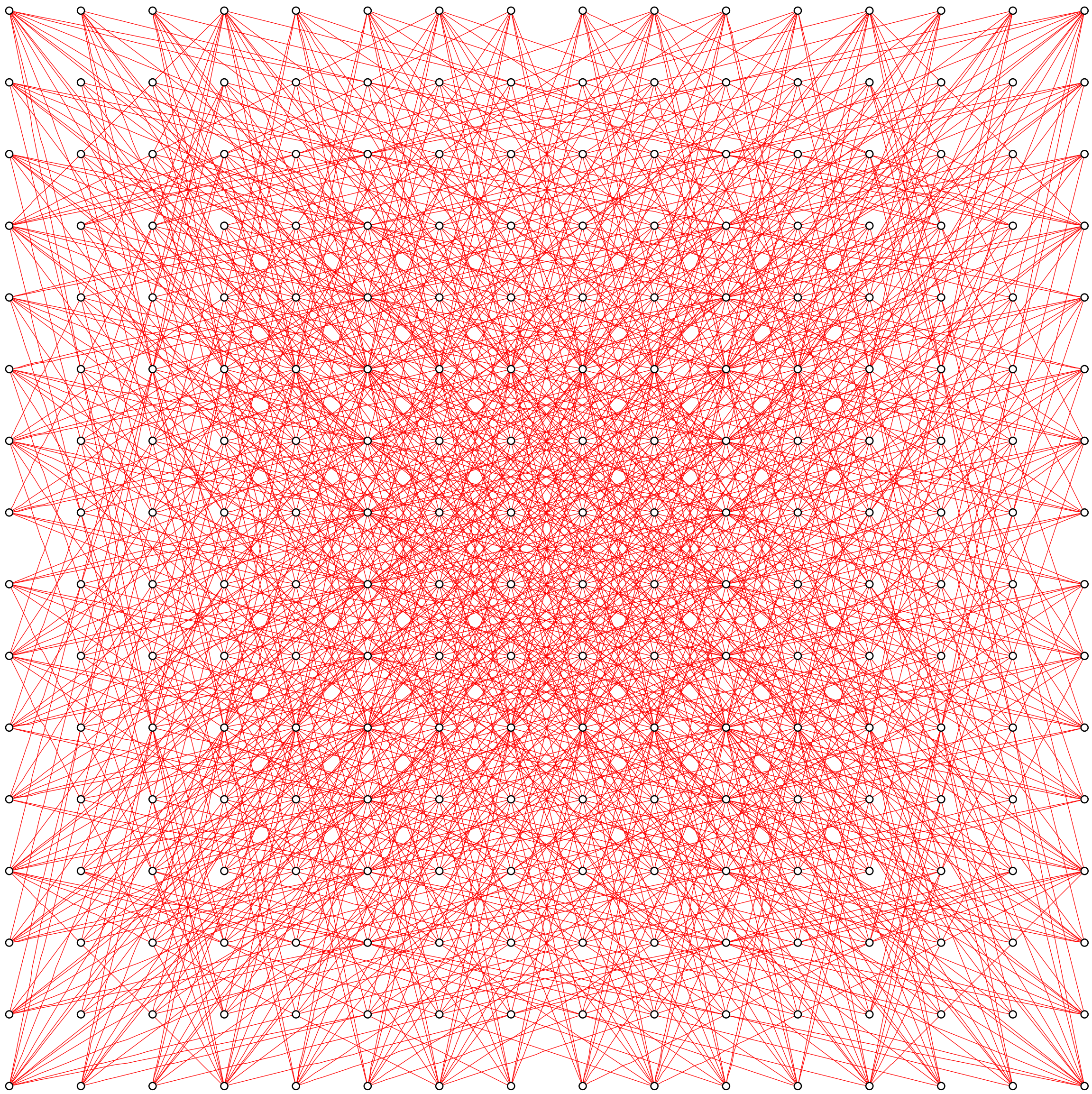
ON LINES THROUGH A GIVEN NUMBER OF POINTS IN A RECTANGULAR GRID OF POINTS

SEPPO MUSTONEN



All $L_4(16, 16) = 548$ lines going through exactly 4 points in a 16×16 grid of points are drawn in red. Each line is represented by a line segment connecting just 4 points.

Date: 12 May 2010.

The studies presented in [3] are continued by examining the number of lines going exactly through j points in an $m \times n$ rectangular grid of points. This number is denoted $L_j(m,n)$ and it has been proved in [3] (p.4) that, by denoting $f(m,n,j) = F_j(m,n)$,

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

where

$$F_j(m,n) = \sum_{\substack{-n<x<n \\ -m<y<m \\ (x,y)=j}} (n-|x|)(m-|y|). \tag{2}$$

The formula (1) is slow in computations when using (2). As in case of the $L(n,n)$ numbers, [1] a much faster recursive formula is available but now for the $F_j(n,n) = F_j(n)$ numbers. It reads

$$\begin{aligned} F_j(n,n) &= 2F_j(n-1,n) - F_j(n-1,n-1) + R_{j1}(n), \\ F_j(n-1,n) &= 2F_j(n-1,n-1) - F_j(n-2,n-1) + R_{j2}(n) \end{aligned} \tag{3}$$

where

$$R_{j1}(n) = R_{j1}(n-1) + 8S(n,j), \tag{4}$$

$$S(n,j) = \begin{cases} \phi((n-1)/j) & \text{if } j \mid n-1; \\ 0 & \text{if } j \nmid n-1 \end{cases} \tag{5}$$

and

$$R_{j2}(n) = 2(n-1)S(n,j) \tag{6}$$

with initial values

$$F(n,n) = F(n-1,n) = R_{j1}(n) = 0, \quad n \le j. \tag{7}$$

This set of formulas was found by a numerical experiment using tools of the Survo system. Let's study the case $j = 3$ as an example. The following setup taken directly from a Survo edit field illustrates how the recursive formula for $F_3(n,n)$ becomes exposed.

[1] The recursive formula for $L(n,n) = \sum_{j=2}^{n} L_j(n,n)$ was found in [3] and it has been proved in [2].

```
------------------------------------------------------------
*DATA F3NN,A,B,N,M
M  11   11111   11111   11111   12  12   12
N   n    F3nn   F3n1n  F3n1n1  Res   d  phi
A   2       0       0       -    -   -    0
*   3       0       0       0    0   -    0
*   4      20       6       0    1   1    1
*   5      56      34      20    1   0    0
*   6     108      78      56    1   0    0
*   7     208     150     108    2   1    1
*   8     340     266     208    2   0    0
*   9     504     414     340    2   0    0
*  10     788     630     504    4   2    2
*  11    1136     946     788    4   0    0
*  12    1548    1326    1136    4   0    0
*  13    2136    1818    1548    6   2    2
*  14    2820    2454    2136    6   0    0
*  15    3600    3186    2820    6   0    0
*  16    4748    4134    3600   10   4    4
*  17    6056    5362    4748   10   0    0
*  18    7524    6750    6056   10   0    0
*  19    9312    8370    7524   12   2    2
*  20   11292   10254    9312   12   0    0
*  21   13464   12330   11292   12   0    0
*  22   16380   14850   13464   18   6    6
*  23   19584   17910   16380   18   0    0
*  24   23076   21258   19584   18   0    0
*  25   27272   25086   23076   22   4    4
*  26   31820   29458   27272   22   0    0
*  27   36720   34182   31820   22   0    0
*  28   42668   39582   36720   28   6    6
*  29   49064   45754   42668   28   0    0
B  30   55908   52374   49064   28   0    0
*
*VAR Res=(F3nn-2*F3n1n+F3n1n1)/8 TO F3NN
*VAR d=Res-Res[-1] TO F3NN
*VAR phi=if(mod(n,3)=1)then(totient((n-1)/3))else(0) TO F3NN
------------------------------------------------------------
```

At first it was detected by linear regression that $F_3(n,n) \approx 2F_3(n-1,n) - F_3(n-1,n-1)$. The columns `F3nn`, `F3n1n`, `F3n1n1` corresponding to $F_3(n,n)$, $F_3(n-1,n)$, and $F_3(n-1,n-1)$, respectively were computed by (2). Actually the calculations were made with larger data sets in order to ensure the results. Since it was observed that all residuals are divisible by 8, they were divided by this number. When seeing them it was natural to take differences `d`. Thereafter it could be seen that `d` values are equal to $\phi((n-1)/3)$ when $n \equiv 1 \pmod 3$ and zero otherwise.

By reversing these calculations, it is easy to confirm the validity of the recursion formula (3) for $F_3(n,n)$. The same structure was found also for j values 2 and 4. The special case is $j = 1$ where the structure prevails unconditionally.

The second part of recursion formulas (3) was established for $j = 3$ by a similar Survo application:

```
____________________________________________________________
*DATA F3N1N,a,b,n,m
m  11   11111   11111   11111  111   111
n   n   F3n1n  F3n1n1  F3n2n1  Res  phi2
a   2       0       -       -    -     0
*   3       0       0       0    0     0
*   4       6       0       0    3     3
*   5      34      20       6    0     0
*   6      78      56      34    0     0
*   7     150     108      78    6     6
*   8     266     208     150    0     0
*   9     414     340     266    0     0
*  10     630     504     414   18    18
*  11     946     788     630    0     0
*  12    1326    1136     946    0     0
*  13    1818    1548    1326   24    24
*  14    2454    2136    1818    0     0
*  15    3186    2820    2454    0     0
*  16    4134    3600    3186   60    60
*  17    5362    4748    4134    0     0
*  18    6750    6056    5362    0     0
*  19    8370    7524    6750   36    36
*  20   10254    9312    8370    0     0
*  21   12330   11292   10254    0     0
*  22   14850   13464   12330  126   126
*  23   17910   16380   14850    0     0
*  24   21258   19584   17910    0     0
*  25   25086   23076   21258   96    96
*  26   29458   27272   25086    0     0
*  27   34182   31820   29458    0     0
*  28   39582   36720   34182  162   162
*  29   45754   42668   39582    0     0
b  30   52374   49064   45754    0     0
*
*VAR Res=(F3n1n-2*F3n1n1+F3n2n1)/2 TO F3N1N
*VAR phi2=if(mod(n,3)=1)then((n-1)*totient((n-1)/3))else(0) TO F3N1N
____________________________________________________________
```

The validity of recursive formulas was confirmed systematically for $n \leq 1000$, $j = 1, 2, 3, 4$ and separately for certain greater values. For example, it was found that $L_3(60000) = 2407391284632795940$ also when the corresponding F_3 values were computed recursively.

By computing values of $L_j(n)$ for huge n values through recursively computed $F_j(n)$ values gives a possibility to study their asymptotic behaviour.

On the basis of the formulas (3) – (7) the following Mathematica code was created:

```
nn=10^6;
j=1;
F[n_]:=F[n]=If[n<=j,0,2*F1[n]-F[n-1]+R1[n]]
F1[n_]:=F1[n]=If[n<=j,0,2*F[n-1]-F1[n-1]+R2[n]]
R1[n_]:=R1[n]=If[n<=j,0,R1[n-1]+8*S[n]]
R2[n_]:=2*(n-1)*S[n]
S[n_]:=If[Mod[n-1,j]==0,EulerPhi[(n-1)/j],0]
Table[F[n],n,1,nn];
F[nn]
```

By changing `nn` and `j` the value of any $F_j(n)$ can be calculated. Thus for $n = 10^6$ the following values were obtained:
F_1: 607927101897802895986964
F_2: 151981775424922694172752
F_3: 67547455720880127685124
giving by (3) $L_2(10^6) = 185755503384418817663292$.

Since in [3] it was found [2] that $L(n)$ is asymptotically equal to $[3/(2\pi)]^2n^4$ or $(c/\pi^2)n^4$ where $c = 9/4$, it is natural to expect that also $L_j(n)$ is asymptotically of the form $(c_j/\pi^2)n^4$ with decreasing rational constants c_j. For $n = 10^6$ an approximation of c_2 will be

$$c_2' = 185755503384418817663292/n^4\pi^2 = 1.8333333337294289$$

and gives good reasons to assume that $c_2 = 11/6$.

For detecting proper values of c_j for greater j it is best to study the asymptotic behaviour of $F_j(n)$. It turns out that the corresponding coefficients, say d_j, for them, according to similar computations, are

j	1	2	3	4	5	6	7	8
d_j	6	3/2	2/3	3/8	6/25	1/6	6/49	6/64

Then it easy to see that each d_j multiplied by j^2 gives a constant 6. Thus it is evident that in general we have $d_j = 6/j^2$ and $F_j(n) = [6/(j\pi)^2]n^4$ asymptotically. It is also plausible that the error term is of the order $O(n^{2.5})$ as it is for $L(n)$.

Thus my conjecture is

$$F_j(n) = [6/(j\pi)^2]n^4 + O(n^{2.5}). \tag{8}$$

[2]My conjecture, based on numerical experiments, was $L(n) = [3/(2\pi)]^2n^4 + O(n^{2.5})$ and this has now been proved in [2] on the condition that the Riemann hypothesis is true.

Corresponding asymptotic results for $L_j(n)$ numbers can then be represented simply by using (1) and we have

$$L_j(n) = 3[1/(j+1)^2 - 2/j^2 + 1/(j-1)^2]/\pi^2 n^4 + O(n^{2.5}), \quad j = 2, 3, \dots. \tag{9}$$

Thus asymptotic values proportionally to $L(n)$ are
$p(j) = 3[1/(j+1)^2 - 2/j^2 + 1/(j-1)^2]/(9/4)$ and numerically

j	$p(j)$
2	0.814815
3	0.120370
4	0.034814
5	0.013704
6	0.006470
7	0.003449
8	0.002005
9	0.001245
10	0.000814
11	0.000554
12	0.000390
13	0.000283
14	0.000210
15	0.000159
sum	0.999282

It may seem almost paradoxical that from lines going through at least two points a great majority (81.5 percent) actually go through two points only.

Proof of the recursive formulas (3)

At first two basic formulas are presented. We have

$$\sum_{\substack{i=1\\(i,n)=j}}^{n} 1 = S(n+1, j) \tag{10}$$

where $S(n+1, j)$ is defined according to (5). If $j \nmid n$, this sum is zero. If $j \mid n$, by denoting $i' = i/j, n' = n/j$, we have

$$\sum_{\substack{i=1\\(i,n)=j}}^{n} 1 = \sum_{\substack{i'=1\\(i',n')=1}}^{n'} 1 = \phi(n') = \phi(n/j)$$

and thus (10) has been proved. The second formula is

$$\sum_{\substack{i=1\\(i,n)=j}}^{n} i = \sum_{\substack{i=1\\(i,n)=j}}^{n} (n-i) = \frac{1}{2}S(n+1, j). \tag{11}$$

Again for $j \nmid n$ this sum is zero. If $j \mid n$, by reversing the order of 'applicable' summands we see that the two sums in (11) are equal. By taking their sum we obtain

$$\sum_{\substack{i=1 \\ (i,n)=j}}^{n} i + \sum_{\substack{i=1 \\ (i,n)=j}}^{n} (n-i) = n \sum_{\substack{i=1 \\ (i,n)=j}}^{n} 1 = n\phi(n/j).$$

Then

$$\sum_{\substack{i=1 \\ (i,n)=j}}^{n} i = \frac{1}{2} n\phi(n/j)$$

and also (11) has been proved.

Now to the actual proof of (3):

Using (2) for $R_{j1}(n)$ a typical summand is

$$a_1(n,x,y) = (n-|x|)(n-|y|) - 2(n-1-|x|)(n-|y|) + (n-1-|x|)(n-1-|y|)$$

with following alternatives

$$a_1(n,x,y) = \begin{cases} 1+|x|-|y|, & |x|,|y| = 0,1,\ldots,n-2; \\ n-|y|, & |x| = n-1; \\ |x|-n+2, & |y| = n-1. \end{cases}$$

Non-positive indexes are avoided by writing (2) in the form (taking cases $|x| = j, y = 0$ and $x = 0, |y| = j$ apart)

$$F_j(m,n) = 2[m(n-j) + (m-j)n] + 4 \sum_{\substack{0<x<n \\ 0<y<m \\ (x,y)=j}} (n-x)(m-y). \tag{12}$$

Then [3]

(13)

$$\begin{aligned} R_{j1}(n) &= F_j(n,n) - 2F_j(n-1,n) + F_j(n-1,n-1) \\ &= 4 + 4 \sum_{\substack{0<x<n-1 \\ 0<y<n-1 \\ (x,y)=j}} (1+x-y) + 4 \sum_{\substack{0<x<n \\ (x,n-1)=j}} (x-n+2) + 4 \sum_{\substack{0<y<n \\ (y,n-1)=j}} (n-y) \\ &= 4 + 4 \times [2 \sum_{x=1}^{n-2} \sum_{\substack{y=1 \\ (x,y)=j}}^{x} 1 - 1] + 4 \sum_{\substack{x=1 \\ (x,n-1)=j}}^{n-1} 2 = 4 \times 2 \sum_{x=1}^{n-2} \sum_{\substack{y=1 \\ (x,y)=j}}^{x} 1 + 4 \sum_{\substack{x=1 \\ (x,n-1)=j}}^{n-1} 2 \\ &= 8 \sum_{x=1}^{n-2} S(x+1,j) + 8S(n,j) \quad \text{according to (10)} \\ &= 8 \sum_{x=1}^{n-1} S(x+1,j) = R_{j1}(n-1) + 8S(n,j). \end{aligned}$$

[3] The $2[m(n-j)+(m-j)n]$ terms when computing $R_{j1}(n)$ give 4 and when computing $R_{j2}(n)$ they give 0. Reduction -1 on the third line of this formula is due to the case (j,j) which otherwise would be included twice.

Similarly, by (2) a typical summand for for $R_{j2}(n)$ is
$a_2(n,x,y) = (n-1-|x|)(n-|y|)-2(n-1-|x|)(n-1-|y|)+(n-2-|x|)(n-1-|y|)$
with following alternatives

$$a_2(n,x,y) = \begin{cases} |y|-|x|, & |x|,|y| = 0,1,\dots,n-2; \\ 0, & |x| = n-1; \\ n-1-|x|, & |y| = n-1. \end{cases}$$

Then by using (12)

$$\begin{aligned} R_{j2}(n) &= F_j(n-1,n) - 2F_j(n-1,n-1) + F_j(n-2,n-1) \\ &= 4 \sum_{\substack{0<x<n-1 \\ 0<y<n-1 \\ (x,y)=j}} (y-x) + 4 \sum_{\substack{x=1 \\ (x,n-1)=j}}^{n-1} (n-1-x) \\ &= 0 + 2(n-1)S(n,j) \quad \text{according to (11).} \end{aligned} \tag{14}$$

Also the recursive formulas (21) in [3] can now be proved [4] by formulas (3) here and (7) in [3] by showing that

$$R_1(n) = \frac{1}{2}[R_{11}(n) - R_{21}(n)] \tag{15}$$

and

$$R_2(n) = \frac{1}{2}[R_{12}(n) - R_{22}(n)]. \tag{16}$$

For example, we have
(17)

$$\begin{aligned} L(n,n) - R_1(n) &= 2L(n-1,n) - L(n-1,n-1) \\ &= F_1(n-1,n) - F_2(n-1,n) - \frac{1}{2}F_1(n-1,n-1) + \frac{1}{2}F_2(n-1,n-1) \\ &= \frac{1}{2}[2F_1(n-1,n) - F_1(n-1,n-1)] - \frac{1}{2}[2F_2(n-1,n) - F_2(n-1,n-1)] \\ &= \frac{1}{2}[F_1(n,n) - R_{11}(n)] - \frac{1}{2}[F_2(n,n) - R_{21}(n)] \\ &= \frac{1}{2}[F_1(n,n) - F_2(n,n)] - \frac{1}{2}[R_{11}(n) - R_{21}(n)] \\ &= L(n,n) - \frac{1}{2}[R_{11}(n) - R_{21}(n)] \end{aligned}$$

giving (15).

It is then easy to show that $R_1(n)$ and $R_2(n)$ obtained by formulas (15) and (16) are the same as those given in [3]. In the second case, identities $\phi(4k) = 2\phi(2k)$ and $\phi(4k+2) = \phi(2k+1)$ are needed. They are special cases of the well-known formula $\phi(mn) = \phi(m)\phi(n)(d/\phi(d))$ where $d = \gcd(m,n)$.

[4] A different proof has been presented in [2].

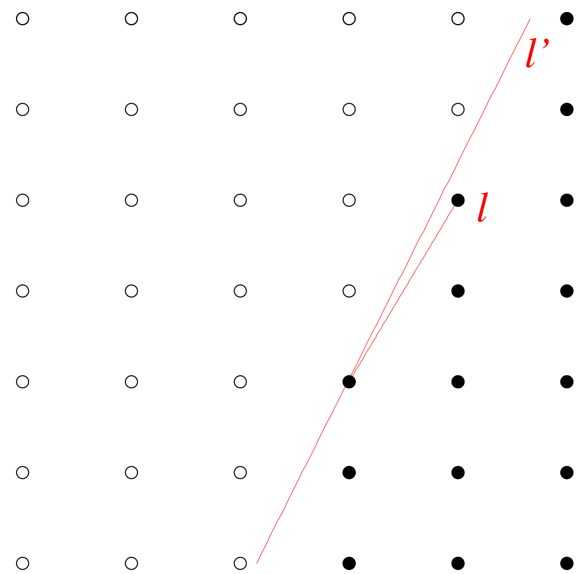
Combinatorial interpretations of the F numbers

According to my considerations in [3] (Section 3) the number of line segments connecting $j+1$ points in an $m \times n$ grid is $F_j(m,n)/2$. Those line segments may be partially overlapped when $j > 1$.

The integer sequence $F_1(n,n)/2, n = 1, 2, \cdots = 0, 6, 28, 86, 200, 418\ldots$ is presented in Sloane's Encyclopedia [4] as a sequence A141255 with a related definition "Total number of line segments between points visible to each other in a square $n \times n$ lattice". [5]

Similarly, $F_j(m,n)/2$ could then be characterized as the total number of line segments between points visible to each other exactly through $j-1$ points in a regular $m \times n$ grid of points.

Also $F_1(m,n)/2$ is the number of ways to divide the points into two nonempty sets using a straight line in an $m \times n$ grid of points. I noticed this relation when seeing in the description of A141255 that A141255(n) =A114043(n) − 1.



In fact, there is an bijection (one-to-one correspondence) between line segments between points visible to each other and such divisions into nonempty subsets in an $m \times n$ grid of points. This bijection is defined by taking any pair of points visible to each other and drawing the line l connecting them. If the slope of the line is non-negative and the direction angle is less than $\pi/2$, a new line l' is made by an infinitesimal rotation of the line l counterclockwise around the leftmost point. The line l' then defines the division of the points to two nonempty subsets corresponding to the selected points so that the only gridpoint on the line l' and points below that line form the first subset. The line l may contain several applicable line segments. The procedure just described attaches each of them to a different division of the points. The line segments on different lines cannot produce any of divisions produced by l. Thus no two line segments cannot determine the same division to subsets. There is one special case. The first horizontal line segment s_1 in the

[5] I have submitted sequences $F_j(n)/2, j = 2, 3, 4, 5, 6, 7, 8, 9$ as A177719 – A177726 to [4] on 13 May 2010.

upper left corner of the grid produces the entire set set of gridpoints. On the other hand, the division d_1 where the other subset consist of the single gridpoint in the lower right corner is generated by no line segment. The required bijection is then completed by letting s_1 correspond to d_1.

The same procedure applies to negative slopes by rotating the entire setup by 90 degrees.

If the number of line segments between points visible to each other is denoted N_1 and the number of divisions by a straight lines by N_2, it thus shows that $N_1 \leq N_2$.

Let l be a line with a non-negative slope and dividing the gridpoints into two nonempty disjoint subsets S_1 and S_2 and let S_2 be the set of gridpoints on l or below it. If no gridpoint locates on l, the line can be moved downwards until it encounters a gridpoint P as a line l'. If P is on the highest level in S_2, l' is rotated counterclockwise around P until it meets another gridpoint. The line thus created still separates subsets S_1 and S_2 and it can go through even more gridpoints. The line segment between the two first of them from left to right is the one corresponding to the division defined by l. The second alternative is that P is on the lowest level in S_2. The the same situation is attained by rotating l clockwise around P. The remaining alternative is that there are gridpoints belonging to S_2 both above P and below P. By a rotation of l' clockwise around P at least one gridpoint above it in S_2 will be met. Then P and the first of those gridpoints define a line segment corresponding to division specified by l. P in this 'middle case' is uniquely determined since if there were another point with a similar property, this would eventually lead to $N_1 > N_2$.

The only exeption is division d_1 (no second point for defining a line segment) which is mapped to s_1.

It is clear that no two divisions cannot correspond to the same line segment.

Since a corresponding construction is possible also for lines with a negative slope, we have $N_2 \leq N_1$. Combining these results leads to $N_1 = N_2$.

The division of a grid by a straight line into two subsets has been studied in [1]. On the basis of the above bijection it is clear that another (maybe simpler?) proof for Theorem 2 in [1] is obtained. It is also evident that $L(m, n)$ corresponds to the number of stable two-dimensional threshold functions and $F_1(m, n)/2 - L(m, n) = F_2(m, n)/2$ to the number of unstable such functions according to (7) in [3].

References

[1] M.Alekseyev, On the number of two-dimensional threshold functions, 2009. `http://arXiv.org/abs/math.CO/0602511`

[2] A-M.Ernvall-Hytönen, K.Matomäki, P.Haukkanen, J.K.Merikoski, Formulas for the number of gridlines, (Unpublished manuscript), 2009.

[3] S.Mustonen, On lines and their intersection points in a rectangular grid of points, 2009. `http://www.survo.fi/papers/PointsInGrid.pdf`

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

The current version of this paper can be downloaded from
http://www.survo.fi/papers/LinesInGrid2.pdf

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