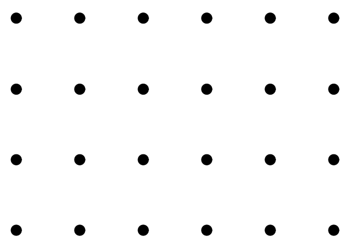
ON LINES AND THEIR INTERSECTION POINTS IN A RECTANGULAR GRID OF POINTS

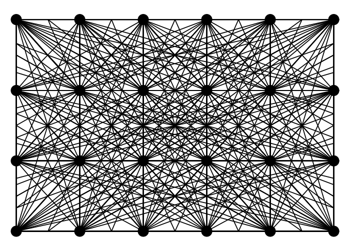
SEPPO MUSTONEN

1. Introduction

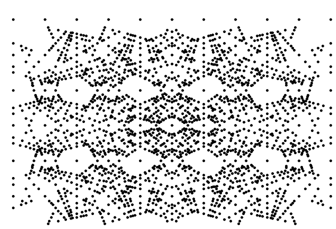
In an $m \times n$ rectangular grid of points (here $m = 4, n = 6$)



lines through at least 2 points of the grid (136 lines)



and points of intersection of these lines inside the grid (1961 points)



will be studied.

Date: 16 April 2009.

In sequel, the number of lines through at least 2 points of the grid is denoted $L(m,n)$ and the number of intersection points of these lines inside (or on the border of) the grid is denoted $S(m,n)$. The number of lines through exactly j points of the grid is denoted $L_j(m,n)$ so that

$$L(m,n) = \sum_{j=2}^{\max(m,n)} L_j(m,n). \tag{1}$$

Furthermore, let $L(n) = L(n,n)$, $L_j(n) = L_j(n,n)$, and $S(n) = S(n,n)$.

I got interested in these lines and points during my earlier study [2]. In particular, I calculated the number of intersection points $S(n)$ for $n = 2, \ldots, 11$:

n	2	3	4	5	6	7	8	9	10	11
$S(n)$	5	37	405	2225	11641	35677	114409	295701	718469	1475709

(Notation *sect*$(n-1)$ was used for $S(n)$ in [2].) These numbers were originally computed by brute force. Thus all possible $n^7(n+1)/2$ points of intersection between lines from (X_1, Y_1) to (X_2, Y_2) and from (X_3, Y_3) to (X_4, Y_4) were listed for $X_1 = 0, \ldots, n-1, X_2 = X_1, \ldots, n-1$, and for remaining 6 coordinates from 0 to $n-1$. Most of the points were then generated multiple times and these multiplicities were then removed after sorting the list of points.

Now, an essentially more effective procedure is adopted by first determining all distinct lines through at least 2 points of the grid and then the required points of intersection by a balanced tree search and insertion algorithm described in [1].

Currently, values of $S(n)$ have been computed for $n = 2, \ldots, 30$ and they together with corresponding $L(n)$ values are given in Table 1.

Integer sequences $L(m,n)$ and $S(m,n)$ will now be examined.

2. Lines through grid points

The $L(n)$ numbers can be found in [3] as sequence A018808. There also a formula

$$L(n) = \frac{1}{2}[f(n,1) - f(n,2)] \tag{2}$$

where

$$f(n,k) = \sum_{\substack{-n<x<n \\ -n<y<n \\ (x,y)=k}} (n-|x|)(n-|y|) \tag{3}$$

is given without any reference about its origin. It will be seen that the formula is correct, in principle, but a better formulation for (3) is

$$f(n,k) = \sum_{\substack{-n<kx<n \\ -n<ky<n \\ (x,y)=1}} (n-|kx|)(n-|ky|) \tag{4}$$

In [3] also some related integer sequences like A018809 (Number of lines through exactly 2 points of an $n \times n$ grid of points) are presented correctly but with an

n	$L(n)$	$S(n)$
2	6	5
3	20	37
4	62	405
5	140	2225
6	306	11641
7	536	35677
8	938	114409
9	1492	295701
10	2306	718469
11	3296	1475709
12	4722	3093025
13	6460	5771929
14	8830	10895273
15	11568	18785841
16	14946	31414269
17	18900	50274501
18	23926	81288641
19	29544	124066161
20	36510	190860537
21	44388	282399889
22	53586	411505049
23	63648	580614301
24	75674	824814797
25	88948	1138709849
26	104374	1570665877
27	121032	2115178249
28	139966	2833746309
29	160636	3732420861
30	184466	4937226173

TABLE 1. Number of lines $L(n)$ and points of intersection $S(n)$

invalid formula

$$L_2(n) = \frac{1}{2}[f(n,4) - 2f(n,3) + f(n,2)].$$

This formula should read

$$L_2(n) = \frac{1}{2}[f(n,3) - 2f(n,2) + f(n,1)]. \tag{5}$$

A similar flaw[1] appears at least in formulas of sequences A018810, A018811, A018812 for $L_j(n), j = 3, 4, 5$, and A119437.

Thus the correct formula for $j = 2, 3, \ldots, n$ is

$$L_j(n) = \frac{1}{2}[f(n,j+1) - 2f(n,j) + f(n,j-1)] \tag{6}$$

and it, for example, satisfies (1).

3. Lines in a rectangular grid

It will be shown first that

[1]Formulas were corrected on April 25, 2009 in [3].

$$L(m,n) = \frac{1}{2}[f(m,n,1) - f(m,n,2)] \tag{7}$$

where

$$f(m,n,k) = \sum_{\substack{-n<kx<n \\ -m<ky<m \\ (x,y)=1}} (n - |kx|)(m - |ky|). \tag{8}$$

Let (u_1, v_1) and (u_2, v_2) be two points in an $m \times n$ grid. Then the line through these two points has the equation

$$(u_2 - u_1)v - (v_2 - v_1)u = v_1(u_2 - u_1) - u_1(v_2 - v_1) \tag{9}$$

where the differences

$$x = u_2 - u_1, \qquad y = v_2 - v_1 \tag{10}$$

determine the slope of the line and $-n < x < n,\ -m < y < m$.

It is easy to see that the number of right triangles with vertices in the grid points and having vertical and horizontal legs x, y is

$$g_{m,n}(x,y) = \begin{cases} (n - |x|)(m - |y|), & \text{if } |x| < n \text{ and } |y| < m; \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

since the leg x can be selected in $n - |x|$ ways and the leg y in $m - |y|$ ways within the $m \times n$ grid of points. If $x = 0$ or $y = 0$, the triangle reduces to a vertical or a horizontal line segment.

However, $g_{m,n}(x, y)$ typically exceeds the number of lines with slope (10) since some of hypotenuses of the triangles locate on the same line. It may now be assumed without loss of generality that $(x, y) = 1$. Then it is important to notice that if a line includes N triangles with legs x, y, it includes $N - 1$ (halfly overlapped) triangles with legs $2x, 2y$. Therefore the number of distinct lines with slope (10) is

$$M(x,y) = g_{m,n}(x,y) - g_{m,n}(2x,2y) \tag{12}$$

where it is assumed that $(x, y) = 1$.

Finally, by observing that legs $-x, -y$ lead to same lines as legs x, y and by summing (12) over all possible slopes, the equation (7) is shown to be valid.

Next it will be shown that the equation (6) generalized for an $m \times n$ grid is

$$L_j(m,n) = \frac{1}{2}[f(m,n,j+1) - 2f(m,n,j) + f(m,n,j-1)] \tag{13}$$

where $f(m, n, k)$ is defined as (8), is valid for $j = 2, 3, \ldots, \max(m, n)$.

Again, let us study lines with a fixed slope x, y where $(x, y) = 1$ within the grid. Assume that such a line goes through exactly h points of the grid, for example, through points

$$(u_0 + ix, v_0 + iy), i = 0, 1, 2, \ldots, h-1.$$

Let $d(i)$ be the number of triangles with legs ix, iy along the line and consider a characteristic

$$p = d(j+1) - 2d(j) + d(j-1). \tag{14}$$

If $h < j$, we have $d(j-1) = d(j) = d(j+1) = 0$ and $p = 0$.

If $h = j$ (i.e. the line goes through exactly j points), we have $d(j-1) = 1$ (only one triangle included with legs $(j-1)x, (j-1)y)$, $d(j) = d(j+1) = 0$ and $p = 1$.

If $h = j+1$, we have $d(j-1) = 2, d(j) = 1, d(j+1) = 0$ and $p = 0$.

If $h = q > j+1$, we have
$d(j-1) = q - j + 1$ triangles with legs $(j-1)x, (j-1)y$,
$d(j) = q - j$ triangles with legs jx, jy,
$d(j+1) = q - j - 1$ triangles with legs $(j+1)x, (j+1)y$,
and $p = (q - j + 1) - 2(q - j) + (q - j - 1) = 0$.

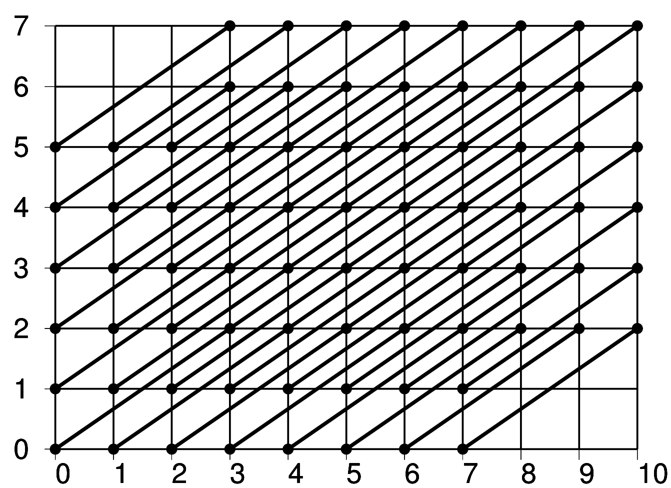
Thus $p = 1$ if $h = j$ and $p = 0$ otherwise.

Since the total number of right triangles with legs ix, iy in the grid is $g_{m,n}(ix, iy)$ (here especially for $i = j-1, j, j+1$), and each triangle is related exactly to one line, the total number of lines with slope $x, y, (x, y) = 1$ and going through exactly j points is according to p values

$$\begin{aligned} M_j(x, y) = g_{m,n}((j+1)x, (j+1)y) - 2g_{m,n}(jx, jy) \\ + g_{m,n}((j-1)x, (j-1)y). \end{aligned} \tag{15}$$

Again, by observing that legs $-x, -y$ lead to same lines as legs x, y and by summing the $M_j(x, y)$ values over all possible slopes, the equation (13) is shown to be valid.

As an example, let us study a grid with $m = 8, n = 11$



where lines with slope $x = 3, y = 2$ are drawn. The number of right triangles with legs 3,2 is $(n - |x|)(m - |y|) = 48$ and the corresponding number for 6,4 is

$(n-|2x|)(m-|2y|)=20$ and so according to (12) the number of lines with slope 3,2 is $48-20=28$. The number of lines with this slope and going through exactly 2 points is according to (15)
$M_2(3,2)=(11-3\cdot 3)(8-3\cdot 2)-2(11-2\cdot 3)(8-2\cdot 2)+(11-3)(8-2)=12.$
Similarly,
$M_3(3,2)=0-2(11-3\cdot 3)(8-3\cdot 2)+(11-2\cdot 3)(8-2\cdot 2)=12.$
and
$M_4(3,2)=0-0+(11-3\cdot 3)(8-3\cdot 2)=4.$

The total number of lines through at least 2 points of the grid is $L(8,11)=1759$ and values for $L_j(8,11)$ are

j	$L_j(8,11)$	
2	1430	
3	200	
4	72	
5	16	
6	10	
7	4	
8	19	vertical and diagonal lines
9	0	
10	0	
11	8	horizontal lines

4. Asymptotic behaviour

No closed or recursive formula for $L(n)$ numbers is known [2]. The behaviour of $L(n)$ seems to be essentially dependent on the divisibility of numbers $2,3,\ldots,n-1$ as one can see from the ratios $L(n)/L(n-1)$ in Table 2.

The $L(n)/L(n-1)$ numbers are typically decreasing, but not obviously when $n-1$ is a prime number or a number with a great smallest prime factor. This phenomenon is a reflection of the fact that if $n-1$ is such an integer, the amount of various new slopes of lines is greater than for integers with small divisors. Then there may be some hope for finding an expression for $L(n)$ depending on some number theoretic function like Euler's totient function $\phi(n-1)$ appearing in the table.

Numeric values of $L(n)$ can be calculated rather efficiently by the formula (2) and it is now done for $n=2,3,\ldots,15000$. These $L(n)$ values are available in
`http://www.survo.fi/papers/Lseq.zip`
and this list extends the table for 100 first values compiled by T.D.Noe in connection of sequence A018808 in [3].

Furthermore, values of $L(n)$ has been calculated even for some greater n.

Since the number of lines going through 2 points in a randomly distorted $n\times n$ grid is $n^2(n^2-1)/2$ with probability 1, it is plausible that

$$L(n)\simeq Cn^4. \tag{16}$$

[2]In fact, I have found (19 April 2009) such a formula by studying $L(n)$ and $L(n-1,n)$ sequences numerically. See Appendix 1.

n	$L(n)$	$L(n)/L(n-1)$	$n-1$	$L(n)/n^4$	$\phi(n-1)$
2	6	-	1	0.37500	1
3	20	3.33333	2	0.24691	1
4	62	3.10000	3	0.24219	2
5	140	2.25806	4	0.22400	2
6	306	2.18571	5	0.23611	4
7	536	1.75163	6	0.22324	2
8	938	1.75000	7	0.22900	6
9	1492	1.59062	8	0.22740	4
10	2306	1.54558	9	0.23060	6
11	3296	1.42931	10	0.22512	4
12	4722	1.43265	11	0.22772	10
13	6460	1.36806	12	0.22618	4
14	8830	1.36687	13	0.22985	12
15	11568	1.31008	14	0.22850	6
16	14946	1.29201	15	0.22806	8
17	18900	1.26455	16	0.22629	8
18	23926	1.26593	17	0.22792	16
19	29544	1.23481	18	0.22670	6
20	36510	1.23578	19	0.22819	18
21	44388	1.21578	20	0.22824	8
22	53586	1.20722	21	0.22875	12
23	63648	1.18777	22	0.22744	10
24	75674	1.18895	23	0.22809	22
25	88948	1.17541	24	0.22771	8
26	104374	1.17343	25	0.22840	20
27	121032	1.15960	26	0.22774	12
28	139966	1.15644	27	0.22771	18
29	160636	1.14768	28	0.22712	12
30	184466	1.14835	29	0.22774	28

TABLE 2. $L(n)$ values related to divisibility of $n-1$

The value $L(40000) = 583610033692337762$ divided by 40000^4 gives an approximation $C' = 0.227972669\ldots$ and it was easy to note that $C'\pi^2 = 2.250000067\cdots \approx 9/4$. Thus it seems likely that

$$C = [3/(2\pi)]^2 = 0.227972663195\ldots \tag{17}$$

and then $C' - C \approx -6.8 \cdot 10^{-9}$.

Still a better approximation was obtained by calculating $L(60000) = 2954525721400635290$ giving $C'\pi^2 = 2.2500000048\ldots$ and $C' - C \approx -4.9 \cdot 10^{-10}$.

Analogously, for an $m \times n$ grid, an asymptotic expression

$$L(m, n) = [3/(2\pi)mn]^2 \tag{18}$$

seems to be valid.

Asymptotic expression (16) with C given by (17) works well also for smaller values of n.

In Fig. 1 differences

$$D(n) = L(n) - Cn^4 \tag{19}$$

are displayed for $n = 2, 3, \ldots, 15000$.

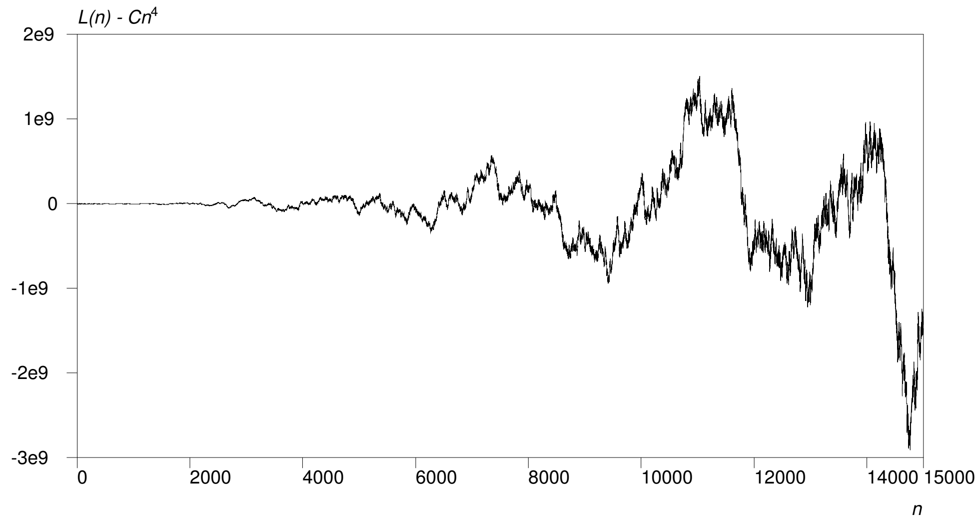


FIGURE 1. Deviances of $L(n)$ from the asymptotic expression

On the basis of this data, it seems that $D(n)$ might have magnitude of $O(n^2\sqrt{n}) = O(n^{2.5})$ as one can see in Fig. 2.

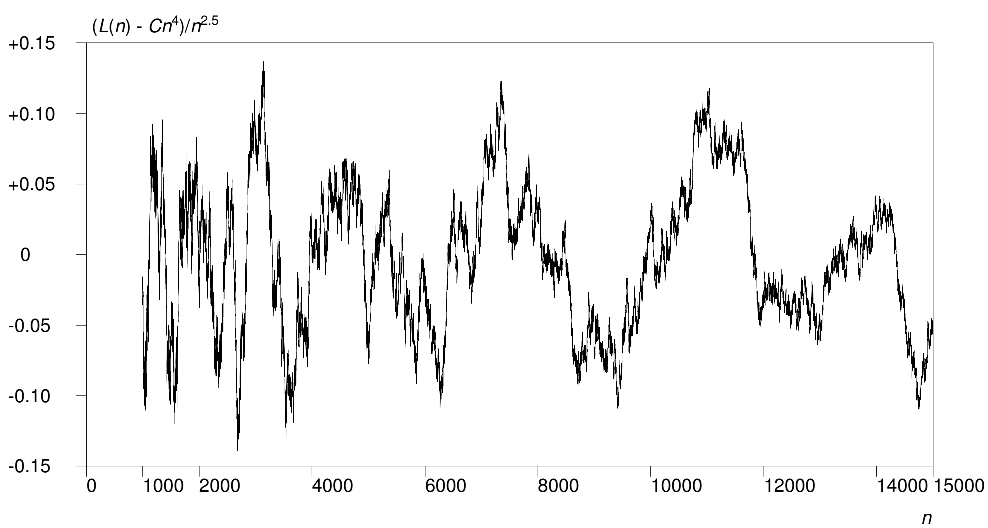


FIGURE 2. Proportional deviances of $(L(n) - Cn^4)/n^{2.5}$ from the asymptotic expression

Thus my conjecture is that a more accurate asymptotic expression for $L(n)$ is [3]

$$L(n) = [3/(2\pi)n^2]^2 + O(n^{2.5}). \tag{20}$$

5. Points of intersection

When studying points of intersection between lines through at least 2 points in an $m \times n$ grid, it is not enough to know the number of lines concerned $L(m,n)$. Now each such a line should be uniquely identified. In order to achieve this goal, in a general equation of a straight line

$$ax + by + c = 0$$

it may be assumed without loss of generality that a, b, c are integers without any common factors, $a \geq 0$, and if $a = 0$, $b > 0$. The point of intersection of two lines

$$a_1x + b_1y + c_1 = 0, \quad a_2x + b_2y + c_2 = 0$$

is (x_0, y_0) where

$$x_0 = (b_1c_2 - b_2c_1)/d, \quad y_0 = (a_2c_1 - a_1c_2)/d$$

and $d = a_1b_2 - a_2b_1$. Thus x_0 and y_0 are rational numbers, say, $x_0 = u_1/v_1$ and $y_0 = u_2/v_2$ where without loss of generality, it is assumed that $(u_1, v_1) = (u_2, v_2) = 1$. When the point of intersection exists (i.e. $d \neq 0$), the 4-tuple (u_1, v_1, u_2, v_2) of integers (for $v_1, v_2 > 0$) accurately and uniquely identifies the point.

The number of points of intersection $S(m,n)$ is determined by making a list of all distinct 4-tuples satisfying the conditions

$$0 \leq u_1 \leq (m-1)v_1, \quad 0 \leq u_2 \leq (n-1)v_2$$

on the basis of all combinations of $L(m,n)$ lines.

This is much harder than calculating $L(m,n)$ since there seems to be no formula of type (7), for example, and the $S(m,n)$ values are essentially greater than $L(m,n)$ values. Currently, $S(n,n) = S(n)$ values have been calculated only for $n = 2, 3, \ldots, 30$ and these values with certain derived 'statistics' are presented in Table 3.

As for $L(n)$ values, the ratios of subsequent values are decreasing except when $n-1$ is a prime number. In this case, it is natural to assume that $S(n) \simeq C_2n^8$ where C_2 is a constant of magnitude 0.0075 as indicated by the last column in Table 3.

[3]Computations have now been extended to $n = 10^8$, see 8.1.

n	$S(n)$	$S(n)/S(n-1)$	$n-1$	$S(n)/n^8$
2	5	-	1	0.0195312500
3	37	7.4000000000	2	0.0056393842
4	405	10.9459459459	3	0.0061798096
5	2225	5.4938271605	4	0.0056960000
6	11641	5.2319101124	5	0.0069307508
7	35677	3.0647710678	6	0.0061887652
8	114409	3.2067998991	7	0.0068193078
9	295701	2.5845956175	8	0.0068693037
10	718469	2.4297144751	9	0.0071846900
11	1475709	2.0539633582	10	0.0068842914
12	3093025	2.0959586206	11	0.0071933876
13	5771929	1.8661113311	12	0.0070757774
14	10895273	1.8876311542	13	0.0073826764
15	18785841	1.7242193931	14	0.0073299425
16	31414269	1.6722311767	15	0.0073142045
17	50274501	1.6003715063	16	0.0072070311
18	81288641	1.6168960285	17	0.0073764911
19	124066161	1.5262422827	18	0.0073050726
20	190860537	1.5383770680	19	0.0074554897
21	282399889	1.4796138240	20	0.0074663813
22	411505049	1.4571714262	21	0.0074988337
23	580614301	1.4109530428	22	0.0074142127
24	824814797	1.4205898745	23	0.0074931859
25	1138709849	1.3805642832	24	0.0074626489
26	1570665877	1.3793380977	25	0.0075213712
27	2115178249	1.3466761327	26	0.0074892247
28	2833746309	1.3397198606	27	0.0075006123
29	3732420861	1.3171330296	28	0.0074611647
30	4937226173	1.3227946035	29	0.0075251123

TABLE 3. $S(n)$ values related to divisibility of $n-1$

REFERENCES

[1] D.E.Knuth, *The Art of Computer Programming, Vol. 3: Sorting and Searching*, 2nd ed. Reading, MA: Addison-Wesley, pp. 462 – 464, 1998.

[2] S.Mustonen, Statistical accuracy of geometric constructions, pp. 61 – 63, 2008. `http://www.survo.fi/papers/GeomAccuracy.pdf`

[3] Sloane, N. J. A. "The On-Line Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences`

6. Appendix 1: Recursive formulas

With the aid of Euler's totient function $\phi(n)$ I have found the following recursive formulas after making some experimental studies with the numerical values of $L(n) = L(n,n)$ and $L(n-1,n)$:

$$\begin{aligned} L(n,n) &= 2L(n-1,n) - L(n-1,n-1) + R_1(n), \\ L(n-1,n) &= 2L(n-1,n-1) - L(n-2,n-1) + R_2(n) \end{aligned} \tag{21}$$

where

$$\begin{aligned} R_1(n) &= R_1(n-1) + 4(\phi(n-1) - e(n)), \\ e(n) = 0 \quad \text{if } n \text{ is even}, &\quad e(n) = \phi((n-1)/2) \quad \text{if } n \text{ is odd} \end{aligned} \tag{22}$$

and

$$R_2(n) = \begin{cases} (n-1)\phi(n-1) & \text{if } n \text{ is even;} \\ (n-1)\phi(n-1)/2 & \text{if } n \equiv 1 \pmod 4; \\ 0 & \text{if } n \equiv 3 \pmod 4) \end{cases} \tag{23}$$

with initial values $L(0,0) = L(0,1) = R_1(1) = 0$.

These formulas are much faster in calculations than earlier ones. For example, all $L(n)$'s for $n = 2, 3, ..., 60000$ are computed and saved in a file in less than 0.3 seconds on my PC by a C program written as a Survo program module and working iteratively.

I have also written Mathematica code

```
L[0]=0;
L1[1]=0;
R1[1]=0;
L[n_]:=L[n]=2*L1[n]-L[n-1]+R1[n]
L1[n_]:=L1[n]=2*L[n-1]-L1[n-1]+R2[n]
R1[n_]:=R1[n]=R1[n-1]+4*(EulerPhi[n-1]-e[n])
e[n_]:=If[Mod[n,2]==0,0,EulerPhi[(n-1)/2]]
R2[n_]:=
If[Mod[n,2]==0,(n-1)*EulerPhi[n-1],
If[Mod[n,4]==1,(n-1)*EulerPhi[n-1]/2,0]]
Table[L[n],n,0,50]
```

working in truly recursive manner and giving

```
{0, 0, 6, 20, 62, 140, 306, 536, 938, 1492, 2306, 3296, 4722, 6460,
8830, 11568, 14946, 18900, 23926, 29544, 36510, 44388, 53586, 63648,
75674, 88948, 104374, 121032, 139966, 160636, 184466, 209944, 239050,
270588, 305478, 342480, 383370, 427020, 475830, 527280, 583338,
642900, 708798, 777912, 854022, 934604, 1021074, 1111368, 1209994,
1313612, 1425770}
```

Also it does the job rather quickly giving values for $n \leq 60000$ in less than 5 seconds.

I found these recursive formulas by making some experiments by using the SURVO MM system. When preparing this paper, I had created certain C programs as Survo modules for computing values of $L(m,n)$. One of them is `LMN` giving $L(m,n)$ values according to (7):

Date 16 April 2009

```
____________________________________________________________
*LMN 4,4
*L(4,4): 62
*LMN 5,5
*L(5,5): 140
*LMN 11,8
*L(11,8): 1759
*LMN 99,100
*L(99,100): 22338303
____________________________________________________________
```

A summary of my experiment related to the first of the recurrence equations (21) is given in the following excerpt from a Survo edit field.

```
____________________________________________________________
*DATA LNN,A,B,N,M
*
N  n  Lnn    Ln1n   Ln1n1  R1p4  d phi1 diff  e
M 11 111111 111111 111111 1111  11   11 11   11
A  3     20     11      6    1   -    1  -    1
*  4     62     35     20    3   2    2  0    0
*  5    140     93     62    4   1    2  1    1
*  6    306    207    140    8   4    4  0    0
*  7    536    405    306    8   0    2  2    2
*  8    938    709    536   14   6    6  0    0
*  9   1492   1183    938   16   2    4  2    2
* 10   2306   1855   1492   22   6    6  0    0
* 11   3296   2757   2306   22   0    4  4    4
* 12   4722   3945   3296   32  10   10  0    0
* 13   6460   5523   4722   34   2    4  2    2
* 14   8830   7553   6460   46  12   12  0    0
* 15  11568  10107   8830   46   0    6  6    6
* 16  14946  13149  11568   54   8    8  0    0
* 17  18900  16807  14946   58   4    8  4    4
* 18  23926  21265  18900   74  16   16  0    0
* 19  29544  26587  23926   74   0    6  6    6
* 20  36510  32843  29544   92  18   18  0    0
* 21  44388  40257  36510   96   4    8  4    4
* 22  53586  48771  44388  108  12   12  0    0
* 23  63648  58401  53586  108   0   10 10   10
* 24  75674  69401  63648  130  22   22  0    0
* 25  88948  82043  75674  134   4    8  4    4
* 26 104374  96353  88948  154  20   20  0    0
* 27 121032 112395 104374  154   0   12 12   12
* 28 139966 130155 121032  172  18   18  0    0
* 29 160636 149945 139966  178   6   12  6    6
* 30 184466 172139 160636  206  28   28  0    0
* 31 209944 196793 184466  206   0    8  8    8
* 32 239050 224025 209944  236  30   30  0    0
* 33 270588 254331 239050  244   8   16  8    8
* 34 305478 287505 270588  264  20   20  0    0
B 35 342480 323451 305478  264   0   16 16   16
____________________________________________________________
```

Since an $n \times n$ grid can be constructed from an $(n-1)\times(n-1)$ grid first by adding a new column of $n-1$ points and then a new row of n points, it seemed natural to see a possible simple relation between $L(n, n)$, $L(n-1, n)$, and $L(n-1, n-1)$. Thus the data set LNN was created by first computing values of them as variables Lnn, Ln1n, and Ln1n1 for $n = 3, 4, \ldots, 35$ using the new LMN command of SURVO MM.

A simple linear regression model gave the following results

```
____________________________________________________________
150 LINREG LNN,CUR+1 / VARS=Lnn(Y),Ln1n(X),Ln1n1(X) RESULTS=0
151 Linear regression analysis: Data LNN, Regressand Lnn       N=33
152 Variable Regr.coeff.         Std.dev.         t          beta
153 Ln1n      2.13682629851       0.00431933063  494.712371  2.01238114
154 Ln1n1    -1.14163994110       0.00458688877 -248.892004 -1.01243795
155 constant  37.1962120274       5.15255468344  7.21898443
156 Variance of regressand Lnn=9882609734.98485 df=32
157 Residual variance=291.690892945110 df=30
158 R=0.999999986 R^2=0.999999972
____________________________________________________________
```

showing very good 'statistical' relation between variables concerned. Since the regression coefficient of Ln1n is close to 2 and that of Ln1n1 is close to -1, we have $L(n, n) \approx 2L(n-1, n) - L(n-1, n-1)$.

All values of the difference $L(n, n) - 2L(n-1, n) + L(n-1, n-1)$ are divisible by 4 on the range $n = 3, 4, \ldots, 35$. Therefore the variable R1p4 was computed by a VAR command as

```
VAR R1p4=(Lnn-2*Ln1n+Ln1n1)/4 TO LNN
```

These residuals are growing monotonously except for every fourth n where the same value appears twice. Then it is natural to take first differences d by

```
VAR d=R1p4-R1p4[-1] TO LNN
```

Now it is easy to note that d values are equal to $n-2$ when $n-1$ is a prime number, i.e. $\mathtt{d}(n) = \phi(n-1)$ when $n-1$ is a prime. Thus these values in general may be related to values $\phi(n-1)$ of Eulers's Totient function. These values are computed as phi1 by

```
VAR phi1=totient(n-1) TO LNN
```

Next, differences $\phi(n-1) - \mathtt{d}(n)$ are computed as a variable diff by

```
VAR diff=phi1-d TO LNN
```

and it can be seen that the diff values are 0 when n is an even integer. It was not difficult to see that for the odd values of n we have $\mathtt{diff}(n) = \phi((n-1)/2)$ which is confirmed by computing the last column as e by

```
VAR e=phi2 TO LNN
phi2=if(mod(n,2)=0)then(0)else(totient((n-1)/2))
```

Thus the columns diff and e are identical and it is easy to deduce by taking the corresponding steps backwards that the first of the recurrence equations (21) is valid for $n = 3, 4, \ldots, 35$.

The recursive equation for $L(n, n)$ depends also on $L(n-1, n)$ and it is rather worthless without knowing a formula for $L(n-1, n)$. Now it is reasonable to expect that a similar recurrence is valid for $L(n-1, n)$ numbers, too. In fact, $R_2(n) = L(n-1, n) - 2L(n-1, n-1) + L(n-2, n-1)$ are suitable residuals in this case and these values are computed for $n = 3, 4, \ldots, 35$ in Survo by a VAR command on line 248 as a variable R2.

```
_____________________________________________________________
210 *DATA LMN,a,b,n,m
211 *
212 n  n  Ln1n   Ln1n1  Ln2n1   R2   e
213 m 11 111111 111111 111111  111 111
214 a  3     11      6      1    0   0
215 *  4     35     20     11    6   6   3*2=6
216 *  5     93     62     35    4   4           4*totient(4)/2=4
217 *  6    207    140     93   20  20   5*4=20
218 *  7    405    306    207    0   0
219 *  8    709    536    405   42  42   7*6=42
220 *  9   1183    938    709   16  16           8*totient(8)/2=16
221 * 10   1855   1492   1183   54  54           9*totient(9)=54
222 * 11   2757   2306   1855    0   0
223 * 12   3945   3296   2757  110 110   11*10=110
224 * 13   5523   4722   3945   24  24           12*totient(12)/2=48
225 * 14   7553   6460   5523  156 156   13*12=156
226 * 15  10107   8830   7553    0   0
227 * 16  13149  11568  10107  120 120           15*totient(15)=120
228 * 17  16807  14946  13149   64  64           16*totient(16)/2=64
229 * 18  21265  18900  16807  272 272   17*16=272
230 * 19  26587  23926  21265    0   0
231 * 20  32843  29544  26587  342 342   18*17=306
232 * 21  40257  36510  32843   80  80           20*totient(20)/2=80
233 * 22  48771  44388  40257  252 252           21*totient(21)=252
234 * 23  58401  53586  48771    0   0
235 * 24  69401  63648  58401  506 506   23*22=506
236 * 25  82043  75674  69401   96  96           24*totient(24)/2=96
237 * 26  96353  88948  82043  500 500           25*totient(25)=500
238 * 27 112395 104374  96353    0   0
239 * 28 130155 121032 112395  486 486           27*totient(27)=486
240 * 29 149945 139966 130155  168 168           28*totient(28)/2=168
241 * 30 172139 160636 149945  812 812   29*28=812
242 * 31 196793 184466 172139    0   0
243 * 32 224025 209944 196793  930 930   31*30=930
244 * 33 254331 239050 224025  256 256           32*totient(32)/2=256
245 * 34 287505 270588 254331  660 660           33*totient(33)=660
246 b 35 323451 305478 287505    0   0
247 *
248 *VAR R2=Ln1n-2*Ln1n1+Ln2n1 TO LMN
249 *
250 *VAR e=D1 TO LMN
251 *D1=if(mod(n,2)=0)then((n-1)*totient(n-1))else(D2)
252 *D2=if(mod(n,4)=3)then(0)else((n-1)/2*totient(n-1))
_____________________________________________________________
```

It is immediately detected that this is a simpler case than the previous one. Surprisingly these residuals are zero for $n = 3, 7, 11, \ldots$ i.e. when $n \equiv 3 \pmod 4$. It is easy to detect general rules also for non-zero residuals. If $n-1$ is a prime number, the residual $R_2(n)$ is equal to $(n-1)(n-2) = (n-1)\phi(n-1)$ and the latter expression is valid also for any even n, i.e. $R_2(n) = (n-1)\phi(n-1)$ when n is even. The remaining case is $n \equiv 1 \pmod 4$ and then we have $R_2(n) = (n-1)\phi(n-1)/2$. These results are checked by a `VAR` command on lines 250 – 252.

Of course, these considerations related to the recursive formula (21) have nothing to do with a strict proof, although the results completely agree with those obtained by formula (2) at least for $n = 2, 3, \ldots, 15000$ and for $n = 40000$ and $n = 60000$. So there is still a challenge to to prove these results generally, although there are hardly any suspects about their validity.

The recurrence takes place by computing $L(n, n)$ and $L(n-1, n)$ values alternatively. It may also be possible to find a recursive formula of type

$$L(n) = 2L(n-1) - L(n-2) + R(n) \tag{24}$$

since computational experiments indicate that $2L(n-1) - L(n-2)$ is a rather good approximation of $L(n)$. At the moment I have no suggestion for a formula of the remainder $R(n)$. Another alternative is to derive a direct formula for $L(n)$ by iterating (21).

By means of recursive formulas (21), it is possible to compute $L(n)$ values for larger n. I have used the Mathematica code presented above since it works with arbitrary precision integers.

I got, for example, $L(2000000) = 3647562610795135871970078$ and this divided by 2000000^4 gave $C' = 0.2279726631746959919981298 75$ and $C' - C \approx -2.1 \cdot 10^{-11}$. Thus C' approximates now $C = [3/(2\pi)]^2$ about 100 times more accurately than for $n = 60000$.

The proportional deviances $(L(n) - Cn^4)/n^{2.5}$ can be seen in a graph
`http://www.survo.fi/papers/DevLn2009.pdf`
for $n = 1000, 1001, 1002, \ldots, 2000000$. The absolute values of these deviances are less that 0.17 so that these new results do not violate the conjecture (20).

7. Appendix 2: Recursive formulas 2

It is logical to study $L(m,n)$ numbers for a fixed m in a similar manner. It turns out that, in general, a recurrence of the form (24), i.e.

$$L(m,n) = 2L(m,n-1) - L(m,n-2) + R(m,n) \tag{25}$$

is valid and the residual terms $R(m,n)$ are *periodic*.

Case $m = 2$:
Trivially we have

$$L(2,n) = n^2 + 2, \quad n = 2,3,\ldots \tag{26}$$

since a $2 \times n$ grid has 2 horizontal lines and the n points on the 'top line' can be connected to points on the 'bottom line' in n^2 ways.

Case $m = 3$:

```
____________________________________________________________
*DATA L3N,A,B,N,M
N  n    L  R3  L3n    D
M 11 1111 111 1111  111
A  2   11   -   11    0
*  3   20   -   20    0
*  4   35   6   35    0
*  5   52   2   52    0
*  6   75   6   75    0
*  7  100   2  100    0
*  8  131   6  131    0
*  9  164   2  164    0
* 10  203   6  203    0
* 11  244   2  244    0
* 12  291   6  291    0
* 13  340   2  340    0
* 14  395   6  395    0
* 15  452   2  452    0
* 16  515   6  515    0
* 17  580   2  580    0
* 18  651   6  651    0
* 19  724   2  724    0
B 20  803   6  803    0
*
*VAR R3=L-2*L[-1]+L[-2] TO L3N
*VAR L3n=2*n^2+3-mod(n,2) TO L3N
*VAR D=L-L' TO L3N
*L3(n):=2*n^2+3-mod(n,2)
*L3(1000)=2000003
*LMN 3,1000
```

Date 18 May 2009

```
128 *L(3,1000): 2000003
____________________________________________________________
```

In the display above from a Survo edit field the $L(3, n)$ values have been computed by the LMN command as Land the residuals $R(3, n)$ by the VAR command on line 122 as R3.
Thus the residuals are

$$R(3, n) = \begin{cases} 6 & \text{if } n \text{ is even;} \\ 2 & \text{if } n \text{ is odd} \end{cases} \tag{27}$$

and the length of the period is 2.
It can be shown (for example, by studying $L(3, n) - L(3, n - 1)$) that

$$L(3, n) = 2n^2 + 3 - \operatorname{mod}(n, 2) \tag{28}$$

and it is also the solution of the difference equation (25). Values of (28) are computed as L3n by the VAR command on line 123 and the difference D=L-L3n is shown to be zero by the command on line 124. The lines 125 – 128 indicate that the formula (28) is valid for $n = 1000$.

Case $m = 4$**:**
Residuals $R(4, n)$ have a period (10,4,12,2,12,4) of length 6.

```
____________________________________________________________
100 *DATA L4N,A,B,N,M
101 N  n      L    R4  Dcheck
102 M 11   1111  1111   1111
103 A  2     18     -      -
104 *  3     35     -      -
105 *  4     62    10     10
106 *  5     93     4      4
107 *  6    136    12     12
108 *  7    181     2      2
109 *  8    238    12     12
110 *  9    299     4      4
111 * 10    370    10     10
112 * 11    445     4      4
113 * 12    532    12     12
114 * 13    621     2      2
115 * 14    722    12     12
116 * 15    827     4      4
117 * 16    942    10     10
118 * 17   1061     4      4
119 * 18   1192    12     12
120 * 19   1325     2      2
121 B 20   1470    12     12
122 *
123 *VAR R4=L-2*L[-1]+L[-2] TO L4N
124 *
125 C 10 4 12 2 12 4
126 *VAR Dcheck=X(C,mod(n+2,6)+1) TO L4N
127 *            X(C,n) is the n'th number on line C.
128 *LMN 4,100
129 *L(4,100): 36670
```

Then $R(4,n)$ can be written as

$$R(4,n) = C(\mathrm{mod}((n+2,6)+1), \quad C = (10,4,12,2,12,4). \tag{29}$$

This formula is 'validated' by Survo commands on lines 125 – 126.

Mathematica code for this case is

```
L4[2]=18;
L4[3]=35;
L4[n_]:=L4[n]=L4[n]=2*L4[n-1]-L4[n-2]+R[n]
c4=10,4,12,2,12,4;
R[n_]:=c4[[Mod[n+2,6]+1]]
Table[L4[n],n,2,100]
```

```
18, 35, 62, 93, 136, 181, 238, 299, 370, 445, 532, 621, 722, 827,
942, 1061, 1192, 1325, 1470, 1619, 1778, 1941, 2116, 2293, 2482,
2675, 2878, 3085, 3304, 3525, 3758, 3995, 4242, 4493, 4756, 5021,
5298, 5579, 5870, 6165, 6472, 6781, 7102, 7427, 7762, 8101, 8452,
8805, 9170, 9539, 9918, 10301, 10696, 11093, 11502, 11915, 12338,
12765, 13204, 13645, 14098, 14555, 15022, 15493, 15976, 16461, 16958,
17459, 17970, 18485, 19012, 19541, 20082, 20627, 21182, 21741, 22312,
22885, 23470, 24059, 24658, 25261, 25876, 26493, 27122, 27755, 28398,
29045, 29704, 30365, 31038, 31715, 32402, 33093, 33796, 34501, 35218,
35939, 36670
```

$L(4,100) = 36670$ is confirmed by the `LMN` command on lines 128 – 129.

The solution of the difference equation (25) for $m = 4$ is

$$L(4,n) = 4n^2 - 3\lfloor n^2/9 \rfloor + C(mod(n-2,18)+1) \tag{30}$$

where

$$C = (2,2,1,-1,4,0,3,2,3,0,4,-1,1,2,2,1,4,1).$$

Similar expressions may be derived for other n values as well, but the length of the C vector grows rapidly with m. Thus a more general approach should be adopted.

General recursion formula for $L(m,n)$:

At first the length of the period may be shortened by observing that, when adding the n^{th} column to an $m \times (n-1)$ grid, the number of lines with *new* slopes, necessarily of form $(n-1,y)$, is $2\psi_1(m,n)$ where

$$\psi_1(m,n) = \sum_{\substack{y=1 \\ (y,n-1)=1}}^{m-1} m - y \tag{31}$$

Then, for example in case $m = 4$ the length of the period in residuals

$$R_1(m,n) = L(m,n) - 2L(m,n-1) + L(m,n-2) - 2\psi_1(m,n) \tag{32}$$

drops to 2 and they seem to be identically zero for even values of n but equal to -4 for odd values.

```
------------------------------------------------------------
*DATA L4N,A,B,N,M
Nm  n       L  PSI1   R4
M1 11 111111 11111  111
A4  2      18     6    -
*4  3      35     4    -
*4  4      62     5    0
*4  5      93     4   -4
*4  6     136     6    0
*4  7     181     3   -4
*4  8     238     6    0
*4  9     299     4   -4
*4 10     370     5    0
*4 11     445     4   -4
*4 12     532     6    0
*4 13     621     3   -4
*4 14     722     6    0
*4 15     827     4   -4
*4 16     942     5    0
*4 17    1061     4   -4
*4 18    1192     6    0
*4 19    1325     3   -4
B4 20    1470     6    0
*
*psi1(M,N):=for(y=1)to(M-1)sum(psi12(M,N,y))
*psi12(M,N,y):=if(gcd(y,N-1)=1)then(M-y)else(0)
*
*VAR PSI1=psi1(m,n) TO L4N
*VAR R4=L-2*L[-1]+L[-2]-2*PSI1 TO L4N
------------------------------------------------------------
```

As a more representative example, let's study the case $m = 5$ in a similar way.

```
------------------------------------------------------------
*DATA L5N,A,B,N,M
N m  n       L  PSI1   R5
M11 11 111111 11111  111
A 5  2      27    10    -
* 5  3      52     6    -
* 5  4      93     8    0
* 5  5     140     6   -6
* 5  6     207    10    0
* 5  7     274     4   -8
* 5  8     361    10    0
* 5  9     454     6   -6
* 5 10     563     8    0
* 5 11     676     6   -8
* 5 12     809    10    0
* 5 13     944     4   -6
* 5 14    1099    10    0
* 5 15    1258     6   -8
* 5 16    1433     8    0
* 5 17    1614     6   -6
* 5 18    1815    10    0
```

```
* 5 19   2016     4   -8
B 5 20   2237    10    0
*
*psi1(M,N):=for(y=1)to(M-1)sum(psi12(M,N,y))
*psi12(M,N,y):=if(gcd(y,N-1)=1)then(M-y)else(0)
*
*VAR PSI1=psi1(m,n) TO L5N
*VAR R5=L-2*L[-1]+L[-2]-2*PSI1 TO L5N
------------------------------------------------------------
```

Again the residuals seem to be identically zero for even n. For odd values residuals of alternative values -6 and -8 appear.

The next figure illustrates the situation in case $m = 5, n = 7$. In this graph, all possible ascending lines connecting at least 2 points in a 5×7 grid of points are drawn separately for different slopes. Lines of the partial 5×5 grid are blue, additional lines possible in the partial 5×6 grid are red, and additional lines in the complete 5×7 grid are black. For any partial graph (with a constant slope) let N_5 be the number of blue lines, N_6 the number of red lines, and N_7 the number of black lines. The last graph of 4 black lines tells that $\psi_1(5, 7) = 4$. In most of the remaining graphs $D_7 = N_7 - 2N_6 + N_5$ is zero as it is trivially for lines with slopes $(1, 0)$ and $(0, 1)$, i.e. horizontal and vertical lines. Only in graphs corresponding to slopes (3,2) and (3,1) D_7 deviates from zero being -1 and -3, respectively. Their sum -4 multiplied by 2 (corresponding cases of descending lines) gives -8 which is the residual `R5` for $m = 5, n = 7$.

In these partial graphs $D_7 = 0$ just indicates that the number of red lines equals to the number of black lines. This can be explained as follows. Let $P_1 = (u_1, v_1)$ and $P_2 = (u_2, v_2)$ be the two first integer points (from left to right) of a line with slope $(u_2 - u_1, v_2 - v_1)$. Let's call points P_1, P_2 *bearing points*. If the line is red, u_2 must be 5 ($n - 2$ in general) and then there exists a parallel black line with $Q_1 = (u_1 + 1, v_1)$ and $Q_2 = (u_2 + 1, v_2)$ as bearing points. However, if D_7 is not zero, in some cases Q_1 and Q_2 are not bearing points, but $Q_0 = (2u_1 - u_2, 2v_1 - v_2)$ and $Q1$ are and thus a potential black line is in fact blue. This happens for slope $(3, 1)$ three times for lines with $P_1 = (2, 3)$, $P_1 = (2, 2)$, $P_1 = (2, 1)$ and for slope $(3, 2)$ once for a line with $P_1 = (2, 2)$.

In a general $m \times n$ grid, let's study a red line with slope (x, y), $x, y > 0$, $(x, y) = 1$ and bearing points $P_1 = (u_1, v_1), P_2 = (u_1 + x, v_1 + y)$. The line is red only if $u_1 + x = n - 2$ and $P_0 = (u_1 - x, v_1 - y)$ is outside the grid. Then we have $P_1 = (n - 2 - x, v_1)$ and $n - 2 - 2x < 0$ and/or $v_1 - y < 0$, i.e.

$$x > (n - 2)/2 \ \vee \ v_1 < y. \tag{33}$$

The potential black line one step to the right of the previous red line has bearing points $Q_1 = (n - 1 - x, v_1), Q_2 = (n - 1, v_1 + y)$ unless $Q_0 = (n - 1 - 2x, v_1 - y)$ is inside or on the border of the grid. Then this line is not a black but a blue line if $n - 1 - 2x \geq 0$ and $v_1 - y \geq 0$, i.e.

$$x \leq (n - 1)/2 \ \wedge \ v_1 \geq y. \tag{34}$$

Then according to inequalities (33) and (34) the line is blue (instead of black) when n is odd and $x = (n - 1)/2$ and $v_1 \geq y$.

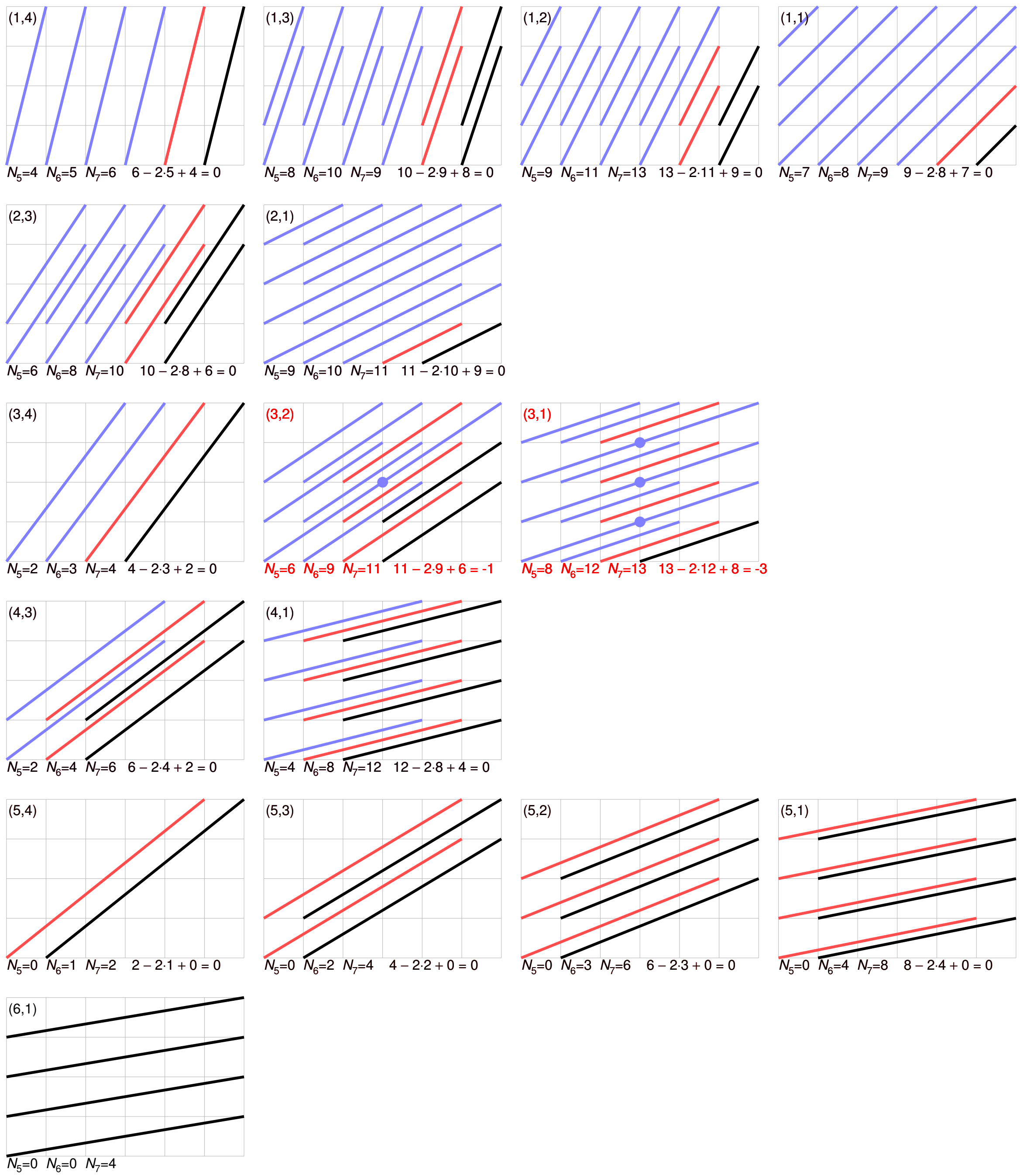


FIGURE 3. Ascending lines in case $m = 5, n = 7$

Then the characteristic $D_n = N_n - 2N_{n-1} + N_{n-2}$ is typically zero and can be non-zero (negative) only when n is odd and the slope of the line is $((n-1)/2, y),\ ((n-1)/2, y) = 1,\ y = 1, 2, \ldots, m-1$.

According to (11) and (12) $N_n = \psi_2(m, n, x, y)$ for a general slope $(x, y),\ x, y > 0$ is

$$\psi_2(m, n, x, y) = (n-x)(m-y) - p(n-2x)p(m-2y), \tag{35}$$

where $p(z) = z$, if $z > 0$ and $p(z) = 0$, if $z \le 0$ and the characteristic $D_n = \psi_3(m, n, x, y)$ is

$$\psi_3(m, n, x, y) = \psi_2(m, n, x, y) - 2\psi_2(m, n-1, x, y) + \psi_2(m, n-2, x, y). \tag{36}$$

Then the residuals (32) are $R_1(m, n) = 2\psi_4(m, n)$ where

$$\psi_4(m, n) = \sum_{\substack{y=1 \\ (\frac{n-1}{2}, y)=1}}^{m-1} \psi_3(m, n, \frac{n-1}{2}, y) \tag{37}$$

when n is odd and $\psi_4(m, n) = 0$ when n is even.

The general formula for $L(m, n)$ is

$$L(m, n) = 2L(m, n-1) - L(m, n-2) + 2\psi_1(m, n) + 2\psi_4(m, n). \tag{38}$$

with initial values $L(m, 2) = L(2, m)$ and $L(m, 3) = L(3, m)$ obtained from (26) and (28).

Although the formula (38) is seemingly more complicated than (7), it is much faster in calculations. For example, if all values of $L(m, n)$ for $m, n = 2, 3, \ldots, 300$ are computed, it is more than 300 times faster. It is also faster when only one value of $L(m, n)$ is needed although it computes also values $L(m, i),\ i = 2, 3, \ldots, n-1$ before getting $L(m, n)$.

This result is applied in case $m = 11$ by means of SURVO MM as follows:

```
____________________________________________________________
*psi1(M,N):=for(y=1)to(M-1)sum(psi12(M,N,y))
*psi12(M,N,y):=if(gcd(y,N-1)=1)then(M-y)else(0)
*
*p(x):=if(x>0)then(x)else(0)
*psi2(M,N,x,y):=(N-x)*(M-y)-p(N-2*x)*p(M-2*y)
*psi3(M,N,x,y):=psi2(M,N,x,y)-2*psi2(M,N-1,x,y)+psi2(M,N-2,x,y)
*psi4(M,N):=if(mod(N,2)=0)then(0)else(psi42(M,N))
*psi42(M,N):=for(y=1)to(M-1)sum(psi43(M,N,y))
*psi43(M,N,y):=if(gcd((N-1)/2,y)=1)then(psi3(M,N,(N-1)/2,y))else(0)
*
*VAR PSI1=psi1(m,n) TO L11N
*VAR R11=L-2*L[-1]+L[-2]-2*PSI1 TO L11N
*
*VAR PSI4=psi4(m,n) TO L11N
*
*VAR RES=L-2*L[-1]+L[-2]-2*PSI1-2*PSI4 TO L11N
*..........................
*VAR Lcheck=2*Lcheck[-1]-Lcheck[-2]+2*PSI1+2*PSI4 TO L11N / IND=n,4,100
*
*DATA L11N,A,B,N,M
N m    n         L  PSI1   R11   PSI4   RES    Lcheck
M11 111    111111  1111  1111   1111  11111   111111
A11   2       123    55     -      0      -      123
*11   3       244    30     -    -25      -      244
```

124	*11	4	445	40	0	0	0	445
125	*11	5	676	30	-30	-15	0	676
126	*11	6	1003	48	0	0	0	1003
127	*11	7	1330	20	-40	-20	0	1330
128	*11	8	1759	51	0	0	0	1759
129	*11	9	2218	30	-30	-15	0	2218
130	*11	10	2757	40	0	0	0	2757
131	*11	11	3296	24	-48	-24	0	3296
132	*11	12	3945	55	0	0	0	3945
133	*11	13	4614	20	-20	-10	0	4614
134	*11	14	5393	55	0	0	0	5393
135	*11	15	6174	26	-50	-25	0	6174
136	*11	16	7021	33	0	0	0	7021
137	*11	17	7898	30	-30	-15	0	7898
138	*11	18	8885	55	0	0	0	8885
139	*11	19	9872	20	-40	-20	0	9872
140	*11	20	10969	55	0	0	0	10969
141	*11	21	12086	24	-28	-14	0	12086
142	*11	22	13275	36	0	0	0	13275
143	*11	23	14474	30	-50	-25	0	14474
144	*11	24	15783	55	0	0	0	15783
145	*11	25	17112	20	-20	-10	0	17112
146	*11	26	18537	48	0	0	0	18537
147	*11	27	19972	30	-50	-25	0	19972
148	*11	28	21487	40	0	0	0	21487
149	*11	29	23024	26	-30	-15	0	23024
150	*11	30	24671	55	0	0	0	24671
...	..	..	...	..	..	..	.	......
211	*11	91	226848	14	-38	-19	0	226848
212	*11	92	231879	51	0	0	0	231879
213	*11	93	236940	30	-30	-15	0	236940
214	*11	94	242081	40	0	0	0	242081
215	*11	95	247232	30	-50	-25	0	247232
216	*11	96	252479	48	0	0	0	252479
217	*11	97	257746	20	-20	-10	0	257746
218	*11	98	263123	55	0	0	0	263123
219	*11	99	268502	26	-50	-25	0	268502
220	B11	100	273961	40	0	0	0	273961

Values of column `L` have been computed by the Survo command `LMN` and the remaining columns by the formulas (31), (35) – (37).

The corresponding Mathematica code with results is

```
m=11;
L[2]=m^2+2;
L[3]=2*m^2+3-Mod[m,2];
psi1[m_,n_]:=Sum[psi2[m,n,y], y,1,m-1]
psi2[m_,n_,y_]:=If[GCD[y,n-1]==1,m-y,0]
p[i_]:=If[i>0,i,0]
psi2[m_,n_,x_,y_]:=psi2[m,n,x,y]=(n-x)*(m-y)-p[n-2*x]*p[m-2*y]
psi3[m_,n_,x_,y_]:=psi2[m,n,x,y]-2*psi2[m,n-1,x,y]+psi2[m,n-2,x,y]
psi4[m_,n_]:=psi4[m,n]=If[Mod[n,2]==0,0,psi42[m,n]]
psi42[m_,n_]:=psi42[m,n]=Sum[psi43[m,n,y], y,1,m-1]
psi43[m_,n_,y_]:=If[GCD[(n-1)/2,y]==1,psi3[m,n,(n-1)/2,y],0]
L[n_]:=L[n]=2*L[n-1]-L[n-2]+2*psi1[m,n]+2*psi4[m,n]
Table[L[n],n,2,100]

123, 244, 445, 676, 1003, 1330, 1759, 2218, 2757, 3296, 3945, 4614,
5393, 6174, 7021, 7898, 8885, 9872, 10969, 12086, 13275, 14474, 15783,
17112, 18537, 19972, 21487, 23024, 24671, 26308, 28055, 29832, 31689,
33556, 35511, 37486, 39571, 41666, 43841, 46036, 48341, 50638, 53045,
55482, 57985, 60498, 63121, 65764, 68509, 71254, 74079, 76934, 79899,
82864, 85925, 89008, 92171, 95344, 98627, 101920, 105323, 108736,
112221, 115736, 119347, 122958, 126679, 130430, 134261, 138084, 142017,
145970, 150033, 154106, 158245, 162414, 166685, 170956, 175337, 179738,
184219, 188710, 193311, 197924, 202633, 207352, 212151, 216980, 221919,
226848, 231879, 236940, 242081, 247232, 252479, 257746, 263123, 268502,
273961
```

8. Appendix 3: Accuracy of asymptotic expression for $L(n)$

In Section 4 it was conjuctured (20) that $L(n) = [3/(2\pi)n^2]^2 + O(n^{2.5})$ according to numerical experiments. Now after my experimental results it has been proved in [1] a weaker result that $L(n) = [3/(2\pi)n^2]^2 + O(n^3 \log n)$. [4] [5]

That result is based on decompositions of (3) for $k = 1, 2$ and, for example,

$$f(n+1,1) = 4(n+1)^2 s_1(n) - 8(n+1)s_2(n) + 4s_3(n) + 4(n+1)n \tag{39}$$

where

$$s_1(n) = \sum_{\substack{x,y=1\\(x,y)=1}} 1, \qquad s_2(n) = \sum_{\substack{x,y=1\\(x,y)=1}} x, \qquad s_3(n) = \sum_{\substack{x,y=1\\(x,y)=1}} xy. \tag{40}$$

By showing that

$$s_i(n) = \frac{n^{i+1}}{2^{i-1}\zeta(2)} + O(n^i \log n), \quad i = 1,2,3, \quad \zeta(2) = \pi^2/6 \tag{41}$$

it is concluded that

$$f(n+1,1) = \frac{n^4}{\zeta(2)} + O(n^3 \log n). \tag{42}$$

A similar argument leads to

$$f(n+1,2) = \frac{n^4}{4\zeta(2)} + O(n^3 \log n) \tag{43}$$

and according to (2)

$$L(n) = \frac{3n^4}{8\zeta(2)} + O(n^3 \log n) = [3/(2\pi)n^2]^2 + O(n^3 \log n). \tag{44}$$

However, although the asymptotic expressions (41) may be the best possible, (42) for $f(n+1,1)$ is not, since, according to (39), it is the difference

$$f(n+1,1) = g_1(n) - g_2(n) \tag{45}$$

of two 'large' positive expressions

$$\begin{aligned} g_1(n) &= 4(n+1)^2 s_1(n) + 4s_3(n) + 4(n+1)n, \\ g_2(n) &= 8(n+1)s_2(n) \end{aligned} \tag{46}$$

and these quantities are strongly related to each other. For $n = 2, 3, \ldots, 10000$ their correlation coefficient is 0.9999999998.

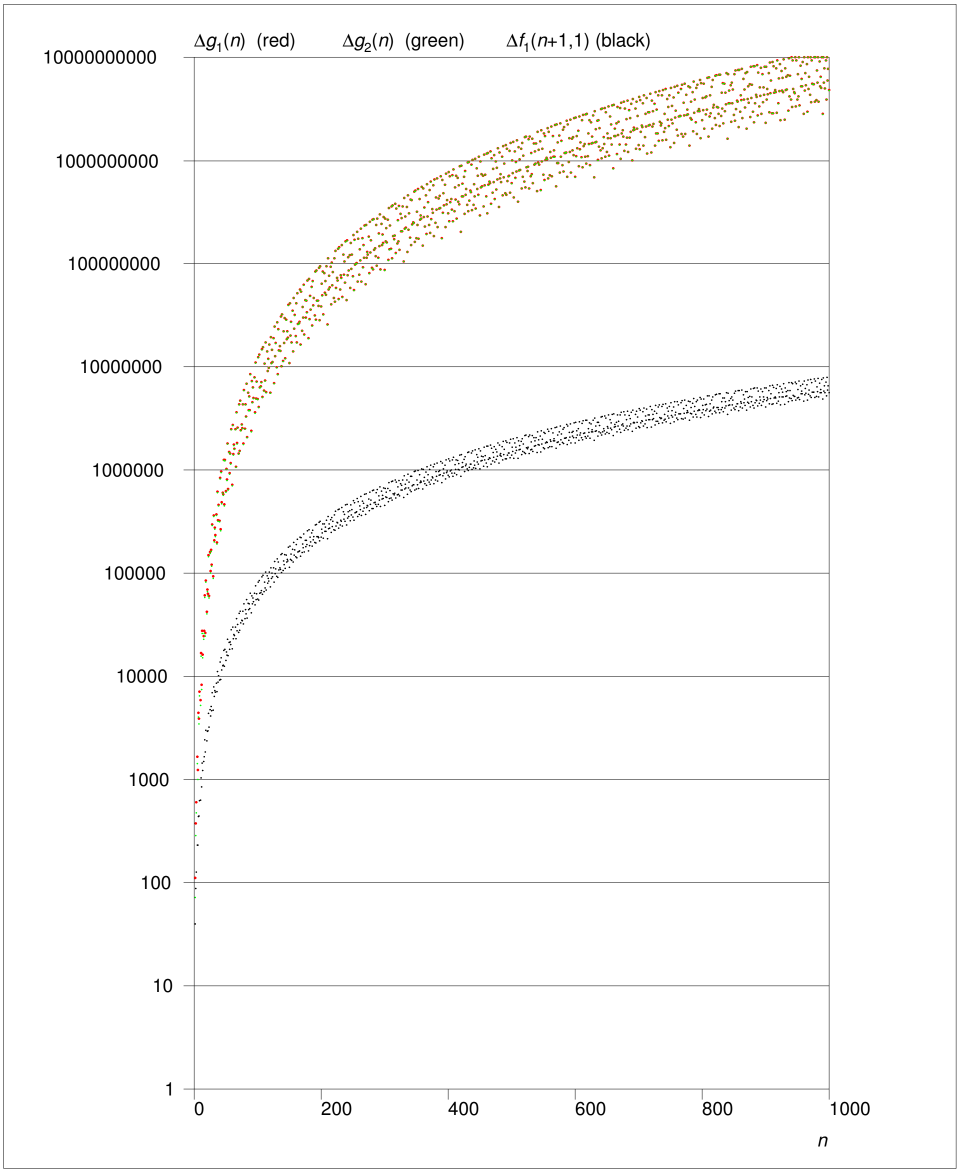
In the next graph, the differences $\Delta f(n,1) = f(n+1,1) - f(n,1)$, $\Delta g_1(n) = g_1(n+1) - g_1(n)$, and $\Delta g_2(n) = g_2(n+1) - g_2(n)$ are plotted for $n = 2, 3, \ldots, 1000$ on a logarithmic scale

---

Date 14 June 2009

[4] Also formulas (21) are proved in [1].

[5] On the basis of results in ([2]), Kaisa Matomäki has now shown that the residual is $O(n^c)$ for any constant $c > 2.5$ if the famous Riemann hypothesis is true. (A personal communication, 2 Aug 2009). Thus my conjecture seems to be essentially correct.



showing how the increments of functions (46) fluctuate almost identically (depending on the divisibility of n) and the variations in increments of $f(n+1,1)$ are of minor magnitude.

In fact, numerical studies indicate that the magnitude of the residual in the asymptotic expression of $f(n,1)$ is $O(n^{2.5})$, i.e. the same as that of $L(n)$.

Since the formula (3) is slow in computations of $f(n,1)$ for large n, I made a similar numerical experiment with Survo as in Appendix 1 (Section 6) and found efficient recursive formulas for $f(n,1) = f(n,n,1)$ and $f(n-1,n,1)$ defined generally in (8) as

$$(47) \qquad \begin{aligned} f(n,n,1) &= 2f(n-1,n,1) - f(n-1,n-1,1) + R_1(n), \\ f(n-1,n,1) &= 2f(n-1,n-1,1) - f(n-2,n-1) + 2(n-1)\phi(n-1), \end{aligned}$$

where

$$R_1(n) = R_1(n-1) + 8\phi(n-1) \tag{48}$$

with initial values $f(0,0,1) = f(0,1,1) = R_1(1) = 0$.

By combining capabilities of Survo and Mathematica, I have computed values of $f(n,1) = f(n,n,1)$ for $n = 2,3,\ldots,10^7$ and found that $|f(n,1) - n^4/\zeta(2)| < 0.22n^{2.5}$ in the interval $(10^6, 10^7)$. In particular,
$f(10^7,1) = 6079271018567762240876092228$ and $f(10^7,1)/(10^7)^4 - 6/\pi^2 \approx 2.7 \cdot 10^{-12}$.

8.1. **Computational details.** I have now computed $L(n)$ values and the proportional deviances $D(n) = (L(n) - Cn^4)/n^{2.5}$ for $n = 2,3,\ldots,10^8$ and the deviances $D(n)$ (100 million dots) are plotted in
`http://www.survo.fi/papers/DevLn2009A.pdf`

The deviances stay within the earlier limits, i.e. $|D(n)| < 0.2$, and give no reason to reject the conjecture (20).

The entire computational process was controlled from Survo. The Mathematica code needed for computing $L(n)$ in portions of a million values was given in a Survo edit field as follows:

```
____________________________________________________________
101 *SAVEP CUR+1,E,INIT.TXT
102 *res = 0, 1000000, 0, 0, 0, 0
103 *res >> resfile
104 E
105 *
106 */MATHRUN INIT.TXT
107 *Out[2]= 0, 1000000, 0, 0, 0, 0
108 *
109 *SAVEP CUR+1,E,LNN.TXT
110 *t1=TimeUsed[];
111 *res = << resfile;
112 *n1=res[[1]];
113 *n2=n1+res[[2]];
114 *L[n1-1]=res[[3]];
115 *L1[n1]=res[[4]];
116 *R1[n1]=res[[5]];
117 *n1=n1+1;
118 *L[n_]:=L[n]=2*L1[n]-L[n-1]+R1[n]
119 *L1[n_]:=L1[n]=2*L[n-1]-L1[n-1]+R2[n]
120 *R1[n_]:=R1[n]=R1[n-1]+4*(EulerPhi[n-1]-e[n])
121 *e[n_]:=If[Mod[n,2]==0,0,EulerPhi[(n-1)/2]]
122 *R2[n_]:=
123 *If[Mod[n,2]==0,(n-1)*EulerPhi[n-1],
124 *If[Mod[n,4]==1,(n-1)*EulerPhi[n-1]/2,0]]
125 *Export["J:/LNN/LN.TXT",Table[n,L[n],n,n1,n2],"Table"]
126 *res=n2,n2-n1+1,L[n2-1],L1[n2],R1[n2],L[n2]
127 *res >> resfile
128 *TimeUsed[]-t1
129 E
____________________________________________________________
```

The $L(n)$ computations were carried out by the code `LNN.TXT` on lines 110 – 128. However, at first the initial values were saved by the code on lines 102 – 103 in `resfile` so that the main code could start properly. These initial values were then updated after each run of the main code. Without this kind of partitioning the process would become almost impossible due to shortage of memory.

The `/MATHRUN` command (Survo macro) on the line 106 just called Mathematica to run the initialization code silently.

The same `/MATHRUN` command was then activated repeatedly to call the main code as follows:

```
____________________________________________________________
131 */MATHRUN LNN.TXT
132 *Out[15]= J:/LNN/LN.TXT
133 *Out[16]= 1000000, 1000000, 227971751347400065430960,
134 *>    227972207291464138543601, 911889250864, 227972663236440100907106
135 *Out[18]= 111.041
136 *
137 *>COPY LNN.TXT+CRLF.TXT LNN1.TXT
138 *
139 */MATHRUN LNN.TXT
140 *Out[15]= J:/LNN/LN.TXT
141 *Out[16]= 2000000, 1000000, 3647555315673185187934080,
142 *>    3647558963232336748458307, 3647562987544, 3647562610795135871970078
143 *Out[18]= 113.833
144 *
145 *>COPY LNN.TXT+CRLF.TXT LNN2.TXT
146 *
147 */MATHRUN LNN.TXT
148 *Out[15]= J:/LNN/LN.TXT
149 *Out[16]= 3000000, 1000000, 18465761097482114973732680,
150 *>    18465773407997468652953031, 8207015539424, 18465785718521029347712806
151 *Out[18]= 115.237
152 *
153 *>COPY LNN.TXT+CRLF.TXT LNN3.TXT
154 *
...
821 */MATHRUN LNN.TXT
822 *Out[15]= J:/LNN/LN.TXT
823 *Out[16]= 100000000, 1000000, 22797265407630329509904737831840,
824 *>    22797265863575644164147398533447, 9118906423118616,
825 *>    22797266319520967937296482353670
826 *Out[18]= 132.82
827 *
828 *>COPY LNN.TXT+CRLF.TXT LNN100.TXT
829 *
830 *SAVEP CUR+1,E,START100.TXT
831 *res= 100000000, 1000000, 22797265407630329509904737831840, \
832 *     22797265863575644164147398533447, 9118906423118616, \
833 *     22797266319520967937296482353670
834 *res >> resfile
835 E
____________________________________________________________
```

The first `/MATHRUN` command on line 131 gave results on lines 132 – 135. The $L(n)$ values, $n = 1, 2, \ldots, 10^6$ were saved in a text file `LN.TXT`. The new initial values are displayed on lines 133 – 134 (saved for the next round internally by Mathematica in `resfile`) and the run took about 111 seconds (line 135).

In subsequent runs, the time seems to increase very slowly which is an indication of the efficiency of the recursive formulas (21) when used in this partitioned way.

The file `LN.TXT` was renamed as `LNN1.TXT` and line end characters were appended at the end of the last line (which for some odd reason is not done by Mathematica although those characters appear at the end of all preceding 999999 lines).

The next activation of `/MATHRUN LNN.TXT` (line 139) produces results until $n = 2 \cdot 10^6$.

After 100 repetitions of this process the results appearing on lines 822 – 826 are obtained and thus the first 100 million values of $L(n)$ have been gathered. [6]

If the process would be continued later, suitable initial values taken from the last results can be saved by first activating `/SAVEP` on line 830 and then `/MATHRUN START100.TXT` (even without knowing anything more about previous results).

The remaining steps for producing the proportional deviances and a graph of them are carried out by Survo in the following manner:

```
____________________________________________________________
101 *FILE CREATE LNN1
102 *FIELDS:
103 *1 N 8   n        (#########)
104 *2 S 32  L        (################################)
105 *3 N 8   D1       (##.##########)
106 *END
107 *
108 *.........................
109 *FILE SAVE LNN1.TXT,LNN1
110 *FIELDS:
111 *1 n  [9]
112 *2 L
113 *END
114 *
115 *pi=3.141592653589793 C=(3/(2*pi))^2
116 *VAR D1=(L-C*n^4)/n^2/sqrt(n) TO LNN1
117 *
118 *..........
119 *FILE LOAD LNN1 / IND=ORDER,1000000
120 *DATA LNN1*,A,B,C
121 C         n L                                         D1
122 B   1000000 227972663236440100907106         0.0411801129
123 *
124 *....................
125 *SIZE=180,250 XDIV=0,1,0 YDIV=0,1,0 HEADER= XLABEL= YLABEL=  D1=-0.2,0.2
126 *FRAME=1 YSCALE=-0.2:,-0.1:,0:,0.1:,0.2: GRID=Y LINE=1 TICKLENGTH=0
127 *HOME=100,100 PEN=[Swiss(7)] TEXTS=T T=1,160,5
128 *
129 *PLOT LNN1,n,D1 / DEVICE=PS,GRAPH1.PS  n=1,1000000 XSCALE=1:,1000000:
____________________________________________________________
```

A Survo data file `LNN1` is created (lines 101 – 106) and results from text file `LNN1.TXT` are saved in it (lines 109 – 113). The proportional deviances `D1` are computed by the `VAR` command (line 116) and the last case $n = 10^6$ is checked (lines 119 – 122). Finally, the first partial graph as a PostScript file `GRAPH1.PS` is drawn by a `PLOT` scheme defined on lines 125 – 129.

The same setup is copied to produce all 100 partial grahs and these graphs are combined to a final one by using the Survo command `EPS JOIN` stepwise.

---

[6] For $n = 10^8$, $L(n) = 227972663195209679372964823536 70$ and $L(n)/n^4 - C \approx -5 \cdot 10^{-14}$.

I have continued (in the beginning of 2015) the empirical study of the asymptotic behaviour by computing consecutive $L(n)$ values to $n = 10^{11}$ and the proportional deviances in this 1000-fold data are presented in the graph
`http://www.survo.fi/papers/DevLn2015.pdf`
still showing no violation of my conjecture (20).
The calculations were carried out in the same way as earlier. Using a recursive formula

$$L(n,n) = L(n-1,n-1) + 2\sum_{i=1}^{n}(R_1(i) + R_2(i)) - R_2(n), \quad n \geq 2 \tag{49}$$

derived as Theorem 2 in [1] from the original formulas (21), speeded up computations to some extent since updating of $L(n-1,n)$ values was avoided.

References

[1] A-M.Ernvall-Hytönen, K.Matomäki, P.Haukkanen, J.K.Merikoski. Formulas for the number of gridlines, *Monatsh. Math.*, 164:157 – 170 (2011).

[2] D.Suryanarayana. On the average order of the function $E(x) = \sum_{n \leq x} \phi(n) - 3x^2/\pi^2$ (II). *Journal of the Indian Mathematical Society*, 42:179 – 195 (1978).

The current version of this paper can be downloaded from
`http://www.survo.fi/papers/PointsInGrid.pdf`

Department of Mathematics and Statistics, University of Helsinki
E-mail address: seppo.mustonen@helsinki.fi