FROM INVARIANCE TO SELF-SIMILARITY:
THE WORK OF MICHAEL HOCHMAN ON FRACTAL DIMENSION
AND ITS AFTERMATH

HILLEL FURSTENBERG
(Communicated by Giovanni Forni)

Abstract. M. Hochman’s work on the dimension of self-similar sets has given impetus to resolving other questions regarding fractal dimension. We describe Hochman’s approach and its influence on the subsequent resolution by P. Shmerkin of the conjecture on the dimension of the intersection of \( \times p \)- and \( \times q \)-Cantor sets.

In the distant background of M. Hochman’s work on self-similar sets [3] is a particular example of “multiparameter rigidity” appearing in [2]. The space in question is the 1-torus \( \mathbb{T} = \mathbb{R}/\mathbb{Z} \), and the action is generated by two endomorphisms, \( T_p \) and \( T_q \), where for \( n \in \mathbb{N}, T_n x = nx \mod 1 \), and \( p, q \) are “multiplicatively independent”, i.e., \( p^k \neq q^l \) for all \( k, l \). (In our discussion, \( p, q \) will always satisfy this condition.) The rigidity claim proved in [2] is that the only closed subsets of \( \mathbb{T} \) invariant under both \( T_p \) and \( T_q \) are certain finite sets of rationals and \( \mathbb{T} \) itself. This implies that for irrational \( \alpha \in \mathbb{T} \), the orbit \( \{ p^m q^n \alpha ; n, m \geq 0 \} \) is dense in \( \mathbb{T} \). This suggests that the individual orbit closures, \( \{ p^m \alpha \} \) and \( \{ q^n \alpha \} \), cannot both be small. One is led to the following quantitative formulation in [1]:

Conjecture 1. For every irrational \( \alpha \in \mathbb{T} \),

\[
\dim \{ p^m \alpha \} + \dim \{ q^n \alpha \} \geq 1. \tag{1}
\]

Here \( \dim \) refers to Hausdorff dimension. An equivalent formulation, more in the spirit of the subsequent discussion is:

Conjecture 2. If \( A \) and \( B \) are two closed subsets of \( \mathbb{T} \) with \( T_p A \subset A, T_q B \subset B \) such that \( \dim A + \dim B < 1 \), then the intersection \( A \cap B \), if not empty, consists only of rationals.

The conclusion in this conjecture implies that under the given conditions \( A \cap B \) is either empty or countable. One now knows:

Theorem 1. If \( A \) and \( B \) are closed subsets of \( \mathbb{T} \), invariant under \( T_p \) and \( T_q \) respectively, then

\[
\dim(A \cap B) \leq \max(\dim A + \dim B - 1, 0). \tag{2}
\]
Here, the empty set is regarded as having negative dimension so that the inequality in (2) is valid when \( A \cap B = \emptyset \). If \( A \cap B \) is countable, then its dimension is 0, so (2) is also verified. When \( A = B \), then either \( \dim A = 0 \) or \( \dim A = 1 \), which for invariant sets implies \( A = \mathbb{T} \). This comes close to the statement in [2].

Theorem 1 was proved independently by P. Shmerkin ([4]) and M. Wu ([5]). Both in the proof of Theorem 1 and in the formulation of Conjecture 2, it suffices to assume that the sets \( A \) and \( B \) are “Cantor-like” sets. To be precise, we will call a subset \( S \subset \mathbb{T} \) a \( b \)-Cantor set if for some subset \( J \subset \{0, 1, \ldots, b-1\} \), \( S = \{ x = \sum_{n=1}^{\infty} a_n b^{-n} \text{ with all } a_n \in J \} \). Suppressing the base \( b \) we call a set of this type a Cantor set. One can show that if \( A \subset \mathbb{T} \) is a closed \( T_p \)-invariant set, and \( \varepsilon > 0 \), there exists a \( p \)-Cantor set \( A' \supset A \) with \( \dim A' < \dim A + \varepsilon \). Using this, it follows that if (2) holds for Cantor sets, it holds for arbitrary invariant sets. Also using the fact that the family of all Cantor sets is countable, it follows from Theorem 1 that \( \bigcup (A \cap B) \cap \dim A + \dim B < 1, T_p A \subset A, T_q B \subset B \) has \( \dim 0 \). We obtain the following result in the spirit of Conjecture 1:

**Theorem 2.** The set \( \{ \alpha \in \mathbb{T}; \dim \overline{\{p_\alpha^n\}} + \dim \{q_\alpha^n\} < 1 \} \) has dimension 0.

Cantor sets are a special case of invariant sets, but they are also examples of self-similar sets which need not be invariant in the usual sense. It is in this more general setting that Shmerkin and Wu give proofs of Theorem 1. Also Hochman’s result in [3] is demonstrated in this setting. We proceed to define self-similarity for subsets of the line. To define such a set we suppose given a finite family of contracting similarity maps \( \Phi = \{ \phi_i \}_{i=1}^k \) when \( \phi_i(x) = \lambda_i x + \alpha_i, |\lambda_i| < 1, \alpha_i \in \mathbb{R} \). It is not hard to show that given \( \Phi \), there is a unique compact set \( X \subset \mathbb{R} \) satisfying

\[
X = \bigcup_{i=1}^k \phi_i(X).
\]

(3)

\( X \) is called the attractor of the “iterated function system” determined by \( \Phi \). (The expression “iterated function system”, or IFS, derives from the fact that the points of \( X \) are limits of sequences \( \phi_{i_1} \phi_{i_2} \cdots \phi_{i_n} (x_0), n \rightarrow \infty \), for any fixed \( x_0 \).) When the \( \lambda_i \) appearing in the definition of the \( \phi_i \) are all equal to \( b^{-1} \) for an integer \( b \), and the \( \alpha_i \in J \), a subset of \( \{0, 1, \ldots, b-1\} \), then the corresponding attractor is exactly the \( b \)-Cantor set determined by \( J \).

**Definition 3.** A set \( S \subset \mathbb{R} \) is called self-similar if it is the attractor of an IFS: \( \Phi = \{ \phi_i \}_{i=1}^k \) with \( \phi_i(x) = \lambda_i x + \alpha_i, |\lambda_i| < 1, 1 \leq i \leq k \). If \( \lambda_i = \lambda \) for all \( i \), then \( S \) is \( \lambda \)-self-similar.

The theorem proved by P. Shmerkin and M. Wu which implies Theorem 1 is the following:

**Theorem 4.** Let \( A \subset \mathbb{R} \) be a \( \lambda \)-self-similar set, \( B \subset \mathbb{R} \) a \( \mu \)-self-similar set such that \( \log \lambda / \log \mu \) is irrational (\( \lambda, \mu \) multiplicatively independent), then

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

(4)
As to the dimension of a self-similar set determined by the data \( \{\lambda_i, \alpha_i, 1 \leq i \leq k\} \) there is no known algorithm covering all cases. An elementary result can be stated when the IFS satisfies a "separation condition".

**Theorem 5.** Let \( X \) be the attractor of \( \Phi = \{\varphi_i\}, \varphi_i(x) = \lambda_i x + \alpha_i, 1 \leq i \leq k, \) and assume that \( X \) is contained in an interval \([a, b]\) satisfying \( \varphi_i([a, b]) \cap \varphi_j([a, b]) = \emptyset \) for \( i \neq j \). Let \( \sigma(\Phi) = s \), where \( s \) is the unique solution to

\[
\sum_{i=1}^{k} \lambda_i^s = 1.
\]

Then \( \dim X = \sigma(\Phi) \).

Note that the condition in the theorem implies \( \sigma(\Phi) \leq 1 \). Without this separation condition the solution \( s \) to (5) may exceed 1. Since the dimension of a set in \( \mathbb{R} \) always has dimension \( \leq 1 \), it will be convenient to define more generally

**Definition 6.** For any IFS \( \Phi = \{\varphi_i\}, \varphi_i(x) = \lambda_i x + \alpha_i, \) we set

\[
\sigma(\Phi) = \min(s, 1)
\]

where \( s \) is the unique solution to (5).

In [3], Hochman gives a weaker condition on \( \Phi \) for which the general statement \( \dim X = \sigma(\Phi) \) is valid. This condition is called "exponential separation". To formulate this condition we define the sets of similarity maps

\[
\Phi_n = \{\varphi_{i_1}\varphi_{i_2}\cdots\varphi_{i_n} \mid \text{each } \varphi_{i_m} \in \Phi\}.
\]

Writing \( \varphi_\alpha = \varphi_{i_1}\varphi_{i_2}\cdots\varphi_{i_n} \) with \( \alpha = i_1i_2\ldots i_n \) we set

\[
\Delta_n = \inf\{||\varphi_\alpha(0) - \varphi_\beta(0)||, \alpha \neq \beta\}.
\]

**Definition 7.** We say that \( \Phi \) satisfies **exponential separation** if \( \exists \delta > 0 \) and \( \Delta_n > \delta^n \) for all \( n \).

(This condition is slightly stronger than that given by Hochman but is adequate for our purposes).

**Theorem 8 ([3]).** If the system \( \Phi \) satisfies exponential separation, then the attractor \( X \) of \( \Phi \) satisfies

\[
\dim X = \sigma(\Phi).
\]

While the issue dealt with in Theorem 8 is quite unlike that of Theorem 4, Hochman's proof of Theorem 8 has had an impact on subsequent work on fractal dimension, and we will illustrate this by highlighting features of Hochman's argument which appear in Shmerkin's proof of Theorem 4. We identify three components of Hochman's proof which have parallels in Shmerkin's work:

i) Reformulating questions dealing with dimension of sets in terms of dimension of probability measures, replacing self-similar sets with self-similar measures,

ii) Using "Shannon entropy" of a probability measure to recover the dimension of a self-similar measure,
iii) Showing the smoothing effect of convolving one probability measure with another, which can be made precise in terms of entropy.

The result actually established by Hochman concerns self-similar measures and implies the corresponding result for self-similar sets. A self-similar measure is determined by an IFS \(\Phi = \{\varphi_i\}_{i=1}^k\) as before, together with a probability vector \(\{p_i\}_{i=1}^k, p_i > 0, \sum_i p_i = 1\). There is now a unique probability measure \(\mu\) on \(\mathbb{R}\) with

\[
\mu = \sum_{i=1}^k p_i \varphi_i \mu,
\]

where we denote by \(\varphi \mu\) the “pushforward measure” under \(\varphi, \varphi \mu(A) = \mu(\varphi^{-1} A)\). When all \(\lambda_i = \lambda\), we can give an explicit expression for \(\mu\) satisfying (8). We define the atomic measure

\[
\rho = \sum_{i=1}^k p_i \delta_{\alpha_i},
\]

and let the operator \(T_\gamma\) for \(\gamma \in \mathbb{R}\) act on measures correspond to the dilation \(T_\gamma x = \gamma x\). We then define \(\mu^{(n)}\) as the convolution

\[
\mu^{(n)} = \rho * T_\lambda \rho * T_{\lambda^2} \rho * \cdots * T_{\lambda^{n-1}} \rho.
\]

The limit \(\mu = \lim \mu^{(n)}\) satisfies \(\mu = \rho * T_\lambda \mu\) and this is equivalent to (8). Moreover we have for each \(n\),

\[
\mu = \mu^{(n)} * T_{\lambda^n} \mu
\]

an equality which will play an important role in the sequel. It is not hard to show that the support of the self-similar measure \(\mu\) is exactly the self-similar set \(X\) determined by \(\Phi\), so information regarding \(\mu\) will provide information about \(X\). For this we need the notion of dimension for a probability measure on \(\mathbb{R}\). The definition is

\[
\dim \mu = \inf \{\dim A, \mu(A) > 0\}.
\]

We will encounter two other definitions of dimension for a measure, and these all coincide for self-similar measures. Note that for any measure \(\nu\), \(\dim \nu \leq \dim(\text{support } \nu)\), so knowing the dimension of a self-similar measure gives a lower bound for the dimension of the self-similar set supporting it. To formulate Hochman’s result, we set for a system \(\Phi\) and vector \(\bar{p} = \{p_1, p_2, \ldots, p_k\}\)

\[
t = \frac{\sum p_i \log p_i}{\sum p_i \log \lambda_i}
\]

and define \(\tau(\Phi, \bar{p}) = \min(t, 1)\). We then have

**Theorem 9.** If the system \(\Phi\) satisfies exponential separation, then for any \(\bar{p}\), the self-similar measure \(\mu\) determined by \((\Phi, \bar{p})\) satisfies

\[
\dim \mu = \tau(\Phi, \bar{p}).
\]
One should observe that in Theorem 8 it is easy to prove the inequality $\dim X \leq \sigma(\Phi)$. This follows by considering the covering of $X$ by unions of intervals $[\varphi_i, \varphi_i, \ldots, \varphi_i, J]$ for each $n$, where $J$ is an interval containing $X$, and letting $n \to \infty$. Thus one is left to prove the opposite inequality. This will follow from Theorem 9 if we can choose $\bar{\mu}$ so that $\tau(\Phi, \bar{\mu}) = \sigma(\Phi)$. This can be done by setting $p_i = \lambda_i^2$ with $s$ given by (5). Thus Theorem 9 implies Theorem 8. We will give a proof of Theorem 9 in the special case: $\lambda_i = \lambda$ for all $i$, $p_i = k^{-1}$ for all $i$ and $k\lambda \geq 1$. The proof will illustrate the main features of Hochman’s proof of the general case. Our hypotheses imply that in (13), $t \geq 1$, so $\tau(\Phi, \bar{\mu}) = 1$, and so we will want to show that $\dim \mu = 1$ for the self-similar measure $\mu$. We should note that if (14) is shown to be valid for some self-similar measure with $\Phi$ as given, this will imply the equality in Theorem 8. So, as regards Theorem 8, the extra hypothesis $p_i = k^{-1}$ is harmless.

The main tool used by Hochman and his proof is the notion of entropy coming from Shannon’s information theory. We start with the definition:

**Definition 10.** If $(X, \mu)$ is a probability space and $\mathcal{P} : X = A_1 \cup A_2 \cup \cdots \cup A_r$ is a partition to disjoint subsets, we set

$$H(\mu, \mathcal{P}) = -\sum \mu(A_i) \log_2 \mu(A_i)$$

and call this the **Shannon entropy** of the measure with respect to the partition. (We take the value of $x \log x$ to be 0 for $x = 0$.)

Given a probability space $(X, \mu)$ and a subset $A \subset X$ with $\mu(A) > 0$, we can define the **conditional measure** $\mu_A$ supported on the set $A$ by

$$\mu_A(B) = \frac{\mu(B \cap A)}{\mu(A)}.$$  

We will say a partition $\mathcal{P}$ **refines** the partition $\mathcal{P}'$ if each set of $\mathcal{P}'$ is a union of sets in $\mathcal{P}$. In this case, for a set $A$ of $\mathcal{P}'$, the partition $\mathcal{P}$ restricts to a partition of $A$ and we can speak of the corresponding entropy $H(\mu_A, \mathcal{P}^A)$ where $\mathcal{P}^A$ represents the induced partition on $A$.

We have the following two lemmas:

**Lemma 11.** Let $\mathcal{P}$ and $\mathcal{P}'$ be two partitions on the space $X$ with $\mathcal{P}$ refining $\mathcal{P}'$. Then

$$H(\mu, \mathcal{P}) = H(\mu, \mathcal{P}') + \sum_{A \in \mathcal{P}^A} \mu(A) H(\mu_A, \mathcal{P}^A).$$  

**Lemma 12.** Let $\mathcal{P} : X = A_1 \cup A_2, \ldots, A_n$ be a partition of $X$, and $\mu$ a probability measure on $X$ with $H(\mu, \mathcal{P}) = h$. Fix $g > h$ and let $G$ be the set of indices $j$ for which $\mu(A_j) < n^{-\delta}$. Then

$$\mu\left( \bigcup_{j \in G} A_j \right) < h/\delta.$$  

We now turn to measures in $\mathbb{R}$ and make definitions for measures supported on the unit interval $I = [0,1]$, but analogous definition can be made for arbitrary intervals. Denote by $\mathcal{D}_n$ the partition of $[0,1]$ to $2^n$ equal intervals. (The number
We will find that we are led to a contradiction if \( \dim \) which for fixed \( m \) we set
\[
\lambda = \frac{1}{n} H(\mu, \mathcal{D}_n)
\]
then we can show

**Proposition 13.** If \( \mu \) is a self-similar measure, then \( \lim \mu(n) \) exists and
\[
(17) \quad \mu(n) \to \dim(\mu), \text{ as } n \to \infty.
\]

If \( \mu \) is an atomic measure; e.g., \( \mu = \sum_{i=1}^{H} p_i \delta_{x_i} \), then \( \dim \mu = 0 \). But \( \mu(n) \) need not vanish for small \( n \). However when \( 2^{-n} \) is less than the minimal distance between atoms, the entropy \( H(\mu, \mathcal{D}_n) = \sum_{i=1}^{H} p_i \log p_i \) which is independent of \( n \).

To prove Theorem 9 under our special hypotheses we note that since all \( \lambda_i \) are equal we have
\[
(11 \text{ bis}) \quad \mu = \mu^{(n)} * T_{\lambda^n} \mu.
\]

We will find that we are led to a contradiction if \( \dim T_{\lambda^n} \mu = \dim \mu < 1 \). Roughly speaking the contradiction comes from the fact that convolution has a smoothing effect on a measure which should increase the entropy. We can't apply this argument directly to (11) since the atomic measure \( \mu^{(n)} \) has dimension 0 and convolving with it has no effect on dimension. However it may have an effect on the approximating Shannon entropies.

To give a precise formulation we consider the “dyadic” subintervals of \([0,1]\):
\[
I_{m,j} = \left[ \frac{j}{2^m}, \frac{j+1}{2^m} \right] \quad 0 \leq j < 2^m
\]
which for fixed \( m \) constitute the partition \( \mathcal{D}_m \). We say that \( I_{m,j} \) is of **level** \( m \). Hochman’s smoothing theorem is the following:

**Theorem 14.** For all \( \epsilon > 0 \) and integers \( m \), there exist \( \eta = \eta(\epsilon, m) > 0 \) and \( n_0(\epsilon, m) \) such that if \( \mu \) is a probability measure on \([0,1]\) and \( n \geq n_0(\epsilon, m) \), and the proportion of dyadic intervals \( I \) of level \( n \) satisfying \( H_m(\mu | I) < 1 - \epsilon \) is greater than \( 1 - \epsilon \), then for any \( \nu \) with entropy \( H_n(\nu) > \epsilon \) we will have
\[
(18) \quad H_n(\mu | \nu) \geq H_n(\mu) + \eta.
\]

We proceed to the proof of Theorem 9. We are given \( \Delta_n > \delta^n \) for \( 0 < \delta < 1 \) and every \( n \). For each \( n \) let \( n' \) be the integer satisfying \( 2^{-n'} \leq \lambda^{-n} < 2^{n'+1} \), and choose \( q \) sufficiently large that \( 2^{-qn'} < \delta^n < \Delta_n \). \( \mu^{(n)} \) is an atomic measure whose atoms are separated by \( \Delta_n \) so by this choice of \( q \), no two atoms of \( \mu^{(n)} \) are in the same dyadic interval of \( \mathcal{D}_{qn'} \). We fix \( q \) and we let \( n \to \infty \) and so \( n' \to \infty \) as well. \( (n'/n) \) is bounded from above and below.) We compare the entropies \( H_{n'}(\mu) \) and \( H_{aqn'}(\mu) \). \( \mathcal{D}_{aqn'} \) refines \( \mathcal{D}_{n'} \), so by (15)
\[
(19) \quad H(\mu, \mathcal{D}_{aqn'}) = H(\mu, \mathcal{D}_{n'}) + \sum_{I \subset \mathcal{D}_n} \mu(I) H(\mu, \mathcal{D}_{aqn'}^I).
\]
By (11) this can be rewritten

\[ H(\mu, \mathcal{D}_{q^m}) = H(\mu, \mathcal{D}_{q^m}) + \sum_{I \in \mathcal{D}_{q^m}} \mu(I) H(\mu^{(n)}_I * T_{\lambda^n} \mu, \mathcal{D}_{q^m}^I). \]

We want to estimate \( H(\mu^{(n)}_I, \mathcal{D}_{q^m}^I). \) By (10), \( \mu^{(n)} \) is a combination of \( k^n \) atoms of equal weight. Inasmuch as \( \lambda^n \leq 2^{-n'} \), the convolution \( \mu^{(n)} * T_{\lambda^n} \mu = \mu \) is close enough to \( \mu^{(n)} \) with respect to \( \mathcal{D}_{q^m} \), so that \( H_{n'}(\mu^{(n)}) - H_{n'}(\mu) \to 0 \) as \( n \to \infty \). In particular, \( H_{n'}(\mu^{(n)}) \to \dim \mu \). We set \( \alpha = \dim \mu \) which we suppose, contrary to the assertion of the theorem, is strictly less than 1. We now apply Lemma 12 to the measure \( \mu^{(n)} \) with respect to the partition \( \mathcal{D}_{q^m} \). For each \( I \in \mathcal{D}_{q^m} \), let \( N(I) \) denote the number of atoms of \( \mu^{(n)} \) that lie inside \( I \). We have \( \mu^{(n)}(I) = N(I)k^{-n} \). Now take \( \beta \) with \( \alpha < \beta < 1 \). Lemma 12 implies that for a set of intervals \( I \) with total mass bounded away from 0, \( \mu^{(n)}(I) > 2^{-\beta n'} \) or \( N(I) > k^n2^{-\beta n'} \). Recall that \( \lambda \geq 1 \) and that \( \lambda^n \sim 2^{-n'} \) so that for some constant \( c \), \( N(I) > c k^n \lambda^{\beta n'} \) and

\[ \log N(I) > \log c + n(\log k\lambda + \log \lambda^{\beta - 1}) \geq \log c + n\log \lambda^{\beta - 1} \]

so that finally

\[ \log N(I) \geq \gamma n' \]

for some \( \gamma > 0 \). Now each interval \( I \) of \( \mathcal{D}_{q^m} \) is divided into \( 2^{(q-1)n'} \) interval of \( \mathcal{D}_{q^m} \). The atoms of \( \mu^{(n)} \) lie in distinct intervals of \( \mathcal{D}_{q^m} \) so that when (21) is valid, we have

\[ H(\mu^{(n)}_I, \mathcal{D}_{q^m}^I) > \gamma n'. \]

Dilating the measure and the partition \( \mathcal{D}_{q^m} \) by \( T_{2^n} \) leaves the entropy intact and we get

\[ H(T_{2^n} \mu^{(n)}_I, \mathcal{D}_{(q-1)n'} \mathcal{D}_{q^m}^I) > \gamma n'. \]

\( T_{2^n} \mu_I \) is a probability measure on an interval \( J \) of unit length for which we now obtain

\[ H_{(q-1)n'}(T_{2^n} \mu^{(n)}_I) > \frac{\gamma}{q-1}. \]

We intend to apply Theorem 14 to the convolution

\[ T_{2^n} \mu^{(n)}_I * T_{2^n} \lambda^n \mu = T_{2^n} \mu^{(n)}_I * T_{\theta n} \gamma \]

where \( \frac{1}{2} < \theta n < 1 \). One more deep fact is needed to verify the hypothesis of Theorem 14. This is the theorem of Hochman to the effect that self-similar measures have uniform entropy dimension which means that in a certain sense, most “dyadic” conditionals measures have entropy dimension close to the dimension of the measure. In our case, assuming \( \alpha < 1 \), together with (24), we find that for some \( \epsilon > 0 \), the hypotheses of Theorem 14 are fulfilled. The result is that for some \( \eta > 0 \),

\[ H_{(q-1)n'}(T_{2^n} \mu^{(n)}_I * T_{\lambda^n} \mu) > H_{(q-1)n'}(T_{\theta n} \mu) + \eta \]
the latter expression approximating $\alpha + \eta$, again by virtue of uniform entropy property of $\mu$. Thus for large $n$,

$$H\{T_{2^n} \mu_I, \mathcal{D}_{(q-1)n'}\} > (q-1)n'\alpha + (q-1)\eta n'$$

which translates to

(26) $$H(\mu_I, \mathcal{D}_{qn'}) \geq (q-1)n'\alpha + (q-1)\eta n'.$$

Recall that (22)—and consequently (26)—was valid for a set of intervals $I$ where total $\mu$-measure was bounded away from 0. For the remaining intervals we still have

$$H(\mu_I \ast T_{\lambda^n} \mu, \mathcal{D}_{qn'}) \geq H(T_{\lambda^n} \mu, \mathcal{D}_{qn'}) = H(T_\theta \mu, \mathcal{D}_{(q-1)n'})$$

because convolution never decreases entropy. As a result every summand in (20) exceeds $(q-1)n'\alpha + o(n')$.

Since the former intervals comprise a set of intervals with $\Sigma \mu(I) > c > 0$, substituting in (20) we obtain

(27) $$qn'\alpha = n'\alpha + (q-1)n'\alpha + c(q-1)\eta n' + o(n')$$

and this is impossible when $n$ is large. This shows that $\alpha = 1$ and establishes Hochman’s theorem under the special hypotheses: $\lambda_i = \lambda, P_i = k^{-1}, 1 \leq i \leq k$, and $k \lambda \geq 1$.

We turn now to Shmerkin’s proof of Theorem 1 in which the influence of Hochman’s proof of Theorem 9 can be seen. On the face of it the two theorems are of a different nature. Theorem 1 gives an upper bound to the dimension of the intersection of special sets whose individual dimensions are known. Theorem 9 gives a precise formula for the dimension of the set in question, but as has been pointed out, the non-trivial part of the proof is showing that the dimension in question is not less than that indicated in the theorem. But in fact the first step in the proof of Theorem 1 is showing that an upper bound of the dimension of an intersection $A \cap B$ will, under certain conditions, be determined by a lower bound on the dimension of the difference set $A - B = \{x - y; x \in A, y \in B\}$. As in Hochman’s paper the lower bound will come from the dimension of a measure on $A - B$. A significant difference between the two presentations is that Hochman uses entropy dimension of measures, whereas Shmerkin used $L^q$-dimension.

We begin with an explanation of the connection between dimensions of $A \cap B$ and $A - B$. For this we need the notion of Frostman exponents.

**Definition 15.** Let $\mu$ be a non-negative measure on a metric space $X$. We say that $s$ is a Frostman exponent for $\mu$, if for some $n_0 > 0$, all the balls $B(x, r)$ of radius $r \leq r_0$ satisfy

$$\mu(B(x, r)) < r^s.$$
Proposition 16. Let $X$ be a compact metric space and $\pi: X \to \mathbb{R}$ a Lipschitz map to the reals. Suppose $\mu$ is a measure on $X$ satisfying $\mu(B(x, r)) \geq c_0 r^h$ for a constant $c_0$ and all $r < r_0$. Assume that the pushforward measure $\pi\mu$ on $\mathbb{R}$ has a Frostman exponent $\alpha$. Then for every $y \in \mathbb{R}$,

$$\dim B \pi^{-1}(y) \leq \beta - \alpha. \tag{28}$$

Given sets $A, B$ as in theorem 1 we take $X = A \times B \subset \mathbb{R}^2$ and define $\pi(x, y) = x - y$. One finds measures $\mu_A, \mu_B$ on $A$ and $B$ respectively so that $\mu = \mu_A \times \mu_B$ satisfies the condition of the proposition with $\beta = \dim A \times \dim B = \dim A \times B$. Since $\pi$ is a Lipschitz map, the dimension of $\pi(A - B)$ cannot exceed $\dim A + \dim B$, nor can it be $> \beta$. So if $\pi(\mu_A \times \mu_B)$ can be shown to have Frostman exponent arbitrarily close to $\alpha = \min(\dim A + \dim B, 1)$ is will follow that $\pi^{-1}(0) = A \times B \cap \{(x, x)\} = A \cap B$ has dimension $\max(\dim A + \dim B - 1, 0)$ as asserted.

We remarked earlier that to prove Theorem 1 for $A$ and $B$ sets invariant under $T_p$ and $T_q$ respectively, it suffices to treat the case where $A$ is a $p$-Cantor set and $B$ a $q$-Cantor set. In this case there are natural measures $\mu_A$ and $\mu_B$ whose dimensions are respectively $\dim A$ and $\dim B$. $A$ and $B$ are also self-similar and $\mu_A, \mu_B$ can also be characterized as self-similar measures with equal weights:

$$\mu_A = r^{-1} \sum_{a \in J_A} \varphi_a \mu_A, \mu_B = s^{-1} \sum_{b \in J_B} \psi_b \mu_B, \text{ where}$$

$$\varphi_a(x) = r^{-1}(x + a), \psi_b(x) = s^{-1}(x + b).$$

Here $J_A \subset \{0, 1, \ldots, p - 1\}, J_B \subset \{0, 1, \ldots, q - 1\}$ and $\dim A = \frac{\log r}{\log p}, \dim B = \frac{\log s}{\log q}$. One checks that $\exists c_1, c_2$ with $\mu_A(B(x, r)) > c_1 r^{\dim A}, \mu_B(B(y, r)) > c_2 r^{\dim B}$, so for the appropriate definition of a metric on $A \times B$ we will have

$$(\mu_A \times \mu_B)(B(x, y), r) \geq c_1 c_2 r^{\dim A + \dim B}.$$ 

The desired result regarding $\pi(\mu_A \times \mu_B)$ is obtained making use of the $L^q$-dimension of this measure whose definition we will recall momentarily, and the evaluation of which for $\pi(\mu_A \times \mu_B)$ is at the core of Shmerkin's argument. The connection of Frostman exponents to the $L^q$-dimension is given by the following

Proposition 17. If $\mu$ is a compactly supported probability measure on $\mathbb{R}$, and for some $q > 1$ the $L^q$-dimension of $\mu$ is greater than $s$, then there is $r_0 > 0$ with

$$\mu([x - r, x + r]) \leq r^{(1 - \frac{1}{q})s} \tag{29}$$

for all $r \leq r_0$.

The $L^q$-dimension of a measure $\mu$ was introduced by Rényi, and, as with the entropy dimension, is the limit of expressions depending on the dyadic partitions $\mathcal{D}_n$ of an interval supporting the measure.
**Definition 18.** The $L^q$-dimension of a probability measure $\mu$ on an interval $[a, b]$ is given by

$$D(\mu, q) = \liminf_{n \to \infty} \left\{ -\frac{\log \left( \sum_{I \in \mathcal{D}_n} \mu(I)^q \right)}{(q - 1)n} \right\}$$

where $\mathcal{D}_n$ is the partition of $[a, b]$ to $2^n$ equal intervals.

We might note that the individual terms

$$-\log \left( \sum_{I \in \mathcal{D}_n} \mu(I)^q \right) \rightarrow -\sum_{I \in \mathcal{D}_n} \mu(I) \log \mu(I) = H(\mu, \mathcal{D}_n)$$

as $q \to 1$. For our use of the inequality (29), we will need arbitrarily large values of $q$. This flexibility is the advantage of the notion of $L^q$-dimension over that of the entropy dimension.

For a fixed $n$ it is useful to introduce the notion of the $L^q$-norm:

**Definition 19.** For a probability measure $\mu$ on an interval,

$$\|\mu(n)\|_q = \left\{ \sum_{I \in \mathcal{D}_n} \mu(I)^q \right\}^{1/q}.$$ (31)

Thus,

$$D(\mu, q) = \liminf_{n \to \infty} \left\{ -\frac{\log \|\mu(n)\|_q^q}{(q - 1)n} \right\}.$$ (32)

As with entropy, there is a smoothing effect in convolving measures which affects the dimension. For $L^q$-norms it can be seen that

$$\|\mu * v(n)\|_q \leq \|\mu(n)\|_q.$$ (33)

In line with Hochman's Theorem 14 and inequality (18), Shmerkin shows how under certain conditions the inequality (33) can be improved to

$$\|\mu * v(m)\|_q \leq 2^{-\varepsilon n} \|\mu(n)\|_q.$$ (34)

Two other points of contact between Shmerkin's argument and Hochman's discussion are

(i) the importance of exponential separation,

(ii) expressing the measure under investigation as an infinite convolution, giving rise to a "self-similarity" property as in the equality (11).

In all this Shmerkin gives credit to Hochman’s exposition. Shmerkin’s treatment of the measure $\pi(\mu_A \times \mu_B)$ parts ways with Hochman’s discussion for a self-similar measure, in that the foregoing measure needs to be seen as one of a family of measures, $\pi_t(\mu_A \times \mu_B)$ where $\pi_t(x, y) = x - ty$. The phenomenon of self-similarity of a single measure is replaced by a “dynamically driven self-similarity” in which each measure in the family is related by convolution to another.
We sketch some of the details. We begin with a description of \( A \times B \) as a “dynamically driven” self-affine set. Assume \( p < q \) where \( A \) is \( p \)-Cantor and \( B \) is \( q \)-Cantor. Underlying Shmerkin’s analysis is a dynamical system on \([1, q]\) where the endpoints are identified, and a transformation \( S \) is defined by

\[
S(t) = \begin{cases} 
pt & \text{for } 1 \leq t < q/p \\
pqt & \text{for } q/p \leq t < q.
\end{cases}
\]

Note that with \( f(t) = \log t/\log q \), the interval \([1, q]\) is identified with the unit interval and \( S \) corresponds to the “irrational rotation”, \( y \to y + \log p/\log q \). It is at this point that the multiplicative independence of \( p, q \) enters, as a result of which the system \([(1, q), S]\) is uniquely ergodic and all orbits are equidistributed.

We define a family of affine transformations in \( A \times B \) depending on the parameter \( t \in [1, q) \), and \( a \in J_A \) and \( b \in J_B \):

\[
\Phi_{ab}^t(x, y) = \begin{cases} 
\left(\frac{x-a}{p}, y\right) & \text{for } 1 \leq t < q/p \\
\left(\frac{x-a}{p} + \frac{a-tb}{pq}\right) & \text{for } q/p \leq t < q.
\end{cases}
\]

For any \( t \in [1, q) \), any point \((x, y) \in A \times B\) can be represented as a limit

\[
(x, y) = \lim \Phi_{a_n b_n}^t \Phi_{a_{n-1} b_{n-1}}^{S_1(t)} \cdots \Phi_{a_1 b_1}^{S_{n-1}(t)} (0, 0)
\]

for some sequence \((a_n, b_n)\) in \((J_A \times J_B)^\mathbb{N}\). Moreover, any such limit gives a point in \( A \times B \). Now define similarity transformations on \( \mathbb{R} \):

\[
\psi_{ab}^t(u) = \begin{cases} 
\frac{u-a}{p} & \text{for } 1 \leq t < q/p \\
\frac{u}{p} + \frac{a-tb}{pq} & \text{for } q/p \leq t < q.
\end{cases}
\]

We define for all \( t, \pi_t(x, y) = x - ty \), and we check the commutation relation

\[
\pi_t \Phi_{ab}^t = \psi_{ab}^t \pi_S(t).
\]

Combining (34) and (36) we find that

\[
\pi_t(x, y) = \lim \pi_t \Phi_{a_1 b_1}^t \Phi_{a_2 b_2}^{S_1(t)} \cdots \Phi_{a_n b_n}^{S_{n-1}(t)} (0, 0)
\]

\[
= \lim \psi_{a_1 b_1}^t \psi_{a_2 b_2}^{S_1(t)} \cdots \psi_{a_n b_n}^{S_{n-1}(t)} (0).
\]

As a result, the set \( A - tB \) consists of points of the form given by (39). We now define a family of atomic measures \( \rho(t) \) by

\[
\rho(t) = \begin{cases} 
\frac{1}{t} \sum_{a \in J_A} \delta_a & \text{for } 1 \leq t \leq q/p \\
\frac{1}{t} \sum_{a \in J_A, b \in J_B} \delta_{a - \frac{p}{q} b} & \text{for } q/p \leq t < q
\end{cases}
\]

and setting \( \lambda = p^{-1} \), we form the convolutions

\[
\mu_{t, n} = T_{\lambda} \rho(t) * T_{\lambda^2} \rho(St) * \cdots * T_{\lambda^n} \rho(S^{n-1}t).
\]

As usual these converge to a limit

\[
\mu_t = T_{\lambda} \rho(t) * T_{\lambda^2} \rho(st) * \cdots * T_{\lambda^n} \rho(s^n t) * \cdots.
\]
Now in analogy with (11) we have

\[ \mu_t = \mu_{t,n} \ast T_{\lambda^n} \mu_{St}. \tag{43} \]

One can verify that \( \mu_t = \pi_t(\mu_A \times \mu_B) \) so that the measure \( \pi(\mu_A \times \mu_B) \) discussed earlier is the measure \( \mu_1 \) defined in (40). The proof of Theorem 1 will follow from a general formula that enables us to evaluate the \( L^q \)-dimension of \( \mu_t \) which turns out to be independent of \( t \). Once we have an explicit expression for the \( L^q \)-dimension of \( \mu \), we invoke Propositions 16 and 17 to give the proof of Theorem 1.

As a consequence of unique ergodicity of \([1, q], S\) one has for any piecewise continuous function \( f(t) \) and \( t \in [1, q] \)

\[
\frac{1}{N} \sum_{n=0}^{N-1} f(S^n t_0) \to \int_1^q f(t) dv(t)
\]
where \( dv \) is the unique invariant measure on \([1, q] \); i.e.,

\[ dv(t) = \frac{dt}{t \log q}. \]

We denote by \( \sigma(t) \) the number of atoms in \( \rho(t) \):

\[ \sigma(t) = \begin{cases} r & \text{for } 1 \leq t < \frac{q}{p} \\ rs & \text{for } \frac{q}{p} \leq t < q \end{cases} \]

\( \log \sigma(t) \) is piecewise continuous, and so

\[
\frac{1}{N} \sum_{n=0}^{N-1} \log \sigma(S^n t_0) \to \frac{1}{\log q} \int_1^q \log \sigma(t) \frac{dt}{t}
\]

for any \( t_0 \in [1, q] \). This becomes

\[ \frac{\log \sigma(t_0) \sigma(S t_0) \ldots \sigma(S^{N-1} t_0)}{N} \to \log r + \frac{\log p}{\log q} \log s. \tag{44} \]

We want to evaluate \( D(\mu_t, q) \) given by

\[ D(\mu_t, q) = \liminf_{n \to \infty} \left\{ \frac{-\log \sum_{I \in \mathcal{F}_{n}} \mu_t(I)^q}{(q-1)n} \right\}. \tag{30 bis} \]

Using the “self-similarity” equations (43) and a version of the subadditive ergodic theorem for uniquely ergodic systems, Shmerkin proves that for values of \( q \to \infty \), the limit in (30 bis) exists for a.e. \( t \in [1, q] \) and is constant, say \( D \). The most intricate section in Shmerkin's proof of Theorem 1 deals with the proof of the following statement:

**Theorem 20.** For every \( n \in \mathbb{N} \) let \( n' \) be defined by \( 2^{n'} \leq \lambda^{-n} < 2^{n'+1} \). Let \( t \) be such that the limit exists in (30 bis) and equals \( D \). Let \( R \in \mathbb{N} \), then

\[
\lim_{n \to \infty} \left\{ \frac{-\log \sum_{I \in \mathcal{F}_{n+R'}} \mu_{t,n}(I)q}{(q-1)n \log \lambda} \right\} = D.
\]
It is for the proof of this theorem that Shmerkin proves a quantitative version of the smoothing phenomenon for convolutions, and specifically for the equation (43). The proof compares \( L^q \)-norms relative to \( D_{R'} \) and \( L^p \)-norms relative to \( D'_{R} \) and the argument can be seen as motivated by Hochman’s treatment of \( H(\mu, D_{qR}) \) vis-a-vis \( H(\mu, D_{R}) \) in equation (19).

Given Theorem 20 we now make use of exponential separation which can be proved for the atoms in the measures \( \rho(t) \ast \rho(St) \ast \cdots \ast \rho(S^n t) \). This implies that for a large enough \( R \), the atoms of \( \mu_{t,n} \) appear in different cells of \( D_{R'} \). Assume \( D < 1 \).

Now, each atom of \( \rho(t) \) carries the weight \( \sigma(t)^{-1} \), so that because of the separation of the atoms in \( D_{R'} \):

\[
\sum_{I \in D_{R'}} \mu_{t,n}(I) = (\sigma(t) \ast \sigma(St) \ast \cdots \ast \sigma(S^{n-1}t))^{-q}.
\]

By Theorem 20

\[
\frac{-\log(\sigma(t) \ast \sigma(St) \ast \cdots \ast \sigma(S^{n-1}t))}{(q-1)n \log \lambda} \to D.
\]

Recall that \( \lambda = p^{-1} \). By (44) this limit is

\[
D = \frac{\log r}{\log p} + \frac{\log s}{\log q} = \dim A + \dim B.
\]

This is a sketch of Shmerkin’s argument which shows that for every \( t \) and for the measure \( \mu_t \) as defined

\[
D(\mu_t, q) = \min(\dim A + \dim B, 1).
\]

In general, had we chosen \( \mu_A \) and \( \mu_B \) such that the measures \( \rho(t) \) are not equidistributed on their supports, the dimension \( D(\mu_t, q) \) will depend on \( q \), and \( \dim A + \dim B \) is replaced by \( \dim \mu_A + \dim \mu_B \).

REFERENCES

[1] H. Furstenberg, Intersections of Cantor sets and transversality of semigroups. In *Problems in Analysis* (Symposium in honor of Salomon Bochner, Princeton University Press, Princeton, N.J. 1969), 41–59, 1970.

[2] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation, *Math. Systems Theory*, 1 (1967), 1–49.

[3] M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy, *Ann. of Math. (2)*, 180 (2014), 773–822.

[4] P. Shmerkin, On Furstenberg’s intersection conjecture, self-similar measures, and the \( L^q \)-norms of convolutions, *Ann. of Math. (2)*, 189 (2019), 319–391.

[5] M. Wu, A proof of Furstenberg’s conjecture on the intersection of \( \times p \) and \( \times q \)-invariant sets, arXiv:1609.08053v3, February 2019.

HILLEL FURSTENBERG <hillel.fursten@mail.huji.ac.il>: Department of Mathematics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel