Combinatorial aspects of tree-like structures
I. Enumerative combinatorics

Christoph Richard
Universität Erlangen-Nürnberg
Population Genetics and Statistical Physics in Synergy
EURANDOM, 26. 08. 2014

references

books

- Random Trees
  (M. Drmota, Springer 2009)
- Analytic Combinatorics
  (P. Flajolet and R. Sedgewick 2009, CUP)
- Polygons, Polyominoes and Polycubes
  (ed A.J. Guttmann 2009, Springer LNP 775)

references

articles

- C. Richard, Limit distributions and scaling functions LNP 775
- U. Schwerdtfeger, Exact solution of two classes of prudent polygons, European J. Combin. 31 (2010), 765–779
- U. Schwerdtfeger, Linear functional equations with a catalytic variable and area limit laws for lattice paths and polygons, European J. Combin. 36 (2014), 608–640
- N. Eisner, Skalenfunktionen von Polygonmodellen, Diploma thesis, Bielefeld (2010)

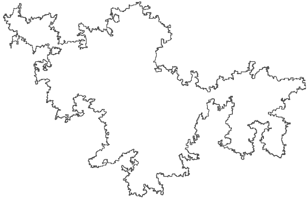
self-avoiding walks and polygons on $\mathbb{Z}^d$

SAW of length m:

- nearest neighbour path $(\omega_1, \ldots, \omega_{m+1})$, vertices pairwise disjoint

SAP of length (perimeter) m:

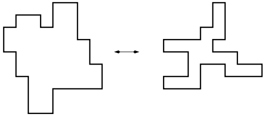
- $(\omega_1, \ldots, \omega_m)$ SAW, ω_1 and ω_m nearest neighbours



- equal exponential growth constants (Madras, Slade 92)
- SAPs "easier" than SAWs

planar SAPs are models for ...

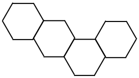
- the vesicle collapse transition in 2d



extended ↔ collapsed polygons (branched polymers, lattice trees)

  - two-dimensional vesicles (perimeter and area)
  - ring polymers (perimeter and number of contacts)

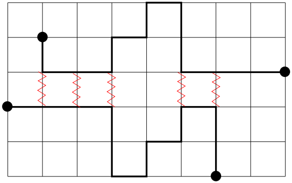
- benzenoid hydrocarbons



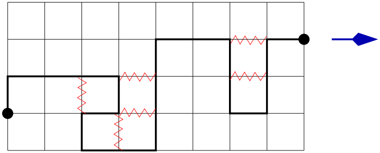
hexagonal lattice SAPs counted by area

SAPs are models for ...

- biopolymers
  - thermal DNA denaturation



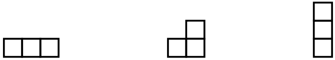
  - force-induced unfolding

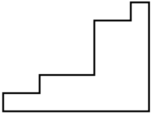


solvable subclasses of SAPs

- partitions and compositions of natural numbers

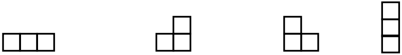
  3: *1+1+1* *1+2* *3*

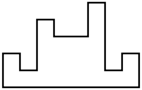
  

  partitions ↔ Ferrers diagrams

  3: *1+1+1* *1+2* *2+1* *3*

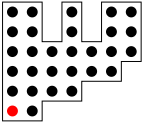
  

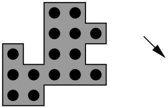
  compositions ↔ bargraph polygons

solvable subclasses of SAPs

- directed polygon:
  all points reachable from a root by $\uparrow$ or $\rightarrow$ steps

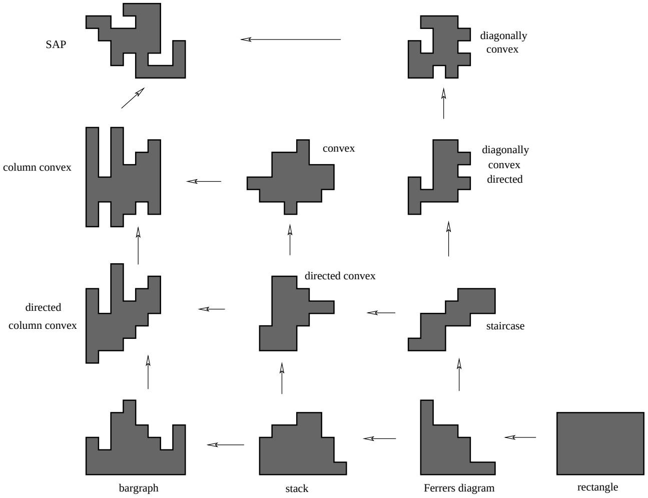


- **v**-convex polygon:
  sections with lines of slope **v** through points convex



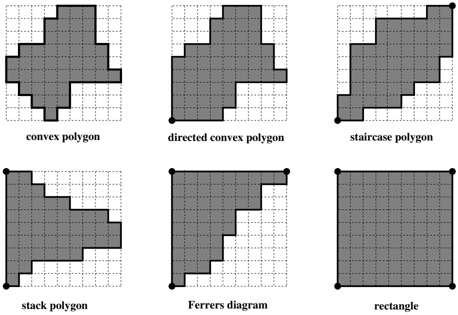
column-convex: $\mathbf{v} = \uparrow$, row-convex: $\mathbf{v} = \rightarrow$, convex=rc $\cap$ cc
diagonally convex: $\mathbf{v} = \searrow$

solvable subclasses of SAPs



convex square lattice polygons

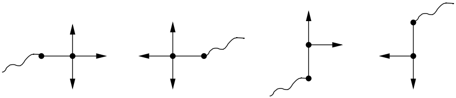
sections with vertical and horizontal lines convex
length equals perimeter of minimal bounding rectangle



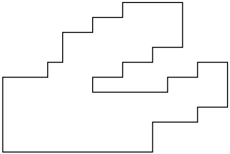
closed expression for perimeter and area generating functions
(e.g. Bousquet-Mélou 96 for horizontally convex polygons)

three-choice polygons

- three-choice walks: Manna 84



- three-choice polygons: Guttmann et al 93

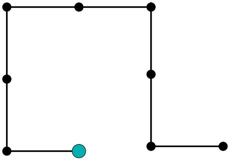


- either staircase polygon or imperfect staircase polygon
- "solution" by Guttmann–Jensen 05

prudent walks and polygons on $\mathbb{Z}^2$

if you aim to be self-avoiding, be prudent:

- Never take a step towards an already occupied vertex!



fundamental property

- each unit step ends on boundary of the current bounding box (minimal bounding rectangle)

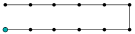
history

- walks introduced by Prea 97, polygons by Guttmann 06
- solution of walk subclasses by Duchi 05, Bousquet-Mélou 10
- solution of polygon subclasses by Schwerdtfeger 10

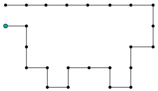
prudent polygons on $\mathbb{Z}^2$

subclasses of prudent polygons

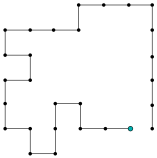
- one-sided walks: every step ends on the top of the box



- two-sided walks: every step ends on the top or right of the box

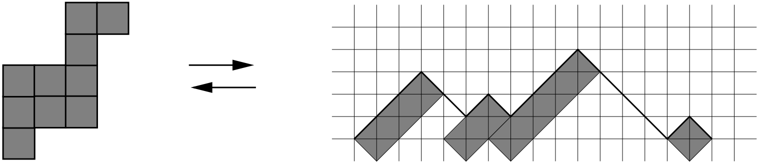


- three-sided walks: every step ends on the top, right or left of the box



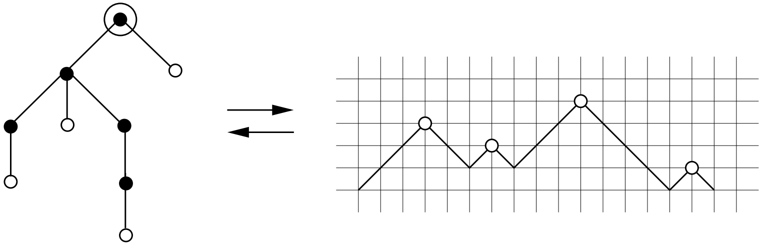
polygons, walks, and trees

staircase polygons:
Dyck path codes height and relative position of polygon columns



polygons, walks, and trees

ordered (plane) rooted trees:
Dyck path codes edge traversal of trees (contour process)

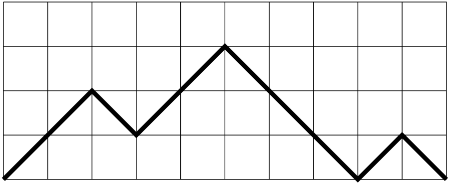


polygons, walks, and trees

- These models allow for certain combinatorial decompositions.
- A decompositon yields a recursion for counting parameters associated to the model.
- On the level of generating functions, this translates into a functional equation for the generating function.
- Sometimes the functional equation yields an explicit solution for the generating function.
- If an explicit solution is absent, manipulation of the functional equation often yields detailed information about the model.

We will discuss this for models of walks and models of trees.

Dyck paths and arches



Dyck paths of length $2n$ ($n \in \mathbb{N}_0$)

- $y : [0, 2n] \to \mathbb{R}_{\geq 0}$ (height map)
- $y(0) = y(2n) = 0$, $|y(j) - y(j-1)| = 1$ ($j \in \mathbb{N}$)
- $y(s)$ for non-integer s by linear extrapolation

arch of length $2n$ ($n \in \mathbb{N}$)

- Dyck path y where $y(s) > 0$ if $s \notin \{0, 2n\}$

combinatorial classes and generating functions

combinatorial classes

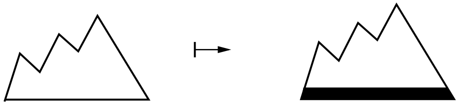
- $\mathcal{D}$ set of Dyck paths
- $\mathcal{A}$ set of arches

generating functions

- weight $w_y(x) = x^n$ of Dyck path y of length $2n$
- (formal) power series

$$D(x) = \sum_{d \in \mathcal{D}} w_d(x), \qquad A(x) = \sum_{a \in \mathcal{A}} w_a(x)$$

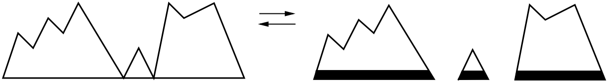
combinatorial constructions: path lifting



- Dyck path with additional bottom layer $\hat{=}$ arch
- relation between generating functions:

$$A(x) = \sum_{d \in \mathcal{D}} w_{\bar{d}}(x) = \sum_{d \in \mathcal{D}} x w_d(x) = xD(x)$$

combinatorial constructions: arch decomposition



- Dyck path $\widehat{=}$ ordered sequence of arches
- length additive w.r.t. sequence construction
- relation between generating functions:

$$D(x) = \sum_{k\geq 0} \sum_{(a_1,\ldots,a_k)\in\mathcal{A}^k} w_{(a_1,\ldots,a_k)}(x)$$

$$= \sum_{k\geq 0} \sum_{(a_1,\ldots,a_k)\in\mathcal{A}^k} w_{a_1}(x)\cdot\ldots\cdot w_{a_k}(x)$$

$$= \sum_{k\geq 0} \left(\sum_{a\in\mathcal{A}} w_a(x)\right)^k = \frac{1}{1-A(x)}$$

Dyck path generating function

- quadratic equation with uniqe power series solution

$$D(x) = \frac{1 - \sqrt{1 - 4x}}{2x} = \sum_{n \geq 0} C_n x^n$$

- Catalan numbers

$$C_n = \frac{1}{n+1} \binom{2n}{n}$$

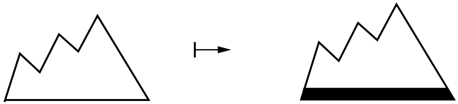
$1, 1, 2, 5, 14, 42, 132, \ldots$

Dyck paths by length and area

(half) length and area generating functions

- Dyck paths $D(x,q) = \sum_{d \in \mathcal{D}} w_d(x,q)$
- arches $A(x,q) = \sum_{a \in \mathcal{A}} w_a(x,q)$
- weight $w_y(x,q) = x^n q^m$ of path y of length $2n$, area m

path lifting:

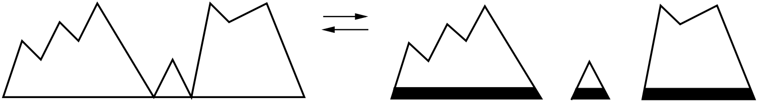


- transformation of weights: q-shift
  $w_{\bar{d}}(x,q) = x^{n+1} q^{m+2n+1} = xq(xq^2)^n q^m = xqw_d(xq^2,q)$
- transformation of generating functions

$$A(x,q) = xqD(xq^2,q)$$

Dyck paths by length and area

arch decomposition:



$$D(x,q) = \frac{1}{1 - A(x,q)}$$

(length, area additive w.r.t. sequence construction)

- q-quadratic functional equation

$$D(x,q) = \frac{1}{1 - xqD(xq^2,q)}$$

Dyck paths by length and area: explicit solution

$$D(x,q) = \frac{1}{1 - xqD(xq^2, q)}$$

- insert $D(x,q) = \frac{A(x,q)}{B(x,q)}$ and equate numerator and denominator

$$A(x,q) = B(xq^2, q), \qquad B(x,q) = B(xq^2, q) - xqA(xq^2, q)$$

- write $B(x,q) = \sum_{n=0}^{\infty} b_n(q)x^n$ and identify

$$b_n = q^{2n} b_n - q^{4n-1} b_{n-1}, \qquad b_0 = 1$$

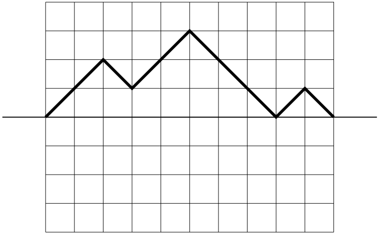
- iterate this to get

$$B(x,q) = \sum_{n=0}^{\infty} \frac{(-x)^n q^{2n^2+n}}{(q^2)_n},$$

  with q-product $(q^2)_n = (1-q^2) \cdot \ldots \cdot (1-q^{2n})$
- q-deformed exponential since $B(x(1-q^2), q) \to e^{-z}$ as $q \to 1$

from Dyck paths to random walks

- Dyck path: non-negative RW starting and ending in $y = 0$



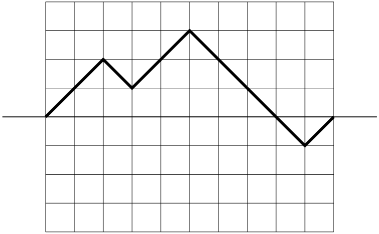
- ordered sequence of Dyck paths with additional bottom layer

$$D(x, q) = \frac{1}{1 - x^2 q D(qx, q)}$$

- counted by length and area (not half-length!)

bilateral Dyck paths

- bilateral Dyck path: RW starting and ending in $y = 0$



- ordered sequence of positive or negative Dyck paths with additional bottom layer

$$B(x, q) = \frac{1}{1 - 2x^2 q D(qx, q)}$$

- counted by length and *absolute* area

meanders

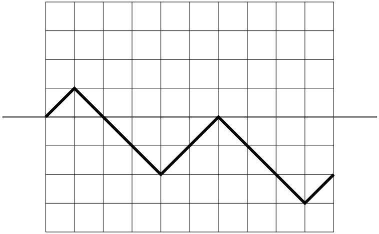
- meander: non-negative RW starting in $y = 0$

- Dyck path or (Dyck path followed by meander with additional bottom layer)

$$M(x, q) = D(x, q)(1 + xqM(qx, q))$$

random walks

- bilateral Dyck path or (bilateral Dyck path followed by a positive or negative meander with additional bottom layer)



- functional equation

$$R(x, q) = B(x, q)(1 + 2xqM(qx, q))$$

- counted by length and *absolute* area

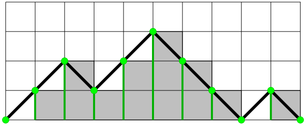
Dyck paths: moments of height (cf Duchon 99)

Dyck path y of length $2n$

- k-th moment of height ($k \in \{1, \ldots, M\}$)

$$n_k = \sum_{i=0}^{2n} y^k(i)$$

- $k = 1$ area, $k = 2$ moment of inertia



- weight

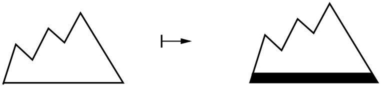
$$w_y(\mathbf{u}) = u_0^{2n} u_1^{n_1} \cdot \ldots \cdot u_M^{n_M}$$

- generating functions

$$D(\mathbf{u}) = \sum_{d \in \mathcal{D}} w_d(\mathbf{u}), \qquad A(\mathbf{u}) = \sum_{a \in \mathcal{A}} w_a(\mathbf{u})$$

Dyck paths: moments of height

- path lifting:



$$A(\mathbf{u}) = u_0 u_1 \cdot \ldots \cdot u_M D(\mathbf{v}(\mathbf{u})),$$

where $\mathbf{v}(\mathbf{u}) = (v_0(\mathbf{u}), v_1(\mathbf{u}), \ldots, v_M(\mathbf{u}))$ is given by

$$v_0(\mathbf{u}) = u_0 u_1^2 \cdot \ldots \cdot u_M^2$$

$$v_k(\mathbf{u}) = \prod_{l=k}^{M} u_l^{\binom{l}{k}} \qquad (k = 1, \ldots, M)$$

Dyck paths: moments of height

moments of height n_ℓ $(1 \leq \ell \leq M)$

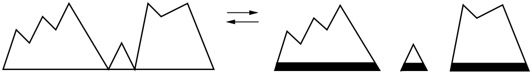
$$\overline{n}_\ell = \sum_{i=0}^{2n+2} \overline{y}^\ell(i) = \sum_{i=0}^{2n} (y(i)+1)^\ell = \sum_{i=0}^{2n} \sum_{k=0}^{\ell} \binom{\ell}{k} y(i)^k$$
$$= (2n+1) + \sum_{k=1}^{\ell} \binom{\ell}{k} \sum_{i=0}^{2n} y(i)^k = (2n+1) + \sum_{k=1}^{\ell} \binom{\ell}{k} n_\ell$$

e.g. for $M = 2$ we obtain

$$w_{\overline{d}}(u_0, u_1, u_2) = u_0^{n+1} u_1^{\overline{n}_1} u_2^{\overline{n}_2} = u_0^{n+1} u_1^{2n+1+n_1} u_2^{2n+1+n_1+n_2}$$
$$= u_0 u_1 u_2 (u_0 u_1^2 u_2^2)^n (u_1 u_2)^{n_1} u_2^{n_2}$$

Dyck paths: moments of height

- arch decomposition:

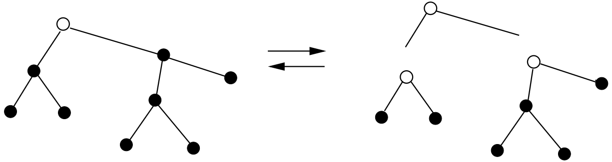


$$D(\mathbf{u}) = \frac{1}{1 - A(\mathbf{u})}$$

(sequence construction: length, height moments additive)

- q-quadratic functional equation
- no explicit expression known

ordered binary rooted trees

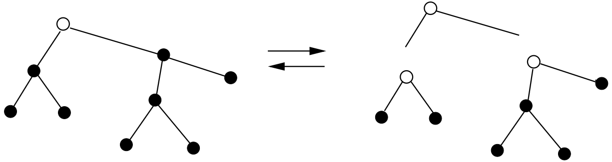


- tree: finite connected graph without cycles
- rooted: marked node, binary: internal nodes outdegree 2
- decomposition: tree $\cong$ ordered pair of trees

counting parameter:

- $n(T)$ # nodes of tree T
- $w_T(x) = x^{n(T)}$ weight of T

ordered binary rooted trees



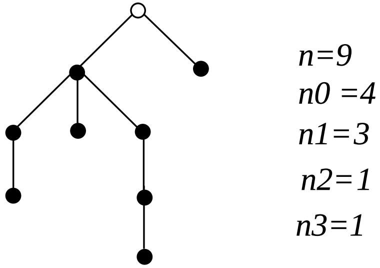
decomposition:

- parameter: $n(T) = n(T_1) + n(T_2) + 1$
- weight: $w_T(x) = x w_{T_1}(x) w_{T_2}(x)$
- quadratic equation for generating function

$$G(x) = \sum_T w_T(x) = x + x \sum_{T_1, T_2} w_{T_1}(x) w_{T_2}(x) = x + xG^2(x)$$

general binary rooted trees: Flajolet–Sedgewick I.5.2

simply generated trees (Meir, Moon 78)



ordered (plane) rooted trees T

- $n(T)$ number of nodes
- $n_k(T)$ number of nodes of outdegree k
- weight $w_T(x) = x^{n(T)} \prod_{k \geq 0} \varphi_k^{n_k(T)}$

simply generated trees

- generating function of φ_k

$$\Phi(t) = \sum_{k\geq 0} \varphi_k t^k$$

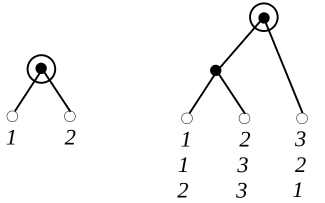
- generating function $G(x) = \sum_T w_T(x)$

$$G(x) = x \sum_{k\geq 0} \varphi_k G(x)^k = x\Phi(G(x))$$

- can be realised as conditioned Galton-Watson trees with offspring distribution $\mathbb{P}(X = k) = \varphi_k t^k / \Phi(t)$
- general rooted trees: Flajolet–Sedgewick I.5.2

phylogenetic trees

- binary rooted trees with labelled leaves



$n(T)$ # leaves of T, $n_k(T)$ # nodes of outdegree k
weight $w_T(x) = x^{n(T)} \prod_{k \geq 2} \varphi_k^{n_k(T)}$

- functional equation for exponential generating function

$$G(x) = \sum_{T \in \mathcal{T}} \frac{w_T(x)}{n(T)!} = x + \frac{x}{2} G(x)^2$$

- general case: Flajolet–Sedgewick II.19

additional counting parameters I

pathlength

$$m(T) = \sum_{v \in T} d(v, o)$$

$d(v, o)$ distance from v to root
$w_T(x, q) = x^{n(T)} q^{m(T)}$ weight of T

decomposition for binary trees:

- parameter: $m(T) = (m(T_1) + n(T_1)) + (m(T_2) + n(T_2))$
- weight: $w_T(x, q) = x w_{T_1}(qx, q) w_{T_2}(qx, q)$

simply generated trees:

$$G(x, q) = x\Phi(G(qx, q))$$

additional counting parameters II

- generalised pathlength

$$m_k(T) := \sum_{v \in T} d(v,o)^k \qquad k \in \{0, \dots, M\}$$

- weight $w_T(\mathbf{u})$ of T as

$$w_T(\mathbf{u}) := u_0^{n(T)} u_1^{m_1(T)} \cdot \ldots \cdot u_M^{m_M(T)} \prod_{k \geq 0} \varphi_k^{n_k(T)}.$$

- generating function $G(\mathbf{u}) = \sum_T w_T(\mathbf{u})$
- q-functional equation

$$G(\mathbf{u}) = u_0 \Phi(G(\mathbf{v}(\mathbf{u}))), \qquad v_k(\mathbf{u}) = \prod_{l=k}^{M} u_l^{\binom{l}{k}} \qquad (k = 0, \dots, M)$$

non-linear parameters: Wiener index (Janson 03)

$$W(T) = \frac{1}{2} \sum_{v,w \in T} d(v, w)$$

- appears in chemistry of acyclic molecules
- analyse the simpler quantity

$$Q(T) = \sum_{v,w \in T} d(v \wedge w, o) = n(T)m(T) - W(T)$$

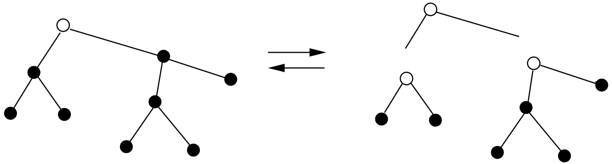
$v \wedge w$ last common ancestor, $n(T)$ # nodes, $m(T)$ pathlength

- this follows from

$$\begin{aligned} d(v, w) &= d(v, v \wedge w) + d(v \wedge w, w) \\ &= d(v, o) - d(v \wedge w, o) + d(w, o) - d(v \wedge w, o) \end{aligned}$$

non-linear parameters: Wiener index (Janson 03)

decomposition for binary trees:



this implies for the counting parameters

- $m(T) = (m(T_1) + n(T_1)) + (m(T_2) + n(T_2))$
- $Q(T) = (Q(T_1) + n(T_1)^2) + (Q(T_2) + n(T_2)^2)$

modified generating function approach possible! (Wagner 12)

discrete meanders with bounded steps (Banderier–Flajolet 02)

discrete meanders

- walks with unit steps in x-direction
- steps in y-direction from $\mathcal{S} \subseteq \{-c, -c+1, \ldots, d-1, d\} \subset \mathbb{Z}$
- c, d positive, $-c, d \in \mathcal{S}$
- meanders require non-negative height

generating functions

- step polynomial $S(u) = \sum_{s_i \in \mathcal{S}} s_i u^i$ with step weights $s_i > 0$
- $F(z, u)$ perimeter and final height gf

discrete meanders with bounded steps

decomposition of meanders

- either empty path or
- meander path with added step, but ...
- correct for steps which fall below $y = 0$

this translates into the functional equation

$$F(z,u) = 1 + zS(u)F(z,u) - z\{u^{<0}\}S(u)F(z,u)$$
$$= 1 + zS(u)F(z,u) - z\sum_{i=0}^{c-1} r_i(u)G_i(z)$$

where $r_i(u) = \left(s_{-c}u^{-c} + \ldots + s_{-(i+1)}u^{-(i+1)}\right)u^i$ and $G_i(z) = [u^i]F(z,u)$

remarks

- single equation with $(c+1)$ unknowns $F(z,u)$, $G_0(z), \ldots, G_{c-1}(z)$
- area by q-shift

discrete meanders: kernel method

$$F(z,u)(1-zS(u)) = 1 - z\sum_{i=0}^{c-1} r_i(u)G_i(z)$$

solve kernel equation $1 - zS(u) = 0$

- c small branches $u_1(z),\dots,u_c(z)$: $u_i(z) \sim c_i z^{\frac{1}{c}}$ as $z \to 0$
- d large branches $v_1(z),\dots,v_d(z)$: $v_i(z) \sim d_i z^{-\frac{1}{d}}$ as $z \to 0$

kernel method: consider $N(z,u) = u^c \cdot rhs(z,u)$

- $N(z,u)$ polynomial in u of deg c with roots $u_1(z),\dots,u_c(z)$
- leading coefficient comparison gives $N(z,u) = \prod_{j=1}^{c}(u - u_j(z))$

we get the (formal) power series solution

$$F(z,u) = \frac{N(z,u)}{u^c(1-zS(u))}$$

prudent polygons by perimeter (Schwerdtfeger 10)

two-sided polygons (bar-graph polygons)

- functional equation solvable by kernel method

$$P_2(x,u) = \frac{1 - x - u(1+x)x - \sqrt{x^2(1-x)^2u^2 - 2x(1-x^2)u + (1-x)^2}}{2xu}$$

three-sided polygons

- functional equation solvable by kernel method

$$P_3(x) = \sum_{k\geq 0} L((xq^2)^k) \prod_{j=0}^{k-1} K((xq^2)^j), \qquad q = \frac{x^2 + 1 - \sqrt{1 - 4x + 2x^2 + x^4}}{2x}$$

$$K(w) = \frac{(1-x)q - 1 - ((1 - x + x^2)q - 1)(P_2(x,qw) + x)w}{1 - x(1+x)q - (x(1 - x - x^3)q + x^2)(P_2(x,qw) + x)w}$$

$$L(w) = \frac{(1 + x^2 - (1 - 2x + 2x^2 + x^4)q)(P_2(x,qw) + x)w}{1 - x(1+x)q - (x(1 - x - x^3)q + x^2)(P_2(x,qw) + x)w}$$

general prudent polygons

- functional equation unsolved (three auxiliary variables)

three-choice polygons (Guttmann–Jensen 05)

- exact enumeration data for half-perimeter gf $P(x)$ suggests

$$\sum_{k=0}^{8} p_k(x) \frac{d^k}{dx^k} P(x) = 0,$$

with

$$p_8(x) = x^3(1-4x)^4(1+4x)(1+4x^2)(1+x+7x^2)q_8(x)$$
$$p_7(x) = x^2(1-4x)^3 q_7(x),\ p_6(x) = x(1-4x)q_6(x)$$
$$p_5(x) = (1-4x)q_5(x),\ p_4(x) = q_4(x)$$
$$p_3(x) = q_3(x),\ p_2(x) = x(1-2x)q_2(x)$$
$$p_1(x) = (1-4x)q_1(x),\ p_0(x) = q_0(x)$$

$q_8(x), q_7(x), \dots, q_0(x)$ known polynomials of degree $25, 31, 32, 33, 33, 32, 29, 29, 29$, which do not factorise

- 206 terms are needed for this equation, 260 terms checked