1. Introduction

1.1. Furstenberg’s intersection conjecture. Let $a \in \mathbb{N}$ and consider the map $T_a$ on the torus $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ which sends $x$ to $ax$. Furstenberg [Fu1, Fu2] introduced the idea that if $a \neq b$, the dynamics of $T_a$ and $T_b$ should not share any common structure unless there is an obvious algebraic reason. More precisely, recall that positive integers $a, b$ are said to be multiplicatively independent if they are not powers of the same integer, or equivalently if $\log a / \log b \notin \mathbb{Q}$. Furstenberg introduced several theorems and conjectures that all capture the idea that if $a$ and $b$ are multiplicatively independent, then the dynamics of $T_a$ and $T_b$ are “transverse” to each other in some senses. The first result along these lines [Fu2] is the statement that if $A$ is a closed subset of $\mathbb{T}$ which is invariant under both $T_a$ and $T_b$, then $A$ is either all of $\mathbb{T}$ or a finite set. This led to the very famous and wide open conjecture that if $\mu$ is any measure on $\mathbb{T}$ invariant under both $T_a$ and $T_b$, then $\mu$ is a linear combination of Lebesgue measure and a purely atomic measure.

There are many variations on this theme that are motivated by classical results in fractal geometry. One major example talks about projections of sets of the form $A \times B$, where $A$ is closed and $T_a$-invariant and $B$ is closed and $T_b$-invariant. A classical theorem of Marstrand (see e.g. [Ma]) states that if $\pi_u$ denotes the orthogonal projection of $\mathbb{R}^2$ onto the line with slope $u$ through the origin, then $\dim_H(\pi_u E) = \min(1, \dim_H(E))$ for Lebesgue almost every $u \in \mathbb{R}$. Furstenberg conjectured that in the special case where $E = A \times B$ as above, the equality in Marstrand’s theorem holds for literally every $u$ (except for $u = 0$ or $u = \infty$, corresponding to projections onto the $x$-axis or $y$-axis). This conjecture was proven in 2012 by Hochman and Shmerkin [HS] and has the aesthetically pleasing corollary that if $A$ and $B$ are as above, then $\dim_H(A + B) = \min(1, \dim_H(A) + \dim_H(B))$.

Another transversality conjecture, which is the focus of this paper, regards slices rather than projections of sets of the form $A \times B$. The slice conjecture is motivated by another classical theorem of Marstrand [Ma] which states that for any Borel set $E \subseteq \mathbb{R}^2$ and almost any line $\ell \subseteq \mathbb{R}^2$, $\dim_H(E \cap \ell) \leq \max(0, \dim_H(E) - 1)$. 

Date: August 12, 2019.
As in the version for projections, Furstenberg conjectured [Fu2] that in the special case of $E = A \times B$ as above, “almost every” is actually “every”.

**Conjecture 1.1** (Furstenberg’s intersection conjecture). Let $a, b \in \mathbb{N}$ be multiplicatively independent integers. Suppose $A \subseteq \mathbb{T}$ is closed and $T_a$-invariant and $B \subseteq \mathbb{T}$ is closed and $T_b$-invariant. Then for any line $\ell \subseteq \mathbb{R}^2$ not parallel to either of the coordinate axes, we have

\[
\dim_H((A \times B) \cap \ell) \leq \max(0, \dim_H(A) + \dim_H(B) - 1).
\]

In particular, $\dim_H(A \cap B) \leq \max(0, \dim_H(A) + \dim_H(B) - 1)$.

This conjecture was essentially proven in the special case $\dim_H(A) + \dim_H(B) < 1/2$ by Furstenberg in 1970 [Fu2] (see section 7 of [Ho3] for more details). However, the full conjecture was recently proven independently and simultaneously by Pablo Shmerkin [Sh] and Meng Wu [Wu]. The purpose of this paper is to give an exposition of the ideas behind Shmerkin’s proof. We will use those ideas to give the proof of a technically easier but morally similar result about the dimension of slices of a different kind of fractal set. In the section 2, we will explain more details and give the precise statements.

### 1.2. Notation and a note about proofs.

Due to the expository nature of this paper, we have elected to present many proofs in less than full detail in order to make the underlying ideas more clear. Many of the proofs require choosing and keeping track of several parameters that depend on each other in different ways, and we have tried to simplify the bookkeeping as much as possible by using imprecise notation. We will write $X \ll Y$ to mean “$X$ is much less than $Y$”, $X \sim Y$ to mean “$X$ is not much bigger than $Y$”, and $X \ll Y$ to mean “$X$ and $Y$ are about the same”. We reserve the symbols $\ll$ and $\sim$ for their usual meanings: $X \ll Y$ means that $X \leq CY$ for some constant $C$, and $X \sim Y$ means that $X \asymp Y$ and $Y \asymp X$. As usual, subscripts on the symbols $\ll$ and $\sim$ denote dependence of the implied constants on any other parameters.

The following chart summarizes some of the common notation used.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{T}$</td>
<td>Torus $\mathbb{R}/\mathbb{Z}$</td>
</tr>
<tr>
<td>$\text{Prob}(X)$</td>
<td>Space of Borel probability measures on a metric space $X$</td>
</tr>
<tr>
<td>$D_m$</td>
<td>Partition of $\mathbb{R}^d$ into dyadic cubes of side length $2^{-m}$</td>
</tr>
<tr>
<td>$D_m(A)$</td>
<td>Elements of $D_m$ meeting the set $A$</td>
</tr>
<tr>
<td>$\dim_H(A)$</td>
<td>Hausdorff dimension of the set $A$</td>
</tr>
<tr>
<td>$\dim_B(A)$</td>
<td>Box (Minkowski) dimension of the set $A$</td>
</tr>
<tr>
<td>$D(\mu, q)$</td>
<td>$L^q$-dimension of the measure $\mu$</td>
</tr>
<tr>
<td>$\pi\mu$</td>
<td>Pushforward of the measure $\mu$ by the map $\pi$</td>
</tr>
<tr>
<td>$\mu</td>
<td>_A$</td>
</tr>
<tr>
<td>$\mu_A$</td>
<td>Measure $\mu$ conditioned on the set $A$ (normalized): $\mu_A(E) = \frac{1}{\mu(A)} \mu(A \cap E)$</td>
</tr>
<tr>
<td>$[k]$</td>
<td>${1, 2, \ldots, k}$</td>
</tr>
<tr>
<td>$B_r(x)$</td>
<td>Ball of radius $r$ centered at $x$</td>
</tr>
</tbody>
</table>

## 2. Background and statement of results

This section summarizes most of the necessary background material necessary for the rest of the paper. Readers already familiar with these topics can safely skip this section and refer back to it later as needed.

### 2.1. Ergodic theory.

In this section, we collect some known results in ergodic theory that will be needed later, and give references for proofs.

**Theorem 2.1** (Kingman’s subadditive ergodic theorem). Let $(X, T, \mathbb{P})$ be an ergodic measure-preserving system and let $\psi_n \in L^1(\mathbb{P})$ be a sequence satisfying $\psi_{n+m}(x) \leq \psi_n(x) + \psi_m(T^n x) + C$ for all $n, m \in \mathbb{N}$, $x \in X$ and some constant $C > 0$. Then for $\mathbb{P}$-almost every $x$,

\[
\lim_{n \to \infty} \frac{\psi_n(x)}{n} = \inf_{n \in \mathbb{N}} \frac{\int_X \psi_n \, d\mathbb{P}}{n}.
\]

See [Ki] for a proof.
Theorem 2.2. Let $X$ be a compact metric space and $T : X \to X$ be continuous. Suppose $(X, T)$ is uniquely ergodic with unique invariant measure $\mathbb{P}$. Then for any continuous $f : X \to \mathbb{R}$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f \, d\mathbb{P}
$$

uniformly in $x \in X$.

See Section 4.3 of [EW] for a proof.

Theorem 2.3. Let $X$ be a compact metric space with $T : X \to X$ continuous, and suppose $(X, T)$ is uniquely ergodic with unique invariant measure $\mathbb{P}$. Let $\psi_n : X \to \mathbb{R}$ be a sequence of functions which are bounded and continuous $\mathbb{P}$-almost everywhere, and suppose that $\psi_{n+m}(x) \leq \psi_n(x) + \psi_m(T^n x)$. Let $L$ denote the $\mathbb{P}$-almost everywhere limit of $\psi_n(x)/n$. Then $\limsup_{n \to \infty} \psi_n(x)/x \leq L$ uniformly in $x$.

See Lemma 4.7 of [Sh] for a proof.

2.2. Iterated function systems. For the purposes of this paper, an iterated function system (IFS) is a collection $\Phi = (\phi_i)_{i=1}^k$ of contracting similarities in $\mathbb{R}^d$, i.e., $\phi_i : \mathbb{R}^d \to \mathbb{R}^d$ is given by $\phi_i(x) = \lambda_i O_i x + t_i$, where $\lambda_i \in (0, 1)$, $O_i$ is an orthogonal $d \times d$ matrix, and $t_i \in \mathbb{R}^d$. It is a well-known fact that given an IFS $\Phi$, there is a unique compact set $K \subseteq \mathbb{R}^d$ satisfying

$$
K = \bigcup_{i=1}^k \phi_i(K).
$$

Definition 2.4. The IFS $\Phi$ is said to satisfy the strong separation condition if the union in (2.3) is disjoint.

We will focus on iterated function systems $\Phi = (\phi_i)_{i=1}^k$ that satisfy the following nice simplifying assumptions.

(IFS0) $\Phi$ is an IFS in $\mathbb{R}^2$

(IFS1) Each $\phi_i$ has common contraction part $\lambda \in (0, 1)$, i.e. each $\lambda_i = \lambda$.

(IFS2) Each $\phi_i$ has common rotation part $2\pi \alpha$ for some $\alpha \notin \mathbb{Q}$, i.e. each $O_i = O_\alpha$ is a rotation by $2\pi \alpha$.

(IFS3) $\Phi$ satisfies the strong separation condition.

In this case, the attractor $K$ can be described explicitly as

$$
K = \left\{ \sum_{n=0}^\infty \lambda^n O_\alpha^n t_n : i = (i_0 i_1 \ldots) \in [k]^\mathbb{N} \right\}.
$$

Here we will collect a few useful properties of iterated function systems satisfying these conditions.

Definition 2.5. Let $\Phi = (\phi_i)_{i=1}^k$ be an IFS in $\mathbb{R}^2$ satisfying (IFS1) – (IFS3). Define the “coding map” $\pi : [k]^\mathbb{N} \to K$ by

$$
\pi(i_0 i_1 \ldots) = \lim_{N \to \infty} (\phi_{i_0} \circ \cdots \circ \phi_{i_N})(0) = \sum_{n=0}^\infty \lambda^n O_\alpha^n t_n.
$$

Let $p = \frac{1}{k} (\delta_1 + \ldots + \delta_k)$ be the uniform measure on $[k]$. The uniform self-similar measure on $K$ is the pushforward measure $\pi(p^{\mathbb{N}})$.

Lemma 2.6. Let $\Phi = (\phi_i)_{i=1}^k$ be an IFS in $\mathbb{R}^2$ satisfying (IFS1) – (IFS3) with attractor $K$. Set $s = \log k / \log(1/\lambda)$. Then $\dim_H(K) = \dim_B(K) = s$. Furthermore, if $\mu$ is the uniform self-similar measure on $K$, then for all $x \in K$ and all sufficiently small $r > 0$, we have $\mu(B_r(x)) \sim r^s$.

A proof of this lemma is contained in the proofs of Theorems 5.13 and 5.16 of [Ho1].

Lemma 2.7. Suppose $\Phi$ is an IFS in $\mathbb{R}^2$ satisfying the strong separation condition (IFS3) with attractor $K$. Let

$$
K_N := \left\{ \sum_{n=0}^{N-1} \lambda^n O_\alpha^n t_n : i = (i_0 i_1 \ldots i_{N-1}) \in [k]^N \right\}
$$

be the $N$th finite approximation to $K$. Then there is a constant $c > 0$ such that for all $N$, $K_N$ has $k^N$ elements and is $c \lambda^N$-separated.
Theorem 2.8 (Corollary 8.3 in [Sh]). Let \( \Phi = (\phi_i)_{i=1}^N \) be an iterated function system in \( \mathbb{R}^2 \) satisfying the conditions (IFS1) – (IFS3). Let \( K \) be its attractor and let \( s = \log k/\log(1/\lambda) = \dim_h(K) = \dim_B(K) \).

Then for any line \( \ell \subseteq \mathbb{R}^2 \),
\[
\dim_B(K \cap \ell) \leq \max(s-1,0).
\]

2.3. Measures on \( \mathbb{R}^d \). Most of this work is concerned with the small-scale structure of probability measures on Euclidean space. In this section we will define the relevant notions and list some basic properties.

Definition 2.9. Let \( \mu, \nu \) be probability measures on \( \mathbb{R}^d \). The convolution \( \mu \ast \nu \) is defined to be the distribution of the random variable \( X + Y \) where \( X, Y \) are independent random variables distributed according to \( \mu, \nu \) respectively. Alternatively, it can be defined by the formula
\[
(\mu \ast \nu)(A) = \iint 1_A(x+y) \, d\mu(x) \, d\nu(y).
\]

Example 2.10. If \( \mu = \sum_i p_i \delta_{x_i} \) and \( \nu = \sum_j q_j \delta_{y_j} \) are purely atomic, then \( \mu \ast \nu \) puts a mass of \( p_i q_j \) at the point \( x_i + y_j \).

Definition 2.11. For \( m \in \mathbb{N} \), let \( \mathcal{D}_m \) be the partition of \( \mathbb{R}^d \) into \( 2^m \)-adic cells, i.e.
\[
\mathcal{D}_m = \left\{ [a_1 2^{-m}, (a_1 + 1)2^{-m}) \times \cdots \times [a_d 2^{-m}, (a_d + 1)2^{-m}) : a_1, \ldots, a_d \in \mathbb{Z} \right\}.
\]

If \( A \) is any subset of \( \mathbb{R}^d \), we also write \( \mathcal{D}_m(A) \) to denote the elements of \( \mathcal{D}_m \) that meet \( A \).

For \( \mu \in \text{Prob}(\mathbb{R}^d) \), let \( \mu^m \) be the probability measure supported on \( 2^{-m} \mathbb{Z}^d \) defined by
\[
\mu^m(x) = \mu\left( x + [0,2^{-m})^d \right).
\]

This is the “discretized version of \( \mu \) at scale \( 2^{-m} \).

Definition 2.12. Let \( q \in [1,\infty) \). If \( \mu \in \text{Prob}(\mathbb{R}^d) \) is purely atomic, then the \( L^q \) norm of \( \mu \) is defined by
\[
||\mu||_q^q = \sum_{x \in \mathbb{R}^d} \mu(x)^q.
\]

It is immediate from the definition that \( ||\mu||_1 = 1 \) and that \( q \mapsto ||\mu||_q^q \) is decreasing and tends to 0 as \( q \to \infty \). We now list some properties of convolutions, discretizations, and \( L^q \) norms that will be useful later.

Lemma 2.13. Let \( \mu, \nu \in \text{Prob}(\mathbb{R}^d) \) and \( q \in (1,\infty). \) For \( a > 0 \), define \( S_a : \mathbb{R}^d \to \mathbb{R}^d \) by \( S_a(x) = ax \). Then the following hold.

(a) \( ||\mu^{m+\ell}||_q \leq ||\mu^m||_q \).

(b) If \( \nu \) is purely atomic, then \( ||\nu^m||_q \geq ||\nu||_q \) with equality if and only if \( 2^{-m} \) is smaller than the smallest distance between any two atoms of \( \nu \).

(c) If \( \mu \) and \( \nu \) are purely atomic, then \( ||\mu \ast \nu||_q \geq ||\mu||_q ||\nu||_q \) with equality if and only if all of the atoms of \( \mu \ast \nu \) are distinct.

(d) \( S_2^m(\mu^{m+\ell}) = (S_2^m \mu)^\ell \).
(e) \( \pi(\mu * \nu) = \pi \mu * \pi \nu \) if \( \pi : \mathbb{R}^d \rightarrow \mathbb{R}^n \) is linear. In particular, for any \( a > 0 \), \( S_a(\mu * \nu) = S_a \mu * S_a \nu \).

(f) Suppose \( \mathcal{P}, \mathcal{Q} \) are finite families of measurable sets such that each element of \( \mathcal{P} \) can be covered by at most \( M \) elements of \( \mathcal{Q} \), and each element of \( \mathcal{Q} \) meets at most \( M \) elements of \( \mathcal{P} \). Then

\[
\sum_{P \in \mathcal{P}} \mu(P)^q \leq M^q \sum_{Q \in \mathcal{Q}} \mu(Q)^q.
\]

(g) Suppose \( \mu_1, \ldots, \mu_n \) are all purely atomic, and suppose each point of \( \mathbb{R} \) is in the support of at most \( M \) of the \( \mu_i \). Then

\[
\left\| \sum_{i=1}^n \mu_i \right\|_q \leq M^{q-1} \sum_{i=1}^n \|\mu_i\|_q^q.
\]

Proof. (a) Follows directly from the elementary inequality \( x_1^q + \ldots + x_n^q \leq (x_1 + \ldots + x_n)^q \).

(b) Same reason as part (a). In the case where \( 2^{-m} \) is smaller than the smallest distance between any two atoms of \( \nu \), then \( \nu \) and \( \nu_m \) are the same measure, so equality obviously holds.

(c) Write \( \mu = \sum_i \mu_i \delta_{x_i} \) and \( \nu = \sum_j \nu_j \delta_{y_j} \). If all of the atoms of \( \mu + \nu \) are distinct, then \( \|\mu + \nu\|_q = \sum_{i,j} (\mu_i \delta_{x_i} \nu_j \delta_{y_j})^q = \sum_i \mu_i^q \nu_j^q \|\mu_i\|_q^q \|\nu_j\|_q^q \). If some of the atoms overlap, then some of the terms are added together before being raised to the \( q \)th power, and then the desired inequality follows from the same elementary inequality as in part (a).

(d) Verified by a direct calculation.

(e) Let \( X, Y \) be independent random variables with distributions \( \mu, \nu \) respectively. Then \( \pi(\mu + \nu) = \pi X + \pi Y \), which has the distribution \( \pi \mu + \pi \nu \).

(f) For each \( P \in \mathcal{P} \), let \( \mathcal{Q}_P \) be a subcollection of \( \mathcal{Q} \) that covers \( P \) and has \( \leq M \) elements. Then

\[
\sum_{P \in \mathcal{P}} \mu(P)^q \leq \sum_{Q \in \mathcal{Q}_P} \left( \sum_{Q \in \mathcal{Q}_P} \mu(Q) \right)^q \leq \sum_{Q \in \mathcal{Q}} \mu(Q)^q \leq M^q \sum_{Q \in \mathcal{Q}} \mu(Q)^q
\]

where the second inequality is Hölder and the third inequality is because each \( Q \in \mathcal{Q} \) appears no more than \( M \) times in all of the \( \mathcal{Q}_P \) combined.

(g) For \( x \in \mathbb{R} \) fixed, apply Hölder’s inequality only to the nonzero terms in the sum \( (\mu_1(x) + \mu_2(x) + \ldots + \mu_n(x))^q \), of which there are at most \( M \), to get \( (\mu_1(x) + \mu_2(x) + \ldots + \mu_n(x))^q \leq M^{q-1}(\mu_1(x)^q + \mu_2(x)^q + \ldots + \mu_n(x)^q) \). Then sum both sides over all \( x \in \mathbb{R} \) to get the desired result.

(h) We have

\[
\left\| (\mu * \nu)^m \right\|_q = \sum_{I \in \mathcal{D}_m} (\mu * \nu)(I)^q = \sum_{I \in \mathcal{D}_m} (\mu * \nu)(P_I)^q
\]

where \( P_I := \{(x, y) \in \mathbb{R}^2 : x + y \in I \} \). We also have

\[
\left\| \mu^m \right\|_q = \sum_{x \in 2^{-m} \mathbb{Z}} (\mu(x))^q = \sum_{x \in 2^{-m} \mathbb{Z}} \left( \sum_{J \in \mathcal{D}_m} (J \mu)(x - J) \right)^q = \sum_{x \in 2^{-m} \mathbb{Z}} (\mu * \nu)(Q_x)^q
\]

where \( Q_x := \bigcup_{J \in 2^{-m} \mathbb{Z}} J \times (x - J) \).

Set \( \mathcal{P} = \{ P_I : I \in \mathcal{D}_m \} \) and \( \mathcal{Q} = \{ Q_x : x \in 2^{-m} \mathbb{Z} \} \). Then \( \mathcal{P} \) and \( \mathcal{Q} \) both cover each other in the way necessary for part (g) to apply, so the claim follows.

The next important notion is that of the \( L^q \) dimension of a measure.

**Definition 2.14.** Let \( \mu \in \operatorname{Prob}(\mathbb{R}^d) \) and \( q \in (1, \infty) \). The upper \( L^q \) dimension of \( \mu \) is defined as

\[
\overline{D}(\mu, q) := \limsup_{m \to \infty} - \frac{\log \|\mu^m\|_q}{(q-1)m} = \limsup_{m \to \infty} - \frac{\log \sum_{I \in \mathcal{D}_m} \mu(I)^q}{(q-1)m}.
\]

The lower \( L^q \) dimension \( \underline{D}(\mu, q) \) is defined analogously, and if the limit exists it is just called the \( L^q \) dimension and is denoted \( D(\mu, q) \).

**Remark 2.15.** It is an easy exercise in Hölder’s inequality to check that \( 0 \leq D(\mu, q) \leq 1 \) for any \( \mu, q \). The factor of \( q - 1 \) in the denominator is only there to ensure this normalization.
Example 2.16. If $\mu$ is Lebesgue measure restricted to the unit cube, then $D(\mu, q) = 1$ for all $q$. If $\nu$ is purely atomic, then $D(\nu, q) = 0$ for all $q$.

The $L^q$ dimension has the following well-known properties.

Lemma 2.17. Let $\mu \in \text{Prob}(\mathbb{R}^d)$ be supported on a bounded interval and define
\begin{equation}
\tau_\mu(q) := \liminf_{m \to \infty} -\frac{\log \sum_{I \in \mathcal{D}_m} \mu(I)^q}{m} = (q-1)D(\mu, q)
\end{equation}
for $q \in [1, \infty)$. Then $\tau_\mu$ is increasing, concave, $\tau_\mu(1) = 0$, and $\tau_\mu(q) \leq q-1$ for all $q$.

Proof. For a proof of the first three assertions, see Proposition 3.2 of [LN]. For the last assertion, simply apply Hölder’s inequality to the equation $1 = \sum_{I \in \mathcal{D}_m} \mu(I)$ and rearrange. \qed

2.4. Self-similar measures. To introduce the language of self-similar measures that will be used in this paper, we consider a generalization in which a different measure can be used at each scale, i.e. we will consider measures of the form
\begin{equation}
\mu \in \{ \text{probability measures on } (X, T) \}.
\end{equation}
However, we still want to have some structure so we will insist that the measures $\mu$ will be supported on a bounded interval and define
\begin{equation}
\tilde{\mu} := \mu \circ \tau_{-1}^{-1},
\end{equation}
where $\mu$ is the uniform measure on the translation parts of the $\phi_i$. This is another expression of the self-similarity of $\mu$ – it says roughly that $\mu$ is obtained by attaching the measure $\tilde{\mu}$ to itself at smaller and smaller scales. Also, if we define the discrete approximation $\mu_n := \sum_{i=1}^{k} \delta_{\phi_i \mu}$, it is easy to check that
\begin{equation}
\mu = \mu_n \ast S^\lambda \mu.
\end{equation}

In this paper we consider a generalization in which a different measure can be used at each scale, i.e. we will consider measures of the form $\mu = \sum_{n=0}^{\infty} S^\lambda \mu_n$ where $0 < \lambda < 1$ and $\mu_n$ are any atomic probability measures on $\mathbb{R}$. However, we still want to have some structure so we will insist that the measures $\mu_n$ are generated by some nice dynamical system.

Definition 2.18. Let $(X, T)$ be a dynamical system and fix $0 < \lambda < 1$. Let $\Delta$ be a function which assigns to each $x \in X$ a purely atomic element of $\text{Prob}(\mathbb{R})$. Then for each $x \in X$, define the measure
\begin{equation}
\mu_x := \sum_{n=0}^{\infty} S^\lambda \Delta(T^n x) \in \text{Prob}(\mathbb{R}).
\end{equation}
The measures $\mu_x$ are called dynamically driven self-similar (DDSS) measures, and the quadruple $(X, T, \Delta, \lambda)$ is called a model generating the measures $\mu_x$. We also define the discrete approximations
\begin{equation}
\mu_{x,n} := \sum_{i=0}^{n-1} S^\lambda \Delta(T^i x).
\end{equation}
The first part of the next lemma is an analog of the self-similarity property (2.22) that justifies calling these measures “dynamically driven self-similar”. The second part is a useful estimate that we will use many times.

Lemma 2.19. Let $(X, T, \Delta, \lambda)$ be a model generating the measures $\mu_x$.

(a) For all $x \in X$, $n \in \mathbb{N}$, we have $\mu_x = \mu_{x,n} \ast S^\lambda \mu_{T^n x}$.

(b) For $n \in \mathbb{N}$, let $m(n) = \lfloor n \log(1/\lambda) \rfloor$ be the smallest integer so that $2^{-m} \leq x$. Then for any $x \in X$, $n \in \mathbb{N}$, $q > 1$, we have $\| \mu_{x,n} \|_q \sim_q \| \mu_{m(n)} \|_q$.

Proof. (a) $\mu_{x,n}$ is the distribution of the random variable $\sum_{i=1}^{n-1} \lambda^i X_i$ where the $X_i$ are independent and $X_i$ has distribution $\Delta(T^i x)$. $S^\lambda \mu_{T^n x}$ is the distribution of the random variable $\sum_{j=0}^{\infty} \lambda^{i+n} X_j$ where the $X_j$ are independent with distribution $\Delta(T^{i+n} x)$. So the right side of the claimed formula is the distribution of the sum of these two variables, which is exactly the sum $\sum_{i=0}^{\infty} \lambda^i X_i$ where the $X_i$ are independent and distributed according to $\Delta(T^i x)$, which is the definition of $\mu_x$.\qed
(b) $\mu_{x,n}$ is a purely atomic measure with minimum spacing between atoms $\approx \lambda^n \approx 2^{-m(n)}$, so the claim follows from part (b) of Lemma 2.13.

In order to prove anything significant about DDSS measures, it is natural to assume some regularity on the model that generates them.

**Definition 2.20.** A model $(X, T, \Delta, \lambda)$ is said to be pleasant if $X$ is a compact metric space, $T$ is a uniquely ergodic transformation on $X$, the DDSS measures $\mu_T$ are all non-atomic and supported on a fixed bounded interval, and the map $x \mapsto \mu_T$ is continuous almost everywhere (where “continuous” means with respect to the weak-* topology on $\text{Prob}(\mathbb{R})$ and “almost everywhere” means with respect to the unique invariant measure for $T$).

**Definition 2.21.** Let $(X, T, \Delta, \lambda)$ be a pleasant model and let $\mathbb{P}$ be the unique invariant measure. The model is said to have exponential separation if for $\mathbb{P}$-almost every $x$, there exists an $R > 0$ such that for infinitely many $n$, all of the atoms of $\mu_{x,n}$ are distinct and $\lambda^{rn}$-separated.

The following theorem is the main tool that drives the proof of Theorem 2.8. The proof is given in Section 4.

**Theorem 2.22** (Theorem 1.11 in [Sh]). Let $(X, T, \Delta, \lambda)$ be a pleasant model with exponential separation generating the family of DDSS measures $\{\mu_x\}_{x \in X} \subseteq \text{Prob}(\mathbb{R})$. Denote by $\mathbb{P}$ the unique invariant measure for $(X, T)$. Assume that the map $x \mapsto \Delta(x)$ is a continuous $\mathbb{P}$-a.e. and the number of atoms of $\Delta(x)$ is a bounded function of $x \in X$. Then for all $x \in X$ and all $q > 1$, $D(\mu_{x,q})$ exists and is equal to

$$
(2.25) \quad \min \left( \frac{1}{(q-1)\log \lambda} \int_X \log ||\Delta(x)||^q d\mathbb{P}(x), 1 \right).
$$

2.5. Small scale structure of sets. Let $A$ be any subset of $[0, 1]$. To analyze the fractal (multiscale) structure of $A$, it is often convenient to identify $A$ with a tree $T_A$ in the following way. First fix a “scale step size” $D \in \mathbb{N}$. Then the nodes of the $r$th level of $T_A$ are the elements of $D_r(A)$, i.e. the $2^{-r}D$-adic intervals which intersect $A$. An interval $J \in D_{r+1}(A)$ is a descendant of $I \in D_r(A)$ if and only if $J \subseteq I$. Each point of $A$ corresponds to a unique infinite path down the levels of $T_A$.

We are most interested in the branching structure of the tree $T_A$. The number of nodes that $T_A$ has on a given level tells us how “spread out” the set $A$ is on a certain scale. For example, take $D = 1$ and consider $A = 2^{-m}[0,1], B = [0,2^{-m})$. Down to the $m$th level, each node of $T_A$ has the maximum number of descendants (two) while each node of $T_B$ has only one descendant, but beyond the $m$th level, the nodes of $T_A$ have one descendant and the nodes of $T_B$ have two.

**Definition 2.23.** Let $A \subseteq [0, 1], D \in \mathbb{N}$, and let $T_A$ be the tree associated to $A$ with scale step size $D$. The set $A$ is said to have regular branching up to level $\ell$ if for all $0 \leq r \leq \ell - 1, |D_{r+1}(A \cap I)|$ is constant as $I$ ranges over $D_r(A)$. In other words, on each level of the tree $T_A$ up to level $\ell$, each node has the same number of descendants (the number of descendants can vary from level to level).

If $A$ has regular branching, it is said to have full branching at level $r$ if $|D_{r+1}(A \cap I)| = 2^D$ for all $I \in D_r(A)$, i.e. each node on the $r$th level has the maximum number of descendants. The set $A$ is said to have no branching at level $r$ if $|D_{r+1}(A \cap I)| = 1$ for all $I \in D_r(A)$.

3. Proof of intersection conjecture assuming dimension formula

In this section, we prove Theorem 2.8 assuming Theorem 2.22. We roughly follow Section 8 of [Sh].

Let $\{\phi_i\}_{i=1}^k$ be an iterated function system in $\mathbb{R}^2$ satisfying the hypotheses of Theorem 2.8 and let $K$ be its attractor. Recall that $s := \dim K = -\log k / \log \lambda$. Let $\bar{\mu} = \frac{1}{k} \sum_{i=1}^k \delta_{\phi_i}$ and let $\mu = *_{n=0}^\infty S_{\lambda^n} R_\alpha^\infty \bar{\mu}$ be the uniform self-similar measure on $K$. For $x \in \mathbb{T}$, define $P_x : K \rightarrow \mathbb{R}$ to be the projection onto the line through the origin in the direction of $e^{2\pi i x}$ followed by a rotation onto the horizontal axis. Let $\mu_x = P_x \mu$. Then we have

$$
(3.1) \quad \mu_x = P_x \left( *_{n=0}^\infty S_{\lambda^n} R_\alpha^\infty \mu \right) = *_{n=0}^\infty P_x S_{\lambda^n} R_\alpha^\infty \mu = *_{n=0}^\infty S_{\lambda^n} P_x R_\alpha^\infty \mu = S_{\lambda^n} P_{R_\alpha^\infty x} \bar{\mu}
$$

where the second and third equalities hold because $P_x : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a linear map and therefore commutes with convolutions and scaling, and the last equality can be checked directly. So $\mu_x$ is the DDSS measure generated by the model $(\mathbb{T}, R_{-\alpha}, \Delta, \lambda)$ where $\Delta(x) := P_x \bar{\mu}$.
Lemma 3.1. The model \((\mathbb{T}, R_{-\alpha}, \Delta, \lambda)\) satisfies all of the hypotheses of Theorem 2.22.

Proof. Since \(\alpha \not\in \mathbb{Q}\), \((\mathbb{T}, R_{-\alpha})\) is uniquely ergodic with Lebesgue measure as the unique invariant measure. It’s clear that every \(\mu_x\) is non-atomic and supported on a fixed compact interval because they are all projections of the same non-atomic, compactly supported measure on \(\mathbb{R}^2\). Furthermore, if \(x_n \to x\), then \(P_{x_n} \to P_x\) uniformly on compact sets, so \(\mu_{x_n} = P_{x_n}\tilde{\mu} \to \mu\tilde{\mu}\) in the weak-* topology on \(P_x\tilde{\mu}\). Therefore \((\mathbb{T}, R_{-\alpha}, \Delta, \lambda)\) is a pleasant model.

To establish exponential separation let \(\mu_n = S_n x_{\alpha} \mu\) and note that \(\mu_{x,n} = P_{x,n}\mu\). By the strong separation condition, the atoms of \(\mu_n\) are all distinct and \(c\lambda^n\)-separated for some \(c > 0\) (see Lemma 2.7). If \(a, b\) are two atoms of \(\mu_n\), then \(|P_x a - P_x b| = |a - b| \cos \theta(x; a, b) \geq c\lambda^n \cos \theta(x; a, b)|\), where \(\cos \theta(x; a, b)\) is the angle between the line through the origin in direction \(e^{2\pi i x}\) and the line connecting \(a\) and \(b\) in \(\mathbb{R}^2\). So if \(|P_x a - P_x b| < \lambda R^m\), then \(\cos \theta(x; a, b) \leq \lambda(R-1)^n\), and for \(a, b\) fixed, the set of \(x\) that satisfy this has Lebesgue measure \(\lesssim \lambda(R-1)^n\). So by taking a union bound over all pairs of atoms \(a, b \in \text{supp}(\mu_n)\), we see that the set of \(x\) for which all of the atoms of \(\mu_{x,n}\) are distinct and \(\lambda R^n\)-separated has Lebesgue measure at least \(1 - CK^{2n}\lambda(R-1)^n\) for some (large) constant \(C\). If we pick \(R\) large enough so that that expression tends to \(1\) as \(n \to \infty\), we conclude that for almost every \(x \in \mathbb{T}\) the atoms of \(\mu_{x,n}\) are distinct and \(\lambda R^n\)-separated for infinitely many \(n\).

Finally, it’s clear that \(x \mapsto \Delta(x) = P_x\tilde{\mu}\) is continuous, and the number of atoms of \(P_x\tilde{\mu}\) is always at most the number of atoms of \(\tilde{\mu}\), completing the proof. \(\square\)

Thus we can apply Theorem 2.22 and conclude that

\[ (3.2) \quad D(P_x\mu, q) = \min \left( \frac{1}{(q-1)\log \lambda} \int_0^1 \log ||P_x\tilde{\mu}||_q^q dx, 1 \right) = \min \left( \frac{\log ||\tilde{\mu}||_q^q}{(q-1)\log \lambda}, 1 \right) = \min(s, 1) \]

for all \(q > 1\) and all \(x \in \mathbb{T}\).

Now we translate the information about \(L^q\) dimension into information about Hausdorff dimension. Let \(D = \min(s, 1)\). By (3.2), if \(m\) is sufficiently large then

\[ (3.3) \quad \sum_{\ell \in D_m} (P_x\mu)(\ell)^q \approx 2^{-D(q-1)m}. \]

For sufficiently small \(r\), if we choose \(m\) so that \(2^{-m} \approx r\), any ball in \(\mathbb{R}\) of radius \(r\) can be covered by \(O(1)\) many elements of \(D_m\), so

\[ (3.4) \quad (P_x\mu)(B_r(y)) \lesssim \left( \sum_{\ell \in D_m} (P_x\mu)(\ell)^q \right)^{1/q} \approx r^{D(1-1/q)}. \]

On the other hand, we know from Lemma 2.6 that \(\mu(B_r(z)) \approx r^s\) for \(z \in K\) and \(r\) sufficiently small. Combined with (3.4), this translates to an upper bound on the box dimension of the fibers of \(P_x\) as follows.

Fix \(a \in \mathbb{R}\), let \(r > 0\) and let \((z_j)_{j=1}^{M(r)}\) be a maximal \(2r\)-separated subset of \(P_x^{-1}(a)\). Let \(U\) be the disjoint union \(\bigcup_j B_r(z_j)\). On one hand, we have

\[ (3.5) \quad \mu(U) = \sum_j \mu(B_r(z_j)) \approx M(r)r^s, \]

and on the other hand, since \(P_x\) is Lipschitz, \(P_x(U)\) is contained in a ball of radius \(\lesssim r\) centered at \(a \in \mathbb{R}\), so

\[ (3.6) \quad \mu(U) \leq \mu(P_x^{-1}(P_x(U))) = (P_x\mu)(P_x(U)) \lesssim r^D. \]

Combining these two equations yields \(M(r) \lesssim r^{D-s}\), so

\[ (3.7) \quad \overline{\dim}_B(P_x^{-1}(a)) = \limsup_{r \to 0} \frac{\log M(r)}{\log 1/r} \leq s - D = \max(s-1, 0). \]

The proof is completed by noting that any slice \(K \cap \ell\) is equal to \(P_x^{-1}(a)\) for some choice of \(a \in \mathbb{R}\), \(x \in \mathbb{T}\). \(\square\)
4. Proof of the dimension formula

4.1. Upper bound. Let \((X,T,\Delta,\lambda)\) be a model satisfying all of the hypotheses of Theorem 2.22. For simplicity, assume from now on that all of the measures \(\mu_x\) are supported on \([0,1]\) (the hypothesis of pleasantness implies that up to translation and scaling, we have not lost any generality). Our goal for this section is to prove the upper bound

\[
\limsup_{m \to \infty} -\frac{\log \|\mu_x^m\|_q^q}{(q-1)m} \leq \min \left\{ \frac{1}{(q-1)\log \lambda} \int_X \log \|\Delta(x)\|_q^q \, d\mathcal{P}(x), 1 \right\}
\]

uniformly in \(x \in X\).

First, by Hölder’s inequality we have

\[
1 = \sum_{I \in \mathcal{D}_m} \mu_x(I) \leq \|\mu_x^m\|_q 2^{m(1-1/q)},
\]

so rearranging gives \(\|\mu_x^m\|_q^q \geq 2^{-m(q-1)}\) and therefore the left side of (4.1) is at most 1.

Let \(m(n) = \lfloor n \log(1/\lambda) \rfloor\), so that \(2^{-m(n)} \approx \lambda^n\). By Lemma 2.13, we have

\[
\|\mu_{x,n}^{m(n)}\|_q^q \approx \|\mu_{x,n}^{m(n)}\|_q^q \geq \|\mu_{x,n}\|_q^q \geq \prod_{i=0}^{n-1} \|\Delta(T^ix)\|_q^q,
\]

so

\[
\limsup_{n \to \infty} -\frac{\log \|\mu_{x,n}^{m(n)}\|_q^q}{(q-1)m(n)} \leq \limsup_{n \to \infty} -\frac{n}{(q-1)m(n)} \sum_{i=0}^{n-1} \log \|\Delta(T^ix)\|_q^q.
\]

Our hypotheses on the model guarantee that \(x \mapsto \log \|\Delta(x)\|_q^q\) is a bounded continuous function of \(x\), so by unique ergodicity, the right side of the above equation converges to

\[
\frac{1}{(q-1)\log \lambda} \int_X \log \|\Delta(x)\|_q^q \, d\mathcal{P}(x)
\]

uniformly in \(x\) as \(n \to \infty\) (see Theorem 2.2).

This establishes the desired result only along the subsequence \(m(n)\) as \(n \to \infty\), but we can upgrade it to convergence as \(m \to \infty\) using a standard interpolation trick as follows. For fixed \(m\), let \(N\) be chosen such that \(m(N-1) \leq m < m(N)\). Then since \(m \mapsto -\log \|\mu_x^m\|_q^q\) is increasing, we have

\[
\frac{-\log \|\mu_x^{m(N-1)}\|_q^q}{m(N-1)} \leq \frac{-\log \|\mu_x^{m(N)}\|_q^q}{m(N)} \leq \frac{-\log \|\mu_x^{m(N)}\|_q^q}{m} \leq \frac{-\log \|\mu_x^{m(N)}\|_q^q}{m(N-1)} = \frac{-\log \|\mu_x^{m(N)}\|_q^q}{m(N)}.
\]

Now since \(m(N) = N \log(1/\lambda)\), \(m(N)/m(N-1) \to 1\) as \(N \to \infty\) and therefore the full sequence tends to the same limit as the subsequence along \(m(N)\) as desired. \(\square\)

4.2. Steps toward a matching lower bound. Let \(\psi_n(x) := \log \|\mu_{x,n}^{m(n)}\|_q^q\) (we leave the dependence on \(q\) implicit). Using Lemmas 2.13 and 2.19, we have the estimate

\[
\left\|\mu_{x,n}^{m(n)+n'}\right\|_q^q = \left\|\mu_{x,n} \ast \lambda_n \ast T^{n'}\right\|_q^q \approx \left\|\mu_{x,n}^{m(n)+n'} \ast \lambda_n \ast T^{n'}\right\|_q^q
\]

\[
\approx \sum_{I \in \mathcal{D}(m,n')} \left\|\mu_{x,n}^{m(n)+n'}\right\|_q^q \left\|\lambda_n \ast T^{n'}\right\|_q^q \approx \sum_{I \in \mathcal{D}(m,n')} \mu_{x,n}(I) \left\|\lambda_n \ast T^{n'}\right\|_q^q
\]

\[
= \left\|\mu_{x,n}^{m(n)}\right\|_q^q \sum_{J \in \mathcal{D}(m,n')} \mu_{T^{n'}(S_{\lambda-n}J)} \approx \left\|\mu_{x,n}^{m(n)}\right\|_q^q \left\|\mu_{T^{n'}x}^{m(n')}\right\|_q^q
\]

where the second inequality in the second line is Young’s convolution inequality. Thus we can apply the subadditive ergodic theorem (Theorem 2.1) to the \(\psi_n\), and therefore the expression

\[
\frac{1}{n} \log \left\|\mu_{x,n}^{m(n)}\right\|_q^q
\]
converges $\mathbb{P}$-almost everywhere to a constant as $n \to \infty$. By the same interpolation trick as in the previous section, we conclude that

$$\lim_{m \to \infty} -\frac{1}{m} \log \|\mu_x^m\|_q^q \quad \text{exists and is equal to a constant } \mathbb{P}\text{-almost everywhere. Call this constant } T(q).$$

Whenever the limit in (4.10) exists, it is by definition equal to $(q-1)D(\mu_x, q)$. So we know that for each $q$, there is a full measure set of $x$ for which $T(q) = (q-1)D(\mu_x, q)$. Apply this to each $q$ in a countable dense subset of $(1, \infty)$ to obtain a single full measure set of $x$ for which $T(q) = (q-1)D(\mu_x, q)$ for all $q \in$ a dense subset of $(1, \infty)$. Fix such an $x$. By Lemma 2.17, the function $q \mapsto (q-1)D(\mu_x, q)$ is increasing and concave and therefore continuous. It is also obvious from the definition that $T(q)$ is an increasing function of $q$. So we have two increasing functions that agree on a dense subset, and one of them is continuous, which implies that they must agree everywhere. We have thus proven that there exists a full measure set of $x$ for which $(q-1)D(\mu_x, q) = T(q)$ for all $q > 1$.

This is not quite strong enough for our purposes because Theorem 2.22 is a statement about $D(\mu_x, q)$ for all $x \in X$. But since we are working with a pleasant model, we may appeal to Theorem 2.3 to upgrade the above result to the stronger statement that

$$\liminf_{m \to \infty} -\frac{1}{m} \log \|\mu_x^m\|_q^q \geq T(q)$$

uniformly in $x \in X$.  \footnote{Actually, Theorem 2.3 only applies to continuous observables, and $x \mapsto \log \|\mu_x^m\|_q^q$ may not be continuous. But one can show that it may be approximated by something continuous (see Corollary 4.8 of [Sh] for details).}

So to complete the proof of Theorem 2.22, we need to prove the formula

$$T(q) = \min \left( \frac{1}{\log \lambda} \int_X \log \|\Delta(x)\|_q^q \, d\mathbb{P}(x), \ q-1 \right)$$

for all $q > 1$. The rest of the paper is devoted to proving (4.12). We can make some simplifying assumptions. First, we have seen that $T(q)$ is an increasing concave function, and therefore it is continuous everywhere and differentiable outside of an at most countable set. So it suffices to prove (4.12) for a fixed $q$ at which $T$ is differentiable. Also, by Lemma 2.17 we know $T(q) \leq q-1$ for all $q$, so from now on we will assume that $T(q) < q-1$ and prove that $T(q)$ is equal to the first term in the min in (4.12).

4.3. Exponential flattening of DDSS measures under convolution. Recall that we are working with a fixed model $(X, T, \Delta, \lambda)$ satisfying all of the hypotheses of Theorem 2.22, generating the DDSS measures $\mu_x, x \in X$. The goal of this section is the following theorem.

**Theorem 4.1** (Theorem 5.1 in [Sh]). Fix $q > 1$ such that $T$ is differentiable at $q$ and $T(q) < q-1$, and let $\sigma > 0$. Then there exists $\epsilon = \epsilon(\sigma, q)$ such that if $m$ is sufficiently large in terms of $\sigma, q, \epsilon$, the following holds: If $\nu$ is a measure supported on $2^{-m}Z$ and $\|\nu\|_q^q \leq 2^{-m}$, then

$$\|\nu * \mu_x^m\|_q^q \leq 2^{-m}2^{-T(q)m}$$

for every $x \in X$.

Before proving Theorem 4.1, we need some detailed information about the small-scale structure of the measures $\mu_x$. For values of $q$ at which $T$ is differentiable, the $L^q$ norms of the measures $\mu_x^m$ display nice regularity at small scales. The Legendre transform of $T$ plays an important role in these results.

**Definition 4.2** (Legendre transform). If $T : \mathbb{R} \to \mathbb{R}$ is a concave function, its Legendre transform $T^* : \mathbb{R} \to (-\infty, \infty)$ is defined by

$$T^*(\alpha) = \inf_{q \in \mathbb{R}} \alpha q - T(q).$$

**Lemma 4.3.** If $T$ is concave and $\alpha = T'(q)$ exists, then $T^*(\alpha) = \alpha q - T(q)$.

**Proof.** By definition,

$$T^*(\alpha) = \inf_{x \in \mathbb{R}} T'(q)x - T(x).$$

$$\lim_{m \to \infty} -\frac{1}{m} \log \|\mu_x^m\|_q^q$$
We have
\begin{equation}
(4.15) \quad T'(q)x - T(x) - (T'(q)q - T(q)) = \left( T'(q) - \frac{T(x) - T(q)}{x - q} \right) (x - q) \geq 0
\end{equation}
for all \( x \in \mathbb{R} \) by the concavity of \( T \), so the inf in (4.14) is achieved at \( x = q \).

\[ \square \]

**Lemma 4.4.** Let \( q > 1 \). If \( T \) is concave, \( \alpha = T'(q) \) exists, \( T(q) < q - 1 \), and \( T(1) = 0 \), then \( T^*(\alpha) \leq \alpha < 1 \).

*Proof.* First we have
\begin{equation}
(4.16) \quad \alpha = T'(q) \leq \frac{T(q) - T(1)}{q - 1} = \frac{T(q)}{q - 1} < 1,
\end{equation}
and then by definition of \( T^* \) we have
\begin{equation}
(4.17) \quad T^*(\alpha) \leq \alpha \cdot 1 - T(1) = \alpha.
\end{equation}
\[ \square \]

The following lemmas are all different manifestations of the heuristic principle “if \( \alpha = T'(q) \) exists, then at very fine scales \( m \), almost all of the contribution to \( ||\mu_{x}^{m}||_{q}^{q} \) comes from \( \approx 2T^*(\alpha)m \) intervals, each of mass \( \approx 2^{-am^m} \).

**Lemma 4.5** (Lemma 4.11 in [Sh]). Suppose \( \alpha = T'(q) \) exists and fix \( \sigma > 0 \). Then there exists \( \epsilon = \epsilon(\sigma,q) > 0 \) such that if \( m \) is sufficiently large in terms of \( \epsilon,\sigma,q \), we have
\begin{equation}
(4.18) \quad \sum_{I \in \mathcal{D}_m : \mu_x(I) \geq 2^{-am^m+\epsilon m}} \mu_x(I)^q \leq 2^{-T(q)m-\epsilon m}
\end{equation}
for all \( x \in X \).

*Proof.* Let \( h > 0 \) be small enough so that \( T(q + h) \approx T(q) + \alpha h \). Now let \( \delta > 0 \) be a very small parameter which will be chosen at the end. For \( j \in \mathbb{N} \) define
\begin{equation}
(4.19) \quad S_j := \{ I \in \mathcal{D}_m : \mu_x(I) \in [2^{-am^m+\delta m},2^{-am^m+(j+1)\delta m}) \}.
\end{equation}
We estimate
\begin{equation}
(4.20) \quad |S_j|2^{-am(q+h)+j\delta m(q+h)} \leq \sum_{I \in S_j} \mu_x(I)^q \leq \sum_{I \in \mathcal{D}_m} \mu_x(I)^q \leq 2^{-T(q+h)m} \approx 2^{-T(q)m-\alpha h m},
\end{equation}
from which we deduce that
\begin{equation}
(4.21) \quad |S_j| \leq 2^{-T(q)m+\alpha m q-j \delta m(q+h)}.
\end{equation}
We now estimate the desired sum by
\begin{equation}
(4.22) \quad \sum_{I \in \mathcal{D}_m : \mu_x(I) \geq 2^{-am^m+\epsilon m}} \mu_x(I)^q = \sum_{j \geq \frac{\epsilon}{\delta}} \sum_{I \in S_j} \mu_x(I)^q \leq \sum_{j \geq \frac{\epsilon}{\delta}} |S_j|2^{-am^m+(j+1)\delta m q}
\end{equation}
\begin{equation}
(4.23) \quad \leq \sum_{j \geq \frac{\epsilon}{\delta}} 2^{-T(q)m+\alpha m q-j \delta m(q+h)-am+\delta m(q+h)} \leq \sum_{j \geq \frac{\epsilon}{\delta}} (2^{-\delta m h})^j
\end{equation}
\begin{equation}
(4.24) \quad \approx 2^{-T(q)m+(\delta q-\sigma h) m}
\end{equation}
where in the last line we used the elementary fact that \( \sum_{j \geq M} r^j \approx r^M \) if \( 0 < r < 1 \). Now picking \( \delta < \sigma h/q \) establishes the desired result.

\[ \square \]

**Lemma 4.6** (Lemma 4.12 in [Sh]). Suppose \( \alpha = T'(q) \) exists and fix \( \kappa > 0 \). Then there exists \( \epsilon = \epsilon(\kappa,q) > 0 \) such that if \( m \) is sufficiently large in terms of \( \epsilon,\kappa,q \), the following holds.

If \( \mathcal{D}' \) is any subcollection of \( \mathcal{D}_m \) with \( |\mathcal{D}'| \leq 2^{T^*(\alpha)m-\epsilon m} \), then
\begin{equation}
(4.26) \quad \sum_{I \in \mathcal{D}'} \mu_x(I)^q \leq 2^{-T(q)m-\epsilon m}
\end{equation}
for all \( x \in X \).
Proof. By Lemma 4.5, those \( I \in \mathcal{D}' \) with \( \mu_x(I) \gg 2^{-\alpha m} \) will contribute \( \ll 2^{-T(q)m} \) to the sum \( \sum_{I \in \mathcal{D}'} \mu_x(I)^q \). So it’s enough to bound
\[
\sum_{I \in \mathcal{D}' : \mu_x(I) \leq 2^{-\alpha m}} \mu_x(I)^q \ll |\mathcal{D}'| 2^{-\alpha q m} \leq 2^{T'(\alpha) m - \alpha q m} = 2^{-T(q) m - km},
\]
as desired (in the last equality we used Lemma 4.3). \( \square \)

Lemma 4.7 (Proposition 4.13 in [Sh]). Suppose \( \alpha = T'(q) \) exists.

1. Let \( \kappa > 0 \). Then there exists \( \epsilon = \epsilon(\kappa,q) > 0 \) such that if \( m \) is sufficiently large in terms of \( \epsilon, \kappa, q \), the following holds.

   Pick any starting scale \( s \in \mathbb{N} \) and any \( I \in \mathcal{D}_s \). If \( \mathcal{D}' \) is any subcollection of \( \mathcal{D}_{s+m}(I) \) with \( |\mathcal{D}'| \leq 2^{T'(\alpha)m - km} \), then
\[
\sum_{J \in \mathcal{D}'} \mu_x(J)^q \leq 2^{-T(q)m - cm} \mu_x(3I)^q
\]
for all \( x \in X \).

2. Let \( \delta > 0 \). Then if \( m \) is sufficiently large in terms of \( \delta, q \), the following holds.

   Pick any starting scale \( s \in \mathbb{N} \) and any \( I \in \mathcal{D}_s \). Then
\[
\sum_{J \in \mathcal{D}_{s+m}(I)} \mu_x(J)^q \leq 2^{-T(q)m + \delta m} \mu_x(3I)^q
\]
for all \( x \in X \).

Remark 4.8. Part (1) of this lemma should be thought of as a “multi-scale” version of Lemma 4.6. Part (2) should be thought of as a multi-scale version of (4.11).

Proof. (1) Let \( s \in \mathbb{N} \), \( I \in \mathcal{D}_s \), and \( \mathcal{D}' \subseteq \mathcal{D}_{s+m}(I) \) be as in the statement of the lemma. Pick \( n \) suitably so that \( \lambda^n \approx 2^{-s} \). Enumerate \( \mu_{x,n} \) as \( \sum_j p_j \delta_{y_j} \). Then using the self-similarity relation \( \mu_x = \mu_{x,n} * S_{\lambda^n} \mu_{T^n x} \), we have
\[
\sum_{J \in \mathcal{D}'} \mu_x(J)^q = \sum_{J \in \mathcal{D}'} \left( \sum_j (p_j \delta_{y_j} * \mu_{T^n x})(J) \right)^q = \sum_{J \in \mathcal{D}'} \left( \sum_j (p_j \mu_{T^n x}(\lambda^{-n}(J - y_j))) \right)^q.
\]
At this point, we note that since \( \mu_{T^n x} \) is supported on \([0,1]\), the only terms in the inner sum which contribute are those \( j \) for which \( y_j \) is within \( \lambda^n \) of the original starting interval \( I \) (since the outer sum only sums over intervals \( J \subseteq I \)). Let \( \mathcal{J} \) be the set of all such \( j \) and let \( p := \sum_{j \in \mathcal{J}} p_j \). Then applying Jensen’s inequality to the above equation (using the convexity of \( t \mapsto t^q \)), we get
\[
\sum_{J \in \mathcal{D}'} \mu_x(J)^q \leq \sum_{J \in \mathcal{D}'} p^{q-1} \sum_{j \in \mathcal{J}} p_j (\mu_{T^n x}(\lambda^{-n}(J - y_j)))^q = p^{q-1} \sum_{j \in \mathcal{J}} p_j \sum_{J \in \mathcal{D}'} \mu_{T^n x}(\lambda^{-n}(J - y_j))^q.
\]
Now since \( \lambda^n \) was chosen to be comparable to \( 2^{-s} \), and each \( J \in \mathcal{D}' \) has length \( 2^{-s-m} \), for each fixed \( j \) the collection of intervals \( \{\lambda^{-n}(J - y_j) : J \in \mathcal{D}'\} \) is comparable to a subset of \( \mathcal{D}_m \) with size \( |\mathcal{D}'| \ll 2^{T'(\alpha)m} \), and so by Lemma 4.6 we have the bound
\[
\sum_{J \in \mathcal{D}'} \mu_x(J)^q \ll p^{q-1} \sum_{j \in \mathcal{J}} p_j 2^{-T(q)m} = p^n \cdot 2^{-T(q)m}.
\]
Finally, we estimate
\[
\mu_x(3I) = \sum_j p_j (\delta_{y_j} * S_{\lambda^n} \mu_{T^n x})(3I) = \sum_j p_j \mu_{T^n x}(\lambda^{-n}(3I - y_j)) \geq \sum_{j \in \mathcal{J}} p_j = p
\]
where the inequality is true because by definition of \( \mathcal{J} \) and choice of \( n \) so that \( \lambda^n \approx 2^{-s} = \) length of \( I, 3I - y_j \supseteq [0,1] \) for all \( j \in \mathcal{J} \), and \( \mu_{T^n x} \) is supported on \([0,1]\). Substituting this estimate into (4.32) yields the desired result.
(2) Repeat the same proof as part (1), except at the beginning sum over all $J \in \mathcal{D}_{s+m}(I)$ instead of $J \in \mathcal{D}'$, and when we reach (4.32) we apply equation (4.11) instead of Lemma 4.6.

The other key thing we need to prove Theorem 4.1 is an inverse theorem for decay of $L^q$ norms under convolution.

We state the full theorem here but omit the proof and refer the reader to [Sh].

**Theorem 4.9** (Theorem 2.1 in [Sh]). Fix $q > 1$, $\delta > 0$, and $D_0 \in \mathbb{N}$. Then there exist $\epsilon > 0$ and $D \geq D_0$ such that if $\ell$ is sufficiently large, then the following holds.

Let $m = \ell D$ and suppose $\mu, \nu$ are measures supported on $2^{-m} \mathbb{Z}$ such that

\[
\|\mu \ast \nu\|_q \geq 2^{-\epsilon m} \|\mu\|_q.
\]

Then, up to appropriate translations of the measures $\mu$, and $\nu$, there exist sets $A \subseteq \text{supp}(\mu)$, $B \subseteq \text{supp}(\nu)$ satisfying the following:

1. $\|\mu\|_q \geq 2^{-\delta m}$, i.e. most of the $L^q$ norm of $\mu$ lives on $A$.
2. $\mu(x) \leq \mu(x)$ for all $x, y \in A$, i.e. $\mu$ is roughly uniform on $A$.
3. There is a sequence $R'_\ell, 0 \leq s \leq \ell - 1$, such that for every $I \in \mathcal{D}_{sD}(A)$, $|\mathcal{D}_{(s+1)D}(A \cap I)| = R'_\ell$, i.e. $A$ has regular branching in steps of size $2^D$ up to scale $2^{\ell D}$.
4. For every $x \in A$, $s \leq \ell - 1$, $x \in \frac{1}{2} \mathcal{D}_{sD}(x)$, i.e. no element of $A$ is too close to the edge of any dyadic intervals.
5. $\nu(B) \geq 2^{-\delta m}$.
6. $\nu(y) \leq 2\nu(x)$ for all $x, y \in B$.
7. There is a sequence $R''_\ell, 0 \leq s \leq \ell - 1$, such that for every $I \in \mathcal{D}_{sD}(B)$, $|\mathcal{D}_{(s+1)D}(B \cap I)| = R''_\ell$.
8. For every $y \in B$, $s \leq \ell - 1$, $y \in \frac{1}{2} \mathcal{D}_{sD}(y)$.

(5) For each $0 \leq s \leq \ell - 1$, either $R''_\ell = 1$ or $R''_\ell \geq 2^{(1-\delta)D}$.

The number $N$ of $0 \leq s \leq \ell - 1$ for which $R''_\ell \geq 2^{(1-\delta)D}$ satisfies

\[
\frac{1}{D} \left( \log(\|\nu\|_{q}^{-q'}) - \delta m \right) \leq N \leq \frac{1}{D} \left( \log(\|\mu\|_{q}^{-q'}) + \delta m \right).
\]

**Remark 4.10.** What this theorem says roughly is that if the norm of $\mu \ast \nu$ is not much smaller than the norm of $\mu$, then $\mu$ and $\nu$ are mostly supported on sets $A$ and $B$ respectively, where $A$ and $B$ have “dual” branching structures in the sense that at every scale, either $A$ has almost full branching or $B$ has no branching. In this sense, this theorem can be viewed as an $L^q$ norm analog of Hochman’s inverse theorem for growth of entropy under convolutions (Theorem 2.7 of [Ho2]).

**Remark 4.11.** For our application to the proof of Theorem 4.1, we will not need the full strength of the theorem. In particular, we will not use the set $B$ at all.

We now have enough tools to sketch the proof of Theorem 4.1.

**Sketch of proof of Theorem 4.1.** Fix an arbitrary $\alpha > 0$. Let $\nu$ be a measure supported on $2^{-m} \mathbb{Z}$ such that $\|\nu\|_q \leq 2^{-\sigma m}$. Note that our goal is to show that $\|\nu \ast \mu_x\|_q$ is exponentially smaller than $2^{-T(q)m}$, not $\|\mu_x\|_q^q$ itself. We know from (4.11) that $\|\mu_x\|_q^q \leq 2^{-T(q)m}$ for all $x$, but a priori there could be some $x \in X$ for which $\|\mu_x\|_q^q$ is far smaller. However if this is the case, we would already be done by Lemma 2.13, so we can assume from now on that $\|\mu_x\|_q^q \approx 2^{-T(q)m}$.

Now apply Lemma 4.7 to get a threshold scale $D_0$ such that for any $s \in \mathbb{N}$, any $I \in \mathcal{D}_s$, and any $D \geq D_0$, the following hold:

1. If $\mathcal{D}' \subseteq \mathcal{D}_{s+D}(I)$ has $\ll 2^{T^*(\alpha)D}$ elements, then

\[
\sum_{J \in \mathcal{D}'} \mu_x(J)^q \ll 2^{-T(q)D} \mu_x(3I)^q.
\]

2. If $\mathcal{D}_s \subseteq \mathcal{D}_{s+D}(I)$, then

\[
\sum_{J \in \mathcal{D}_{s+D}(I)} \mu_x(J)^q \lesssim 2^{-T(q)D} \mu_x(3I)^q.
\]
Next, apply Theorem 4.9 to obtain a scale $D \geq D_0$ and a small $\epsilon > 0$ as in the statement of the theorem. Then $\epsilon := \epsilon'/q$ will be the $\epsilon$ claimed in the statement of Theorem 4.1. From this point on we will also assume that $m = \ell D$ where $\ell$ is the parameter we allow to be sufficiently large. Suppose for contradiction that $\|\nu * \mu^m||_q \geq 2^{-em}2^{-T(q)m} \approx 2^{-\epsilon'm}||\mu^m||_q$. Then the inverse theorem (Theorem 4.9) applies and we obtain a set $A \subseteq 2^{-m}\mathbb{Z}$ (from now on we will slightly abuse notation and identify $A$ with a subset of $\mathcal{D}_m$ in the obvious way) satisfying the following:

(a) $||\mu^m_x||_q \approx ||\mu^m_x||_q^q \approx 2^{-T(q)m}$.
(b) $\mu_x(I) \approx \mu_x(J)$ for all $I, J \in A$.
(c) For $0 \leq s \leq \ell - 1$, $A$ has regular branching from scale $2^{-sD}$ to $2^{-(s+1)D}$ with branching number $R'_s$, i.e. for every $I \in \mathcal{D}_s(D(A))$, $|\mathcal{D}_{s+1}(A \cap I)| = R'_s$.
(d) The number $N$ of $0 \leq s \leq \ell - 1$ for which $R'_s \approx 2^D$ satisfies

$$\frac{T(\nu)^{\ell}}{q-1} \approx \frac{1}{D} \log ||\mu^m_x||_{q'} \geq N \geq \frac{1}{D} \log ||\nu||_{q'} \geq \sigma \ell.$$  

The first step is to get an upper bound on the size of $A$. This is heuristically easy because we know that $\mu_x$ is roughly uniform on $A$ and we also know roughly the size of the $L^q$ norm of $\mu^m_x|_A$. The fact that $T$ is differentiable at $q$ will also provide us with some extra quantitative information about the size of $\mu_x(I)$ for $I \in A$, leading to a good bound on $|A|$. Let $\alpha = T'(q)$. Pick a suitable $\beta$ so that $\mu_x(I) \approx 2^{-\beta m}$ for all $I \in A$. Now if $h$ is very small, then by the differentiability of $T$ we have $T(q+h) \approx T(q) + \alpha h$, so the following two things hold.

$$|A| \cdot 2^{-\beta m(q+h)} \approx \sum_{I \in A} \mu_x(I)^{q+h} \approx \|\mu^m_x|_A\|_{q+h}^{q+h} \approx 2^{-T(q+h)m} \approx 2^{-T(q)m - \alpha h m}$$

$$|A| \cdot 2^{-\beta m q} \approx \sum_{I \in A} \mu_x(I)^q = \|\mu^m_x|_A\|_q^q \approx 2^{-T(q)m}$$

Rearranging both equations to isolate $|A|$ and then cancelling terms, we obtain $\alpha \approx \beta$. Then substituting this back into either of the above equations and rearranging yields $|A| \approx 2^{(\alpha q - T(q)m)} = 2^{T(\alpha)m}$ by Lemma 4.3. This result is another manifestation of the heuristic principle that motivated Lemmas 4.5–4.7.

Since $A$ has regular branching, we know that $|A| = \prod_{s=0}^{\ell-1} R'_s$. We have just shown that $|A| \approx 2^{T(\alpha)D\ell}$, which means that “on average” the branching numbers $R'_s$ are around $2^{T(\alpha)D}$. But by Lemma 4.4 we know that $T'(\alpha) < 1$, and by conclusion (d) of the inverse theorem we know that $A$ has approximately full branching ($R'_s \approx 2^D$) for a positive proportion of steps $0 \leq s \leq \ell - 1$. This implies that we must also have $R'_s \ll 2^{T(\alpha)D}$ for a positive proportion of steps $s$ as follows. Let $S$ be the set of $0 \leq s \leq \ell - 1$ for which $R'_s \approx 2^D$. We have

$$2^{T(\alpha)D\ell} \approx |A| = \prod_{s=0}^{\ell-1} R'_s \gtrsim 2^{D|S|} \prod_{s \in S^c} R'_s,$$

so rearranging gives

$$\prod_{s \in S^c} R'_s \lesssim 2^{T(\alpha)D\ell - D|S|} = 2^{T(\alpha)D(|S| + |S^c|) - D|S|} = 2^{D(|S| - T(\alpha)) + D|S^c| T(\alpha)}.$$  

This implies that

$$\sum_{s \in S^c} \log R'_s \lesssim D|S|(-1 + T(\alpha)) + D|S^c| T(\alpha)$$

$$\lesssim \sigma D(-1 + T(\alpha)) + D T(\alpha)|S^c|$$

$$\leq \sigma D|S^c|(-1 + T(\alpha)) + D T(\alpha)|S^c|$$

$$= (T(\alpha) - (1 - T(\alpha))\sigma D|S^c| =: (T(\alpha) - 2\sigma)D|S^c|.$$


where in the second line we used the right inequality from condition (d) (recall that $T^*(\alpha) < 1$). This inequality says that the average of $\log R'_s$ over $s \in S^c$ is strictly less than $T^*(\alpha) D$, so by (essentially) Markov’s inequality we will be able to find a positive proportion of scales $s$ on which $\log R'_s$ is bounded below $T^*(\alpha) D$.

Let $S_1 = \{s \in S^c : \log R'_s \leq (T^*(\alpha) - \kappa) D\}$. We estimate

$$|S^c \setminus S_1| (T^*(\alpha) - \kappa) D \leq \sum_{s \in S^c \setminus S_1} \log R'_s \leq \sum_{s \in S^c} \log R'_s \lesssim (T^*(\alpha) - 2\kappa) D |S^c|,$$

and rearranging gives

$$|S^c \setminus S_1| \lesssim \frac{T^*(\alpha) - 2\kappa}{T^*(\alpha) - \kappa} |S^c|,$$

so combined with the left inequality from condition (d), we get

$$|S_1| \geq |S^c| - \frac{T^*(\alpha) - 2\kappa}{T^*(\alpha) - \kappa} |S^c| \gtrsim \left(1 - \frac{T(q)}{q - 1}\right) \left(1 - \frac{T^*(\alpha) - 2\kappa}{T^*(\alpha) - \kappa}\right) \ell =: \gamma \ell$$

where $\gamma > 0$ (recall our standing assumption that $T(q) < q - 1$) is a constant depending ultimately only on $\sigma$ and $q$, as desired.

Finally, we will obtain a contradiction from the fact that $A$ has slower-than-average branching a positive proportion of the time. Heuristically, this is due to Lemma 4.7 – every time $A$ has slow branching, the lemma gives us a strong upper bound on the $L^q$ norm at that scale. But by condition (a) of the inverse theorem, the $L^1$ norm at the final scale $2^{-TqD}$ needs to be comparable to $2^{-T(q)\ell D}$, and we will see that if $A$ has too many scales with slow branching, Lemma 4.7 won’t allow the final $L^q$ norm to be big enough.

For $0 \leq s \leq \ell$, define

$$c_s := \sum_{I \in D_{s+1}D(A)} \mu_x(I)^q.$$

For those scales $s$ for which $R'_s \ll 2^{T^*(\alpha)D}$, (4.36) guarantees that

$$c_{s+1} = \sum_{I \in D_{s+1}D(A)} \mu_x(I)^q = \sum_{I \in D_{s+1}D(A)} \sum_{J \in D_{s+1}D(A \setminus I)} \mu_x(J)^q \ll \sum_{I \in D_{s}D(A)} 2^{-T(q)D} \mu_x(3I)^q \approx 2^{-T(q)D} c_s,$$

and for all other scales $s$, (4.37) guarantees by the same calculation that

$$c_{s+1} \lesssim 2^{-T(q)D} c_s.$$

We also know that we must have $c_1 = \sum_{I \in D_1D(A)} \mu_x(I)^q = ||\mu_x^m||_q^q \approx 2^{-T(q)\ell D}$ by condition (a) given by the inverse theorem. To summarize: at each step, $c_s$ decays at least as fast as by a factor of $2^{-T(q)\ell D}$, and by the final (7th) step it should be no smaller than $2^{-T(q)\ell D}$. This is only possible if it decays by exactly $2^{-T(q)\ell D}$ at each step, but we have shown by (4.51) that whenever $A$ has slower-than-average branching, which is a positive proportion of the time, $c_s$ decays much faster than $2^{-T(q)\ell D}$. This is a contradiction. $\square$

4.4. **Finishing the proof.** Theorem 4.1 has the following important consequence which will allow us to complete the proof of Theorem 2.22.

**Theorem 4.12** (Proposition 5.2 in [Sh]). Fix $q > 1$ such that $T$ is differentiable at $q$ and $T(q) < q - 1$. Suppose $x \in X$ is such that

$$\lim_{m \to \infty} \frac{1}{m} \log ||\mu_x^m||_q^q = T(q).$$

Let $m(n) = \lceil n \log(1/\lambda) \rceil$ be the integer such that $2^{-m(n)} \leq \lambda^n < 2 \cdot 2^{-m(n)}$. Then for any $R \in \mathbb{N}$, we have

$$\lim_{n \to \infty} \frac{1}{n \log \lambda} \log ||\mu_{Rm(n)}^R||_q^q = T(q).$$

2Note that it may be possible that $T^*(\alpha) - 2\kappa < 0$ because both $q$ and $\sigma$ are allowed to be arbitrary. When this proof is done with more precise bookkeeping, this issue doesn’t arise. Alternatively, we can assume $\sigma$ is sufficiently small in terms of $q$ so that this number is positive, at the cost of proving a slightly weaker statement than what Theorem 4.1 claims. However, for our application of this theorem in the next section, allowing $\sigma$ to depend on $q$ doesn’t present any problems.
This says roughly that the $L^q$ norm of $\mu_{x,n}^m$ doesn’t change as it is discretized over finer and finer scales $m$. This will be useful because by the exponential separation assumption, once the scale $m$ is fine enough, $\mu_{x,n}^m$ is the same as the undiscretized $\mu_{x,n}$.

**Proof of Theorem 4.12.** First, by Lemmas 2.13 and 2.19 we have
\[
\left\| \mu_{x,n}^m \right\|^q_q \leq \left\| \mu_{x,n}^m \right\|^q_q \approx \left\| \mu_{x,n}^m \right\|^q_q \approx 2^{-T(q)m},
\]
so we just need to establish the reverse inequality.

We now apply Lemmas 2.13 and 2.19 several times and perform the following manipulations.

\[
(4.54) \quad \left\| \mu_{x,n}^m \right\|^q_q = \left\| \left( \mu_{x,n}^m \right) \left( S_{\lambda^m} \mu_{T^m}^n \right) \right\|^q_q = \left\| \sum_{I \in D_m} \mu_{x,n}(I) \left( \mu_{x,n}^m \right) \left( S_{\lambda^m} \mu_{T^m}^n \right) \right\|^q_q
\]

\[
(4.55) \quad \lesssim \sum_{I \in D_m} \mu_{x,n}(I)^q \left\| \left( \mu_{x,n}^m \right) \left( S_{\lambda^m} \mu_{T^m}^n \right) \right\|^q_q
\]

\[
(4.56) \quad = \sum_{I \in D_m} \mu_{x,n}(I)^q \left\| S_{\lambda^m} \left( \mu_{x,n}^m \right) \left( S_{\lambda^m} \mu_{T^m}^n \right) \right\|^q_q
\]

\[
(4.57) \quad = \sum_{I \in D_m} \mu_{x,n}(I)^q \left\| S_{\lambda^m} \left( \mu_{x,n}^m \right) \left( \mu_{x,n}^m \right) \right\|^q_q
\]

where $\rho_{x,I} := S_{\lambda^m}(\mu_{x,n})_I$.

Fix a small parameter $\sigma > 0$ and set $D' = \{ I \in D_m : \left\| \mu_{x,n}^m \right\|^q_q \leq 2^{-\sigma m} \}$. Now in the above sum, for $I \in D'$ Theorem 4.1 says that $\left\| \mu_{x,n}^m \right\|^q_q \leq 2^{-T(q)m}$, and for all other $I$, Lemma 2.13 gives $\left\| \mu_{x,n}^m \right\|^q_q \approx \left\| \mu_{x,n}^m \right\|^q_q \approx 2^{-T(q)m}$. By the hypothesis on $x$, we thus have

\[
(4.59) \quad 2^{-T(q)(R+1)m} \approx \left\| \mu_{x,n}^m \right\|^q_q \lesssim (\text{something } \lesssim 2^{-T(q)m}) \sum_{I \in D'} \mu_{x,n}(I)^q + (\text{something } \lesssim 2^{-T(q)m}) \sum_{I \notin D'} \mu_{x,n}(I)^q
\]

\[
(4.60) \quad \lesssim (\text{something } \lesssim 2^{-T(q)m}) \left\| \mu_{x,n}^m \right\|^q_q + (\text{something } \lesssim 2^{-T(q)m}) \sum_{I \notin D'} \mu_{x,n}(I)^q
\]

\[
(4.61) \quad \lesssim (\text{something } \lesssim 2^{-T(q)(R+1)m}) + (\text{something } \lesssim 2^{-T(q)m}) \sum_{I \notin D'} \mu_{x,n}(I)^q.
\]

By rearranging this implies that

\[
(4.62) \quad \sum_{I \notin D'} \mu_{x,n}(I)^q \gtrsim 2^{-T(q)m}.
\]

By another series of manipulations using Lemmas 2.13 and 2.19, we also estimate

\[
(4.63) \quad \left\| \mu_{x,n}^m \right\|^q_q = \left\| \mu_{x,n}^m \right\|^q_q = \left\| \sum_{I \in D_m} \mu_{x,n}(I) \left( \mu_{x,n}^m \right) \left( S_{\lambda^m} \mu_{T^m}^n \right) \right\|^q_q \lesssim \left\| \mu_{x,n}^m \right\|^q_q \lesssim \left\| \mu_{x,n}^m \right\|^q_q \approx \left\| \mu_{x,n}^m \right\|^q_q
\]

\[
(4.64) \quad = \sum_{I \in D_m} \mu_{x,n}(I)^q \left\| S_{\lambda^m} \left( \mu_{x,n}^m \right) \right\|^q_q = \sum_{I \in D_m} \mu_{x,n}(I)^q \left\| \rho_{x,I}^m \right\|^q_q
\]

Restricting to only the intervals outside $D'$ and using the definition of $D'$ as well as (4.62), we get

\[
(4.65) \quad \left\| \mu_{x,n}^m \right\|^q_q \gtrsim \sum_{I \notin D'} \mu_{x,n}(I)^q \left\| \rho_{x,I}^m \right\|^q_q \gtrsim 2^{-\sigma m} \sum_{I \notin D'} \mu_{x,n}(I)^q \gtrsim 2^{-\sigma m} 2^{-T(q)m}.
\]

Since $\sigma > 0$ was arbitrary, letting it tend to 0 gives the desired inequality. \qed
We now complete the proof of Theorem 2.22. Recall that we fix \( q > 1 \) such that \( T \) is differentiable at \( q \) and \( T(q) < q - 1 \), and our goal is to show

\[
T(q) = \frac{1}{\log \lambda} \int_X \log ||\Delta(x)||_q^q \, d\mathbb{P}(x).
\]

By (4.10), \( \mathbb{P} \)-almost every \( x \) satisfies

\[
\lim_{m \to \infty} -\frac{1}{m} \log ||\mu^m_x||_q^q = T(q).
\]

By the exponential separation assumption, there is some \( R \in \mathbb{N} \) such that \( \mathbb{P} \)-almost every \( x \) has the property that the atoms of \( \mu_{x,n} \) are \( \lambda^R \)-separated. Fix an \( x \in X \) for which both of the above hold. Then, when \( n \) is large, we have

\[
T(q) \approx -\frac{1}{m(n)} \log ||\mu^{m(n)}_x||_q^q \approx -\frac{1}{m(n)} \log ||\mu^{m(n)}_{x,n}||_q^q \approx -\frac{1}{m(n)} \log ||\mu^{Rm(n)}_{x,n}||_q^q \approx \frac{1}{n \log \lambda} \log ||\mu_{x,n}||_q^q
\]

where the second equality is by Lemma 2.19, the third is by Theorem 4.12, and the fourth is by exponential separation and Lemma 2.13. Finally, again since the atoms of \( \mu_{x,n} \) are distinct, the rightmost term above is equal to

\[
\frac{1}{n \log \lambda} \log \prod_{i=0}^{n-1} ||\Delta(T^i x)||_q^q,
\]

and by unique ergodicity this converges as \( n \to \infty \) to

\[
\frac{1}{\log \lambda} \int_X \log ||\Delta(x)||_q^q \, d\mathbb{P}(x),
\]

as desired. \( \Box \)

References


