

Accuracy Analysis of the Box-Counting Algorithm

A.Z. GÓRSKI^a, S. DROŹDŹ^{a,b}, A. MOKRZYCKA^{a,c} AND J. PAWLIK^{a,c}

^aH. Niewodniczański Institute of Nuclear Physics, Polish Academy of Sciences

Radzikowskiego 152, Kraków, 31-342, Poland

^bFaculty of Physics, Mathematics and Computer Science, Cracow University of Technology, 31-155 Kraków, Poland

^cAGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland

Accuracy of the box-counting algorithm for numerical computation of the fractal exponents is investigated. To this end several sample mathematical fractal sets are analyzed. It is shown that the standard deviation obtained for the fit of the fractal scaling in the log-log plot strongly underestimates the actual error. The real computational error was found to have power scaling with respect to the number of data points in the sample (n_{tot}). For fractals embedded in two-dimensional space the error is larger than for those embedded in one-dimensional space. For fractal functions the error is even larger. Obtained formula can give more realistic estimates for the computed generalized fractal exponents' accuracy.

PACS: 05.45.Df, 02.60.Gf, 47.52.+j

1. Introduction

In last decades computations of fractal dimensions (exponents) have become very popular in various areas of physics, as well as in interdisciplinary research. Fractal structures have been found in wide spectrum of problems, ranging from high energy physics [1] to cosmology [2] and from medicine [3] to econophysics [4]. In spite of its popularity accuracy of obtained results is usually either not discussed or overestimated. Moreover, it has been found that in quite a few papers wrong numerical results and conclusions have been published [3, 5–8].

The aim of this paper is to calculate fractal exponents for several well known mathematical fractals with the box-counting algorithm to estimate real accuracy of these computations. The dependence of accuracy with respect to the size of the available data set (n_{tot}) is also discussed and its simple scaling properties are found.

The important point is that the statistical errors for computed fractal dimensions is negligible in comparison to the systematic error. This is true not only for different computational algorithms but also within a given algorithm for different size of the data set. Even in this case, one gets systematic error about order of magnitude bigger than the statistical error. This will be shown in the following Sections. Surprisingly, these practically important issues were not sufficiently analyzed in the literature.

It should be stressed that accuracy of fractal exponent computations in principle depends on many factors. For example, the accuracy can be degraded by presence of noise in the data. Also, one can get different results using different box-counting algorithms (see *e.g.* [9, 10]), another computational algorithms [11] or using different digital representations of the investigated physical ob-

ject (picture). Furthermore, one should be very careful translating Hurst exponents into fractal exponents as, in general, there is no simple relation of both [12].

In Sec. 2 we calculate the box-counting fractal exponents for six different fractal sets for various numbers of data points (n_{tot}). The generalized fractal exponent is defined in the standard way [6]

$$d(q) = \frac{1}{1-q} \lim_{N \rightarrow \infty} \frac{\sum_i \log p_i^q(N)}{\log N}, \quad (1)$$

where N denotes the total number of boxes and $p_i(N)$ is the measure of the subset in the i -th box for the given division N . Where the box size $\varepsilon = 1/N$. To find accuracy estimates the obtained results are compared with precise mathematical values of the exponents determined analytically. Furthermore, we calculate standard errors for the linear fits in the log-log plots used to calculate the exponents. Finally, the inverse power fits were found to give fair approximation of accuracy dependence on the size of the data set (n_{tot}). The final Section contains summary and conclusions.

2. Accuracy estimates

To start with we calculate fractal exponents for fractal sets embedded in one-dimensional space, namely the classical Cantor set (CS) [13] and the (multifractal) asymmetric Cantor set (ACS) [14]. The calculations have been repeated for different sizes of the sets, ranging from less than 10^2 up to 10^5 data points. The final results are given in Fig. 1. Crosses indicate the real accuracy of the box-counting algorithm computations, *i.e.* the absolute value of the difference between the calculated exponent and exact analytical result. Circles give standard errors

obtained for the linear fits in the corresponding log-log plots. In addition, the inverse power fit for accuracy as the function of the number of available data points (n_{tot}) is given by the dashed line. The parameter α denotes the exponent of the inverse power fit

$$\text{real error} \sim \frac{1}{n_{tot}^\alpha}. \quad (2)$$

At first glimpse it is clear that the standard error, that is often treated as the accuracy of the algorithm, considerably (up to the order of magnitude) underestimates the real error. In the case of Cantor set (A) we have $\alpha_{CS} \approx 0.50$. Similar result was obtained for the case of asymmetric Cantor set, $\alpha_{ACS} \approx 0.48$. Hence, in these cases one can expect $\alpha \sim 1/2$ and the error of size $\sim 1/\sqrt{n_{tot}}$.

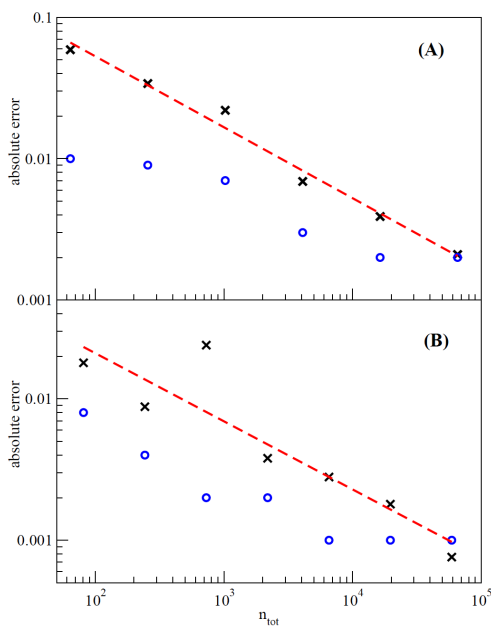


Fig. 1. Real accuracy of the fractal exponent $d(0)$ (crosses) and the standard errors obtained for the linear fit in the log-log plots used to determine fractal exponents (circles). The dashed line is the inverse power fit. The upper plot (A) is for the Cantor set and the lower plot (B) is for the ACS.

As the second step we analyze fractal sets embedded in two-dimensional space: the Sierpiński triangle and the Koch curve. The results are given in Fig. 2, with the same notation as for the Fig. 1. In this case we have $\alpha_{ST} \approx 0.31$ and $\alpha_{KC} \approx 0.18$. Hence, the error scales approximately with $\alpha \sim 1/4$ and the error size is $\sim 1/n_{tot}^{1/4}$. This results is intuitively clear, as to have the same accuracy as for the one-dimensional case the squared number of data points has to be used. In these examples the actual error is also much bigger than the estimated standard error.

For fractals embedded in two-dimensional space special attention should be paid to fractals that are of the function type, *i.e.* there exists a reference frame in which for a given value of one coordinate there is at most one point of the fractal set. Hence, in one direction number

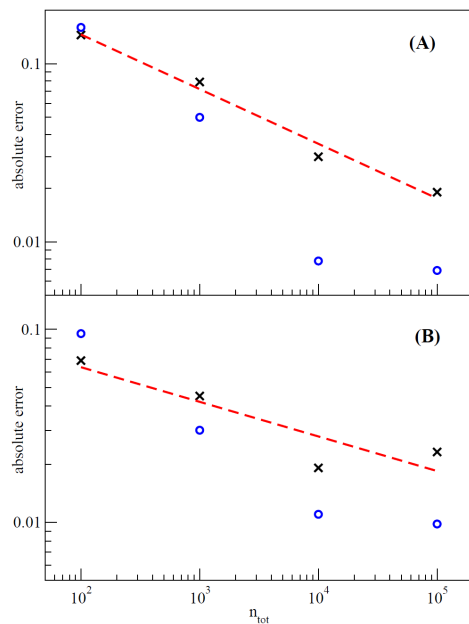


Fig. 2. Real accuracy of the fractal exponent $d(0)$ (crosses) and the standard errors obtained for the linear fit in the log-log plots used to determine fractal exponents (circles). The dashed line is the inverse power fit. The upper plot (A) is for the Sierpiński triangle and the lower plot (B) is for the Koch curve.

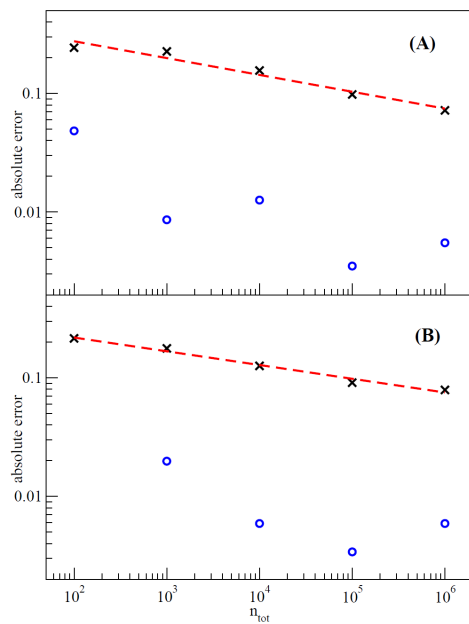


Fig. 3. Real accuracy of the fractal exponent $d(0)$ (crosses) and the standard errors obtained for the linear fit in the log-log plots used to determine fractal exponents (circles). The dashed line is the inverse power fit. The upper plot (A) is for the Weierstrass-Mandelbrot curve with parameter $D = 1.5$ and the lower plot (B) is for the curve with $D = 1.8$.

of points in a given box is limited and this may cause slower convergence of the box-counting scheme (scaling can be observed in one direction only) resulting in bigger errors, slower convergence. Because this type of fractals has wide applicability, *e.g.* in the time series analysis, we will consider this case separately. A good example of such fractal sets with precisely known fractal dimensions are the Weierstrass-Mandelbrot (WM) functions [15]

$$W(t) = \sum_{n=-\infty}^{n=+\infty} \frac{1}{\gamma^{(2-D)n}} [1 - \cos(\gamma^n t)]. \quad (3)$$

We investigate two WM fractal functions with dimensions $d = 1.5$ (A) and $d = 1.8$ (B). The results are shown in Fig. 3. Again one can find a fair inverse power fit for the error with the exponents α_{WM} equal to 0.14 and 0.12, respectively ($\alpha \sim 1/8$ and the error size is $\sim 1/t_{tot}^{1/8}$). Hence, to have similar accuracy as for ordinary fractals embedded in two-dimensional space one has to use squared number of data points (n_{tot}). This is intuitively clear, as in this case the fractal scaling can be observed in only one (instead of two) dimensions. In effect, to have reasonable accuracy for the fractal exponent, a very large number of data points has to be taken into account ($> 10^5$) even though we deal with perfect mathematical fractals without any external noise.

4. Summary and conclusions

It has been shown that for the box-counting algorithm there is a fair inverse power scaling of the actual error of the computed fractal exponents of the type (2). The standard error calculated for the fit in the log-log plot strongly underestimates the actual error leading to the overestimated accuracy. Furthermore, to obtain a given level of accuracy number of data points (n_{tot}) used for fractals embedded in two-dimensional space should be the squared number of data points sufficient for fractals embedded in one-dimensional space. The corresponding exponents α are roughly equal to 1/4 and 1/2, respectively.

Similar phenomenon occurs for fractal functions embedded in two-dimensional space, where the exponent α was found around 1/8. Hence, up to an overall factor, to have the same accuracy as for the Cantor set and 10^2 data points one has to use of order 10^8 data points for the fractal WM function.

The formula (2) and plots in Figs. 1–3 can be used to estimate accuracy of such computation in much higher accuracy than the estimated standard errors for the linear fits. The estimated accuracy of the box-counting algorithm for various sizes of the data sets (n_{tot}) is also presented in Table. It should be stressed that for fractals with additional external noise one can expect even worse results — the errors will be greater.

References

- [1] A. Białas, R. Peszanski, *Nucl. Phys. B* **308**, 803 (1988).
- [2] T. Chmaj, W. Doroba, W. Słomiński, *Z. Phys. C* **50**, 333 (1991).
- [3] A. Z. Górski, J. Skrzat, *J. Anat.* **208**, 353 (2006).
- [4] A.Z. Górski, S. Drożdż, J. Speth, *Physica A* **316**, 496 (2002).
- [5] A. Z. Górski, *Comment on fractality of quantum mechanical energy spectra, preprint arXiv chaos/9804034*.
- [6] A. Z. Górski, *J. Phys.* **A34**, 7933 (2001).
- [7] J. L. McCauley, *Physica A* **309**, 183 (2002).
- [8] P. Castiglioni, *Comput. Biol. Med.* **40**, 950 (2010).
- [9] J. Theiler, *J. Opt. Soc. Am.* **A7**, 1055 (1990).
- [10] A. Roy, E. Perfect, W. M. Dunne, L. D. McKay, *J. Geophys. Res.*, **112**, B12201 (2007).
- [11] J. Brewer, L. D. Girolamo *Atm. Res.*, 433 (2006).
- [12] S. Jaffard, *SIAM J. Math. Anal.* **28** 944; 971, (1997).
- [13] B. B. Mandelbrot, *Fractals: Form, Chance, and Dimension* (San Francisco: Freeman 1977).
- [14] T. Tel, T. Vicsek, *J. Phys.* **A20**, L835 (1987).
- [15] M. V. Berry, Z. V. Lewis, *Proc. R. Soc. Lond.* **A370**, 459 (1980).

TABLE

Absolute errors

n_{tot}	1000	10 000	100 000
1-D fractals	± 0.020	± 0.006	± 0.002
2-D fractals	± 0.060	± 0.030	± 0.020
2-D W-M curve	± 0.200	± 0.150	± 0.100