Fractal dimension computation from equal mass partitions

Yui Shiozawa, Bruce N. Miller, and Jean-Louis Rouet

Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science **24**, 033106 (2014); doi: 10.1063/1.4885778
View online: http://dx.doi.org/10.1063/1.4885778
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/24/3?ver=pdfcov
Published by the AIP Publishing

Fractal dimension computation from equal mass partitions

Yui Shiozawa,[1,a)] Bruce N. Miller,[1,b)] and Jean-Louis Rouet[2,c)]
[1]*Department of Physics and Astronomy, Texas Christian University, Fort Worth, Texas 76129, USA*
[2]*Univ d'Orléans, ISTO, UMR 7327, 45071 Orléans, France; CNRS/INSU, ISTO, UMR 7327, 45071 Orléans, France; and BRGM, ISTO, UMR 7327, BP 36009, 45060 Orléans, France*

(Received 29 December 2013; accepted 18 June 2014; published online 8 July 2014)

Numerical methods which utilize partitions of equal-size, including the box-counting method, remain the most popular choice for computing the generalized dimension of multifractal sets. However, it is known that mass-oriented methods generate relatively good results for computing generalized dimensions for important cases where the box-counting method is known to fail. Here, we revisit two mass-oriented methods and discuss their strengths and limitations. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4885778]

Fractal sets are characterized by self-similarity, and power laws can be associated with them. Examples of fractals in nature are ubiquitous. Their discovery led to the extension of the notion of dimension. For monofractals, the scaling pattern is homogeneous and the set can be characterized by a single dimension. In contrast, multifractals are inhomogeneous and require a spectrum of dimensions D_q to capture their geometry. In finding the generalized dimensions, the box-counting method has been by far the most popular choice among researchers across various fields. However, it is known that the class of methods which deal with partitions into cells of equal size, including the box-counting method, is ill-suited for computing the generalized dimensions on some domain of q. In this paper, two alternative methods which utilize mass-oriented partitions, rather than partitions of equal-size, are investigated.

I. INTRODUCTION

The box-counting method has been the most popular among researchers despite its difficulty to accurately compute the generalized dimension D_q[1] in the negative q range.[2] In this work, we revisit two known alternative numerical methods for obtaining generalized fractal dimensions and discuss their strengths and difficulties. Unlike the box-counting method[3] and the related correlation method,[3] which employ partitions composed of equal-sized cells, the two methods examined in this paper employ mass-oriented partitions. The nearest neighbor method[4] utilizes partitions composed of equal-mass cells while the k-neighbor method[5] uses partitions composed of cells with cumulative mass. These alternative approaches enable one to compute the generalized dimension on the domain where the box-counting method encounters difficulty. The comparison between the correlation method and the nearest neighbor method has been made by Kostelich and Swinney.[6]

This work was originally motivated by the emergence of fractal patterns on the one-dimensional expanding universe model.[7,8] Therefore, our focus is on one-dimensional sets although the numerical methods used in this paper can be applied to higher dimensional spaces. In particular, we applied the methods to the generalized Cantor set.[3] The generalized dimensions of the generalized Cantor set can be readily derived analytically, thus permitting the accuracy of the numerical methods to be verified. Moreover, numerical methods need to deal with finite samples which often give rise to technical difficulties. While a true mathematical fractal is characterized by an infinite nesting structure, fractal-like objects found in nature have a limited hierarchal structure and the range of scales where a power law is observed is finite. Accordingly, when employing a numerical method, one is required to determine the applicability of the method in relation to a finite sampling process. The generalized Cantor set is an ideal set in that the degree of hierarchy can be readily controlled.

The paper is organized as follows: In Sec. II, the important definitions and notations are stated. In Sec. III, we explain the nearest neighbor method and the k-neighbor method in depth. Section IV includes an overview of our results and various raw data obtained using the aforementioned methods. Mathematical methods are employed to analyze the results in Sec. V. In Sec. VI, a summary and conclusions are provided.

II. DEFINITIONS

A. Generalized Cantor set

Starting with a set consisting of a fixed interval of size l on the real line, consider a process whereby a smaller interval is removed from the center leaving intervals of size l_0 and l_1 on the left and right. Further, assign weights p_0 and p_1 to each subinterval, respectively. Now consider repeating the process indefinitely to the remaining interval with the same ratios l_i/l and weights p_i for $\{i = 0, 1\}$. The surviving set is referred to as the generalized Cantor set. For numerical simulation, the procedure is terminated after m^{th} iteration, yielding a finite representation of the Cantor set with degree of hierarchy m. In particular, the uniform Cantor set is obtained by setting $l_0 = l_1 = l/3$, where l is the original length, so the

a)Electronic mail: yui.shiozawa@tcu.edu
b)Electronic mail: b.miller@tcu.edu
c)Electronic mail: jean-louis.rouet@univ-orleans.fr

middle $\frac{1}{3}$ is removed and the ratios and weights are taken equal. Another special case, referred to as the multiplicative binomial process, or MBP, is defined by $l_0 = l_1 = l/2$, so there is no central interval, and the weights are arbitrary.[9]

B. Rényi dimension

As mentioned in the Introduction, the traditional notion of dimension can be extended to generate a spectrum of dimensions for a given set. Suppose $C = \{U_i\}$ is a cover of a set $A \subset \mathbb{R}^n$. Let n_i denote the number of points in U_i among n randomly chosen points from A. Then p_i is associated with U_i for each i by $p_i = \lim_{n\to\infty} \frac{n_i}{n}$. For any real number $q \neq 1$, the generalized dimension D_q, also known as Rényi Dimension, for a set A is given by[10]

$$D_q = -\frac{1}{1-q}\lim_{\epsilon\to 0}\frac{\ln\sum_{i=1}^{N(\epsilon)} p_i^q}{\ln\epsilon}, \tag{1}$$

where $N(\epsilon)$ is the number of sets with diameter $d(U_i) = \epsilon$ required to cover the set A. For $q = 1$, the limiting case where $q \to 1$ is used. There is no explicit formula for D_q when $l_1 \neq l_2$, but the dimension D_q can be found from an implicit relationship that employs the spectrum of scaling indices $f(\alpha)$ and the Legendre transform.[11] For a general set, it is often difficult, if not impossible, to find appropriate covers. Thus methods which permit numerical simulations are important.

III. NUMERICAL METHODS

In this section, three numerical methods for computing the Rényi Dimensions, namely, the box-counting, the nearest-neighbor, and the k-neighbor methods are discussed. The latter two belong to a class of methods with mass-oriented partitions whereas the box-counting method employs a size-oriented partition. It is briefly explained here to illustrate the different approaches in the choice of partitions and their limitations.

A. Box-counting method

This method is probably the most well-known and is closely related to the original definition of the Rényi Dimensions. There are a few slightly different versions that fall under the name "box-counting methods," using "spheres" instead of "boxes," for example,[12] but the underlying ideas are similar: generally, the number of cells required to cover the points in a given set, N, changes as the size of the partitions ϵ changes. The scaling relation can be extracted for a fractal set as the size of the partitions decreases. Namely, for $q = 0$, Eq. (1) reduces to

$$D_0 = -\lim_{\epsilon\to 0}\frac{\ln N(\epsilon)}{\ln\epsilon}. \tag{2}$$

Due to the simplicity of the method, it is widely used among researchers. However, it has been pointed out by many that this method and, more generally, methods that involve partitions of the same size such as the correlation method, do not work well for $q < 1$.[2]

B. Nearest neighbor method

The approach called the "nearest neighbor method" was first introduced by Badii and Politi.[4] This method is essentially based on their observation that there is an exponent D such that

$$\langle\delta\rangle \sim n^{-\frac{1}{D}}, \tag{3}$$

where $\langle\delta\rangle$ denotes the mean distance from each point to its nearest neighbor among n randomly chosen points from a given test set and, as discussed earlier. By naturally extending the premise, the Dimension Function $D(\gamma)$ can be computed by using the moments of order γ of the distribution function $P(\delta, n)$ generated by an ensemble of n randomly chosen points

$$\langle\delta^\gamma\rangle \equiv M_\gamma(n) \equiv \int_0^\infty \delta^\gamma P(\delta, n)d\delta = Kn^{-\frac{\gamma}{D(\gamma)}}, \tag{4}$$

where K is some function of n and γ which asymptotically remains bounded as n becomes large. Here, the meaning of γ should be clear; for positive values of γ, the contribution of high-density regions is suppressed since they generate smaller values of δ, the distance to the nearest neighbor, and vice-versa. The proof of a more general relation is provided by van de Water and Schram.[5] Therefore, the nearest-neighbor approach may be regarded as a special case of more general scaling relations which will be discussed in Sec. III C. From Eq. (4), it follows that the Dimension Function $D(\gamma)$ can be obtained by

$$D(\gamma) = -\lim_{n\to\infty}\frac{\gamma\ln n}{\ln M_\gamma(n)}. \tag{5}$$

The function K generally depends on n and γ but K should be, by definition, irrelevant in the limiting case as in Eq. (5). In numerical analysis, the value of $K(n, \gamma)$ does affect the numerical result since n is finite. The scaling property of Eq. (5) for the uniform Cantor set is shown in Fig. 1. The simulated results for $\gamma\ln(n)$ vs. $-\ln(M_\gamma(n))$ are plotted for a

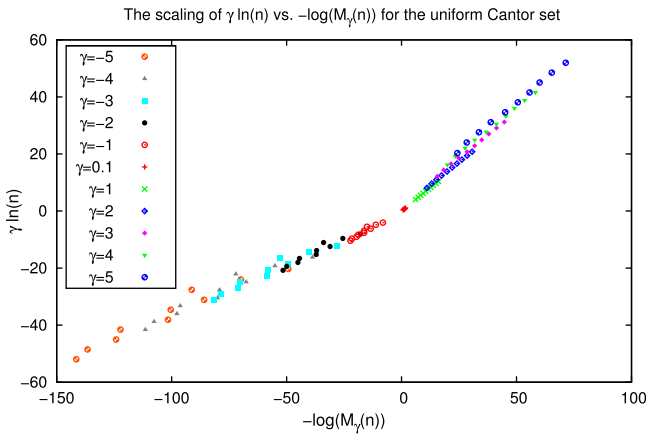

FIG. 1. For the uniform Cantor set, $\gamma\ln(n)$ vs. $-\ln(M_\gamma(n))$ is plotted for each γ as n is increased. According to Eq. (5), the slope converges to $D(\gamma)$. The corresponding result for $D(\gamma)$ is shown in Fig. 2.

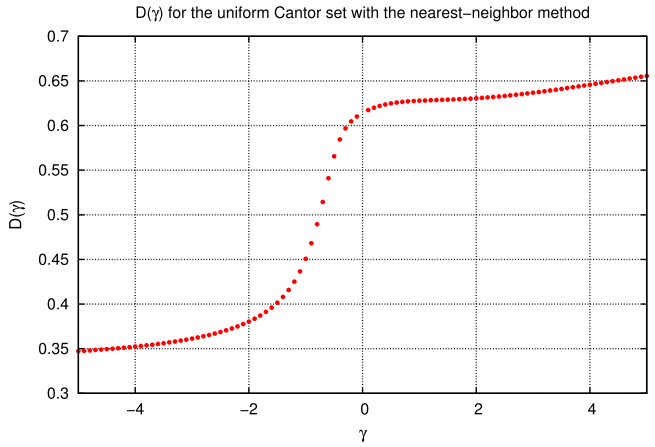

FIG. 2. In this graph, the Dimension Function $D(\gamma)$ for the Uniform Cantor set was computed as the slope of the best-fit line to the corresponding data set which is partially plotted in Fig. 1. $D(\gamma)$ diverges strongly from the analytical result which is log2/log3 for negative γ.

selected set of γ as n increases. For the 10 different values of n selected, the scaling property is clearly observed. In Fig. 2, the value of $D(\gamma)$ was extracted as the slope of the best-fit line in Fig. 1 for each corresponding γ. The slope values in the positive γ range agree well with the analytical results.

The Dimension Function $D(\gamma)$ can be thought of as an alternative generalized dimension and is related to the Rényi Dimension by[4]

$$D[\gamma = (1-q)D_q] = D_q. \tag{6}$$

As the equation suggests, once $D(\gamma)$ is obtained, the generalized dimension D_q can be found as the intersection of $D(\gamma)$ and the straight line with slope $(1-q)^{-1}$ which passes through the origin as illustrated in Fig. 3.

For most cases, the generalized dimension D_q is uniquely determined from $D(\gamma)$. Note that a larger q does not correspond to a larger γ but rather, due to the negative sign in the equation, the limit $q \to \infty$ corresponds to $\gamma \to -\infty$, and vice-versa. Therefore, the index γ plays a similar role as q in that it discriminates the range of density of a given set that most

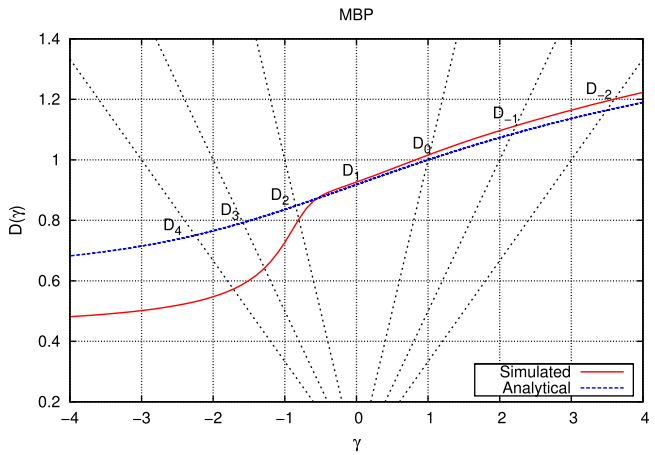

FIG. 3. The solid curve is the simulated result of the Dimension Function for MBP using the nearest neighbor method. Note how D_q can be obtained by locating the corresponding intersections. For example, the box-counting dimension D_0 can be found at the intersection of $D(\gamma)$ and $y = \gamma$.

strongly contributes to $D(\gamma)$. In simulations, the Dimension Function $D(\gamma)$ is obtained using Eq. (5). The formula can, in principle, be applied to sets with any embedding dimension. In the case of a one-dimensional set, sample points are prepared in a way that δ is bounded from above by 1. Therefore, the integral in Eq. (4) can be taken from 0 to 1. Unlike the box-counting method, this algorithm does not make use of partitions of the same size but, rather, of the same "mass" for it can be considered that each element of the partition contains two points, a reference point and its nearest neighbor. Badii and Politi used a slightly improved version of the method, namely, "near-neighbor" method, which uses partitions containing three or four points to smooth out local statistical anomalies.[4] Broggi used partitions containing up to 300 points for systems of large dimensionality.[13] For all these approaches, the number of sample points in a cell is fixed while the total number of sample points n is increased when extracting the Dimension Function. Therefore, these methods differ from the k-neighbor method mainly in that the scaling of cell-size with n is used.

C. *k*-neighbor method

Another method, called "k-neighbor," is similar to the nearest neighbor method in that its partitions are taken according to the number of points inside. However, it is based on the scaling of a moment generating function with k and therefore incorporates a cumulative collection of partitions, one for each value of k selected. Therefore, the scaling property is obtained through the global structure of a given set. A similar global approach with size-oriented partitions was introduced by Tél *et al.*[14] using elements of different size, rather than different mass, and some literature misleadingly refers to it as the "cumulative mass" method.[15] The k-neighbor method records the distance $\delta(k, n)$ from a reference point to the k^{th} neighbor point among $n-1$ randomly chosen points from a given set. van de Water and Schram formulated a technique for evaluating $D(\gamma)$ from the average of $\delta(k, n)^\gamma$ by using the local dimension introduced in Sec. II.[5] The average of $\delta(k, n)^\gamma$ is defined as follows:

$$\Delta^{(\gamma)}(k,n) = \frac{1}{n}\sum_{j=1}^{n} \delta_j^\gamma(k,n), \tag{7}$$

where $\delta_j(k, n)$ represents the k^{th} neighbor distance from the j^{th} reference point when n points are randomly chosen from a test set. Here, all n sample points are used as reference points. When n is large, it can be shown that[5]

$$\langle \Delta^\gamma(k,n)\rangle^{1/\gamma} \cong n^{-1/D(\gamma)}\left[\alpha D(\gamma)\frac{\Gamma(k+\gamma/D(\gamma))}{\Gamma(k)}\right]^{1/\gamma}, \tag{8}$$

where α is some constant independent of γ. Note that the average of δ_j^γ from a single set is used in Eq. (7), whereas the derivation of Eq. (8) is based on the ensemble probability. For large k, a simple approximate relation can be obtained[5]

$$\left[\Delta^{(\gamma)}(k,n)\right]^{1/\gamma} \cong n^{-1/D(\gamma)}k^{1/D(\gamma)}G(k,\gamma), \tag{9}$$

where $G(k, \gamma)$ is a correction function close to unity for large k. The Dimension Function $D(\gamma)$ can be estimated by setting $G(k, \gamma) = 1$ in the first iteration. The dependence of the correction function $G(k, \gamma)$ on k and γ can be obtained from Eq. (8) with the value of $D(\gamma)$ from the first iteration. The Dimension Function $D(\gamma)$ then will be updated using this $G(k, \gamma)$. After a few iterations, the numerical results for $D(\gamma)$ will converge to a single value for each γ. The correction function $G(k, \gamma)$ generally exhibits a periodic pattern as a direct consequence of the self-similarity of fractals as seen in Fig. 12. According to Eq. (9), the Dimension Function $D(\gamma)$ can, in principle, be obtained from the slope of the best-fit straight line in the log-log plot with either a fixed n or k. When k is fixed to 1, the equation reduces to the key relation in Eq. (4) for the nearest neighbor method. For the near-neighbor method, $k = 3$ or 4 may be used. With the k-neighbor method, we used a fixed value of n. By fixing n instead of k, we can extract a global property of a given set. The use of scaling with k makes the k-neighbor method less sensitive to local anomalies which often arise from a finite sampling process.

IV. RESULTS

Generally, with a small amount of computational time, both of the methods in the fixed-mass class give good indications of the Rényi Dimension in the vicinity of the box-counting dimension ($q = 0$) on various generalized Cantor sets. This is a major advantage over the box-counting method if one seeks to find D_0, the box-counting dimension itself. Around the box-counting dimension, the nearest neighbor method yields a result closest to the analytical solutions. However, as q moves away from 0 and hence γ moves away from D_0, the k-neighbor method generally produces more accurate results. Therefore, at this point, no single method seems reliable enough for an extended domain q of the generalized dimension. However, the combination of the aforementioned methods reveals the essential features of a given set such as whether it is a monofractal or multifractal. For a multifractal set, how the dimension changes over the domain q is a key property. The k-neighbor seems to be the best method to start with as it can provide an estimate of the generalized dimension over an extended region, albeit not too accurately especially for $q < 1$. To obtain the dimension to a higher accuracy for a particular q or γ, the box-counting or the nearest neighbor method may be used. For $q > 1$, the box-counting method should be employed and for $q < 1$, the nearest neighbor, provided that q is not a very large negative number. Therefore, if possible, the results obtained from these methods should be compared and examined to see if they are consistent within the uncertainty of each method.

A. Nearest neighbor method

In the nearest neighbor method, the Dimension Function $D(\gamma)$ was extracted from Eq. (5), where the right hand side reads $-\frac{\gamma \ln n}{\ln M_\gamma(n)}$ before taking the limit. To investigate how it approaches to the limit, $\ln n / \ln M_\gamma$ versus $\ln n$ for the uniform Cantor set was plotted in Fig. 4. The points in the plot indicate how $-\gamma \ln n / \ln M_\gamma$ seemingly approaches the theoretical

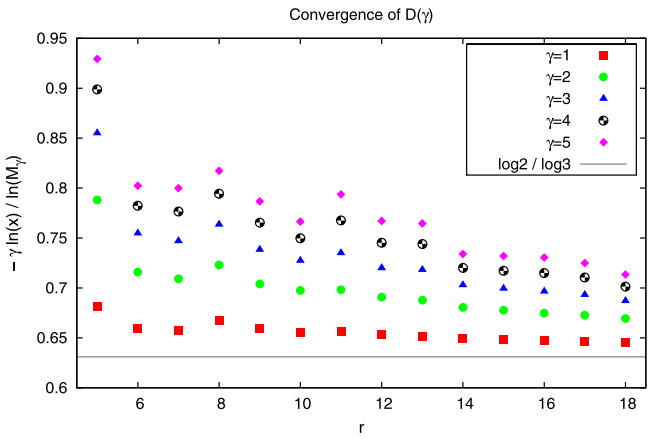

FIG. 4. This figure shows how increasing $n = 2^r$ affects the value of $-\gamma \ln n / \ln M_\gamma$. The plot was generated for the uniform Cantor set. The analytical value for $D(\gamma)$ for all γ is $\log 2 / \log 3 = 0.630...$ which corresponds to the horizontal line in the plot.

limit of $\ln 2 / \ln 3 = 0.63...$ as $\ln(n)$ increases in the case of uniform Cantor set. However, it can be seen that the convergence rate is rather slow. Given that the hierarchy degree m is large enough, increasing n can almost always guarantee a higher accuracy around the box-counting dimension. However, since the convergence rate is rather slow, determining the limit is not a trivial task. For $\gamma = 1$, the number of sample points $n = 2^9 = 512$ was required to obtain the result within 5% accuracy and $n = 2^{17}$ to obtain the result within 3%. For quick simulations, we typically used $n = 2^{16}$ and 10 ensembles. In general, we employed the linear regression technique and obtained the limit from the slope of the appropriate log-log plot. While the overall qualitative features of the Dimension Function, such as the non-decreasing property, are properly reflected on the domain, where γ is positive, the deviations and the fluctuations around $\gamma = -1$ seem sudden and uncontrolled. The difficulty of obtaining a sensible result for $\gamma < -1$ seems persistent throughout the set we have tested. In Fig. 5, the results for various generalized Cantor sets are shown; the domain of γ on which the simulated $D(\gamma)$ agrees well with the analytical results is between 0 and 2. For a multifractal, as γ increases, the numerical results start to diverge from the analytical result as well.

B. *k*-neighbor method

Unlike the nearest neighbor method, where the choice of n is often limited by a finite sample size and the available computation time, the k-neighbor method can utilize a larger data set from which the slope is extracted to estimate $D(\gamma)$. In general, fine structure occurs in the log-log plots, which injects arbitrariness in a slope-fitting process. This point is covered in detail in Sec. V. For a fixed value of n, $D(\gamma)$ or, to be precise, the corresponding $1/D(\gamma)$ in Eq. (9), is taken as the slope of $\log \delta^\gamma(k, n)$ versus $\log(k/n)$. As clearly shown in Fig. 12, the $\delta^\gamma(k, n)$ obtained exhibits a periodic pattern, so all approaches for obtaining the slope seem to inject ambiguity. We have used the standard linear regression technique[16] using sample points equally spaced in the logarithmic scale of k rather than in the linear k scale. Another consideration is

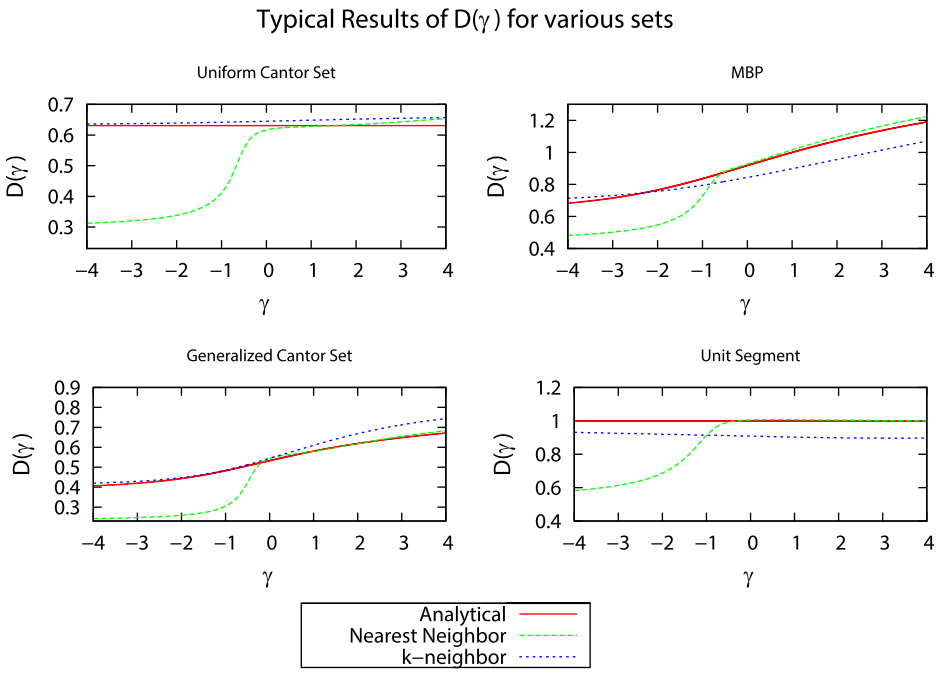

FIG. 5. These plots show typical results of $D(\gamma)$ vs. γ for the nearest neighbor method and the k-neighbor method applied to four different sets. The corresponding analytical results are shown for comparison. "Unit Segment" here means an interval of unit length and can be thought as the 0^{th} finite representation of the Cantor set. For negative γ, numerical results persistently deviate from the analytical results for the nearest neighbor method. While the k-neighbor method works relatively well for all γ, the outcome may not be as accurate as the nearest neighbor method for small positive γ.

that the slope, and therefore the result for $D(\gamma)$, depends on the range over which the linear regression is applied. It turns out that the best range seems to depend on the value of γ as shown in Fig. 6. The plot shows how $D(\gamma)$ varies when the upper bound of the slope range increases for the case of the uniform Cantor set with the analytical dimension of $\log 2/\log 3 = 0.63...$ for all γ values considered.

As a result of these findings, we have used two different boundaries for computing the slope, one for positive γ and the other for negative γ, to produce the final results. Since the inaccuracy inherited from these ambiguities cannot be entirely removed by increasing n as in the nearest neighbor method, it is more difficult for the k-neighbor method to be adjusted to obtain a better result before knowing the theoretical values. Nevertheless, aside from these ambiguities in the method, the k-neighbor works for both positive and negative ranges of q, and therefore, is a good candidate as an initial method to investigate a given set. In the simulation, the ordering of the $n-1$ points for each reference point according to their relative

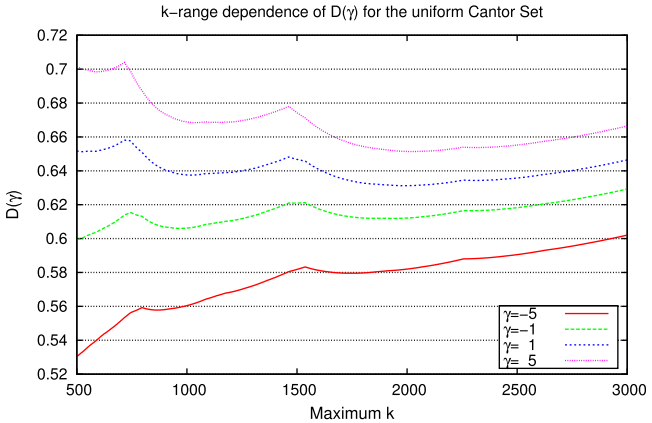

FIG. 6. This plot shows how $D(\gamma)$ differs when a different range is used to extract the slope in the k-neighbor method. For the uniform Cantor set, increasing the upper bound of k generally seems to produce better results. However, this is not a general result.

position takes most of the computational time. Since the ordering takes more time as the embedding dimension increases, the method is said to be especially suited for one-dimensional sets. Furthermore, in contrast with the nearest neighbor method, the hierarchy degree m can be substantially smaller. The size of the scaling region expectedly diminishes as m decreases. However, the Dimension Function deduced from the best-linear-fit from the appropriate scaling region produces acceptable results. For the uniform Cantor set, when m is as small as 5, we obtained $D(\gamma)$ on the order of 0.6 as shown in Fig. 7. This shows that to estimate the fractal dimension from the k-neighbor method, the finite representation does not necessarily require a large degree of hierarchy. Hence, the k-neighbor method is a good candidate for estimating the fractal dimensions when only a limited hierarchy degree is available.

V. ANALYSIS

A. Range and stability

In the nearest neighbor method, the probability distribution of $P(\delta, n)$ plays a key role as seen in Eq. (4). Hence, it is worthwhile to investigate the nature of probability distributions associated with fractal sets. Starting with the conjecture for the mathematical form for the cumulative distribution function for the uniform Cantor set,

$$S(\delta, n) = 1 - \exp[-n(2\delta)^{D_0}], \tag{10}$$

Badii and Politi argue that the correct form of the probability density distribution of uniform Cantor set for $n \gg 1$ is given by[4]

$$P(\delta, n) = 2D_0 n(2\delta)^{D_0 - 1}\exp[-n(2\delta)^{D_0}]. \tag{11}$$

Note that there is a singularity in the gamma function, Eq. (12), for nonpositive integer z[17]

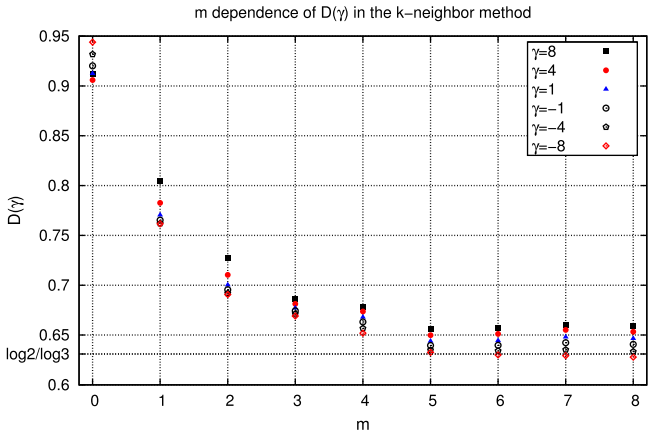

FIG. 7. These plots show how the results for $D(\gamma)$ change as m varies when the k-neighbor method is applied to the m^{th} finite representation of the uniform standard Cantor set. The theoretical value for $D(\gamma)$ is $\log(2)/\log(3)$ for all γ. For all iterations, the value of n is fixed at 10 000. The k-neighbor method provides relatively good results even when m is as small as 5.

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \tag{12}$$

By substituting Eq. (11) into (4), a simple computation yields that

$$M_\gamma(n) = \left(\frac{1}{2n}\right)^{\gamma/D_0} \int_0^\infty x^{\frac{\gamma}{D_0}} e^{-x} dx, \tag{13}$$

$$= \left(\frac{1}{2n}\right)^{\gamma/D_0} \Gamma(\gamma/D_0 + 1), \tag{14}$$

where $x = n(2\delta)^{D_0}$. Therefore, the function $M_\gamma(n)$ involves singularities for $\gamma < -D_0$. This means that, for the generalized Cantor set, the nearest neighbor method is ill-suited for obtaining the correlation dimension ($q = 2$) or D_q for larger q. The function $D(\gamma)$ was calculated for each of four different data sets using the nearest neighbor method and is shown in Fig. 5. In each plot, the numerical results are compared to the corresponding analytical results. The influence of the singularity is observed for a variety of sets. Note that the k-neighbor method does not suffer from this kind of singularity. For the k-neighbor method, the corresponding singularity can be found in Eq. (8). However, this time, the singularity can be avoided by taking a sufficiently large k. Accordingly, the k-neighbor method could generate sensible results in the entire range of γ we have investigated.

It is worth noting that the simulated probability distribution functions did not completely converge to the theoretical distribution of Eq. (11). The Komologov-Smirnov goodness-of-fit test[18] measures the maximum discrepancy between two sample cumulative distributions and was employed to compare the theoretical distribution given by Eq. (10) with different values for D_0 and the distribution obtained in simulations.[18] As seen in Fig. 8, as m increases, the simulated distribution for the uniform Cantor set initially approaches the theoretical distribution when $D_0 = \frac{\ln 2}{\ln 3}$ is inserted in Eq. (10). When the number of points $n = 2^r$ exceeds the number of intervals 2^m, the nearest point for each reference point is likely to fall in the same interval which contains the reference point. This means that the nearest neighbor statistic does not reflect the property of the Cantor set but rather that of a line. Therefore, when m is increased, the result of the K-S goodness-of-fit test constantly decreases as long as $m < r$. One would rationally expect the convergence to improve when m is increased further but this was not observed. The maximum discrepancy reaches a plateau when $m = r$, suggesting that there is a constant disparity between the two distributions which does not diminish even when the finite representation of the Cantor set has a large hierarchy degree. Among the values used, the theoretical distribution with $D = D_0 = \ln 2/\ln 3$ showed the best fit for $m > 14$. In the following, unless otherwise noted, we used the hierarchy degree of $m = 30$ when generating the finite representation of the Cantor set. Using a larger value does not significantly improve the results for the number of sample points we typically used, and would not be consistent with double precision arithmetic employed in the computations.

The effective domain is also related to the stability of the method. For both methods, as $|\gamma|$ increases, the nearest

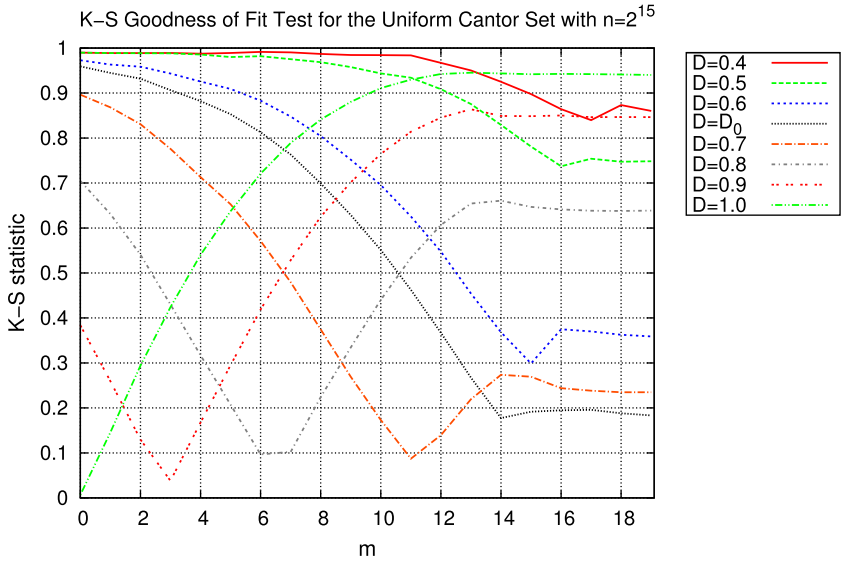

FIG. 8. The Kolmogorov-Smirnov goodness-of-fit test was used to compare the simulated probability density distribution and the theoretical distribution proposed by Badii and Politi for the uniform Cantor set with $n = 2^{15}$. According to Eq. (10), various values between 0 and 1 were substituted for D_0 for the purpose of this test. Smaller values of the outcome indicate a better fit. The finite representation of the Cantor set with $m = 1$ is the unit interval. Therefore, expectedly, the test function with $D = 1$ exhibits the best fit among others. As m increases, the K-S statistic decreases for $D = D_0 = \ln 2/\ln 3$ and similar values. However, they reach plateaus after $m = 15$.

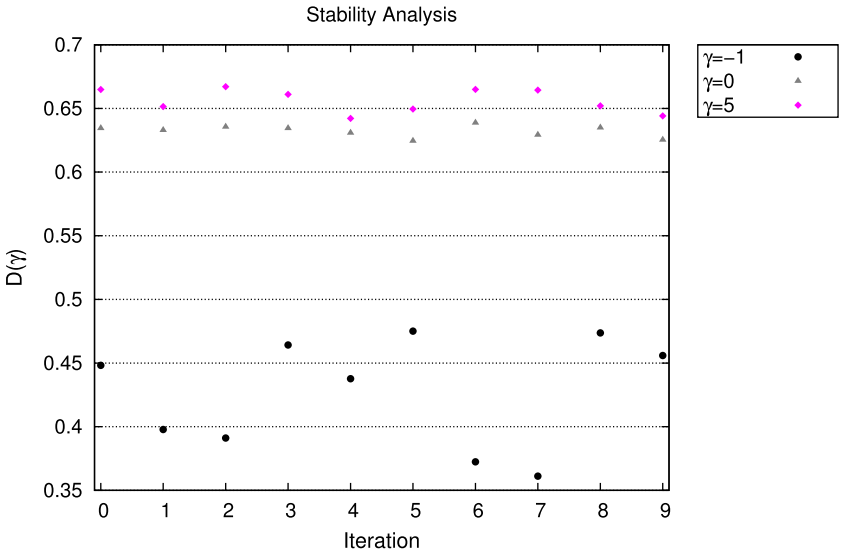

FIG. 9. This figure shows how each iteration of the simulation generates a different outcome for $D(\gamma)$. Each iteration is numbered on the horizontal axis. Sample sets were taken from the uniform Cantor set. In the range where Eq. (4) does not exhibit singularities, the results fluctuate more as γ increases. Larger fluctuation indicates more sensitive dependence on the particular choice of a sample set. The result for $\gamma = -1$ is also included to illustrate the difficulty of the method in the negative range of γ. The outcome in this range fluctuates even more, and the average of the outcome is significantly smaller than the theoretical prediction which is roughly 0.63.

distance, δ, is either amplified or attenuated. Consequently, the contribution from only a few sample points among n chosen points starts to dominate the integral or sum in the equations. Unlike the nearest neighbor method, however, the effect of a few sample points is relatively small in the k-neighbor method due to the global feature. For the nearest neighbor method, simulations require a large number of ensembles and therefore, an extensive amount of computational time and memory for a relatively large (in magnitude) negative γ. How the Dimension Function $D(\gamma)$ varies in each implementation when the nearest neighbor method is applied is shown in Fig. 9. Each iteration is numbered on the horizontal axis. In the positive range of γ, the values of $D(\gamma)$ fluctuate more when computed under the same number of sample points as γ increases.

This difficulty can be partially overcome by employing the "near" neighbor instead of the nearest neighbor as it makes the simulation less dependent on the local property of a single reference point. However, it eventually suffers from the same difficulty as the magnitude of γ increases. The results for $D(\gamma)$ are shown in Fig. 10 when the near neighbor method is used. The integer i denotes the i^{th} neighbor points included in the partitions with $i = 1$ being the nearest neighbor method. Moreover, as i increases, all the relevant equations need to be modified accordingly but the dependence on i is not obvious. Overall, the k-neighbor method has an advantage for large $|\gamma|$.

B. The limitation of numerical methods

As shown in Figs. 11 and 12, plots of the probability distribution $P(\delta, n)$ of δ for the nearest neighbor method or the k^{th} neighbor distance $\delta^{\gamma}(k, n)$ typically exhibit self-similar fine structure which arises from the original fractal geometry. However, unless a construction recipe is known in advance, as in the case of the generalized Cantor set, the exact nature of the fine structure is difficult to obtain. Moreover, to find its exact nature is essentially redundant for it would be

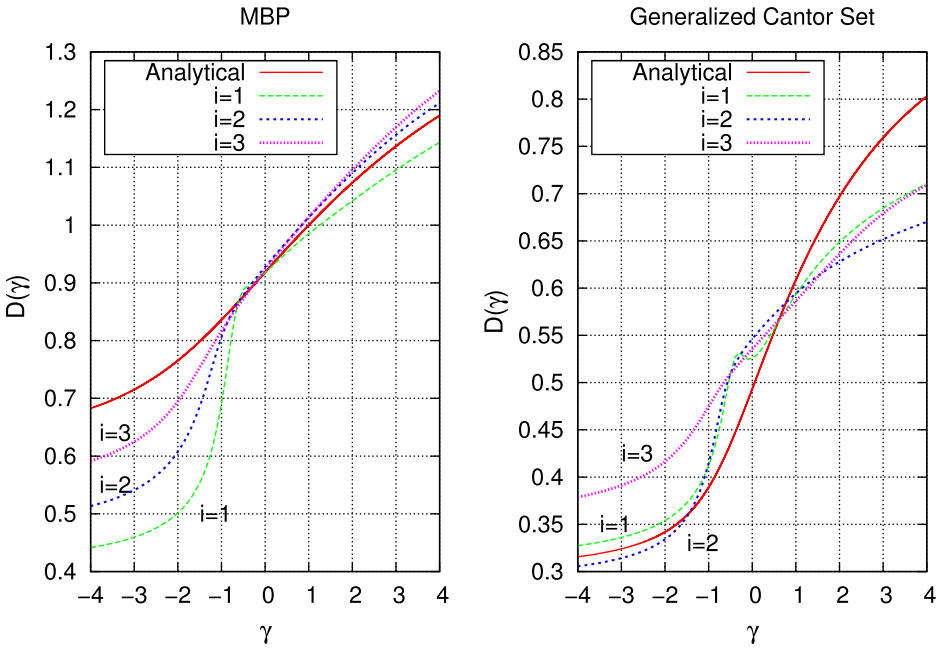

FIG. 10. These plots show how using near neighbor instead the nearest neighbor affects the result. The integer i denotes the i^{th} neighbor. While increasing i generally makes $D(\gamma)$ smoother, one cannot expect that the results improve when i is increased.

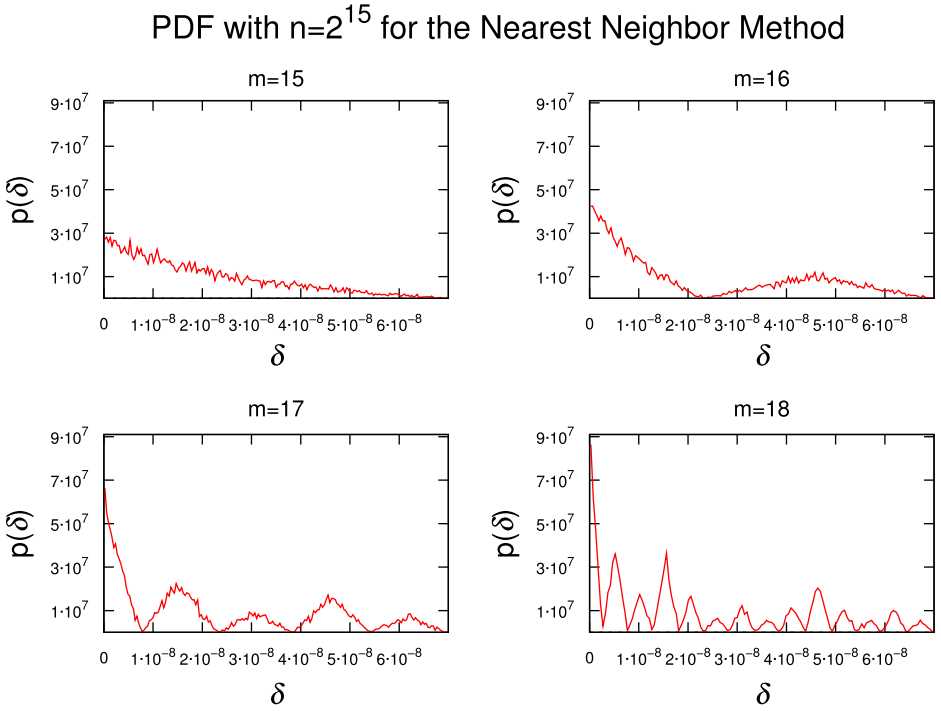

FIG. 11. These plots show how the hierarchy degree m affects the probability distribution of the nearest neighbor method. The sample sets were taken from the uniform Cantor set. While the cumulative distribution is somewhat more stable, as m increases, the fine structure of the probability distribution of δ emerges, exhibiting self-similar patterns. A limited horizontal range from 0 to 3^{-15} is plotted.

another fractal set which is as complex as the original fractal set. Hence, numerical methods are typically developed based on an assumption that these fine structures will not affect their output in any substantial way. Nevertheless, we should not simply ignore the effect of the fine structure as a set would not be a fractal without them. In the equations such as Eqs. (4) and (9), the fine structures are absorbed by the constant or correction term. In general, these correction terms depend on the hierarchy degree used in creating a test set as well as other parameters of these methods. However, it is difficult to estimate the error attributed to the correction term, and therefore this raises a question concerning the reliability of the method.

In principle, the largest possible m should be used to reflect the infinite hierarchical self-similarity. For the nearest neighbor method, the number of reference points, n, needs to be smaller than 2^m. Therefore, to increase n to obtain more accurate results, one needs to increase m as well. However, unlike the case of sample points where increasing n generally guarantees a more accurate result, increasing m does not necessarily do so. As we can see in Fig. 7, once m reaches a certain threshold, increasing m will not produce a better result.

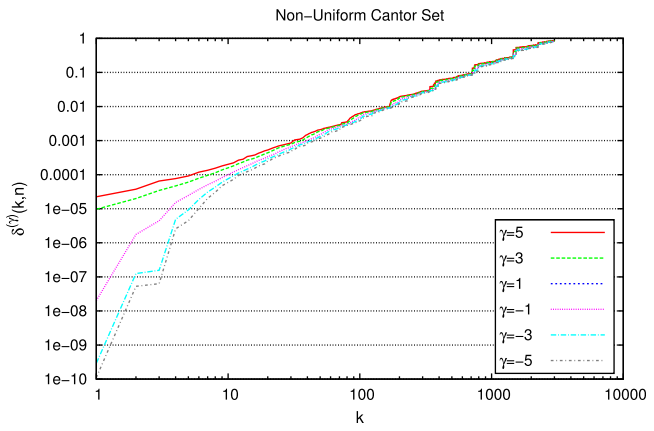

FIG. 12. $\delta^{\gamma}(k,n) = \left(\Delta^{\gamma}(k,n)\right)^{(1/\gamma)}$ is plotted versus k in a log-log plot. The fine structure inherited from the non-uniform Cantor set is observed.

VI. CONCLUSION

In contrast with the box-counting method, or similar methods which utilize partitions into cells of equal size, the nearest neighbor method, which employs partitions of equal mass, as well as the k-neighbor method, which employs partitions of distributed mass, are good candidates for estimating the generalized fractal dimension for negative q. The k-neighbor method works for the complete range of q and no serious deviations were found. By choosing an appropriate scaling region, it is possible to estimate the generalized dimensions even with a small hierarchy degree. However, the method involves linear regression and the results depend on how the best-fit line is obtained. Therefore, the k-neighbor method is a good option for a starting point and to investigate the general outlook of D_q.

If the sample size is large, the nearest neighbor method can be the best method for small negative q. Although the result is sensitive to the local anomalies, one can choose the size of n according to one's required precision to extract the dimension. However, in contrast with the k-neighbor method, the hierarchy degree, m, also needs to be sufficiently large in order to obtain a desirable probability distribution. Therefore, if the sample size of a finite representation is small, the nearest neighbor method is not a practical choice. For positive q, the methods with partitions of equal size may be used.

In general, a few different methods should be applied before one determines if the results from different methods are consistent. The k-neighbor method should provide the overall features of D_q. Given that the subjective choice of the best-fit line affects the result, it is important to determine the window of ambiguity. If the sample size is adequate,

apply the nearest-neighbor method for negative q and box-counting or similar method for positive q. The results from these two different methods should lie within the window of ambiguity. Given the strengths and the limitations of these methods, it would be interesting to apply them to a set with unknown fractal dimensions.

In any simulation of the kind worked out in this paper, the finite sample correction needs to be taken care of. Although a number of correction terms have been proposed over the years,[5,19] many of them add extra complications to the simulation without achieving a dramatic increase in accuracy.[5,13,20] In the process of exploring the form of the nearest neighbor distribution of the generalized Cantor set, some interesting properties have been obtained; the order of taking m and n to infinity may not commute as usually assumed. Since a numerical sample only possesses a finite hierarchy, a new algorithm which does not assume an infinite hierarchy may be useful. In future work, it will be shown that a new analysis of generalized dimension may be based on some quantities that are independent of the hierarchy.

ACKNOWLEDGMENTS

We would like to thank Dr. Igor Prokhorenkov for his insightful and valuable advice.

[1]A. Renyi, "Dimension, entropy and information," in *Transactions of the First Prague Conference on Information Theory* (1960), pp. 545–556.
[2]R. Riedi, "An improved multifractal formalism and self-similar measures," J. Math. Anal. Appl. **189**, 462–490 (1995).
[3]J. Feder, *Fractals* (Plenum Press, New York, 1988).
[4]R. Badii and A. Politi, "Statistical description of chaotic attractors: The dimension function," J. Stat. Phys. **40**, 725–750 (1985).
[5]W. van de Water and P. Schram, "Generalized dimensions from near-neighbor information," Phys. Rev. A **37**, 3118–3125 (1988).
[6]E. J. Kostelich and H. L. Swinney, "Practical considerations in estimating dimension from time series data," Phys. Scr. **40**, 436 (1989).
[7]B. N. Miller and J.-L. Rouet, "Ewald sums for one dimension," Phys. Rev. E **82**, 066203 (2010).
[8]B. N. Miller and J.-L. Rouet, "Cosmology in one dimension: Fractal geometry, power spectra and correlation," J. Stat. Mech. Theor. Exp. **2010**, P12028.
[9]P. M. E. Altham, "Two generalizations of the binomial distribution," J. R. Stat. Soc. Ser. C **27**, 162–167 (1978).
[10]H. Hentschel and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," Physica D **8**, 435–444 (1983).
[11]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**, 1141 (1986).
[12]W. Caswell and J. Yorke, "Invisible errors in dimension calculations: Geometric and systematic effects," in *Dimensions and Entropies in Chaotic Systems* (Springer, 1986), pp. 123–136.
[13]G. Broggi, "Evaluation of dimensions and entropies of chaotic systems," J. Opt. Soc. Am. B **5**, 1020–1028 (1988).
[14]T. Tél, Á. Fülöp, and T. Vicsek, "Determination of fractal dimensions for geometrical multifractals," Physica A **159**, 155–166 (1989).
[15]R. Lopes and N. Betrouni, "Fractal and multifractal analysis: A review," Med. Image Anal. **13**, 634–649 (2009).
[16]W. Press, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, *Cambridge, UK New York*, 2007).
[17]M. Abramowitz, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1970).
[18]W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968).
[19]R. Badii and G. Broggi, "Measurement of the dimension spectrum $f(\alpha)$: Fixed-mass approach," Phys. Lett. A **131**, 339–343 (1988).
[20]P. Grassberger, "Generalizations of the Hausdorff dimension of fractal measures," Phys. Lett. A **107**, 101–105 (1985).