



ELSEVIER

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

Exact solution of two classes of prudent polygons

Uwe Schwerdtfeger

Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, 33501 Bielefeld, Germany

ARTICLE INFO

Article history:

Received 14 January 2009

Accepted 23 July 2009

Available online 29 September 2009

ABSTRACT

Prudent walks are self-avoiding walks on a lattice which never step into the direction of an already occupied vertex. We study the closed version of these walks, called prudent polygons, where the last vertex of the walk is adjacent to its first one. More precisely, we give the half-perimeter generating functions of two subclasses of prudent polygons on the square lattice, which turn out to be algebraic and non-D-finite, respectively.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The enumeration of self-avoiding walks (SAW) and polygons (SAP) on a lattice by their number of steps [11] is a long standing problem in combinatorics. Extrapolation of series data from exact enumeration has led to high precision estimates of the exponential growth rate and subexponential corrections but an exact solution of either problem (i.e. finding the generating function, see below) seems out of reach. Rechnitzer [16] has shown that the anisotropic generating function of SAPs on the square lattice is not D-finite. A (possibly multivariate) function $f(\mathbf{z})$ is *D-finite*, if the vector space over $\mathbb{C}(\mathbf{z})$ spanned by its derivatives is finite dimensional. In the univariate case this means that f is a solution of a homogeneous linear ordinary differential equation with polynomial coefficients. At present one tries to find solvable subclasses with large exponential growth rates. This approach is particularly successful in two dimensions. We will restrict to the square lattice in our paper. One promising example is the class of *prudent walks* (PW) [5, 14]. These are SAWs which never step towards an already occupied vertex. Note that a general prudent walk is not *reversible*, i.e. the walk traversed backwards from its terminal vertex to its initial vertex may not be prudent. Since SAWs are counted modulo translation, we may choose the initial vertex of a PW to be the origin $(0, 0)$. The full problem of PW is unsolved, but recently Bousquet-Mélou [2] succeeded in enumerating a substantial subclass. We adopt the terminology of her paper and use the same methods to obtain the generating functions for the corresponding polygon models defined below. Every nearest neighbour walk on the square lattice has a minimal bounding rectangle containing it, referred to as the *box* of the walk. It is easy

E-mail address: uschwerd@math.uni-bielefeld.de.

to see that each unit step of a prudent walk ends on the boundary of its current box. (This is *not* a characterisation of PWs, e.g. the walk $(0, 0) \rightarrow (0, 1) \rightarrow (1, 1) \rightarrow (2, 1) \rightarrow (2, 0) \rightarrow (1, 0)$ is not prudent.) This property allows the definition of the following subclasses. Call a PW *one-sided*, if every step ends on the *top* side and *two-sided*, if every step ends on *top* or on the *right* side of the current box. Similarly, a PW is referred to as *three-sided* if every step ends on the *left*, *top* or the *right* side of the current box and additionally each *left step* and each *right step* that ends on the *bottom side* of the current box *inflates* the box.

Remark. (i) As soon as the width of the box of a PW is greater than one, the latter additional condition is redundant. It rules out certain configurations which can occur only if the box width is equal to one, namely “downward zig-zags” of width one, e.g. $\dots \rightarrow (1, 0) \rightarrow (1, -1) \rightarrow (0, -1) \rightarrow (0, -2) \rightarrow (1, -2) \rightarrow (1, -3) \rightarrow \dots$ which needlessly complicate the computations below.
(ii) Duchi [5] introduced two-sided and three-sided PWs as *type-1* and *type-2* PWs, respectively. In [4, 7] the authors also employ her notation.

Explicit expressions for the generating functions of one-, two- and three-sided PWs have been found so far confirming the data obtained in [4,7] by computer enumeration. The first class consists of partially directed walks and has a rational generating function. The second class was shown to have an algebraic generating function by Duchi [5] and recently in [2] the third class was solved and the generating function was found not to be D-finite. Guttmann [4,7,9] proposed to study the polygon version of the problem, meaning walks, whose last vertex is adjacent to the starting vertex. We exclude single edges from this definition. As above, the property of being prudent demands a starting vertex and a terminal vertex. So prudent polygons are *rooted* polygons with a directed root edge. Note further that a prudent polygon (PP) which ends, say, to the right of the origin (i.e. in the vertex $(1, 0)$) may never step right of the line $x = 1$, and furthermore if the walk hits that line it has to head directly to the vertex $(1, 0)$. So prudent polygons are *directed* in the sense that they contain a corner of their box. Moreover, a k -sided PP can be interpreted as a $(k - 1)$ -sided PW confined in a half-plane, see also Section 6. In this paper we deal with the polygon versions of the two-sided and three-sided walks, referred to as two-sided and three-sided PPs. Enumeration of one-sided PPs is trivial, since these are simply rows of unit cells. We give explicit expressions for the half-perimeter generating functions of two-sided PPs and three-sided PPs and show that the latter is not D-finite, as expected on numerical grounds [4,7]. To our knowledge three-sided PPs are the first *exactly solved* polygon model with a non-D-finite *half-perimeter* generating function. Enumeration of the full class of PPs remains an open problem, as for the walk case.

Outline: In Section 2 we give functional equations for the generating functions which are based on decompositions of the classes in question, in Section 3 we solve those by the kernel method [1–3, 12] and in Section 4 we study the analytic behaviour of the generating functions of two-sided and three-sided prudent polygons. Section 5 is dedicated to the random generation of three-sided PPs.

2. Functional equations

In combinatorial enumeration of objects from a class \mathcal{P} (say PPs) with respect to counting parameters $c_1, \dots, c_n : \mathcal{P} \rightarrow \mathbb{Z}_{\geq 0}$ (say perimeter, area etc.) the (multivariate) power series

$$P(x_1, \dots, x_n) = \sum_{\alpha \in \mathcal{P}} x_1^{c_1(\alpha)} \dots x_n^{c_n(\alpha)}$$

is called the *generating function* of \mathcal{P} . In the following we will count prudent polygons by half-perimeter and other, so-called *catalytic counting parameters*. The variables in the generating function marking the latter are called *catalytic variables*. Their introduction allows us to translate certain combinatorial decompositions into non-trivial functional equations for the associated generating functions [8]. Furthermore, we identify a PP (a “closed” PW) with the collection of unit cells it encloses and build larger PPs from smaller ones by attaching unit cells in a prudent fashion, i.e. the new boundary walk with the same initial vertex remains prudent. A two-sided prudent polygon either ends at the vertex above the origin or at the vertex to the right of it. This partitions two-sided PPs into two

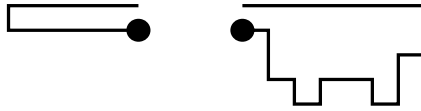


Fig. 1. Degenerate (left) and generic 2-sided PPs ending on the top of the box.

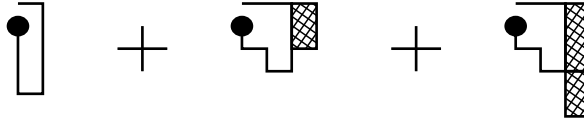


Fig. 2. Illustration of the decomposition underlying functional equation (2.1).

subsets, which can be transferred into each other by the reflection in the diagonal $x = y$. So it suffices to enumerate prudent polygons ending on the top of their box. Here two cases occur, namely the “degenerate” case when the first steps of the walk are left. The resulting PP is simply a row of unit cells pointing to the left. These have a half-perimeter generating function $t^2 + t^3 + \dots = t^2/(1 - t)$. In the “generic case” such a PP is a bar graph turned upside down, i.e. a column convex polyomino containing the top side of its bounding box, cf. Fig. 1. Denote by $B(t, u, w)$ the generating function of bar graphs counted by half-perimeter, width and height of the rightmost column (catalytic parameter), marked by t, u and w respectively. Here w is the catalytic variable. The width parameter is not a catalytic parameter. However, it will be important in the study of three-sided PPs. We follow the lines of [1].

Lemma 2.1. *The generating function $B(t, u, w)$ of bar graphs satisfies the functional equation*

$$B(t, u, w) = u \left(\frac{t^2 w}{1 - wt} + \frac{wt (B(t, u, 1) - B(t, u, w))}{1 - w} + \frac{B(t, u, w) t^2 w}{1 - wt} \right). \tag{2.1}$$

Proof. A bar graph is either a single column, or it is obtained by attaching a new column to the right side of a bar graph. The decomposition is sketched in Fig. 2. Single columns of height ≥ 1 contribute $ut^2 w/(1 - wt)$ to the generating function. The polygons obtained by adding a column which is shorter than or equal to the old rightmost column contribute the second summand. This is seen as follows. A polygon of half-perimeter n , width k and rightmost column height l contributing $t^n u^k w^l$ to $B(t, u, w)$ gives rise to l polygons whose rightmost column is shorter or equal. Their contribution sums up to

$$tu \sum_{j=1}^l t^n u^k w^j = tuw \frac{t^n u^k 1^l - t^n u^k w^l}{1 - w}. \tag{2.2}$$

Summing this over all polygons gives the second summand. The third summand corresponds to adding a larger column. To this end duplicate the rightmost column and attach a non-empty column below the so obtained new rightmost column. A so obtained bar graph can be viewed as an ordered pair of a bar graph and a column. The generating function of those pairs is the third summand of the rhs. This finishes the proof. \square

The walk constituting the boundary of a three-sided PP has $(0, 0)$ as its initial vertex and $(1, 0)$ or $(-1, 0)$ or $(0, 1)$ as its terminal vertex. Those walks with terminal vertex $(0, 1)$ may not step above the line $y = 1$ and they have to move directly to the vertex $(0, 1)$ as soon as they step upon that line. This leads to two sorts of bar graphs either rooted on their left or on their right side, see Fig. 3.

So only those three-sided PPs are of further interest, which end in $(1, 0)$ or $(-1, 0)$. Both classes are transformed into each other by a reflection in the line $x = 0$. We study those ending to the right of the origin in the vertex $(1, 0)$. Again a degenerate and a generic case are distinguished, according to whether such a PP reaches its terminal vertex from below via the vertex $(1, -1)$ (“counterclockwise around the origin”) or from above, via the vertex $(1, 1)$ (“clockwise”). In the degenerate case we



Fig. 3. Three-sided PPs with terminal vertex (0, 1) are bar graphs.

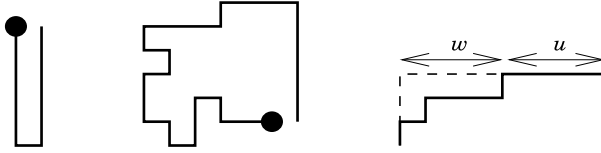


Fig. 4. Degenerate and generic three-sided PPs, catalytic variables.

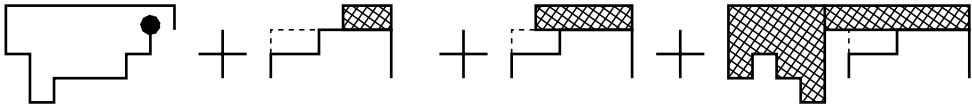


Fig. 5. Illustration of the decomposition underlying functional equation (2.3).

simply obtain a single column. In the generic case, a (possibly empty) sequence of initial down steps is followed by a left step. So denote by $R(t, u, w)$ the generating function of the generic three-sided PPs ending in the vertex $(1, 0)$ counted by half-perimeter, the length of the top row and the distance of the top left corner of the top row and the top left corner of the box, marked by t, u , and w , respectively, cf. Fig. 4. Here both u and w are catalytic variables.

Lemma 2.2. *The generating function $R(t, u, w)$ of generic three-sided PPs satisfies the functional equation*

$$R(t, u, w) = ut (B(t, u) + t) + \frac{ut (R(t, w, w) - R(t, u, w))}{w - u} + \frac{ut^2 (R(t, u, w) - R(t, u, ut))}{w - ut} + R(t, u, ut) ut (B(t, u) + t), \tag{2.3}$$

where $B(t, u) := B(t, u, 1)$ is the generating function of bar graphs counted by half-perimeter and width.

Proof. The decomposition we use is sketched in Fig. 5. The polygons in question contain the top right corner of their box. This corner is some point $(1, y)$. If $y = 1$, then the PP is either the unit square containing $(0, 0)$ and $(1, 1)$ or a bar graph as above with that unit square glued to the right. This yields the first summand. A PP with $y > 1$ is obtained in one of the following three ways from a PP with top right corner $(1, y - 1)$. The first is to add a new row on top, which is shorter than or equal to the original top row. A similar computation as in (2.2) (with some additional book keeping on w) yields the second summand. The second way to obtain a larger PP from a smaller one is by adding a new row on top, which is longer than the original top row, but does not inflate the box to the left. Again a treatment similar to the computation in (2.2) yields the third summand. The third way to extend a PP is to add a row on top of length equal to the width of the box plus one and possibly an arbitrary bar graph. This finally yields the fourth summand and the functional equation is complete. \square

Remark. As in the case of general SAPs [10] we can define a concatenation of two three-sided PPs. Roughly speaking, the one PP can be enlarged by inserting the other one at the top corner of the leftmost column, see Fig. 6.

The numbers $pp_3^{(m)}$ of three-sided PPs hence satisfy $pp_3^{(m+n)} \geq pp_3^{(m)} \cdot pp_3^{(n)}$. This implies the existence of a connective constant β , i.e. a representation $pp_3^{(m)} = \exp(\beta m + o(m))$. The precise value for β and the subexponential corrections are given in Section 4. The converse inequality holds

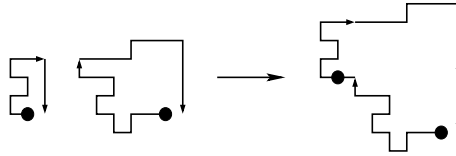


Fig. 6. Concatenating two 3-sided PPs.

for prudent walks, since breaking an $m + n$ step PW after m steps leaves one with a pair of prudent walks of respective lengths m and n .

3. Solution by the kernel method

The following result has already been obtained in [15] as the solution of an algebraic equation which arises from a different decomposition of the class. We derive it here for completeness and to recall the “classic” kernel method as applied in [1].

Theorem 3.1. *The generating function $B(t, u) := B(t, u, 1)$ of bar graphs counted by half-perimeter and width is equal to*

$$B(t, u) = \frac{1 - t - u(1 + t)t - \sqrt{t^2(1 - t)^2u^2 - 2t(1 - t^2)u + (1 - t)^2}}{2tu}. \tag{3.1}$$

Proof. The functional equation (2.1) is equivalent to

$$0 = (t^2uw(1 - w) - uwt(1 - wt) - (1 - w)(1 - wt))B(t, u, w) + tuw(1 - wt)B(t, u, 1) + t^2uw(1 - w). \tag{3.2}$$

The kernel equation

$$0 = (t^2uw(1 - w) - uwt(1 - wt) - (1 - w)(1 - wt))$$

is a quadratic equation in the catalytic variable w and has the following unique power series solution $q(t, u)$

$$q(t, u) = \frac{1 + (1 - u)t + ut^2 - \sqrt{t^2(1 - t)^2u^2 - 2t(1 - t^2)u + (1 - t)^2}}{2t}. \tag{3.3}$$

Upon substituting $w = q(t, u)$ into Eq. (3.2), the terms with $B(t, u, w)$ are cancelled and we can solve for $B(t, u, 1)$, which leads to (3.1). □

Remark. In principle, $B(t, u, w)$ can also be computed, by substituting the result for $B(t, u, 1)$ into Eq. (3.2).

By setting $u = w = 1$ in the bar graph generating function, adding the contribution of the degenerate two-sided PPs and multiplication by 2, we obtain the following result so far conjectured by series extrapolation from exact enumeration data [4].

Corollary 3.2. *The generating function of two-sided prudent polygons is equal to*

$$\begin{aligned} PP_2(t) &= \frac{1}{t} \left(\frac{1 - 3t + t^2 + 3t^3}{1 - t} - \sqrt{(1 - t)(1 - 3t - t^2 - t^3)} \right) \\ &= 4z^2 + 6z^3 + 12z^4 + 28z^5 + 72z^6 + 196z^7 + 552z^8 + 1590z^9 \\ &\quad + 4656z^{10} + 13812z^{11} + 41412z^{12} + 125286z^{13} + 381976z^{14} \\ &\quad + 1172440z^{15} + 3620024z^{16} + 11235830z^{17} + 35036928z^{18} \\ &\quad + 109715014z^{19} + 344863872z^{20} + \dots \end{aligned} \tag{3.4}$$

Now we turn to the three-sided case. Note that the sum of the catalytic counting parameters, namely the length of the top row and the distance of its top left corner to the top left corner of the box, is equal to the width of the polygon. We have the following result for the generic three-sided PPs ending on the right. It is derived in a similar way as the corresponding result on PWs in [2].

Theorem 3.3. *The functional equation (2.3) has a unique power series solution. For the generating function $R(t, w, w)$ of generic three-sided prudent polygons ending on the right and counted by half-perimeter and width we have an explicit expression as an infinite sum of formal power series*

$$R(t, w, w) = \sum_{k \geq 0} L\left((tq^2)^k w\right) \prod_{j=0}^{k-1} K\left((tq^2)^j w\right). \tag{3.5}$$

Here

$$q := q(t, 1) = \frac{t^2 + 1 - \sqrt{1 - 4t + 2t^2 + t^4}}{2t}, \tag{3.6}$$

with $q(t, u)$ as in (3.3) in the proof of Theorem 3.1. K and L are given by

$$K(w) = \frac{(1-t)q - 1 - ((1-t+t^2)q - 1)(B(t, qw) + t)w}{1-t(1+t)q - (t(1-t-t^3)q + t^2)(B(t, qw) + t)w} \tag{3.7}$$

and

$$L(w) = \frac{(1+t^2 - (1-2t+2t^2+t^4)q)(B(t, qw) + t)w}{1-t(1+t)q - (t(1-t-t^3)q + t^2)(B(t, qw) + t)w}, \tag{3.8}$$

where $B(t, u)$ is the generating function of bar graphs as in (3.1).

Proof. The functional equation (2.3) is equivalent to

$$\begin{aligned} 0 &= (ut^2(w-u) - ut(w-ut) - (w-u)(w-ut))R(t, u, w) \\ &\quad + (ut(B(t, u) + t)(w-u)(w-ut) - t^2u(w-u))R(t, u, ut) \\ &\quad + ut(w-ut)R(t, w, w) + ut(w-u)(w-ut)(B(t, u) + t). \end{aligned} \tag{3.9}$$

We first solve the kernel equation

$$(ut^2(w-u) - ut(w-ut) - (w-u)(w-ut)) = 0$$

for u and w . The unique power series solutions are $U(t, w) = q(t)w$ resp. $W(t, u) = q(t)tu$, with $q(t)$ as in (3.6). We substitute $w = W(t, u)$ in Eq. (3.9) and obtain an expression for $R(t, u, ut)$ in terms of $R(t, qtu, qtu)$, namely

$$R(t, u, ut) = \frac{(q-1)R(t, qtu, qtu) + (1-q)(1-tq)(B(t, u) + t)u}{1-qt - (1-q)(1-tq)(B(t, u) + t)u}. \tag{3.10}$$

Substitute this into Eq. (3.9) and set $u = U(t, w)$. This relates $R(t, w, w)$ and $R(t, wtq^2, wtq^2)$ as follows:

$$R(t, w, w) = K(w) \cdot R(t, wtq^2, wtq^2) + L(w), \tag{3.11}$$

with $K(w)$ and $L(w)$ as in (3.7) and (3.8). $K(w)$ and $L(w)$ are a formal power series in t , which is seen as follows: $B(t, qw)$ is well-defined as a formal power series in t as $[t^N]B(t, u)$ is a polynomial in u of degree at most $N-1$. Furthermore by the definition of B we see $B(t, u) = t^2u + O(t^3)$. The denominator is now easily seen to be $1 + O(t)$, so both $K(w)$ and $L(w)$ are well-defined as formal power series in t . Inspecting the first few coefficients we see $(1-t)q - 1 = O(t^3)$ and $1 - (1-t+t^2)q = O(t^2)$, so the numerator of $K(w)$ is $O(t^3)$. In a similar way the numerator of $L(w)$ is seen to be $w \cdot O(t^2)$. Moreover we have $tq^2 = t + O(t^2)$. So we can iterate Eq. (3.11) and obtain formula (3.5). \square

Remark. (i) We have the following alternative expressions for $K(w)$ and $L(w)$:

$$K(w) = \frac{((1 - q)(1 - qt)qwt (B(t, qw) + t) + t^2q(q - 1)) (q - 1)}{q(1 - qt)^2 ((1 - q)qwt (B(t, qw) + t) + t)} \tag{3.12}$$

and

$$L(w) = \frac{(1 - qt)(1 - q^2t)(q - 1)qt w (B(t, qw) + t)}{q(1 - qt)^2 ((1 - q)qwt (B(t, qw) + t) + t)}. \tag{3.13}$$

The expressions (3.7) and (3.8) were obtained by expressing powers of q in terms of q , e.g.

$$\begin{aligned} q^2 &= (t(t^2 + 1)q - t) / t^2, \\ q^3 &= (t(t^4 + 2t^2 - t + 1)q - t^3 - t) / t^3, \\ q^4 &= (tq(t^6 + 3t^4 - 2t^3 + 3t^2 - 2t + 1) - t + t^2 - 2t^3 - t^5) / t^4. \end{aligned}$$

(ii) In principle one could also compute $R(t, u, w)$. To obtain the generating function of all three-sided PPs we sum up the contributions of the degenerate PPs and those ending on top, multiply by two and obtain

$$PP_3(t) = 2 \left(\frac{t^2}{1 - t} + B(t, 1) + R(t, 1, 1) \right).$$

The first few terms of the series $PP_3(t)$ are

$$\begin{aligned} PP_3(t) &= 6t^2 + 10t^3 + 24t^4 + 66t^5 + 198t^6 + 628t^7 + 2068t^8 + 7004t^9 + 24260t^{10} \\ &\quad + 85596t^{11} + 306692t^{12} + 1113204t^{13} + 4085120t^{14} \\ &\quad + 15131436t^{15} + 56495170t^{16} + 212377850t^{17} + 803094926t^{18} \\ &\quad + 3052424080t^{19} + 11653580124t^{20} + \dots \end{aligned}$$

4. Analytic properties of the generating functions

So far we have considered the generating functions in question as *formal* power series. A crude estimate on the number of SAPs of half-perimeter n is 4^{2n} which is the total number of all nearest neighbour walks on the square lattice of length $2n$. So the series $PP_2(t)$ and $PP_3(t)$ converge at least in the open disc $\{|t| < 1/16\}$ and represent analytic functions there. This section deals with the analytic properties of these functions. We first discuss the analytic structure of the generating function of two-sided PPs.

Proposition 4.1. *The generating function $PP_2(t)$, cf. (3.4), is algebraic of degree 2, with its dominant singularity a square root singularity at $t = \rho$, where ρ is the unique real root of the equation*

$$\frac{1 - 4t + 2t^2 + t^4}{1 - t} = 1 - 3t - t^2 - t^3 = 0. \tag{4.1}$$

With $\theta = \sqrt[3]{26 + 6\sqrt{33}}$ the exact value for ρ can be written as

$$\rho = \frac{\theta^2 - \theta - 8}{3\theta} = 0.2955977 \dots$$

The number $pp_2^{(m)}$ of two-sided PPs of half-perimeter m is asymptotically

$$pp_2^{(m)} \sim A \cdot \rho^{-m} \cdot m^{-3/2} \quad (m \rightarrow \infty),$$

where

$$A = \frac{\sqrt{(-37 + 11\sqrt{33})\theta^2 + (-152 + 8\sqrt{33})\theta + 32}}{4\sqrt{6\pi\rho}} = 0.8548166\dots$$

- Remark.** (i) The generating function of two-sided prudent walks is algebraic with its dominant singularity a simple pole at $\bar{\sigma} = 0.403\dots$. Its coefficients are asymptotically equal to $\kappa \cdot \bar{\sigma}^{-m}$, where $\kappa = 2.51\dots$, cf. [2].
- (ii) The asymptotic number of bar graphs as well as staircase polygons (counted by half-perimeter) and Dyck paths (by half-length) is of the form $\kappa \cdot \mu^n \cdot n^{-3/2}$. Furthermore, the area random variables in the fixed-perimeter (fixed-length) ensembles of all three models are known to converge weakly to the Airy distribution [6,17,19].

The analytic structure of $PP_3(t)$ is far more complicated due to the analytic structure of $R(t, 1, 1)$, which is stated in the main result **Theorem 4.4**. In what follows we make frequent use of the following facts about the series q :

Lemma 4.2. *The series q , $(1-t)q - 1$, q^2t , $t(1+t)q$ and $t(1-t-t^3)q + t^2$ have non-negative integer coefficients. For $|t| \leq \rho$ we have the estimates*

$$|q| \leq \frac{|t|^2 + 1}{2|t|}, \quad |q^2t| \leq 1, \quad |(1-t)q - 1| \leq \rho, \quad |1 - t(1+t)q| \geq \rho. \tag{4.2}$$

Equality holds if and only if $t = \rho$. Furthermore

$$q(\rho) = \frac{\rho^2 + 1}{2\rho} = \frac{1}{\sqrt{\rho}}. \tag{4.3}$$

The singular behaviour of $B(t, q(q^2t)^N)$ and $B(t, qw)$ plays an important role in the study of $R(t, 1, 1)$.

Lemma 4.3. *For $N \geq 0$ the dominant singularity of $B(t, q(q^2t)^N)$ is σ_N , which is the unique solution in the interval $[0, \rho)$ of the equation*

$$u(t) - q(q^2t)^N = \frac{1}{t} \cdot \frac{1 - \sqrt{t}}{1 + \sqrt{t}} - q(q^2t)^N = 0.$$

In particular, $\sigma := \sigma_0 = \tau^2 = 0.2441312\dots$, where τ is the unique real root of the polynomial $t^5 + 2t^2 + 3t - 2$. The sequence $\{\sigma_N, N \geq 0\}$ is monotonically increasing and converges to ρ . Furthermore $B(t, qw)$ is analytic in the polydisc $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$.

Proof. $B(t, u)$ is singular if and only if

$$t^2(1-t)^2u^2 - 2t(1-t^2)u + (1-t)^2 = 0.$$

The relevant solution $u(t)$ with $u(\rho) = 1$ is

$$u(t) = \frac{1}{t} \cdot \frac{1 - \sqrt{t}}{1 + \sqrt{t}}. \tag{4.4}$$

$B(t, q(q^2t)^N)$ is singular if $q(q^2t)^N = u(t)$. This equation has a solution σ_N in the interval $(0, \rho)$, as $u(t) \rightarrow 1$ and $q(q^2t)^N \rightarrow (\rho^2 + 1)/2\rho = 1/\sqrt{\rho} > 1$, for $t \rightarrow \rho$. Here u is strictly decreasing and $q(q^2t)^N$ strictly increasing. We further see that σ_N converges to ρ , as for arbitrary fixed t with

$0 < t < \rho$ we can chose N sufficiently large, such that $u(t) > q (q^2t)^N$, see Lemma 4.2. So $\sigma_N \geq t$, which shows the convergence. Monotonicity follows, as $q (q^2t)^{N+1} < q (q^2t)^N$ for $t \in (0, \rho)$. All these singularities are square root singularities, as the expressions under the root are analytic in $|t| < \rho$. $B(t, qw)$ is singular, if $w = u(t)/q$ and hence

$$|w| = \frac{|u(t)|}{|q|} \geq \sqrt{\rho}u(\rho) = \sqrt{\rho},$$

with equality if and only if $t = \rho$. So there is no singularity inside the polydisc. \square

Now we are ready to state the main result, which is proven in the subsequent lemmas.

Theorem 4.4. *The function $R(t, 1, 1)$ is analytic in the disc $\{|t| < \sigma\}$ with its unique dominant singularity a square root singularity at σ . Moreover it is meromorphic in the slit disc*

$$D_{\sigma, \rho} = \{|t| < \rho\} \setminus [\sigma, \rho),$$

and it has infinitely many square root singularities in the set $\{\sigma_N, N = 0, 1, 2, \dots\}$. In particular, $R(t, 1, 1)$ is not D -finite.

- Remark.** (i) The number $pp_3^{(m)}$ of three-sided PPs of half-perimeter m is asymptotically equal to $\kappa \cdot \sigma^{-m} \cdot m^{-3/2}$ for some positive constant κ . In particular, two-sided PPs are exponentially rare among three-sided PPs.
 (ii) The generating function of three-sided prudent walks has its dominant singularity a simple pole at $\bar{\sigma} = 0.403\dots$, as in the two-sided case. It is meromorphic in some larger disc of radius $\bar{\rho} = \sqrt{2} - 1$ with infinitely many simple poles in the interval $[\bar{\sigma}, \bar{\rho})$. Its coefficients grow like $\kappa \cdot \bar{\sigma}^{-m}$, for some $\kappa > 0$ [2].

Possible singularities of $R(t, 1, 1)$ in $D_{\sigma, \rho}$ are zeroes of the denominators of $K(w)$ and $L(w)$, places, where the representation (3.5) diverges, and square root singularities of $B(t, q (q^2t)^N)$. Now we investigate the analytic properties of the single summands in the representation (3.5).

Lemma 4.5. 1. $K((q^2t)^N)$ and $L((q^2t)^N)$ are analytic in $\{|t| < \sigma_N\}$.

2. $K((q^2t)^N w)$ and $L((q^2t)^N w)$ are analytic in $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$.

Proof. With the above definition of $u(t)$ and a short computation we obtain the estimate

$$\left| B(t, q (q^2t)^N) \right| < B(|t|, u(|t|)) = \sqrt{|t|}. \tag{4.5}$$

The denominator of $K(w)$ and $L(w)$ is

$$1 - T(t, w) = 1 - t(1+t)q - (t(1-t-t^3)q + t^2)(B(t, qw) + t)w.$$

$T(t, w)$ is a power series in t and w with non-negative coefficients and $T(0, w) = 0$. Hence we have the estimate

$$T(t, (q^2t)^N) \leq T(\sigma_N, (q(\sigma_N)^2 \sigma_N)^N) \leq T\left(\sigma_N, \frac{u(\sigma_N)}{q(\sigma_N)}\right).$$

A computation shows that the function $1 - T(t, u(t)/q(t))$

$$1 - T\left(t, \frac{u(t)}{q(t)}\right) = 1 - t(1+t)q - (t(1-t-t^3)q + t^2)\left(\sqrt{t} + t\right)\frac{u(t)}{q(t)}$$

has no zeroes in $[\sigma, \rho]$. This finishes the proof of the first assertion, as $K((q^2t)^N)$ and $L((q^2t)^N)$ do not have poles inside $\{|t| < \sigma_N\}$. Furthermore, the denominator $1 - T(t, (q^2t)^N w)$ is analytic in the

polydisc $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$, with the only singular point $(t, w) = (\rho, \sqrt{\rho})$ on its boundary. As above we see

$$\left| T\left(t, (q^2t)^N w\right) \right| \leq T(|t|, |w|) \leq T\left(\rho, \frac{u(\rho)}{q(\rho)}\right) = T(\rho, \sqrt{\rho}),$$

and hence the denominator is non-zero in the domain in question and $K\left((q^2t)^N w\right)$ and $L\left((q^2t)^N w\right)$ are both analytic in the polydisc. \square

- Lemma 4.6.** 1. *The series representation (3.5) of $R(t, 1, 1)$ is a series of algebraic functions, which converges compactly in the slit disc $D_{\sigma, \rho} = \{|t| < \rho\} \setminus [\sigma, \rho)$ to a meromorphic function.*
 2. *Furthermore the corresponding representation of $R(t, w, w)$ converges compactly in the polydisc $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$ to an analytic function.*
 3. *The Taylor expansion of $R(t, w, w)$ about $(t, w) = (0, 0)$ converges absolutely in $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$.*

Proof. For the first assertion choose $0 < r < \rho$. We look at the disc $\{|t| \leq r\}$. The term independent of w in the numerator of $K(w)$ is strictly less than ρ for $|t| \leq r$ and the corresponding term in the denominator is strictly larger than ρ , see Lemma 4.2. So we can choose N large such that $\sigma_N > r$ and $\left|K\left((q^2t)^N\right)\right| < 1$ for $|t| \leq r$. Split the series at N . The summands for $k = 0, \dots, N - 1$ sum up to a function which is meromorphic in the slit disc $\{|t| \leq r\} \setminus [\sigma, r]$. In the rest of the series take out the common factors to obtain

$$\prod_{j=0}^{N-1} K\left((tq^2)^j\right) \sum_{k \geq 0} L\left((tq^2)^{N+k}\right) \prod_{j=0}^{k-1} K\left((tq^2)^{N+j}\right). \tag{4.6}$$

The first product is a meromorphic function in the slit disc. $L\left((tq^2)^{N+k}\right)$ is easily seen to converge uniformly to 0 in $|t| \leq r$ as $k \rightarrow \infty$. In $|t| \leq r$ all summands are holomorphic (see the above discussion) and the sum can be estimated by a geometric series and hence converges uniformly in the compact disc $\{|t| \leq r\}$. By Montel’s theorem the limit of the sum is again analytic. This finishes the proof for the first assertion. The second assertion is proven along the similar lines. By the multivariate version of Montel’s theorem [18] the limit function is also analytic in the domain in question and thus the third assertion follows. \square

Lemma 4.7. *$R(t, 1, 1)$ is singular at infinitely many of the σ_N . Furthermore, $R(t, 1, 1)$ is singular at σ .*

Proof. Terms singular at σ_N only show up in the summands for $k \geq N$. The sum of these (4.6) is equal to

$$\prod_{j=0}^{N-1} K\left((tq^2)^j\right) \left[L\left((tq^2)^N\right) + K\left((tq^2)^N\right) R\left(t, (tq^2)^{N+1}, (tq^2)^{N+1}\right) \right]. \tag{4.7}$$

In order to show that the singularity σ_N does not cancel, only the term in square brackets is of interest. Singular terms show up in the numerators and the common denominator of $K\left((tq^2)^N\right)$ and $L\left((tq^2)^N\right)$. We now manipulate the expressions (3.12) and (3.13) for $K(w)$ and $L(w)$ in order to get rid of singular terms in the denominator, where the factor

$$(1 - q)qwt (B(t, qw) + t) + t$$

leads to a singularity at σ_N for $w = (tq^2)^N$. Write

$$qwt (B(t, qw) + t) = A(w) - \phi(w),$$

where

$$A(w) = \frac{1}{2} (1 + t - qw(1 + t)t)$$

$$\phi(w) = \frac{1}{2} \sqrt{t^2(1 - t)^2(qw)^2 - 2t(1 - t^2)qw + (1 - t)^2}.$$

Then $A\left((tq^2)^N\right)$ is analytic in $\{|t| < \rho\}$. After multiplication of the numerator and denominator with $(1 - q)A(w) + t + (1 - q)\phi(w)$ there is no more occurrence of ϕ in the denominator. We now have to collect the terms involving $\phi(w)$ in the numerators of $K(w)$ and $L(w)$. In the numerator of $K(w)$ the terms involving $\phi(w)$ sum up to

$$P_K(w)\phi(w) := t(q - 1)^2(1 - q^2t)\phi(w).$$

The terms involving $\phi(w)$ in the numerator of $L(w)$ sum up to

$$P_L(w)\phi(w) := (1 - qt)(1 - q^2t)(1 - q)t\phi(w).$$

So the singularity at σ_N can only cancel if

$$-\frac{P_L\left((\sigma_N q(\sigma_N)^2)^N\right)}{P_K\left((\sigma_N q(\sigma_N)^2)^N\right)} = R\left(\sigma_N, (\sigma_N q(\sigma_N)^2)^{N+1}, (\sigma_N q(\sigma_N)^2)^{N+1}\right). \tag{4.8}$$

In order to prove that this equation can hold for at most finitely many of the σ_N , we show that for σ_N sufficiently close to ρ the lhs of Eq. (4.8) is strictly decreasing while the rhs is strictly increasing. Since (σ_N) is monotonically increasing and converges to ρ this will finish the proof. We first prove the assertion on the rhs. The Taylor expansion of $R(t, w, w)$ about $(0, 0)$ has non-negative coefficients and represents $R(t, w, w)$ in the polydisc $\{|t| < \rho\} \times \{|w| < \sqrt{\rho}\}$ by Lemma 4.6. By the definition of σ_N and $u(t)$ we have

$$(\sigma_N q(\sigma_N)^2)^{N+1} = u(\sigma_N) q(\sigma_N) \sigma_N.$$

The rhs of the last equation is strictly increasing for sufficiently large N and converges to $\sqrt{\rho}$ as $N \rightarrow \infty$. The sequence σ_N is also strictly increasing by Lemma 4.3. So for large enough N the sequence $R\left(\sigma_N, (\sigma_N q(\sigma_N)^2)^{N+1}, (\sigma_N q(\sigma_N)^2)^{N+1}\right)$ is strictly increasing.

Now we turn to the lhs of Eq. (4.8). A computation yields

$$-\frac{P_L\left((\sigma_N q(\sigma_N)^2)^N\right)}{P_K\left((\sigma_N q(\sigma_N)^2)^N\right)} = \frac{1 - \sigma_N q(\sigma_N)}{q(\sigma_N) - 1},$$

which easily seen to be ultimately strictly decreasing. This finishes the proof of Lemma 4.7. \square

The Lemmas 4.5–4.7 together constitute a proof of Theorem 4.4.

5. Random three-sided prudent polygons

In [2] prudent walks of a given fixed length are generated uniformly at random with a refined version of a method proposed in [13]. We briefly describe a version of the method tailored to our particular needs. The main ingredient are *generating trees*. These are trees with their nodes labelled in such a way that if two nodes bear the same label, then the multisets of the labels of their children are the same. In this section we present a generating tree for generic three-sided prudent polygons.

The decomposition underlying the functional equation (2.3) (cf. Fig. 5) yields a rule according to which a larger three-sided prudent polygon can be constructed starting from a smaller one. We refine this to a step-by-step procedure which allows to generate any three-sided PP of half-perimeter m in a unique way, starting from the unit square, such that after the k th construction step we have a PP of half-perimeter $k + 2$, $k = 0, 1, \dots, m - 2$. The four types of steps used in the construction are

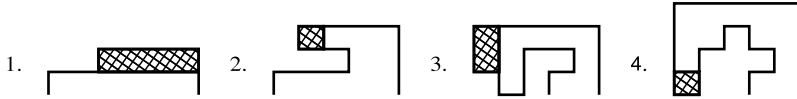


Fig. 7. The types of steps used to obtain a generating tree.

1. attaching a new top row which is shorter than or equal to the current top row,
2. attaching a unit square to the left side of the current top row,
3. attaching a new leftmost column which is shorter than or equal to the present leftmost column,
4. attaching a unit square to the bottom side of the leftmost column.

Any of these steps, if admissible, increases the half-perimeter by one, see Fig. 7. In the proof of the functional equation (2.3) the steps of types 3 and 4 are encapsulated in the “attaching a bargraph to the left” operation. Hence any generic three-sided PP can be generated starting from the unit square by using only these steps.

To each polygon we associate a label encoding the admissible steps which can be applied to enlarge it. This label is a four-tuple $(A, B, k, l) \in \{L, T\} \times \{y, n\} \times \mathbb{Z}_{\geq 0}^2$. A encodes the last building step. It is equal to T (top), if the last step was of type 1 which inflated the box to the top, or of type 2 but *without* inflating the box to the left. A is equal to L if the last step was of type 2 *and* thereby inflating the box to the left, or of types 3 or 4. If $A = T$, the parameter $B \in \{y, n\}$ (yes/no) indicates if the current top row is longer than or equal to the second row from the top, and hence if a step of type 2 is applicable. Similarly, if $A = L$, B decides if a step of type 4 can be performed, i.e. if the leftmost column is shorter than or equal to the second but leftmost one. The parameter k always denotes the length of the top row, and l is either the length of the leftmost column or the distance of the left end of the top row to the left side of the box, depending on whether $A = L$ or $A = T$, respectively. Finally, the unit square receives the label $(L, n, 1, 1)$.

Labels for the generating tree of 3-sided PPs			
A	B	k	l
T	Top row extendable?	Length of top row	Distance of box to top row
L	Left col. extendable?	Length of top row	Length of leftmost col

Remark. The polygons with $A = L$ are precisely those corresponding to the first and last term on the rhs of functional equation (2.3), see also Fig. 5 (“attaching bar graphs”).

The construction steps yield the following rewriting rules for the labels.

$$(T, n, k, l) \rightarrow \begin{cases} (T, n, i, l + k - i), & i = 1, \dots, k - 1 \\ (T, y, k, l) \end{cases} \tag{5.1}$$

$$(T, y, k, l) \rightarrow \begin{cases} (T, n, i, l + k - i), & i = 1, \dots, k - 1 \\ (T, y, k, l) \\ (T, y, k + 1, l - 1), & \text{if } l \geq 1 \\ (L, n, k + 1, 1), & \text{if } l = 0 \end{cases} \tag{5.2}$$

$$(L, n, k, l) \rightarrow \begin{cases} (T, n, i, k - i), & i = 1, \dots, k - 1 \\ (T, y, k, 0) \\ (L, n, k + 1, i), & i = 1, \dots, l - 1 \\ (L, y, k + 1, l) \end{cases} \tag{5.3}$$

$$(L, y, k, l) \rightarrow \begin{cases} (T, n, i, k - i), & i = 1, \dots, k - 1 \\ (T, y, k, 0) \\ (L, n, k + 1, i), & i = 1, \dots, l - 1 \\ (L, y, k + 1, l) \\ (L, y, k, l + 1). \end{cases} \tag{5.4}$$

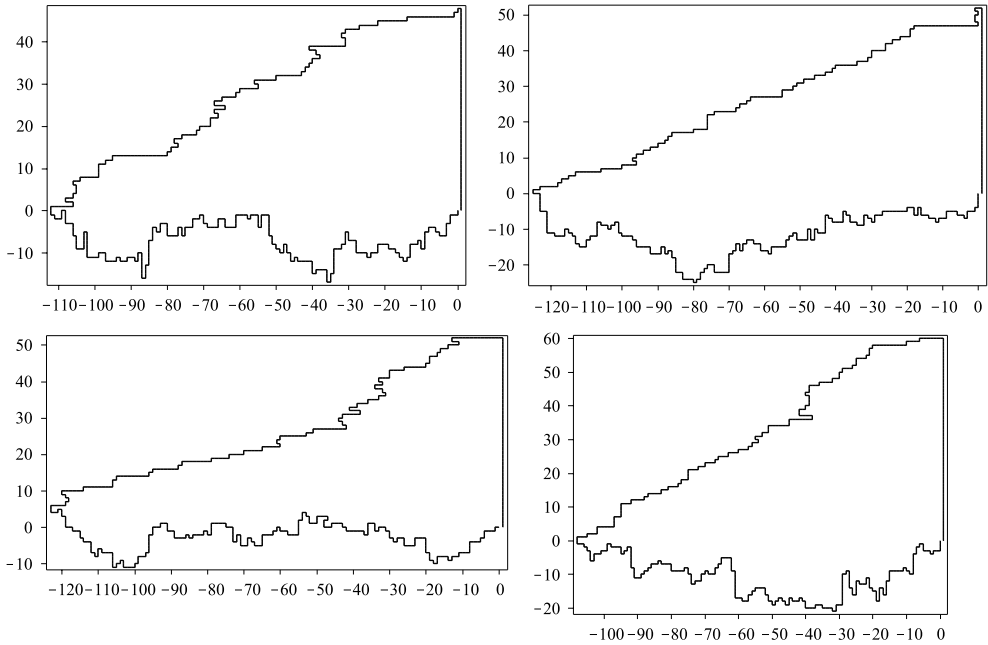


Fig. 8. Random 3-sided PPs of half-perimeter 250.

The labelled rooted tree generated according to these rewriting rules with its root labelled $(L, n, 1, 1)$ is a generating tree for three-sided prudent polygons.

The idea is that polygons of half-perimeter m correspond bijectively to paths of length $m - 2$ starting in the root (and hence to vertices on the level $m - 2$). In order to generate such a polygon (or path) uniformly at random, each step in the path has to be chosen according to an appropriate probability. This probability can be expressed in terms of *extension numbers*. If π is a polygon of half-perimeter $m - s$ (a path of length $m - s - 2$), then $EX(\pi, s)$ denotes the number of polygons of half-perimeter m which can be reached from π in s construction steps, or equivalently of extensions of length s of the path. Denote by $Ch(\pi)$ the set of polygons obtained from π in one step, i.e. the children of π in the generating tree. Now the right probability to choose $\alpha \in Ch(\pi)$ in our random sampling procedure is equal to

$$\mathbb{P}(\alpha|\pi) = \frac{EX(\alpha, s - 1)}{EX(\pi, s)}.$$

The numbers $EX(\pi, s)$ can be computed recursively, namely

$$EX(\pi, s) = \begin{cases} 1 & \text{if } s = 0, \\ \sum_{\alpha \in Ch(\pi)} EX(\alpha, s - 1) & \text{otherwise.} \end{cases}$$

The crucial observation is that $EX(\pi, \cdot)$ only depends on the label of π . In the first $m - 2$ levels of the tree $O(m^2)$ different labels occur since none of the parameters exceeds m . It hence takes $O(m^3)$ operations to compute the all required extension numbers. We have implemented the procedure and computed these numbers up to $m = 250$. See Fig. 8 for some samples.

6. Conclusion

We have solved the class of two-sided and three-sided prudent polygons, the generating function being algebraic in the former and non-D-finite in the latter case. The analysis shows that two-sided PPs

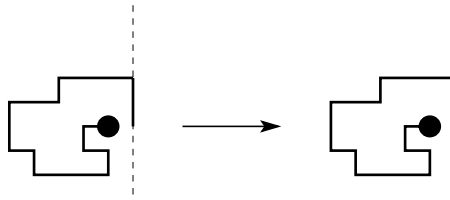


Fig. 9. Embedding of k -sided PPs into $k - 1$ -sided PWs.

are exponentially rare among three-sided PPs which is different from the corresponding walk models where the growth rates are equal.

It would be nice to solve the class of general prudent polygons. We expect that the involved functional equations require three or more catalytic variables, which is the case for the equation found for the walk model.

Since the exponential growth rates of SAWs and SAPs are known to be equal [10] it is also interesting to compare the exponential growth rates of k -sided PWs and PPs. To that end it suffices to study PPs ending in $(1, 0)$. As already mentioned in the introduction, a k -sided PP ending in $(1, 0)$ may never step right of the line $x = 1$ and it heads towards the vertex $(1, 0)$ as soon as it hits that line for the first time in a point $(1, y_0)$. Up to that step the boundary walk of that k -sided PP is genuinely $k - 1$ -sided. This yields an injective map sending a k -sided PP to a $k - 1$ -sided PW simply by reflecting the segment joining $(1, y_0)$ and $(1, 0)$ in the line $y = y_0$, see Fig. 9. We denote the so obtained subclass of $k - 1$ -sided PWs by “embedded k -sided PPs”. If we count PPs by full perimeter, their exponential growth rates become $1/\sqrt{\rho} = 1.83\dots$ for two-sided PPs and $1/\sqrt{\sigma} = 2.02\dots$ for three-sided PPs.

It is known that the exponential growth rate of PWs is equal to $1 + \sqrt{2} = 2.41\dots$ in the one-sided case and equal to $2.48\dots$ in the two- and three-sided cases [2]. The latter rate is also expected for unrestricted PWs [4,7]. Consequently, for $k = 2, 3$, our results imply that k -sided PPs are exponentially rare among k -sided PWs and, via embedding, among $k - 1$ -sided PWs. Furthermore, the rate of three-sided PPs is even smaller than that of one-sided PWs. This is not surprising looking at the pictures in Fig. 8, as such a PP roughly consists of two “almost” one-sided PWs, one heading to the far left followed by one “almost directed” walk up and to the right (and the closing tail). We expect that a similar heuristic argument also applies in the general case, which is also supported by an estimated value of approximately $2.1 < 1 + \sqrt{2}$ for the growth rate of general PPs [4,7].

Acknowledgements

Part of this work was done during the author’s stay at the University of Melbourne. He would like to thank MASCOS and the members of the Statistical Mechanics and Combinatorics group for their kind hospitality. In particular he would like to thank Tony Guttmann for introducing the problem to him. He thanks Christoph Richard for helpful comments on the manuscript. He would also like to acknowledge the financial support by the German Research Council (DFG) within CRC 701.

References

- [1] M. Bousquet-Mélou, A method for the enumeration of various classes of column-convex polygons, *Discrete Math.* 154 (1996) 1–25. doi:10.1016/0012-365X(95)00003-F.
- [2] M. Bousquet-Mélou, Families of prudent self-avoiding walks (2008), arXiv:math.CO/0804.4843.
- [3] M. Bousquet-Mélou, A. Jehanne, Polynomial equations with one catalytic variable, algebraic series and map enumeration, *J. Combin. Theory, Ser. B* 96 (5) (2006) 623–672. doi:10.1016/j.jctb.2005.12.003.
- [4] J. Dethridge, A.J. Guttmann, I. Jensen, Prudent self-avoiding walks and polygons, in: *Random Polymers*, EURANDOM, Eindhoven, The Netherlands, June, 2007.
- [5] E. Duchi, On some classes of prudent walks, in: *FPSAC’05*, Taormina, Italy, 2005.
- [6] P. Duchon, Q -grammars and wall polyominoes, *Ann. Comb.* 3 (2–4) (1999) 311–321. doi:10.1007/BF01608790.
- [7] T.M. Garoni, M. Timothy, A.J. Guttmann, I. Jensen, J. Dethridge, Prudent walks and polygons, *J. Phys. A* 42 (9) (2009) doi:10.1088/1751-8113/42/9/095205.
- [8] I.P. Goulden, D.M. Jackson, *Combinatorial Enumeration*, Wiley, New York, 1983.

- [9] A.J. Guttmann, Some solvable, and as yet unsolvable, polygon and walk models, *IOP, J. Phys.: Conference Series* 42 (2006) 98–110. doi:10.1088/1742-6596/42/1/011.
- [10] J.M. Hammersley, The number of polygons on a lattice, *Proc. Camb. Philos. Soc.* 57 (1961) 516–523.
- [11] N. Madras, G. Slade, The self-avoiding walk, in: *Probability and Its Applications*, vol. xiv, Birkhäuser, Boston, MA, 1993, p. 425.
- [12] M. Mishna, A. Rechnitzer, Two non-holonomic lattice walks in the quarter plane, *Theoret. Comput. Sci.* 410 (38–40) (2009) 3616–3630. doi:10.1016/j.tcs.2009.04.008.
- [13] A. Nijenhuis, H. Wilf, Combinatorial algorithms for computers and calculators, in: *Computer Science and Applied Mathematics*, vol. xv, 2nd ed., Academic Press, New York, San Francisco, London, 1978, p. 302.
- [14] P. Préa, Exterior self-avoiding walks on the square-lattice, unpublished manuscript, 1997.
- [15] T. Prellberg, R. Brak, Critical exponents from nonlinear functional equations for partially directed cluster models, *J. Stat. Phys.* 78 (3–4) (1995) 701–730. doi:10.1007/BF02183685.
- [16] A. Rechnitzer, Haruspicy 2: The anisotropic generating function of self-avoiding polygons is not D-finite, *J. Combin. Theory, Ser. A* 113 (3) (2006) 520–546. doi:10.1016/j.jcta.2005.04.010.
- [17] C. Richard, Limit distributions and scaling functions, in: *Polygons, Polyominoes and Polycubes*, in: *Lecture Notes in Physics*, Springer, Berlin/Heidelberg, 2009, pp. 247–299. doi:10.1007/978-1-4020-9927-4_11.
- [18] V. Scheidemann, *Introduction to Complex Analysis in Several Variables*, Birkhäuser, Basel, 2005.
- [19] L. Takács, A Bernoulli excursion and its various applications, *Adv. Appl. Probab.* 23 (3) (1991) 557–585. doi:10.2307/1427622.