March 1997, IFTUWr 910/97

Generalized Brun's constants

Marek Wolf

Institute of Theoretical Physics, University of Wrocław
Pl.Maxa Borna 9, PL-50-204 Wrocław, Poland
e-mail: mwolf@ift.uni.wroc.pl

Abstract

It is argued, that the sums of reciprocals of all consecutive primes separated by gaps of the length d for $d \geq 6$ are equal to $\frac{4c_2}{d}\prod_{p|d,p>2}\frac{p-1}{p-2}$, where $c_2 = \prod_{p>2}\left(1-\frac{1}{(p-1)^2}\right)$. Besides some heuristic arguments leading to this formula, there is also comparison with the results of computer investigation up to $2^{42} \approx 4.4 \times 10^{12}$. These "experimental" data provides supports for the conjectured formulae. It is also shown, how the guessed formula reproduces the well known fact, that the sum of reciprocals of all primes $p < x$ grows like $\ln(\ln(x))$.

1. In 1919 Brun [1] has shown that the sum of the reciprocals of all twin primes is finite:

$$\mathcal{B}_2 = \left(\frac{1}{3}+\frac{1}{5}\right)+\left(\frac{1}{5}+\frac{1}{7}\right)+\left(\frac{1}{11}+\frac{1}{13}\right)+\ldots<\infty. \tag{1}$$

Sometimes 5 is included only once, but here I will adopt the above convention. The analytical formula for $\mathcal{B}_2$ is unknown[1] and the sum (1) is called the Brun constant [2]. The numerical estimations give [3] $\mathcal{B}_2 = 1.9021605\ldots$. In this paper I am going to generalize the above $\mathcal{B}_2$ to the sums of reciprocals of all consecutive primes separated by gap d and to propose a compact expression giving the values of these sums for $d \geq 6$.

2. Let $\mathcal{T}_d$ denote the set of consecutive primes separated by distance d:

$$\mathcal{T}_d = \{p_{n+1}, p_n \mid p_{n+1} - p_n = d\} \tag{2}$$

where $d = 2k,\ k = 1, 2, \ldots$. I shall call these primes "twins of genus d". For $d = 2$ the set $\mathcal{T}_2$ contains twin primes, while for $d = 4$ the set $\mathcal{T}_4$ contains cousin primes.

The generalized Brun's constants are defined by the equation:

$$\mathcal{B}_d = \sum_{p\in\mathcal{T}_d} 1/p. \tag{3}$$

For $d = 2$ we have the usual Brun's constant $\mathcal{B}_2$ for twins.

If $\pi_d(x)$ denotes the number of pairs (p, p') less than x such, that p and $p' = p+d$ are both prime, then it can be proved (see [4], chapter IV) that:

$$\pi_d(x) = \mathcal{O}(x(\ln(\ln(x)))^2 \ln^{-2}(x)). \tag{4}$$

Let us remark, that the r.h.s does not involve d explicitly, it is hidden in implicit constant. Because the series (3) contain only *consecutive* primes separated by distance d, they are majorized by the series over all prime pairs $p, p+d$, not necessary successive. From these facts it follows, that the sum in (3) is convergent for every d, so generalized Brun's constants are finite.

3. Let us define partial (finite) sums:

$$\mathcal{B}_d(x) = \sum_{p\in\mathcal{T}_d, p<x} 1/p. \tag{5}$$

The dependence of $\mathcal{B}_2(x)$ on x is usually (see [2], [5]), obtained by appealing to the Hardy–Littlewood conjecture [6] on the number of primes separated by the distance d. It states that there are infinitely many pairs of primes

$$p, p' = p + d, \tag{6}$$

[1] It can be remarked, that E. Karst has conjectured (see [2]), that $4/3 + \mathcal{B}_2 = \pi$, but today it seems to be ruled out by the modern estimations of the Brun's constant.

for every even d and

$$\pi_d(x) \sim 2c_2 \frac{x}{\ln^2(x)} \prod_{p|d} \frac{p-1}{p-2}. \tag{7}$$

Here the constant c_2 (sometimes called "twin–prime " constant, see [7]) is defined in the following way:

$$c_2 \equiv \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) = 0.66016\ldots \tag{8}$$

The pairs of primes separated by $d = 2$ and $d = 4$ are special as they always have to be consecutive primes (with the exception of the pair (3,7) containing 5 in the middle) and for $d \geq 6$ the function $\pi_d(x)$ counts *all* pairs $p, p' = p+d$, not *necessarily successive*. For $d = 2$ equation (7) gives that the probability to find a pair of twins in the vicinity of x is $2c_2/\ln^2(x)$, so the expected value of the finite approximation to the Brun constant can be estimated as follows:

$$\mathcal{B}_2(x) = \mathcal{B}_2(\infty) - \sum_{p \in \mathcal{T}_2, p>x} \frac{1}{p} \approx \mathcal{B}_2 - 4c_2 \int_x^\infty \frac{du}{u \ln^2(u)} = \mathcal{B}_2 - \frac{4c_2}{\ln(x)}. \tag{9}$$

It means that the plot of finite approximations $\mathcal{B}_2(x)$ to the original Brun constant is a linear function of $1/\ln(x)$. The same reasoning applies *mutatis mutandis* to the gap $d = 4$. In the Fig.1 I have plotted the values of $\mathcal{B}_2(x)$ and $\mathcal{B}_4(x)$ found during the computer scan up to $x = 2^{22}$, $x = 2^{23}, \ldots, x = 2^{42}$ versus $1/\ln(x)$ and indeed both plots are parallel with the slope -2.64073 for twins and -2.64069 for cousins, while $4c_2 \approx 2.64064$; intercepts are 1.90216 and 1.19705 respectively for $\mathcal{B}_2$ and $\mathcal{B}_4$.

To repeat the above reasoning for $d = 2,\ 4$ for larger d an analog of the Hardy–Littlewood conjecture for the pairs of *consecutive* primes separated by distance d is needed. In [8] it was conjectured, that the number $h_d(x)$ of *successive* primes $(p_{n+1},\ p_n)$ smaller than x and separated by d $(= p_{n+1} - p_n)$ is diminished by the factor $\exp(-d/\ln(x))$ in comparison with *all* pairs of primes (p, p') apart in the distance d $(= p' - p)$:

$$h_d(x) = \pi_d(x) e^{-d/\ln(x)} = 2c_2 \frac{x}{\ln^2(x)} \prod_{p|d} \frac{p-1}{p-2} e^{-d/\ln(x)}. \tag{10}$$

Nothing is known about the error term in the above equation. For $d \geq 6$ equation (10) gives:

$$\mathcal{B}_d(x) = \mathcal{B}_d(\infty) - \sum_{p \in \mathcal{T}_d, p>x} \frac{1}{p} \approx \mathcal{B}_d - 4c_2 \prod_{p|d} \frac{p-1}{p-2} \int_x^\infty \frac{e^{-d/\ln(u)}}{u \ln^2(u)} du. \tag{11}$$

The integral can be calculated explicitly:

$$\mathcal{B}_d(x) = \mathcal{B}_d(\infty) + \frac{4c_2}{d} \prod_{p|d} \frac{p-1}{p-2} \left(e^{-d/\ln(x)} - 1\right). \tag{12}$$

From this it follows, that the partial sums $\mathcal{B}_d(x)$ for $d \geq 6$ should depend linearly on $\exp(-d/\ln(x))$ instead of $1/\ln(x)$ for $\mathcal{B}_2(x)$ and $\mathcal{B}_4(x)$.

4. I have made the computer search for prime numbers up to $N = 2^{42} \approx 4.4 \times 10^{12}$ and calculated the finite approximations to the generalized Brun's constants $\mathcal{B}_d(x)$. The computer program was implemented on the DEC Alpha 300X/175MHz workstation in DECFortran which allows some assembly level instructions and it took 101 days of CPU time to reach 2^{42}. The partial generalized Brun's constants $\mathcal{B}_d(x)$ were stored at $x = 2^{22}, x = 2^{23}, \ldots, 2^{41}, 2^{42}$ and a sample of the obtained results is presented in Table I.

TABLE **I**

A part of the $\mathcal{B}_d(x)$ data obtained by the computer for $x = 2^{42}$

d	$\mathcal{B}_d(x)$	d	$\mathcal{B}_d(x)$	d	$\mathcal{B}_d(x)$	d	$\mathcal{B}_d(x)$
2	1.81145E+00	62	3.98920E–03	102	1.09505E–03	142	7.92544E–05
4	1.10634E+00	64	4.04969E–03	104	5.19325E–04	144	1.40257E–04
6	9.61169E–01	66	7.52082E–03	106	4.53440E–04	146	6.22615E–05
8	2.57759E–01	68	3.10471E–03	108	8.09941E–04	148	6.82308E–05
10	3.12936E–01	70	4.51534E–03	110	5.07854E–04	150	1.43922E–04
12	2.98231E–01	72	4.64803E–03	112	4.04087E–04	152	4.88808E–05
14	1.74150E–01	74	2.29100E–03	114	6.26933E–04	154	6.04450E–05
16	9.81150E–02	76	2.07442E–03	116	2.72974E–04	156	8.65953E–05
18	1.66470E–01	78	4.01778E–03	118	2.68305E–04	158	3.68855E–05
20	8.22241E–02	80	2.08683E–03	120	5.94943E–04	160	4.46518E–05
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

In Fig.2 I have plotted actual data obtained by a computer for a sample of gaps d to test equation (12). Indeed, the points follow straight lines with different slopes, supposed to be equal to $\frac{4c_2}{d}\prod_{p|d}\frac{p-1}{p-2}$. On the other hand, Fig. 3 presents somehow inverse of Fig.2: here a plot of $\mathcal{B}_d(x)$ is given for all d, but only for 3 values of $x = 2^{22}, x = 2^{32}$ and $x = 2^{42}$ which span over the 6 orders range. In Fig.4 there is an enlargement of Fig.3 showing that the overall dependence of $\mathcal{B}_d(x)$ is described by $e^{-d/\ln(x)}/d$. The oscillations in the "experimental" values of $\mathcal{B}_d(x)$ are caused by overimposing the product

$$\mathcal{J}(d) = \prod_{p|d}\frac{p-1}{p-2} \tag{13}$$

onto the overall factor $e^{-d/\ln(x)}/d$.

The equation (12) can be written in the form:

$$\mathcal{B}_d(x) = \alpha(d) + \beta(d)e^{-d/\ln(x)} \tag{14}$$

The fitting of the parameters $\alpha(d)$ and $\beta(d)$ by means of the least square method shows, that the curves $\mathcal{B}_d(x)$ tend to cross the point $(0,0)$ in the variables $(\exp(-d/\ln(x)),\ \mathcal{B}_d(x))$ for increasing d, so in other words $\alpha(d) = 0$, what is obvious, because

$\mathcal{B}_d(x=1)=0$ and $\exp(-d/\ln(x))=0$ for $x=1$. It means that there is in (12), in fact, the cancellation of two terms not involving the variable x and:

$$\mathcal{B}_d(x) = \frac{4c_2}{d}\prod_{p|d}\frac{p-1}{p-2}e^{-d/\ln(x)} + error\ term(d,x). \tag{15}$$

The computer results suggest that the error term in the above conjecture decreases rather fast with d and x. From (15) the main conjecture of this paper follows:

$$\mathcal{B}_d(\infty) \equiv \mathcal{B}_d = \frac{4c_2}{d}\prod_{p|d}\frac{p-1}{p-2} \quad \text{for } d \geq 6. \tag{16}$$

However, the above conclusion is based on the assumption, that the dependence of $\mathcal{B}_d(x)$ on x of the form (14) is valid down to such small values as $x \approx 1$. But in the next paragraph we will give arguments that there is no room for additional dependence on x.

The comparison of the formula (16) with the values extrapolated from the partial approximations $\mathcal{B}_d(2^{42})$ using the equation (12) is shown in the Fig.5 and Fig.6 for $d \geq 6$ — predicted by (16) values for $d = 2$ and $d = 4$ are skipped. However, remarkably good approximations to $\mathcal{B}_2$ and $\mathcal{B}_4$ are provided by $4/d$ (solid line). The Fig.6 shows explicitly, that for small d predicted by (16) values are smaller than the values of $\mathcal{B}_d$ extrapolated from $\mathcal{B}_d(2^{42})$ via the equation (12). For $d = 6$ even the partial sum $\mathcal{B}_6(2^{42}) = 0.96116$ is larger than $\mathcal{B}_6 = 0.88021\ldots$ given by (16). It can be linked to the fact that the gap $d = 6$ is a so called "champion": up to roughly 10^{36} it is the most often occuring gap between consecutive primes, see [8]. I guess that all these differences between the formula (16) and actual or extrapolated values can be incorporated into the unknown error term:

$$\mathcal{B}_d(\infty) \equiv \mathcal{B}_d = \frac{4c_2}{d}\prod_{p|d}\frac{p-1}{p-2} \ + \ error\ term(d), \quad \text{for } d \geq 6. \tag{17}$$

The data shown on the Fig.6 suggest that the error term should decrease with d.

5. It is well known, that the sum of reciprocals of all primes smaller than x is given by [10] (see also [11], Theorem 427 and 428):

$$\sum_{p<x}\frac{1}{p} = \ln(\ln(x)) + M + o(1); \tag{18}$$

here $M = \gamma - \sum_{k=2}^{\infty} 1/k\Sigma(k) = 0.2614972\ldots$, where $\Sigma(k) = \sum_p 1/p^k$and $\gamma = 0.5772157\ldots$ is the Euler constant. On the other hand, it can be expressed by finite approximations to the generalized Brun's constants:

$$\sum_{p<x}\frac{1}{p} = \frac{1}{2} + \frac{1}{6} + \frac{1}{2}\sum_d \mathcal{B}_d(x) = M' + \frac{2}{3} + 2c_2\sum_{d=2}^{G(x)}\frac{1}{d}\prod_{p|d}\frac{p-1}{p-2}e^{-d/\ln(x)} \tag{19}$$

I have introduced here the constant M' which accounts the sum of the unknown errors terms in (15) as well as incorporates the fact, that the dependence of $B_2(x)$

and $B_4(x)$ on x is not described by the formula (15). The sum in (19) runs over even d and extends up to the greatest gap $G(x)$ between two consecutive primes smaller than x. It was conjectured by Cramer [12], see also [13], [14], [15],[8], that:

$$G(x) \sim \ln^2(x). \tag{20}$$

The product $\mathcal{J}(d)$ behaves very erratically, but it takes values of the order 1. In fact, its mean value is equal $1/c_2$ [9]:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \prod_{p|2k,p>2} \frac{p-1}{p-2} = \frac{1}{\prod_{p>2}(1-\frac{1}{(p-1)^2})} \quad (= \ 1.51478\ldots). \tag{21}$$

Replacing the product $\mathcal{J}(d)$ under the sum (19) by its mean value $1/c_2$ we end up with the series:

$$\sum_{p<x} \frac{1}{p} = M' + \frac{2}{3} + 2\sum_{d=2}^{G(x)} \frac{1}{d} e^{-d/\ln(x)} = M' + \frac{2}{3} + \sum_{k=1}^{\frac{1}{2}G(x)} \frac{1}{k} q^k, \quad q = e^{-2/\ln(x)}. \tag{22}$$

Expanding $\ln(1-q)$, where $0 < q < 1$ into the series we obtain

$$\sum_{k=1}^{n} \frac{1}{k} q^k = -\ln(1-q) + \int_0^q \frac{u^n}{u-1} du \tag{23}$$

For large x the term with logarithm goes into:

$$\ln(1-e^{-2/\ln(x)}) = -\ln(\ln(x)) + \ln(2) + \mathcal{O}(1/\ln(x)) \tag{24}$$

Now, by the mean value theorem, we calculate the integral:

$$\mathcal{I} = \int_0^q \frac{u^n}{u-1} du = \frac{1}{(\theta q - 1)} \frac{q^{n+1}}{(n+1)}, \quad 0 < \theta < 1. \tag{25}$$

But $q = \exp(-2/\ln(x)) < 1$ and:

$$\mid \frac{1}{\theta q - 1} \mid < \frac{1}{1-q} = \frac{e^{2/\ln(x)}}{e^{2/\ln(x)} - 1} < \frac{\ln(x)}{2} e^{2/\ln(x)} = \mathcal{O}(\ln(x)) \tag{26}$$

For $x \gg 1$ we have on the virtue of the Cramer conjecture that in our case $n \sim \frac{1}{2}\ln^2(x)$, thus:

$$\mid \mathcal{I} \mid = \mathcal{O}(1/x\ln(x)) \tag{27}$$

Finally we obtain that from (12) and (19) there follows:

$$\sum_{p<x} \frac{1}{p} = \ln(\ln(x)) + M' + \frac{2}{3} - \ln(2) + \mathcal{O}(1/\ln(x)) \tag{28}$$

Because $2/3$ is practically equal to $\ln(2)$ to require consistency with the Merten's theorem we have to postulate that $M' \approx M$. The comparison of the Mertens estimation for $\sum_{p<x} 1/p$ with data obtained by a computer is shown in Fig.7.

Because the dependence of $\mathcal{B}_d(x)$ on x of the form $\exp(-d/\ln(x))$ reproduces exactly the leading term in $\sum_{p<x} 1/p$ it suggests that if there is additional dependence on x in (15), it has to decrease with x very fast. Another possibility of such a modification of (15), which may not change the leading term in (28), namely the changing sign with d factor, seems to be not excluded by data in Fig.6, where for small d the oscillations can be seen, which however for larger d disappear and the curve tend *monotonically* from below to the assymptotic value 1.

Acknowledgment: The author receives support from the KBN research grant No 2 P302 057 07

References

[1] V.Brun, "La serie 1/5 + 1/7 +... est convergente ou finie", *Bull. Sci.Math.* **43** (1919), pp.124-128

[2] D.Shanks and J.W. Wrench, "Brun's Constant", Math. Comp. **28** (1974), p.293-299

[3] T. R. Nicely, "Enumeration to 10^{14} of the Twin Primes and Brun's constant", *Virginia Journal of Science* **46** (1995), pp.195-204

[4] W. Narkiewicz, *Number Theory*, (World Scientific, Singapore, 1993)

[5] R.P.Brent, "Irregularities in the Distribution of Primes and Twin Primes", *Math. Comp.* **29** (1975), pp.43-56

[6] G.H.Hardy and J.E. Littlewood, "Some problems of 'Partitio Numerorum' III: On the expression of a number as a sum of primes", *Acta Mathematica* **44** (1922), p.1-70

[7] J.W.Wrench, "Evaluation of Artin's constant and the Twin Prime constant", *Math.Comp.***15** (1961), p.396-398

[8] M.Wolf, "Some conjectures on the gaps between consecutive primes", preprint IFTUWr 894/95, submitted to *J.Number Theory*; summary of the results appeared as "Unexpected Regularities in the Distribution of Prime Numbers", in Proc. of the 8th Joint EPS-APS Int.Conf. Physics Computing'96, edts. P.Borcherds et al, (Krak/ow, 1996), s. 361-367; available at www.ift.uni.wroc.pl/~mwolf.

[9] E. Bombieri and H. Davenport, "Small differences between prime numbers", *Proc. Royal Soc.*, **A293**, (1966), 1-18

[10] F.Mertens, "Ein Beitrag zur analytischen Zahlentheorie" *J.Reine Ang. Math.* **77** (1874), p.289 and **78** (1874), p. 46 – 62

[11] G.H.Hardy and E.M.Wright, *An Introduction to the Theory of Numbers*, (Clarendon Press, Oxford, 1960)

[12] H.Cramer, "On the order of magnitude of difference between consecutive prime numbers", *Acta Arith.* **2** (1937), p.23-46

[13] D.Shanks, "On Maximal Gaps between Succesive Primes", *Math.Comp.* **18** (1964), p.464

[14] H.Maier, C.Pomerance, "Unusually large gaps between primes", *Trans. Amer. Math.Soc.* **322** (1990), p.201-237

[15] A.Granville,"Unexpected irregularities in the Distribution of Prime Numbers", in Proceedings of the International Congress of Mathematicians (Zurich, Switzerland, 1994), Vol. I (1995), 388–399.

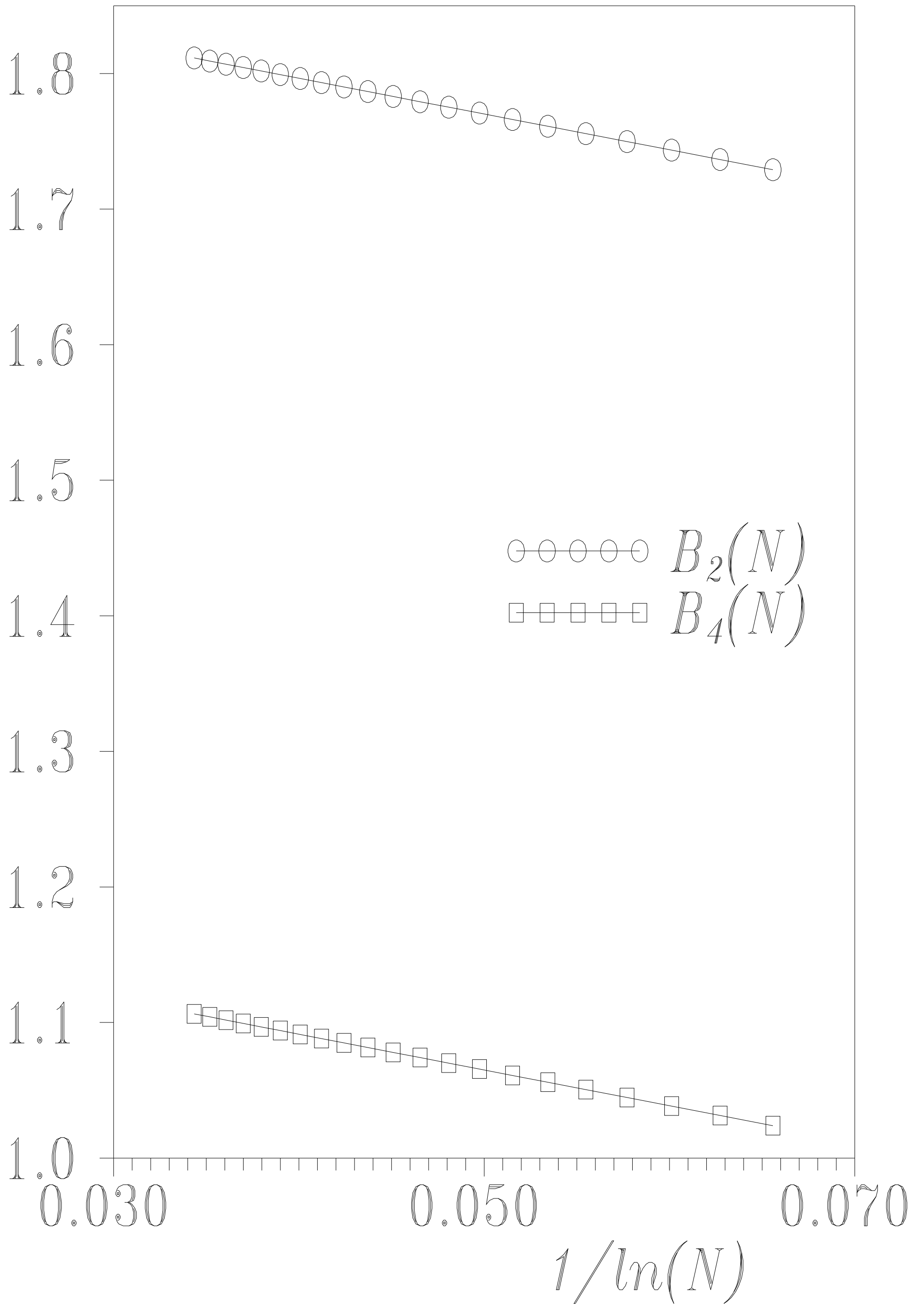

Fig.1 The plot showing the dependence of the finite approximations to the Brun constant for twins (circles) and cousins (squares). On the horizontal axis values of $1/\ln(x)$ are used.

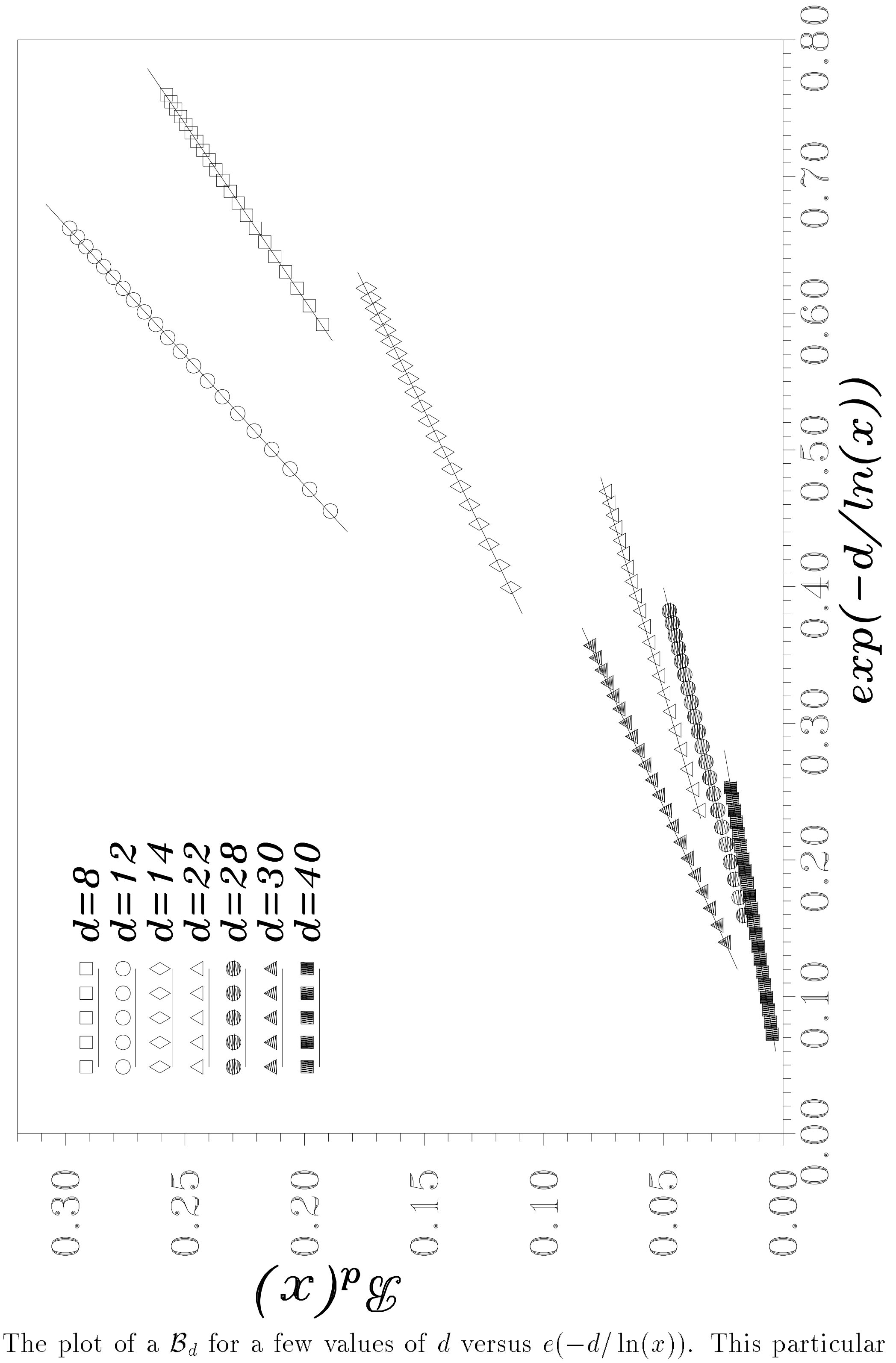

Fig. 2 The plot of a $\mathcal{B}_d$ for a few values of d versus $e(-d/\ln(x))$. This particular choice of d was chosen to present various factorizations of d and simultaneously to avoid overlapping of graphs. Let us remark that the plots for $d = 14$ and $d = 30$ are almost parallel, what agrees with the value of $4c_2/d\prod_{p|d}\frac{p-1}{p-2}$, which is equal to 0.2263 and 0.2347 for $d = 14$ and $d = 30$ respectively.

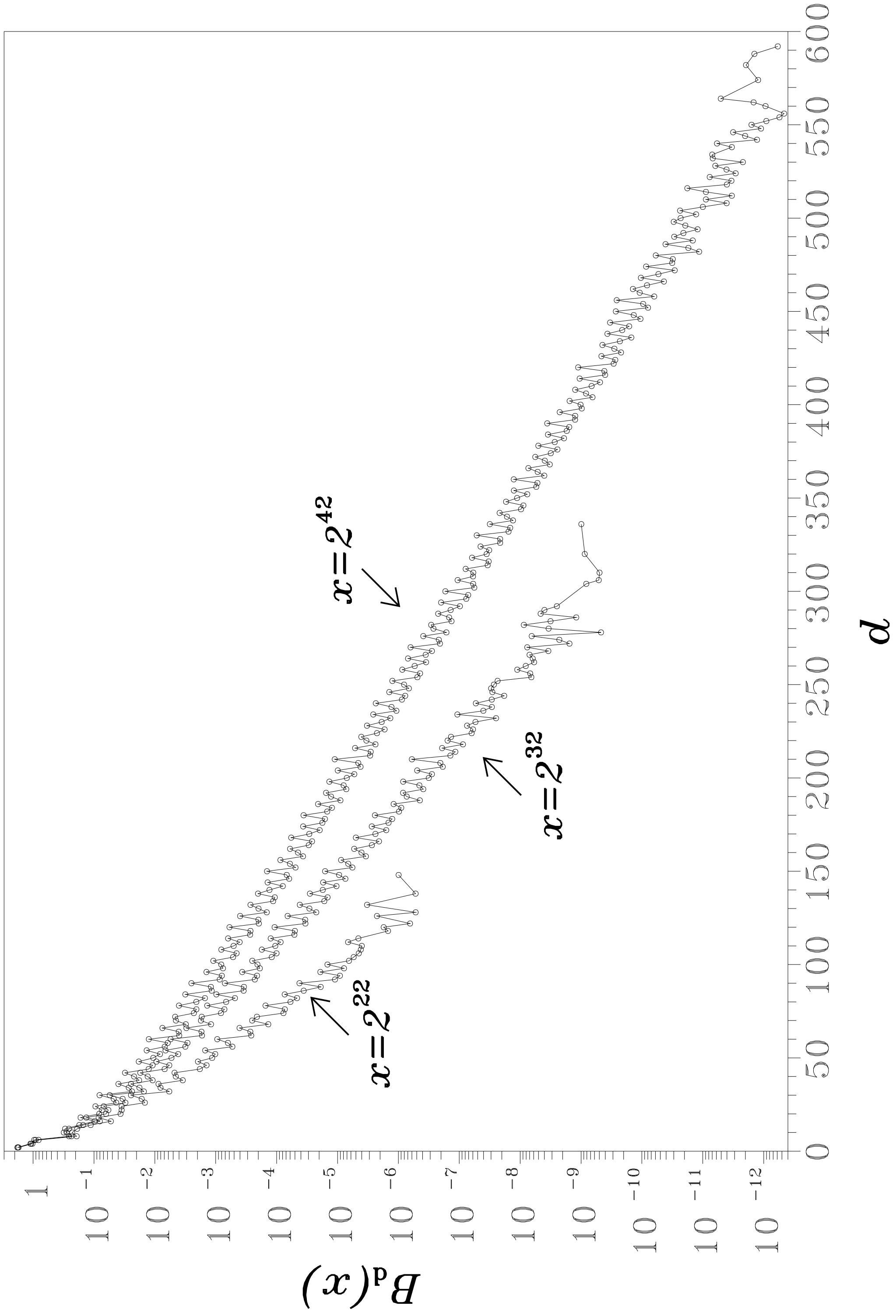

Fig.3 The plot showing the dependence of the finite approximations to the Brun constant $\mathcal{B}_d(x)$ for $x = 2^{22}, 2^{32}$ and 2^{42}. The characteristic pattern of the circles is caused by the product $\mathcal{J}(d)$ and reflects occuring frequency of the gaps of the length d, see [8]. Because the product $\mathcal{J}(d)$ displays the oscillations of the period 6, every six-th circle is at the top of "jigsaw" like plot.

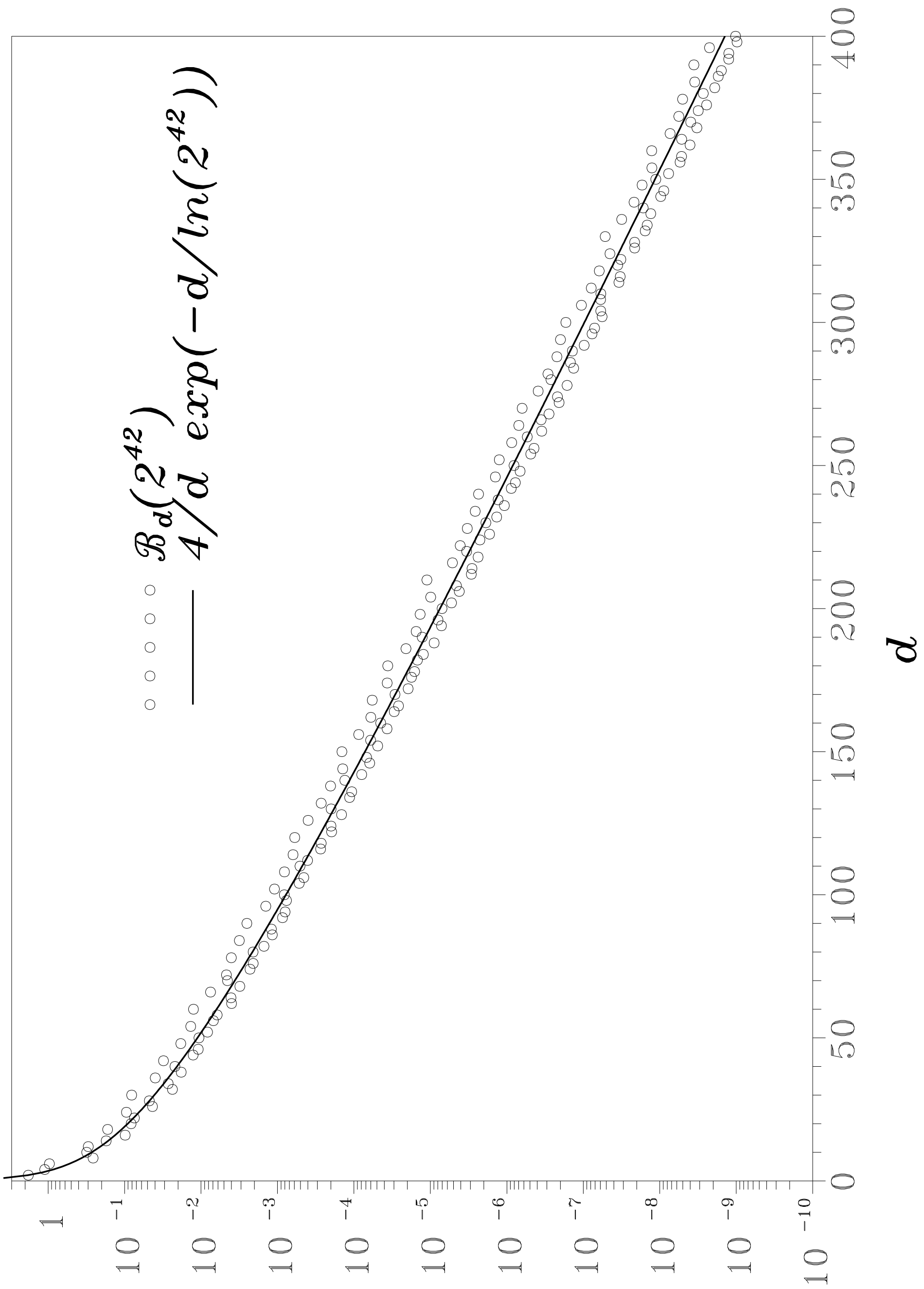

Fig.4 The actual points for $\mathcal{B}_d(x = 2^{42})$ plotted for $d = 2, \ldots 400$ and the plot of the function $4/d \exp(-d/\ln(x))$ (bold line) describing the overall shape displayed by partial values of the generalized Brun's constants. In fact the slope in the exponent was not taken to be $\ln(2^{42}) = 0.03435\ldots$, but the used value $0.03945\ldots$ was determined by trials. I have chacked that the "experimental" slope converges to $1/\ln(x)$ for increasing x.

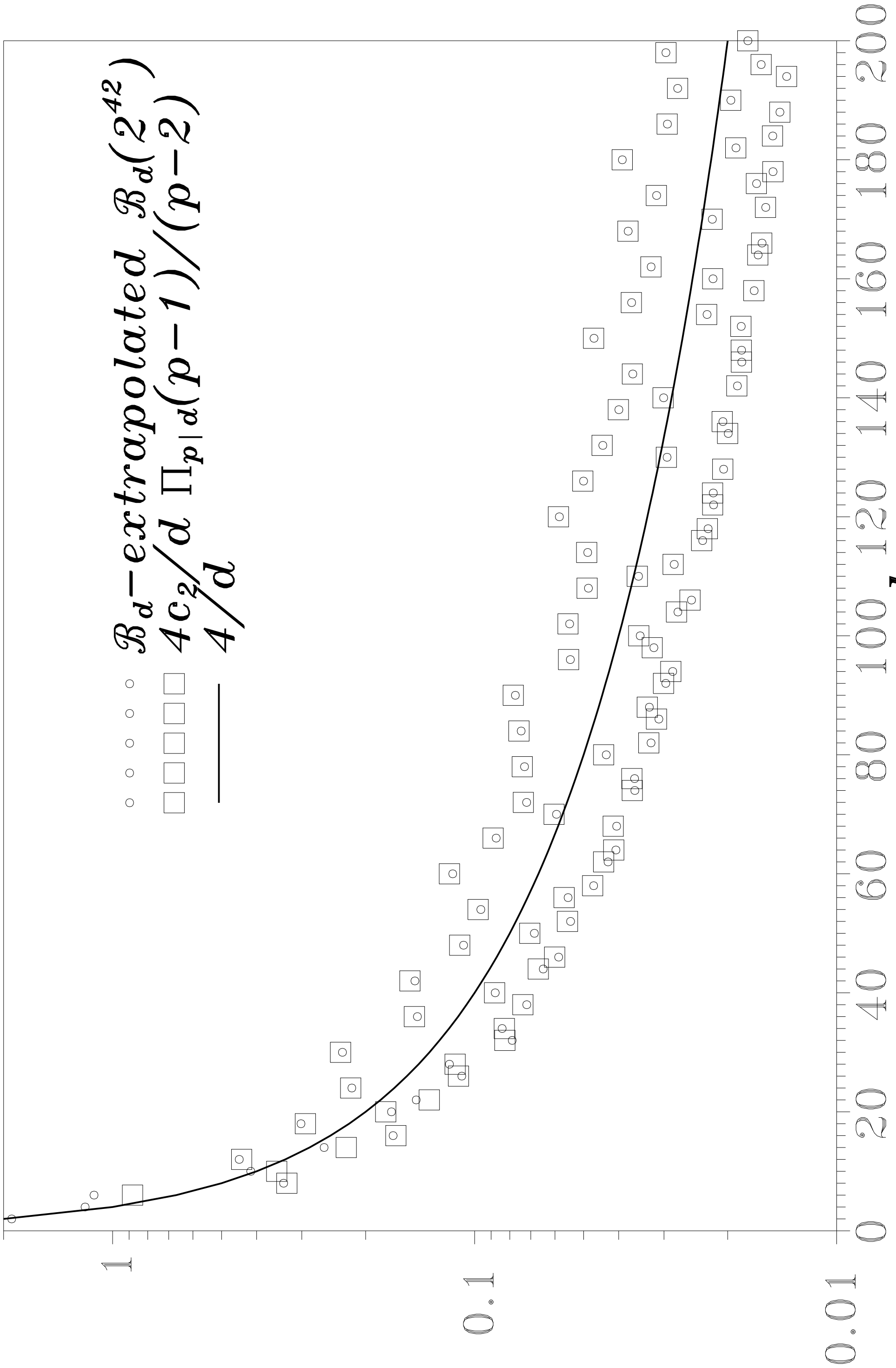

Fig.5 Plot of the values of $\mathcal{B}_d$ extrapolated from finite approximations $\mathcal{B}_d(2^{42})$ by means of the equation (12) (circles) and obtained from the conjecture (16) (squares). There is logarithmic scale on the y-axis to allow better "resolution" — on both axis linear or logarithmic many points collapse to the same place. Because the product $\mathcal{J}(d)$ behaves on the average as $1/c_2$, see (21), the rough approximation to $\mathcal{B}_d$ is given simply by $4/d$ and it is plotted by a bold line. The squares representing values obtained from (16) for $d = 2$ and $d = 4$ were skipped. Hovewer, remarkably good approximations to $\mathcal{B}_2$ and $\mathcal{B}_4$ is provided by $4/d$.

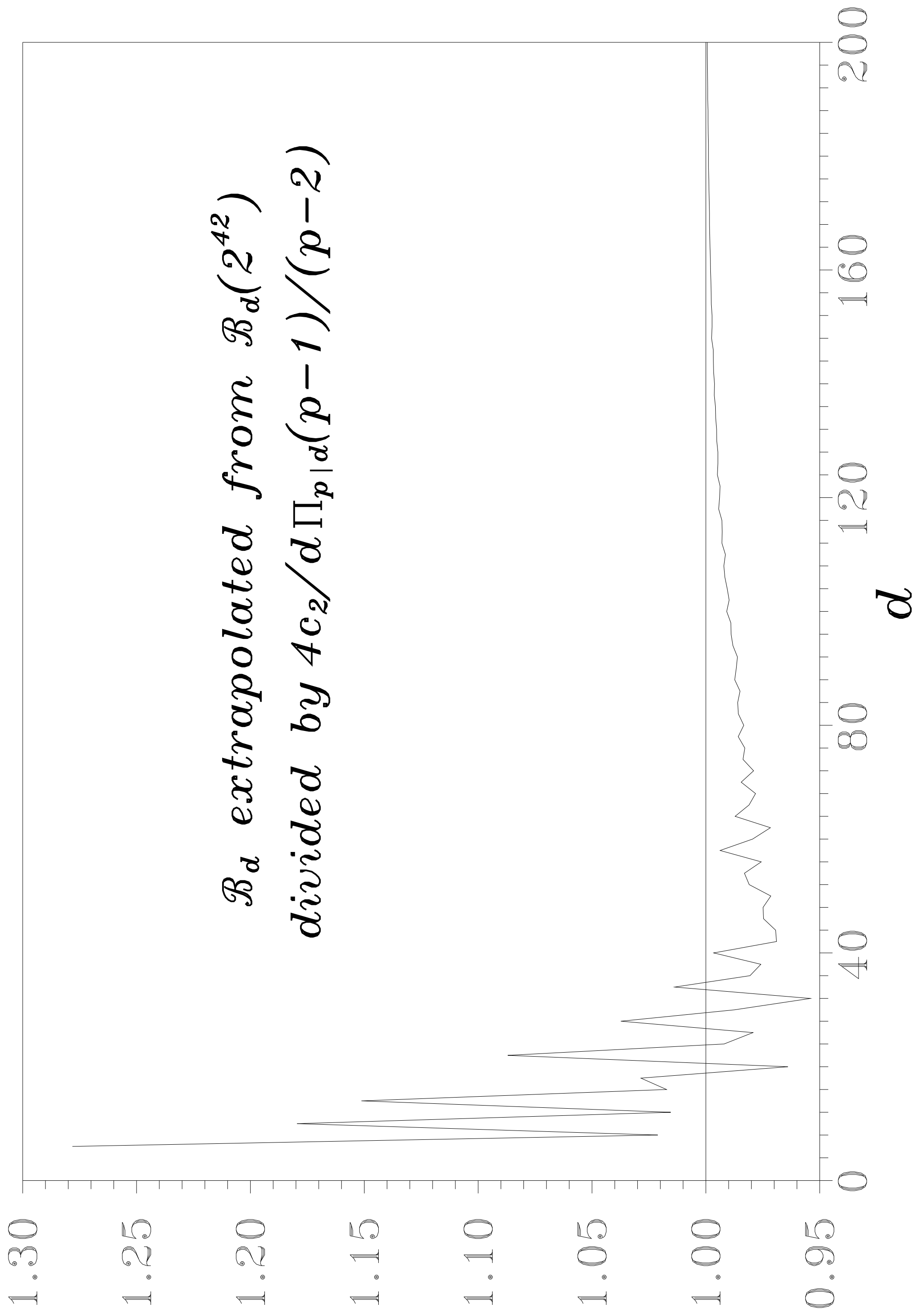

Fig.6 The plot of the ratio of $\mathcal{B}_d$ extrapolated from finite approximations $\mathcal{B}_d(2^{42})$ by means of the equation (12) divided by the values obtained from the conjecture (16).

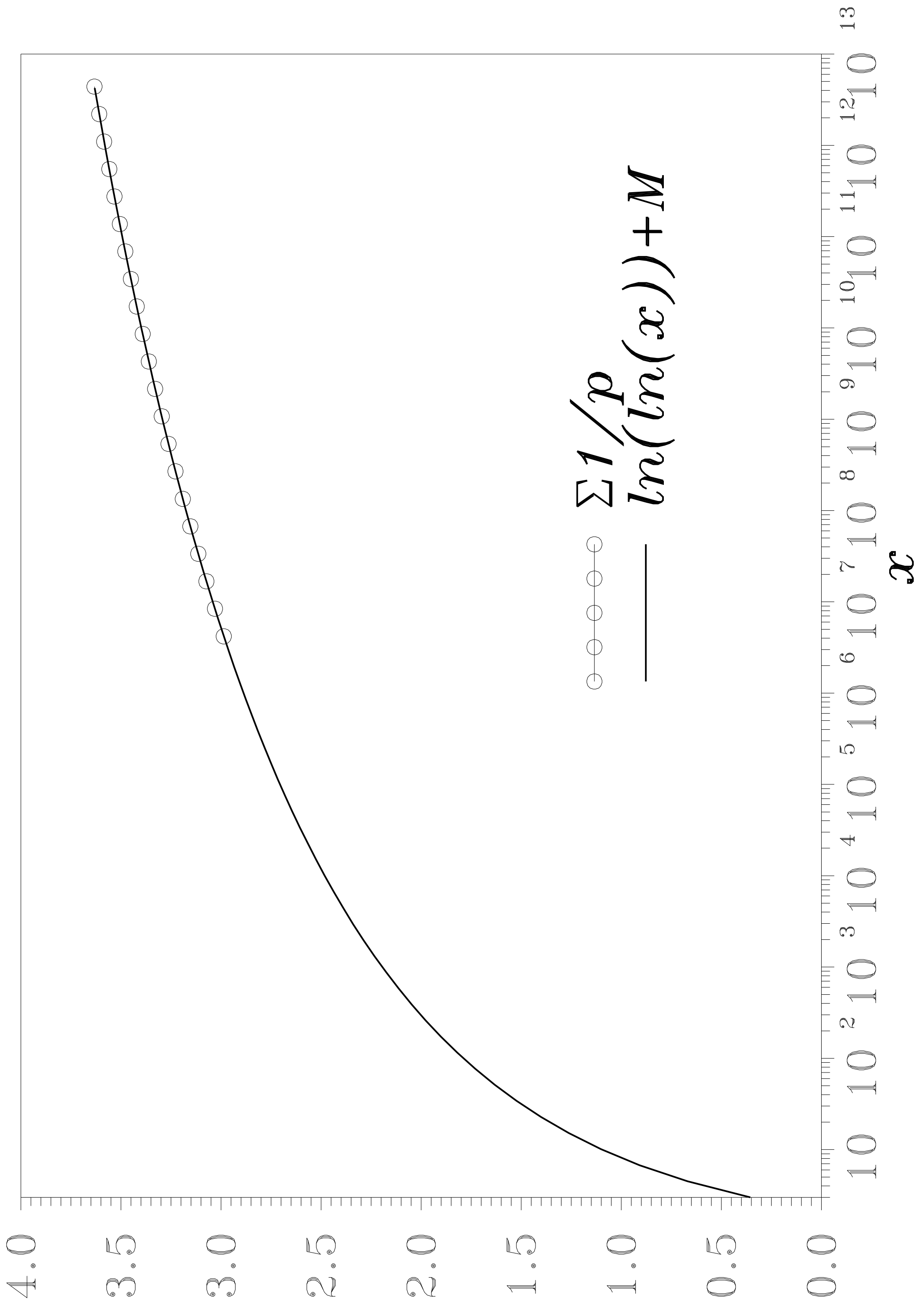

Fig.7 The plot of the actual values of $\sum_{p<x} 1/p$ obtained from the computer data for $\mathcal{B}_d(x)$ (circles) and the Mertens estimation (18) plotted by solid line. The original of this figure has y axis of the length 16 cm, so if the x axis would be plotted in the linear scale, instead of logarithmic, then it should be 400 000 000 km long — a half of the distance between Jupiter and Sun.