = i_{n'} + 1, \dots, n$, the construction ends with a rotation in $[t_{i_{n'-1}}, 1]$ (as in the case of the estimate (5.19)).

In particular, for each subsquare κ_0 we find a sequence of times $\{t_{i_j}(\kappa_0)\}$, $j = 1, \dots, n'(\kappa_0)$, where we are performing a rotation. There are two cases to be considered: if $T_0(\kappa_0)$ is empty then

$$\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) \leq \frac{\mathcal{O}(1)}{D^2},$$

otherwise

$$\text{Tot.Var.}(\hat{b}_t^D - \tilde{b}_t^D)(\kappa_0) \leq \mathcal{O}(1) \|\text{Tot.Var.}(\tilde{b}^D)(\kappa_0)\|_\infty + \frac{\mathcal{O}(n\wp)}{D^3}.$$

Summing over all possible κ_0 we find that

$$\begin{aligned}\text{Tot.Var.}(\hat{b}_s^D)(K) &\leq \text{Tot.Var.}(\tilde{b}_s^D)(K) + D^2\biggl(\frac{\mathcal{O}(n\wp)}{D^3} + \frac{\mathcal{O}(1)}{D^2}\biggr) + C_2\|\text{Tot.Var.}(\tilde{b}^D)(K)\|_\infty \\ &\leq \text{Tot.Var.}(\tilde{b}_s^D)(K) + C_1 + C_2\|\text{Tot.Var.}(\tilde{b}^D)(K)\|_\infty\end{aligned}$$

if $D \gg 1$, therefore we can conclude the proof finding a positive constant $C > 0$ such that

$$\|\text{Tot.Var.}(\hat{b}^D)(K)\|_\infty \leq C_1 + C_2\|\text{Tot.Var.}(\tilde{b}^D)(K)\|_\infty$$

which is the desired estimate. □

Remark 5.18. The same result can be obtained for the d-dimensional case, by using the maps of Section 5.1.1 and Remarks 5.10 and 5.13.

6 Appendix

Proof of Lemma 3.7. By the Ergodic Theorem, $T = X_{t=1}$ is ergodic iff

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_{T^i(A)} \to_{L^1} |A|$$

In particular, if T is ergodic, then by writing

$$\frac{1}{n}\int_0^n \chi_{X_t(A)} dt = \int_0^1 \frac{n-1}{n}\biggl(\frac{1}{n-1}\sum_{i=0}^{n-1}\chi_{T^i(X_s(A))}\biggr) ds$$

we see that

$$\fint_0^t \chi_{X_s(A)} ds \to_{L^1} |A|.$$

It is immediate to find a counterexample to the converse implication: just consider rotation of the unit circle with period 1.

The proof of the implication $\Rightarrow$ in the second point is analogous. For the converse, let $A, B \in \Sigma$ such that

$$\frac{1}{n}\sum_{i=0}^{n}\bigl[|T^i(A)\cap B| - |A||B|\bigr]^2 > \epsilon.$$

By the continuity of $s \mapsto X_s$ in the neighborhood topology we have that there exists $\bar{s}$ such that for $0 \leq s \leq \bar{s}$ it holds

$$\bigl|X_s(B)\triangle B\bigr| = \bigl|B\triangle (X_s)^{-1}(B)\bigr| < \frac{\epsilon}{2}.$$

Hence we can write

$$\begin{aligned}\fint_0^n \bigl[|X_t(A)\cap B| - |A||B|\bigr]^2 dt &\geq \frac{n}{T}\int_0^{\bar{s}} \frac{1}{n}\sum_{i=0}^{n-1}\bigl[\bigl|X_s(T^i(A))\cap B\bigr| - |A\cap B|\bigr]^2 ds \\ &= \int_0^{\bar{s}} \frac{1}{n}\sum_{i=0}^{n-1}\bigl[\bigl|T^i(A)\cap (X_s)^{-1}(B)\bigr| - |A\cap B|\bigr]^2 ds \\ &\geq \int_0^{\bar{s}} \frac{1}{n}\sum_{i=0}^{n-1}\bigl[\bigl|T^i(A)\cap B\bigr| - |A\cap B|\bigr]^2 ds - \bar{s}\frac{\epsilon}{2} > \bar{s}\frac{\epsilon}{2}\end{aligned}$$

for $n \gg 1$. Hence

$$\liminf \fint_0^T \bigl[|X_t(A)\cap B| - |A||B|\bigr]^2 dt \neq 0.$$

Finally, if T is strongly mixing, the continuity of $s \mapsto X_s$ in the neighborhood topology gives that $s \mapsto X_s^n = X_s \circ T^n$ is a family of equicontinuous functions, and since for all s fixed

$$\lim_{n\to\infty} |X_s(T^n(A))\cap B| = |A||B|$$

we conclude that X_s^n converges to 0 uniformly in s. The opposite implication is trivial. □