Chapter 2

Some basic ergodic theory

Ergodic theory deals with groups or semigroups of measure preserving transformations T^g on a probability space $(X, \mathcal{B}, \mu)$. In later sections X will be a compact metric space endowed with its Borel σ-algebra $\mathcal{B}$, but most of the theory presented in this chapter is independent of this topological background.

We start with a section on Birkhoff's ergodic theorem, which generalizes the classical formulation from Theorem 1.4.2 to arbitrary measure preserving $\mathbb{Z}_+^d$-actions. In order to characterize those systems for which the limit in the ergodic theorem is constant, we introduce the concept of ergodicity in Section 2.2 and discuss the related notions of weak mixing and mixing. In Section 2.3 we show how an arbitrary measure preserving system can be decomposed into ergodic components. We finish this chapter with two brief sections on return times and return maps and on factors and extensions. Both may be skipped on a first reading, because we make no explicit use of them except in Chapter 6.

There are several textbooks on general ergodic theory which the reader might wish to consult for further information, for example [60], [13], [46], [37].

2.1 Birkhoff's ergodic theorem

The basic objects of ergodic theory are measure preserving dynamical systems, a notion that embraces as special cases lattice systems discussed at the end of Section 1.2 and the interval maps from Section 1.4. In this section we state and prove Birkhoff's ergodic theorem for $\mathbb{Z}_+^d$-actions that applies to such systems.

2.1.1 Definition *A measure preserving dynamical system (m.p.d.s.) is a quadruple $(X, \mathcal{B}, \mu, \mathcal{T})$ consisting of*

a) *a probability space $(X, \mathcal{B}, \mu)$ and*

b) *a $\mathcal{B}$-measurable action $\mathcal{T} = (T^g : g \in G)$ of the semigroup $G = \mathbb{Z}_+^d$ or of the group $G = \mathbb{Z}^d$ on X which leaves the measure μ invariant (symbolically: $\mathcal{T}\mu = \mu$). This means:*

1) $T^g : X \to X$ is $\mathcal{B}$-measurable and $T^g\mu = \mu$ for all $g \in G$,
2) $T^0 = Id_X$ and $T^{g+g'} = T^g \circ T^{g'}$ for all $g, g' \in G$.

2.1.2 Remarks
a) Observe that $\mathcal{T}\mu = \mu$ implies

$$\int_{T^{-g}A} f \circ T^g \, d\mu = \int_A f \, d\mu, \text{ in particular } \|f \circ T^g\|_1 = \|f\|_1$$

for all $A \in \mathcal{B}$, $f \in L^1_\mu$ and $g \in G$.

b) If $G = \mathbb{Z}^d$, then all $T^g \in \mathcal{T}$ are invertible, because $-g \in G$ for all $g \in G$ and $T^g \circ T^{-g} = T^0 = Id_X$.

c) Let $e_1, \ldots, e_d$ be the canonical basis for the lattice $\mathbb{Z}^d_+$ and consider $g = \sum_{i=1}^d g_i e_i \in G$. Then $T^g = (T^{e_1})^{g_1} \circ \cdots \circ (T^{e_d})^{g_d}$ where $(T^{e_i})^{g_i}$ denotes the g_i-fold iterate of T^{e_i}. Hence, if $d = 1$, then T^g is the g-fold iterate of the map $T = T^1$. In this case we also write $(X, \mathcal{B}, \mu, T)$ instead of $(X, \mathcal{B}, \mu, \mathcal{T})$. If $d > 1$, the transformations T^{e_i} do not necessarily commute.

d) Recent contributions to the ergodic theory of $\mathbb{Z}^d$-actions are collected in [48]. Actions of more general groups are studied in [42], [54]. ◇

2.1.3 Definition *A measurable function $f : X \to \mathbb{R}$ is $\mathcal{T}$-invariant* mod μ *if $f \circ T^g = f$ μ-a.s. for each $g \in G$. Accordingly a set $A \in \mathcal{B}$ is $\mathcal{T}$-invariant* mod μ *if $\mu(T^{-g}A \triangle A) = 0$ for all $g \in G$. We write $\mathcal{I}_\mu(\mathcal{T}) := \{A \in \mathcal{B} : \mu(T^{-g}A \triangle A) = 0 \; \forall g \in G\}$. If $\mathcal{T}$ is generated by a single transformation T, we write also $\mathcal{I}_\mu(T)$ instead of $\mathcal{I}_\mu(\mathcal{T})$.*

2.1.4 Remark Let $\mathcal{I}(\mathcal{T}) := \{B \in \mathcal{B} : T^{-g}B = B \; \forall g \in G\}$. Then $\mathcal{I}(\mathcal{T}) = \mathcal{I}_\mu(\mathcal{T})$ mod μ, because for each $B \in \mathcal{I}_\mu(\mathcal{T})$ the set $B' := \bigcup_{g \in G} \bigcap_{h \in G} T^{-(g+h)}B$ belongs to $\mathcal{I}(\mathcal{T})$ and $\mu(B \triangle B') = 0$.

It is easy to show that $\mathcal{I}(\mathcal{T})$ and $\mathcal{I}_\mu(\mathcal{T})$ are σ-algebras and that a $\mathcal{B}$-measurable function f is $\mathcal{T}$-invariant ($\mathcal{T}$-invariant mod μ) if and only if it is $\mathcal{I}(\mathcal{T})$-measurable ($\mathcal{I}_\mu(\mathcal{T})$-measurable). ◇

For $n \in \mathbb{Z}_+$ let

$$\Lambda_n := \{g = \sum_i g_i e_i \in G : |g_i| < n \; \forall i = 1, \ldots, d\} \quad \text{and} \quad \lambda_n := |\Lambda_n| .$$

One of the corner-stones of ergodic theory is

urable and $T^g\mu = \mu$ for all $g \in G$,
$T^g \circ T^{g'}$ for all $g, g' \in G$.

ies

$\int_A f\,d\mu$, in particular $\|f \circ T^g\|_1 = \|f\|_1$

G.

$\mathcal{T}$ are invertible, because $-g \in G$ for all $g \in G$ and

onical basis for the lattice $\mathbb{Z}^d_+$ and consider $g =$ $(T^{e_1})^{g_1} \circ \cdots \circ (T^{e_d})^{g_d}$ where $(T^{e_i})^{g_i}$ denotes the if $d = 1$, then T^g is the g-fold iterate of the map write $(X, \mathcal{B}, \mu, T)$ instead of $(X, \mathcal{B}, \mu, \mathcal{T})$. If $d > 1$, t necessarily commute.

e ergodic theory of $\mathbb{Z}^d$-actions are collected in [48]. ps are studied in [42], [54]. ◇

le function $f : X \to \mathbb{R}$ is $\mathcal{T}$-invariant mod μ if $f \circ$ G. Accordingly a set $A \in \mathcal{B}$ is $\mathcal{T}$-invariant mod μ if G. We write $\mathcal{I}_\mu(\mathcal{T}) := \{A \in \mathcal{B} : \mu(T^{-g}A \triangle A) =$ d by a single transformation T, we write also $\mathcal{I}_\mu(T)$

$\{B \in \mathcal{B} : T^{-g}B = B\ \forall g \in G\}$. Then $\mathcal{I}(\mathcal{T}) = \mathcal{I}_\mu(\mathcal{T})$ $\mathcal{I}_\mu(\mathcal{T})$ the set $B' := \bigcup_{g \in G} \bigcap_{h \in G} T^{-(g+h)}B$ belongs

$\mathcal{T})$ and $\mathcal{I}_\mu(\mathcal{T})$ are σ-algebras and that a $\mathcal{B}$-measurable -invariant mod μ) if and only if it is $\mathcal{I}(\mathcal{T})$-measurable ◇

$G : |g_i| < n\ \forall i = 1, \dots, d\}$ and $\lambda_n := |\Lambda_n|$.

ergodic theory is

2.1.5 Theorem (Birkhoff's ergodic theorem)
Let $(X, \mathcal{B}, \mu, \mathcal{T})$ be a m.p.d.s. For each $f \in L^1_\mu$ the limit

$$\bar{f}(x) := \lim_{n \to \infty} \frac{1}{\lambda_n} \sum_{g \in \Lambda_n} f(T^g x) \tag{2.1}$$

exists μ-a.s. and in L^1_μ. The function $\bar{f}$ is $\mathcal{T}$-invariant mod μ *and for each set $A \in \mathcal{I}_\mu(\mathcal{T})$*

$$\int_A \bar{f}\,d\mu = \int_A f\,d\mu\,. \tag{2.2}$$

2.1.6 Remarks
a) The limit $\bar{f}$ is $\mathcal{I}_\mu(\mathcal{T})$-measurable, and in view of identity (2.2) it is a version of $E_\mu[f \mid \mathcal{I}_\mu(\mathcal{T})]$, the conditional expectation of f under μ given $\mathcal{I}_\mu(\mathcal{T})$, see A.4.32.

b) It suffices to prove the theorem for $G = \mathbb{Z}^d_+$, because the case $G = \mathbb{Z}^d$ can be reduced to $G = \mathbb{Z}^d_+$ in the following way:

Let $\mathcal{T} = (T^g : g \in \mathbb{Z}^d)$ be a $\mathbb{Z}^d$-action. For each $\sigma \in \{+1, -1\}^d$ denote by $\mathcal{T}_\sigma = (T^g_\sigma : g \in \mathbb{Z}^d_+)$ the $\mathbb{Z}^d_+$-action $T^g_\sigma := T^{(\sigma_1 g_1, \dots, \sigma_d g_d)}$. Since a measurable set is T^g-invariant mod μ if and only if it is T^{-g}-invariant mod μ, the σ-algebras $\mathcal{I}_\mu(\mathcal{T}_\sigma)$ are all identical and coincide with $\mathcal{I}_\mu(\mathcal{T})$. Now apply the ergodic theorem for the action $\mathcal{T}_\sigma$ to a function $f \in L^1_\mu$ and denote the resulting limit by $\bar{f}_\sigma$. By the previous remark, all these $\bar{f}_\sigma$ are versions of the conditional expectation $E_\mu[f \mid \mathcal{I}_\mu(\mathcal{T})]$.

Denote now by Λ_n the n-box in $\mathbb{Z}^d$ and by Λ^+_n the one in $\mathbb{Z}^d_+$. If it were true that $\sum_{g \in \Lambda_n} f \circ T^g = \sum_\sigma \sum_{g \in \Lambda^+_n} f \circ T^g_\sigma$, it would follow immediately that $\bar{f} = 2^{-d} \sum_\sigma \bar{f}_\sigma = E_\mu[f \mid \mathcal{I}_\mu(\mathcal{T})]$ μ-as. However, this decomposition of the summation over Λ_n counts elements g with exactly one $g_i = 0$ twice and, more generally, those with exactly k indices i for which $g_i = 0$ are counted 2^k times, see Figure 2.1. Therefore the sum $\sum_\sigma \sum_{g \in \Lambda^+_n} f \circ T^g_\sigma$ must be modified by averages over sets of elements $g \in \Lambda^+_n$ for which certain g_i vanish. But this means that these averages obey the ergodic theorem for a $\mathbb{Z}^{d'}_+$-action with $d' < d$. In particular, the corresponding norming constants are $n^{-d'}$ so that these contributions vanish asymptotically if they are normed by $\lambda_n^{-1} = n^{-d}$.

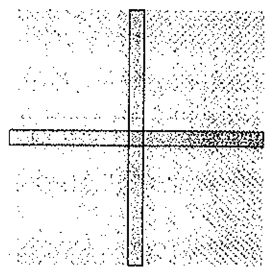

Figure 2.1: The decomposition of Λ_n

c) In the one-dimensional case the ergodic theorem takes the form

$$\bar{f}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \quad \text{for } \mu\text{-a.e. } x.$$

By the previous remark one has for $G = \mathbb{Z}$

$$\bar{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k} x) \quad \mu\text{-a.s.}$$

d) Ergodic averages along more general families of "parameter intervals" than Λ_n are also of interest. Theorem 6.2.8 in [34] shows that these averages converge to essentially the same $\bar{f}$ for a broad class of d-dimensional index sets. ◇

Proof of the ergodic theorem: This proof comes in several steps. Some of them are independent of the dimension, others are much simpler for $d = 1$ than for $d > 1$. In order not to overload the proof with technicalities we therefore perform some of the steps only for $d = 1$ first and provide the necessary modifications for $d > 1$ later.

For $d = 1$ we are following closely the proof in [4, Theorem 2.2]. Steps 1, 2, and 4 of our exposition are written in such a way that we can refer to them later in our proof for the case $d > 1$, which is close to the proof of [34, Theorem 6.2.8]. Also, for $d > 1$, we use the ergodic theorem for $\mathbb{Z}_+^{d-1}$-actions in Step 1. This introduces some inductive structure to the proof.

We use the following notation: for measurable $f : X \to \mathbb{R}$ let

$$S_n f := \sum_{g\in\Lambda_n} f \circ T^g , \quad A_n f := \frac{1}{\lambda_n} S_n f ,$$
$$A^- f := \liminf_{n\to\infty} A_n f , \quad A^+ f := \limsup_{n\to\infty} A_n f .$$

Step 1: As $f = f^+ - f^-$, it suffices to prove the theorem for $0 \le f \in L^1_\mu$. For $i = 1, \ldots, d$ we have

$$\begin{aligned} A_n f \circ T^{e_i} &= A_n f - \frac{1}{n^d} \sum_{g\in\Lambda_n, g_i=0} f \circ T^g + \frac{1}{n^d} \sum_{g\in(\Lambda_n+e_i)\setminus\Lambda_n} f \circ T^g \\ &\ge A_n f - \frac{1}{n} \left(\frac{1}{n^{d-1}} \sum_{g\in\Lambda_n, g_i=0} f \circ T^g \right) . \end{aligned}$$

If $d = 1$ (and hence $i = 1$), the term in brackets is identically equal to f. If $d > 1$ we apply the ergodic theorem for $\mathbb{Z}_+^{d-1}$-actions to this term and conclude that it converges almost surely to some finite value. Therefore, in any case

$$A^+ f \circ T^{e_i} = \limsup_{n\to\infty} A_n f \circ T^{e_i} \ge \limsup_{n\to\infty} A_n f = A^+ f$$

μ-a.s. As at the same time $S_n f \circ T^{e_i} \le S_{n+1} f$ and hence

$$A^+ f \circ T^{e_i} = \limsup_{n\to\infty} \frac{1}{\lambda_n} S_n f \circ T^{e_i} \le \limsup_{n\to\infty} \frac{(n+1)^d}{n^d} \frac{1}{\lambda_{n+1}} S_{n+1} f = A^+ f ,$$

as for $G = \mathbb{Z}$

$$f(T^k x) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^{-k}x) \quad \mu\text{-a.s.}$$

ore general families of "parameter intervals" than Λ_n 6.2.8 in [34] shows that these averages converge to road class of d-dimensional index sets. ◇

This proof comes in several steps. Some of them nsion, others are much simpler for $d = 1$ than for ad the proof with technicalities we therefore perform $= 1$ first and provide the necessary modifications for

ng closely the proof in [4, Theorem 2.2]. Steps 1, 2, ritten in such a way that we can refer to them later in , which is close to the proof of [34, Theorem 6.2.8]. e ergodic theorem for $\mathbb{Z}_+^{d-1}$-actions in Step 1. This ructure to the proof.

ation: for measurable $f : X \to \mathbb{R}$ let

$$\sum_{g\in\Lambda_n} f\circ T^g, \quad A_n f := \frac{1}{\lambda_n} S_n f,$$

$$\inf A_n f, \quad A^+ f := \limsup_{n\to\infty} A_n f.$$

it suffices to prove the theorem for $0 \le f \in L^1_\mu$. For

$$-\frac{1}{n^d}\sum_{g\in\Lambda_n, g_i=0} f\circ T^g + \frac{1}{n^d}\sum_{g\in(\Lambda_n+e_i)\setminus\Lambda_n} f\circ T^g$$

$$-\frac{1}{n}\left(\frac{1}{n^{d-1}}\sum_{g\in\Lambda_n, g_i=0} f\circ T^g\right).$$

the term in brackets is identically equal to f. If $d > 1$ em for $\mathbb{Z}_+^{d-1}$-actions to this term and conclude that some finite value. Therefore, in any case

$$\limsup_{n\to\infty} A_n f\circ T^{e_i} \ge \limsup_{n\to\infty} A_n f = A^+ f$$

$S_n f\circ T^{e_i} \le S_{n+1} f$ and hence

$$S_n f\circ T^{e_i} \le \limsup_{n\to\infty} \frac{(n+1)^d}{n^d}\frac{1}{\lambda_{n+1}} S_{n+1} f = A^+ f,$$

it follows that

$$A^+ f \circ T^g = A^+ f \quad \mu\text{-a.s.} \tag{2.3}$$

for all $g \in G$.

Step 2: The idea of the proof is to show rather directly that $A^- f \ge A^+ f$ μ-a.s. We will find, however, that such a naive approach does not yield a pathwise estimate. Instead we are going to prove that

$$\int A^+ f \, d\mu \le \gamma_d \int f \, d\mu$$

where $\gamma_1 = 1$ and $\gamma_d = 2^d$ for $d > 1$. For the case $d = 1$ it is then an easy task to deduce from this in Step 3 that $\int A^+ f \, d\mu \le \int A^- f \, d\mu$ and hence $A^+ f = A^- f$ μ-a.s.

The basic idea is to decompose the sum $S_n f(x)$ into sums over small cubes Λ_k (uniformly bounded in size) in such a way that the averages of f over these subcubes are nearly equal to $A^+ f(x)$.

For two real-valued functions g, h on X let $g \wedge h := \min\{g, h\}$ be the pointwise minimum. Let $r > 1$, $0 < \epsilon < 1$, and define $H = H_{r,\epsilon} := (A^+ f \wedge r)(1 - \epsilon)$. Then $0 \le H < r$, and as $(A^+ f) \circ T^g = A^+ f$ by (2.3), also $H \circ T^g = H$ for all $g \in G$. In order to perform the above mentioned decomposition let

$$\tau : X \to \mathbb{N}, \quad x \mapsto \min\{n \ge 1 : S_n f(x) \ge \lambda_n \cdot H(x)\}.$$

Observe that $\tau(x) < \infty$ for all $x \in X$:

if $A^+ f(x) = 0$, then $S_1 f(x) = f(x) \ge 0 = H(x)$ so that $\tau(x) = 1$,
if $0 < A^+ f(x) < \infty$, then $A^+ f(x) > H(x)$ so that $\tau(x) < \infty$, and
if $A^+ f(x) = \infty$, then $A^+ f(x) > r(1 - \epsilon) = H(x)$ so that $\tau(x) < \infty$.

The definition of τ suggests that the subblocks of Λ_n we are looking for are of the form $\Lambda_{\tau(T^g x)} + g$ for suitable $g \in \Lambda_n$. If τ is uniformly bounded we can actually work with such blocks. Otherwise we replace f by a function $\tilde{f}$ close to f such that the associated $\tilde{\tau}$ is bounded. More precisely, for $M > 0$ consider $\tilde{f} := f + H \cdot 1_{\{\tau > M\}}$ and $\tilde{\tau}(x) := \min\{n \ge 1 : S_n \tilde{f}(x) \ge \lambda_n \cdot H(x)\}$. Observe that

if $\tau(x) > M$, then $S_1 \tilde{f}(x) = \tilde{f}(x) \ge H(x)$, i.e., $\tilde{\tau}(x) = 1 \le M$, and
if $\tau(x) \le M$, then $S_{\tau(x)} \tilde{f}(x) \ge S_{\tau(x)} f(x) \ge \lambda_{\tau(x)} H(x)$, i.e., $\tilde{\tau}(x) \le \tau(x) \le M$.

Hence $\tilde{\tau} \le M$. Suppose now that for each $x \in X$ there is some set $\Lambda'_n(x) \subseteq \Lambda_n$ with the property that $(\Lambda_{\tilde{\tau}(T^g x)} + g : g \in \Lambda'_n(x))$ is a family of pairwise disjoint "cubes" contained in Λ_n such that

$$\sum_{g \in \Lambda'_n(x)} |\Lambda_{\tilde{\tau}(T^g x)}| \ge \gamma_d^{-1} (n - M)^d. \tag{2.4}$$

This condition ensures that the union of these cubes covers approximately a fraction γ_d^{-1} of Λ_n at least, see Figure 2.2. For general d this family is provided by Lemma 2.1.7 below; in the case $d = 1$ the set $\Lambda'_n(x)$ is simply the set of all integers $\tau_k(x) \in \Lambda_{n-M}$ defined recursively by

$$\tau_0 = 0 \quad \text{and} \quad \tau_{k+1} = \tau_k + \tilde{\tau} \circ T^{\tau_k} \text{ for } k \geq 0 .$$

In view of the definition of $\tilde{\tau}$ and the $\mathcal{T}$-invariance of H we can estimate

$$\begin{aligned} S_n\tilde{f}(x) &\geq \sum_{h\in\Lambda'_n(x)} \sum_{g\in\Lambda_{\tilde{\tau}(T^h x)}+h} (\tilde{f}\circ T^g)(x) = \sum_{h\in\Lambda'_n(x)} \sum_{g\in\Lambda_{\tilde{\tau}(T^h x)}} (\tilde{f}\circ T^g)(T^h x) \\ &= \sum_{h\in\Lambda'_n(x)} (S_{\tilde{\tau}(T^h x)}\tilde{f})(T^h x) \geq \sum_{h\in\Lambda'_n(x)} \lambda_{\tilde{\tau}(T^h x)} \cdot H(T^h x) \\ &= \sum_{h\in\Lambda'_n(x)} |\Lambda_{\tilde{\tau}(T^h x)}| \cdot H(x) \geq \gamma_d^{-1}(n-M)^d \cdot H(x) . \end{aligned}$$

Hence, observing that $H \leq r$,

$$S_n f = S_n(\tilde{f} - H\,1_{\{\tau>M\}}) \geq \gamma_d^{-1}(n-M)^d \cdot H - r\, S_n 1_{\{\tau>M\}} . \tag{2.5}$$

Dividing this inequality by λ_n and passing to the limit $n \to \infty$ this yields at once that $A^- f \geq \gamma_d^{-1} H - rA^+ 1_{\{\tau>M\}} = \gamma_d^{-1}(1-\epsilon)(A^+ f \wedge r) - rA^+ 1_{\{\tau>M\}}$. The hope is now that $A^+ 1_{\{\tau>M\}}$ is small because $\mu\{\tau > M\}$ is small if M is large enough. But this term is of the same type as the term $A^+ f$ that we want to estimate. Furthermore, we have made no use of the $\mathcal{T}$-invariance of μ so far, and it is exactly at this point where we have to use it. Instead of estimating pathwise averages $A_n 1_{\{\tau>M\}}(x)$, we integrate inequality (2.5) and observe that $\mathcal{T}\mu = \mu$:

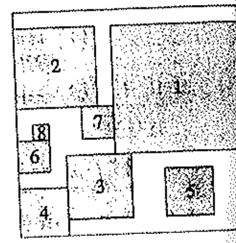

Figure 2.2: An exampl
for the cubes $\Lambda_{\tau(T^g x)} +$

$$\int f\, d\mu = \frac{1}{n^d} \int S_n f\, d\mu \geq \gamma_d^{-1} \left(\frac{n-M}{n}\right)^d \int H\, d\mu - r\,\mu\{\tau > M\} .$$

In the limit $n \to \infty$ this yields

$$\int f\, d\mu + r\,\mu\{\tau > M\} \geq \gamma_d^{-1} \int H\, d\mu .$$

As $\lim_{M\to\infty} \mu\{\tau > M\} = 0$ and as $H = (1-\epsilon)(A^+ f \wedge r)$, this yields

$$\gamma_d \int f\, d\mu \geq (1-\epsilon) \int (A^+ f \wedge r)\, d\mu .$$

he union of these cubes covers approximately a frac-
Figure 2.2. For general d this family is provided by
ase $d = 1$ the set $\Lambda'_n(x)$ is simply the set of all integers
ively by

d $\quad \tau_{k+1} = \tau_k + \tilde{\tau} \circ T^{\tau_k}$ for $k \geq 0$.

and the $\mathcal{T}$-invariance of H we can estimate

$$\sum (\tilde{f} \circ T^g)(x) = \sum_{h \in \Lambda'_n(x)} \sum_{g \in \Lambda_{\tilde{\tau}(T^h x)}} (\tilde{f} \circ T^g)(T^h x)$$

$$\ldots \tilde{f})(T^h x) \geq \sum_{h \in \Lambda'_n(x)} \lambda_{\tilde{\tau}(T^h x)} \cdot H(T^h x)$$

$$\ldots | \cdot H(x) \geq \gamma_d^{-1} (n - M)^d \cdot H(x) .$$

r,

$$\ldots) \geq \gamma_d^{-1} (n - M)^d \cdot H - r\, S_n 1_{\{\tau > M\}} . \qquad (2.5)$$

λ_n and passing to the limit $n \to \infty$ this yields at once
$1_{\{\tau > M\}} = \gamma_d^{-1}(1 - \epsilon)(A^+ f \wedge$
ope is now that $A^+ 1_{\{\tau > M\}}$ is
is small if M is large enough.
e type as the term $A^+ f$ that we
more, we have made no use of
ar, and it is exactly at this point
Instead of estimating pathwise
e integrate inequality (2.5) and

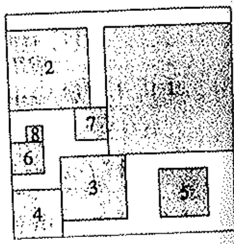

Figure 2.2: An exampl
for the cubes $\Lambda_{\tau(T^g x)} + $

$$f\, d\mu \geq \gamma_d^{-1} \left(\frac{n - M}{n} \right)^d \int H\, d\mu - r\, \mu\{\tau > M\} .$$

ields

$$d\mu + r\, \mu\{\tau > M\} \geq \gamma_d^{-1} \int H\, d\mu .$$

$= 0$ and as $H = (1 - \epsilon)(A^+ f \wedge r)$, this yields

$$\int f\, d\mu \geq (1 - \epsilon) \int (A^+ f \wedge r)\, d\mu .$$

Passing with $\epsilon \to 0$ and then with $r \to \infty$ it follows that

$$\int A^+ f\, d\mu \leq \gamma_d \int f\, d\mu . \qquad (2.6)$$

(The last limit uses the monotone convergence theorem, see A.4.19.)

Step 3 (only for $d = 1$): Here we derive the almost sure convergence in (2.1) from estimate (2.6). Our first observation is that for bounded f ($0 \leq f \leq M$) we have $A^- f = -A^+(-f) = M - A^+(M - f)$ so that

$$\int A^- f\, d\mu = M - \int A^+(M - f)\, d\mu \geq M - \int (M - f)\, d\mu = \int f\, d\mu$$

by (2.6) as $\gamma_1 = 1$. By a simple truncation argument this inequality extends to unbounded $f \geq 0$: Let $M > 0$. Then

$$\int A^- f\, d\mu \geq \int A^-(f \wedge M)\, d\mu \geq \int f \wedge M\, d\mu \nearrow \int f\, d\mu$$

as $M \to \infty$ by the monotone convergence theorem (see A.4.19), so that for any measurable $f \geq 0$

$$\int A^- f\, d\mu \geq \int f\, d\mu \geq \int A^+ f\, d\mu . \qquad (2.7)$$

As $A^- f \leq A^+ f$ it follows that $\bar{f} = A^- f = A^+ f$ μ-a.s.

Step 4: We prove the remaining assertions of the theorem. For the L^1_μ-convergence of $A_n f$ to $\bar{f}$ it suffices to show that the sequence $(A_n f)_n$ is uniformly integrable (see A.4.21): as $A_n f - r = A_n(f - r) \leq A_n((f - r)^+)$ where $A_n((f - r)^+) \geq 0$, we have also $(A_n f - r)^+ \leq A_n((f - r)^+)$ so that

$$\int (A_n f - r)^+\, d\mu \leq \int A_n((f - r)^+)\, d\mu = \int (f - r)^+ d\mu ,$$

and this tends to zero as $r \to \infty$.

The $\mathcal{T}$-invariance mod μ of $\bar{f}$ follows from (2.3), and (2.2) is now a direct consequence of the $\mathcal{T}$-invariance of μ: if $A \in \mathcal{I}_\mu(\mathcal{T})$, then

$$\int_A \bar{f}\, d\mu = \lim_{n \to \infty} \frac{1}{\lambda_n} \int_A S_n f\, d\mu = \lim_{n \to \infty} \frac{1}{\lambda_n} \sum_{g \in \Lambda_n} \int_{T^{-g} A} f \circ T^g\, d\mu = \int_A f\, d\mu .$$

This finishes the proof of the ergodic theorem for the case $d = 1$. Reviewing the four steps we see that the following modifications and additions are necessary for $d > 1$:

1. a different choice of the cubes in inequality (2.4), and
2. a new proof of $A^+f = A^-f$ μ-a.s. in Step 3.

We start with the first point. In the case $d > 1$, the cube Λ_n can no longer be paved without gaps using cubes $\Lambda_{\tau(T^h x)}+h$ as in one dimension. Still the following lemma provides a sufficient (though fragmentary) paving:

2.1.7 Lemma *Given positive integers $n > M$ and a map $\ell : \Lambda_n \to \{1,\dots,M\}$, there is a set $\Lambda'_n \subseteq \Lambda_n$ such that $(\Lambda_{\ell(g)}+g : g \in \Lambda'_n)$ is a family of pairwise disjoint "cubes" contained in Λ_n and*

$$\sum_{g\in\Lambda'_n} |\Lambda_{\ell(g)}| \geq 2^{-d}(n-M)^d\,. \tag{2.8}$$

Proof: Let $\mathcal{D}_M$ be a maximal collection of disjoint sets of the form $\Lambda_M + g$ with $\ell(g) = M$ and $g \in \Lambda_{n-M}$. When $\mathcal{D}_M,\dots,\mathcal{D}_i$ have been constructed for some $i > 1$, let $\mathcal{D}_{i-1}$ be a maximal collection of sets of the form $\Lambda_{i-1} + g$ with $\ell(g) = i-1$ and $g \in \Lambda_{n-M}$, for which all sets in $\mathcal{D}_{i-1}\cup\ldots\cup\mathcal{D}_M$ are disjoint. Let

$$\Lambda'_n := \{g \in \Lambda_{n-M} : \Lambda_{\ell(g)} + g \in \mathcal{D}_1 \cup \ldots \cup \mathcal{D}_M\}\,.$$

Observe that the sets $D_g := \Lambda_{\ell(g)} + g$, $g \in \Lambda'_n$, are pairwise disjoint subsets of Λ_n.

Take any $h \in \Lambda_{n-M}$. By the maximality of $\mathcal{D}_{\ell(h)}$ there exist some $j \geq \ell(h)$ and $\Lambda_j + g \in \mathcal{D}_j$ with $(\Lambda_{\ell(g)} + g) \cap (\Lambda_{\ell(h)} + h) \neq \emptyset$. Then $\ell(g) = j \geq \ell(h)$ and $\Lambda_{\ell(g)} \supseteq \Lambda_{\ell(h)}$. Hence $h \in \tilde{D}_g := g + \Lambda_{\ell(g)} - \Lambda_{\ell(g)}$, and therefore $\Lambda_{n-M} \subseteq \bigcup_{g\in\Lambda'_n} \tilde{D}_g$. Now (2.8) follows from the inequality $|\tilde{D}_g| \leq 2^d \cdot |D_g|$. $\square$

It remains to replace Step 3 in the proof of the ergodic theorem by a more elaborate argument. For the following lemma we need the concept of an orthogonal complement in the Hilbert space L^2_μ: For a linear subspace N of L^2_μ let $N^\perp := \{h \in L^2_\mu : \langle f, h\rangle = 0\ \forall f \in N\}$. Then $N^\perp$ is a closed linear subspace and $\mathrm{clos}(N) \oplus N^\perp = L^2_\mu$.

2.1.8 Lemma *Let $\mathcal{T} = \{T^g : g \in G\}$, $F := \{f \in L^2_\mu : f\circ T = f\ \forall T \in \mathcal{T}\}$ and $N := \{h - h\circ T : h \in L^2_\mu, T \in \mathcal{T}\}$. Then $F = N^\perp$.*

Proof: To show $F \subseteq N^\perp$, let $f \in F$, $h \in L^2_\mu$ and $T \in \mathcal{T}$. Then

$$\langle f, h - h\circ T\rangle = \langle f, h\rangle - \langle f\circ T, h\circ T\rangle = \langle f, h\rangle - \langle f, h\rangle = 0\,,$$

i.e., $f \in N^\perp$.

For the reverse inclusion let $f \in N^\perp$. Then $\langle f, f - f\circ T\rangle = 0$ so that $\langle f, f\rangle = \langle f\circ T, f\rangle$. It follows that

$$\|f - f\circ T\|^2 = \langle f, f - f\circ T\rangle - \langle f\circ T, f\rangle + \langle f\circ T, f\circ T\rangle = -\|f\|^2 + \|f\|^2 = 0$$

i.e., $f \in F$. □

Continuation of the proof of the ergodic theorem for $d > 1$:
For $f \in L^1_\mu$ let $\Delta f := A^+ f - A^- f$. We must show that $\Delta f = 0$ μ-a.s. To this end we decompose f into a sum of functions from either the space F or N of the previous lemma plus a small remainder term. Let $\epsilon > 0$.

1. $f = u_1 + r_0$ where $u_1 \in L^2_\mu$ and $\int |r_0|\, d\mu < \epsilon$. This is possible because L^2_μ is $\|.\|_1$-dense in L^1_μ.
2. $u_1 = f_1 + u_2 + \cdots + u_k + r_1$ for some $k \geq 2$ where $f_1 \in F$, $u_j = h_j - h_j \circ T^{g_j} \in N$, $g_j \in G$ $(j = 2, \dots, k)$ and $\int |r_1|\, d\mu \leq \|r_1\|_2 < \epsilon$, see Lemma 2.1.8.
3. $h_j = f_j + r_j$ where f_j is bounded and $\int |r_j|\, d\mu < \frac{\epsilon}{2k}$ $(j = 2, \dots, k)$. For example one can choose $f_j = h_j \cdot 1_{\{|h_j| \leq M\}}$ for some sufficiently large $M > 0$.

Let $R := r_0 + r_1 + \sum_{j=2}^k (r_j - r_j \circ T^{g_j})$. Then $\int |R|\, d\mu < 3\epsilon$ and

$$f = f_1 + \sum_{j=2}^{k} (f_j - f_j \circ T^{g_j}) + R$$

so that

$$\Delta f \leq \Delta f_1 + \sum_{j=2}^{k} \Delta(f_j - f_j \circ T^{g_j}) + \Delta R\,.$$

Now $\Delta f_1 = A^+ f_1 - A^- f_1 = f_1 - f_1 = 0$, and for $j = 2, \dots, k$ we observe that

$$\begin{aligned} |A_n(f_j - f_j \circ T^{g_j})| &= \left| \frac{1}{\lambda_n} \sum_{g \in \Lambda_n} f_j \circ T^g - \frac{1}{\lambda_n} \sum_{g \in \Lambda_n + g_j} f_j \circ T^g \right| \\ &\leq \frac{1}{\lambda_n} \sum_{g \in \Lambda_n \Delta (\Lambda_n + g_j)} |f_j \circ T^g| \\ &\leq \frac{1}{\lambda_n} \cdot 2d\, |g_j|\, n^{d-1} \cdot \|f_j\|_\infty \\ &= 2dn^{-1} |g_j|\, \|f_j\|_\infty\,, \end{aligned}$$

so that $A^-(f_j - f_j \circ T^{g_j}) = A^+(f_j - f_j \circ T^{g_j}) = 0$ and hence $\Delta(f_j - f_j \circ T^{g_j}) = 0$. It follows that

$$0 \leq \Delta f \leq \Delta R = A^+ R - A^- R \leq 2 A^+ |R|\,.$$

In view of (2.6) this implies

$$\int \Delta f\, d\mu \leq 2 \int A^+ |R|\, d\mu \leq 2\gamma_d \int |R|\, d\mu \leq 2\gamma_d \cdot 3\epsilon\,.$$

As $\epsilon > 0$ was arbitrary, we conclude that $\int \Delta f\, d\mu = 0$ and hence $\Delta f = 0$ μ-a.s. □

2.1.9 Remark The proof of the ergodic theorem given above goes back to articles by Kamae [30] and Katznelson and Weiss [32]. More traditional proofs, based on a so called maximal ergodic theorem, can be found in many textbooks on ergodic theory, in particular in Krengel's book [34] devoted completely to ergodic theorems. $\diamond$

2.1.10 Exercise Give a short proof of the L^2-*ergodic theorem* due to von Neumann:

$$\lim_{n\to\infty}\left\|\lambda_n^{-1}\sum_{g\in\Lambda_n} f\circ T^g - \bar{f}\right\|_2 = 0 \quad \forall f \in L^2_\mu ,$$

where $\bar{f}$ is the orthogonal projection of f on the subspace F from Lemma 2.1.8. $\diamond$

2.2 Ergodicity and mixing

The ergodic theorem is a rather general Law of Large Numbers for measure preserving dynamical systems, except that the limit $\bar{f}$ is not necessarily constant. Quite often, however, the limit turns out to be just $\mu(f)$ (as in Example 1.4.5). More generally, as $\bar{f}$ is a version of $E_\mu[f \mid \mathcal{I}_\mu(\mathcal{T})]$, the σ-algebra $\mathcal{I}_\mu(\mathcal{T})$ must be studied, and cases where it contains only null sets and sets of full measure are of particular interest, because in such instances $\bar{f}$ is constant μ-a.s.

2.2.1 Definition *The m.p.d.s. $(X, \mathcal{B}, \mu, \mathcal{T})$ is ergodic if $\mu(A) = 0$ or $\mu(A) = 1$ for all $A \in \mathcal{I}_\mu(\mathcal{T})$.*

2.2.2 Lemma *For a m.p.d.s. $(X, \mathcal{B}, \mu, \mathcal{T})$ the following statements are equivalent:*

1. *$(X, \mathcal{B}, \mu, \mathcal{T})$ is ergodic.*
2. *For all $f \in L^p_\mu$ $(1 \le p \le \infty)$ we have: if $f \circ T^g = f \bmod \mu$ for all $g \in G$, then $f = \text{const} \bmod \mu$.*
3. *$\bar{f} = \int f\, d\mu \bmod \mu$ for each measurable $f : X \to \mathbb{R}$ with $f^+ \in L^1_\mu$.*
4. *For all $A, B \in \mathcal{B}$*

$$\lim_{n\to\infty}\frac{1}{\lambda_n}\sum_{g\in\Lambda_n}\mu(T^{-g}A\cap B) = \mu(A)\mu(B)\,. \tag{2.9}$$

5. *There is an $\cap$-stable family $\mathcal{A} \subseteq \mathcal{B}$ generating $\mathcal{B}$ such that (2.9) holds for all $A, B \in \mathcal{A}$.*
6. *If ν is another $\mathcal{T}$-invariant probability measure and if $\nu \ll \mu$, then $\nu = \mu$.*