ACCURACY OF THE BOX COUNTING ALGORITHM

A. Z. Górski, S. Drożdż, A. Mokrzycka, J. Pawlik

Complex Systems Theory Lab., IFJ PAN, Kraków

INTRODUCTION

MOTIVATION:

- ✓ fractal dimensions are very often computed (also in econophysics)
- ✓ usually accuracy of these computations is not discussed or misunderstood (overestimated!)

GOALS:

- ✓ to find an estimate of the accuracy
- ✓ dependence on the number of available data points (n_{tot}) in the sample
- ✓ is there a simple scaling of the error as a function of n_{tot} ?
- ✓ dependence on dimensionality (and other factors)

INTRODUCTION

MATHEMATICAL SUBTLETIES

- ✓ different mathematical definitions of fractal exponents
  [J. Theiler, J. Opt. Soc. Am A7 (1990) 1055]
- ✓ problems for "wild" sets – pseudofractals [AZG, J. Phys. A34 (2001) 7933]
- ✓ problem of equivalence of different computational algorithms
- ✓ indirect computations via generalized Hurst exponents and its traps
  [S. Jaffard, SIAM J. Math. Anal. 28 (1997) 944; 971]

INTRODUCTION

PRACTICAL PROBLEMS

- ✓ real computations ➔ finite samples ➔ such sample can correspond to different fractal sets! ("physical fractal" vs. mathematical fractal)
- ✓ infinite limit in the fractal exponent definition ➔ problems to obtain proper results [AZG, J. Phys. A34 (2001) 7933]
- ✓ linear fit/scaling problems (arbitrary choice of fitted points) [McCauley, Physica A309 (2002) 183; AZG, J. Skrzat, J. Anat. (2006) 208]
- ✓ is "σ" of the log-log fit adequate to estimate the computational accuracy?

- ✓ various computational algorithms
- ✓ various representations of physical objects (e.g. digitalization of continuous quantities like colors, different shapes of covering sets etc.)

INTRODUCTION

BOX COUNTING ALGORITHM

- ✓ n_{tot} — number of data points in a sample
- ✓ $N = 2^1, 2^2, 2^3, \ldots$ — enumerates successive divisions of the scale ($\varepsilon = 1/N$)
- ✓ choice of points that are to be fitted to the power curve – linear log-log fit (quite arbitrary, k – points are fitted, selected by „visual inspection”)
- ✓ determination of the scaling exponent (α) and the standard deviation for the fit (σ)

$$d(q) = \frac{1}{1-q} \lim_{N \to \infty} \frac{\ln \sum_i p_i^{\,q}(N)}{\ln N}$$

- ➢ for small sets (small n_{tot}) k is small and the result can be quite erratic and σ relatively large

COMPUTATIONS – 1D fractals

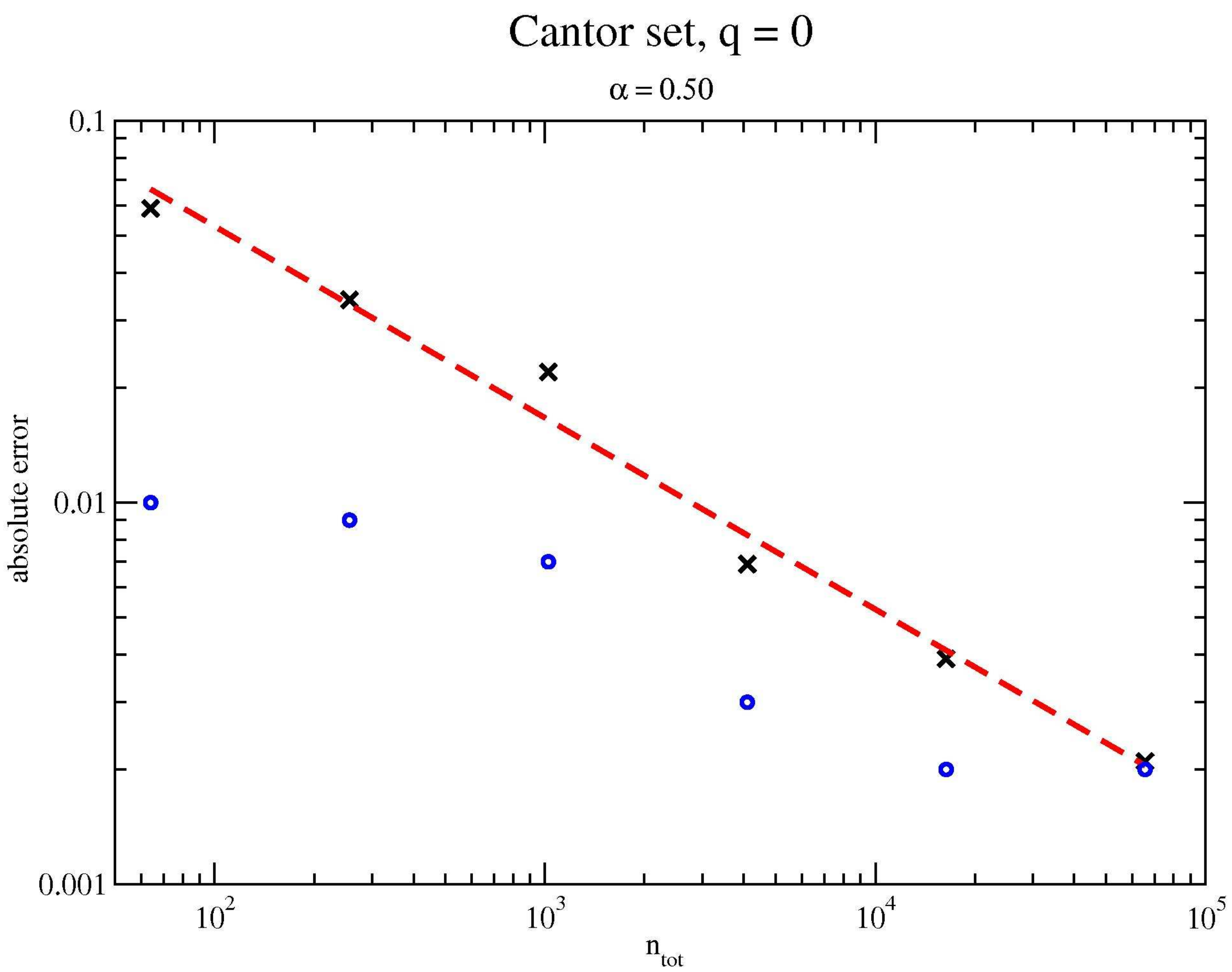
CANTOR SETS

The parameter n_{tot} (set size) is taken form 2^6 to 2^{16} . The mathematical dimension is

$$d = \log 2 / \log 3 = 0.630929\ldots$$

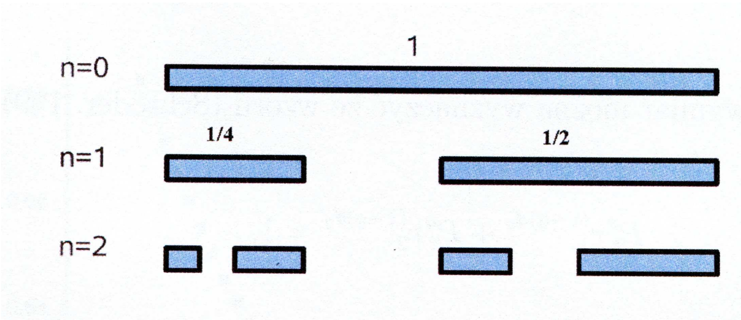
In the plots there are displayed actual absolute errors = |computed result – mathematical result| (black crosses) and standard errors (blue circles) calculated for each fit (value of α) for a given set (n_{tot}).

COMPUTATIONS – 1D fractals



COMPUTATIONS – 1D fractals

ASYMMETRIC CANTOR SETS

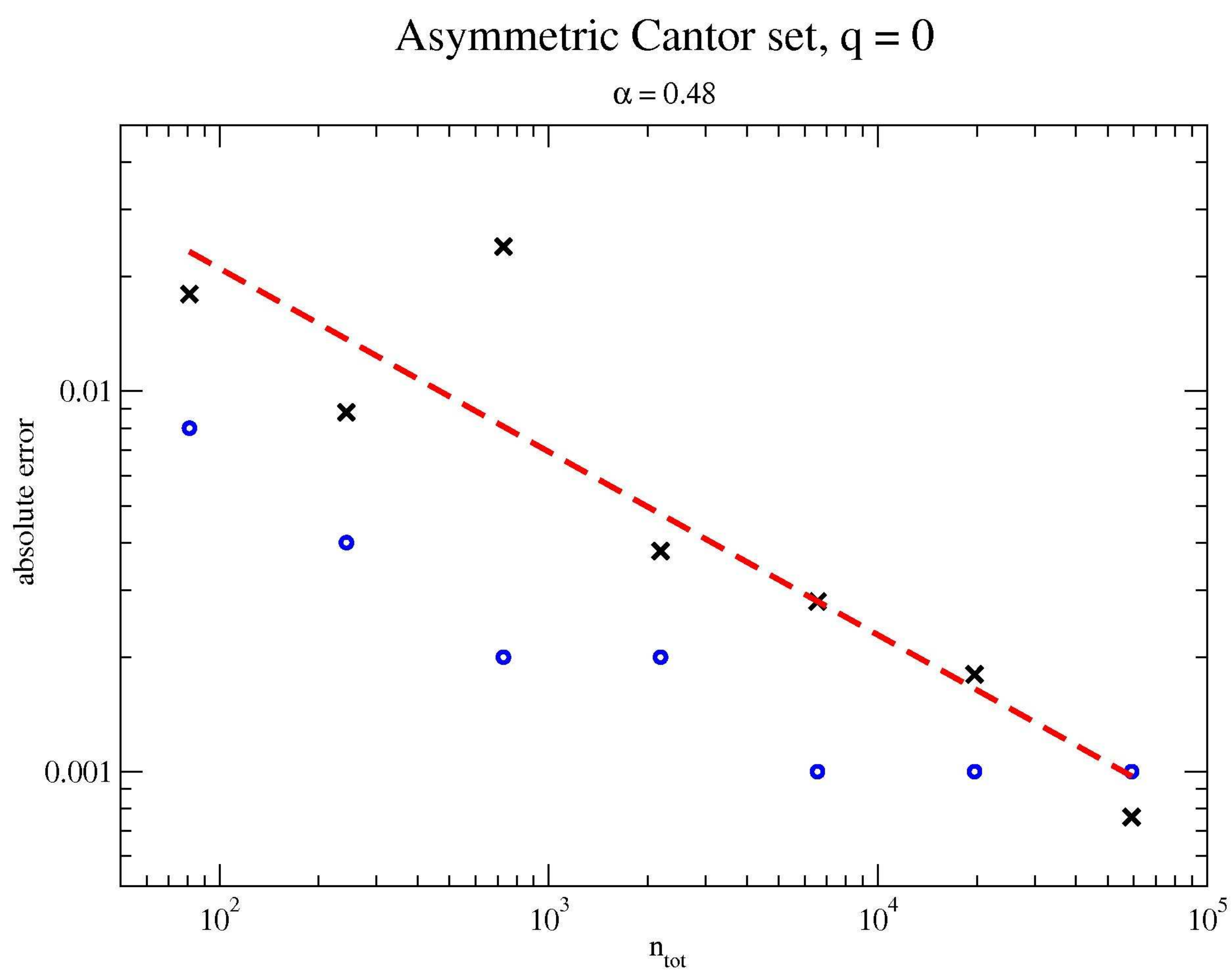


Similar analysis was done for the (multifractal) asymmetric Cantor set for dimensions

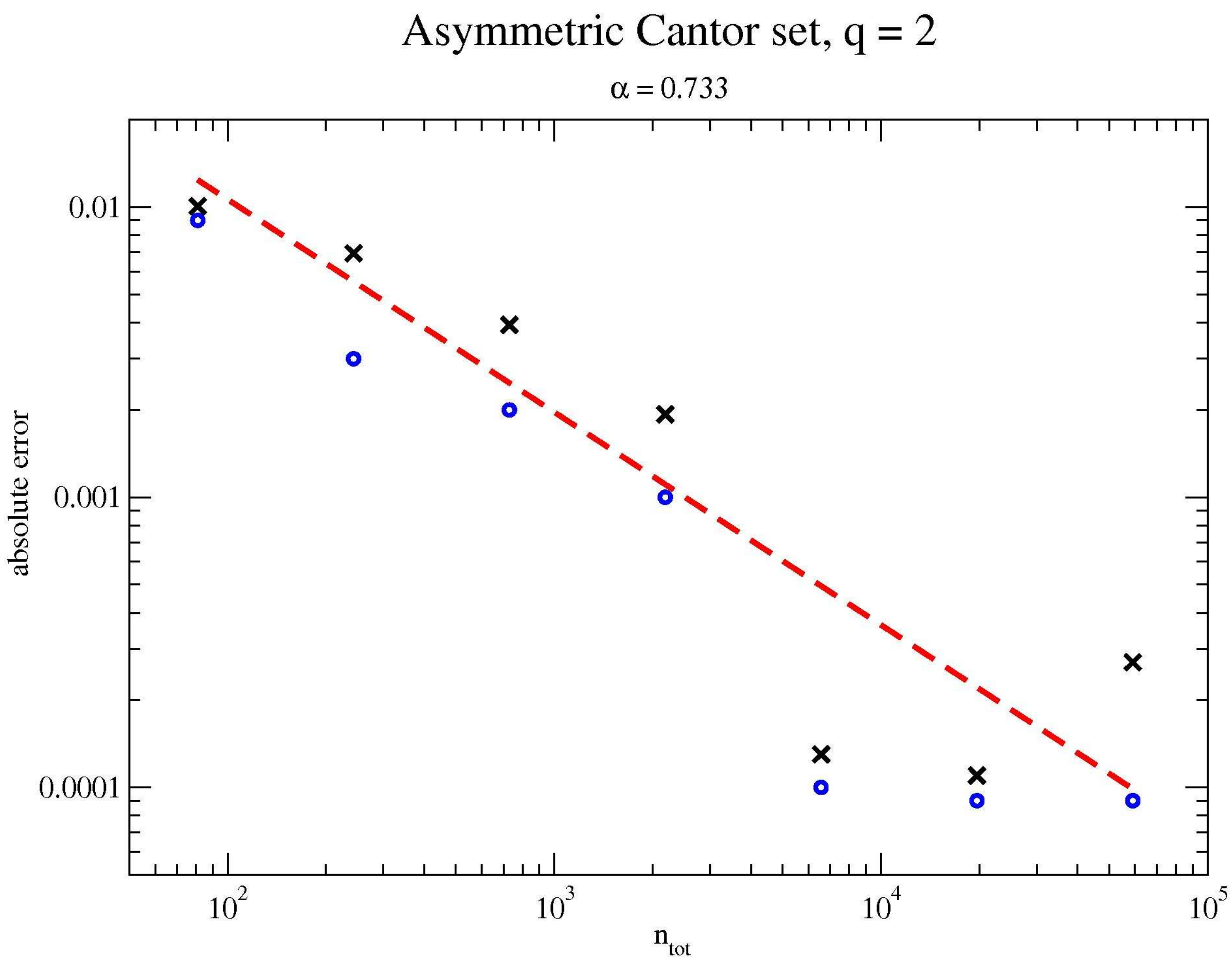
d(0) = 0.6942 and d(2) = 0.6831

The parameter n_{tot} was taken form 3^4 to 3^{10} .

COMPUTATIONS – 1D fractals



COMPUTATIONS – 1D fractals



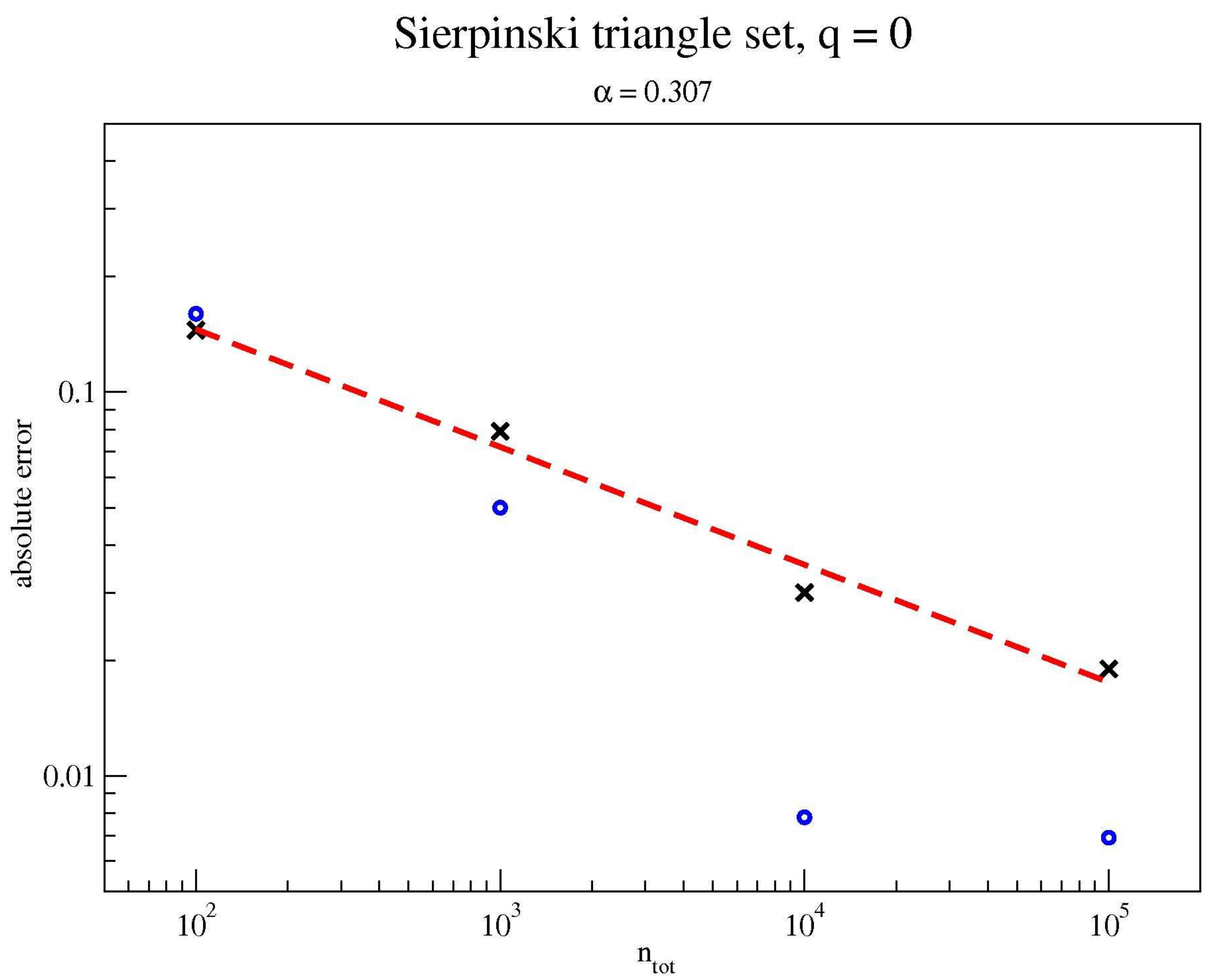
COMPUTATIONS – 2D fractals

SIERPIŃSKI TRIANGLE

The parameter n_{tot} (set size) is taken form 10^2 to 10^5 . The mathematical dimension is

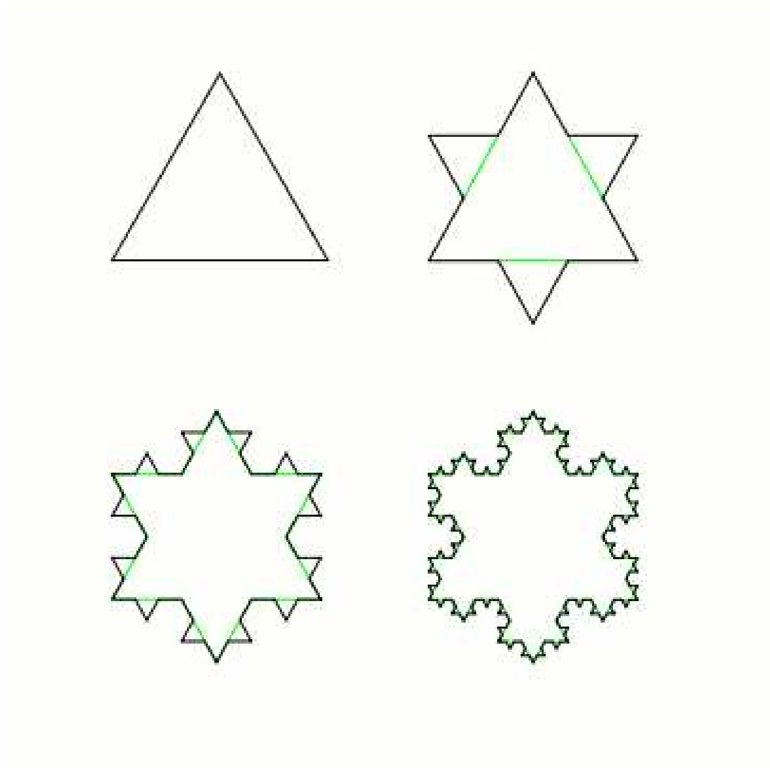
$$d = \log 3 / \log 2 = 1.58496$$

COMPUTATIONS – 2D fractals



COMPUTATIONS – 2D fractals

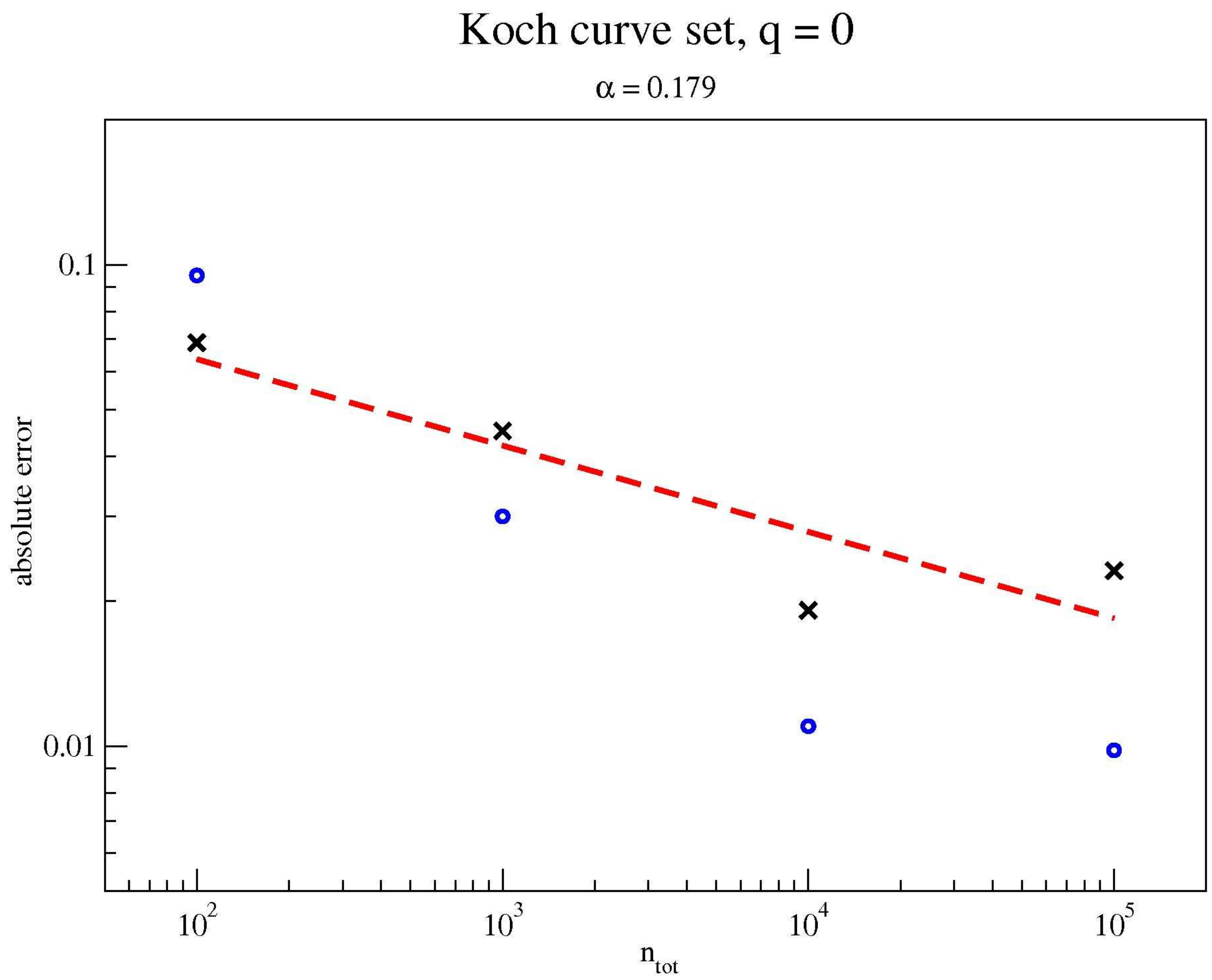
KOCH CURVE



The parameter n_{tot} (set size) is taken form 10^2 to 10^5 . The mathematical dimension is

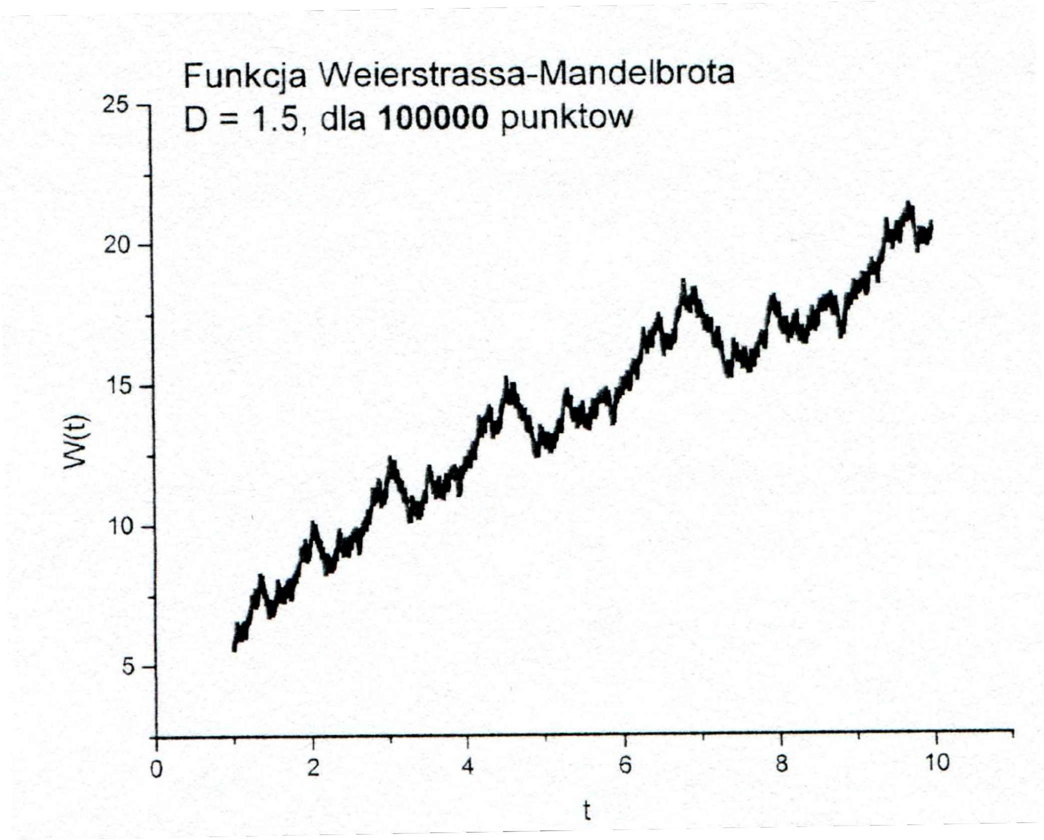
$$d = \log 4 / \log 3 = 1.26186$$

COMPUTATIONS – 2D fractals



COMPUTATIONS – 2D fractals

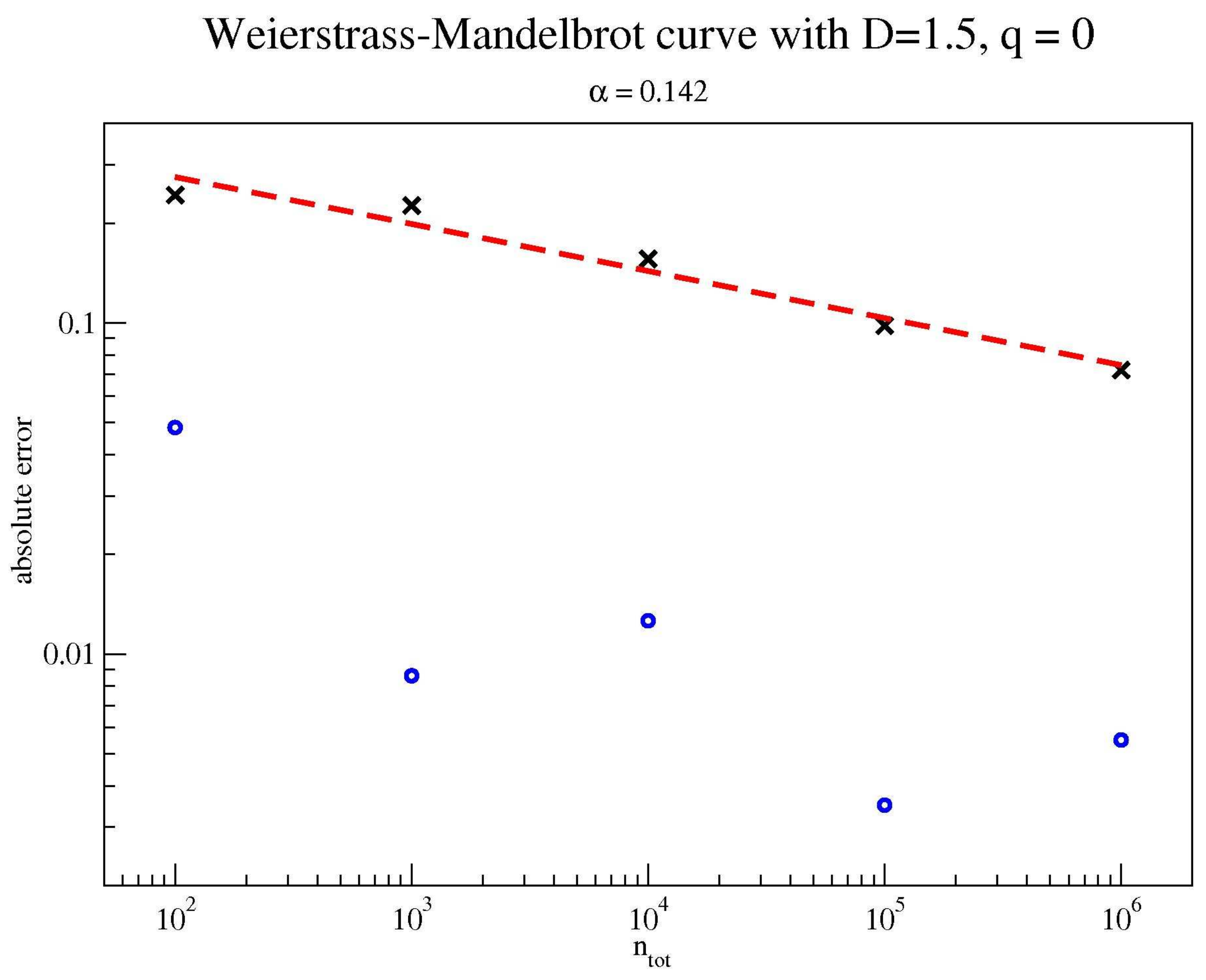
WEIERSTRASS-MANDELBROT CURVE – D=1.5



The parameter n_{tot} (set size) is taken form 10^2 to 10^6 . The mathematical dimension is

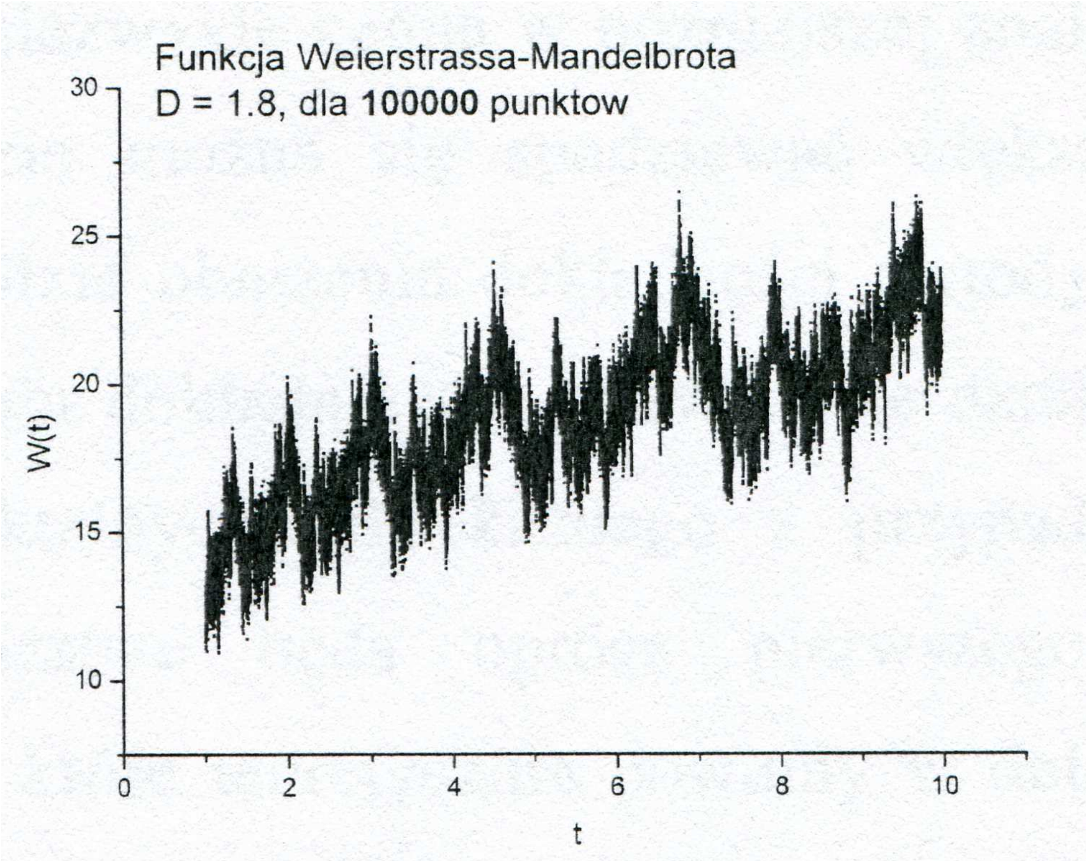
$$d = 1.5$$

COMPUTATIONS – 2D fractals



COMPUTATIONS – 2D fractals

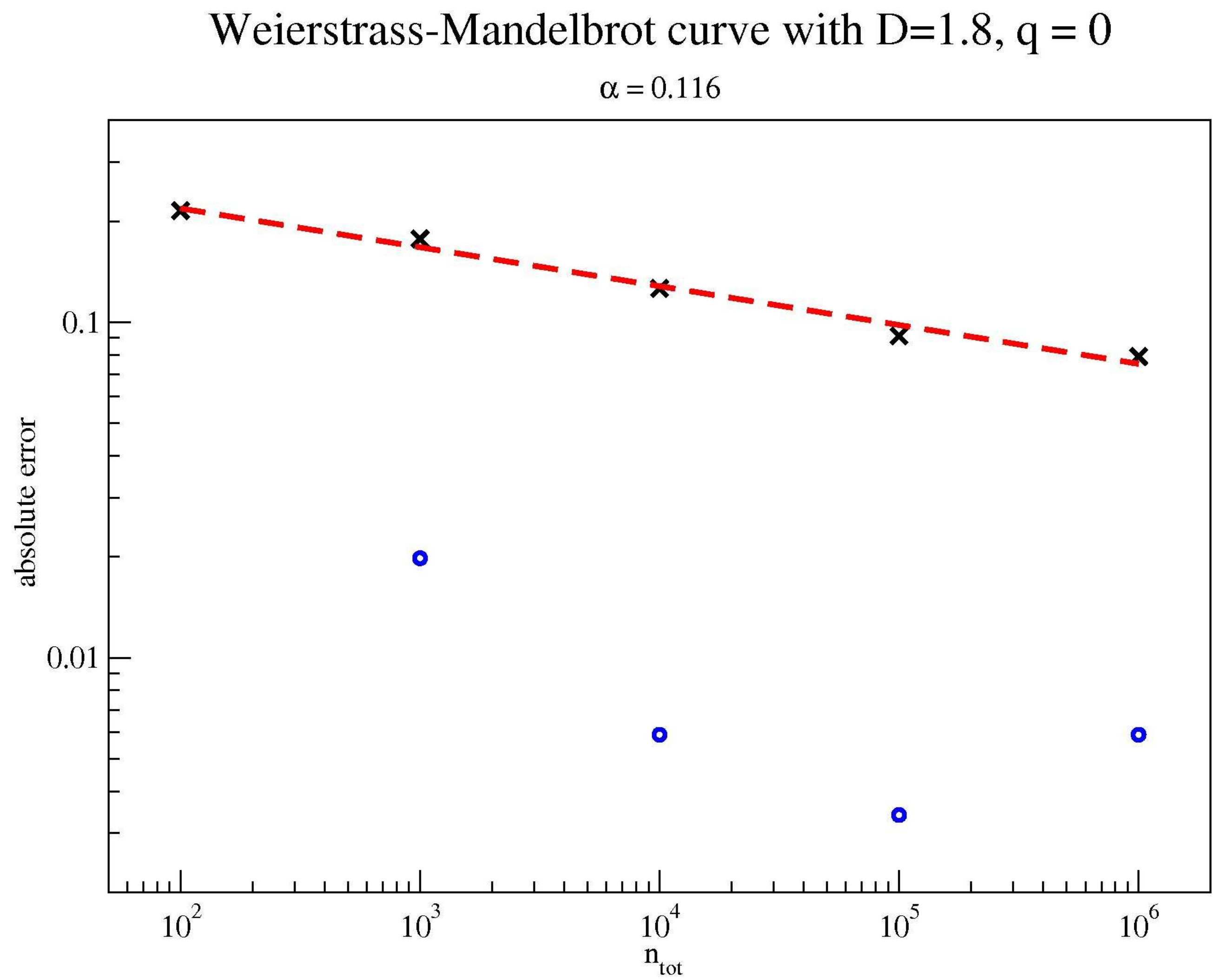
WEIERSTRASS-MANDELBROT CURVE – D=1.8



The parameter n_{tot} (set size) is taken form 10^2 to 10^6 . The mathematical dimension is

d = 1.8

COMPUTATIONS – 2D fractals



CONCLUSIONS

BASIC RESULTS:

- **the computational accuracy (error) is several times larger than the standard error estimated for the linear fit in the log-log plot (σ)**
- **accuracy of computations scales with n_{tot} according to the inverse power law:**

$$\text{error} \sim 1 / n_{tot}^{\alpha}$$

- **for 1D fractals the scaling exponent $\alpha \approx 1/2$**
- **for 2D fractals the scaling exponent $\alpha \approx 1/4$ (Koch curve & Sierpiński carpet)**
- **for W-M functions $\alpha \approx 1/8 \div 1/6$ (W-M curves)**

CONCLUSIONS

ABSOLUTE ACCURACY:

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

CONCLUSIONS

SUMMARY:

- standard error of the linear fit considerably overestimates accuracy of the algorithm
- for typical fractals errors in the 2D case are square roots of the errors for 1D fractals – as one can expect (n_{tot}^2 points necessary to cover the square)
- for W-M functions convergence of error is slower – fractal structure is present in one dimension only (along y-axis)
- for realistic, non-ideal fractals (with noise) one can expect greater errors
- differences between adjacent sets of the linear fit are better error estiamte than σ

THANKS FOR YOUR ATTENTION !