FRACTAL DIMENSION, APPROXIMATION AND DATA SETS

L. BETTI, I. CHIO, J. FLEISCHMAN, A. IOSEVICH, F. IULIANELLI, S. KIRILA, M. MARTINO, A. MAYELI, S. PACK, Z. SHENG, C. TALIANCIC, A. THOMAS, N. WHYBRA, E. WYMAN, U. YILDIRIM, AND K. ZHAO

ABSTRACT. The purpose of this paper is to study the fractal phenomena in large data sets and associated questions of dimension reduction. We examine situations where the classical Principal Component Analysis is not effective in identifying the salient underlying fractal features of the data set. Instead, we employ the discrete energy, a technique borrowed from geometric measure theory, to limit the number of points of a given data set that lie near a k-dimensional hyperplane or, more generally, near a set of a given upper Minkowski dimension. Concrete motivations stemming from naturally arising data sets are described, and future directions are outlined.

1. Introduction

One of the basic questions in data science, which is to determine the "effective" dimension of a large data set. If a data set has 10^6 points in 1000-dimensional space, it is extremely useful to be able to detect if a significant proportion of these points live on a lower-dimensional plane, or a more complicated surface, reflecting the hidden relationships between the features present in the data set. One of the main tools in this area is Principal Component Analysis (PCA) (see e.g. [5]). This method has been and will remain a fundamental tool in the study of dimensionality of data sets. We propose to complement this method with a set of tools that capture important dimensionality phenomena that PCA does not see.

As a simple synthetic example, let us use the stages of construction of a Cantor-type subset of $[0,1]$ consisting of real numbers that have only 0s and 2s in their base 4 expansions. More precisely, divide $[0,1]$ in four equal segments, and remove the segments $(1/4,1/2)$ and $(3/4,1]$. We keep the endpoints of the remaining segments, obtaining $\{0,1/4,1/2,3/4\}$. Repeating the same procedure with the remaining two intervals of length $1/4$, we obtain 16 points at the next stage, and so on. Let $C_{2^{k+1}}$ denote the resulting set and note that it has 2^{k+1} points. Let $n=2^{2(k+1)}$ and define $P_n=C_{2^{k+1}}\times C_{2^{k+1}}$. From the point of view of PCA, this set is two-dimensional, and yet the fractal pattern that is present may be of considerable practical significance. We are going to explore this idea from a numerical point of view below.

In order to illustrate the ubiquity of fractional dimension phenomena in large data sets, let's consider the following hypothetical example. Suppose that we have a time series representing sales of a retail store going back 40 years. Suppose that one wished to look at the times when the sales were in the top 5% of all sales

The work of the fourth listed author was supported in part by the NSF HDR Tripods 1934962 grant and the NSF DMS 2154232 grant. The work of the fourteenth listed author was supported in part by the NSF DMS 2204397 grant.

in a given year, and it turned out that this happened every July, December and April, and that during those months it happened during the first week, and during that week it happened on Fridays and Saturdays, and that on those days, the sales peaked in the mornings. As the reader can see, this structure is highly reminiscent of the Cantor type construction from the previous paragraph. While this type of phenomenon has been extensively studied in terms of seasonality, we believe that a fractal perspective can be of considerable value in view of the fact that if the specific months, weeks, days, and times in the hypothetical above change, while their relative number remains roughly the same, the seasonality considerations no longer apply, while the fractal dimension analysis, as we shall see, is still valid and effective.

Another manifestation of the fractional dimensional phenomena in large data sets comes from the stock price data (see, e.g., [11]). It has been noted by several authors that fractional Brownian motion can be used to model stock price volatility. The fractional Brownian path has the upper Minkowski dimension > 1, reflecting the volatility of the data (see also [2]). Combined with the discussion in the previous paragraph about the variants of seasonality, we arrive at a very interesting situation where we have a set of effective dimensions < 1 on the time axis, combined with the volatile data modeled by a function whose graph has dimension > 1. A proper understanding of a situation of this type calls for advanced tools and perspectives from geometric measure theory, frame theory and harmonic analysis.

The fractal phenomenon in data sets has been studied before. We have been particularly influenced by the investigations by Smaller Jr., Turcotte and others in [13], [14] and [15].

1.1. **Structure of the paper.** In order to describe our results, we need to set up the notion of fractional dimension of finite points. In Section 2, we develop the notion of Hausdorff dimension for families of finite point sets living in the d-dimensional unit cube $[0,1]^d$, $d \geq 2$. We place particular focus on point sets that are given as graphs of a function from $[0,1]^{d-1}$ to $[0,1]$, with the idea of modeling real-life data sets where output, such as sales figures, is viewed as a function of the inputs that may include, the date, location, inflation rate, and other figures.

Theoretical results: After developing the notion of Hausdorff dimension of point sets by introducing the discrete s-energy

$$I_s(P_n) = n^{-2} \sum_{p \neq p'; p,p' \in P_n; |P_n| = n} |p - p'|^{-s},$$

we show (Theorem 4) that if the discrete s-energy of a point set P_n is suitably small, then P_n cannot contain too many points on k-dimensional surfaces, $k < s$, thus showing that small s energy limits the extent to which an effective dimension reduction can be implemented on a given data set. Note that $|p - p'|$ above denotes the Euclidean distance between p and p', whereas $|P_n|$ denotes the size of the finite set P_n.

For a comprehensive treatment of discrete energy, see [3]. We also note that discrete energy has previously been studied in the context of geometric combinatorics and geometric measure theory. See, e.g. [8] and [9].

Numerical experiments: In order to illustrate the utility of the ideas described above, we apply the standard PCA (Principal Component Analysis) to a discretized version of the Cartesian product of two Cantor sets and show that PCA does not fundamentally distinguish between this sparse example and a scale integer grid. We then show that the discrete dimensionality of the same set is quickly and efficiently estimated using a Python implementation of the discrete energy function described above. These results are described in Subsection 2.2 below and carried out later in the paper.

Future directions: After identifying discrete energy as an effective tool for determining the concentration of points of a numerical data set, we shall turn our attention towards understanding how the fractal nature of a data set can be exploited to make accurate forecasts using neural network models. More precisely, we shall ask how the architecture of a neural network should be influenced by the complexity of the data set as measured by discrete energy and other analytic tools. These ideas will be explored thoroughly in a sequel.

2. Discrete fractional dimension and concentration of data sets

2.1. **Definitions and core results.**

In this subsection, we develop some basic definitions of fractional dimension in a discrete setting and state some results that will be proven later in the paper.

Let $P_n \subset [0,1]^d$, $d \geq 2$ be a finite set of size n. We consider a family of such sets,

$$\mathcal{P} = \{P_n\},$$

where n ranges over some subset of the positive integers. For $s \in [0,d]$, define the *discrete s-energy* of P_n as

$$I_s(P_n) = n^{-2} \sum_{p \neq p'; p,p' \in P_n} |p - p'|^{-s}, \tag{1}$$

where here the sum is over pairs (p,p') with $p,p' \in P_n$ and $p \neq p'$.

Definition 1. Let $\mathcal{P}$ be as above. Define the discrete Hausdorff dimension of $\mathcal{P}$ as

$$dim_{\mathcal{H}}(\mathcal{P}) = \sup \left\{ s \in [0,d] : \sup_n I_s(P_n) < \infty \right\}.$$

We shall sometimes restrict our attention to point sets of the form

$$G_n(f) = \left\{ (j/q, f(j/q)) : j \in \mathbb{Z}^{d-1} \cap [0,q)^{d-1} \right\}, \qquad n = q^{d-1}, \tag{2}$$

for some $f : [0,1]^{d-1} \to [0,1]$. We shall refer to these as point set graphs. However, whenever possible, we shall establish results for the more general point sets defined above.

Before we start stating and proving our results, we describe the practical motivation for the point sets considered above. Suppose that a company X has d branches around the country, and we wish to describe the sales total for each branch on each of n dates. The resulting point set is a 2-dimensional array

$$\{p_{k,i}\}_{1 \leq k \leq n; 1 \leq i \leq d}.$$

We can regard this point set in two equivalent ways. We can fix a branch, labeled by i, and consider the time series array

$$\{p_{1i}, p_{2i}, \ldots, p_{ni}\}. \tag{3}$$

We have d such time series, so in this way we may regard our point set as d vectors in $\mathbb{R}^n$. Alternatively, we may fix a date and consider a vector of sales figures at the different branches on a given day:

$$\{p_{k1}, p_{k2}, \ldots, p_{kd}\}. \tag{4}$$

In this way, we may regard the point set as a collection of n vectors in $\mathbb{R}^d$, as we have it set up above.

Each of the two perspectives described above has its practical advantages from the point of view of the results described below. We are going to show that if the discrete s-energy of a point set is suitably small, then the point set cannot be too concentrated on a lower dimensional hyper-plane determining a certain linear relationship between the data points. If we look at this from the point of view of the time series in (3), the linear relationship signifies the possibility that sales figures for different branches are strongly related to one another across the board in an easy-to-describe fashion. From the point of view of (4), a concentration of data on a hyper-plane would say that for a significant number of different dates, there is a simple linear relationship between the sales figures for the various branches.

Since the number of dates is likely to be considerably larger than the number of branches of a company, the analysis in each case is of a different nature, as we shall see below.

We now turn our attention to the technical results. We begin with considering n points in d-dimensional space, as we set up at the beginning of the section, but we shall describe the "flipped" perspective afterward.

Lemma 2. *Let $\mathcal{P} = \{G_n\}$ be a family of point set graphs as in* (2) *above. Then $\dim_{\mathcal{H}}(\mathcal{P}) \geq d - 1$.*

We want to be able to describe the support of a family $\mathcal{P} = \{P_n\}$ of point sets of a given discrete Hausdorff dimension. To do so, we introduce a little structure. Take a compact set $E \subset \mathbb{R}^d$ and a positive number $\delta \leq 1$. We let $N_E(\delta)$ denote the minimal number of closed balls of radius δ needed to cover E. Recall

$$\underline{\dim}_{\mathcal{M}}(E) = \liminf_{\delta \to 0} \frac{\log(N_E(\delta))}{\log(1/\delta)} \qquad \text{and} \qquad \overline{\dim}_{\mathcal{M}}(E) = \limsup_{\delta \to 0} \frac{\log(N_E(\delta))}{\log(1/\delta)}$$

are the lower and upper Minkowski dimension of E, respectively. When the lower and the upper Minkowski dimensions agree, we call their common value the Minkowski dimension of E.

We will typically assume that, for some dimension k (which need not be an integer),

$$N_E(\delta) \leq \max(C_E \delta^{-k}, 1) \qquad \text{for } \delta > 0 \tag{5}$$

for some constant C_E. Here, C_E depends on E, but also on k as well. If such a constant exists, we must have $\overline{\dim}_{\mathcal{M}}(E) \leq k$. Not only this, but E must have finite k-dimensional upper Minkowski content. The condition (5) includes a wide range of essential examples, such as planar regions, piecewise-smooth manifolds, and a range of fractal sets including all of the Cantor sets we discuss in later sections.

We are ready for our core statement about the relationship between the discrete Hausdorff dimension and the Minkowski dimension of the set on which it is supported.

Theorem 3. *Let P_n be a point set of size n contained in a compact subset $E \subset \mathbb{R}^d$ satisfying (5) with $k > 0$. If $s > k$, we have the lower bound*

$$I_s(P_n) \geq \frac{k}{s-k} C_E^{-\frac{s}{k}} (n^{\frac{s}{k}-1} - 1). \tag{6}$$

Consequently, if $\mathcal{P} = \{P_n\}$ is a family of point sets contained in E of upper Minkowski dimension $\overline{\dim}_{\mathcal{M}}(E) \leq k$, then $\dim_{\mathcal{H}} \mathcal{P} \leq k$.

We now turn our attention to the question of dimension reduction. Suppose that P_n is as above and $\dim_{\mathcal{H}}(\mathcal{P}) = s$. The question we ask is whether it is possible that the points of $\mathcal{P}$ concentrate on k-dimensional hyper-planes or, more generally, on smooth k-dimensional surfaces. Our next result says that the fractal dimension places significant limitations on such possibilities.

Theorem 4. *Let E be a compact subset of $\mathbb{R}^d$ satisfying (5), and let P_n be a point set in $\mathbb{R}^d$ of size n. Then, for $s > k$,*

$$|P_n \cap E| \leq \left(1 + C_E^{\frac{s}{k}} \left(\frac{s}{k} - 1\right) I_s(P_n)\right)^{\frac{1}{\frac{s}{k}+1}} n^{\frac{2}{\frac{s}{k}+1}}. \tag{7}$$

Moreover, the exponent $\frac{2}{1+\frac{s}{k}}$ is, in general, best possible.

Remark 5. Since P_n is finite, so is $I_s(P_n)$. Thus, in view of Theorem 4, it is imperative to minimize the quantity

$$\left(1 + C_E^{\frac{s}{k}} \left(\frac{s}{k} - 1\right) I_s(P_n)\right)^{\frac{1}{\frac{s}{k}+1}} n^{\frac{2}{\frac{s}{k}+1}}.$$

To see the interplay between the various quantities above, we assume first that P_n has a diameter not exceeding 1. Then, observe that for a fixed n, $I_s(P_n)$ is an increasing function of s, as is

$$\left(\frac{s-k}{k}\right)^{\frac{1}{1+\frac{s}{k}}}.$$

On the other hand, $n^{\frac{2}{1+\frac{s}{k}}}$ is a decreasing function of s (with k fixed). Ultimately, an effective algorithm is needed to choose s given k and the point set P_n.

It is important, in practice, that we introduce a bit of "wiggle room" into Theorems 3 and 4, and instead consider point sets that are in some ϵ-neighborhood of E. Thankfully, both theorems have thickened versions with the same exponents. This flexibility is quite important for potential applications.

Theorem 6. *Let E be a compact subset of $\mathbb{R}^d$ satisfying (5) with $k > 0$. Let P_n be a point set of size n such that each point is within a distance of $\epsilon > 0$ from E. Then, if $s > k$, we have the discrete energy bound*

$$I_s(P_n) \geq \frac{k}{s-k} C_E^{-\frac{s}{k}} \left(\left(1 - \frac{\epsilon A_{s,k}}{C_E^{\frac{1}{k}} n^{-\frac{1}{k}}}\right) n^{\frac{s}{k}-1} - 1 \right)$$

with constant

$$A_{s,k} = \frac{s(s+1)(s-k)}{k(s-k+1)}.$$

In particular

$$I_s(P_n) \geq \frac{k}{s-k} C_E^{-\frac{s}{k}} \left(\frac{1}{2} n^{\frac{s}{k}-1} - 1\right) \qquad \text{if } \epsilon \leq \frac{C_E^{\frac{1}{k}}}{2A_{s,k}} n^{-\frac{1}{k}}.$$

Theorem 7. *Let P_n be a point set in $\mathbb{R}^d$ and let E be a compact subset of $\mathbb{R}^d$ satisfying* (5)*. Then, if E^ϵ is the ϵ-thickening of E with*

$$\epsilon = \frac{C_E^{\frac{1}{k}}}{2A_{s,k}} n^{-\frac{1}{k}},$$

then

$$|P_n \cap E^\epsilon| \leq 2^{\frac{1}{\frac{s}{k}+1}} \left(1 + C_E^{\frac{s}{k}} \left(\frac{s}{k} - 1\right) I_s(P_n)\right)^{\frac{1}{\frac{s}{k}+1}} n^{\frac{2}{\frac{s}{k}+1}}.$$

We will turn our attention to the case where our family of point sets lies along the graph of a function as above. To apply the results from above, we will need to connect the regularity of the function to the dimension of its graph. Recall, we say that $f : [0,1]^{d-1} \to \mathbb{R}$ is Hölder continuous of order $\alpha \in (0,1]$ if there exists a constant $\rho > 0$ such that

$$|f(x) - f(y)| \leq \rho |x - y|^\alpha$$

for all $x, y \in [0,1]^{d-1}$.

Lemma 8. *Let G_n be as above, and consider the corresponding $\mathcal{P}$. Suppose that $f : [0,1]^{d-1} \to [0,1]$ is Hölder continuous of order $\alpha \in (0,1]$. Then*

$$\dim_{\mathcal{H}}(\mathcal{P}) \leq d - \alpha.$$

Our last result of this section shows that every possible discrete dimension in $[d-1, d]$ is possible for the point set $G_n(f)$.

Theorem 9. *For each $s \in [d-1, d]$ there exists $f : [0,1]^{d-1} \to [0,1]$, such that if $\mathcal{P} = \{G_n(f)\}$ (with $G_n(f)$ defined as above), then $dim_{\mathcal{H}}(\mathcal{P}) = s$.*

2.2. **Computational results and examples.**

2.2.1. *Discrete dimension of Cartesian products of Cantor-type sets.* Start with the interval $[0,1]$ and positive integers $m, n \in \mathbb{Z}^+$ s.t. $m < n$. Divide $[0,1]$ into n subintervals of equal length, and choose m of those intervals. Let $C^1_{m,n}$ be the set of endpoints of the intervals chosen. At the k-th step, split each of the remaining intervals into n equal subintervals, choose m of those intervals, and let $C^k_{m,n}$ be the set of endpoints of the intervals chosen. At each step, we pick the same m subintervals from the remaining intervals. Since the m subintervals we choose to keep are arbitrary, the set $C^k_{m,n}$ is not unique and merely denotes one set satisfying these conditions.

Theorem 10. *Let $d \in \mathbb{Z}^+$. Let $C^k_{m_1,n_1}, \ldots, C^k_{m_d,n_d}$ be discrete Cantor sets. Set $A_k = \prod_{i=1}^d C^k_{m_i,n_i}$. Then*

$$dim_{\mathcal{H}}(A_k) = \frac{\ln(m_1)}{\ln(n_1)} + \cdots + \frac{\ln(m_d)}{\ln(n_d)}.$$

Now that we have this method for computing the dimension of discrete Cantor set products, we can compute the discrete s-energy of the sets in these families to better observe their behavior. For instance, if we compute the discrete s-energy of the discrete Cantor set products from Figure 6 below and other Cantor-type data distribution, where for each set, s is the dimension of that set computed by

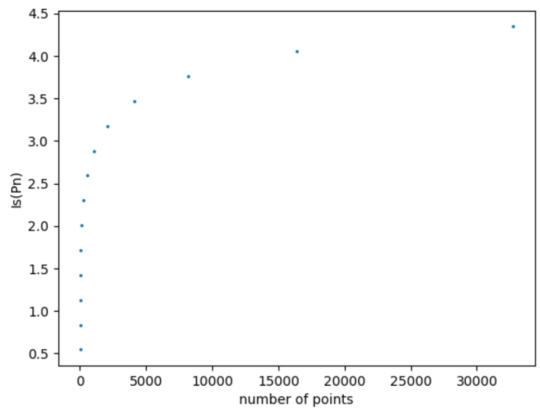

FIGURE 1. Discrete s-energy of $C_{2,4}^k$ for $1 \leq k \leq 15$ with $s = \frac{1}{2}$.

Theorem 10, we can see in Figure 1, Figure 3, Figure 4 and 5 that the discrete s-energy of the sets in each case increases slowly. This is what we would expect given that for each family of sets, s is the critical value, and for any $s' < s$, the discrete s'-energy of the sets are all bounded by a single constant.

Meanwhile, if we compute the discrete 2-energy of a family of lattices in $[0,1]^2$, then we can see from Figure 4 that the discrete 2-energy of the lattices increases slowly as well, possibly suggesting that the family of lattices is 2-adaptable. This would make sense intuitively because we would expect the family of lattices to be 2-dimensional, suggesting that this new notion of dimension may align with our general notion of dimension.

2.2.2. *PCA analysis of fractal and non-fractal objects.* As we denote in the introduction, the classical PCA method does not distinguish some key types of data sets. For example, it is not great at distinguishing a grid from, say, a Cartesian product of two Cantor sets, as we shall discuss below.

Recall the process of PCA: suppose there is a data set after preprocessing with the mean subtraction and standardization: $X = \{x_1, x_2, x_3, \ldots, x_n\}, x_i \in \mathbb{R}^D$. Mean subtraction will compute the mean of the dataset and subtract it from every single data point. Standardization will divide the data points by the standard deviation for every dimension [5]. To reduce set X's dimension to a lower one without losing too much information, find the data covariance matrix XX^T, calculating the m-th largest eigenvalues of that covariance matrix. Name the m eigenvectors $w_1, w_2, \ldots, w_m$ as m principle components, corresponding to the m largest eigenvalues. After generating a new matrix $W = \{w_1, w_2, w_3, \ldots. w_m\}, w_i \in R^D$, the data can be compressed to $y_i = W^T x_i$ for each x_i. The dimension of the original data set is reduced and represented as $X' = \{y_1, y_2, y_3, \ldots. y_n\}, y_i \in \mathbb{R}^M$, where $M < D$.

To illustrate some limitations of PCA, we test it with the grid and Cartesian product of two Cantor sets. The PCA analysis described above shows that the covariance matrix of the grid and Cartesian product of two Cantor sets are diagonal, and the eigenvalues for both data distributions are identical. Such results indicate that the two corresponding eigenvectors explain the same weight of the variance

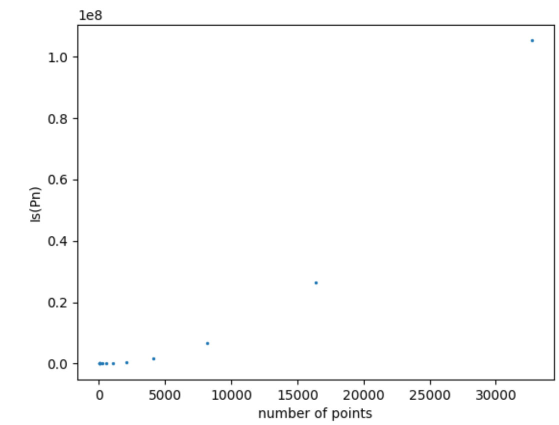

FIGURE 2. Discrete s-energy of $C_{2,4}^k$ with $2^{\mathrm{k}+1}$ points for $1 \leq k \leq 15$ with s=1.5.

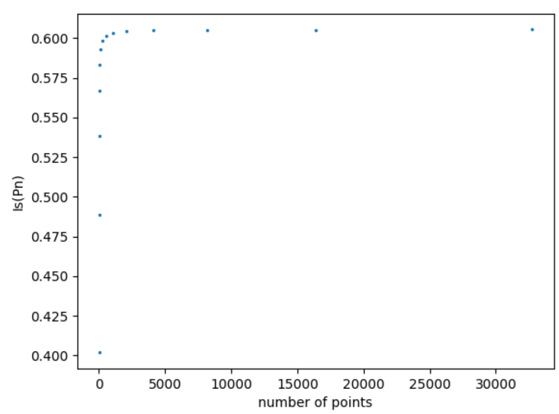

FIGURE 3. Discrete s-energy of $C_{2,4}^k$ with $2^{\mathrm{k}+1}$ points for $1 \leq k \leq 15$ with s=0.1.

of the data set, meaning that there is no need to remove any components or further reduce the dimensions. Therefore, PCA regards the 2-dimensional lattice and Cartesian product of two Cantor sets as 2-dimensional sets.

Nevertheless, the discrete s-energy remains robust to identify the dimension of a cantor set is, in fact, less than 1. For instance, our codes and output figures 1-3 show that if choosing s such that $s < \frac{ln(2)}{ln(4)}$, the discrete s-energy of $C_{2,4}^k$ converges to a relatively small number as the number of points (depend on k) increases. Otherwise, discrete s-energy diverges. The results are consistent with Theorem 10.

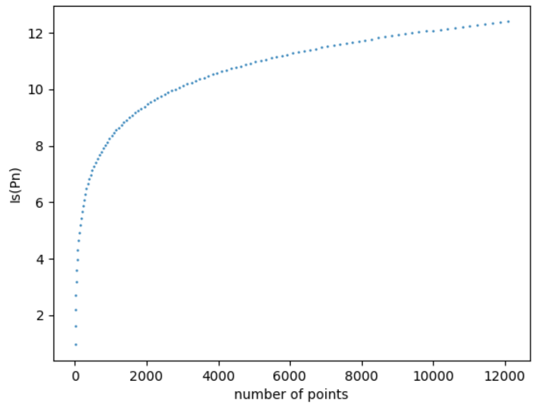

FIGURE 4. Discrete s-energy of 2-dimensional lattice with $(k+1)^2$ points for $1 \leq k \leq 110$, $s = 2$.

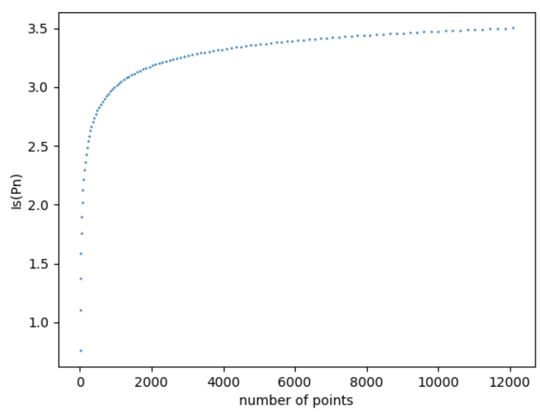

FIGURE 5. Discrete s-energy of 2-dimensional lattice with $(k+1)^2$ points for $1 \leq k \leq 110$, $s = 1.5$.

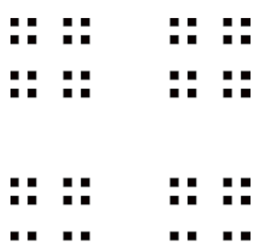

FIGURE 6. The Cartesian product of two middle-half Cantor sets (otherwise known as the Garnett set).

3. Proof of results from Section 2

3.1. **Proof of Lemma 2.** We will show the s-discrete energy of G_n is uniformly bounded if $0 \le s < d-1$. As usual, we let $n = q^{d-1}$ for a positive integer q. Let $p = (j/q, f(j/q)) \in G_n$, where here $j \in \mathbb{Z}^{d-1} \cap [0,q)^{d-1}$. Note, for each such pair p, p',

$$|p - p'|^{-s} \le |j/q - j'/q|^{-s} = q^s |j - j'|^{-s}.$$

Hence, we have

$$I_s(P_n) = n^{-2} \sum_{p \neq p'} |p - p'|^{-s} \le q^{-2(d-1)+s} \sum_{j \neq j'} |j - j'|^{-s},$$

where here j and j' range over $\mathbb{Z}^{d-1} \cap [0,q)^{d-1}$. By a change of variables, the above is bounded by

$$q^{-(d-1)+s} \sum_{\substack{j \in \mathbb{Z}^{d-1} \cap (-q,q)^{d-1} \\ j \neq 0}} |j|^{-s} \le C q^{-(d-1)+s} q^{d-1-s}$$

by the integral test, where here C is a constant depending on d and s only. The conclusion of Lemma 2 follows.

3.2. **Proofs of Theorems 3 and 6.** We start by only assuming that E is a compact subset of $[0,1]^d$ and P_n is any set of n points in E. We will not introduce the regularity condition (5) until a bit later.

We write

$$|p - p'|^{-s} = s \int_0^\infty \mathbf{1}_{[0,\infty)}(r - |p - p'|) r^{-s-1}\, dr,$$

so that

$$\begin{aligned} I_s(P_n) &= n^{-2} \sum_{p \neq p'} |p - p'|^{-s} \\ &= s n^{-2} \sum_{p \neq p'} \int_0^\infty \mathbf{1}_{[0,\infty)}(r - |p - p'|) r^{-s-1}\, dr \\ &= s n^{-2} \int_0^\infty \left(|\{(p,p') : |p - p'| \le r\}| - n \right) r^{-s-1}\, dr. \end{aligned}$$

The subtracted n is here to eliminate the contribution to $|\{(p,p') : |p - p'| \le r\}|$ of the pairs (p,p') with $p = p'$. To estimate $I_s(P_n)$, we will need a lower bound on the size of this set.

Lemma 11. *Let E be a compact subset of $\mathbb{R}^d$ and let $P_n \subset E$ be a point set in E of size n. Then,*

$$|\{(p,p') \in P_n \times P_n : |p - p'| \le r\}| \ge \frac{n^2}{N_E(r)}.$$

Proof. Take a minimal cover of E by $N_E(r)$ balls of radius r. Select a partition of E into $N_E(r)$ pairwise disjoint sets

$$E = E_1 \cup \cdots \cup E_{N_E(r)}$$

where each part E_i is contained in its respective ball in the cover. Note,

$$|\{(p,p'):|p-p'|\le r\}| \ge \sum_j |E_j \cap P_n|^2 \ge \frac{1}{N_E(r)} \left(\sum_j |E_j \cap P_n|\right)^2 = \frac{n^2}{N_E(r)},$$

where the second inequality is an application of Cauchy-Schwarz. This concludes the proof of the lemma. □

Remark 12. The Cauchy-Schwarz inequality used above is optimal when $|E_j \cap P_n|$ is constant across j, i.e. when P_n is well-distributed in E. If, say, P_n is contained in a plane E in $[0,1]^d$, then the energy $I_s(P_n)$ is minimized (or nearly minimized) if P_n is a lattice in E.

We now resume the proof of Theorem 3. Take δ to be any number with $0 < \delta \le 1$. Then, the lemma gives

$$|\{(p,p'):|p-p'|\le r\}| - n \ge \begin{cases} 0 & r<\delta \\ N_E(r)^{-1}n^2 - n & r \ge \delta. \end{cases}$$

It now follows that

$$I_s(P_n) \ge sn^{-2} \int_\delta^\infty \left(N_E(r)^{-1}n^2 - n\right) r^{-s-1}\, dr.$$

After some minor calculations, it follows:

Lemma 13. *Let E be a compact subset of $[0,1]^d$ and P_n a set of n points in E. Then,*

$$I_s(P_n) \ge s\left(\int_\delta^\infty N_E(r)^{-1} r^{-s-1}\, dr\right) - n^{-1}\delta^{-s}$$

where $0 < \delta \le 1$.

We are now ready to prove Theorem 3.

Theorem 3. We now can assume (5) from which we have

$$N_E(r)^{-1} \ge \min(C_E^{-1} r^k, 1).$$

Hence by the lemma, we have

$$\begin{aligned} I_s(P_n) &\ge C_E^{-1} s \int_\delta^{C_E^{\frac{1}{k}}} r^{k-s-1}\, dr + s\int_{C_E^{\frac{1}{k}}}^\infty r^{-s-1}\, dr - n^{-1}\delta^{-s} \\ &= \frac{C_E^{-1}s}{k-s}\left(C_E^{1-\frac{s}{k}} - \delta^{k-s}\right) + C_E^{-\frac{s}{k}} - n^{-1}\delta^{-s}. \end{aligned}$$

Taking $\delta = (C_E^{-1}n)^{-\frac{1}{k}}$ yields the bound, after some simplification. □

Theorem 6 is also now within easy reach.

Theorem 6. Let E^ϵ denote the ϵ-thickening of E. Now, if E satisfies (5), then

$$N_{E^\epsilon}(r) \le N_E(r-\epsilon) \le \max(C_E(r-\epsilon)^{-k}, 1).$$

By the lemma, we have

$$\begin{aligned} I_s(P_n) &\geq s\int_{\delta+\epsilon}^{\infty}(N_E(r-\epsilon))^{-1}r^{-s-1}\,dr - n^{-1}(\delta+\epsilon)^{-s} \\ &\geq s\int_{\delta}^{\infty} N_E(r)^{-1}(r+\epsilon)^{-s-1}\,dr - n^{-1}\delta^{-s}. \end{aligned}$$

Since r^{-s-1} is convex, we have

$$(r+\epsilon)^{-s-1} \geq r^{-s-1} - (s+1)\epsilon r^{-s-2} \qquad \text{for } r>0.$$

Hence,

$$I_s(P_n) \geq \left(s\int_{\delta}^{\infty} N_E(r)^{-1}r^{-s-1}\,dr - n^{-1}\delta^{-s}\right) - s(s+1)\epsilon\int_{\delta}^{\infty} N_E(r)^{-1}r^{-s-2}\,dr.$$

Taking $\delta = C_E^{\frac{1}{k}}n^{-\frac{1}{k}}$ and repeating the argument in the proof of Theorem 3 yields

$$\begin{aligned} I_s(P_n) \geq{} & \frac{k}{s-k}C_E^{-\frac{s}{k}}(n^{\frac{s}{k}-1}-1) \\ & - \epsilon s(s+1)\left(C_E^{-1}\int_{C_E^{\frac{1}{k}}n^{-\frac{1}{k}}}^{C_E^{\frac{1}{k}}} r^{k-s-2}\,dr + \int_{C_E^{\frac{1}{k}}}^{\infty} r^{-s-2}\,dr\right). \end{aligned}$$

By change of variables, we have

$$C_E^{-1}\int_{C_E^{\frac{1}{k}}n^{-\frac{1}{k}}}^{C_E^{\frac{1}{k}}} r^{k-s-2}\,dr = \frac{C_E^{-\frac{s+1}{k}}}{s-k+1}\left(n^{-1+\frac{s+1}{k}}-1\right)$$

and

$$\int_{C_E^{\frac{1}{k}}}^{\infty} r^{-s-2}\,dr = \frac{C_E^{-\frac{s+1}{k}}}{s+1}.$$

Their sum is then bounded above by $\frac{C_E^{-\frac{s+1}{k}}}{s-k+1}n^{-1+\frac{s+1}{k}}$, and hence we have the lower bound

$$I_s(P_n) \geq \frac{k}{s-k}C_E^{-\frac{s}{k}}(n^{\frac{s}{k}-1}-1) - \frac{\epsilon}{C_E^{\frac{1}{k}}n^{-\frac{1}{k}}}\frac{C_E^{-\frac{s}{k}}s(s+1)}{s-k+1}n^{\frac{s}{k}-1}.$$

The bound in the theorem follows. □

3.3. **Proofs of Theorems 4 and 7.**

Theorem 4. For fixed n, let $m = |P_n \cap E|$ and $P'_m = P_n \cap E$. On one hand, we have an upper bound

$$I_s(P'_m) = m^{-2}\sum_{\substack{p\neq p' \\ p,p'\in P_{m'}}} |p-p'|^{-s} \leq m^{-2}\sum_{p\neq p';p,p'\in P_n} |p-p'|^{-s} = m^{-2}n^2 I_s(P_n)$$

on $I_s(P'_m)$. At the same time, since $P'_m \subset E$, (6) gives us a lower bound

$$I_s(P'_m) \geq \frac{k}{s-k}C_E^{-\frac{s}{k}}(m^{\frac{s}{k}-1}-1).$$

By comparing the upper and lower bounds, we obtain

$$(m^{\frac{s}{k}-1}-1)m^2 \leq C_E^{\frac{s}{k}}\left(\frac{s}{k}-1\right)I_s(P_n)n^2.$$

The left side is bounded below by $m^{\frac{s}{k}+1} - n^2$, and hence we have

$$m^{\frac{s}{k}+1} \leq \left(1 + C_E^{\frac{s}{k}}\left(\frac{s}{k}-1\right) I_s(P_n)\right) n^2.$$

The bound in the theorem follows.

Theorem 4 is, in general, best possible. To see this, we start with a family $\mathcal{P}$ of lattices

$$P_n = \left\{\frac{j}{n^{\frac{1}{d}}} : j \in \mathbb{Z}^d, 0 \leq j_i \leq n^{\frac{1}{d}}, 1 \leq i \leq d\right\}.$$

It is not difficult to check that $I_s(P_n) \leq C$ independently of n if $s < d$, from which we have $\dim_{\mathcal{H}} \mathcal{P} = d$. Now let us modify P_n by replacing the $n^{\frac{k}{d}}$ points in the plane $x_1 = \cdots = x_k = 0$ with a lattice $Z := m^{-1/k}\mathbb{Z}^k \cap [0, m^{1/k})^k$ of m points. We will take $m \gg n^{\frac{k}{d}}$, which effectively increases the number of points on the plane. However, we will not add so many as to disturb the bounds on the s-energy of $\mathcal{P}$. To this end, we require that if $s < k$, there is a constant C for which

$$I_s(P_n) \leq C \qquad \text{for each } n.$$

The left side reads as a bounded term plus

$$n^{-2} \sum_{\substack{u,u' \in Z \\ u \neq u'}} \left|\frac{u-u'}{m^{\frac{1}{k}}}\right|^{-s} \approx n^{-2} m^{\frac{s}{k}+1} \sum_{w \in Z} |w|^{-s} \approx n^{-2} m^{\frac{s}{k}+1},$$

and hence the desired bounds hold if and only if $m^{\frac{s}{k}+1} \leq Cn^2$, i.e. $m \leq Cn^{\frac{2}{1+\frac{s}{k}}}$. Here and throughout, $X \approx Y$ means that there exists a uniform constant C such that $C^{-1}Y \leq X \leq CY$. □

Next, we proceed with the proof of Theorem 7, the thickened version of Theorem 4.

Theorem 7. We follow the proof of Theorem 4 above, except we take $P'_m = P_n \cap E^\epsilon$ where here

$$\epsilon = \frac{C_E^{\frac{1}{k}}}{2A_{s,k}} n^{-\frac{1}{k}} \leq \frac{C_E^{\frac{1}{k}}}{2A_{s,k}} m^{-\frac{1}{k}}$$

with notation as in Theorem 6. The same theorem then tells us

$$I_s(P'_m) \geq \frac{k}{s-k} C_E^{-\frac{s}{k}} \left(\frac{1}{2} m^{\frac{s}{k}-1} - 1\right).$$

We similarly have an upper bound

$$I_s(P'_m) \leq m^{-2} n^2 I_s(P_n).$$

Proceeding as before, we obtain

$$\frac{1}{2} m^{\frac{s}{k}+1} - n^2 \leq \frac{1}{2} m^{\frac{s}{k}+1} - m^2 \leq C_E^{\frac{s}{k}} \left(\frac{s}{k}-1\right) I_s(P_n) n^2.$$

It follows

$$m \leq 2^{\frac{1}{\frac{s}{k}+1}} \left(1 + C_E^{\frac{s}{k}}\left(\frac{s}{k}-1\right) I_s(P_n)\right)^{\frac{1}{\frac{s}{k}+1}} n^{\frac{2}{\frac{s}{k}+1}}$$

as needed. □

3.4. **Proof of Lemma 8.** We claim the graph of f satisfies (5) for $k = d - \alpha$; afterward, Theorem 3 completes the proof. We do this instead by covering the graph of f with cubes of length δ. Passing from cubes to balls will only affect the constant in (5).

For fixed q, let Q_j denote the cube $\prod_{i=1}^{d-1}[\frac{j_i}{q}, \frac{j_i+1}{q}]$ where here j ranges over $\mathbb{Z}^{d-1} \cap [0,q)^{d-1}$. Since f is Hölder continuous of order α, $f(Q_j)$ has diameter at most $\rho(\sqrt{d-1}q^{-1})^{\alpha}$. Hence, we can cover the graph of f over Q_j by $\rho\sqrt{d-1}q^{1-\alpha}$ d-dimensional cubes of side length q^{-1}. Repeating for each of the q^{d-1} cubes Q_j, we cover the whole graph of f by $q^{d-1} \cdot \rho\sqrt{d-1}q^{1-\alpha} = \rho\sqrt{d-1}q^{d-\alpha}$ cubes of length q^{-1}, and our claim is proved after taking $\delta = q^{-1}$.

4. Proof of Theorem 9

4.1. **2-dimensional case.**

Proof. In order to prove the case for $d = 2$, the construction used by B.Hunt is followed [17]. In such construction, we consider Weierstrass functions $f_\theta : [0,1] \to [0,1]$ with random phases of the form:

$$f_\theta(x) = \sum_{n=0}^{\infty} a^n cos\,(2\pi(b^n x + \theta_n)) \tag{8}$$

where $0 < a < 1 < b$, $ab < 1$, and $\theta = (\theta_1, \theta_2, \dots) \in [0,1]^\infty = H$ is randomly chosen by sampling each of its entries according to the uniform distribution on $[0,1]$.

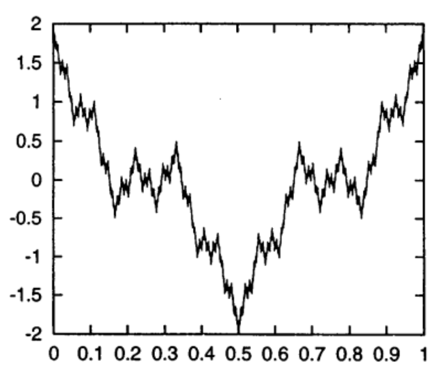

FIGURE 7. Graph of $f(x)$ with $a = 0.5$ and $b = 3$ and $\theta_n = 0$ for any n [17].

Note that, for any $\theta \in H$, f_θ is an Holder continuous function with exponent $\alpha = -\frac{\log(a)}{\log(b)}$. In fact, for any $x, y \in [0,1]$, substituting $a = b^{-\alpha}$:

$$\begin{aligned}
|f_\theta(x) - f_\theta(y)| &= \left|\sum_{n=0}^{\infty} b^{-n\alpha}\cos(2\pi(b^n x + \theta_n)) - \sum_{n=0}^{\infty} b^{-n\alpha}\cos(2\pi(b^n y + \theta_n))\right| \\
&\leq \sum_{n=0}^{\infty} b^{-n\alpha}\min(2, 2\pi b^n |x-y|) \\
&\leq \sum_{n=0}^{m-1} 2\pi b^{n(1-\alpha)} |x-y| + \sum_{n=m}^{\infty} 2b^{-n\alpha} \\
&= 2\pi \frac{b^{(1-\alpha)m} - 1}{b^{(1-\alpha)} - 1} |x-y| + 2\frac{b^{-\alpha m}}{1 - b^{-\alpha}}
\end{aligned}$$

for any integer $m > 0$. The first inequality followed, thanks to the fact that the cosine function is Lipschitz (an immediate consequence of the mean value theorem). Set m to be the positive integer such that $b^{-m} < |x-y| \leq b^{-(m-1)}$. Then, resuming the estimate:

$$\begin{aligned}
|f_\theta(x) - f_\theta(y)| &\leq 2\pi \frac{\left(\frac{b}{|x-y|}\right)^{(1-\alpha)}}{b^{(1-\alpha)} - 1} |x-y| + 2\frac{|x-y|^\alpha}{1-b^{-\alpha}} \\
&\leq \left(\frac{2\pi b^{(1-\alpha)}}{b^{(1-\alpha)} - 1} + \frac{2}{1-b^{-\alpha}}\right) |x-y|^\alpha .
\end{aligned} \tag{9}$$

Thus, combining (9) and Lemma 8, we have that $dim_{\mathcal{H}}(\mathcal{P}) \geq 2 - \alpha$. To prove the other inequality, the random phases are very useful as they allow us to estimate the energy integral indirectly. In fact, the key point of the proof is to show that, for any $s < 2 - \alpha$, $\mathbb{E}_H(I_s(G_n)) < \infty$ independently of n (i.e., independently of q). Consequently, this implies that for almost every random sequence of phases $\theta \in H$, f_θ satisfies $dim_{\mathcal{H}}(\mathcal{P}) \leq 2 - \alpha$ (completing the proof for $d = 2$). To perform the estimate by linearity of expectation:

$$\mathbb{E}_H(I_s(G_n)) = q^{-2} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z}}} \mathbb{E}_H\left(\left|\left(\frac{j}{q}, f_\theta\left(\frac{j}{q}\right)\right) - \left(\frac{j'}{q}, f_\theta\left(\frac{j'}{q}\right)\right)\right|^{-s}\right).$$

Next, the following lemma (proved by B.Hunt [17]) provides an upper bound for the expectations inside the sum when $\frac{j}{q}$ and $\frac{j}{q}$ are close enough.

Proposition 14. *For any x, y and any such that $|x-y| < 1/2b^2$:*

$$\mathbb{E}_H\left(|(x, f_\theta(x)) - (y, f_\theta(y))|^{-s}\right) \leq C\,|x-y|^{(1-\alpha-s)} \tag{10}$$

for any $s \in (1, 2-\alpha)$

Then, resuming the proof, we can decompose the sum into two terms:

$$\mathbb{E}_H(I_s(G_n)) = I + II,$$

where

$$I = q^{-2} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z} \\ |j-j'| \leq \frac{q}{2b^2}}} \mathbb{E}_H \left(\left| \left(\frac{j}{q}, f_\theta \left(\frac{j}{q} \right) \right) - \left(\frac{j'}{q}, f_\theta \left(\frac{j'}{q} \right) \right) \right|^{-s} \right),$$

$$II = q^{-2} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z} \\ |j-j'| > \frac{q}{2b^2}}} \mathbb{E}_H \left(\left| \left(\frac{j}{q}, f_\theta \left(\frac{j}{q} \right) \right) - \left(\frac{j'}{q}, f_\theta \left(\frac{j'}{q} \right) \right) \right|^{-s} \right).$$

The second term is easy to be bounded as:

$$II \leq q^{-2} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z} \\ |j-j'| > \frac{q}{2b^2}}} \left| \frac{j}{q} - \frac{j'}{q} \right|^{-s} \leq q^{-2} q^2 (2b^2)^s = (2b^2)^s.$$

For the first term, using proposition (14):

$$\begin{aligned} I &\lesssim q^{-2} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z} \\ |j-j'| \leq \frac{q}{2b^2}}} \left| \frac{j}{q} - \frac{j'}{q} \right|^{1-\alpha-s} \lesssim q^{-3+\alpha+s} \sum_{\substack{j \neq j' \\ j,j' \in [0,q) \cap \mathbb{Z}}} |j - j'|^{1-\alpha-s} \\ &\lesssim q^{-2+\alpha+s} \sum_{\substack{j \in (-q,q) \cap \mathbb{Z} \\ j \neq 0}} |j|^{1-\alpha-s} \lesssim q^{-2+\alpha+s} q^{2-\alpha-s} \lesssim 1. \end{aligned}$$

completing the proof. Note that the third and fourth inequality are, respectively, due to the change of variable $j - j'$ to j and due to the integral test (where the integral converges for values of s strictly smaller than $2 - \alpha$, as desired). The 2-dimensional case is fully proved since for any given value $x \in (0, 1]$, we can find values of a and b such that $\alpha = x$. This is a consequence of the fact that $\alpha = -\frac{\log(a)}{\log b}$ and $0 < a < 1 < b$ with $ab \geq 1$. In fact, fixing a positive $a < 1$, $b = \frac{1}{a^x} \geq \frac{1}{a}$ for any value of $x \in (0, 1]$ and $\alpha = x$. □

4.2. **Higher dimensional case for $d \geq 3$.**

Proof. To generalize to higher dimensions (i.e. $d \geq 3$), we consider the following functions $g_\theta : [0, 1]^{d-1} \to [0, 1]$ of the form:

$$g_\theta(x) = g_\theta(x_1, \ldots, x_{d-1}) = f_\theta(x_1) \tag{11}$$

where the functions f_θ are defined as in (8).

For any $\theta \in H$, g_θ is again Holder continuous with exponent α. This is because, using (9) and (11), for any $x, y \in [0, 1]^{d-1}$:

$$\begin{aligned} |g_\theta(x) - g_\theta(y)| = |f_\theta(x_1) - f_\theta(y_1)| &\leq \left(\frac{2\pi b^{(1-\alpha)}}{b^{(1-\alpha)} - 1} + \frac{2}{1 - b^{-\alpha}} \right) |x_1 - y_1|^\alpha \\ &\leq \left(\frac{2\pi b^{(1-\alpha)}}{b^{(1-\alpha)} - 1} + \frac{2}{1 - b^{-\alpha}} \right) |x - y|^\alpha . \end{aligned} \tag{12}$$

Thus, combining (12) and Lemma 8, $dim_{\mathcal{H}}(\mathcal{P}) \geq d - \alpha$.

To prove the other inequality, looking at the expectation over all random phases as in the case $d = 2$, we can perform the following decomposition:

$$\begin{aligned}
\mathbb{E}_H(I_s(G_n)) &= q^{-2(d-1)}\,\mathbb{E}_H\left(\sum_{\substack{j\neq j'\\ j,j'\in[0,q)^{d-1}\cap\mathbb{Z}^{d-1}}}\left|\left(\frac{j}{q},g_\theta\left(\frac{j}{q}\right)\right)-\left(\frac{j'}{q},g_\theta\left(\frac{j'}{q}\right)\right)\right|^{-s}\right)\\
&= q^{-2(d-1)}\,\mathbb{E}_H\left(\sum_{\substack{j\neq j'\\ j,j'\in[0,q)^{d-1}\cap\mathbb{Z}^{d-1}}}\left|\left(\frac{j}{q},f_\theta\left(\frac{j_1}{q}\right)\right)-\left(\frac{j'}{q},f_\theta\left(\frac{j'_1}{q}\right)\right)\right|^{-s}\right)\\
&= I + II
\end{aligned}$$

where

$$I = q^{-2(d-1)}\sum_{\substack{j_1\neq j'_1\\ j_1,j'_1\in[0,q)\cap\mathbb{Z}}}\mathbb{E}_H\left(\sum_{\bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}\left(\left|\frac{j_1}{q}-\frac{j'_1}{q}\right|^2+\left|f_\theta\left(\frac{j_1}{q}\right)-f_\theta\left(\frac{j'_1}{q}\right)\right|^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\right)$$

$$II = q^{-2(d-1)}\sum_{\substack{j_1=j'_1\\ j_1,j'_1\in[0,q)\cap\mathbb{Z}}}\mathbb{E}_H\left(\sum_{\substack{\bar{j}\neq\bar{j}'\\ \bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}}\left(\left|\frac{j_1}{q}-\frac{j'_1}{q}\right|^2+\left|f_\theta\left(\frac{j_1}{q}\right)-f_\theta\left(\frac{j'_1}{q}\right)\right|^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\right)$$

where $\bar{j} = (j_2, ..., j_{d-1})$. We just need to worry about bounding term I since it is greater than term II. To see this, note that any addend in the inner sum of II is comparable to an addend of the inner sum of II but, at the same time, the outer sum of I contains way more terms than the outer sum of II (due to the difference between the conditions $j_1 = j'_1$ and $j_1 \neq j'_1$). To bound the first term, we can further decompose it into two pieces $I = III + IV$, where

$$III = q^{-2(d-1)}\sum_{\substack{j_1\neq j'_1\\ j_1,j'_1\in[0,q)\cap\mathbb{Z}\\ |j_1-j'_1|\leq\frac{q}{2b^2}}}\mathbb{E}_H\left(\sum_{\bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}\left(\left|\frac{j_1}{q}-\frac{j'_1}{q}\right|^2+\left|f_\theta\left(\frac{j_1}{q}\right)-f_\theta\left(\frac{j'_1}{q}\right)\right|^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\right)$$

$$IV = q^{-2(d-1)}\sum_{\substack{j_1\neq j'_1\\ j_1,j'_1\in[0,q)\cap\mathbb{Z}\\ |j_1-j'_1|>\frac{q}{2b^2}}}\mathbb{E}_H\left(\sum_{\bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}\left(\left|\frac{j_1}{q}-\frac{j'_1}{q}\right|^2+\left|f_\theta\left(\frac{j_1}{q}\right)-f_\theta\left(\frac{j'_1}{q}\right)\right|^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\right)$$

The last term is bounded as follows:

$$\begin{aligned}
IV &\leq q^{-2(d-1)}\sum_{\substack{j_1\neq j'_1\\ j_1,j'_1\in[0,q)\cap\mathbb{Z}\\ |j_1-j'_1|>\frac{q}{2b^2}}}\ \sum_{\bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}\left|\frac{j_1}{q}-\frac{j'_1}{q}\right|^{-s}\\
&\leq q^{-2(d-1)}q^2q^{2(d-2)}(2b^2)^s = (2b^2)^s.
\end{aligned}$$

In the inner sum of III, the quantity $\left|\frac{j_1}{q}-\frac{j_1'}{q}\right|^2+\left|f_\theta(\frac{j_1}{q})-f_\theta(\frac{j_1'}{q})\right|^2$ is constant (i.e. does not depend on $\bar{j},\bar{j}'$). For the sake of clarity, set $\beta^2=\left|\frac{j_1}{q}-\frac{j_1'}{q}\right|^2+\left|f_\theta(\frac{j_1}{q})-f_\theta(\frac{j_1'}{q})\right|^2$. We can then find an upper bound for such inner sum:

$$
\begin{aligned}
&\sum_{\bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}\left(\beta^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\\
&\qquad= q^{d-2}\beta^{-s}+\sum_{\substack{\bar{j}\neq\bar{j}'\\ \bar{j},\bar{j}'\in[0,q)^{d-2}\cap\mathbb{Z}^{d-2}}}\left(\beta^2+\left|\frac{\bar{j}}{q}-\frac{\bar{j}'}{q}\right|^2\right)^{-\frac{s}{2}}\\
&\qquad\le q^{d-2}\beta^{-s}+q^{d-2+s}\sum_{\substack{\bar{j}\neq\bar{j}'\\ \bar{j}\in(-q,q)^{d-2}\cap\mathbb{Z}^{d-2}}}\left((q\beta)^2+|\bar{j}|^2\right)^{-\frac{s}{2}}\\
&\qquad\lesssim q^{d-2}\beta^{-s}+q^{d-2+s}\int_{x\in[0,q)^{d-2}}\left((q\beta)^2+|x|^2\right)^{-\frac{s}{2}}\,dx\\
&\qquad\lesssim q^{d-2}\beta^{-s}+q^{d-2+s}(q\beta)^{d-2-s}\int_{x\in\mathbb{R}^{d-2}}\left(1+|x|^2\right)^{-\frac{s}{2}}\,dx\\
&\qquad\lesssim q^{d-2}\beta^{-s}+q^{d-2+s}(q\beta)^{d-2-s}\\
&\qquad\lesssim q^{2(d-2)}\beta^{d-2-s}.
\end{aligned}
\tag{13}
$$

In evaluating the integral, the change of variable x to $(q\beta)x$ is performed. Note that the above integral is convergent for any value $s>d-2$ (which is fine since we consider s such that $d-1<s<d-\alpha$). The first inequality follows by the change of variable $\bar{j}-\bar{j}'$ to $\bar{j}$ and the second one by integral test. Additionally, since we consider $j_1\neq j_1'$ and then $\beta^2\geq q^{-2}$, the last inequality follows because

$$
\begin{aligned}
&q^{d-2+s}(q\beta)^{d-2-s}=\beta^{-s}q^{d-2+s}(q)^{d-2-s}(\beta)^{d-2}\\
&=\beta^{-s}q^{2(d-2)}(\beta)^{d-2}\geq\beta^{-s}q^{2(d-2)}(q^{-1})^{d-2}\\
&=\beta^{-s}q^{(d-2)}
\end{aligned}
$$

for $d\geq 2$.

Consequently, using (13) and by substituting back for β:

$$III \lesssim q^{-2(d-1)} q^{2(d-2)} \sum_{\substack{j_1 \neq j_1' \\ j_1, j_1' \in [0,q) \cap \mathbb{Z} \\ |j_1 - j_1'| \leq \frac{q}{2b^2}}} \mathbb{E}_H \left(\left(\left| \frac{j_1}{q} - \frac{j_1'}{q} \right|^2 + \left| f_\theta\left(\frac{j_1}{q}\right) - f_\theta\left(\frac{j_1'}{q}\right) \right|^2 \right)^{\frac{d-2-s}{2}} \right)$$

$$= q^{-2} \sum_{\substack{j_1 \neq j_1' \\ j_1, j_1' \in [0,q) \cap \mathbb{Z} \\ |j_1 - j_1'| \leq \frac{q}{2b^2}}} \mathbb{E}_H \left(\left| \left(\frac{j_1}{q}, f_\theta\left(\frac{j_1}{q}\right) \right) - \left(\frac{j_1'}{q}, f_\theta\left(\frac{j_1'}{q}\right) \right) \right|^{d-2-s} \right)$$

$$\lesssim q^{-2} q^{-1+\alpha-d+2+s} \sum_{\substack{j_1 \neq j_1' \\ j_1, j_1' \in [0,q) \cap \mathbb{Z}}} |j_1 - j_1'|^{1-\alpha+d-2-s}$$

$$\lesssim q^{-1+\alpha-d+s} q \sum_{j_1 \in [0,q) \cap \mathbb{Z}} |j_1|^{\alpha+d-1-s} \lesssim q^{-d+\alpha+s} q^{d-\alpha-s} \lesssim 1$$

completing the proof also for the case $d \geq 3$. The second inequality follows thanks to Proposition 14, the third one by the change of variable $j_1 - j_1'$ to j_1 and the fourth one by integral test (where the integral converges for values s strictly smaller than $d - \alpha$, as desired). Note that we could use proposition 14 successfully since $1 = -d + (d-1) + 2 < -d + s + 2 < -d + (d - \alpha) + 2 = 2 - \alpha$. Again, the proof is complete since for any given value $x \in (0, 1]$, we can find values of a and b such that $\alpha = x$. This is a consequence of the fact that $\alpha = -\frac{\log(a)}{\log b}$ and $0 < a < 1 < b$ with $ab \geq 1$. In fact, fixing a positive $a < 1$, $b = \frac{1}{a^x} \geq \frac{1}{a}$ for any value of $x \in (0, 1]$ and $\alpha = x$. □

4.3. **Proof of Theorem 10.** We will proceed by approximating the sum $|A_k|^{-2} \sum_{a \neq a'} |a - a'|^{-s}$ by using sets centered at each $a' \in A_k$ such that for each point a in a given set, $|a_i - a_i'| \approx n_i^{-j_i}$ for all i, where $j_1, \ldots, j_d$ all range from 1 to k. In this case, $|a_i - a_i'| \approx {n_i}^{-j_i}$ means that ${n_i}^{-j_i - 1} < |a_i - a_i'| \leq {n_i}^{-j_i}$. We do not need to consider side lengths with dimensions $n_i^{-j_i}$ for $j_i > k$ because by construction, the i-th discrete Cantor set $C^k_{m_i, n_i}$ does not have any distinct points whose distance is less than n_i^{-k}. Approximating the sum this way gives us

$$|A_k|^{-2} \sum_{a \neq a'} |a - a'|^{-s} \approx |A_k|^{-2} \sum_{j_1, \ldots, j_d = 1}^{k} \sum_{\substack{a \neq a' \\ |a_1 - a_1'| \approx n_1^{-j_1} \\ \vdots \\ |a_d - a_d'| \approx n_d^{-j_d}}} |a - a'|^{-s}. \tag{14}$$

Next, note that for any $a \neq a'$,

$$|a - a'|^{-s} \approx (|a_1 - a_1'| + \cdots + |a_d - a_d'|)^{-s} \leq \max_{1 \leq i \leq n} \{|a_i - a_i'|\}^{-s}. \tag{15}$$

Putting (14) and (15) together, we get

$$
\begin{aligned}
&|A_k|^{-2}\sum_{j_1,\ldots,j_d=1}^{k}\sum_{\substack{a\neq a'\\ |a_1-a_1'|\approx n_1^{-j_1}\\ \vdots\\ |a_d-a_d'|\approx n_d^{-j_d}}}|a-a'|^{-s}\\
(16)\qquad &\approx |A_k|^{-2}\sum_{j_1,\ldots,j_d=1}^{k}\sum_{\substack{a\neq a'\\ |a_1-a_1'|\approx n_1^{-j_1}\\ \vdots\\ |a_d-a_d'|\approx n_d^{-j_d}}}\max_{1\le i\le n}\{|a_i-a_i'|\}^{-s}\\
&= |A_k|^{-2}\sum_{j_1,\ldots,j_d=1}^{k}\sum_{\substack{a\neq a'\\ |a_1-a_1'|\approx n_1^{-j_1}\\ \vdots\\ |a_d-a_d'|\approx n_d^{-j_d}}}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s}.
\end{aligned}
$$

Temporarily fix $a'\in A_k$. For any i, the number of elements $a_i\in C^k_{m_i,n_i}$ such that $|a_i-a_i'|\approx n_i^{-j_i}$ is approximately $\frac{|C^k_{m_i,n_i}|}{m_i^{j_i}}$. Therefore, it follows that the number of $a\in A_k$ such that $|a_i-a_i'|\approx n_i^{-j_i}$ for all i is

$$
\frac{|C^k_{m_1,n_1}|}{m_1^{j_1}}\cdots\frac{|C^k_{m_d,n_d}|}{m_d^{j_d}}=\frac{|A_k|}{m_1^{j_1}\cdots m_d^{j_d}}.
$$

Finally, since there are $|A_k|$ possible choices for a', we can see that

$$
\begin{aligned}
&|A_k|^{-2}\sum_{j_1,\ldots,j_d=1}^{k}\sum_{\substack{a\neq a'\\ +|a_1-a_1'|\approx n_1^{-j_1}\\ \vdots\\ |a_d-a_d'|\approx n_d^{-j_d}}}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s}\\
(17)\qquad &= |A_k|^{-2}\sum_{j_1,\ldots,j_d=1}^{k}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s}\sum_{\substack{a\neq a'\\ +|a_1-a_1'|\approx n_1^{-j_1}\\ \vdots\\ |a_d-a_d'|\approx n_d^{-j_d}}}1\\
&\approx |A_k|^{-2}|A_k||A_k|\sum_{j_1,\ldots,j_d=1}^{k}m_1^{-j_1}\cdots m_d^{-j_d}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s}\\
&= \sum_{j_1,\ldots,j_d=1}^{k}m_1^{-j_1}\cdots m_d^{-j_d}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s}
\end{aligned}
$$

Now, consider the case where $\max_{1\le i\le n}\{n_i^{-j_i}\} = n_1^{-j_1}$. This would mean that for all i,

$$n_1^{-j_1} \ge n_i^{-j_i} \implies n_1^{j_1} \le n_i^{j_i} \implies j_1\frac{\ln(n_1)}{\ln(n_i)} \le j_i.$$

Note that the inequalities in (17) also hold for $n_i^{-j_i}$ when it is maximal. Therefore, we can further split the inner sum in (16) by considering the different cases in which each $n_i^{-j_i}$ is maximal to get

$$\sum_{j_1,\dots,j_d=1}^{k} m_1^{-j_1}\cdots m_d^{-j_d}\max_{1\le i\le n}\{n_i^{-j_i}\}^{-s} = \sum_{j_1\frac{\ln(n_1)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_1^{j_1 s}+\cdots$$
$$+\sum_{j_d\frac{\ln(n_d)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_d^{j_1 s}.$$

We restrict our focus to the first term of the above sum in order to determine for which s the sum converges. For $2 \le i \le n$, the term $n_i^{-j_i}$ in the sum is a geometric series starting at $j_1\frac{\ln(n_1)}{\ln(n_i)}$, so we can approximate each of these terms by $m_i^{j_1\frac{\ln(n_1)}{\ln(n_i)}}$. This gives us that

$$\sum_{j_1\frac{\ln(n_1)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_1^{j_1 s} \approx \sum_{j_1=1}^{n} m_1^{-j_1}\cdots m_d^{-j_1\frac{\ln(n_1)}{\ln(n_d)}}n_1^{j_1 s} = \sum_{j_1=1}^{n}\left(m_1^{-1}\cdots m_d^{-\frac{\ln(n_1)}{\ln(n_d)}}n_1^{s}\right)^{j_1}.$$

Now we have a geometric series, so we know that for this series to converge as $n\to\infty$, we must have that

$$m_1^{-1}\cdots m_d^{-\frac{\ln(n_1)}{\ln(n_d)}}n_1^{s} < 1 \iff -\left(\ln(m_1)+\ln(m_2)\frac{\ln(n_1)}{\ln(n_2)}+\cdots+\ln(m_d)\frac{\ln(n_1)}{\ln(n_d)}\right)+s\ln(n_1) < 0$$
$$\iff s < \frac{\ln(m_1)}{\ln(n_1)}+\cdots+\frac{\ln(m_d)}{\ln(n_d)}.$$

We can use an identical argument to show that all of the other sums

$$\sum_{j_l\frac{\ln(n_1)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_l^{j_l s}$$

also converge exactly when

$$s < \frac{\ln(m_1)}{\ln(n_1)}+\cdots+\frac{\ln(m_d)}{\ln(n_d)}.$$

Furthermore, since

$$I_s(A_k) = |A_k|^{-2}\sum_{a\ne a'}|a-a'|^{-s} \approx \sum_{j_1\frac{\ln(n_1)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_1^{j_1 s}+\cdots+\sum_{j_d\frac{\ln(n_d)}{\ln(n_i)}\le j_i\le k} m_1^{-j_1}\cdots m_d^{-j_d}n_d^{j_1 s},$$

this implies that $\{I_s(A_k)\}_{k\ge1}$ converges exactly when

$$s < \frac{\ln(m_1)}{\ln(n_1)}+\cdots+\frac{\ln(m_d)}{\ln(n_d)}.$$

Hence, $\{A_k\}_{k\geq 1}$ is s-adaptable for exactly these values of s, so by definition, we may conclude that

$$s_{critical} = \frac{\ln(m_1)}{\ln(n_1)} + \cdots + \frac{\ln(m_d)}{\ln(n_d)}$$

as desired.

With this result, if we revisit the discrete Cantor set products from before, then we get that the dimension of $\{C^k_{2,3} \times C^k_{2,3}\}_{k\geq 1}$ is $\frac{\ln(2)}{\ln(3)} + \frac{\ln(2)}{\ln(3)} = 2\frac{\ln(2)}{\ln(3)}$, the dimension of $\{C^k_{2,4} \times C^k_{2,4}\}_{k\geq 1}$ is $\frac{\ln(2)}{\ln(4)} + \frac{\ln(2)}{\ln(4)} = 1$ and the dimension of $\{C^k_{2,4} \times C^k_{2,3} \times C^k_{2,3}\}_{k\geq 1}$ is $\frac{\ln(2)}{\ln(4)} + \frac{\ln(2)}{\ln(3)} + \frac{\ln(2)}{\ln(3)} = \frac{1}{2} + 2\frac{\ln(2)}{\ln(3)}$, which makes sense given how sparse these sets are. This demonstrates that our alternate notion of dimension is better at capturing the dimensionality of discrete Cantor set products than PCA.

References

[1] Alexander Bain, *Mind and Body: The Theories of Their Relation*, New York: D. Appleton and Company (1873).

[2] E. Alos and J. Lion, *An Intuitive Introduction to Fractional and Rough Volatilities*, MDPI, (2021).

[3] A. Borodachov, D. Hardin, and E. Saff, *Discrete energy on rectifiable sets*, Springer Monographs in Mathematics. Springer, New York, (2019).

[4] H. Cohn and N. Elkies, *New upper bounds on sphere packings* I. Ann. of Math. (2) **157** (2003), no. 2, 689-714.

[5] M. Deisenroth, A. Faisal, and C. Ong, *Mathematics for machine learning*, Cambridge University Press, (2020).

[6] G. Folland, *Fourier Analysis and Its Applications*, Pure and Applied undergraduate text, American Mathematical Society 1992.

[7] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, (2014).

[8] L. Guth, A. Iosevich, Y. Ou and H. Wang, *On Falconer's distance problem in the plane*, arXiv:1808.09346 Invent. Math. 219 (2020), no. 3, 779-830.

[9] A. Iosevich, M. Rudnev, and I. Uriarte-Tuero, *Theory of dimension for large discrete sets and applications* (English summary) Math. Model. Nat. Phenom. **9** (2014), no. 5, 148-169.

[10] W. James, *The Principles of Psychology*, New York: H. Holt and Company, (1890).

[11] B. Mandelbrot, *How fractals can explain what's wrong with Wall Street*, Scientific American, September 15, (2008).

[12] F. Rosenblatt, *The Perceptron: A Probalistic Model For Information Storage And Organization In The Brain*, Psychological Review. **65** (6): 386-408, (1958).

[13] R. Smalley Jr, J. Chatelain, D. Turcotte, and R. Prevot, *A fractal approach to the clustering of earthquakes: applications to the seismicity of the new hebrides*, Bulletin of the Seismological Society of America, Vol. **77**, No. 4, pp. 1368-1381, (1987).

[14] D. Turcotte, *Fractals and fragmentation*, J. Geophys. Res. **91**, 1921-1926, (1986).

[15] D. Turcotte, *A fractal model for crustal deformation*, Tectonophysics **132**, 261-269, (1987).

[16] P. Werbos, *Beyond regression: New tools for prediction and analysis in the behavioral sciences*, Thesis (Ph.D.)- Harvard University, (1975).

[17] Brian R. Hunt,*The Hausdorff dimension of graphs of Weierstrass functions*, Proceedings of the American Mathematical Society, (1998)