Exact solution of the staircase and row-convex polygon perimeter and area generating function

R Brak† and A J Guttmann†
†Department of Mathematics,
University of Melbourne, Parkville,
Victoria 3052,
Australia

July 9, 2001

ShortTitle: Row-convex and staircase polygons
PACS numbers: 0210,0550

Abstract

An explicit expression is obtained for the perimeter and area generating function $G(y,z) = \sum_{n>=2} \sum_{m>=1} c_{n,m} y^n z^m$, where $c_{n,m}$ is the number of row-convex polygons with area m and perimeter n. A similar expression is obtained for the area-perimeter generating function for staircase polygons. Both expressions contain q-series.

1 Introduction

The solution of the "self-avoiding polygon" problem in greater than one dimension remains elusive, however progress has been made in solving various simpler, but non-trivial, self-avoiding polygon (SAP) problems. In particular there are three classes of polygon problems where exact solutions now exist, these are the staircase polygons, convex polygons and row-convex polygons.

The most relevant quantity of interest is the perimeter generating function

$$G(y) = \sum_{n=1}^{\infty} c_n y^n,$$

where c_n is the number of polygons of the appropriate type which have a perimeter of n steps. If possible, the two variable area and perimeter generating function $G(y, z)$ should be calculated. Here,

$$G(y, z) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_{n,m} y^n z^m,$$

where $c_{n,m}$ is the number of the appropriate polygons with perimeter n and area m.

The polygon generating function is analogous to the free energy of a magnetic spin system. In particular the single variable (perimeter) generating function is analogous to the zero field free energy, whilst the two variable generating function is analogous to the free energy of a spin system in a non-zero external field. The y variable is a temperature-like coordinate, whilst z is an external field-like coordinate. An alternative analogy is to a fluid system where the generating function is analogous to the grand potential and z to the activity. This analogy suggests the polygon system should possess a phase diagram, which is obtained in §4.

A SAP is any self-avoiding walk whose final site is a nearest-neighbour of the initial site, augmented by the bond joining the final site to the initial site. For the square lattice, staircase polygons are a subset of the SAPs whose steps consist only of a sequence of north and east steps followed by a sequence consisting only of south and west steps. The staircase perimeter generating function was first obtained by Temperley (1956) and the perimeter and area generating function was obtained by Pólya (1969).

The SAPs whose perimeter is the same as that of their minimum bounding rectangle are called convex. The convex perimeter generating function was obtained independently by Delest and Viennot (1984), Guttmann and Enting (1988), Lin and Chang (1988) and Kim (1988). Althought the two variable generating function for convex polygons has not been solved, various area weighted moments have been obtained. The r^{th} area-weighted moment is defined by

$$G_r(y) = \frac{1}{r!}\frac{d^r}{dz^r}G(y,z)\,|_{z=1}$$

The first two area-weighted moments were obtained by Enting and Guttmann (1989) and the first, second and third by Lin (1990) who provided a general method for obtaining the r^{th} moment.

Row-convex polygons on the square lattice are SAPs satisfying the following constraint: the number of *vertical* steps in the minimum bounding rectangle has the same number of vertical steps as the polygon. Thus, loosely speaking, the row-convex polygons are convex in the row (horizontal) direction but unconstrained in the vertical direction. The row-convex class of polygons is larger then the staircase and convex polygons but smaller the the SAPs. The first explicit solution of the single variable generating function was obtained by Brak et al (1990).

In this paper we obtain the two variable generating function for row-convex polygons in §3. In §2 we obtain the two variable generating function for the staircase polygons. Although this generating function has already been obtained by Pòlya (1969), the result we obtain is of a very different form to that obtained by Pòlya.

2 Staircase Polygon Generating function

We follow the method of Temperley (1956) and consider a sequence of generating functions. Let h_n be the generating function for staircase polygons, on the square lattice, whose first row contains exactly n squares. Then h_n

satisfies the following equations

$$\begin{aligned}
h_1 =& zy^4 + y^2 z\Big[h_1 + h_2 + h_3 + h_4 + \dots\Big],\\
h_2 =& z^2y^6 + y^2z^2\Big[y^2h_1 + (1+y^2)h_2 + (1+y^2)h_3\\
&+(1+y^2)h_4 + \dots\Big],\\
h_3 =& z^3y^8 + y^2z^3\Big[y^4h_1 + (y^2+y^4)h_2 + (1+y^2+y^4)h_3\\
&+(1+y^2+y^4)h_4 + \dots\Big],\\
h_4 =& z^4y^{10} + y^2z^4\Big[y^6h_1 + (y^4+y^6)h_2 + (y^2+y^4+y^6)h_3\\
&+(1+y^2+y^4+y^6)h_4 + \dots\Big], \qquad (1)
\end{aligned}$$

from which the following recurrence relation may be derived:

$$h_{n+2} - z(1+y^2 - y^2z^{n+1})h_{n+1} + y^2z^2h_n = 0. \qquad (2)$$

This recurrence relation is not a difference equation with constant (ie. n independent) coefficients and thus not amenable to the "text book" methods of solution. We try the following ansatz (Privman and Švrakić 1988);

$$h_n = \lambda^n \sum_{m=0}^{\infty} r_m(z) z^{mn} \qquad (3)$$

where $r_m(z)$ is an arbitrary function of z for $m > 0$ and $r_0 = 1$ (in order to obtain the correct $z \to 0$ limit). Substituting (3) into (2) and rearranging gives

$$\begin{aligned}
&\lambda^2 - z(1+y^2)\lambda + z^2y^2\\
&\quad + \sum_{m=1}^{\infty} [y^2\lambda z^{m+1} r_{m-1} + r_m(\lambda^2 z^{2m} - \lambda z(1+y^2)z^m + z^2y^2)]z^{mn} = 0.
\end{aligned}$$

(4)

We then choose λ and r_m such that

$$\lambda^2 - z(1+y^2)\lambda + z^2y^2 = 0 \qquad (5)$$

and

$$y^2\lambda z^{m+1} r_{m-1} + r_m(\lambda^2 z^{2m} - \lambda z(1+y^2)z^m + z^2y^2) = 0. \qquad (6)$$

Equation (5) has two solutions $\lambda_1 = z$ and $\lambda_2 = zy^2$. As (6) is just a first order difference equation for r_m it is readily solved to give

$$\begin{aligned} r_m &= \prod_{k=1}^{m} \frac{-y^2\lambda z^{k+1}}{\lambda^2 z^{2k} - \lambda z(1+y^2)z^k + y^2z^2} \qquad z \neq 1 \\ &= \frac{(-zy^2\lambda)^m z^{m(m+1)/2}}{\prod_{k=1}^{m}(\lambda z^k - zy^2)(\lambda z^k - z)}. \end{aligned} \tag{7}$$

Thus the general solution is $h_n = A_1 h_n^{(1)} + A_2 h_n^{(2)}$, where

$$h_n^{(i)} = \lambda_i^n + \lambda_i^n \sum_{m=1}^{\infty} \frac{(-zy^2\lambda_i)^m z^{m(m+1+2n)/2}}{\prod_{k=1}^{m}(\lambda_i z^k - zy^2)(\lambda_i z^k - z)} \tag{8}$$

and where A_1 and A_2 are arbitrary functions of x and y determined by the initial conditions.

The generating function $G(y,z)$ is given by

$$G(y,z) = \sum_{n=1}^{\infty} (A_1 h_n^{(1)} + A_2 h_n^{(2)}). \tag{9}$$

In the limit $y \to 0$, $z = 1$ we have $\lambda_1 \sim O(1)$ and $\lambda_2 \sim O(y^2)$ whilst $h_n \sim O(y^{2n+2})$. This implies that $A_1 = 0$. The second constant A_2 is obtained by substituting h_2 and h_1 obtained from (8) into the recurrence relation (2) (with $n = 0$, $h_0 \equiv y^2$) and solving for A_2. If this is done and the result substituted into (9) then, after some rearranging the following result is obtained

$$G(y,z) = \frac{y^2}{\Delta}\Big[\frac{y^2 z}{1-y^2 z} + \sum_{m=1}^{\infty} \frac{y^2 z^{m+1}}{1-y^2 z^{m+1}} R_m z^{m(m+1)/2}\Big] \tag{10}$$

where

$$\Delta = 1 - y^2 z + \sum_{m=1}^{\infty} (1 - y^2 z^m + y^2(1-z)) R_m z^{m(m+3)/2} \tag{11}$$

and

$$R_m = \frac{(-y^2)^m}{\prod_{k=1}^{m}(1-z^k)(1-y^2 z^k)}. \tag{12}$$

Equation (10) has been expanded to order y^{20} using *Mathematica* (Wolfram 1980) and the coefficients of positive powers of z agree with those obtained by expanding the result obtained by Pólya (1969). We note the form of (10) is very different to that given by Pólya's result. His result is also invariant under the transformation $z \to 1/z$ (ie. it generates negative powers of z). It would be of some interest to relate the two results, particularly as the Pólya result contains the more "natural" Gaussian polynomials which generate areas under "zig-zag" paths.

3 Row-convex Polygon Generating function

Let g_n be the generating function for row-convex polygons whose first row contains exactly n squares. Then, as shown by Temperley (1956), g_n satisfies the following equations

$$\begin{aligned}
g_1 =& zy^4 + zy^2\Big[g_1 + 2g_2 + 3g_3 + 4g_4 + \dots\Big], \\
g_2 =& z^2y^6 + z^2y^2\Big[2y^2g_1 + (1+2y^2)g_2 + (2+2y^2)g_3 + (3+2y^2)g_4 + \dots\Big], \\
g_3 =& z^3y^8 + z^3y^2\Big[3y^4g_1 + (2y^2+2y^4)g_2 + (1+2y^2+2y^4)g_3 \\
&+ (2+2y^2+2y^4)g_4 + \dots\Big], \\
g_4 =& z^4y^{10} + z^4y^2\Big[4y^6g_1 + (3y^4+2y^6)g_2 + (2y^2+2y^4+2y^6)g_3 \\
&+ (1+2y^2+2y^4+2y^6)g_4 + \dots\Big] \qquad (13)
\end{aligned}$$

and similarly for $n > 4$. From these it is easily shown that g_n satisfies the recurrence relation,

$$g_{n+4} + p_1 g_{n+3} + (p_2 - p_3 z^{n+2}) g_{n+2} + p_4 g_{n+1} + p_5 g_n = 0 \qquad (14)$$

where

$$\begin{aligned}
p_1 =& -2z(1+y^2), \\
p_2 =& z^2(1+4y^2+y^4), \\
p_3 =& z^2y^2(1-y^2)^2, \\
p_4 =& -2z^3y^2(1+y^2), \\
p_5 =& z^4y^4. \qquad (15)
\end{aligned}$$

The solution to (14) is rather more complex that that for the staircase polygons. To see why consider the following: Consider (14) in the limit

$n \to \infty$ with z *fixed*, then the z^n factor tends to zero (as $|z| < 1$) and hence in this limit, $g_n \to \hat{g}$ and (14) becomes

$$\hat{g}_{n+4} + p_1 \hat{g}_{n+3} + p_2 \hat{g}_{n+2} + p_4 \hat{g}_{n+1} + p_5 \hat{g}_n = 0, \qquad n \to \infty, \tag{18}$$

which has the characteristic equation

$$(\lambda - zy^2)^2 (\lambda - z)^2 = 0. \tag{19}$$

Thus (19), unlike the staircase polygons characteristic equation, has two pairs of *equal* roots and hence the four independent solutions are

$$\begin{aligned} \hat{g}_n =& \lambda^n, \qquad \lambda = zy^2 \quad \text{or} \quad z \\ \hat{g}_n =& n\lambda^n, \qquad \lambda = zy^2 \quad \text{or} \quad z. \end{aligned} \tag{20}$$

This suggests that g_n and ng_n provide the four independent solutions to (14) rather than all four coming from g_n. Before considering all four solutions we solve (14) for g_n. Trying again an ansatz (Privman and Švrakić 1988) of the form

$$g_n = \lambda^n \sum_{m=0}^{\infty} r_m(z) z^{mn} \tag{21}$$

with $r_0 = 1,\ n \geq 1$ gives rise to the characteristic equation (19) and to a first order difference equation for r_m similar to (6). The difference equation has the solution

$$r_m = \frac{\lambda^{2m} (zy)^{2m} (1 - y^2)^{2m} z^{m(m+1)}}{\prod_{k=1}^{m} (\lambda z^k - z)^2 \prod_{k=1}^{m} (\lambda z^k - y^2 z)^2}. \tag{22}$$

Thus for the characteristic value $\lambda_1 = y^2 z$, we have one solution,

$$g_n^{(1)} = (y^2 z)^n + (y^2 z)^n \sum_{m=1}^{\infty} \frac{y^{2m} (1 - y^2)^{2m} z^{m(m+n+1)}}{\prod_{k=1}^{m} (1 - y^2 z^k)^2 \prod_{k=1}^{m} (1 - z^k)^2}. \tag{23}$$

The second solution with $\lambda = z$ we reject on the same grounds that A_1 of (9) was rejected.

We now try to obtain the remaining two independent solutions. Returning to (14) we try a solution of the form

$$g_n^{(2)} = n g_n^{(1)} + q_n \tag{24}$$

which, after substituting into (14), gives

$$\begin{aligned} &q_{n+4} + p_1 q_{n+3} + (p_2 - p_3 z^{n+2}) q_{n+2} + p_4 q_{n+1} + p_5 q_n \\ &+ 4 g_{n+4}^{(1)} + 3 p_1 g_{n+3}^{(1)} + 2 (p_2 - p_3 z^{n+2}) g_{n+2}^{(1)} + p_4 g_{n+1}^{(1)} = 0. \end{aligned} \tag{25}$$

To solve (25) we try

$$q_n = (y^2 z)^n \sum_{m=0}^{\infty} \frac{y^{2m}(1-y^2)^{2m} z^{m(m+n+1)} u_m(y,z)}{\prod_{k=1}^{m}(1-y^2 z^k)^2 \prod_{k=1}^{m}(1-z^k)^2} \tag{26}$$

with $u_0 = 0$. Substituing into (25), and after some work we get the following simple first order difference equation for u_m:

$$u_m - u_{m-1} = \frac{2(1-y^2 z^{2m})}{(1-z^m)(1-y^2 z^m)}, \tag{27}$$

which has the solution

$$u_m = 2\sum_{k=1}^{m} \frac{(1-y^2 z^{2k})}{(1-z^k)(1-y^2 z^k)} \qquad m \geq 1. \tag{28}$$

Substituting (28) into (26) and with (23) gives $g_n^{(2)}$ explicitly. The fourth solution given by taking $\lambda = z$ is also rejected as it gives the wrong limiting behaviour.

The general solution is thus

$$g_n = A_1 g_n^{(1)} + A_2 g_n^{(2)}. \tag{29}$$

where A_1 and A_2 are arbitrary functions of y and z (but independent of n). What we eventually require is $G(y,z) = \sum_{n=1}^{\infty} g_n$. However in order to eliminate A_1 and A_2 we shall also require $H = \sum_{n=1}^{\infty} n g_n$.

From (13) we obtain

$$\begin{aligned} g_2 - z g_1 &= z^2 y^4 (y^2 - 1) - z^2 y^2 (1 - 2y^2) G \\ g_1 &= z y^4 + z y^2 H \end{aligned} \tag{30}$$

which provide the "initial" conditions for the recurrence relation (14). Substituting the expressions obtained for g_1, g_2, H and G into (30) gives two linear simultaneous equations for the two unknowns A_1 and A_2. Solving these two equations gives the final solution for the two variable generating function $G(y,z)$ as

$$G(y,z) = \sum_{m=0}^{\infty} R_m z^{m(m+2)} \frac{y^2 z}{1 - y^2 z^{m+1}} \Big[A_1 + A_2((1-y^2 z^{m+1})^{-1} + 2S_m)\Big], \tag{31a}$$

where

$$
\begin{aligned}
A_1 =& \frac{1}{\Delta}\sum_{m=0}^{\infty} R_m z^{m(m+2)}\Big[y^2(y^2-1)Q_m^{(4)} - y^2 Q_m^{(2)}\Big],\\
A_2 =& \frac{1}{\Delta}\sum_{m=0}^{\infty} R_m z^{m(m+2)}\Big[y^2 Q_m^{(1)} - y^2(y^2-1)Q_m^{(3)}\Big],\\
\Delta =& \sum_{m=0}^{\infty}\sum_{\ell=0}^{\infty} R_m R_\ell \Big[Q_m^{(1)}Q_\ell^{(4)} - Q_m^{(2)}Q_\ell^{(3)}\Big] z^{m(m+2)+\ell(\ell+2)},\\
Q_m^{(1)} =& y^2 z^m - 1 + y^2 z(1-2y^2)/T_m,\\
Q_m^{(2)} =& 2S_m Q_m^{(1)} + 2y^2 z^m - 1 + y^2 z(1-2y^2)/T_m^2,\\
Q_m^{(3)} =& 1 - y^2 z/T_m^2,\\
Q_m^{(4)} =& 2S_m(1 - y^2 z/T_m^2) + 1 - y^2 z(1 + y^2 z^{m+1})/T_m^3,\\
T_m =& 1 - y^2 z^{m+1},\\
S_m =& \sum_{k=1}^{m} \frac{(1-y^2 z^{2k})}{(1-z^k)(1-y^2 z^k)},\\
R_m =& \frac{y^{2m}(1-y^2)^{2m}}{\prod_{k=1}^{m}(1-y^2z^k)^2 \prod_{k=1}^{m}(1-z^k)^2}.
\end{aligned} \tag{31b}
$$

The above solution has been expanded as a two variable Taylor series using *Mathematica* to order $y^{18}z^{16}$ and all the terms agree with the exact enumeration results of Enting and Guttmann (unpublished). We note that because of the natural boundary of singularities at $|z| = 1$, coming from R_m, it is not possible to take the limit $z \to 1$ of (31) in order to obtain the perimeter generating function of Brak et. al. (1990).

4 Numerical analysis of row-convex result

Because of the analogy of the polygon problem to a fluid system we should expect to be able to construct a “phase diagram” from the “grand potential” $G(y, z)$. The general form of the phase diagram has been suggested by M E Fisher (private communication). The phase boundary is a line of critical points $y = y(z)$.

As a prelude to an analytical analysis of the two variable generating functions, which we hope to publish, we have carried out a numerical analysis of both the staircase and row-convex two variable generating functions (10) and (31), which consists of the following steps: the equations have been used to generate a series expansion of $G(y, z)$. For a range of numerical values of y, the single variable series in z were then analysed using differential approximants to obtain the critical value of z and the corresponding critical exponent. The line of points $y = y(z)$ is the "phase boundary " which is plotted for the row-convex and staircase polygons in figure 1 (as y^2 against z). The vertical line from (1,0) to $(1, y_c^2)$ is the "condensation line", or the "zero-field" phase boundary. The critical value y_c is the critical value of the perimeter generating function and is given by $y_c = \sqrt{2} - 1$ for the row-convex polygons and $y_c = 1/2$ for the convex polygons (Brak et al 1990). The critical exponents along the line are 1 (corresponding to a simple pole) except when the boundary is approached along the line $z = 1$ where the exponents are 1/2 (a branch zero) for both row-convex and staircase polygons.

The limiting behaviour of the row-convex phase boundary as $z \to 0$ is found to be

$$y^2 \sim \frac{1}{2.91z} \qquad z \to 0, \qquad \text{(row-convex)}$$

which is seen, in figure 2, from the "straight line" behaviour of a plot of $1/y^2$ against z, in the region of $z = 0$. For the staircase polygons the behaviour is

$$y^2 \sim \frac{1}{2z} \qquad z \to 0, \qquad \text{(staircase)}$$

as shown in figure 2. A numerical fit of a rational function to each of the phase boundary curves gave the following results:

$$y^2 = \frac{0.1151 + z}{z(0.0402 + 1.9983z)} \qquad \text{(staircase)}$$

$$y^2 = \frac{-0.0761 + z}{z(0.0567 + 2.9131z)}. \qquad \text{(row-convex)}$$

When these rational approximations are plotted on the same graph as the phase boundaries of figure 1, the curves are indistinguishable.

5 References

Brak R, Guttmann A J and Enting I G 1990J Phys. A: Math. Gen.- to appear

Delest M P and Viennot G 1984 Theor. Comput. Sci. 343377.

Enting I G and Guttmann A J 1989 J Phys. A: Math. Gen.222639

Guttmann A J and Enting I G 1988 J Phys. A: Math. Gen.21L165

– 1988b J Phys. A: Math. Gen. 21L467

Kim D 1988 Disct. Math.7047

Lin K YPreprint 1990

Lin K Y and Chang S J 1988 J Phys. A: Math. Gen.212635

Privman V and Švrakić 1988J. Stat. Phys.511091

Pólya G 1969J. Comb. Theor.6102

Temperley H N V 1956Phys. Rev.1031

Wolfram S 1988 Mathematica^{TM}, A System for Doing Mathematics by Computer (New York: Addison-Wesley)

6 Figures

A plot of the phase boundary for the row-convex polygons (solid diamonds) and the staircase polygons (dotted square). The z variable is the area generating variable (analogous to an external field), and y^2 is the *square* of the perimeter generating variable (analogous to a temperature variable).

A plot of the phase boundary with plot axis $1/y^2$ against z. The linear behaviour of the plots near the origin shows that $y^2 \sim 1/2.91z, \quad z \to 0$ for the row-convex polygons and that $y^2 \sim 1/2z, \quad z \to 0$ for the staircase polygons