On infinite-volume mixing

MARCO LENCI *

June 2009

Abstract

In the context of the long-standing issue of mixing in infinite ergodic theory, we introduce the idea of mixing for observables possessing an infinite-volume average. The idea is borrowed from statistical mechanics and appears to be relevant, at least for extended systems with a direct physical interpretation. We discuss the pros and cons of a few mathematical definitions that can be devised, testing them on a prototypical class of infinite measure-preserving dynamical systems, namely, the random walks.

Mathematics Subject Classification: 37A40, 37A25, 82B41, 60G50.

1 Historical introduction

The textboox definition of mixing for a transformation $T : \mathcal{M} \longrightarrow \mathcal{M}$ preserving a probability measure μ is

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \tag{1.1}$$

for all measurable sets $A, B \subset \mathcal{M}$ [W]. Extending this definition to the case where μ is a σ-finite measure with $\mu(\mathcal{M}) = \infty$ is a fundamental issue in infinite ergodic theory. References to this problem can be found in the literature at least as far back as 1937, when Hopf devoted a section of his famous *Ergodentheorie* [H] to an example of a dynamical system that he calls 'mixing'. It consists of a set $\mathcal{M} \subset \mathbb{R}^2$, of infinite Lebesgue measure, and a map $T : \mathcal{M} \longrightarrow \mathcal{M}$ preserving μ, the Lebesgue measure on $\mathcal{M}$. He proved a property that is equivalent to this one: there exists a sequence $\{\rho_n\}_{n\in\mathbb{N}}$ of positive numbers such that

$$\lim_{n\to\infty} \rho_n\, \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \tag{1.2}$$

*Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40126 Bologna, Italy. E-mail: `lenci@dm.unibo.it`

for all squarable sets $A, B \subset \mathcal{M}$ (a squarable set is a bounded set whose boundary has measure zero).

For a long time, mathematicians did not seem to act on this suggestion, perhaps due in part to the impossibility, in *any* reasonable dynamical system, of verifying (1.2) for all finite-measure sets A, B. (This fact, which might not have been clear to Hopf himself, can be derived from a famous 1964 paper by Hajian and Kakutani [HK].)

Work in this direction, however, picked up rather intensively in the 1960's [Or, KP, KO, Pr, Kr, Pa], to the point that Krickeberg in 1967 [Kr] proposed (1.2) as the definition of mixing for almost-everywhere continuous endomorphisms of a Borel space $(\mathcal{M}, \mu)$ (with some extra, inessential, conditions on the sets A, B). Krickeberg applied his definition to Markov chains with an infinite state space and an infinite invariant measure, which is very interesting in the context of this paper because our main examples will be the prototypical infinite-state Markov chains, namely the random walks (cf. Sections 2.2, 4 and 5).

Krickeberg's definition has been studied by several researchers since then [Fr, To1, To2] and, in recent times, it was independently rediscovered by Isola, who uses (1.2) with $A = B$ and calls $\{\rho_n\}$ the *scaling rate* [I1, I2]. It failed, however, to establish itself as the ultimate definion of mixing in infinite measure. In my opinion, this is not so much because of the less-than-perfect requirement of a topological structure in a measure-theoretic problem, but rather for its inherent inability to describe the "global" infinite-measure aspects of a dynamics: after all, (1.2) only involves finite-measure sets. Related to this, it is unclear how this definition may be specified towards stronger and more physically relevant chaotic properties, such as, for example, the correlation decay rate. A little thinking convinces one that the speed of convergence in (1.2) cannot in general be uniform, even for uniformly nice sets (A and B can be arbitrarily far from each other so that the l.h.s. of (1.2) is negligible for arbitrarily long times).

At any rate, by the end of the 1960's, Krengel and Sucheston [KS] approached the problem from a more measure-theoretic point of view and devised the following two definitions: A discrete-time, nonsingular dynamical system $(\mathcal{M}, \mu, T)$ is called mixing if and only if the sequence $\{T^{-n}A\}_{n\in\mathbb{N}}$ is *semiremotely trivial* for all measurable $A \subset \mathcal{M}$ with $\mu(A) < \infty$; it is called completely mixing if the condition holds for *all* measurable A. (A nonsingular map is one for which $\mu(A) = 0$ implies $\mu(T^{-1}A) = 0$. As for the definition of semiremotely trivial, which is unimportant here, we refer the reader to [KS].) Both definitions reduce to the standard definition (1.1) for maps preserving a probability measure [Su].

However, Krengel and Sucheston themselves proved results that imply that many reasonable (including all invertible) measure-preserving maps cannot be completely mixing [KS, Thms. 3.1 and 5.1]. As for mixing, again specializing to measure-preserving maps, their definition is equivalent to

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = 0 \tag{1.3}$$

for all finite-measure sets A, B [KS, §2]. This is a rather brutal weakening of (1.2): for instance, it would classify a translation in $\mathbb{R}^d$ as mixing! Therefore, however illuminating [Sa], the Krengel–Sucheston definitions are of little applicability to most simple extensive systems that mathematicians would like to study.

Aaronson in 1997 [A, §2.5] wrote

> [...] the discussion in [KS] indicates that there is no reasonable generalisation of mixing.

Be that as it may, the drive to produce a general definition of mixing in infinite ergodic theory had apparently ceased by the mid 1970's.

In this paper I will not try to give a universal and firm definition of mixing—in which I am not a keen believer myself—for σ-finite measure-preserving dynamical systems, but rather a few very general *notions*, that can be completely specified on a case-by-case basis, depending on what type of information one wants to extract from the dynamical system under scrutiny.

To do so, I will borrow some ideas and a little terminology from physics, in particular from statistical mechanics. The key concept this work is based on is that of *infinite-volume average*, which I illustrate with an example.

Let us consider a measurable $A \subset \mathbb{R}^2$ and ask, what is the probability that a random $x \in \mathbb{R}^2$ will belong to A? The probabilist in the room will immediately say that this question fully depends on what we mean by random. Suppose we reply that random means that each x can be drawn with equal probability. The probabilist will likely give us a condescending look and explain that the Lebesgue measure m on $\mathbb{R}^2$, which is the only uniform measure on $\mathbb{R}^2$, cannot be normalized.

However, if we have a physicist in the room as well, she might answer that the sought probability is something like

$$\lim_{r \to +\infty} \frac{m(A \cap [-r, r]^2)}{4r^2}, \tag{1.4}$$

provided the limit exists. Of course, such an answer is riddled with issues, but it does capture the idea that in physics one only looks at finite quantities. Infinity is a mental construct to fit an endless amount of situations, and a finite limit at infinity is the formal way to say that most of these situations will look alike.

More generally, for a given dynamical system $(\mathcal{M}, \mathscr{A}, \mu, \{T^t\})$, where $(\mathcal{M}, \mathscr{A})$ is a measure space and $\{T^t\}$ is a (semi-)group of transformations $\mathcal{M} \longrightarrow \mathcal{M}$ preserving the infinite measure μ, we will choose a family of ever-larger sets V, with $\mu(V) < \infty$, that "approximate $\mathcal{M}$". Using the language of physics, one might say that choosing these sets will define how we *measure* our infinite system—more precisely, how we measure its *observables*.

We will deal with two types of observables: The *global*, or *macroscopic*, observables will be a suitable class of functions $F \in L^\infty(\mathcal{M}, \mathscr{A}, \mu)$ for which

$$\overline{\mu}(F) := \lim_{V \nearrow \mathcal{M}} \frac{1}{\mu(V)} \int_V F \, d\mu \tag{1.5}$$

exists. What the above limit means and what class of functions it applies to will be clarified in Section 2. The *local*, or *microscopic*, observables will be essentially the elements of $L^1(\mathcal{M}, \mathscr{A}, \mu)$.

Then, skipping many necessary details and all-important specifications which are found in Sections 2 and 3, our notions of mixing will basically reduce to the two limits:

$$\lim_{t\to\infty} \overline{\mu}((F \circ T^t)G) = \overline{\mu}(F)\overline{\mu}(G), \tag{1.6}$$

for any two global observables F, G; and

$$\lim_{t\to\infty} \mu((F \circ T^t)g) = \overline{\mu}(F)\mu(g), \tag{1.7}$$

for F global and g local (with the obvious notation $\mu(g) := \int g \, d\mu$).

To the extent to which the above notions can be made into rigorous definitions—and they can, cf. Sec. 3—they seem to improve on the attempted definitions that we have recalled earlier, chiefly because they involve observables which can be supported all over the phase space (think, for example, of the velocity of a particle in an aperiodic Lorentz gas, or the potential energy of a small mass in a formally infinite celestial conglomerate, etc.). So they are more apt to reveal the large-scale aspects of a given dynamics.

In particular, (1.6) may be called *global-global mixing*, because it somehow expresses the vanishing of the correlation coefficient between two global observables, while (1.7) may be called *global-local mixing*, because the coupling is between a global and a local observable. The latter notion can be quite useful if one takes $g \geq 0$ with $\int g \, d\mu = 1$. Then μ_g, the probability measure defined by $d\mu_g := g \, d\mu$, can be considered an initial state for the system. In this interpretation, the l.h.s. of (1.7) reads $\mu_g(F \circ T^t) =: T^t_* \mu_g(F)$, that is, the expected value of the observable F relative to the state at time t, and (1.7) asserts that such quantity converges to $\overline{\mu}(F)$. Hence, $\overline{\mu}$ acts as a sort of equilibrium state for the system.

In Section 4 we will apply the above ideas to certain basic yet representative examples of infinite-measure dynamical systems, the random walks in $\mathbb{Z}^d$. We will specify all the mathematical objects needed to obtain rigorous definitions out of (1.6)-(1.7) and we will check that, under reasonable conditions, all these definitions are verified. The proofs, given in Section 5, use basic harmonic analysis on groups [R] and an estimate for a certain Fourier norm. The latter is presented in the Appendix, together with other technical results.

Acknowledgments. This research began while I was a faculty at Stevens Institute of Technology, Hoboken, NJ, USA. There I was partially supported by NSF Grant DMS-0405439.

2 Infinite-volume limit and observables

A *measure-preserving dynamical system* is the quadruple $(\mathcal{M}, \mathscr{A}, \mu, \{T^t\})$, where $\mathcal{M}$ is a measure space with the σ-algebra $\mathscr{A}$, endowed with the σ-finite measure μ, while $\{T^t\}_{t\in\mathbb{G}}$ is a group (respectively, semigroup) of automorphisms (respectively, endomorphisms) $\mathcal{M} \longrightarrow \mathcal{M}$, labeled by the free additive parameter $t \in \mathbb{G}$ (without significant loss of generality, we assume $\mathbb{G} = \mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$). This means that

$$\mu(T^{-t}A) = \mu(A), \qquad \forall A \in \mathscr{A}, \forall t \in \mathbb{G}. \tag{2.1}$$

If $\mathbb{G} = \mathbb{R}$, $\{T^t\}$ is called *the flow*, whereas if $\mathbb{G} = \mathbb{Z}$ or $\mathbb{N}$, the generator $T := T^1$ is called *the map*, and one usually denotes the dynamical system by $(\mathcal{M}, \mu, T)$.

In this paper we are interested in the case $\mu(\mathcal{M}) = \infty$. The measure-theoretic properties of dynamical systems preserving an infinite measure are the subject of *infinite ergodic theory* [A], whose most basic definition, perhaps, is the following [HK]:

Definition 2.1 *The measure-preserving dynamical system $(\mathcal{M}, \mathscr{A}, \mu, \{T^t\})$ is called* ergodic *if every measurable invariant set (i.e., any $A \in \mathscr{A}$ such that $\mu(T^{-t}A \,\triangle\, A) = 0\ \forall t$), has either zero measure or full measure (the latter meaning $\mu(\mathcal{M} \setminus A) = 0$).*

In the case where μ is a probability measure, the above is one of the several equivalent formulations of ergodicity, including among others: the equivalence of the Birkhoff average with the phase average, for all $f \in L^1(\mathcal{M}, \mathscr{A}, \mu)$ (the definition generally ascribed to Boltzmann); the absence of nontrivial integrals of motion in $L^1(\mathcal{M}, \mathscr{A}, \mu)$; the strong law of large numbers for the random variables $\{f \circ T^t\}$. In infinite measure, these definitions are no longer equivalent and, among those that keep making sense, Definition 2.1 is in some sense the strongest.

A notion that is much harder to transport to infinite ergodic theory, as we have discussed in the introduction, is that of *mixing*. In terms of *observables*, i.e., scalar functions on $\mathcal{M}$, it reads:

$$\forall f, g \in L^2(\mathcal{M}, \mathscr{A}, \mu), \qquad \lim_{t\to\infty} \mu((f \circ T^t)g) = \mu(f)\mu(g). \tag{2.2}$$

In other words, the correlation coefficient between the two random variables $f \circ T^t$ and g vanishes asymptotically. This phrasing makes it apparent that the notion of finite-measure mixing is intrinsically probabilistic.

2.1 Infinite-volume limit

The discussion in the introduction suggests that one should find an asymptotic decorrelation formula, similar to (2.2), which applies to observables that, unlike L^2 functions, "see" a nonnegligible portion of the space. For this, one needs to define a sort of "normalized measure" for these observables. This is how one comes to think of averaging a function over $\mathcal{M}$ by means of an *infinite-volume limit*.

The idea is borrowed from statistical mechanics where the question arises of measuring the *intensive* quantities of a system (for example, the temperature or the density of a gas). These are represented by sequences of functions defined over larger and larger phase spaces, corresponding to larger and larger portions of the physical system, normalized in such a way that their integral converges to a finite number. This number is supposed to predict the outcome of an experimental measurement.

Coming back to our math, we introduce a notation that is going to be used often in the remainder:

$$\mathscr{A}_f := \{A \in \mathscr{A} \mid \mu(A) < \infty\} . \tag{2.3}$$

Definition 2.2 *The family* $\mathscr{V} \subset \mathscr{A}_f$ *is called* exhaustive *if it contains a sequence* $\{V_n\}_{n\in\mathbb{N}}$*, increasing w.r.t. inclusion, such that* $\bigcup_n V_n = \mathcal{M}$*.*

Definition 2.3 *Let* $\mathscr{V}$ *be exhaustive. If* ϕ *is defined on* $\mathscr{A}_f$ *and has values in some topological space, we write*

$$\lim_{V \nearrow \mathcal{M}} \phi(V) := \lim_{\substack{V\in\mathscr{V} \\ \mu(V)\to\infty}} \phi(V) = L,$$

when, for every neighborhood $\mathcal{U}$ *of* L*, there exists an* $M \in \mathbb{R}^+$ *such that*

$$V \in \mathscr{V}, \mu(V) \geq M \implies \phi(V) \in \mathcal{U}.$$

This will be called the μ-uniform infinite-volume limit w.r.t. the family $\mathscr{V}$*, or simply the* infinite-volume limit*.*

It is apparent that the above definition depends decisively on the choice of $\mathscr{V}$. In the example discussed in the introduction, cf. (1.4), $\mathscr{V} = \{[-r, r]^2\}_{r>0}$, and it is easy to think of other exhaustive families of subsets of $\mathbb{R}^2$ for which the infinite-volume limit for the "probability" of A differs from (1.4).

In general, all the results that we are going to discuss in this paper will depend on the choice of $\mathscr{V}$ in Definition 2.3. This is no shortcoming! In fact, we *want* to retain this choice, because this is how we incorporate in the mathematical description of an extended system the way we *observe* the system, that is, how we measure the observables of that system. In other words, the choice of $\mathscr{V}$ defines what it means to pick a large region of $\mathcal{M}$, which we assume—or rather *declare*—represents the whole space.

The first property we assume of our systems may be called 'compatibility of the infinite-volume limit with the dynamics':

(A1) For any fixed $t \in \mathbb{G}$, for $V \nearrow \mathcal{M}$, $\mu(T^{-t}V \triangle V) = o(\mu(V))$.

This means that the scale of the dynamics is local and not global: If V is a very large set "approximating $\mathcal{M}$", then its evolution at a fixed time should differ little from V, in relative terms. (In statistical mechanics one would say that, over a large region of the space, the dynamics can only produce negligible *surface effects*—we will come back to this point in Section 3.1.)

One can expect **(A1)** to hold in most situations. It does, for istance, when $\mathcal{M}$ is a metric space such that the μ-measure of a ball grows like a power of the radius, uniformly in the center of the ball; the dynamics is bounded, i.e., $\forall t \in \mathbb{G}$, $\exists K = K(t)$ such that, for μ-a.e. $x \in \mathcal{M}$, $\mathrm{dist}(T^t x, x) \le K$; and the elements of $\mathscr{V}$ are balls.

2.2 A couple of examples

A very simple example of this setup is the dynamical system defined as follows. Set

$$\varphi(x) := \begin{cases} 3x - 1, & \text{for } x \in [0,1); \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

Then, on $\mathcal{M} = \mathbb{R}$, define the self-map

$$Tx := \sum_{j \in \mathbb{Z}} \varphi(x - j) + j. \tag{2.5}$$

The definition is well-posed because, for any given $x \in \mathbb{R}$, only one term of the sum is nonzero. Furthermore, each x has three distinct counterimages via T, where the derivative of the map is constantly equal to 3. This shows that T is a noninvertible map which preserves the Lebesgue measure μ.

$(\mathcal{M}, \mu, T)$ describes a random walk in $\mathbb{Z}$, in the following sense: Suppose an initial condition $x \in [0,1)$ is randomly chosen according to μ_0, the Lebesgue measure in $[0,1)$ (it is no loss of generality to restrict to $[0,1)$ because the dynamical system is clearly invariant for the action of $\mathbb{Z}$). Then Tx will land in one of the intervals $[-1,0)$, $[0,1)$, $[1,2)$, with probability $1/3$ in each case. More generally, if $[x] := \max\{m \in \mathbb{Z} \mid m \le x\}$ denotes the integer part of $x \in \mathbb{R}$, we have

$$\mu_0\left(\left\{x \in [0,1) \;\middle|\; [T^n x] = k_n, [T^{n-1}x] = k_{n-1}, \ldots, [Tx] = k_1\right\}\right) = 3^{-n}, \tag{2.6}$$

provided that $|k_j - k_{j-1}| \le 1$, for $j = 1, \ldots, n$ (with $k_0 = 0$). This implies that, using the notation of conditional probability,

$$\begin{aligned} &\mu_0\left([T^n x] = k_n \;\middle|\; [T^{n-1}x] = k_{n-1}, \ldots, [Tx] = k_1\right) = \\ &\mu_0\left([T^n x] = k_n \;\middle|\; [T^{n-1}x] = k_{n-1}\right) = 1/3. \end{aligned} \tag{2.7}$$

Hence, $\{[T^n x]\}_n$ is a Markov chain in $\mathbb{Z}$ with same-site and nearest-neighbor jumps, each with probability $1/3$; namely, it is a (space-)homogeneous random walk.

As for the choice of $\mathscr{V}$, the example of the introduction would seem to suggest that we pick sets of the type $V = [-r, r]$ (with $r \in \mathbb{R}$) or, to fully exploit the $\mathbb{Z}$-structure of this dynamical system with no appreciable loss of generality, sets of the type $V = [-k, k]$ (with $k \in \mathbb{N}$). Although this is a legitimate choice, we will see later that a better option is

$$\mathscr{V} := \{[k, \ell] \mid k, \ell \in \mathbb{Z}, k < \ell\}. \tag{2.8}$$

(Actually, this will be a crucial part of our discussion, and we refer the reader to Section 3.1.) For $V = [k, \ell]$, $n \in \mathbb{N}$, and $\ell - k > 2n$, it is easy to verify that

$$[k+n, \ell-n] \subset T^{-n}V \subset [k-n, \ell+n], \tag{2.9}$$

which implies **(A1)**.

A less trivial system that fits well the framework we are describing, and to which I have developed a certain attachment in the last few years, is the (aperiodic) Lorentz gas [L1, L2].

In $\mathbb{R}^2$ (just to fix the smallest interesting dimension) a Lorentz gas is the billiard system in $C := \mathbb{R}^2 \setminus \bigcup_{n\in\mathbb{N}} \mathcal{O}_n$, where $\{\mathcal{O}_n\}$ is a countable collection of pairwise disjoint, convex, bounded regular sets. This means that a point particle moves with constant unit velocity in C until it hits an obstacle $\mathcal{O}_n$, which reflects the particle according to the Fresnel law: the angle of reflection equals the angle of incidence (the modulus of the velocity remains equal to 1). The phace space for this system is then $\mathcal{M} = C \times S^1$, where $q \in C$ represents the position and $v \in S^1$ the velocity of the particle. If T^t denotes the flow just described (which is unambiguously defined at all noncollision times), it is well known that T^t preserves the Liouville measure μ, which turns out to be the product of the Lebesgue measure on C and the Haar measure on S^1. Clearly, save for bizarre situations, $\mu(\mathcal{M}) = \infty$.

Since sufficient conditions for the ergodicity of $(\mathcal{M}, \mu, T)$ are known [L1], and since, for the finite-measure version of the Lorentz gas (the so-called *Sinai billiard*), mixing and stronger stochastic properties essentially follow from ergodicity, it is of interest to devise one or more sound definitions of mixing for this dynamical system.

As for the exhaustive family $\mathscr{V}$, in analogy to the previous system, a reasonable choice would be

$$\mathscr{V} := \left\{ (C \cap R) \times S^1 \mid R = [a,b] \times [c,d], \text{ with } a < b,\ c < d \right\}. \tag{2.10}$$

In Section 4 we will present a third example, which generalizes, in more than one way, the first system introduced above. It is a class of invertible dynamical systems representing all homogeneous random walks in $\mathbb{Z}^d$. One of its points of relevance is that it is designed to retain the most essential features of the Lorentz gas discussed above. It is thus a greatly simplified toy model, which we are able to study in depth. As a matter of fact, it will be the testing ground for our new notions of mixing, cf. Sections 4 and 5.

2.3 Global and local observables

In order to define a surrogate probability measure for our system, we need to declare *what* we intend to measure. In other words, we need to specify the *observables*, namely, the functions $\mathcal{M} \longrightarrow \mathbb{R}$ which represent the (sole) information that we can get on the state of the system.

In finite ergodic theory, this class of functions is $L^1(\mathcal{M}, \mathscr{A}, \mu)$, or sometimes $L^2(\mathcal{M}, \mathscr{A}, \mu)$. In virtually every situation, both are amply sufficient to give a full description of the state of the system (in fact, quite generally, the position itself of $x \in \mathcal{M}$ is given by a finite number of square-integrable functions). This is conspicuosly not true in infinite ergodic theory. Indeed, the forthcoming discussion will try to convince the reader that the choice of the observables is precisely at the heart of the matter in infinite-measure mixing (a point that, after all, was already contained in [KS]).

We deal with two categories of observables: the *global*, or *macroscopic*, observables and the *local*, or *microscopic*, observables.

Let us start by introducing the former, whose space we denote by $\mathcal{G}$. We will not give a definition, but rather a presentation of the minimal features that $\mathcal{G}$ should have. A precise definition only makes sense on a case-by-case basis and, indeed, the choice of $\mathcal{G}$ is part of our description of the system, just like the choice of $\mathscr{V}$. As a typographical rule, we indicate a global observable with an upper-case Roman letter, as in $F : \mathcal{M} \longrightarrow \mathbb{R}$.

We require at least the following conditions:

(A2) $\qquad \mathcal{G} \subset L^\infty(\mathcal{M}, \mathscr{A}, \mu)$.

(A3) $\qquad \forall F \in \mathcal{G}, \qquad \exists\, \overline{\mu}(F) := \lim_{V \nearrow \mathcal{M}} \mu_V(F) := \lim_{V \nearrow \mathcal{M}} \frac{1}{\mu(V)} \int_V F \, d\mu.$

We call $\overline{\mu}(F)$ the *average* of F (w.r.t. μ and $\mathscr{V}$). This functional is dynamics-invariant:

Lemma 2.4 *Under assumptions* **(A1)-(A3)**, $\overline{\mu}(F) = \overline{\mu}(F \circ T^t)$, $\forall t \in \mathbb{G}$.

Proof. Using the invariance of μ and then **(A1)-(A2)**, we have

$$\frac{1}{\mu(V)} \int_V F \, d\mu = \frac{1}{\mu(V)} \int_{T^{-t}V} (F \circ T^t) \, d\mu = \frac{1}{\mu(V)} \int_V (F \circ T^t) \, d\mu + o(1). \tag{2.11}$$

Applying **(A3)** gives the assertion. Q.E.D.

As for the class of local observables, denoted by $\mathcal{L}$, this can be generally taken to be $L^1(\mathcal{M}, \mathscr{A}, \mu)$. As we will see below, this choice is much less delicate than the choice of $\mathcal{G}$. Nonetheless, some results may require occasional restrictions on L^1, so, in the same spirit as **(A2)**, we only require

(A4) $\qquad \mathcal{L} \subseteq L^1(\mathcal{M}, \mathscr{A}, \mu)$.

Local observables are indicated with a lower-case Roman letter, as in $g : \mathcal{M} \longrightarrow \mathbb{R}$.

3 Definitions and related questions

3.1 Global-global mixing

On the basis of Lemma 2.4, one might attempt the following definition of mixing:

$$\textbf{(M2)} \qquad \forall F, G \in \mathcal{G}, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)G) = \overline{\mu}(F)\,\overline{\mu}(G),$$

provided that $\overline{\mu}((F \circ T^t)G)$ exists for all $t \in \mathbb{G}$, or at least for t large enough. This last point represents a problem, because it is not easy, in general, to devise a space $\mathcal{G}$ with the ring property (if $\mathcal{G}$ is defined to be T^t-invariant, which one may expect in many reasonable cases, the ring property suffices to prove that $F, G \in \mathcal{G}$ implies $(F \circ T^t)G \in \mathcal{G}$).

Generally speaking, there are only two solutions to this problem, both of which are rather cheap, honestly speaking. The first solution is to declare that this question should be dealt with on a case-by-case basis. The second solution is to devise another definition of mixing which just does away with the problem:

$$\textbf{(M1)} \qquad \forall F, G \in \mathcal{G}, \qquad \lim_{\substack{t\to\infty \\ V \nearrow \mathcal{M}}} \mu_V((F \circ T^t)G) = \overline{\mu}(F)\,\overline{\mu}(G),$$

having adopted the notation $\mu_V(\cdot) = \int_V (\cdot) d\mu / \mu(V)$, as introduced in **(A3)**. The above means that, $\forall \varepsilon > 0$, $\exists M > 0$ such that, for all $t \geq M$ and $V \in \mathscr{V}$ with $\mu(V) \geq M$, $|\mu_V((F \circ T^t)G) - \overline{\mu}(F)\overline{\mu}(G)| < \varepsilon$.

(M1) is evidently weaker than **(M2)**, but it still retains a great deal of the physical meaning of mixing because it prescribes that, if the region V is big enough and the time t is large enough, the two observables $F \circ T^t$ and G are practically uncorrelated on V.

We refer to **(M2)** and **(M1)** as definitions of *global-global mixing* because they consider the coupling of two global observables.

Let us now focus on a couple of less technical and more substantial questions concerning both **(M1)** and **(M2)**. The first has to do with the importance on $\mathscr{V}$ too, not just $\mathcal{G}$, for either condition to function as a sound definition of mixing.

Let us exemplify the question by means of the dynamical system defined by (2.4)-(2.5). This is a system that should be classified as mixing by any reasonable definition (cf. also Section 4). Suppose that for that system we had made the first, more restrictive, choice of $\mathscr{V}$ presented in Section 2.2, that is, we had chosen sets of the type $V = [-k, k]$, with $k \in \mathbb{N}$. The function $F : \mathbb{R} \longrightarrow \mathbb{R}$, defined by

$$F(x) := \begin{cases} -1, & \text{for } x < 0; \\ \;\;\,1, & \text{for } x \geq 0, \end{cases} \tag{3.1}$$

is bounded and has average $\overline{\mu}(F) = 0$. Now, fix $n > 0$ and consider $F \circ T^n$. Given the action of the map T and its interpretation as a random walk, it is not hard to see that, for $x < -n$, $F(T^n x) = -1$ and, for $x \geq n$, $F(T^n x) = 1$ (determining $F(T^n x)$

for $x \in [-n, n)$ is more complicated and irrelevant here). This and (3.1) imply that, for $|x| > n$, $F(T^n x)F(x) = 1$. Therefore, for k much larger than n and $V = [-k, k]$, $\mu_V((F \circ T^n)F)$ is very close to 1 and indeed $\overline{\mu}((F \circ T^n)F) = 1$. Since $\overline{\mu}(F) = 0$, this shows that, for F as in (3.1) and $G = F$, both limits in **(M1)** and **(M2)** fail to hold!

But this is reasonable: after all, F has variations (causing it to be nonconstant) only on a negligible set, namely $\{x = 0\} \subset \mathcal{M}$. By negligible set we mean, in this context, a set whose ρ-neighborhoods, for all $\rho > 0$, have "measure" zero w.r.t. $\overline{\mu}$. Therefore this is an instance of the phenomenon, which is well known in statistical mechanics, whereby the infinite-volume limit does not see *surface effects*. (When the dynamics is bounded, in the sense specified in the last paragraph of Section 2.1, the evolution of F can produce no more than surface effects.)

So we must avoid global observables that have significant variations on negligible sets. But this does not mean that we should cherry-pick our observables (although there is nothing wrong with that)! In the case at hand, for instance, the unusable observables are automatically eliminated by a smarter choice of $\mathscr{V}$, the one given in (2.8). That exhaustive family is translation invariant, which seems right for a system that is translation invariant.

An analogous discussion can be made for the other example presented in Section 2.2, the Lorentz gas, and for most extended dynamical systems one can imagine.

We may conclude that $\mathscr{V}$ should be not too small as to make the verification of **(M1)** impossible while, at the same time, not too large as to make the class of global observables satistying **(A3)** too meager. A right balance seems for $\mathscr{V}$ to include all the symmetries, or "quasi-symmetries", of the system and no more.

The second, and more critical, question concerning **(M1)**-**(M2)** is that these definitions are completely blind to the local aspects of the dynamics. For instance, they not able to detect an invariant set A, if $\mu(A) < \infty$. One can easily produce a system that is mixing in the sense of **(M1)**, or even **(M2)**, but not ergodic as per Definition 2.1. (For example, take a Lorentz gas and make one scatterer hollow: points inside that scatterer will stay confined there, thus breaking ergodicity, while all other trajectories will be the same as in the unperturbed system, which one believes to be at least **(M1)**-mixing, with the right choice of $\mathcal{G}$ and $\mathscr{V}$.)

We have no fix for this issue, other then giving a few more definitions which take into account local observables as well.

3.2 Global-local mixing

The most natural way to couple global and local observables in a definition of mixing is this:

(M4) $\qquad \forall F \in \mathcal{G},\ \forall g \in \mathcal{L}, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)g) = \overline{\mu}(F)\mu(g),$

where we have used the convenient notation $\mu(g) := \int_{\mathcal{M}} g\, d\mu$. This is the first notion of *global-local mixing* we give. Since for some systems, such as the Lorentz gas, this can be rather hard to prove, we give a weaker version as well:

(M3) $\forall F \in \mathcal{G},\ \forall g \in \mathcal{L}$ with $\mu(g) = 0, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)g) = 0.$

(Notice that **(M3)** and **(M4)** are equivalent in ordinary ergodic theory, because one can always substract a constant function from any observable in order to make its integral vanish. Not so in infinite measure!)

We will see momentarily that, for systems for which a uniform version of **(M4)** can be established, it is possible to pass from global-local mixing to global-local mixing. So we give one last definition:

(M5) $\forall F \in \mathcal{G}, \qquad \lim_{t\to\infty} \sup_{g\in\mathcal{L}\setminus 0} \frac{1}{\mu(|g|)} \left| \overline{\mu}((F \circ T^t)g) - \overline{\mu}(F)\mu(g) \right| = 0.$

3.3 Summary of assumptions and definitions

For the convenience of the reader, we summarize here all the assumptions we have made and all the definitions of mixing we have given, listing the latter in the correct hyerarchical order, as clarified by Propositions 3.1 and 3.2 below.

The following are the minimal requirements on the dynamical system $(\mathcal{M}, \mathscr{A}, \mu, \{T^t\})$, the exhaustive family $\mathscr{V}$, the space of the global observables $\mathcal{G}$, and the space of the local observables $\mathcal{L}$:

(A1) For any fixed $t \in \mathbb{G}$, for $V \nearrow \mathcal{M},\ \ \mu(T^{-t}V \triangle V) = o(\mu(V))$.

(A2) $\mathcal{G} \subset L^\infty(\mathcal{M}, \mathscr{A}, \mu)$.

(A3) $\forall F \in \mathcal{G}, \qquad \exists\, \overline{\mu}(F) := \lim_{V\nearrow\mathcal{M}} \mu_V(F) := \lim_{V\nearrow\mathcal{M}} \frac{1}{\mu(V)} \int_V F\, d\mu.$

(A4) $\mathcal{L} \subseteq L^1(\mathcal{M}, \mathscr{A}, \mu)$.

The definitions of global-global mixing are:

(M1) $\forall F, G \in \mathcal{G}, \qquad \lim_{\substack{t\to\infty \\ V\nearrow\mathcal{M}}} \mu_V((F \circ T^t)G) = \overline{\mu}(F)\,\overline{\mu}(G).$

(M2) $\forall F, G \in \mathcal{G}, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)G) = \overline{\mu}(F)\,\overline{\mu}(G).$

The definitions of global-local mixing are:

(M3) $\forall F \in \mathcal{G},\ \forall g \in \mathcal{L}$ with $\mu(g) = 0, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)g) = 0.$

(M4) $\forall F \in \mathcal{G},\ \forall g \in \mathcal{L}, \qquad \lim_{t\to\infty} \overline{\mu}((F \circ T^t)g) = \overline{\mu}(F)\mu(g).$

(M5) $\forall F \in \mathcal{G}, \qquad \lim_{t\to\infty} \sup_{g\in\mathcal{L}\setminus 0} \frac{1}{\mu(|g|)} \left| \overline{\mu}((F \circ T^t)g) - \overline{\mu}(F)\mu(g) \right| = 0.$

Proposition 3.1 *Under all the assumptions made so far,*

$$
\begin{aligned}
(\mathbf{M2}) &\implies (\mathbf{M1}); \\
(\mathbf{M5}) &\implies (\mathbf{M4}) \implies (\mathbf{M3}).
\end{aligned}
$$

On the other hand, **(M1)** *implies that the limit in* **(M2)** *holds for all pairs* $F, G \in \mathcal{G}$ *such that* $\overline{\mu}((F \circ T^t)G)$ *exists for all* t *large enough.*

PROOF. The two chains of implications are obvious. The last assertion follows directly from the definition of the double limit $t \to \infty$, $V \nearrow \mathcal{M}$; cf. Section 3.1. Q.E.D.

With reasonable hypotheses, the strongest version of global-local mixing implies the weakest version of global-global mixing:

Proposition 3.2 *Suppose that every* $G \in \mathcal{G}$ *can be written* μ*-almost everywhere as*

$$
G(x) = \sum_{j \in \mathbb{N}} g_j(x), \qquad \text{with } g_j \in \mathcal{L},
$$

and, for every $V \in \mathscr{V}$*, there exists a finite subset* $\mathbb{J}_V$ *of* $\mathbb{N}$*, such that*

$$
\begin{aligned}
\mu \left(\left| G \, 1_V - \sum_{j \in \mathbb{J}_V} g_j \right| \right) &= o(\mu(V)); &\quad (3.2) \\
\sum_{j \in \mathbb{J}_V} \| g_j \|_{L^1} &= O(\mu(V)). &\quad (3.3)
\end{aligned}
$$

Then **(M5)** $\implies$ **(M1)***.*

Remark 3.3 The hypotheses of Proposition 3.2 above are less cumbersome than they appear. One should think of the very common situation in which $\mathcal{M}$ admits a partition of unity, $\sum_j \psi_j(x) \equiv 1$, where the ψ_j are nonnegative integrable functions which are roughly translations of one another. In many such cases one can expect $g_j := G\psi_j$ to verify all of the above conditions. At any rate, if the g_j are all positive or all negative then (3.3) follows from (3.2).

PROOF OF PROPOSITION 3.2. Fix $F, G \in \mathcal{G}$ and take any $\varepsilon > 0$. We denote for short $F^t := F \circ T^t$ and $g_V := \sum_{j \in \mathbb{J}_V} g_j$. (3.2) and **(A3)** imply that, for $\mu(V)$ large enough,

$$
\left| \frac{\mu(g_V)}{\mu(V)} - \overline{\mu}(G) \right| \le \left| \frac{\mu(g_V)}{\mu(V)} - \mu_V(G) \right| + |\mu_V(G) - \overline{\mu}(G)| \le \frac{\varepsilon}{3 \, |\overline{\mu}(F)|} \tag{3.4}
$$

and, since F is bounded,

$$
\left| \mu_V(F^t G) - \frac{\mu(F^t g_V)}{\mu(V)} \right| \le \frac{\varepsilon}{3}. \tag{3.5}
$$

On the other hand, **(M5)** implies that

$$\left|\mu(F^t g_j) - \overline{\mu}(F)\mu(g_j)\right| \le \vartheta(t)\|g_j\|_{L^1}, \tag{3.6}$$

where $\lim_{t\to\infty} \vartheta(t) = 0$ and $\vartheta(t)$ does not depend on j or V. Summing over $j \in \mathbb{J}_V$ and using (3.3), one gets

$$\left|\frac{\mu(F^t g_V)}{\mu(V)} - \frac{\overline{\mu}(F)\mu(g_V)}{\mu(V)}\right| \le \vartheta(t)\,\frac{\sum_{j\in\mathbb{J}_V}\|g_j\|}{\mu(V)} \le \frac{\varepsilon}{3}, \tag{3.7}$$

for both $\mu(V)$ and t large enough.

Putting together (3.5), (3.7) and (3.4), in that order, we conclude that there exists $M = M(\varepsilon)$ such that, for $\mu(V) \ge M$ and $t \ge M$,

$$\left|\mu_V(F^t G) - \overline{\mu}(F)\overline{\mu}(G)\right| \le \varepsilon, \tag{3.8}$$

which is precisely **(M1)**. Q.E.D.

4 Mixing for random walks

In this section we see how the previous definitions play out for a fairly representative family of infinite measure-preserving dynamical systems. These systems, which we introduce momentarily, generalize the random walk of Section 2.2 in two ways. First and foremost, they represent *all* the random walks in $\mathbb{Z}^d$. Secondly, they are invertible dynamical systems, which can be reduced, for example, to the noninvertible dynamical system of Section 2.2 by a mere restriction of the σ-algebra.

To begin with, let a random walk in $\mathbb{Z}^d$ be defined by the transition probabilities $\{p_\beta\}_{\beta\in\mathbb{Z}^d}$, with $p_\beta \ge 0$ and $\sum_\beta p_\beta = 1$. This means that, if the walker is in $\alpha \in \mathbb{Z}^d$, he will have probability p_β to move to $\alpha + \beta$ in the next step. We introduce some notation that will be useful later:

$$\mathbb{D} := \left\{\beta \in \mathbb{Z}^d \mid p_\beta > 0\right\} =: \{\beta^{(j)}\}_{j\in\mathbb{Z}_N} \tag{4.1}$$

is the set of the "active" directions for the random walk, endowed with some enumeration $\mathbb{Z}_N \ni j \mapsto \beta^{(j)} \in \mathbb{D}$. If $\mathbb{D}$ is infinite, then $N := \infty$ and $\mathbb{Z}_N := \mathbb{Z}^+$; if $\mathbb{D}$ is finite, then N denotes its cardinality and $\mathbb{Z}_N := \mathbb{Z}^+ \cap [1, N]$.

We view this random walk as a dynamical system $(\mathcal{M}, \mathscr{A}, \mu, T)$, where:

- $\mathcal{M} := \mathbb{Z}^d \times [0,1)^2$. If we denote $S_\alpha := \{\alpha\} \times [0,1)^2$, for $\alpha \in \mathbb{Z}^d$, then $\mathcal{M} = \bigcup_\alpha S_\alpha$ can be interpreted as the disjoint union of $\mathbb{Z}^d$ copies of the unit square.

- $\mathscr{A}$ is the natural σ-algebra for $\mathcal{M}$, i.e., the σ-algebra generated by all the Lebesgue-measurable subsets of S_α, $\forall \alpha \in \mathbb{Z}^d$, with the natural identification $S_\alpha \simeq [0,1)^2$.

- μ is the infinite measure that coincides with the Lebesgue measure when restricted to each S_α.

- In order to define T, set $q_0 = 0$ and, for $k \in \mathbb{Z}_N$,

$$q_k := \sum_{j=1}^{k} p_{\beta^{(j)}}; \qquad R_k := [q_{k-1}, q_k) \times [0,1). \tag{4.2}$$

Clearly $\{R_k\}_{k\in\mathbb{Z}_N}$ is a partition of $[0,1)^2$ into adjacent rectangles of height 1 and width, respectively,

$$q_k - q_{k-1} = p_{\beta^{(k)}} = \mu(R_k). \tag{4.3}$$

For $x = (\alpha, y) = (\alpha; y_1, y_2) \in \mathbb{Z}^d \times [0,1)^2$, let $k \in \mathbb{Z}_N$ be the unique positive integer such that $y \in R_k$ (equivalently, $q_{k-1} \le y_1 < q_k$). One defines

$$Tx = T(\alpha; y_1, y_2) := \left(\alpha + \beta^{(k)} \, ; \, p^{-1}_{\beta^{(k)}}(y_1 - q_{k-1}) \, , \, p_{\beta^{(k)}} y_2 + q_{k-1}\right). \tag{4.4}$$

Therefore T is a piecewise linear, hyperbolic, invertible map $\mathcal{M} \longrightarrow \mathcal{M}$ which preserves μ (because its determinant, in the variables (y_1, y_2), is 1). Denoting $R_{\alpha,k} := \{\alpha\} \times R_k$, it is easy to see that T is a Markov map for the partition $\{R_{\alpha,k}\}_{\alpha,k}$.

Now define $\psi : \mathcal{M} \longrightarrow \mathbb{Z}^d$ as $\psi(\alpha, y) = \alpha$. It is evident that, chosen $x = (0, y)$ at random in S_0 w.r.t. μ, the stochastic process $\{\psi(T^n x)\}_{n\in\mathbb{N}}$ is the random walk introduced at the beginning of the section, with initial position in the origin.

Moving on, we need to specify $\mathscr{V}$, the exhaustive family of sets that determines the infinite-volume limit: For all $\gamma = (\gamma_1, \ldots, \gamma_d) \in \mathbb{Z}^d$ and $r \in \mathbb{Z}^+$, the set

$$B_{\gamma,r} = \left\{\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}^d \mid \gamma_i - r \le \alpha_i \le \gamma_i + r, \ \forall i = 1, \ldots, d\right\} \tag{4.5}$$

is called a *square box in* $\mathbb{Z}^d$. Then we pose

$$\mathscr{V} := \left\{V = B_{\gamma,r} \times [0,1)^2 \;\middle|\; \gamma \in \mathbb{Z}^d, \ r \in \mathbb{Z}^+\right\} \tag{4.6}$$

Remark 4.1 Clearly, definition (4.5) does not capture all the square boxes in $\mathbb{Z}^d$, but only those whose side length is odd. This choice, made on grounds of simplicity, does not really limit the generality of $\mathscr{V}$, and indeed the forthcoming results can be proven even in the case when (4.6) is modified to include all square boxes.

Lemma 4.2 *The dynamical system and the exhaustive family defined above verify* **(A1)**.

When $\mathbb{D}$ is finite, the proof of Lemma 4.2 is rather straightforward, along the same lines as (2.9) for the first example of Section 2.2. In the general case, it has inessential technical complications, so we postpone it to Section A.1 of the Appendix.

In order to introduce our observables, we need some preliminary notation. Let $\mathscr{B} \subset \mathscr{A}$ be the σ-algebra on $\mathcal{M}$ generated by the partition $\{S_\alpha\}$ and, as is customary, $T^m \mathscr{B} = \{T^{-m} A \mid A \in \mathscr{B}\}$. Then, for $\ell, m \in \mathbb{Z}$ with $\ell \le m$, define

$$\mathscr{B}_{\ell,m} := T^\ell \mathscr{B} \vee T^{\ell-1} \mathscr{B} \cdots \vee T^{m+1} \mathscr{B} \vee T^m \mathscr{B}. \tag{4.7}$$

To fix the ideas, $\mathscr{B}_{0,1}$ is the σ-algebra corresponding to the partition $\{R_{\alpha,k}\}$. More generally, consider $\ell < 0 < m$: Recalling that $N = \#\mathbb{D}$ is the number of rectangles in each partition $\{R_k\}$ of S_α, one can see that the fundamental partition of $\mathscr{B}_{\ell,m}$ is made up of $N^{|\ell|+m}$ rectangles whose widths are bounded by λ^m and whose heights are bounded by $\lambda^{|\ell|}$, where $\lambda := \max\{p_\beta\}$. Finally, set

$$\begin{aligned} \mathscr{B}_{\ell,+\infty} &:= \bigvee_{m \in \mathbb{N}} \mathscr{B}_{\ell,m}; \\ \mathscr{B}_{-\infty,m} &:= \bigvee_{-\ell \in \mathbb{N}} \mathscr{B}_{\ell,m}; \\ \mathscr{B}_{-\infty,+\infty} &:= \bigvee_{-\ell, m \in \mathbb{N}} \mathscr{B}_{\ell,m}. \end{aligned} \tag{4.8}$$

If we exclude, now and for the remainder of the section, the trivial case where $N = 1$ (i.e., $p_\beta = 0$, $\forall \beta \ne \beta^{(1)}$), one has that $\lambda < 1$. This and the previous observation on the fundamental sets of $\mathscr{B}_{\ell,m}$ imply that the sets of $\mathscr{B}_{0,+\infty}$ are measurable unions of segments of the type $\{\alpha\} \times \{y_1\} \times [0, 1)$. We call those the *local stable manifolds* (LSMs) of the system and $\mathscr{B}_{0,+\infty}$ the *stable σ-algebra*, also denoted by $\mathscr{A}_s$. Analogously, $\mathscr{A}_u := \mathscr{B}_{-\infty,0}$ is called the *unstable σ-algebra* and its sets are measurable unions of *local unstable manifolds* (LUMs) $\{\alpha\} \times [0, 1) \times \{y_2\}$. Clearly, then, $\mathscr{B}_{-\infty,+\infty} = \mathscr{A}$.

We define several classes of global observables:

$$\mathcal{G}_m := \{F \in L^\infty(\mathcal{M}, \mathscr{B}_{-m,m}, \mu) \mid \exists\, \overline{\mu}(F) \text{ as per definition } \textbf{(A3)}\}; \tag{4.9}$$

$$\mathcal{G} := \overline{\bigcup_{m \in \mathbb{N}} \mathcal{G}_m}, \tag{4.10}$$

where the closure is meant in the L^∞ norm. One should notice that

Lemma 4.3 *Given the definitions* (4.9)-(4.10),

$$\mathcal{G} = \left\{ F \in \overline{\bigcup_{m \in \mathbb{N}} L^\infty(\mathcal{M}, \mathscr{B}_{-m,m}, \mu)} \;\middle|\; \exists\, \overline{\mu}(F) \right\}.$$

Proof. Let us prove the left-to-right inclusion. Given $F \in \mathcal{G}$ and $n \ge 1$, there exists an $F_n \in \bigcup_m \mathcal{G}_m$ such that $\|F_n - F\|_{L^\infty} \le 1/n$. This implies that

$$\overline{\mu}(F_n) - \frac{1}{n} \le \liminf_{V \nearrow \mathcal{M}} \mu_V(F) \le \limsup_{V \nearrow \mathcal{M}} \mu_V(F) \le \overline{\mu}(F_n) + \frac{1}{n}. \tag{4.11}$$

On the other hand, $\{\overline{\mu}(F_n)\}$ is a Cauchy sequence, because $\{F_n\}$ is. Its convergence thus proves the existence of $\overline{\mu}(F)$.

Conversely, if F is an observable in the closure of $\bigcup_m L^\infty(\mathcal{M}, \mathscr{B}_{-m,m}, \mu)$ for which $\overline{\mu}(F)$ exists, then, for any $\varepsilon > 0$, there exist $m \in \mathbb{N}$ and $F' \in L^\infty(\mathcal{M}, \mathscr{B}_{-m,m}, \mu)$ such that

$$\|F' - F\|_{L^\infty} \le \varepsilon/2. \tag{4.12}$$

Denoting $F'' := \mathbb{E}(F|\mathscr{B}_{-m,m})$, (4.12) implies that $\|F' - F''\|_{L^\infty} \le \varepsilon/2$, hence $\|F - F''\|_{L^\infty} \le \varepsilon$. Furthermore, since $\mathscr{V} \subset \mathscr{B} \subset \mathscr{B}_{-m,m}$, then $\overline{\mu}(F'')$ exists and equals $\overline{\mu}(F)$. This shows that $F'' \in \mathcal{G}_m$. Therefore $F \in \mathcal{G}$. Q.E.D.

Remark 4.4 In view of Lemma 4.3, one might wonder why we do not consider the more natural class

$$\mathcal{G}_\infty := \{F \in L^\infty(\mathcal{M}, \mathscr{A}, \mu) \mid \exists\, \overline{\mu}(F)\} \tag{4.13}$$

instead of $\mathcal{G}$. (The inclusion $\mathcal{G} \subset \mathcal{G}_\infty$ is clearly strict.) The reason, which will hardly surprise the "hyperbolic" dynamicist, is that we need to approximate a global observable with locally constant functions *uniformly* over $\mathcal{M}$, cf. (4.9)-(4.10). At any rate, $\mathcal{G}$ does not lack generality: for instance, any uniformly continuous F verifying **(A3)** belongs in that class.

As for the local observables, we also introduce countably many classes:

$$\begin{aligned} \mathcal{L}_m &:= L^1(\mathcal{M}, \mathscr{B}_{-m,m}, \mu); && (4.14)\\ \mathcal{L} &:= L^1(\mathcal{M}, \mathscr{A}, \mu). && (4.15) \end{aligned}$$

Prior to stating the main theorem of this section, we give a lemma that will help appreciate its statement. The proof of the lemma is given in Section A.2 of the Appendix.

Lemma 4.5 *Let $\{\beta^{(j)}\}_{j \in \mathbb{Z}_N} \subset \mathbb{Z}^d$ and $j' \in \mathbb{Z}_N$. Then*

$$\mathrm{span}_{\mathbb{Z}}\{\beta^{(j)} - \beta^{(j')}\}_{j \in \mathbb{Z}_N}$$

does not depend on j'.

Theorem 4.6 *Let $(\mathcal{M}, \mathscr{A}, \mu, T)$ be the dynamical system described above. Set $\nu := \max\{2, [d/2] + 1\}$, where $[\cdot]$ the integer part of a positive number, and suppose that*

(i) the probability distribution p has a finite ν-th momentum:

$$\sum_{\beta \in \mathbb{Z}^d} |\beta|^\nu p_\beta < \infty;$$

(ii) for a given $j' \in \mathbb{Z}_N$, $\mathrm{span}_{\mathbb{Z}}\{\beta^{(j)} - \beta^{(j')}\}_{j \in \mathbb{Z}_N} = \mathbb{Z}^d$.

Then the system is mixing in the following senses:

(a) **(M5)** *relative to* $\mathcal{G}_m$ *and* $\mathcal{L}_m$*, for all* $m \in \mathbb{N}$*;*

(b) **(M4)-(M3)** *relative to* $\mathcal{G}$ *and* $\mathcal{L}$*;*

(c) **(M2)** *relative to* $\mathcal{G}_m$*, for all* $m \in \mathbb{N}$*, with the extra requirement that* F *be* $\mathbb{Z}^d$*-periodic, i.e.,* $F(\alpha, y) = F(0, y)$*,* $\forall \alpha \in \mathbb{Z}^d$*,* $\forall y \in [0,1)^2$*;*

(d) **(M1)** *relative to* $\mathcal{G}$*.*

Remark 4.7 Condition *(ii)* is essential as it has to do with the irreducibility of the random walk [S]. In fact, assuming for simplicity that $0 \in \mathbb{D}$, if $\mathrm{span}_{\mathbb{Z}}(\mathbb{D}) \neq \mathbb{Z}^d$, then the random walk is reducible and the system cannot be mixing in any sense, as is ascertained via the global observable

$$F(\alpha, y) := \begin{cases} 1, & \text{for } \alpha \in \mathrm{span}_{\mathbb{Z}}(\mathbb{D}); \\ 0, & \text{otherwise.} \end{cases} \tag{4.16}$$

One last observation that may be of interest is that statement *(c)* is far from optimal. **(M2)** holds for a much larger class of global observables, depending especially on the distribution $\{p_\beta\}$. In the formulation of Theorem 4.6, however, I was mainly interested in a nontrivial case in which **(M2)** could be verified easily.

5 Proof of Theorem 4.6

Since the proof is rather lengthy, we will divide it into pieces, or *stages*, as follows:

Stage 1: We prove *(a)* using three extra assumptions.

Stage 2: We remove one of the extra assumptions.

Stage 3: We remove the remaning two extra assumptions.

Stage 4: We prove *(b)-(d)*.

5.1 Stage 1: Extra assumptions

Let us initially assume that:

(E1) $F \in \mathscr{A}_s = \mathscr{B}_{0,+\infty}$;

(E2) $\mathcal{L}_m$ only comprises indicator functions of the type $g = 1_Q$, where Q is a fundamental set of $\mathscr{B}_{-m,0}$ and $Q \subseteq S_0$;

(E3) the random walk has zero drift, i.e., $\displaystyle\sum_{\beta \in \mathbb{Z}^d} \beta \, p_\beta = 0$.

From (E2), Q is a rectangle of the type $\{0\} \times [0,1) \times I$. We denote the length of I by

$$h := |I| = \mu(Q). \tag{5.1}$$

In this setting, **(M5)** amounts to showing that

$$\lim_{n\to\infty} \frac{1}{\mu(|g|)} \int_{\mathcal{M}} (F \circ T^n) g \, d\mu = \lim_{n\to\infty} \frac{1}{h} \int_{T^n Q} F \, d\mu = \overline{\mu}(F), \tag{5.2}$$

uniformly in Q, that is, with a speed of convergence that does not depend on the choice of I. (In the above we have used the invariance of μ and the fact that $\mu(|g|) = h = \mu(g)$.) Achieving this will be Stage 1 of the proof.

Q can be thought of as partitioned into $\{Q \cap R_{0,j}\}_{j \in \mathbb{Z}_N}$, which are rectangles of width $p_{\beta^{(j)}}$ and height h. By construction, T acts on each such rectangle by stretching it horizontally and shrinking it vertically by a factor $p_{\beta^{(j)}}^{-1}$, and then mapping the resulting rectangle, of width 1, rigidly into $S_{\beta^{(j)}}$.

Iterating this procedure n times, we obtain

$$T^n Q = \bigcup_{j_1,\ldots,j_n \in \mathbb{Z}_N} Q_{j_1,\ldots,j_n} := \bigcup_{j_1,\ldots,j_n \in \mathbb{Z}_N} \{\alpha_{j_1,\ldots,j_n}\} \times [0,1) \times I_{j_1,\ldots,j_n}, \tag{5.3}$$

which is a disjoint union of N^n thin rectangles of width 1. Each $Q_{j_1,\ldots,j_n}$ is the set of all the points $T^n x$ for which the trajectory of $x \in Q$ followed the *itinerary* $S_{\beta^{(j_1)}}$, $S_{\beta^{(j_1)}+\beta^{(j_2)}}$, ..., $S_{\alpha_{j_1,\ldots,j_n}}$, where

$$\alpha_{j_1,\ldots,j_n} := \beta^{(j_i)} + \beta^{(j_2)} + \cdots + \beta^{(j_n)}. \tag{5.4}$$

Therefore, recalling (5.1), the height of $Q_{j_1,\ldots,j_n}$ is

$$h_{j_1,\ldots,j_n} := |I_{j_1,\ldots,j_n}| = \mu(Q_{j_1,\ldots,j_n}) = h \, p_{\beta^{(j_1)}} p_{\beta^{(j_2)}} \cdots p_{\beta^{(j_n)}}. \tag{5.5}$$

Since $F \in \mathscr{A}_s$, we can write, with a slight abuse of notation, $F(\alpha; y_1, y_2) = F_\alpha(y_1)$. Then

$$\begin{aligned}
\int_{T^n Q} F \, d\mu &= \sum_{j_1,\ldots,j_n \in \mathbb{Z}_N} \int_{Q_{j_1,\ldots,j_n}} F \, d\mu = \\
&= \sum_{j_1,\ldots,j_n \in \mathbb{Z}_N} \int_{I_{j_1,\ldots,j_n}} \int_0^1 F_{\alpha_{j_1,\ldots,j_n}}(y_1) \, dy_1 \, dy_2 = \\
&= \sum_{j_1,\ldots,j_n \in \mathbb{Z}_N} h \, p_{\beta^{(j_1)}} p_{\beta^{(j_2)}} \cdots p_{\beta^{(j_n)}} \, f_{\alpha_{j_1,\ldots,j_n}},
\end{aligned} \tag{5.6}$$

having used (5.3), (5.5) and the following definition:

$$f_\alpha := \int_0^1 F_\alpha(y_1) \, dy_1. \tag{5.7}$$

In view of (5.4) and (4.1), another way to write (5.6) is

$$\int_{T^n Q} F \, d\mu = h \sum_{\beta^{(1)},\ldots,\beta^{(n)} \in \mathbb{Z}^d} p_{\beta^{(1)}} p_{\beta^{(2)}} \cdots p_{\beta^{(n)}} \, f_{\beta^{(1)}+\ldots+\beta^{(n)}}. \tag{5.8}$$

5.2 Fourier analysis

The technical backbone of the proof is Fourier analysis on $\mathbb{Z}^d$, for which we proceed to establish the necessary notation [K, R]. Let $a := \{a_\alpha\}_{\alpha \in \mathbb{Z}^d} \in \ell^s(\mathbb{Z}^d; \mathbb{C})$, with $s \in [1, \infty]$. Its Fourier transform is denoted

$$\widetilde{a}(\theta) = \widetilde{a}(\theta_1, \dots, \theta_d) := \sum_{\alpha \in \mathbb{Z}^d} a_\alpha \, e^{\imath \alpha \cdot \theta} = \sum_{\alpha \in \mathbb{Z}^d} a_\alpha \, e^{\imath(\alpha_1 \theta_1 + \ldots + \alpha_d \theta_d)}, \tag{5.9}$$

where $\imath$ is the imaginary unit and $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{T}^d := (\mathbb{R}/2\pi\mathbb{Z})^d$. If $s > 2$, (5.9) must be intended in the weak sense, cf. (5.11). The corresponding inverse transform is given by

$$a_\alpha = \int_{\mathbb{T}^d} \widetilde{a}(\theta) \, e^{-\imath \alpha \cdot \theta} \, \underline{d}\theta, \tag{5.10}$$

where $\underline{d}\theta := (2\pi)^{-d} d\theta$. The Parseval formula, in this setting, reads as follows: If $b = \{b_\alpha\} \in \ell^{s'}(\mathbb{Z}^d; \mathbb{C})$, with $1/s + 1/s' = 1$, then

$$\langle a, b \rangle := \sum_{\alpha \in \mathbb{Z}^d} \overline{a_\alpha} \, b_\alpha = \int_{\mathbb{T}^d} \overline{\widetilde{a}(\theta)} \, \widetilde{b}(\theta) \, \underline{d}\theta, \tag{5.11}$$

the bar denoting complex conjugation. Another standard result that we need is the duality between convolution and product: If

$$(a * b)_\alpha := \sum_{\beta \in \mathbb{Z}^d} a_\beta \, b_{\alpha - \beta} = \sum_{\beta \in \mathbb{Z}^d} a_{\alpha - \beta} \, b_\beta \tag{5.12}$$

is well-defined in the proper or weak sense, then

$$\widetilde{(a * b)}(\theta) = \widetilde{a}(\theta) \, \widetilde{b}(\theta). \tag{5.13}$$

Applying these concepts to our case, we see that $f := \{f_\alpha\} \in \ell^\infty$ by construction (because $F \in \mathcal{G} \subset L^\infty(\mathcal{M})$) so $\widetilde{f}$ is a distribution on $\mathbb{T}^d$. On the other hand, $p := \{p_\alpha\} \in \ell^1$, which makes $\widetilde{p}$ continuous (it is actually much more than that, cf. Section 5.3). In particular, since p is a probability distribution on $\mathbb{Z}^d$, $\widetilde{p}(0) = 1$. Defining

$$p^{(n)} := \underbrace{p * \cdots * p}_{n \text{ times}} \tag{5.14}$$

we can rearrange (5.8) into

$$\frac{1}{h} \int_{T^{-n}Q} F \, d\mu = \left\langle f, p^{(n)} \right\rangle = \int_{\mathbb{T}^d} \overline{\widetilde{f}(\theta)} \, \widetilde{p^{(n)}}(\theta) \, \underline{d}\theta = \int_{\mathbb{T}^d} \overline{\widetilde{f}(\theta)} \, \widetilde{p}^n(\theta) \, \underline{d}\theta, \tag{5.15}$$

where we have used (5.13), (5.14) and the fact that $f_\alpha \in \mathbb{R}$. Hence Stage 1 of the proof, cf. (5.2), reduces to showing that

$$\lim_{n \to \infty} \left\langle f, p^{(n)} \right\rangle = \overline{\mu}(F). \tag{5.16}$$

Now recall definition (4.5). For $r \in \mathbb{N}$, let

$$q_\alpha^{(r)} := \begin{cases} (2r+1)^{-d}, & \text{if } \alpha \in B_{0,r}; \\ 0, & \text{otherwise,} \end{cases} \tag{5.17}$$

define a function $q^{(r)} : \mathbb{Z}^d \longrightarrow \mathbb{R}$. Its Fourier transform is easily computed to be

$$\widetilde{q^{(r)}}(\theta_1, \ldots, \theta_d) = \prod_{i=1}^{d} \frac{\sin\left((r+1/2)\theta_i\right)}{(r+1/2)\sin\theta_i}. \tag{5.18}$$

In view of (4.6), let us denote $V_{\alpha,r} := B_{\alpha,r} \times [0,1)^2$. Since F verifies **(A3)** and the infinite-volume limit is μ-uniform (cf. Definition 2.3), we have that

$$\lim_{r\to\infty} \frac{1}{\mu(V_{\alpha,r})} \int_{V_{\alpha,r}} F \, d\mu = \lim_{r\to\infty} \frac{1}{(2r+1)^d} \sum_{\beta \in B_{\alpha,r}} f_\beta =$$

$$\lim_{r\to\infty} \left(f * q^{(r)}\right)_\alpha = \overline{\mu}(F), \tag{5.19}$$

uniformly in α. (We have used notation (5.7).) This, in turn, yields

$$\lim_{r\to\infty} \left\langle f, q^{(r)} * p^{(n)} \right\rangle = \lim_{r\to\infty} \left\langle f * q^{(r)}, p^{(n)} \right\rangle = \overline{\mu}(F) \tag{5.20}$$

uniformly in n, because $p^{(n)}$ is a probability distribution on $\mathbb{Z}^d$. (In the first equality we have used the fact that $q_\alpha^{(r)} = q_{-\alpha}^{(r)}$, by construction.) This fact implies that, for any sequence $\{r_n\} \subset \mathbb{N}$ with $r_n \to \infty$,

$$\lim_{n\to\infty} \left\langle f, q^{(r_n)} * p^{(n)} \right\rangle = \overline{\mu}(F). \tag{5.21}$$

Therefore, comparing (5.21) with (5.16), we see that Stage 1 is achieved once we have shown that there exists a diverging sequence $\{r_n\}$ of natural numbers such that

$$\lim_{n\to\infty} \left\langle f, p^{(n)} - q^{(r_n)} * p^{(n)} \right\rangle = 0. \tag{5.22}$$

For an *a fortiori* choice of $\{r_n\}$, let us set

$$g_\alpha^{(n)} := p_\alpha^{(n)} - \left(q^{(r_n)} * p^{(n)}\right)_\alpha, \tag{5.23}$$

which gives

$$\widetilde{g^{(n)}}(\theta) = \left(1 - \widetilde{q^{(r_n)}}(\theta)\right) \widetilde{p}^n(\theta). \tag{5.24}$$

A convenient estimate in view of (5.22) is

$$\left|\left\langle f, g^{(n)} \right\rangle\right| \le \|f\|_{\ell^\infty} \left\|g^{(n)}\right\|_{\ell^1} = \|f\|_{\ell^\infty} \left\|\widetilde{g^{(n)}}\right\|_{\mathcal{A}} \le C \left\|\widetilde{g^{(n)}}\right\|_{H^\nu}, \tag{5.25}$$

where the norms $\|\cdot\|_{\mathcal{A}}$ and $\|\cdot\|_{H^\nu}$ are introduced in Section A.3 of the Appendix (cf. in particular Lemma A.2 and notice that $\|\cdot\|_{H^{\overline{\nu}}} \le \|\cdot\|_{H^\nu}$).

Therefore (5.22) will be proved once we establish that

$$\left\|\widetilde{g^{(n)}}\right\|_{H^\nu} \le \left\|\widetilde{g^{(n)}}\right\|_{L^1} + \sum_{i=1}^{d} \left\|\partial_i^\nu \widetilde{g^{(n)}}\right\|_{L^2} \to 0, \quad \text{as } n \to \infty, \tag{5.26}$$

for a suitable choice of $\{r_n\}$ in definition (5.23). Here $\partial_i := \partial/\partial\theta_i$, acting on functions $\mathbb{T}^d \longrightarrow \mathbb{C}$.

5.3 Properties of $\widetilde{q^{(r)}}$ and $\widetilde{p}$

In view of the above goal we need to study some properties of the functions $\widetilde{q^{(r)}}$ and $\widetilde{p}$.

Remark 5.1 None of the proofs in this section will use (E3).

As a start, let us notice that $\widetilde{q^{(r)}}$ is C^∞ by construction and $\widetilde{p}$ is C^ν, with $\nu \geq 2$, by hypothesis *(i)* of Theorem 4.6.

Lemma 5.2 *Fix $r \in \mathbb{Z}^+$. On $\mathbb{T}^d$, $\widetilde{q^{(r)}}(0) = \widetilde{p}(0) = 1$ and, for $\theta \neq 0$,*

$$|\widetilde{q^{(r)}}(\theta)| < 1, \qquad |\widetilde{p}(\theta)| < 1.$$

Proof. We first prove the assertions on $\widetilde{p}$. Since p is a probability distribution, $\widetilde{p}(0) = 1$ and $|\widetilde{p}(\theta)| \leq 1$, $\forall \theta$. Suppose by contradiction that $\exists \theta' \in \mathbb{T}^d$, $\theta' \neq 0$, such that $|\widetilde{p}(\theta')| = 1$, that is,

$$\widetilde{p}(\theta') = \sum_{j \in \mathbb{Z}_N} p_{\beta^{(j)}} \, e^{\imath \theta' \cdot \beta^{(j)}} = e^{\imath a}, \tag{5.27}$$

for some $a \in \mathbb{R}$. Since $p_{\beta^{(j)}} > 0$ and $\sum_j p_{\beta^{(j)}} = 1$, necessarily $e^{\imath \theta' \cdot \beta^{(j)}} = e^{\imath a}$, $\forall j \in \mathbb{Z}_N$, whence, $\forall j, j'$,

$$e^{\imath \theta' \cdot (\beta^{(j)} - \beta^{(j')})} = 1. \tag{5.28}$$

Let us define the *character* (i.e., the homomorphism $\mathbb{Z}^d \longrightarrow S^1 \subset \mathbb{C}$) $\eta_{\theta'}(\alpha) := e^{\imath \theta' \cdot \alpha}$. It is easy to see that $\eta_{\theta'}$ is not the trivial character (which instead corresponds to $\theta' = 0$; this is a particular case of the so-called Pontryagin Duality [R, Thm. 2.1.2]). On the other hand, (5.28) reads $\eta_{\theta'}(\beta^{(j)} - \beta^{(j')}) = 1$ and hypothesis *(ii)* of Theorem 4.6 implies that $\eta_{\theta'} \equiv 1$, thereby creating a contradiction.

As for the assertions on $\widetilde{q^{(r)}}$, there is nothing more to prove, because $\widetilde{q^{(r)}}$ satisfies the same properties as $\widetilde{p}$, insofar as the above argument is concerned. Q.E.D.

Notational convention. *From now on, C will denote a generic universal constant. This means that its actual value will vary from formula to formula but will never depend on n, r, or θ.*

Lemma 5.3 *If we think of $\widetilde{p}$ as a periodic function on $\mathbb{R}^d$ (as opposed to a function on $\mathbb{T}^d$), there exists a neighborhood $\mathcal{U}$ of $\theta = 0$ and a positive constant C such that, for $\theta \in \mathcal{U}$,*

$$|\widetilde{p}(\theta)| \leq 1 - C|\theta|^2 = 1 - C\left(\theta_1^2 + \cdots + \theta_d^2\right).$$

PROOF. As $\theta \to 0$,

$$\widetilde{p}(\theta) = 1 + \imath v \cdot \theta + \mathcal{O}(|\theta|^2), \tag{5.29}$$

where $v \in \mathbb{R}^d$ is the *drift* of the random walk, defined as

$$v := \sum_{\beta \in \mathbb{Z}^d} \beta p_\beta. \tag{5.30}$$

The Lagrange remainder in (5.29) holds because $\widetilde{p}$ is at least C^2. Hence $|\widetilde{p}(\theta)|^2 = 1 + \mathcal{O}(|\theta|^2)$, which implies the assertion. Q.E.D.

Lemma 5.4 *Regarding $\widetilde{q^{(r)}}$ as a periodic function on $\mathbb{R}^d$, one has that the following expansion,*

$$\widetilde{q^{(r)}}(\theta_1, \ldots, \theta_d) = \prod_{i=1}^{d} \left(1 + \sum_{j=1}^{\infty} \xi_j(r)\, \theta_i^{2j} \right),$$

holds uniformly on the compact subsets of $\mathbb{R}^d$. Furthermore, as $r \to \infty$,

$$|\xi_j(r)| \le C \frac{r^{2j}}{(2j)!}.$$

PROOF. By the factorizability of $q^{(r)}$, it is sufficient to treat the case $d = 1$.

Since $q^{(r)}$ is compactly supported in $\mathbb{Z}^d$, $\widetilde{q^{(r)}}(\theta)$ is an entire function of θ, which we have already calculated in (5.17). Its Taylor expansion at the origin is even and its (even) terms are

$$\xi_j(r) = \frac{1}{(2j)!} \, \partial^{2j} \widetilde{q^{(r)}}(0) = \frac{1}{(2j)!} \, \frac{(-1)^j}{2r+1} \sum_{\alpha=-r}^{r} \alpha^{2j}. \tag{5.31}$$

This gives $\xi_0(r) \equiv 1$ and the desired estimates. Q.E.D.

We estimate the norms in the r.h.s. of (5.26) by splitting the corresponding integrals into two parts: one over $\mathcal{B}_n$, which is the ball of center 0 and radius $n^{-(1-\varepsilon)/2}$ in $\mathbb{T}^d$, and the other over $\mathbb{T}^d \setminus \mathcal{B}_n$. Here $\varepsilon > 0$ is a small constant to be fixed later and n is a large integer.

Lemma 5.5 *There exists a $\kappa > 0$ such that, for n sufficiently large,*

$$\max_{\mathbb{T}^d \setminus \mathcal{B}_n} |\widetilde{p}|^n \le e^{-\kappa n^\varepsilon}.$$

PROOF. By elementary Taylor approximations, Lemma 5.3 implies that there exists a constant $\kappa > 0$ such that, for $\theta \in \mathcal{U}$,

$$|\widetilde{p}(\theta)| \le e^{-\kappa |\theta|^2}. \tag{5.32}$$

For n large enough, by Lemma 5.2, the continuity of $\widetilde{p}$ and the compactness of $\mathbb{T}^d$,

$$\max_{\theta \in \mathbb{T}^d \setminus \mathcal{B}_n} |\widetilde{p}(\theta)| = \max_{\theta \in \partial \mathcal{B}_n} |\widetilde{p}(\theta)| \le e^{-\kappa n^{-1+\varepsilon}} \tag{5.33}$$

(in the last inequality we have used (5.32), which applies because, for n large, $\mathcal{B}_n \subset \mathcal{U}$). Q.E.D.

5.4 End of Stage 1

Let us begin to attack (5.26) by estimating $\|\widetilde{g^{(n)}}\|_{L^1}$. First of all,

$$\left\|\widetilde{g^{(n)}}\right\|_{L^1(\mathbb{T}^d\setminus\mathcal{B}_n)} \leq (2\pi)^d 2e^{-\kappa n^\varepsilon}, \tag{5.34}$$

by Lemma 5.2 applied to $\widetilde{q^{(r)}}$ and Lemma 5.5—see (5.24). Again, applying Lemma 5.2 to both $\widetilde{q^{(r)}}$ and $\widetilde{p}$,

$$\left\|\widetilde{g^{(n)}}\right\|_{L^1(\mathcal{B}_n)} \leq \frac{C}{n^{d(1-\varepsilon)/2}} \tag{5.35}$$

where C does not depend on r, that is, r_n.

As for the remaining terms in the r.h.s. of (5.26), clearly

$$\partial_i^\nu \widetilde{g^{(n)}} = \sum_{k+l=\nu} \binom{\nu}{k} \partial_i^k \left(1-\widetilde{q^{(r)}}\right) \partial_i^l \widetilde{p}^n. \tag{5.36}$$

Let us estimate (5.36) on $\mathbb{T}^d\setminus\mathcal{B}_n$. Fixing $l \geq 1$, one verifies by repeated differentiation that

$$\partial_i^l \widetilde{p}^n = \sum_{w=1}^{l} \sum_{j_1+\ldots+j_w=l} C^l_{j_1,\ldots,j_w} \frac{n!}{(n-w)!} \widetilde{p}^{n-w} (\partial_i^{j_1}\widetilde{p})(\partial_i^{j_2}\widetilde{p}) \cdots (\partial_i^{j_w}\widetilde{p}), \tag{5.37}$$

where the combination of the two sums above represents the sum over all the partitions $\{j_1, j_2, \ldots, j_w\}$ of l (i.e., $j_u \geq 1$ and $j_1 + j_2 + \ldots + j_w = l$), with any cardinality w, and $C^l_{j_1,\ldots,j_w} \in \mathbb{N}$ is a combinatorial coefficient independent of n. Since $\widetilde{p} \in C^\nu$, all the derivatives that appear on the r.h.s. of (5.37) are continuous functions of θ. Therefore, by Lemma 5.5,

$$\max_{\mathbb{T}^d\setminus\mathcal{B}_n} \left|\partial_i^l \widetilde{p}^n\right| \leq Cn^l e^{-\kappa(n-l)^\varepsilon}. \tag{5.38}$$

As for the other factors in the r.h.s. of (5.36), for $k \geq 1$ we use definition (5.17) to estimate

$$\max_{\mathbb{T}^d} \left|\partial_i^k \widetilde{q^{(r)}}\right| \leq \sum_{\alpha\in\mathbb{Z}^d} \left|\alpha_i^k q_\alpha^{(r)}\right| = \frac{1}{2r+1} \sum_{\alpha_i=-r}^{r} |\alpha_i|^k \leq Cr^k. \tag{5.39}$$

(For $k = 0$, we already know that $1 - \widetilde{q^{(r)}}$ is bounded.) Using (5.38)-(5.39) into (5.36), we get

$$\left\|\partial_i^\nu \widetilde{g^{(n)}}\right\|^2_{L^2(\mathbb{T}^d\setminus\mathcal{B}_n)} \leq Cr^{2\nu} n^{2\nu} e^{2\kappa(n-\nu)^\varepsilon}, \tag{5.40}$$

which tends to zero, as $n \to \infty$, provided that $r = r_n$ grows no faster than a power of n. This will be verified *a fortiori*, see (5.41).

The estimation of the last term, namely $\|\partial_i^\nu \widetilde{g^{(n)}}\|_{L^2(\mathcal{B}_n)}$, is the most delicate, therefore we organize most of the computations involved in the following

Lemma 5.6 *For $\nu \in \mathbb{Z}^+$ (not necessarily as in the statement of Theorem 4.6), assume $\widetilde{p} \in C^\nu$. Then take any sequence of positive numbers Λ_n, with $\Lambda_n \to 0$. For n large enough and uniformly for*

$$1 \le r \le \Lambda_n \, n^{(1-\varepsilon)/2},$$

one has

$$\max_{\mathcal{B}_n} \left| \partial_i^\nu \widetilde{g^{(n)}} \right| \le C r^{2\nu} n^{(-2+2\varepsilon+\nu(1+\varepsilon))/2}.$$

Once Lemma 5.6 is established, which will happen momentarily, we can finally choose both the sequence $\{r_n\}$ and the parameter ε: the former will be any diverging sequence such that

$$r_n \le C n^\varepsilon, \tag{5.41}$$

whereas

$$\varepsilon < \min \left\{ \frac{1}{3}, \frac{1}{2(5\nu + 2 + d/2)} \right\}. \tag{5.42}$$

The condition $\varepsilon < 1/3$ is necessary in order to apply Lemma 5.6 to $r = r_n$. In fact, via (5.41), $r_n n^{-(1-\varepsilon)/2} \le C n^{-(1-3\varepsilon)/2} =: \Lambda_n$. The other inequality is needed when we use the assertion of Lemma 5.6 to estimate

$$\begin{aligned} \left\| \partial_i^\nu \widetilde{g^{(n)}} \right\|^2_{L^2(\mathcal{B}_n)} &\le C r_n^{4\nu} \, n^{-2+2\varepsilon+\nu(1+\varepsilon)} \, n^{-d(1-\varepsilon)/2} \le \\ &\le C n^{-2+\nu-d/2+(5\nu+2+d/2)\varepsilon}. \end{aligned} \tag{5.43}$$

For ε so small that $(5\nu + 2 + d/2)\varepsilon < 1/2$ the above vanishes, as $n \to \infty$, because $\nu - d/2 \le 3/2$ (for $d = 1$, it equals $3/2$; for $d \ge 2$ it equals 1 or 1/2, depending on d being even or odd).

In conclusion, estimates (5.34), (5.35), (5.40) and (5.43) prove (5.26), thus (5.22), thus (5.16) and, lastly, (5.2), uniformly in Q as requested. This ends Stage 1 of the proof.

Proof of Lemma 5.6. Throughout the proof it is understood that $\theta \in \mathcal{B}_n$, i.e., $|\theta| \le n^{-(1-\varepsilon)/2}$. The condition on r ensures that

$$\left| \xi_j(r) \theta^{2j} \right| \le C r^{2j} n^{-(1-\varepsilon)j} \le C \Lambda_n^{2j}, \tag{5.44}$$

which implies that, for n so large that $\Lambda_n < 1$, the expansion of Lemma 5.4 and all those derived by it by differentiation w.r.t. θ_i are meaningful. More importantly, as $n \to \infty$ and uniformly in r as described in the statement of the lemma, each such expansion can be approximated by a convenient upper bound on its first term. We indicate this with the symbol $\sim$: for example,

$$1 - \widetilde{q^{(r)}}(\theta) \;\sim\; |\xi_1(r)\theta|^2 \le C r^2 n^{-(1-\varepsilon)} \tag{5.45}$$

and, for $k \geq 1$,

$$\left|\partial_i^k \widetilde{q^{(r)}}(\theta)\right| \sim \left\{ \begin{array}{lll} (k+1)!\,|\xi_{(k+1)/2}(r)\theta_i| & \leq & Cr^{k+1}n^{-(1-\varepsilon)/2}, \quad \text{if } k \text{ is odd,} \\ k!\,|\xi_{k/2}(r)| & \leq & Cr^k, \quad \text{if } k \text{ is even.} \end{array} \right. \tag{5.46}$$

Furthermore, (E3) is equivalent to $\nabla\widetilde{p}(0) = 0$, which implies

$$|\partial_i \widetilde{p}(\theta)| \leq C|\theta| \leq Cn^{-(1-\varepsilon)/2}. \tag{5.47}$$

We will proceed by induction on $\nu \geq 1$. When $\nu = 1$ our function reads

$$\partial_i \widetilde{g^{(n)}} = -\partial_i \widetilde{q^{(r)}}\, \widetilde{p}^n + \left(1 - \widetilde{q^{(r)}}\right) n\widetilde{p}^{n-1}\, \partial_i \widetilde{p}. \tag{5.48}$$

Hence, from (5.45)-(5.47), and Lemma 5.2 applied to $\widetilde{p}$,

$$\max_{\mathcal{B}_n} \left|\partial_i \widetilde{g^{(n)}}\right| \leq Cr^2 n^{(-1+3\varepsilon)/2}, \tag{5.49}$$

proving the assertion for $\nu = 1$.

Now we assume the assertion with ν and set out to prove the one with $\nu + 1$. In practice, this means that increasing the order of the derivative by one must worsen the inequality of Lemma 5.6 at most by a factor $r^2 n^{(1+\varepsilon)/2}$.

We apply ∂_i to (5.36). On the r.h.s., ∂_i can either hit $\partial_i^k(1 - \widetilde{q^{(r)}})$ or $\partial_i^l \widetilde{p}^n$. Let us analyze the two cases separately.

In the first case, assuming for the moment $k \geq 1$, we see via (5.46) that

$$\left|\partial_i^{k+1} \widetilde{q^{(r)}}(\theta)\right| \leq \left\{ \begin{array}{ll} Cr^{k+1} = \left(Cr^{k+1}\, n^{-(1-\varepsilon)/2}\right) n^{(1-\varepsilon)/2}, & \text{if } k \text{ is odd,} \\ Cr^{k+2} n^{-(1-\varepsilon)/2} = \left(Cr^k\right) r^2\, n^{-(1-\varepsilon)/2}, & \text{if } k \text{ is even.} \end{array} \right. \tag{5.50}$$

The terms within parentheses in the above represent the estimates for $\partial_i^k \widetilde{q^{(r)}}$, respectively for k odd and even, coming from (5.46). Also, for $k = 0$,

$$\partial_i \widetilde{q^{(r)}} \leq Cr^2\, n^{-(1-\varepsilon)/2} = \left(Cr^2\, n^{-(1-\varepsilon)}\right) n^{(1-\varepsilon)/2}. \tag{5.51}$$

Again, the term in the parentheses is the estimate for $1 - \widetilde{q^{(r)}}$ coming from (5.45). In any event, applying another derivative to the term $\partial_i^k(1 - \widetilde{q^{(r)}})$ will change our estimate at most by a factor $r^2 n^{(1-\varepsilon)/2}$, which is consistent with our inductive step.

In the second case, we use the expansion (5.37): ∂_i can either hit $\widetilde{p}^{n-w}$ or one of the $\partial_i^{j_u} \widetilde{p}$. In this first sub-case,

$$\left|\partial_i\, \widetilde{p}^{n-w}(\theta)\right| = (n-w)\left|\widetilde{p}^{n-w-1}(\theta)\, \partial_i \widetilde{p}(\theta)\right| \leq Cn|\theta| \leq Cn^{(1+\varepsilon)/2}, \tag{5.52}$$

via (5.47). As for the second sub-case, without loss of generality, we assume the worst-case estimate for $\partial_i^{j_u} \widetilde{p}$ on $\mathcal{B}_n$, that is,

$$\left|\partial_i^{j_u} \widetilde{p}(\theta)\right| \leq \left\{ \begin{array}{ll} C|\theta| \leq Cn^{-(1-\varepsilon)/2}, & \text{if } j_u = 1, \\ C, & \text{if } j_u \geq 2. \end{array} \right. \tag{5.53}$$

This implies that increasing by one the order of the derivative in the l.h.s. of (5.53) will worsen our most conservative estimate at most by a factor $n^{(1-\varepsilon)/2}$. Considering (5.52) as well, we conclude that applying another derivative to $\partial_i^l \widetilde{p}^n$ will change its estimate at most by a factor $n^{(1+\varepsilon)/2}$, which is again consistent with our inductive step. Q.E.D.

Remark 5.7 The careful reader might worry that the unrigorous use of the symbol C for a generic constant may jeopardize the above proof. It does not, since all the constants that have been used do not depend on r or n. In principle, they may depend on k (though it is easy to see that they do not), or i, or j_u, but these integers only take on a finite number of values, so bounds can be found that do not depend on any of the variables.

5.5 Stage 2: Removing (E3)

If, contrary to assumption (E3), $v = -\imath \nabla \widetilde{p}(0) \neq 0$, cf. (5.30), we define $\delta^{(n)} \in \mathbb{Z}^d$ to be the (not necessarily unique) lattice point for which $\delta^{(n)}/n$ best approximates $v \in \mathbb{R}^d$. One clearly has

$$\left| \frac{\delta_i^{(n)}}{n} - v_i \right| \leq \frac{1}{2n}, \tag{5.54}$$

where the subscript i denotes, as usual, the i-th component of a d-dimensional vector. Now, for $\theta \in (-\pi, \pi)^d$, set

$$\widetilde{\pi}_n(\theta) := \widetilde{p}(\theta)\, e^{-\imath(\delta^{(n)} \cdot \theta)/n}. \tag{5.55}$$

We want to interpret $\widetilde{\pi}_n$ as a generally discontinuous function $\mathbb{T}^d \longrightarrow \mathbb{C}$. On the other hand, $\widetilde{\pi}_n^n$ is a smooth function of $\mathbb{T}^d$ and, by (5.24) and Lemma A.1 of the Appendix,

$$\begin{aligned} \left\| \widetilde{g^{(n)}} \right\|_{\mathcal{A}} &= \left\| \left(1 - \widetilde{q^{(r_n)}}\right) \widetilde{p}^n \right\|_{\mathcal{A}} = \\ &= \left\| \left(1 - \widetilde{q^{(r_n)}}\right) \widetilde{p}^n \, \widetilde{\omega}_{-\delta^{(n)}} \right\|_{\mathcal{A}} = \\ &= \left\| \left(1 - \widetilde{q^{(r_n)}}\right) \widetilde{\pi}_n^n \right\|_{\mathcal{A}} \end{aligned} \tag{5.56}$$

(having used notation (A.9) as well). Comparing the above with (5.25), in view of our goal (5.26), we see that it is sufficient to repeat all the estimations of Sections 5.3–5.4, replacing $\widetilde{p}$ with $\widetilde{\pi}_n$. This is no problem, except for estimate (5.47)—which is also reflected in (5.52) and (5.53). (Consider Remark 5.1 and the fact that Lemmas 5.2 and 5.5 cannot distinguish between $\widetilde{p}$ and $\widetilde{\pi}_n$.)

In order to find an effective substitute for (5.47), we write, for $\theta \in \mathcal{B}_n$,

$$\partial_i \widetilde{\pi}_n(\theta) = \partial_i \widetilde{\pi}_n(0) + u_n(\theta') \cdot \theta, \tag{5.57}$$

where $u_n(\theta')$ is the i-th row of the Hessian of $\widetilde{\pi}_n$ evaluated at some $\theta' \in \mathcal{B}_n$. This has a finite limit, for $n \to \infty$, as one can easily verify by direct computation on (5.55) (it is in fact, up to a minus sing, the i-th row of the covariance matrix of p). Therefore

$$|u_n(\theta') \cdot \theta| \le C|\theta| \le Cn^{-(1-\varepsilon)/2}. \tag{5.58}$$

On the other hand, by (5.54),

$$|\partial_i \widetilde{\pi}_n(0)| = \left| \partial_i \widetilde{p}(0) - \imath \frac{\delta_i^{(n)}}{n} \right| = \left| v_i - \frac{\delta_i^{(n)}}{n} \right| \le Cn^{-1}. \tag{5.59}$$

Thus, using (5.58)-(5.59) in (5.57),

$$|\partial_i \widetilde{\pi}_n(\theta)| \le Cn^{-(1-\varepsilon)/2}, \tag{5.60}$$

which is the same bound as (5.47). This proves that the H^ν-norm of $\widetilde{g^{(n)}}$ can be estimated as in Section 5.4 even when $\widetilde{p}$ is replaced by $\widetilde{\pi}_n$. That is, (5.2) holds even when (E3) does not, which completes Stage 2.

5.6 Stage 3: Removing (E1) and (E2)

It is easy to realize that the convergence rate in (5.2) is not only independent of the choice of $Q \subseteq S_0$, it is also independent of the fact that Q is contained in S_0, as long as it remains an element of the countable partition associated to $\mathscr{B}_{-m,0}$. In fact, if we take $Q \subseteq S_\alpha$, with $\alpha \neq 0$, we can shift, via the natural action of $\mathbb{Z}^d$ onto $\mathcal{M}$, both Q and F by the quantity $-\alpha$. (5.2) continues to hold with the same convergence rate because all the properties of F that were used in Stages 1 and 2 are translation invariant.

In formula, there exists a positive vanishing sequence $\{\vartheta_n^{(m)}\}_{n\in\mathbb{N}}$ such that, if $g = 1_Q$ and Q is a fundamental set of $\mathscr{B}_{-m,0}$,

$$|\mu((F \circ T^n)g) - \overline{\mu}(F)\mu(g)| \le \|g\|_{L^1}\, \vartheta_n^{(m)}. \tag{5.61}$$

Since (5.61) depends continuously on $g \in L^1$, it is immediate to extend it to $g = \sum_{j\in\mathbb{N}} a_j 1_{Q_j}$, with $a_j > 0$, that is, to a generic positive function in $L^1(\mathcal{M}, \mathscr{B}_{-m,0}, \mu)$. If g is such that both the positive part g^+ and the negative part g^- are nonzero, we apply (5.61) twice to g^+ and g^-. An easy estimate proves that the formula holds in this case as well.

Therefore **(M5)** holds w.r.t. $\widehat{\mathcal{G}} := \{F \in L^\infty(\mathcal{M}, \mathscr{A}_s, \mu) \mid \exists\, \overline{\mu}(F)\}$ and $\widehat{\mathcal{L}}_m := L^1(\mathcal{M}, \mathscr{B}_{-m,0}, \mu)$, for all $m \in \mathbb{N}$.

Now, if $F \in \mathcal{G}_m$ and $g \in \mathcal{L}_m$, the invariance of μ and Lemma 2.4 give that

$$\begin{aligned} & \mu((F \circ T^n)g) - \overline{\mu}(F)\mu(g) = \\ = \;& \mu((F \circ T^{n-m})(g \circ T^{-m})) - \overline{\mu}(F \circ T^m)\mu(g \circ T^{-m}). \end{aligned} \tag{5.62}$$

Since $g \circ T^{-m} \in \widehat{\mathcal{L}}_{2m}$ and $F \circ T^m \in \widehat{\mathcal{G}}$ (because $\mathscr{B}_{0,2m} \subset \mathscr{A}_s$), we apply the previous result and see that *(a)* holds with a convergence rate $\vartheta_{n-2m}^{(2m)}$.

5.7 Stage 4: Proof of the remaining assertions

Statement *(a)* immediately implies **(M4)** relative to $\mathcal{G}_m$ and $\mathcal{L}_m$ (Proposition 3.1). One readily extends it to $\mathcal{G}$ and $\mathcal{L}$, thus proving *(b)*, by means of the following obvious lemma:

Lemma 5.8 *If $\mathcal{G}'$ is a dense subset of $\mathcal{G}$ in the L^∞-norm and $\mathcal{L}'$ is a dense subset of $\mathcal{L}$ in the L^1-norm, then* **(M4)** *for $\mathcal{G}'$ and $\mathcal{L}'$ implies* **(M4)** *for $\mathcal{G}$ and $\mathcal{L}$.*

As concerns *(d)*, it is easy to verify that Proposition 3.2 applies, w.r.t. the family of local observables $g_\alpha := G\, 1_{S_\alpha}$, so *(a)* implies **(M1)** relative to $\mathcal{G}_m$. We extend it to $\mathcal{G}$ by means of another obvious result.

Lemma 5.9 *If $\mathcal{G}'$ is a dense subset of $\mathcal{G}$ in the L^∞-norm, then* **(M1)** *for $\mathcal{G}'$ implies* **(M1)** *for $\mathcal{G}$.*

Finally, let us consider *(c)*. By the second part of Proposition 3.1, it suffices to show that, if $F, G \in \mathcal{G}_m$ and F is $\mathbb{Z}^d$-periodic, then $\overline{\mu}((F \circ T^n)G)$ exists for n large enough. By the same arguments as in the proof of Lemma 2.4, when $V \nearrow \mathcal{M}$,

$$\mu_V((F \circ T^n)G) = \mu_V((F \circ T^{n-m})(G \circ T^{-m})) + o(\mu(V)). \tag{5.63}$$

So we can reduce to proving the existence of the infinite-volume limit of the above r.h.s., for all $n \geq 2m$. Since $G \circ T^{-m}$ is measurable w.r.t. $\mathscr{B}_{-2m,0} \subset \mathscr{A}_u$, with a slight abuse of notation we can define, for $\alpha \in Z^d$,

$$b_\alpha := \int_{S_\alpha} G \circ T^{-m}\, d\mu = \int_0^1 G \circ T^{-m}(\alpha; y_2)\, dy_2. \tag{5.64}$$

An analogous definition can be made for $F \circ T^{n-m}$, which is measurable w.r.t. $\mathscr{B}_{0,2m} \subset \mathscr{A}_s$. In this case, notice that $F \circ T^{n-m}$ is also $\mathbb{Z}^d$-periodic, so we can write

$$a := \int_{S_\alpha} F \circ T^{n-m}\, d\mu = \int_0^1 F \circ T^{n-m}(y_1)\, dy_1. \tag{5.65}$$

Clearly, then,

$$\int_{S_\alpha} (F \circ T^{n-m})\, (G \circ T^{-m})\, d\mu = a\, b_\alpha \tag{5.66}$$

and, for $V = \bigcup_{\alpha \in B_{\gamma,r}} S_\alpha$,

$$\mu_V((F \circ T^{n-m})(G \circ T^{-m})) = \frac{a}{(2r+1)^d} \sum_{\alpha \in B_{\gamma,r}} b_\alpha. \tag{5.67}$$

Since $\overline{\mu}(G)$ exists, by Lemma 2.4

$$\overline{\mu}(G \circ T^{-m}) = \lim_{r \to \infty} \frac{1}{(2r+1)^d} \sum_{\alpha \in B_{\gamma,r}} b_\alpha \tag{5.68}$$

exists and the limit is uniform in γ. Also, it is obvious that $\overline{\mu}(F \circ T^{n-m}) = a$. Hence, as $V \nearrow \mathcal{M}$, the r.h.s. of (5.67) tends to $\overline{\mu}(F \circ T^{n-m}) \overline{\mu}(G \circ T^{-m})$, which is what we wanted to prove.

This concludes the proof of Theorem 4.6. Q.E.D.

A Appendix

We collect here a few technical results which would have been distracting in the body of the paper. The most important of them is an estimate on a certain Fourier norm that is pivotal in the proof of Theorem 4.6. This is presented in Section A.3.

A.1 Proof of Lemma 4.2

Since T is an automorphism, it is no loss of generality to prove the assertion for $t = -1$. Also, since **(A1)** is invariant for the action of $\mathbb{Z}^d$ on V, we may assume that all $V \in \mathscr{V}$ are of the form $V_r := B_{0,r} \times [0,1)^2$. Thus, the infinite-volume limit becomes the limit $r \to \infty$.

Let $r' = r'(r) := [r^{1/2}]$ ($[\cdot]$ is the integer part of a positive number) and

$$\varphi(r') := \mu(TS_0 \setminus V_{r'}) = \sum_{\beta \notin B_{0,r'}} p_\beta. \tag{A.1}$$

Clearly, $\varphi(r') \to 0$, as r' and r tend to infinity. This and the translation invariance of T imply that

$$\mu(TV_r \setminus V_{r+r'}) \leq \mu(V_r)\, \varphi(r'), \tag{A.2}$$

whence

$$\mu(TV_r \cup V_r) \leq \mu(V_{r+r'}) + \mu(TV_r \setminus V_{r+r'}) = (2r+1)^d + o((2r+1)^d). \tag{A.3}$$

With a dual argument, considering that $\mu(V_{r-r'} \setminus TV_r) = \mu(T^{-1}V_{r-r'} \setminus V_r)$ and that T^{-1} acts essentially as T (after a swapping of the coordinates y_1 and y_2, the map T^{-1} becomes of the same type as T), we obtain

$$\mu(TV_r \cap V_r) \geq \mu(V_{r-r'}) - \mu(V_{r-r'} \setminus TV_r) = (2r+1)^d + o((2r+1)^d). \tag{A.4}$$

Taking the difference of (A.3) and (A.4) yields **(A1)** with $t = -1$. Q.E.D.

A.2 Proof of Lemma 4.5

We must show that, if $j', j'' \in \mathbb{Z}_N$ with $j' \neq j''$, then

$$\mathrm{span}_{\mathbb{Z}}\{\beta^{(j)} - \beta^{(j')}\}_{j \neq j'} = \mathrm{span}_{\mathbb{Z}}\{\beta^{(j)} - \beta^{(j'')}\}_{j \neq j''}. \tag{A.5}$$

The generic element of the l.h.s. of (A.5) is

$$\gamma = \sum_{j \neq j'} n_j (\beta^{(j)} - \beta^{(j')}) = \sum_{j \neq j'} n_j \beta^{(j)} - \left(\sum_{j \neq j'} n_j \right) \beta^{(j')}, \tag{A.6}$$

where $\{n_j\}_{j \neq j'}$ are free variables, i.e., are arbitrarily chosen integers. Upon defining $n_{j'} := -\sum_{j \neq j'} n_j$, which implies $n_{j''} = -\sum_{j \neq j''} n_j$, (A.6) becomes

$$\gamma = \sum_{j \in \mathbb{Z}_N} n_j \beta^{(j)} = \sum_{j \neq j''} n_j \beta^{(j)} - \left(\sum_{j \neq j''} n_j \right) \beta^{(j'')} = \sum_{j \neq j''} n_j (\beta^{(j)} - \beta^{(j'')}), \tag{A.7}$$

which is the generic element of the r.h.s. of (A.5), if we consider $\{n_j\}_{j \neq j''}$ to be the free variables and $n_{j''}$ to depend on them. Q.E.D.

A.3 Absolutely convergent Fourier series

In this section we present a convenient estimate for the space $\mathcal{A}$ of functions $\widetilde{a} : \mathbb{T}^d \longrightarrow \mathbb{C}$ with an absolutely convergent Fourier series $\{a_\beta\} = a$ [K, §6]. This functional space is defined as the maximal domain of the norm

$$\|\widetilde{a}\|_{\mathcal{A}} := \|a\|_{\ell^1} := \sum_{\beta \in \mathbb{Z}^d} |a_\beta|. \tag{A.8}$$

This norm has a couple of straightforward invariances. For $\gamma \in \mathbb{Z}^d$, $\zeta \in \mathbb{T}^d$, and $\widetilde{a} \in \mathcal{A}$, let

$$\begin{aligned} \widetilde{\omega}_\gamma(\theta) &:= e^{\imath \gamma \cdot \theta}; && \text{(A.9)} \\ (\tau_\zeta \widetilde{a})(\theta) &:= \widetilde{a}(\theta + \zeta). && \text{(A.10)} \end{aligned}$$

Lemma A.1 *Given $\widetilde{a} \in A$, for all $\gamma \in \mathbb{Z}^d$ and $\zeta \in \mathbb{T}^d$,*

$$\|\widetilde{a}\widetilde{\omega}_\gamma\|_{\mathcal{A}} = \|\tau_\zeta \widetilde{a}\|_{\mathcal{A}} = \|\widetilde{a}\|_{\mathcal{A}}.$$

PROOF. Trivial verification upon computation of the Fourier series of $\widetilde{a}\widetilde{\omega}_\gamma$ and $\tau_\zeta \widetilde{a}$. Q.E.D.

The following estimate is a modification—mostly, a simplification—of a 1984 result by Nowak [No]. The proof, which we give for completeness, is practically copied from that article.

Lemma A.2 *Let $\bar{\nu} = [d/2] + 1$ be the smallest integer strictly bigger than $d/2$. There exists a constant $C_d > 0$ such that*

$$\|\widetilde{a}\|_{\mathcal{A}} \le C_d \|\widetilde{a}\|_{H^{\bar{\nu}}},$$

where

$$\begin{aligned}\|\widetilde{a}\|_{H^{\bar\nu}} &:= |a_0| + \sum_{i=1}^{d} \left(\sum_{\beta\in\mathbb{Z}^d} \left|\beta_i^{\bar\nu} a_\beta\right|^2 \right)^{1/2} = \\ &= \left| \int_{\mathbb{T}^d} \widetilde{a}(\theta) \underline{d}\theta \right| + \sum_{i=1}^{d} \left(\int_{\mathbb{T}^d} \left| \frac{\partial^{\bar\nu}\widetilde{a}}{\partial\theta_i^{\bar\nu}}(\theta) \right|^2 \underline{d}\theta \right)^{1/2}.\end{aligned}$$

Proof. Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_d)$ be a permutation of $(1, 2, \ldots, d)$. Let us define

$$Z_\sigma := \left\{ (\beta_1, \beta_2, \ldots, \beta_d) \in \mathbb{Z}^d \;\middle|\; 0 < |\beta_{\sigma_1}| \le |\beta_{\sigma_2}| \cdots \le |\beta_{\sigma_d}| \right\}. \tag{A.11}$$

Clearly,

$$\mathbb{Z}^d = \{0\} \cup \bigcup_\sigma Z_\sigma, \tag{A.12}$$

although the union is not disjoint. Using the Cauchy-Schwartz inequality,

$$\left(\sum_{\beta\in Z_\sigma} |a_\beta| \right)^2 \le C_{\bar\nu} \sum_{\beta\in Z_\sigma} \beta_{\sigma_d}^{2\bar\nu} |a_\beta|^2 \le C_{\bar\nu} \sum_{\beta\in\mathbb{Z}^d} \left|\beta_{\sigma_d}^{\bar\nu} a_\beta\right|^2, \tag{A.13}$$

where we have denoted

$$\begin{aligned}C_{\bar\nu} := \sum_{\beta\in Z_\sigma} \beta_{\sigma_d}^{-2\bar\nu} &= \sum_{|\alpha_1|>0} \sum_{|\alpha_2|\ge|\alpha_1|} \cdots \sum_{|\alpha_{d-1}|\ge|\alpha_{d-2}|} \sum_{|\alpha_d|\ge|\alpha_{d-1}|} \alpha_d^{-2\bar\nu} \le \\ &\le C \sum_{|\alpha_1|>0} \sum_{|\alpha_2|\ge|\alpha_1|} \cdots \sum_{|\alpha_{d-1}|\ge|\alpha_{d-2}|} \alpha_{d-1}^{-2\bar\nu+1} \le \\ &\le \cdots\cdots \le C \sum_{|\alpha_1|>0} \alpha_1^{-2\bar\nu+d-1} < \infty\end{aligned} \tag{A.14}$$

(as in Section 5, C represents a generic constant). In view of (A.12), summing the square root of (A.13) over all the permutations σ, we obtain

$$\sum_{\beta\neq 0} |a_\beta| \;\le\; (d-1)!\sqrt{C_{\bar\nu}} \sum_{i=1}^{d} \left(\sum_{\beta\in\mathbb{Z}^d} \left|\beta_i^{\bar\nu} a_\beta\right|^2 \right)^{1/2}. \tag{A.15}$$

whence the assertion of the lemma. Q.E.D.

References

[A] J. Aaronson, *An introduction to infinite ergodic theory*, Mathematical Surveys and Monographs, 50. American Mathematical Society, Providence, RI, 1997.

[Fr] N. A. Friedman, *Mixing transformations in an infinite measure space*, in: Studies in probability and ergodic theory, pp. 167–184, Adv. in Math. Suppl. Stud., 2. Academic Press, New York-London, 1978.

[HK] A. B. Hajian and S. Kakutani, *Weakly wandering sets and invariant measures*, Trans. Amer. Math. Soc. **110** (1964), 136–151.

[H] E. Hopf, *Ergodentheorie*, Springer-Verlag, Berlin, 1937.

[I1] S. Isola, *Renewal sequences and intermittency*, J. Statist. Phys. **97** (1999), no. 1-2, 263–280.

[I2] S. Isola, *On systems with finite ergodic degree*, Far East J. Dyn. Syst. **5** (2003), no. 1, 1–62.

[KP] S. Kakutani and W. Parry, *Infinite measure preserving transformations with "mixing"*, Bull. Amer. Math. Soc. **69** (1963), 752–756.

[K] Y. Katznelson, *An introduction to harmonic analysis*, 3rd ed. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2004.

[KO] J. F. C. Kingman and S. Orey, *Ratio limit theorems for Markov chains*, Proc. Amer. Math. Soc. **15** (1964), 907–910.

[KS] U. Krengel and L. Sucheston, *Mixing in infinite measure spaces*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **13** (1969), 150–164.

[Kr] K. Krickeberg, *Strong mixing properties of Markov chains with infinite invariant measure*, in: 1967 Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, CA, 1965/66), Vol. II, Part 2, pp. 431–446. Univ. California Press, Berkeley, CA, 1967.

[L1] M. Lenci, *Aperiodic Lorentz gas: recurrence and ergodicity*, Ergodic Theory Dynam. Systems **23** (2003), no. 3, 869–883.

[L2] M. Lenci, *Typicality of recurrence for Lorentz gases*, Ergodic Theory Dynam. Systems **26** (2006), no. 3, 799-820.

[No] Z. Nowak, *Criteria for absolute convergence of multiple Fourier series*, Ark. Mat. **22** (1984), no. 1, 25–32.

[Or] S. Orey, *Strong ratio limit property*, Bull. Amer. Math. Soc. **67** (1961), 571–574.

[Pa] F. Papangelou, *Strong ratio limits, R-recurrence and mixing properties of discrete parameter Markov processes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **8** (1967), 259–297.

[Pr] W. E. Pruitt, *Strong ratio limit property for R-recurrent Markov chains*, Proc. Amer. Math. Soc. **16** (1965), 196–200.

[R] W. Rudin, *Fourier analysis on groups*, Interscience Tracts in Pure and Applied Mathematics, no. 12, Interscience Publishers (a division of John Wiley and Sons), New York-London, 1962.

[Sa] U. Sachdeva, *On category of mixing in infinite measure spaces*, Math. Systems Theory **5** (1971), 319–330.

[S] F. Spitzer, *Principles of random walk*, 2nd ed. Graduate Texts in Mathematics, 34. Springer-Verlag, New York-Heidelberg, 1976.

[Su] L. Sucheston, *On mixing and the zero-one law*, J. Math. Anal. Appl. **6** (1963), 447–456.

[Th] M. Thaler, *The asymptotics of the Perron-Frobenius operator of a class of interval maps preserving infinite measures*, Studia Math. **143** (2000), no. 2, 103–119.

[To1] S. Tomatsu, *Uniformity of mixing transformations with infinite measure*, Bull. Fac. Gen. Ed. Gifu Univ. No. **17** (1981), 43–49 (1982).

[To2] S. Tomatsu, *Local uniformity of mixing transformations with infinite measure*, Bull. Fac. Gen. Ed. Gifu Univ. No. **20** (1984), 1–5 (1985).

[W] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982.