Multifractal analysis via Lagrange duality

ALEX RUTAR

ABSTRACT. We provide a self-contained exposition of the well-known multifractal formalism for self-similar measures satisfying the strong separation condition. At the heart of our method lies a pair of quasiconvex optimization problems which encode the parametric geometry of the Lagrange dual associated with the constrained variational principle. We also give a direct derivation of the Hausdorff dimension of the level sets of the upper and lower local dimensions by exploiting certain weak uniformity properties of the space of Bernoulli measures.

CONTENTS

1. INTRODUCTION

1.1. Prelude on multifractal analysis. Many naturally-occurring systems can be modelled by a sequence of independent and identically distributed random variables, or more generally by a dynamical system supporting an invariant measure. In these situations, one is often interested in the long-term behaviour given *typical* initial conditions.

In *multifractal analysis*, one is interested in the *non-typical* behaviour of a dynamical system. One of the oldest examples which hints at multifractal analysis can be found in the study of *Besicovitch–Eggleston sets* [Bes34; Egg51]. Fix an integer $b \geq 2$ and consider the base-b expansion of a point $x \in [0,1]$. Suppose x has base-b expansion $x = 0.a_1a_2a_3\ldots$ where $a_i \in \{0,1,\ldots,b-1\}$. Then the *digit frequencies* of x exist almost surely: there is a set $E \subset [0,1]$ of full Lebesgue measure such that for each $x = 0.a_1a_2a_3\ldots \in E$ and $i = 0,1,\ldots,b-1$,

$$\lim_{n\to\infty} \frac{\#\{j : a_j = i, 1 \leq i \leq n\}}{n} = \frac{1}{b}.$$

But what can be said about non-typical behaviour—the complement of E is non-empty, but what can be said about its structure?

In order to understand non-typical behaviour, it is reasonable to associate with the original system different invariant measures, for which typical points of the new measure will reveal new properties of the underlying system. For instance, we may define a measure on $[0,1]$ by taking points with base-b expansion chosen randomly according to some alternative distribution $(p_0,\ldots,p_{b-1})$ where, independently for each n, we choose a_n to be the digit i with probability p_i. Typical points for this measure will not be elements of the set E (unless $p_0 = \cdots = p_{b-1} = 1/b$), but the measure still has positive dimension: in fact, the set of typical points for this measure has Hausdorff dimension precisely

$$\frac{-\sum_{i=0}^{b-1} p_i \log p_i}{\log b}.$$

In particular, even though the complement of E has measure 0, it has Hausdorff dimension 1. Moreover, this gives us a good understanding of the size of the sets of points where the digit frequencies exist and are given by some distribution. But beyond this, there are still many points (in fact also a set of full Hausdorff dimension) for which the digit frequencies do not exist at all and one can ask a variety of questions about such points.

In this document, we will focus on the multifractal analysis of measures, of which the above consideration is a special case. Give a compactly supported Borel probability measure in $\mathbb{R}^d$ and $x \in \operatorname{supp}\mu$, we wish to study the local scaling properties of the measure μ at x from the perspective of the *local dimension*

$$\dim_{\mathrm{loc}}(\mu,x) = \lim_{r\to 0} \frac{\log\big(\mu(B(x,r))\big)}{\log r},$$

when the limit exists (or taking a limit infimum or limit supremum when it does not). If the measure μ is sufficiently nice, then $\dim_{\mathrm{loc}}(\mu,x)$ almost surely exists and

attains some constant value. Moreover, this average scaling property is intimately related to the *dimension* of the measure μ, since a measure which has larger local dimensions is more concentrated in neighbourhoods of points in its support. But what about non-typical points? A reasonable question to ask is how large such sets are: we will be most interested in the quantity

$$f_\mu(\alpha) = \dim\{x \in \operatorname{supp}\mu : \dim_{\mathrm{loc}}(\mu, x) = \alpha\} \tag{1.1}$$

where $\dim$ is either Hausdorff or packing dimension; or perhaps some variant by considering only the points where the lower or upper limit converge to some value α. We refer to this function as the *multifractal spectrum* of the measure μ. For more detail on the Hausdorff and packing dimensions, we refer to reader to the books [BP17; Fal97]. The level sets of local dimensions will often be dense subsets of the support of the measure, so other notions of dimension such as the box dimensions (which are stable under closure) do not typically yield meaningful information.

Note that the function f_μ appears to be defined in a somewhat pathological way: it is the Hausdorff dimension of a level set of an asymptotically-defined pointwise function, so there appears to be no *a priori* reason to expect that the dependence on α is sensible. Nonetheless, the *multifractal miracle* (perhaps first observed non-rigorously in the physics literature [HJK+86]) is that the quantity f_μ, for the class of measures that we will consider, is in fact a non-trivial concave analytic function of α. We will study the multifractal quantity f_μ, in conjunction with other quantitative measures of smoothness such as the L^q-spectrum which will be introduced in later sections.

For digit frequency measures, and more generally for the family of measures that we will consider, we emphasize that the multifractal formalism has been rigorously established in [AP96; CM92]. Expositions of the standard proofs can be found in [Fal97, §11.2] and [BSS23, Chapter 5]. Our goal in this document is to provide a novel approach to and exposition of these results in order to motivate the multifractal analysis of measures and highlight the relationship with Lagrange duality, constrained optimization, and large deviations theory.

1.2. Constrained optimization and duality. Recall our general goal: we have fixed a measure μ and we wish to understand the multifractal spectrum $f(\alpha)$ defined in (1.1). Suppose that we could choose a measure ν such that $\dim_{\mathrm{loc}}(\mu, x) = \alpha$ for ν-a.e. $x \in \operatorname{supp}\nu$. Then we would immediately obtain the lower bound

$$f(\alpha) \geq \dim_{\mathrm{H}}\mu \coloneqq \sup\{\dim_{\mathrm{H}} E : E \subset \operatorname{supp}\mu : \mu(E) = 1\}$$

(see §2.1 for details on dimensions of measures).

Unfortunately, this idea is not particularly helpful in general, since on any set with given Hausdorff dimension one can find measures with dimension arbitrarily close to the dimension of the original set. However, measures defined on digit sets (and more generally the measures that we will consider in this document) have more structure: they are projections of *invariant* measures on some compact metric space Ω under a coding map $\pi\colon \Omega \to \mathbb{R}^d$. Given a continuous transformation $T\colon \Omega \to \Omega$, recall that a measure ν on Ω is *T-invariant* if for all Borel sets $E \subset \Omega$,

$\nu(E) = \nu(T^{-1}(E))$, and *ergodic* if $\nu(E) \in \{0, 1\}$ for any set E with $E = T^{-1}(E)$. Let $\mathcal{M}$ denote the space of T-invariant and ergodic measures. Then one might ask for the following *constrained variational principle*: is the general inequality

$$f(\alpha) \geq \sup_{\nu \in \mathcal{M}} \{\dim_{\mathrm{H}} \pi_* \nu : \dim_{\mathrm{loc}}(\mu, x) = \alpha \text{ for } \pi_*\nu\text{-a.e. } x\} \tag{1.2}$$

sharp? Here, $\pi_*\nu$ is the pushforward measure of ν under the projection map π. Moreover, the *dimension* of $\pi_*\nu$ is determined in a loose sense by the *entropy* of ν as well as the geometry of the projection map π. If such a variational principle holds, the situation seems more hopeful! In fact, assuming that the map π is at least reasonably well-behaved, our domain and objective function satisfies important forms of regularity: the space $\mathcal{M}$ is a compact and convex metric space (with topology precisely the weak-star topology) and the entropy map on $\mathcal{M}$ is upper semicontinuous [Wal82, Proposition 5.2 and Theorem 8.2]. We can therefore replace the general question of understanding the multifractal spectrum with two new questions:

1. Can we understand the variational lower bound in (1.2); and
2. Can we prove that equality in (1.2) actually holds?

Associated with the constrained optimization problem which gives a lower bound to $f(\alpha)$ is the corresponding *unconstrained Lagrange dual* minimization problem $\tau(q)$, which encodes the Lagrange multipliers problem associated with the constrained maximization. These dual optimization problems are fundamentally related by the concave conjugate. When the underlying space and the objective function are particularly well-behaved, the relationship between $f(\alpha)$ and $\tau(q)$ is well-understood (a particularly elegant case occurs when the domain and objective function are convex, see for instance [Roc70, Part III]). However, the key difference in our perspective is to make minimal assumptions on the domain and objective functions, and to instead understand the *parametric geometry* of the maximization problem *as a function of* α. This allows us to convert topological properties of the Lagrange dual or generic properties of the function τ to information about the primal function $f(\alpha)$ (for instance, connectedness of the set of minimizers or differentiability of τ at a value q implies the concave conjugate relation at values of α corresponding to q). This approach is introduced in an abstract setting in §3.1.

We then need to prove that the inequality in (1.2) is in fact an equality. To do this, we make a key observation: in our setting, the unconstrained dual corresponds precisely to the *L^q-spectrum* of the measure (see §2.2 for a formal definition). This proof can be found in Theorem 2.8, which uses the *method of types* from large deviations theory to establish a variational formula for the function $\tau(q)$. The author was motivated to take this approach by the recent work of Kolossváry [Kol23]. In order to complete the proof, it then remains to establish a general concave conjugate relationship between the L^q-spectrum and the multifractal spectrum through a *geometric large deviations bound* (see Proposition 3.7). This part of the argument has much more of a geometric flavour, since we must handle points which are "invisible" to the dynamics (in that they are non-typical for any pushforward of an ergodic measure on Ω). In Proposition 3.15 we also give a direct argument using a variant of a classical density theorem for Hausdorff dimension to establish equality in (1.2).

In fact, the method of types argument and the density argument are precisely the tools required to reduce a covering argument for general points to a covering argument for "typical" points, which allows us to take advantage of the underlying dynamical structure of the sets and measures under consideration. Such reductions are necessary since covering arguments must work for all points of a set simultaneously.

In this document, we focus on the multifractal analysis of the local dimensions of a probability measure μ. One can ask analogous versions of this question, such as for Birkhoff averages on a dynamical system. In this case, the constrained and unconstrained optimization problems considered above are intimately related to the notion of *topological pressure* (see [Cli14] for a survey on this dynamical perspective to multifractal analysis). The general approach to duality is certainly very applicable in this dynamical context as well. In fractal geometry, however, a key role is played by the projection map π and the geometry of Euclidean space which prevents us from taking a purely dynamical perspective.

Beyond the scope of this article, there are many new difficulties but also many recent breakthroughs, such as the establishment of a *dynamical dimension gap* for self-affine sets [DS17] (in some sense the strongest possible failure of equality in (1.2)) or the formula for the L^q-spectrum of an overlapping self-similar measure [Shm19] for $q > 1$, which draws on deep ideas from discretized sum-product theory and additive combinatorics.

Many of the ideas in these notes can be extended and used in more general settings. Some results (such as the general setup in §3.1, or the approach to the variational principle in §3.4) have been stated with this use-case in mind. However, generally speaking, for clarity of exposition and in order to highlight the many connections between these diverse areas of mathematics while minimizing the technical complexity, we will restrict our attention to a special setting: the case of self-similar measures satisfying the strong separation condition.

1.3. Self-similar sets and measures. We now introduce the class of measures which we will focus on for the remainder of this document. This class of measures generalizes the "digit frequency measures" discussed in the introduction. A key distinction, which only causes minor difficulties in our setting but is a critical difficulty in developing this theory in a more general context, is that we do not require the subdivision into parts to all be of the same size.

Fix a finite index set $\mathcal{I}$. For each $i \in \mathcal{I}$, let $S_i \colon \mathbb{R}^d \to \mathbb{R}^d$ be a *contracting similarity*, that is for each $i \in \mathcal{I}$ there is some $r_i \in (0,1)$ so that

$$|S_i(x) - S_i(y)| = r_i \cdot |x - y| \text{ for all } x, y \in \mathbb{R}^d .$$

We then call the finite family $\{S_i\}_{i\in\mathcal{I}}$ an *iterated function system of similarities*, or *IFS* for short.

Let $\mathcal{P} = \mathcal{P}(\mathcal{I}) \subset \mathbb{R}^{\mathcal{I}}$ denote the set of probability vectors

$$\mathcal{P} = \big\{(p_i)_{i\in\mathcal{I}} : p_i \in [0,1], \textstyle\sum_{i\in\mathcal{I}} p_i = 1\big\}.$$

We equip the space $\mathcal{P}$ with the metric from $\mathbb{R}^{\mathcal{I}}$: note that $\mathcal{P}$ is compact. Now to

each self-similar IFS, there is a unique compact set K satisfying

$$K = \bigcup_{i \in \mathcal{I}} S_i(K)$$

and, given a probability vector $\boldsymbol{p} = (p_i)_{i \in \mathcal{I}} \in \mathcal{P}$, a unique Borel probability measure $\mu_{\boldsymbol{p}}$ satisfying

$$\mu_{\boldsymbol{p}} = \sum_{i \in \mathcal{I}} p_i \cdot \mu_{\boldsymbol{p}} \circ S_i^{-1}.$$

We call the set K the *attractor* and the measure $\mu_{\boldsymbol{p}}$ the *invariant measure* associated with the IFS $\{S_i\}_{i \in \mathcal{I}}$ and probabilities $(p_i)_{i \in \mathcal{I}}$. If each $p_i > 0$, we note that $\operatorname{supp} \mu_{\boldsymbol{p}} = K$.

A classic example of an iterated function system is the Cantor IFS, which is the system $\{x \mapsto x/3, x \mapsto x/3 + 2/3\}$. Then the attractor K is the middle-thirds Cantor set (i.e. the set of points in $[0, 1]$ with ternary expansion consisting of only 0s and 2s). If we take the probability vector $(1/2, 1/2) \in \mathcal{P}$, then the corresponding measure μ is just Hausdorff $\frac{\log 2}{\log 3}$-measure restricted to K and normalized to have measure 1.

We say that the IFS $\{S_i\}_{i \in \mathcal{I}}$ satisfies the *strong separation condition*, or *SSC* for short, if $S_i(K) \cap S_j(K) = \varnothing$ for all $i \neq j$ in $\mathcal{I}$. Note that, formally, the SSC disallows measures which arise from a digit frequency set as discussed in the introduction. However, those measures satisfy the following slightly weaker condition. We say that the IFS $\{S_i\}_{i \in \mathcal{I}}$ satisfies the *open set condition*, or *OSC* for short, if there is an open set U such that $S_i(U) \subset U$ for all $i \in \mathcal{I}$, and $S_i(U) \cap S_j(U) = \varnothing$ for $i \neq j$.

The main results in this document also hold under the OSC, though there are some additional technical complications related to non-injectivity of the projection map π that we can bypass under the SSC. For simplicity of exposition, we will for the remainder of the document assume that the SSC holds.

1.4. Shift space and symbolic coding. As a result of the iterative nature of the construction of self-similar sets and measures, it is natural to introduce notation formalizing this iterative procedure. One natural setting for such a construction is the *full shift* on symbols $\mathcal{I}$. Let $\Omega = \mathcal{I}^{\mathbb{N}}$ denote the set of sequences on symbols $\mathcal{I}$ equipped with the product topology, and let $\pi \colon \Omega \to K$ denote the natural coding map defined by

$$\left\{\pi\left((i_n)_{n=1}^{\infty}\right)\right\} = \lim_{n \to \infty} S_{i_1} \circ \cdots \circ S_{i_n}(K).$$

Note that if the IFS satisfies the SSC, the map $\pi \colon \Omega \to K$ is a homeomorphism.

It will be useful to introduce some notation to handle sequences in Ω. First, let $\mathcal{I}^* = \bigcup_{n=0}^{\infty} \mathcal{I}^n$ denote the space of all finite words on the alphabet $\mathcal{I}$, equipped with the operation of concatenation. Given $\mathtt{i} \in \mathcal{I}^*$ and $\mathtt{k} \in \mathcal{I}^*$, we say that $\mathtt{i}$ is a *prefix* of $\mathtt{k}$ if there is a $\mathtt{j} \in \mathcal{I}^*$ so that $\mathtt{k} = \mathtt{ij}$. Concatenation also extends naturally to expressions of the form $\mathtt{i}\gamma$ for $\gamma \in \Omega$, and allows us to define the notion of a prefix of an infinite word. In particular, for $\gamma \in \Omega$, we let $\gamma|_n$ denote the unique prefix of γ of length n.

We equip the sequence space Ω with the *shift map* $\sigma\colon \Omega \to \Omega$ defined by

$$\sigma(i_1, i_2, i_3, \ldots) = (i_2, i_3, \ldots).$$

For example, for all $n \in \mathbb{N}$, $\gamma = \gamma|_n \sigma^n(\gamma)$. A probability vector $\boldsymbol{p} \in \mathcal{P}$ then naturally defines a shift-invariant and ergodic product measure $\boldsymbol{p}^{\mathbb{N}}$. We call such a measure *Bernoulli*. The self-similar measure $\mu_{\boldsymbol{p}}$ is simply the pushforward of $\boldsymbol{p}^{\mathbb{N}}$ by the coding map π.

Finally, we set for $\boldsymbol{p} = (p_i)_{i\in\mathcal{I}}$

$$\Omega_{\boldsymbol{p}} = \Big\{(i_n)_{n=1}^{\infty} \in \Omega : \lim_{n\to\infty} \frac{\#\{\ell : i_\ell = j \text{ for } 1 \le \ell \le n\}}{n} = p_j \text{ for } j \in \mathcal{I}\Big\},$$

in other words the collection of sequences in Ω where the digit frequencies exist and are given by the probability vector $\boldsymbol{p}$.

It will turn out, for our purposes, that it suffices to consider only Bernoulli measures, and so we will not consider the more general class of ergodic measures on Ω for the remainder of the document. Note that for more general iterated function systems (with corresponding geometric complexities associated with the map π), this is no longer the case.

2. DIMENSIONS AND L^q-SPECTRA OF SELF-SIMILAR MEASURES

In this section, we begin with some classical theory: the notion of the Hausdorff dimension of a measure. We will also derive a variational formula for the L^q-spectrum of a self-similar measure satisfying the SSC using the *method of types* from large deviations theory. Such a strategy was motivated by the main results and techniques used in [Kol23].

There are substantial advantages to deriving a formula for the L^q-spectrum in terms of a variational formula, in contrast to classical proofs such as the one given in [Fal97, §11.2] which work directly with an analytic closed formula. Firstly, it suffices to consider continuous functions rather than smooth functions, which gives much more flexibility in more complex settings. Secondly, as we will see in §3.1 and §4.2, there are many indirect techniques to understand the geometry of a general optimization problem, which makes it easier to obtain precise information instead of directly using an implicit formula. Such techniques become essential in settings where a closed formula for the L^q-spectrum may not even exist.

2.1. Hausdorff dimension of self-similar measures.

Given a Borel probability measure μ, we denote the *local dimension* of μ at x by

$$\dim_{\mathrm{loc}}(\mu, x) = \lim_{r\to 0} \frac{\log \mu(B(x, r))}{\log r}$$

when the limit exists. If the limit does not exist, we write $\underline{\dim}_{\mathrm{loc}}(\mu, x)$ (resp. $\overline{\dim}_{\mathrm{loc}}(\mu, x)$) to denote the respective quantity except with a limit infimum (resp. limit supremum) in place of the limit. We can use the local dimensions to give a formulation of the Hausdorff dimension of μ.

Definition 2.1. Given a compactly supported Borel probability measure μ, we write

$$\dim_{\mathrm{H}} \mu = \operatorname*{ess\,sup}_{x \in \operatorname{supp} \mu} \underline{\dim}_{\mathrm{loc}}(\mu, x).$$

We say μ is *exact-dimensional* if there is an $\alpha \in \mathbb{R}$ so that $\dim_{\mathrm{loc}}(\mu, x) = \alpha$ for μ-a.e. $x \in \operatorname{supp} \mu$.

Equivalently, as explained in [Fal97, Proposition 10.1],

$$\dim_{\mathrm{H}} \mu = \sup\{\dim_{\mathrm{H}} E : E \subset \operatorname{supp} \mu : \mu(E) = 1\}.$$

Now given a self-similar IFS $\{S_i\}_{i \in \mathcal{I}}$ and $\boldsymbol{p} = (p_i)_{i \in \mathcal{I}}, \boldsymbol{w} = (w_i)_{i \in \mathcal{I}} \in \mathcal{P}(\mathcal{I})$, denote the *cross entropy* and *entropy* by

$$H(\boldsymbol{w}, \boldsymbol{p}) = \sum_{i \in \mathcal{I}} w_i \log(1/p_i) \qquad \text{and} \qquad H(\boldsymbol{w}) = H(\boldsymbol{w}, \boldsymbol{w}).$$

We also denote the *Kullback–Leibler divergence* (also known as the *relative entropy*) by

$$D_{\mathrm{KL}}(\boldsymbol{w} \,\|\, \boldsymbol{p}) = H(\boldsymbol{w}, \boldsymbol{p}) - H(\boldsymbol{w}).$$

We recall by Jensen's inequality applied to the logarithm, $D_{\mathrm{KL}}(\boldsymbol{w} \,\|\, \boldsymbol{p}) \geq 0$ with equality if and only if $\boldsymbol{w} = \boldsymbol{p}$. For a general introduction to these concepts in information theory, we refer the reader to [CT06]. Finally, we define the *Lyapunov exponent*

$$\chi(\boldsymbol{w}) = \sum_{i \in \mathcal{I}} w_i \log(1/r_i).$$

The Lyapunov exponent captures the asymptotic contraction rate at points typical for the measure $\mu_{\boldsymbol{w}}$. The key role of the Lyapunov exponent in dimension theory is highlighted, for instance, by the seminal works of Ledrappier–Young [LY85a; LY85b].

We will now determine the dimensions of self-similar measures. We first observe the following technical lemma, which allows a reduction to balls which intersect precisely one image $S_{\mathtt{i}}(K)$ for $\mathtt{i} \in \mathcal{I}^*$. In essence, this will allow us to treat the geometry of $\mu_{\boldsymbol{p}}$ in a purely symbolic way. Note that $\operatorname{diam}(K) \cdot r_{\mathtt{i}} = \operatorname{diam}(S_{\mathtt{i}}(K))$.

Lemma 2.2. *For all sufficiently small $\delta > 0$ and for all $\boldsymbol{p} \in \mathcal{P}$ with $p_i > 0$ for all $i \in \mathcal{I}$, there is a constant $c = c(\boldsymbol{p}, \delta) > 0$ so that for all $\mathtt{i} \in \mathcal{I}^*$ and $x \in S_{\mathtt{i}}(K)$,*

$$c \cdot p_{\mathtt{i}} \leq \mu_{\boldsymbol{p}}\big(B(x, \delta \cdot r_{\mathtt{i}})\big) \leq p_{\mathtt{i}}.$$

Proof. Since the IFS satisfies the strong separation condition, for all sufficiently small $\delta > 0$ the $\delta \cdot r_{\mathtt{i}}$-neighbourhood of $S_{\mathtt{i}}(K)$ in K is again $S_{\mathtt{i}}(K)$. Thus

$$\mu_{\boldsymbol{p}}\big(B(x, \delta \cdot r_{\mathtt{i}})\big) \leq p_{\mathtt{i}}.$$

On the other hand, since $\delta > 0$ is fixed, there is a uniform $N \in \mathbb{N}$ and a word $\mathtt{j} \in \mathcal{I}^N$ so that

$$S_{\mathtt{i}} \circ S_{\mathtt{j}}(K) \subseteq \mu_{\boldsymbol{p}}\big(B(x, \operatorname{diam}(S_{\mathtt{i}}(K)) \cdot \delta)\big).$$

Taking $c = \min\{p_{\mathtt{j}} : \mathtt{j} \in \mathcal{I}^N\}$ gives the desired result. □

We also obtain a simple lemma which gives information about local dimensions in terms of the digit frequencies of the symbolic representation.

Lemma 2.3. *Let $\boldsymbol{w} \in \mathcal{P}(\mathcal{I})$ and let $\gamma = (i_n)_{n=1}^{\infty} \in \Omega_{\boldsymbol{w}}$. Then*

$$\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, \gamma) = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}.$$

Proof. By Lemma 2.2, there is a $\delta > 0$ and a $c > 0$ so that for any $n \in \mathbb{N}$,

$$c \cdot p_{\gamma|_n} \leq \mu\big(B(\pi(\gamma), \delta \cdot r_{\gamma|_n})\big) \leq p_{\gamma|_n}.$$

Thus

$$\dim_{\mathrm{loc}}(\mu, \pi(\gamma)) = \lim_{n\to\infty} \frac{\log p_{\gamma|_n}}{\log r_{\gamma|_n}} = \lim_{n\to\infty} \frac{\log \prod_{i\in\mathcal{I}} p_i^{nq_i}}{\log \prod_{i\in\mathcal{I}} r_i^{nq_i}} = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}$$

from the definition of $\Omega_{\boldsymbol{w}}$, as claimed. □

In particular, we obtain following well-known dimension formula for self-similar measures.

Proposition 2.4. *Let $\{S_i\}_{i\in\mathcal{I}}$ be an IFS satisfying the SSC and let $\boldsymbol{p}, \boldsymbol{w} \in \mathcal{P}$. Then for $\mu_{\boldsymbol{w}}$-a.e. $x \in \operatorname{supp} \mu_{\boldsymbol{w}}$,*

$$\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, x) = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}.$$

In particular, $\mu_{\boldsymbol{p}}$ is exact-dimensional with dimension

$$\dim_{\mathrm{H}} \mu_{\boldsymbol{p}} = \frac{H(\boldsymbol{p})}{\chi(\boldsymbol{p})}.$$

Proof. By Kolmogorov's Strong Law of Large Numbers, $\boldsymbol{w}^{\mathbb{N}}(\Omega_{\boldsymbol{w}}) = 1$. Thus the dimensional result follows from Lemma 2.3. □

2.2. L^q-spectra of self-similar measures. Let μ be a compactly supported Borel probability measure and let $q \in \mathbb{R}$. We write for $r \in (0, 1)$

$$G_\mu(r, q) = \sup\left\{ \sum_i \mu\big(B(x_i, r)\big)^q : \{x_i\}_i \text{ is a } 2r\text{-separated subset of } \operatorname{supp}\mu \right\}.$$

We then denote the *L^q-spectrum* of μ at q by

$$\tau_\mu(q) = \liminf_{r\to 0} \frac{\log G_\mu(r,q)}{\log r}.$$

We recall the following standard result.

Lemma 2.5. *Let μ be a compactly supported Borel probability measure. Then $\tau_\mu(q)$ is a concave and increasing function of q. Moreover, $\tau_\mu(0) = -\dim_{\mathrm{B}}(\operatorname{supp}\mu)$ and $\tau_\mu(1) = 0$.*

Proof. Let $q_1 < q_2$. The observation that $\tau_\mu(q)$ is increasing follows since $\mu(B(x_i,r))^{q_1} \geq \mu(B(x_i,r))^{q_2}$ for any $x_i \in \operatorname{supp}\mu$.

In addition, concavity is a standard application of Hölder's inequality: let $0<\lambda<1$, and then with Hölder's inequality with exponents $1/\lambda$ and $1/(1-\lambda)$,

$$\sum_i \mu\big(B(x_i,r)\big)^{\lambda q_1+(1-\lambda)q_2} \leq \left(\sum_i \mu\big(B(x_i,r)\big)^{q_1}\right)^{\lambda} \left(\sum_i \mu\big(B(x_i,r)\big)^{q_2}\right)^{1-\lambda} \tag{2.1}$$

and taking suprema and logarithms yields concavity.

Finally, using the equivalence of packings and coverings, we see that $\tau_\mu(0) = -\dim_{\mathrm{B}}(\operatorname{supp}\mu)$, and since μ is a probability measure there is a constant $c>0$ so that $0<c\leq G_\mu(r,1)\leq 1$ for all $r>0$, so $\tau_\mu(1)=0$. □

In our special setting, we can rewrite the L^q-spectrum essentially as a symbolic sum. For notational simplicity, we will write $\tau_{\boldsymbol{p}}$ in place of $\tau_{\mu_{\boldsymbol{p}}}$, where the dependency on the underlying IFS is implicit. Since $(r_{\gamma|_n})_{n=1}^\infty$ is a strictly decreasing sequence for each $\gamma\in\Omega$, setting

$$\Lambda_r \coloneqq \{\mathtt{i}\in\mathcal{I}^* : r_{\mathtt{i}} \leq r < r_{\mathtt{i}^-}\},$$

we have that Λ_r is a section of $\mathcal{I}^*$ for each $r>0$ (that is, for all n sufficiently large and $\mathtt{i}\in\mathcal{I}^n$, $\mathtt{i}$ has a unique prefix in Λ_r). Heuristically, Λ_r is a symbolic representation of the collection of images $S_{\mathtt{i}}(K)$ with diameter approximately r.

We now have the following immediate application of Lemma 2.2.

Lemma 2.6. *Let $\boldsymbol{p}\in\mathcal{P}$ be arbitrary and $q\in\mathbb{R}$. Then there are constants $c_1,c_2>0$ (depending on q) and a constant $\delta>0$ so that for all $r\in(0,1)$*

$$c_1 G_{\mu_{\boldsymbol{p}}}(\delta r, q) \leq \sum_{\mathtt{i}\in\Lambda_r} p_{\mathtt{i}}^q \leq c_2 G_{\mu_{\boldsymbol{p}}}(\delta r, q).$$

In particular,

$$\tau_{\boldsymbol{p}}(q) = \liminf_{r\to 0} \frac{\log \sum_{\mathtt{i}\in\Lambda_r} p_{\mathtt{i}}^q}{\log r}.$$

Proof. By Lemma 2.2, there are constants $\delta>0$ and $c>0$ so that for each $r\in(0,1)$ and $\mathtt{i}\in\Lambda_r$, there is some $x_{\mathtt{i}}\in S_{\mathtt{i}}(K)$ so that

$$c\cdot p_{\mathtt{i}} \leq \mu\big(B(x_{\mathtt{i}},\delta r)\big) \leq p_{\mathtt{i}} \tag{2.2}$$

and moreover $B(x, \delta r) \cap K \subseteq S_{\mathtt{i}}(K)$, so that $\{x_{\mathtt{i}}\}_{\mathtt{i}\in\Lambda_r}$ is a $2\delta r$-separated subset of K.

On the other hand, there is some $N \in \mathbb{N}$ so that if $\{x_i\}_i$ is any $2\delta r$-separated subset of K, then there is some unique $\mathtt{i}_i \in \Lambda_r$ so that $B(x_i, \delta r) \cap K \subseteq S_{\mathtt{i}_i}(K)$, and some $\mathtt{j}_i \in \mathcal{I}^n$ so that $S_{\mathtt{i}_i} \circ S_{\mathtt{j}_i}(K) \subseteq B(x_i, \delta r)$. Thus with $c' = \min\{p_{\mathtt{j}} : \mathtt{j} \in \mathcal{I}^n\}$,

$$c' p_{\mathtt{i}_i} \leq \mu\big(B(x_i, \delta r)\big) \leq p_{\mathtt{i}_i}. \tag{2.3}$$

The desired inequalities now follow from (2.2) and (2.3). □

In our next result, we will use the method of types from large deviations theory: we refer the reader to the book by Dembo & Zeitouni [DZ10] for more background and detail. To this end, for $n \in \mathbb{N}$, $i \in \mathcal{I}$, and $\mathtt{i} = (i_1, \ldots, i_n) \in \mathcal{I}^n$, let $\xi_i(\mathtt{i})$ denote the *frequency of letter i in $\mathtt{i}$*, that is

$$\xi_i(\mathtt{i}) = \frac{\#\{j : 1 \leq h \leq n, i_j = i\}}{n}. \tag{2.4}$$

We refer to the probability vector $\boldsymbol{\xi}(\mathtt{i}) \coloneqq (\xi_i(\mathtt{i}))_{i\in\mathcal{I}} \in \mathcal{P}(\mathcal{I})$ as the *type* associated with $\mathtt{i}$.

Now given an $r \in (0, 1)$, we let

$$\mathcal{T}_r = \big\{\boldsymbol{\xi}(\mathtt{i}) : \mathtt{i} \in \Lambda_r\big\}$$

denote the set of all types at scale r. Conversely, given a $\boldsymbol{w} \in \mathcal{T}_r$, we let

$$\mathcal{C}_r(\boldsymbol{w}) = \big\{\mathtt{i} \in \Lambda_r : \boldsymbol{\xi}(\mathtt{i}) = \boldsymbol{w}\big\}.$$

The main point of introducing types is that the values $r_{\mathtt{i}}$, $|\mathtt{i}|$, and $p_{\mathtt{i}}$ (for $\boldsymbol{p} \in \mathcal{P}$) are constant on each type class. Moreover, there are few types, and we can determine the size of each type class up to a sub-exponential error. Here, for notational simplicity, we use Landau's asymptotic notation, i.e. given real-valued functions f, g defined on some domain A, we say $f = O(g)$ if there is a constant $C > 0$ so that $|f(a)/g(a)| \leq C$ for all $a \in A$.

Lemma 2.7. *Let $r \in (0, 1)$ be arbitrary. Then*

$$\frac{\#\mathcal{T}_r}{\log r} = O\left(\frac{\log\log(1/r)}{\log(1/r)}\right). \tag{2.5}$$

and, for fixed $\boldsymbol{w} \in \mathcal{T}_r$ with $\boldsymbol{w} = \boldsymbol{\xi}(\mathtt{i})$ for $\mathtt{i} \in \Lambda_r$,

$$\frac{\#\mathcal{C}_r(\boldsymbol{w})}{|\mathtt{i}|} = H(\boldsymbol{w}) + O\left(\frac{\log\log(1/r)}{\log(1/r)}\right). \tag{2.6}$$

Proof. First, since $r_i \in (0, 1)$ for all $i \in \mathcal{I}$, for all $r \in (0, 1)$ and $\mathtt{i} \in \Lambda_r$, $|\mathtt{i}| \asymp \log(1/r)$.

Since the type $\boldsymbol{\xi}(\mathtt{i})$ does not depend on the order of the letters in $\mathtt{i}$, a direct count (see [DZ10, Lemma 2.1.2] for more detail) gives for each $m \in \mathbb{N}$

$$\#\{\boldsymbol{\xi}(\mathtt{i}) : \mathtt{i} \in \mathcal{I}^m\} \leq (m + 1)^{\#\mathcal{I}}.$$

Thus there is a fixed constant $M > 0$ so that

$$\begin{aligned}\#\mathcal{T}_r &\leq \sum_{m=0}^{\lceil M\log(1/r)\rceil} \#\{\boldsymbol{\xi}(\mathtt{i}) : \mathtt{i} \in \mathcal{I}^m\} \\ &\leq \big(M\log(1/r)+1\big)\cdot\big(M\log(1/r)+2\big)^{\#\mathcal{I}}.\end{aligned}$$

Taking logarithms and dividing through by $\log(1/r)$ gives (2.5).

Finally, (2.6) is precisely [DZ10, Lemma 2.1.8], again using the fact that $|\mathtt{i}| \asymp \log(1/r)$. Alternatively, one can directly apply Stirling's formula—that $\log(n!) = n\log n - n + O(\log n)$—and use the observation that the number of distinct permutations of a word $\mathtt{i} = (i_1, \dots, i_n)$ with $k_i = n\cdot\boldsymbol{\xi}(\mathtt{i})$ for $i \in \mathcal{I}$ is $n!\cdot\big(\prod_{i\in\mathcal{I}} k_i!\big)^{-1}$. □

We can now establish our formula for the L^q-spectrum of a self-similar measure.

Theorem 2.8. *Let $\boldsymbol{p} \in \mathcal{P}$. Then the limit defining $\tau_{\boldsymbol{p}}$ exists, and moreover*

$$\tau_{\boldsymbol{p}}(q) = \inf_{\boldsymbol{w}\in\mathcal{P}}\left\{\frac{qH(\boldsymbol{w},\boldsymbol{p}) - H(\boldsymbol{w})}{\chi(\boldsymbol{w})}\right\}.$$

Proof. We fix $\boldsymbol{p} \in \mathcal{P}$ and $q \in \mathbb{R}$. Suppose $r \in (0,1)$, fix $\boldsymbol{w} \in \mathcal{T}_r$, and let m denote the common value of $|\mathtt{i}|$ for $\mathtt{i} \in \mathcal{C}_r(\boldsymbol{w})$. Then for any $\mathtt{i} \in \mathcal{C}_r(\boldsymbol{w})$,

$$p_{\mathtt{i}}^q = \prod_{i\in\mathcal{I}} p_i^{q\cdot n\cdot w_i} \qquad\qquad r_{\mathtt{i}} = \prod_{i\in\mathcal{I}} r_i^{n\cdot w_i}.$$

Therefore

$$\sum_{\mathtt{i}\in\Lambda_r} p_{\mathtt{i}}^q = \sum_{\boldsymbol{w}\in\mathcal{T}_r} \#\mathcal{C}_r(\boldsymbol{w}) \prod_{i\in\mathcal{I}} p_i^{q\cdot n\cdot w_i},$$

and moreover, for fixed $\boldsymbol{w} \in \mathcal{T}_r$, by (2.6) in Lemma 2.7,

$$\frac{\log \#\mathcal{C}_r(\boldsymbol{w}) \prod_{i\in\mathcal{I}} p_i^{q\cdot n\cdot w_i}}{\log r} = \frac{qH(\boldsymbol{w},\boldsymbol{p}) - H(\boldsymbol{w})}{\chi(\boldsymbol{w})} + O\left(\frac{\log\log(1/r)}{\log(1/r)}\right). \tag{2.7}$$

Thus by (2.5), with $\boldsymbol{w}^{(r)} \in \mathcal{T}_r$ chosen to minimize the quantity in (2.7),

$$\frac{\log \sum_{\mathtt{i}\in\Lambda_r} p_{\mathtt{i}}^q}{\log r} = \frac{qH(\boldsymbol{w}^{(r)},\boldsymbol{p}) - H(\boldsymbol{w}^{(r)})}{\chi(\boldsymbol{w}^{(r)})} + O\left(\frac{\log\log(1/r)}{\log(1/r)}\right).$$

But the map $\boldsymbol{w} \mapsto \frac{qH(\boldsymbol{w},\boldsymbol{p})-H(\boldsymbol{w})}{\chi(\boldsymbol{w})}$ is continuous, so by compactness of $\mathcal{P}$ and since $\mathcal{T}_r$ becomes arbitrarily dense in $\mathcal{P}$ as r goes to 0, it follows that the limit defining $\tau_{\boldsymbol{p}}(q)$ in fact exists and is equal to the desired expression. □

2.3. An explicit formula for the L^q-spectrum. We now have a formula for $\tau_{\boldsymbol{p}}(q)$ as the solution to an optimization problem, but it would be nice to have an explicit formula. Let's work out a reasonable guess for the formula for $\tau_{\boldsymbol{p}}(q)$ using Lagrange multipliers, for now sweeping any technical issues under the rug.

The space of probability vectors $\mathcal{P}$ is a subset of the affine hyperplane $\{\boldsymbol{w} \in \mathbb{R}^{\mathcal{I}} : \|\boldsymbol{w}\| = 1\}$. Write

$$\phi(q, \boldsymbol{w}) = qH(\boldsymbol{w}, \boldsymbol{p}) - H(\boldsymbol{w}). \tag{2.8}$$

Suppose $\boldsymbol{w}$ is a minimizing vector for $\tau_{\boldsymbol{p}}(q)$. Informally applying the method of Lagrange multipliers, such a vector $\boldsymbol{w} \in \mathcal{P}$ must satisfy

$$\frac{\partial}{\partial w_i}\frac{\phi(q, \boldsymbol{w})}{\chi(\boldsymbol{w})} = \lambda \frac{\partial}{\partial w_i}(\|\boldsymbol{w}\| - 1) = \lambda$$

for some $\lambda \in \mathbb{R}$. In particular, for each $i \in \mathcal{I}$,

$$\chi(\boldsymbol{w}) \cdot \frac{\partial}{\partial w_i}\phi(q, \boldsymbol{w}) + \phi(q, \boldsymbol{w}) \cdot \frac{\partial}{\partial w_i}\chi(\boldsymbol{w}) = \lambda \cdot \chi(w)^2$$

which, after rearranging and computing the derivatives, yields

$$q \log(1/p_i) + \log w_i + 1 = \lambda \cdot \chi(\boldsymbol{w}) + \tau_{\boldsymbol{p}}(q) \log(1/r_i). \tag{2.9}$$

Multiplying by w_i and summing over all $i \in \mathcal{I}$,

$$\phi(q, \boldsymbol{w}) + 1 = \lambda \cdot \chi(\boldsymbol{w}) + \chi(\boldsymbol{w})\tau_{\boldsymbol{p}}(q).$$

But $\boldsymbol{w}$ is a minimizing vector, so $\phi(q, \boldsymbol{w})/\chi(\boldsymbol{w}) = \tau_{\boldsymbol{p}}(q)$ so that $\lambda = 1/\chi(\boldsymbol{w})$. Substituting this back into (2.9) and rearranging for w_i yields

$$w_i = p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)}.$$

Therefore, we would guess that $\tau_{\boldsymbol{p}}(q)$ must be the solution to the equation

$$\sum_{i \in \mathcal{I}} p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)} = 1.$$

and the minimization is attained at the probability vector $\left(p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)}\right)_{i \in \mathcal{I}}$.

While the derivation above is certainly not rigorous (though it can be made rigorous without too much trouble), it helps to motivate the following formula for the L^q-spectrum which we will prove is the correct formula through a very simple information-theoretic argument. First, define the function

$$\psi(q, t) = \sum_{i \in \mathcal{I}} p_i^q r_i^{-t}.$$

Note that for fixed q, $t \mapsto \psi(q, t)$ is a strictly decreasing function of t with $\lim_{t \to \infty} \psi(q, t) = 0$ and $\lim_{t \to \infty} \psi(q, t) = \infty$. Thus there is a unique value $T(q)$ so that $\psi(q, T(q)) = 1$; and moreover, by the analytic implicit function theorem, $T(q)$ is an analytic function of q. Finally, we define the vector

$$\boldsymbol{z}(q) = \left(p_i^q r_i^{-T(q)}\right)_{i \in \mathcal{I}}$$

which is indeed a probability vector by the definition of $T(q)$. This was the original formula established for the L^q-spectrum by Arbeiter & Patzschke [AP96].

Proposition 2.9 ([AP96]). *Fix $\boldsymbol{p} \in \mathcal{P}$ and $q \in \mathbb{R}$: then $\tau_{\boldsymbol{p}}(q)$ is the unique solution to*

$$\sum_{i \in \mathcal{I}} p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)} = 1. \tag{2.10}$$

In particular, $\tau_{\boldsymbol{p}}$ is analytic on $\mathbb{R}$.

Proof. Observe for $\boldsymbol{w} \in \mathcal{P}$ that

$$\frac{qH(\boldsymbol{w}, \boldsymbol{p}) - H(\boldsymbol{w})}{\chi(\boldsymbol{w})} = T(q) + \frac{D_{\mathrm{KL}}(\boldsymbol{w} \,\|\, \boldsymbol{z}(q))}{\chi(\boldsymbol{w})}.$$

But $D_{\mathrm{KL}}(\boldsymbol{w} \,\|\, \boldsymbol{z}(q)) \geq 0$ with equality if and only if $\boldsymbol{w} = \boldsymbol{z}(q)$, so the minimization is attained uniquely at the probability vector $\boldsymbol{z}(q)$ with value $T(q)$. Thus the result follows by Theorem 2.8. □

2.4. Asymptotes of the L^q-spectra. To conclude this section, let's analyze the asymptotic behaviour of $\tau_{\boldsymbol{p}}(q)$, as well as the limiting values of the corresponding optimization vector, which we recall is given by

$$\boldsymbol{z}(q) = \left(p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)} \right)_{i \in \mathcal{I}}.$$

First, for each $i \in \mathcal{I}$, let κ_i be such that $r_i^{\kappa_i} = p_i$, that is

$$\kappa_i = \frac{\log p_i}{\log r_i}.$$

In particular, $z_i(q) = r_i^{q\kappa_i - \tau_{\boldsymbol{p}}(q)}$. But for $q > 0$, $q\kappa_i - \tau_{\boldsymbol{p}}(q) \geq q\kappa_{\min} - \tau_{\boldsymbol{p}}(q)$, and the $z_i(q)$ cannot all converge to 0 as q diverges to infinity. The same argument holds for $q < 0$, so that $\lim_{q \to \infty} z_i(q) > 0$ if and only if $\kappa_i = \kappa_{\min}$, and $\lim_{q \to -\infty} z_i(q) > 0$ if and only if $\kappa_i = \kappa_{\max}$. Thus the limits $\boldsymbol{z}(\pm\infty) \coloneqq \lim_{q \to \pm\infty} \boldsymbol{z}(q)$ exist and are given by

$$z_i(-\infty) = \begin{cases} 0 & : \kappa_i \neq \kappa_{\max} \\ r_i^{s_{\max}} & : \kappa_i = \kappa_{\max} \end{cases} \qquad \text{where} \qquad \sum_{\substack{i \in \mathcal{I} \\ \kappa_i = \kappa_{\max}}} r_i^{s_{\max}} = 1.$$

and similarly

$$z_i(\infty) = \begin{cases} 0 & : \kappa_i \neq \kappa_{\min} \\ r_i^{s_{\min}} & : \kappa_i = \kappa_{\min} \end{cases} \qquad \text{where} \qquad \sum_{\substack{i \in \mathcal{I} \\ \kappa_i = \kappa_{\min}}} r_i^{s_{\min}} = 1.$$

Equivalently,

$$s_{\max} = \lim_{q \to -\infty} (q\kappa_i - \tau_{\boldsymbol{p}}(q)) \qquad \text{and} \qquad s_{\min} = \lim_{q \to \infty} (q\kappa_i - \tau_{\boldsymbol{p}}(q)).$$

This implies that $\tau_{\boldsymbol{p}}(q)$ has asymptotes given by

$$\ell_{-\infty}(q) = q\kappa_{\max} - s_{\max} \qquad \text{and} \qquad \ell_{\infty}(q) = q\kappa_{\min} - s_{\min}.$$

A general depiction of the function $\tau_{\boldsymbol{p}}(q)$ along with its asymptotes is given in Figure 1.

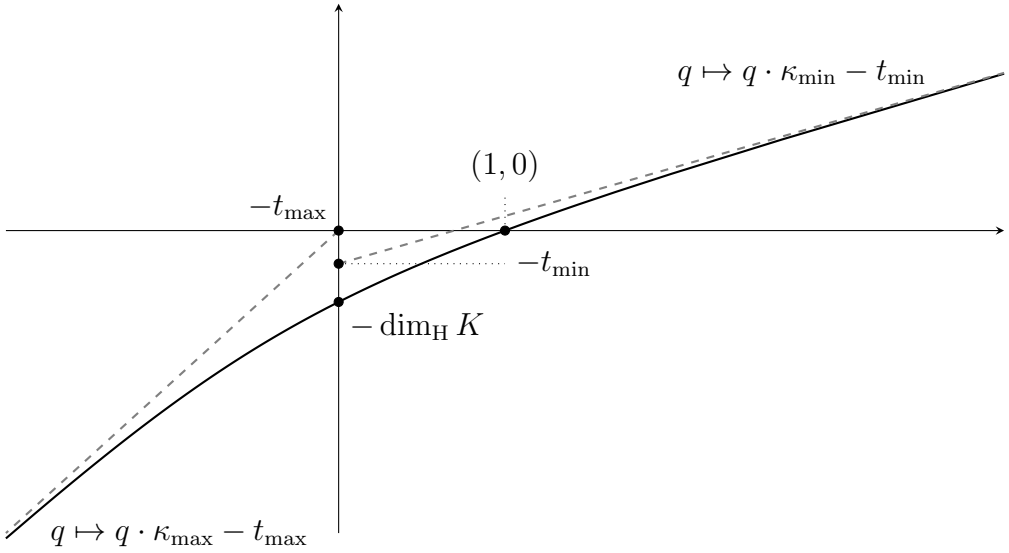

FIGURE 1. Plot of the L^q-spectrum of the measure $\mu_{\boldsymbol{p}}$, along with its asymptotes.

3. LAGRANGE DUALITY AND MULTIFRACTAL FORMALISM

Before proceeding with the remaining results of this section, let us summarize the current situation. We are attempting to obtain a formula for the multifractal spectrum $f_{\boldsymbol{p}}(\alpha)$, for which we have the variational lower bound

$$f_{\boldsymbol{p}}(\alpha) \geq \sup_{\boldsymbol{w} \in \mathcal{P}} \{\dim_{\mathrm{H}} \mu_{\boldsymbol{w}} : \dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, x) = \alpha \text{ for } \mu_{\boldsymbol{p}}\text{-a.e. } x\}.$$

By Proposition 2.4, the optimization on the right hand side is precisely

$$\sup_{\boldsymbol{w} \in \mathcal{P}} \left\{ \frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})} : \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})} = \alpha \right\}. \tag{3.1}$$

In order to understand this optimization problem, taking motivation from the method of Lagrange multipliers, we want to study the *Lagrange dual*, which is optimization problem

$$\inf_{\boldsymbol{w} \in \mathcal{P}} \left\{ \frac{q \cdot H(\boldsymbol{w}, \boldsymbol{p}) - H(\boldsymbol{w})}{\chi(\boldsymbol{w})} \right\}. \tag{3.2}$$

But this optimization problem is precisely the formula for the L^q-spectrum of $\mu_{\boldsymbol{p}}$, as proven in Theorem 2.8! Therefore, in order to complete the proof, we need two main ingredients:

1. We need to understand the relationship between the dual optimization problems (3.1) and (3.2).
2. We need to obtain bounds on $f_{\boldsymbol{p}}(\alpha)$ by using the L^q-spectrum of $\mu_{\boldsymbol{p}}$.

These two ingredients are handled in the subsequent sections. In §3.1, we address the duality in 1 in a general context—real-valued continuous functions on a compact Hausdorff topological space. Then in §3.2 we study the relationship between $f_{\boldsymbol{p}}$ and $\tau^*_{\boldsymbol{p}}$: in fact, we will prove that $f_{\boldsymbol{p}}(\alpha) \leq \tau^*_{\boldsymbol{p}}(\alpha)$ in general.

Finally, we provide an alternative proof of the variational formula which avoids the machinery of duality in §3.4. This proof is conceptually useful for different reasons, since it also immediately generalizes to different types of level sets of local dimensions.

3.1. Continuous optimization and duality. Suppose Δ is a compact Hausdorff topological space and suppose we are given a continuous function $u\colon \Delta \to \mathbb{R}$ and an upper semicontinuous function $v\colon \Delta \to \mathbb{R}$. We consider the constrained optimization

$$f(\alpha) = \max_{\boldsymbol{w}\in\Delta}\big\{v(\boldsymbol{w}) : u(\boldsymbol{w}) = \alpha\big\}$$

with corresponding unconstrained dual

$$\tau(q) = \min_{\boldsymbol{w}\in\Delta}\big\{q\cdot u(\boldsymbol{w}) - v(\boldsymbol{w})\big\}. \tag{3.3}$$

Here, the maximum over the empty set is $-\infty$. Note that both the maximum and minimum are attained on compact sets since v is upper semicontinuous. Of course, τ is a concave function of q since it is an infimum of affine functions. On the other hand, f need not be concave.

In some sense, one can think of the function $\tau(q)$ as encoding the geometry of the Lagrange multiplier problem associated with $f(\alpha)$. However, this is only a motivating heuristic since in our abstract setup, there is no differentiable structure in sight.

Before we continue, let's recall some basic facts from convex optimization. For a more in-depth introduction, we refer the reader to [Roc70]. First, for a general function $g\colon \mathbb{R} \to \mathbb{R}\cup\{-\infty\}$, we denote the *concave conjugate* by

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

Note that g^* is always concave, and g^{**} is the concave hull of g.

Now suppose moreover that $g\colon \mathbb{R} \to \mathbb{R}\cup\{-\infty\}$ is a concave function. Then for $q \in \mathbb{R}$, we write $\partial g(q)$ to denote the *subdifferential* of g at q, i.e.

$$\partial g(q) = \{\alpha : \alpha(y-q) + g(q) \geq g(y) \text{ for any } y \in \mathbb{R}\}.$$

Equivalently,

$$g^*(\alpha) + g(q) \leq \alpha q \tag{3.4}$$

with equality if and only if $\alpha \in \partial g(q)$. We also let $\partial^- g(q)$ (resp. $\partial^+ g(q)$) denote the left (resp. right) derivative of g of q. Then $\partial g(q) = [\partial^+ g(q), \partial^- g(q)]$. In particular, g is differentiable at q if and only if $\partial g(q) = \{\alpha\}$, in which case $g'(q) = \alpha$. Since the subdifferentials form an ordered family of intervals of $\mathbb{R}$ which overlap only on their endpoints, there can be at most countably many points with non-singleton subdifferential: in particular, g is differentiable at all but countably many $q \in \mathbb{R}$.

Now, we say that a line

$$\ell(t) = a \cdot t + b$$

is a *supporting line* for τ at q if $\tau(q) = \ell(q)$ and $\ell(t) \geq \tau(t)$ for all $t \in \mathbb{R}$. Equivalently, $\partial\tau(q)$ is precisely the set of possible slopes of supporting lines for τ at q.

Now consider specifically the function τ from (3.3). We say that τ is *supported* at (q, α) for $q \in \mathbb{R}$ if there is a $\boldsymbol{w} \in \Delta$ so that $\tau(q) = q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$ and $u(\boldsymbol{w}) = \alpha$. Equivalently, the line $t \mapsto t \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$ is a supporting line for τ at q with slope α. Therefore, the problem of determining the values of α for which τ is supported at (q, α) is precisely the problem of determining the slopes of supporting lines which appear from the minimization defining $\tau(q)$.

For the remainder of this section, we establish some elementary facts concerning the dual problems $\tau(q)$ and $f(\alpha)$. We begin with the following basic fact about supports of the function τ.

Lemma 3.1. *For any $q \in \mathbb{R}$ and $\alpha \in \{\partial^-\tau(q), \partial^+\tau(q)\}$, τ is supported at (q, α).*

Proof. Let $(q_n)_{n=1}^{\infty}$ be a sequence converging to q monotonically from the left. For each $n \in \mathbb{N}$, write $\tau(q_n) = q_n \cdot u(\boldsymbol{w}_n) - v(\boldsymbol{w}_n)$ for some $\boldsymbol{w}_n \in \Delta$. By compactness of Δ, passing to a subsequence if necessary, we may assume that $\lim_{n\to\infty} \boldsymbol{w}_n = \boldsymbol{w}$. Next, observe that the line

$$\ell_n(t) \coloneqq t \cdot u(\boldsymbol{w}_n) - v(\boldsymbol{w}_n)$$

is a supporting line for τ at q_n and therefore $u(\boldsymbol{w}_n) \in \partial\tau(q_n)$. Moreover, since $\lim_{n\to\infty} \partial^-\tau(q_n) = \partial^-\tau(q)$, it follows that

$$u(\boldsymbol{w}) = \lim_{n\to\infty} u(\boldsymbol{w}_n) = \partial\tau^-(q).$$

But then by upper semicontinuity of v and continuity of τ,

$$\tau(q) \leq q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w}) \leq \lim_{n\to\infty} (q \cdot u(\boldsymbol{w}_n) - v(\boldsymbol{w}_n)) = \lim_{n\to\infty} \tau(q_n) = \tau(q)$$

so that equality holds and τ is supported at $(q, \partial^-\tau(q))$. The same argument works for $\partial^+\tau(q)$, giving the result. □

Remark 3.2. Since u is continuous and Δ is compact, it follows that the left and right derivatives are uniformly bounded away from $\pm\infty$.

Using Lemma 3.1, we can now characterize concavity of the function f.

Proposition 3.3. *For any $\alpha \in \mathbb{R}$, $f(\alpha) \leq \tau^*(\alpha)$, and if τ is supported at (q, α) for some $q \in \mathbb{R}$, then $f(\alpha) = q\alpha - \tau(q) = \tau^*(\alpha)$. Moreover, the following are equivalent:*

(i) For all $q \in \mathbb{R}$ and $\alpha \in \partial\tau(q)$, τ is supported at (q, α).
(ii) $f(\alpha) = \tau^*(\alpha)$ for all $\alpha \in \mathbb{R}$.
(iii) f is a concave function.

Proof. First, let $\alpha \in \mathbb{R}$ be arbitrary. If $f(\alpha) = -\infty$, we are done; otherwise, since Δ is compact and u and v are continuous, there is some $\boldsymbol{w} \in \Delta$ so that $u(\boldsymbol{w}) = \alpha$ and $v(\boldsymbol{w}) = f(\alpha)$. Then for any $q \in \mathbb{R}$,

$$\tau(q) \leq q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w}) = q\alpha - f(\alpha).$$

But q was arbitrary, so $f(\alpha) \leq \tau^*(\alpha)$.

Next, if τ is supported at (q, α) for $q \in \mathbb{R}$, get $\boldsymbol{w}$ so that $u(\boldsymbol{w}) = \alpha$ and $\tau(q) = q\alpha - v(\boldsymbol{w})$. But then

$$\tau^*(\alpha) \geq f(\alpha) \geq v(\boldsymbol{w}) = q\alpha - \tau(q) \geq \tau^*(\alpha).$$

as claimed.

Now, (i) implies (ii) was proven above, and (ii) immediately implies (iii). It remains to verify that (iii) implies (i). Let $q \in \mathbb{R}$. Then if $\alpha \in \partial\tau(q)$, write $\alpha = \lambda\partial^+\tau(q) + (1-\lambda)\partial^-\tau(q)$ for some $\lambda \in [0,1]$. Then by concavity of f and Lemma 3.1,

$$\begin{aligned}
\tau^*(\alpha) &\geq f(\alpha) \\
&\geq \lambda f\big(\partial^+\tau(q)\big) + (1-\lambda) f\big(\partial^-\tau(q)\big) \\
&= \lambda(q\partial^+\tau(q) - \tau(q)) + (1-\lambda)(q\partial^-\tau(q) - \tau(q)) \\
&= q\alpha - \tau(q) \\
&\geq \tau^*(\alpha).
\end{aligned}$$

Thus all the inequalities are in fact equalities. In particular, taking $\boldsymbol{w}$ so that $u(\boldsymbol{w}) = \alpha$ and $f(\alpha) = v(\boldsymbol{w})$, substituting this into the previous equation implies that $v(\boldsymbol{w}) = q \cdot u(\boldsymbol{w}) - \tau(q)$, as required. □

To round off our results concerning the concave conjugate relationship, let's describe what happens as q diverges to $\pm\infty$. We first extend the derivatives to $\pm\infty$ by defining

$$\partial^-\tau(\infty) = \lim_{q\to\infty} \frac{\tau(q)}{q} \qquad \text{and} \qquad \partial^+\tau(-\infty) = \lim_{q\to-\infty} \frac{\tau(q)}{q}.$$

That these limits exist and take finite values follows by monotonicity of the partial derivatives combined with the observation in Remark 3.2. Thus τ has affine asymptotes at $\pm\infty$ with slopes $\partial^-\tau(\infty)$ and $\partial^+\tau(-\infty)$. We then say that τ is supported at $(\infty, \partial^-\tau(\infty))$ if there is a $\boldsymbol{w}$ so that the line $q \mapsto q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$ is the affine asymptote of τ at ∞. The same definition holds at $-\infty$ as well. Then, the same proof as Lemma 3.1 (along a sequence $(q_n)_{n=1}^\infty$ diverging to $\pm\infty$) combined with a modified Proposition 3.3 yields the following result.

Proposition 3.4. *The function τ is supported at $(\infty, \partial^+\tau(-\infty))$ and $(-\infty, \partial^-\tau(\infty))$. In particular, $\tau^*(\partial^+\tau(-\infty)) = f(\partial^+\tau(-\infty))$ and $\tau^*(\partial^-\tau(\infty)) = f(\partial^-\tau(\infty))$.*

We conclude this section with two explicit situations in which we can establish the concave conjugate relationship. The first situation, even without any knowledge of the underlying optimization, occurs when $\tau(q)$ is differentiable.

Corollary 3.5. *If $q \in \mathbb{R}$ is such that $\alpha = \tau'(q)$ exists, then $f(\alpha) = \tau^*(\alpha)$. In particular, $\tau(q) = f^*(q)$ for all $q \in \mathbb{R}$.*

Proof. If $q \in \mathbb{R}$ and $\tau'(q) = \alpha$ exists, by Lemma 3.1, τ is supported at q. Thus by Proposition 3.3, $f(\alpha) = \tau^*(\alpha)$. In particular, $\tau(q) = f^*(q)$ for all $q \in \mathbb{R}$ for which $\tau'(q)$ exists. But f^* and τ are both continuous functions and $\tau'(q)$ exists for a dense set of $q \in \mathbb{R}$, so in fact $\tau(q) = f^*(q)$ for all $q \in \mathbb{R}$. □

As our second (and final) application, we can also use information about the structure of the set on which the optimization is attained to abstractly establish the concave conjugate relationship. For each $q \in \mathbb{R}$, denote the set of minimizers by

$$M(q) = \{\boldsymbol{w} \in \Delta : q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w}) = \tau(q)\}.$$

Since u is continuous and v is upper semicontinuous, $M(q)$ is a compact set for all q. Moreover, we obtain the following result.

Corollary 3.6. *Suppose $q \in \mathbb{R}$ and $M(q)$ is connected. Then $f(\alpha) = \tau^*(\alpha)$ for all $\alpha \in \partial\tau(q)$. Moreover, if $M(q)$ is a singleton, then τ is differentiable at q.*

Proof. By Lemma 3.1, τ is supported at $(q, \partial^-\tau(q))$ and $(q, \partial^+\tau(q))$. Thus get $\boldsymbol{w}_-, \boldsymbol{w}_+ \in M(q)$ so that

$$u(\boldsymbol{w}_-) = \partial^-\tau(q) \qquad \text{and} \qquad u(\boldsymbol{w}_+) = \partial^+\tau(q).$$

Since $M(q)$ is connected and u is continuous, $u(M(q)) \subset \mathbb{R}$ is an interval containing $\partial^-\tau(q)$ and $\partial^+\tau(q)$, and therefore $\partial\tau(q) \subset u(M(q))$. In particular, for any $\alpha \in \partial\tau(q)$, there is a $\boldsymbol{w} \in M(q)$ so that $u(\boldsymbol{w}) = \alpha$, so that τ is supported at (q, α) and therefore $f(\alpha) = \tau^*(\alpha)$ by Proposition 3.3.

If moreover $M(q)$ is a singleton, then we must have $\boldsymbol{w}_- = \boldsymbol{w}_+$, forcing $\partial^-\tau(q) = \partial^+\tau(q)$ so that τ is differentiable at q. □

3.2. Multifractal spectrum and general upper bound.

For $\alpha \in \mathbb{R}$, let

$$F_\mu(\alpha) = \{x \in \operatorname{supp}\mu : \dim_{\mathrm{loc}}(\mu, x) = \alpha\}$$

and we define the *multifractal spectrum* $f_\mu \colon \mathbb{R} \to \{-\infty\} \cup [0, d]$ by

$$f_\mu(\alpha) = \dim_{\mathrm{H}} F_\mu(\alpha)$$

using the convention that $\dim_{\mathrm{H}} \varnothing = -\infty$. We say that the measure μ satisfies the *multifractal formalism* if

$$f_\mu(\alpha) = \tau_\mu^*(\alpha)$$

for all $\alpha \in \mathbb{R}$.

We first establish the general upper bound for the multifractal spectrum from the L^q-spectrum. Such a bound has been known for a long time; see, for example, [LN99, Theorem 4.1]. However, many of the proofs in the literature do not precisely address the cases when α corresponds to a slope of the asymptote of τ_μ. We clarify this case explicitly in the following result.

Proposition 3.7. *Let μ be a compactly supported Borel probability measure. Then*

$$f_\mu(\alpha) \leq \tau_\mu^*(\alpha)$$

for all $\alpha \in \mathbb{R}$.

Proof. For $\alpha \in \mathbb{R}$, $r \in (0,1)$ and $\epsilon > 0$, let

$$\mathcal{M}_{r,\epsilon}(\alpha) = \Big\{ x \in \operatorname{supp}\mu : r^{\alpha+\epsilon} \leq \mu\big(B(x,r)\big) \leq r^{\alpha-\epsilon} \Big\}.$$

Our strategy is to control the size of packings with centres in $\mathcal{M}_{r,\epsilon}(\alpha)$ using the L^q-spectrum of μ, to which we can then apply the Hausdorff dimension version of the Vitali covering theorem (see [Fal85, Theorem 1.10]) to obtain bounds on $f_\mu(\alpha)$.

First, suppose we fix a disjoint family of balls $\{B(x_i,r)\}_{i=1}^N$ where $x_i \in \mathcal{M}_{r,\epsilon}$. Suppose $q \geq 0$ is arbitrary. Then

$$G_\mu(r,q) \geq \sum_{i=1}^N \mu\big(B(x_i,r)\big)^q \geq N \cdot r^{q(\alpha+\epsilon)}. \tag{3.5}$$

Moreover, since $\tau_\mu(q) = \liminf_{r\to 0}(\log G_\mu(r,q))/(\log r)$, there is some $r_\epsilon \in (0,1)$ so that for all $r \leq r_\epsilon$, $G_\mu(r,q) \leq r^{\tau_\mu(q)-\epsilon}$. Combining this with (3.5),

$$\frac{\log N}{\log(1/r)} \leq q\alpha - \tau_\mu(q) + (|q|+1)\epsilon \tag{3.6}$$

for all $r \geq r_\epsilon$. The same argument for $q < 0$ also yields (3.6) using the bound $\mu(B(x_i,r)) \leq r^{\alpha-\epsilon}$.

Now fix $\alpha \in \mathbb{R}$ and let $\zeta > 0$ be arbitrary: we will show for all r sufficiently small (depending only on ζ and α) that

$$\frac{\log N}{\log(1/r)} \leq \tau_\mu^*(\alpha) + \zeta. \tag{3.7}$$

To do this, we must handle a few different cases depending on the choice of α. Recall that $\tau_\mu(q)$ is concave with asymptotes at $\pm\infty$ given respectively by

$$q \mapsto q \cdot a_{-\infty} - \tau_\mu^*(a_{-\infty}) \qquad \text{and} \qquad q \mapsto q \cdot a_\infty - \tau_\mu^*(a_\infty).$$

If $a_\infty < \alpha < a_{-\infty}$, there is a $q \in \mathbb{R}$ so that $q \in \tau_\mu^*(\alpha)$ so that $q\alpha - \tau_\mu(q) = \tau^*(\alpha)$. Taking ϵ sufficiently small so that $(|q|+1)\epsilon \leq \zeta$ then yields the claim. Next, if $\alpha = a_\infty$ (the case $\alpha = a_{-\infty}$ is analogous), since

$$\lim_{q\to\infty}(q\alpha - \tau(q)) = \tau_\mu^*(\alpha),$$

first get q so that $q\alpha - \tau(q) \leq \tau_\mu^* + \zeta/2$, and then get $\epsilon > 0$ so that $(|q|+1)\epsilon \leq \zeta/2$. Combining this bounds also yields the claim. Finally, if $\alpha < -a_\infty$ or $\alpha > a_\infty$, taking an infimum over all $q \in \mathbb{R}$ yields $N = 0$, for ϵ sufficiently small depending on α so in fact $\mathcal{M}_{r,\epsilon}(\alpha) = \varnothing$ for all ϵ sufficiently small and r sufficiently small (depending on ϵ).

Now for each $x \in F_\mu(\alpha)$, we can find some $N_x \in \mathbb{N}$ so that for all $n \geq N_x$, $\mu\big(B(x, 2^{-n})\big) \geq 2^{-n(\alpha+\epsilon)}$. In particular, for any $N \in \mathbb{N}$,

$$\mathcal{G}_\epsilon := \bigcup_{n=N}^{\infty} \mathcal{M}_{2^{-n},\epsilon}(\alpha)$$

is a Vitali cover for $F_\mu(\alpha)$.

To conclude the proof, fix $\alpha \in \mathbb{R}$, let $\xi > 0$ be arbitrary and get $M \in \mathbb{N}$ so that (3.7) holds for all for all $r \leq 2^{-M}$. Suppose $\{B(x_j, 2^{-n_j})\}_{j=1}^{\infty}$ is a disjoint subcollection of $\mathcal{G}_\epsilon$: then with $s = \tau^*(\alpha) + 2\xi$,

$$\sum_{j=1}^{\infty} 2^{-n_j s} \leq \sum_{n=M}^{\infty} 2^{-ns} 2^{-n(\tau^*(\alpha)+\xi} = \sum_{n=N_\epsilon}^{\infty} (2^{-\xi})^n < \infty$$

by (3.7). Thus by the Vitali covering theorem for Hausdorff measure, there is a cover $\{E_i\}_{i=1}^{\infty}$ for $F_\mu(\alpha)$ such that

$$\mathcal{H}^s(F_\mu(\alpha)) \leq \sum_{i=1}^{\infty} (\operatorname{diam} E_i)^s < \infty$$

and thus $f_\mu(\alpha) \leq \tau^*(\alpha) + 2\xi$. Since $\xi > 0$ was arbitrary, the result follows. □

Remark 3.8. Why is this bound not sharp in general? In the proof, we are controlling the number of balls at resolution r which have measure larger (or smaller) than the expected quantity. However, the set of points with local dimension α might always be hard to cover at some fixed scale r, even though some parts of the set are easier to cover at some scales than others. This is a similar phenomenon to sets for which the Hausdorff dimension and lower box dimension differ—which is precisely the case here when $q = 0$.

In fact, one might think of the concave conjugate $\tau_\mu^*(\alpha)$ as a *geometric large deviations bound* for the multifractal spectrum. Instead of bounding the set of points with non-typical scaling, the L^q-spectrum bounds the asymptotic size of the set of points with non-typical scaling at a fixed scale r, as r converges to zero. This analogy is particularly relevant for a measure with dynamical origin (this is highlighted, for instance, in the proof of Theorem 2.8).

3.3. Multifractal formalism for self-similar measures. We can now establish the multifractal formalism for self-similar measures satisfying the SSC by combining the various tools that we introduced in the previous sections. Again, for short, given a self-similar measure $\mu_{\boldsymbol{p}}$ we write $f_{\mu_{\boldsymbol{p}}} = f_{\boldsymbol{p}}$. A depiction of the multifractal spectrum $f_{\boldsymbol{p}}$ is given in Figure 2, with the same parameters from Figure 1. The formula for the multifractal spectrum $f_{\boldsymbol{p}}(\alpha)$ is originally due to Cawley & Mauldin [CM92], and the relationship with the L^q-spectrum established by Arbeiter & Patzschke [AP96]

Theorem 3.9 ([AP96; CM92]). *Let $\{S_i\}_{i\in\mathcal{I}}$ be an IFS satisfying the SSC and let $\boldsymbol{p} \in \mathcal{P}$. Then $\mu_{\boldsymbol{p}}$ satisfies the multifractal formalism.*

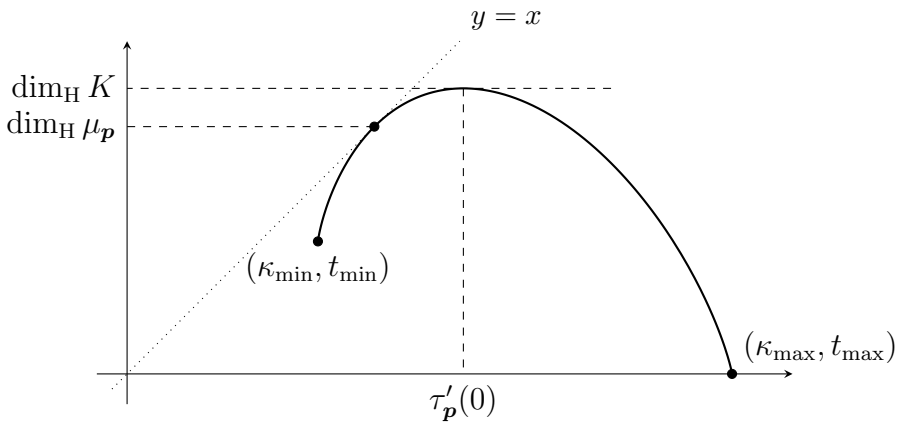

FIGURE 2. Plot of the multifractal spectrum $f_{\boldsymbol{p}}(\alpha)$.

Proof. A direct application of Proposition 2.4 gives that $f_{\boldsymbol{p}} \geq f$ where

$$f(\alpha) = \sup_{\boldsymbol{w} \in \mathcal{P}} \left\{ \frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})} : \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})} = \alpha \right\}.$$

Moreover, by Theorem 2.8 and Corollary 3.5 combined with Proposition 2.9, $f = \tau^*_{\boldsymbol{p}}$. Finally, $f_{\boldsymbol{p}} \leq \tau^*_{\boldsymbol{p}}$ holds in general by Proposition 3.7, giving the desired result. □

Remark 3.10. Alternatively, one can use Proposition 3.15 in the following section, in place of the general upper bound from Proposition 3.7. In particular, if one only wishes to obtain the dimensional formula for $f_{\boldsymbol{p}}(\alpha)$, this provides a method to entirely bypass the results concerning L^q-spectra.

Remark 3.11. In our proof of the multifractal formalism, we obtain a formula for $f_{\boldsymbol{p}}$ which also includes the endpoints. Many proofs of the multifractal formalism either ignore the behaviour at the endpoint entirely, or handle it by explicit arguments which are qualitatively different than the main proof of the multifractal formalism. In contrast, this optimization technique gives the result at the endpoint for free: the endpoint behaviour is simply attained as the limit (in the compact space $\mathcal{P}$) of the optimizing vector for the L^q-spectrum. The geometry of this parametrization is discussed in more detail in §4.

3.4. Alternative proof of the variational principle. In this section, we prove a general covering theorem for the Hausdorff dimension of sets. This bound is similar to the general upper bound for the Hausdorff dimension using measures (see, for example, [Fal97, Proposition 10.1]), though we will state and prove an extension which permits the choice of measure to depend on the point in the set. This additional flexibility will allow us to provide an alternative proof of the variational formula for the level sets of local dimensions.

Definition 3.12. Let $K \subset \mathbb{R}^d$ be an arbitrary compact set and Δ a set of Borel probability measures with $\operatorname{supp} \mu \subset K$ for all $\mu \in \Delta$. We say that Δ has *uniform densities* if for every $\epsilon > 0$, there is a set of Borel measures $\mathcal{E}$ on $\mathbb{R}^d$ with $\sum_{\nu \in \mathcal{E}} \nu(\mathbb{R}^d) < \infty$

and constants $C > 0$, $\eta > 0$ so that for all $\mu \in \Delta$, $r \in (0, \eta)$, and $x \in K$, there is a $\nu = \nu_{\mu,x,r} \in \mathcal{E}$ so that

$$\frac{\log \nu\big(B(x, Cr)\big)}{\log r} \leq \frac{\log \mu\big(B(x, r)\big)}{\log r} + \epsilon.$$

Intuitively, having uniform densities combines a uniform semi-continuity condition on the local dimensions of measures in Δ with a pre-compactness condition on Δ. We emphasize that we do not require the measures in $\mathcal{E}$ to belong to Δ.

A trivial example is any finite set of compactly supported Borel probability measures. Less trivially, and more usefully for us, the set of Bernoulli measures associated with a self-similar IFS (not necessarily satisfying the SSC) also has uniform densities.

Proposition 3.13. *Let $\{S_i\}_{i\in\mathcal{I}}$ be a self-similar IFS with attractor K. Let $\Delta \subset \mathcal{P}$ be arbitrary. Then $\{\mu_{\boldsymbol{p}} : \boldsymbol{p} \in \Delta\}$ has uniform densities.*

Proof. Let $\epsilon > 0$ be arbitrary. First, we construct the family of measures $\mathcal{E}$ by perturbing each $\boldsymbol{p} \in \Delta$ away from the boundary of the simplex $\mathcal{P}$. Let δ be chosen so that $\max\{r_i^\epsilon : i \in \mathcal{I}\} < \delta < 1$, and set

$$\beta = \max\left\{\frac{1}{\#\mathcal{I}}, \frac{\delta}{\#\mathcal{I}^2}\right\}.$$

Then given $\boldsymbol{z} \in \mathcal{P}$ with $z_i \geq \beta$ for all $i \in \mathcal{I}$, write

$$U(\boldsymbol{z}) \coloneqq \{\boldsymbol{p} \in \mathcal{P} : z_i > p_i r_i^\epsilon\}.$$

Note that $U(\boldsymbol{z})$ is an open subset of $\mathcal{P}$. Moreover, suppose $\boldsymbol{p} \in \mathcal{P}$ is arbitrary. Partition $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$ where $\mathcal{I}_1 = \{i \in \mathcal{I} : p_i \geq 1/\#\mathcal{I}\}$. Let $t = \sum_{i\in\mathcal{I}_1} p_i(1-\delta)$ and note that $t \geq \delta/\#\mathcal{I} > 0$. We then define $\boldsymbol{z} = (z_i)_{i\in\mathcal{I}} \in \mathcal{P}$ by the rule

$$z_i = \begin{cases} \delta p_i : i \in \mathcal{I}_1 \\ p_i + \frac{t}{\#\mathcal{I}_2} : i \in \mathcal{I}_2 \end{cases}.$$

Note that $z_i \geq \beta$ for all $i \in \mathcal{I}$ and moreover $z_i \geq \delta p_i > p_i r_i^\epsilon$. In particular, $\{U(\boldsymbol{z}) : \boldsymbol{z} \in \mathcal{P}, z_i \geq \beta \text{ for all } i \in \mathcal{I}\}$ is an open cover for $\mathcal{P}$, and therefore has a finite subcover $\{U(\boldsymbol{z}_1), \ldots, U(\boldsymbol{z}_m)\}$. Let

$$\mathcal{E} = \{\mu_{\boldsymbol{z}_1}, \ldots, \mu_{\boldsymbol{z}_m}\}.$$

We verify that $\mathcal{E}$ has the desired properties.

First, for $0 < r < 1$ and $x \in K$, write

$$\Lambda_r = \big\{\mathtt{i} = (i_1, \ldots, i_n) \in \mathcal{I}^* : n \geq 1, r_{i_1} \cdots r_{i_n} \leq r < r_{i_1} \cdots r_{i_{n-1}}\big\}$$
$$\Lambda_r(x) = \{\mathtt{i} \in \Lambda_r : S_{\mathtt{i}}(K) \cap B(x, r) \neq \emptyset\}.$$

Observe that for all $\mathtt{i} \in \Lambda_r(x)$, with $C \coloneqq \operatorname{diam} K + 1$,

$$S_{\mathtt{i}}(K) \subset B\big(x, Cr\big). \tag{3.8}$$

Next, let $\eta > 0$ be sufficiently small so that for $0 < r < \eta$,

$$r^{\epsilon} \leq \min\{r_i^{\epsilon} : i \in \mathcal{I}\} \tag{3.9}$$

Finally, let $\boldsymbol{p} \in \mathcal{P}$ and let $\boldsymbol{z} \in \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_m\}$ be chosen so that $\boldsymbol{p} \in U(\boldsymbol{z})$. Let $0 < r < \eta$ and $x \in K$ be arbitrary. Then

$$\begin{aligned} r^{2\epsilon}\mu_{\boldsymbol{p}}\big(B(x,r)\big) &\leq \sum_{\mathtt{i} \in \Lambda_r(x)} r^{2\epsilon} p_{\mathtt{i}} \\ &\leq \sum_{\mathtt{i} \in \Lambda_r(x)} r_{\mathtt{i}}^{\epsilon} p_{\mathtt{i}} && \text{by (3.9)} \\ &\leq \sum_{\mathtt{i} \in \Lambda_r(x)} z_{\mathtt{i}} && \text{since } \boldsymbol{p} \in U(\boldsymbol{z}) \\ &\leq \mu_{\boldsymbol{z}}\big(B(x, Cr)\big) && \text{by (3.8)} \end{aligned}$$

where $\epsilon > 0$ is arbitrary, so the desired claim follows. □

In the following result, the case where Δ is a singleton is standard: the additional flexibility where the measure μ can be chosen to depend on the point is useful in applications.

Proposition 3.14. *Let $K \subset \mathbb{R}^d$ be a compact set and let Δ be a family of measures on K with uniform densities. Suppose t is such that for all $x \in K$, there is a $\mu_x \in \Delta$ so that*

$$\liminf_{r \to 0} \frac{\log \mu_x\big(B(x,r)\big)}{\log r} \leq t.$$

Then $\dim_{\mathrm{H}} K \leq t$.

Proof. Fix $\epsilon > 0$ and $\eta > 0$. For each $x \in K$, by assumption, there is a measure $\mu_x \in \Delta$ and a ball $B(x, r_x)$ with $r_x \in (0, \eta)$ such that

$$r_x^{t+\epsilon} \leq \mu_x\big(B(x, r_x)\big) \tag{3.10}$$

Next, since Δ has uniform densities, there is a constant $C > 0$ and family $\mathcal{E}$ of finite Borel probability measures on $\mathbb{R}^d$ with $\sum_{\nu \in \mathcal{E}} \nu(\mathbb{R}^d) < \infty$ so that, taking η to be sufficiently small, for any $\mu \in \Delta$, $x \in K$ and $r \in (0, \eta)$, there is a $\nu = \nu_{\mu,x,r} \in \mathcal{E}$ so that

$$r^{\epsilon}\mu\big(B(x,r)\big) \leq \nu\big(B(x, Cr)\big). \tag{3.11}$$

Let $\mathcal{B} = \{B(x, Cr_x)\}_{x \in K}$. By the Besicovitch covering theorem (see [Mat95, Theorem 2.7]), there is a constant c_d depending only on the ambient dimension d and families of balls $\mathcal{B}_i \subset \mathcal{B}$ for $i = 1, \dots, c_d$ such that

$$K \subset \bigcup_{i=1}^{c_d} \bigcup_{B \in \mathcal{B}_i} B$$

and moreover the balls in each $\mathcal{B}_i$ are disjoint. Thus for each $i = 1, \dots, c_d$ and $B(x,r) \in \mathcal{B}_i$, by (3.10) combined with (3.11), there is a $\nu \in \mathcal{E}$ so that

$$r^{t+2\epsilon} \leq \nu\big(B(x, Cr)\big).$$

But the balls in each $\mathcal{B}_i$ are pairwise disjoint, so

$$\begin{aligned}\sum_{i=1}^{c_d} \bigcup_{B\in\mathcal{B}_i} (\operatorname{diam} B)^{t+2\epsilon} &\leq \sum_{i=1}^{c_d} \sum_{B(x,r)\in\mathcal{B}_i} (2Cr)^{t+2\epsilon} \\ &\leq \sum_{i=1}^{c_d} \sum_{\nu\in\mathcal{E}} \nu\left(\bigcup_{B\in\mathcal{B}_i} B\right) \\ &\leq c_d(2C)^{t+2\epsilon} \cdot \left(\sum_{\nu\in\mathcal{E}} \nu(\mathbb{R}^d)\right) < \infty\end{aligned}$$

which, since $\eta > 0$ and $\epsilon > 0$ were arbitrary, gives that $\dim_{\mathrm{H}} K \leq t$. □

We now obtain the following variational upper bound. In fact, our result is a bit stronger since we can actually bound any point such that the local dimension α is attained along a subsequence. More precisely, given a Borel probability measure μ and $x \in \operatorname{supp}\mu$, we write

$$\mathcal{D}(\mu, x) = \left\{\alpha : \exists (r_n)_{n=1}^{\infty} \searrow 0 \text{ s.t. } \lim_{n\to\infty} \frac{\log \mu(B(x, r_n))}{\log r_n} = \alpha\right\}.$$

Of course, $\dim_{\mathrm{loc}}(\mu, x) = \alpha$ if and only if $\mathcal{D}(\mu, x) = \{\alpha\}$. The following formula was first established in [OW03] via a direct covering argument.

Proposition 3.15 ([OW03]). *Let $\{S_i\}_{i\in\mathcal{I}}$ be a self-similar IFS satisfying the SSC. Then*

$$\begin{aligned} f_{\boldsymbol{p}}(\alpha) &= \dim_{\mathrm{H}}\{x \in K : \alpha \in \mathcal{D}(\mu_{\boldsymbol{p}}, x)\} \\ &= \sup_{\boldsymbol{w}\in\mathcal{P}} \left\{\frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})} : \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})} = \alpha\right\}. \end{aligned} \tag{3.12}$$

Proof. Recalling Proposition 2.4, it suffices to prove that

$$\dim_{\mathrm{H}}\{x \in K : \alpha \in \mathcal{D}(\mu_{\boldsymbol{p}}, x)\} \leq \sup_{\boldsymbol{w}\in\mathcal{P}} \left\{\frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})} : \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})} = \alpha\right\}.$$

To this end, fix $\gamma \in \Omega$ and $\alpha \in \mathcal{D}(\mu_{\boldsymbol{p}}, x)$. By (the proof of) Lemma 2.3, there is a subsequence $(n_k)_{k=1}^{\infty}$ so that $\lim_{k\to\infty} \boldsymbol{\xi}(\gamma|_{n_k}) = \boldsymbol{w}$ and

$$\lim_{k\to\infty} \frac{\log \mu_{\boldsymbol{p}}\big(B(x, r_{\gamma|_{n_k}})\big)}{\log r_{\gamma|_{n_k}}} = \alpha.$$

In particular,

$$\frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})} = \alpha;$$

and

$$\liminf_{r\to 0}\frac{\log \mu_{\boldsymbol{p}}\big(B(x,r)\big)}{\log r}\leq \lim_{k\to\infty}\frac{H(\boldsymbol{\xi}(\gamma|_{n_k}),\boldsymbol{w})}{\chi(\boldsymbol{\xi}(\gamma|_{n_k}))}=\frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})}.$$

Thus the desired result follows by Proposition 3.13 and Proposition 3.14. □

Remark 3.16. As discussed in the introduction, both the technique of uniform densities, as well as the method of types, are useful tools for reducing a general covering argument to a covering argument for "typical" points of a well-chosen measure. Heuristically, one might interpret the method of types argument as the "box-dimension" variant of this reduction, whereas the uniform densities argument is the "Hausdorff" variant of this reduction.

Remark 3.17. A similar argument as presented in this section will allow us to instead bound the *packing* dimension of the set $\{x\in K:\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}},x)=\alpha\}$. In order to bound the packing dimension, in Proposition 3.14, instead of providing a cover at infinitely many scales, one must provide a cover at *all* scales. Therefore if one replaces the limit infimum with a limit supremum in Proposition 3.14, one instead obtains a bound for the packing dimension of the underlying set. Then a similar argument as given in Proposition 3.15 shows that

$$\dim_{\mathrm{P}}\{x\in K:\alpha\in\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}},x)\}=\alpha\leq\sup_{\boldsymbol{w}\in\mathcal{P}}\left\{\frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})}:\frac{H(\boldsymbol{w},\boldsymbol{p})}{\chi(\boldsymbol{w})}=\alpha\right\}:$$

the key difference is that the subsequence $(n_k)_{k=1}^{\infty}$ must now be chosen to realize the limit supremum, which necessitates the assumption that the local dimension exists. In fact, it is no longer true that the packing dimension of the set $\{x\in K:\alpha\in\mathcal{D}(\mu_{\boldsymbol{p}},x)\}$ satisfies the same bound (see, for instance, [BOS07]).

4. CONCLUDING THOUGHTS: THE GEOMETRY OF OPTIMIZATION

4.1. Recalling duality. Recall the setting from §3.1: Δ is a compact set and we are given a continuous function $u\colon\Delta\to\mathbb{R}$ and an upper semicontinuous function $v\colon\Delta\to\mathbb{R}$. In fact, we now have a bit more context since in the multifractal analysis of self-similar measures, $\Delta=\mathcal{P}$ is the set of probability vectors, the function v is the dimension of the corresponding pushforward measure

$$v(\boldsymbol{w})=\dim_{\mathrm{H}}\mu_{\boldsymbol{w}}=\frac{H(\boldsymbol{w})}{\chi(\boldsymbol{w})}$$

and the function u gives the $\mu_{\boldsymbol{w}}$-typical local dimensions of $\mu_{\boldsymbol{p}}$

$$u(\boldsymbol{w})=\frac{H(\boldsymbol{w},\boldsymbol{p})}{\chi(\boldsymbol{w})}.$$

We then considered the constrained optimization problem

$$f(\alpha)=\max_{\boldsymbol{w}\in\Delta}\big\{v(\boldsymbol{w}):u(\boldsymbol{w})=\alpha\big\}$$

with unconstrained dual

$$\tau(q) = \min_{\boldsymbol{w}\in\Delta}\big\{q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})\big\}.$$

In our particular situation $f(\alpha) = f_{\boldsymbol{p}}(\alpha)$ is precisely the multifractal spectrum, and $\tau(q) = \tau_{\boldsymbol{p}}(q)$ is the L^q-spectrum.

In Proposition 2.9, we proved that $\sum_{i\in\mathcal{I}} p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)} = 1$. In particular, by the analytic implicit function theorem, this gives that $\tau_{\boldsymbol{p}}$ is differentiable; this was used in conjunction with Corollary 3.5 to prove the concave conjugate relationship. In fact, we proved more: we showed that

$$\tau_{\boldsymbol{p}}(q) = q \cdot u(\boldsymbol{z}(q)) - v(\boldsymbol{z}(q))$$

where $\boldsymbol{z}(q) = \big(p_i^q r_i^{-\tau_{\boldsymbol{p}}(q)}\big)_{i\in\mathcal{I}}$ defines a continuous path in $\mathcal{P}$. Moreover, the minimization giving $\tau_{\boldsymbol{p}}(q)$ is attained uniquely at the vector $\boldsymbol{z}(q)$.

We note that one can also implicitly verify that the minimization is attained uniquely at a vector $\boldsymbol{z}(q)$. Fix some $q \in \mathbb{R}$ and for each $t \in \mathbb{R}$, consider the lower level set

$$E(t,q) = \left\{\boldsymbol{w} \in \mathcal{P} : \frac{q \cdot H(\boldsymbol{w},\boldsymbol{p}) - H(\boldsymbol{w})}{\chi(\boldsymbol{w})} \le t\right\}.$$

Rearranging the condition on $E(t,q)$, we equivalently see that

$$E(t,q) = \big\{\boldsymbol{w} \in \mathcal{P} : H(\boldsymbol{w}) - q \cdot H(\boldsymbol{w},\boldsymbol{p}) + t \cdot \chi(\boldsymbol{w}) \ge 0\big\}.$$

But $H(\boldsymbol{w})$ is a strictly convex function and $\boldsymbol{w} \mapsto t \cdot \chi(\boldsymbol{w}) - q \cdot H(\boldsymbol{w},\boldsymbol{p})$ is a linear function in $\boldsymbol{w}$, so $E(t,q)$ is a convex set. Therefore, noting that

$$\tau_{\boldsymbol{p}}(q) = \inf\{t : E(t,q) \neq \varnothing\},$$

we must have that $E(\tau_{\boldsymbol{p}}(q)) = \{\boldsymbol{z}(q)\}$. Since all the relevant functions are continuous in q, we also conclude that $\boldsymbol{z}$ is continuous in q. The simplex in $\mathbb{R}^3$, the curve $\boldsymbol{z}(q)$, and the level sets of the function $\boldsymbol{w} \mapsto q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$, are depicted in Figure 3. The parameters are the same as the parameters used in Figure 1 and Figure 2,

This technique is useful to determine uniqueness of the minimization, or continuous path optimization, even in situations where it may be difficult to determine an explicit formula for the minimization. In fact, we just proved that the objective function $\phi(\boldsymbol{w}) = q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$ is *strictly quasiconvex*: that is, for all $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{P}$ and $\lambda \in (0,1)$,

$$\phi\big(\lambda\boldsymbol{w}_1 + (1-\lambda)\boldsymbol{w}_2\big) < \max\{\phi(\boldsymbol{w}_1), \phi(\boldsymbol{w}_2)\}.$$

We say that ϕ is *quasiconvex* if the above inequality holds with a regular inequality in place of a strict inequality. Quasiconvexity is a useful property since it follows directly from the definition that for continuous quasiconvex functions defined on a compact convex set, the minimum must be attained on a convex subset. Moreover,

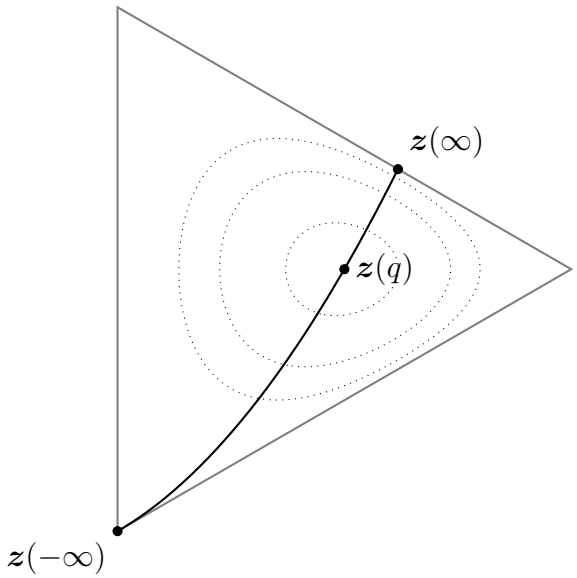

FIGURE 3. Parametrization of the optimization vector in the simplex $\mathcal{P} \subset \mathbb{R}^3$ and contour lines of $\boldsymbol{w} \mapsto q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$.

strict quasiconvexity implies that the minimum must be attained at a singleton. In particular, (strict) quasiconvexity of the objective function $\boldsymbol{w} \mapsto q \cdot u(\boldsymbol{w}) - v(\boldsymbol{w})$ at some $q \in \mathbb{R}$ allows us to immediately conclude the validity of the variational formula (or even differentiability) at q, by using Corollary 3.6.

While quasiconvex functions share many properties with convex functions, one important exception is that a sum of quasiconvex functions is *not* necessarily quasiconvex. For instance, this is why the higher-dimensional versions of the optimization problems discussed in this document are poorly behaved in general, such as non-uniqueness of measures of maximal dimensions [BF11].

4.2. Parametrizing the optimization. It turns out that even abstractly, having a continuously parametrized family of minimizers yields differentiability of the function $\tau(q)$ along with other geometric information. Suppose there is a continuous function $\boldsymbol{z} \colon \mathbb{R} \to \Delta$ such that $\tau(q) = q \cdot u(\boldsymbol{z}(q)) - b(\boldsymbol{z}(q))$. Using compactness of Δ, extend $\boldsymbol{z}$ to take values at $\pm\infty$ by taking any limit along a subsequence of q diverging to $\pm\infty$. We establish the following properties.

Proposition 4.1. *Suppose $\boldsymbol{z} \colon \mathbb{R} \to \Delta$ satisfies the above conditions. Then:*

(i) The function $\tau(q)$ is differentiable with derivative $\tau'(q) = u(\boldsymbol{z}(q))$.

(ii) The function $q \mapsto u(\boldsymbol{z}(q))$ is monotonically decreasing.

(iii) If $\alpha \in [u(\boldsymbol{z}(\infty)), u(\boldsymbol{z}(-\infty))]$, then

$$f(\alpha) = v(\boldsymbol{z}(q_\alpha))$$

where $q_\alpha \in \mathbb{R}$ is any value so that $u(\boldsymbol{z}(q_\alpha)) = \alpha$.

(iv) If $\alpha \notin [u(\boldsymbol{z}(\infty)), u(\boldsymbol{z}(-\infty))]$, then $f(\alpha) = -\infty$.

Proof. First, to see (i), for each $q \in \mathbb{R}$, since $\boldsymbol{z}(q)$ attains the minimum for $\tau(q)$, it follows that $u(\boldsymbol{z}(q)) \in \partial\tau(q)$. Moreover, by concavity of $\tau(q)$, if $q_1 < q_2$, then

$\inf \partial\tau(q_1) > \sup \partial\tau(q_2)$. Since $\boldsymbol{z}$ is continuous, this forces $\partial\tau(q) = \{u(\boldsymbol{z}(q))\}$, so that $\tau'(q) = u(\boldsymbol{z}(q))$, as claimed.

Now, (ii) follows from (i) since τ is concave and therefore has monotonically decreasing derivative.

To see (iii), just as in the proof of Proposition 3.3, fix $\alpha \in [u(\boldsymbol{z}(\infty)), u(\boldsymbol{z}(-\infty))]$ and let $q_\alpha \in \mathbb{R} \cup \{-\infty, \infty\}$ be such that $u(\boldsymbol{z}(q_\alpha)) = \alpha$. Then

$$f(\alpha) \geq v(\boldsymbol{z}(q_\alpha)) = q_\alpha \alpha - \tau(q) = \tau^*(\alpha)$$

since $\alpha \in \partial\tau(q_\alpha)$. We recall that the inequality $f(\alpha) \leq \tau^*(\alpha)$ holds in general from Proposition 3.3.

Finally, (iv) follows since

$$u(\boldsymbol{z}(\pm\infty)) = \lim_{q \to \pm\infty} \frac{\tau(q)}{q}$$

so for any $\alpha \notin [u(\boldsymbol{z}(\infty)), u(\boldsymbol{z}(-\infty))]$, $f(\alpha) \leq \tau^*(\alpha) = -\infty$. □

As a quick example, one setting in which such a function $\boldsymbol{z}$ must necessarily exist occurs the optimization defining $\tau(q)$ has a unique minimum for all $q \in \mathbb{R}$.

Corollary 4.2. *Suppose for each $q \in \mathbb{R}$, the minimization defining $\tau(q)$ is attained uniquely at a value $\boldsymbol{z}(q)$. Then $\boldsymbol{z} \colon \mathbb{R} \to \Delta$ is continuous.*

Proof. Since Δ is compact, we may use the sequential characterization of continuity along sequences for which the limit $\boldsymbol{z}(q_n)$ exists. Let $q_0 \in \mathbb{R}$ be arbitrary and let $(q_n)_{n=1}^\infty$ be any sequence of real numbers converging to q_0 such that the limit $\boldsymbol{w} = \lim_{n\to\infty} \boldsymbol{z}(q_n)$ exists. Then since τ is a continuous, u is continuous, and v is upper semicontinuous,

$$\tau(q_0) \leq q_0 \cdot u(\boldsymbol{w}) - v(\boldsymbol{w}) \leq \lim_{n\to\infty} (q_n \cdot u(\boldsymbol{z}(q_n)) - v(\boldsymbol{z}(q_n))) = \lim_{n\to\infty} \tau(q_n) = \tau(q_0)$$

so that all the inequalities are in fact equalities. But $\boldsymbol{z}(q_0)$ is unique with this property, so $\boldsymbol{z}(q_0) = \boldsymbol{w}$ and continuity follows. □

ACKNOWLEDGEMENTS

The author thanks Kenneth Falconer for many comments on various draft versions of this document, which have helped to improve the exposition greatly. He also thanks Amlan Banaji for pointing out various typos and suggesting some improvements. These notes have benefited greatly from conversations with István Kolossváry (who introduced the author to the method of types) and with Thomas Jordan.

REFERENCES

[AP96] M. Arbeiter and N. Patzschke. *Random self-similar multifractals*. Math. Nachr. **181** (1996), 5–42. zbl:0873.28003.

[BOS07] I. S. Baek, L. Olsen, and N. Snigireva. *Divergence points of self-similar measures and packing dimension*. Adv. Math. **214** (2007), 267–287. zbl:1141.28006.

[BSS23] B. Bárány, K. Simon, and B. Solomyak. *Self-similar and self-affine sets and measures*. Vol. 276. Providence, RI: American Mathematical Society, 2023. zbl:07759243.

[BF11] J. Barral and D.-J. Feng. *Non-uniqueness of ergodic measures with full Hausdorff dimensions on a Gatzouras–Lalley carpet*. Nonlinearity **24** (2011), 2563–2567. zbl:1228.37025.

[Bes34] A. S. Besicovitch. *Sets of fractional dimensions. IV: On rational approximation to real numbers*. J. Lond. Math. Soc. **9** (1934), 126–131. zbl:0009.05301.

[BP17] C. J. Bishop and Y. Peres. *Fractals in probability and analysis*. Vol. 162. Cambridge University Press, 2017. zbl:1390.28012.

[CM92] R. Cawley and R. D. Mauldin. *Multifractal decompositions of Moran fractals*. Adv. Math. **92** (1992), 196–236. zbl:0763.58018.

[Cli14] V. Climenhaga. *The thermodynamic approach to multifractal analysis*. Ergodic Theory Dynam. Systems **34** (2014), 1409–1450. zbl:1319.37016.

[CT06] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2006. zbl:1140.94001.

[DS17] T. Das and D. Simmons. *The Hausdorff and dynamical dimensions of self-affine sponges: a dimension gap result*. Invent. Math. **210** (2017), 85–134. zbl:1387.37026.

[DZ10] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Vol. 38. Berlin: Springer, 2010. zbl:1177.60035.

[Egg51] H. G. Eggleston. *Sets of fractional dimensions which occur in some problems of number theory*. Proc. Lond. Math. Soc. **54** (1951), 42–93. zbl:0045.16603.

[Fal85] K. J. Falconer. *The geometry of fractal sets*. Vol. 85. Cambridge: Cambridge University Press, 1985. zbl:0587.28004.

[Fal97] K. Falconer. *Techniques in fractal geometry*. Chichester: John Wiley & Sons, 1997. zbl:0869.28003.

[HJK+86] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman. *Fractal measures and their singularities: the characterization of strange sets.* Phys. Rev. A **33** (1986), 1141–1151. zbl:1184.37028.

[Kol23] I. Kolossváry. *The L^q spectrum of self-affine measures on sponges*. J. Lond. Math. Soc. **108** (2023), 666–701. zbl:07731132.

[LN99] K.-S. Lau and S.-M. Ngai. *Multifractal measures and a weak separation condition*. Adv. Math. **141** (1999), 45–96. zbl:0929.28007.

[LY85a] F. Ledrappier and L.-S. Young. *The metric entropy of diffeomorphisms. I: Characterization of measures satisfying Pesin's entropy formula.* Ann. Math. **122** (1985), 509–539. zbl:0605.58028.

[LY85b] F. Ledrappier and L.-S. Young. *The metric entropy of diffeomorphisms. II: Relations between entropy, exponents and dimension*. Ann. Math. **122** (1985), 540–574. zbl:1371.37012.

[Mat95] P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*. Vol. 44. Cambridge University Press, 1995. zbl:0819.28004.

[OW03] L. Olsen and S. Winter. *Normal and non-normal points of self-similar sets and divergence points of self-similar measures*. J. Lond. Math. Soc. **67** (2003), 103–122. zbl:1040.28014.

[Roc70] R. T. Rockafellar. *Convex analysis*. Vol. 28. Princeton University Press, Princeton, NJ, 1970. zbl:0193.18401.

[Shm19] P. Shmerkin. *On Furstenberg's intersection conjecture, self-similar measures, and the L^q norms of convolutions*. Ann. Math. **189** (2019), 319–391. zbl:1426.11079.

[Wal82] P. Walters. *An introduction to ergodic theory*. Vol. 79. Springer, Cham, 1982. zbl:0475.28009.

Alex Rutar
Mathematical Institute, University of St Andrews, St Andrews KY16 9SS, Scotland
Email: alex@rutar.org