Combinatorics, Probability and Computing (2014) 23, 725–748. © Cambridge University Press 2014
doi:10.1017/S0963548313000643

On the Importance Sampling of Self-Avoiding Walks

MIREILLE BOUSQUET-MÉLOU

MBM: CNRS, LaBRI, Université Bordeaux 1, 351 Cours de la Libération, 33405 Talence, France
(e-mail: mireille.bousquet@labri.fr)

Received 6 September 2012; revised 14 February 2013; first published online 20 January 2014

In a 1976 paper published in *Science*, Knuth presented an algorithm to sample (non-uniform) self-avoiding walks crossing a square of side k. From this sample, he constructed an estimator for the number of such walks. The quality of this estimator is directly related to the (relative) variance of a certain random variable X_k. From his experiments, Knuth suspected that this variance was extremely large (so that the estimator would not be very efficient). But how large? For the analogous Rosenbluth algorithm, which samples *unconfined* self-avoiding walks of length n, the variance of the corresponding estimator is believed to be exponential in n.

A few years ago, Bassetti and Diaconis showed that, for a sampler *à la Knuth* that generates walks crossing a $k \times k$ square and consisting of North and East steps, the relative variance is only $O(\sqrt{k})$. In this note we take one step further and show that, for walks consisting of North, South and East steps, the relative variance jumps to $2^{k(k+1)}/(k+1)^{2k}$. This is exponential in the average length of the walks, which is of order k^2. We also obtain partial results for general self-avoiding walks crossing a square, suggesting that the relative variance could be exponential in k^2 (which is again the average length of these walks).

Knuth's algorithm is a basic example of a widely used technique called *sequential importance sampling*. The present paper, following the paper by Bassetti and Diaconis, is one of very few examples where the variance of the estimator can be found.

2010 *Mathematics subject classification*: Primary 05A15
Secondary 60C05

1. Introduction

A *self-avoiding walk* (SAW) on a graph is a walk that never visits the same vertex twice. Let $\mathcal{W}_k$ be the set of SAWs on a $k \times k$ square grid, going from the South-West vertex to the North-East vertex (Figure 2). In his paper 'Coping with finiteness' [14, 15], Knuth described the following algorithm to generate a (non-uniform) random walk of $\mathcal{W}_k$. Start from the South-West corner, and at each time, choose with equal probability (which can be 1/3, 1/2 or 1) one of the *eligible* steps. A step is *eligible* if, once appended to the current walk, it gives a self-avoiding walk that can be extended so as to end at the

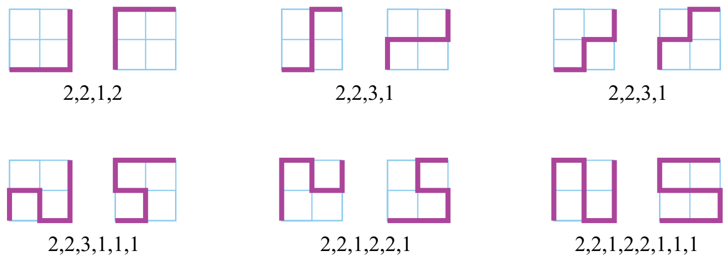


Figure 1. (Colour online) The twelve self-avoiding walks crossing the 2×2 square. For each one we give the sequence $1/p_1, 1/p_2, \dots$ where p_i is the probability of the ith step. The probability of the walk is thus the reciprocal of the product of the terms in the list. Two walks have probability 1/8, six have probability 1/12, and four have probability 1/16. Two walks that differ by a diagonal symmetry have the same probability.

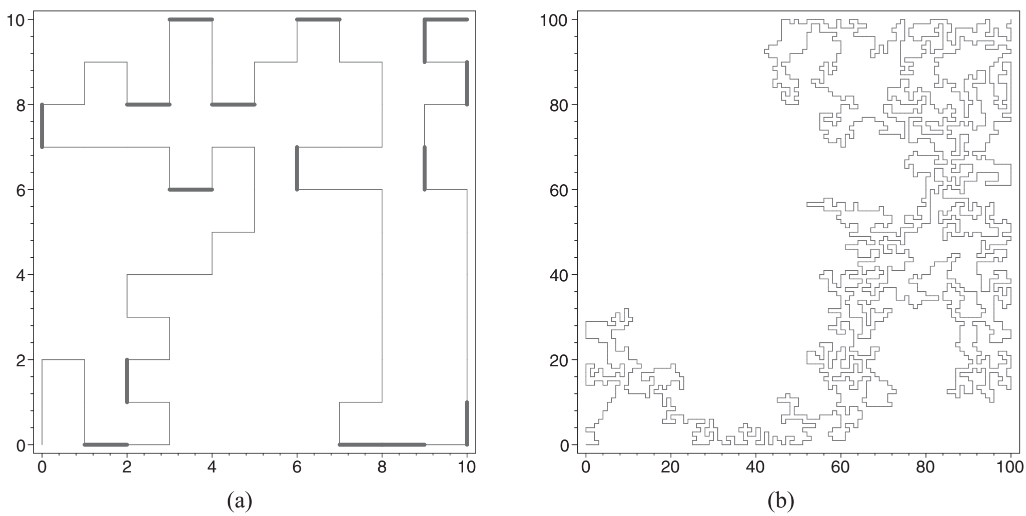


Figure 2. (a) A SAW crossing the 10×10 square. The thick steps have probability 1, that is, each of them is the only eligible step at the time when it is taken. (b) A SAW crossing the 100×100 square, obtained via Knuth's algorithm.

North-East corner. In this way the walk is never trapped and the algorithm always succeeds.[1] Figure 1 shows the probabilities of the 12 possible walks when $k = 2$. Two bigger examples ($k = 10$, $k = 100$) are shown in Figure 2. This procedure is a basic example of a widely used technique called *sequential importance sampling* [5, 10, 11]. It is also a variant, for walks confined to a square, of the Rosenbluth algorithm, which generates unconfined SAWs [21].

Let $p(w)$ denote the probability of drawing the walk $w \in \mathcal{W}_k$. Consider the random variable $X_k = 1/p(w)$, where w is a random walk of $\mathcal{W}_k$ drawn according to the distribution $p(\cdot)$. Clearly,

$$\mathbb{E}(X_k) = \sum_{w \in \mathcal{W}_k} p(w) \frac{1}{p(w)} = |\mathcal{W}_k|.$$

[1] We describe in Section 5.3 how to detect algorithmically when a new step traps the walk.

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

Hence one can unbiasedly estimate the number of SAWs crossing a $k \times k$ square by generating N walks $w^{(1)}, \ldots, w^{(N)}$ of $\mathcal{W}_k$, and computing

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{p(w^{(i)})}. \tag{1.1}$$

By generating 'several thousand' walks for $k = 10$, Knuth obtained

$$|\mathcal{W}_{10}| \simeq (1.6 \pm 0.3) \times 10^{24},$$

which is quite good given the now-known exact value, $1,568,758,030,464,750,013,214,100$ (see [7, 15]). We have reproduced Knuth's experiment, and found, with a first group of 10 000 walks, the estimate 1.78×10^{24}, and with a second group, the estimate 1.38×10^{24}. As observed by Knuth, the values $1/(p(w^{(i)}))$ vary a lot (a small sample of ten walks gave us values ranging from 10^{11} to 10^{24}), and one may suspect that the variance of X_k is probably much larger than $\mathbb{E}(X_k)^2$, or, in other words, that the *relative variance*

$$\mathrm{Var}\left(\frac{X_k}{\mathbb{E}(X_k)}\right)$$

is large. Note that

$$\mathrm{Var}(X_k) = \mathbb{E}(X_k^2) - \mathbb{E}(X_k)^2 = \sum_{w \in \mathcal{W}_k} \frac{1}{p(w)} - |\mathcal{W}_k|^2.$$

Also, observe that the variance of the estimator (1.1) is $\mathrm{Var}(X_k)/N$, so that $\mathrm{Var}(X_k)$ is a measure of the quality of this estimator. Let us mention that, even though sequential importance sampling is widely used, no general bounds on the variance of the estimators are available.

Knuth's observation led Bassetti and Diaconis to study a simpler algorithm, in which a step, in order to be eligible, has to go North (N) or East (E) [5]. The resulting walk is called a *directed* walk, or NE-walk. Each step has probability $1/2$ unless it follows the North or East side of the square, in which case it has probability 1. Let $\mathcal{D}_k$ denote the set of directed walks of $\mathcal{W}_k$, and let $p(w)$ be the probability of generating the directed walk w with this new algorithm. Define the random variable $X_k = 1/p(w)$ as above. Then

$$\mathbb{E}(X_k) = |\mathcal{D}_k| = \binom{2k}{k} \sim \frac{4^k}{\sqrt{\pi k}}.$$

Of course, since $|\mathcal{D}_k|$ is known exactly, there is no point in using importance sampling to estimate this cardinality, but it is interesting to know that the variance of the estimator can be determined, as follows. By the above argument, a walk of $\mathcal{D}_k$ that hits the North or East side of the square for the first time at time $k+i$ has probability $1/2^{k+i}$. Since there are $2\binom{k+i-1}{i}$ such walks,

$$\mathbb{E}(X_k^2) = \sum_{w \in \mathcal{D}_k} \frac{1}{p(w)} = \sum_{i=0}^{k-1} 2^{k+i+1} \binom{k+i-1}{i}.$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

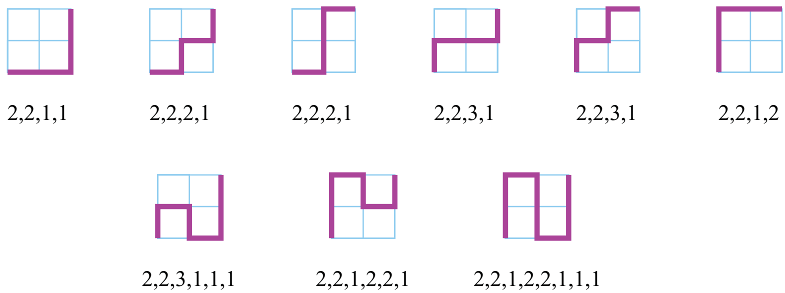


Figure 3. (Colour online) The nine NES-walks crossing a 2×2 square, with the reciprocals of the probabilities of their steps.

The corresponding generating function is

$$\sum_{k \geqslant 1} \mathbb{E}(X_k^2) x^k = \frac{2\,x}{1+2x}\left(\frac{3}{\sqrt{1-16\,x}} - 1\right), \tag{1.2}$$

and an elementary singularity analysis [12, Chapter VI] gives

$$\mathbb{E}(X_k^2) \sim \frac{16^k}{3\sqrt{\pi k}},$$

which is roughly $\sqrt{k}$ times larger than

$$\mathbb{E}(X_k)^2 \sim \frac{16^k}{\pi k}.$$

In this note, we first take one more step in the direction of the general problem by declaring that South steps are also eligible. The resulting walks are *partially directed* walks, or NES-walks. The probabilities of the 9 walks obtained when $k = 2$ are shown in Figure 3. Of course, these probabilities are not the same as those obtained from Knuth's original algorithm.

We will prove that one outcome of this increased generality is that the ratio between $\mathbb{E}(X_k^2)$ and $\mathbb{E}(X_k)^2$ becomes much larger:

$$\mathbb{E}(X_k^2) \sim \frac{3}{2}\, 2^{k(k+1)} \quad \text{while} \quad \mathbb{E}(X_k)^2 = (k+1)^{2k}.$$

We will also prove that the average length of a (uniform) NES-walk confined to the $k \times k$ square is quadratic in k, so that the variance of X_k is exponential in the length, as predicted for SAWs generated by the Rosenbluth algorithm [6].

Since the x/y symmetry is lost with partially directed walks, it is natural to generalize the original question by enclosing walks in a rectangle R of height k and width ℓ. Thus, let $\mathcal{P}_{k,\ell}$ be the set of partially directed walks that start from the South-West corner of R and end at the North-East corner. A walk of $\mathcal{P}_{k,\ell}$ contains exactly ℓ East steps, and choosing the heights of these steps determines the walk completely. Hence the number of walks in $\mathcal{P}_{k,\ell}$ is $(k+1)^\ell$.

Thus, defining the random variable $X_{k,\ell} = 1/p(w)$ as above, we have

$$\mathbb{E}(X_{k,\ell}) = |\mathcal{P}_{k,\ell}| = (k+1)^\ell. \tag{1.3}$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

We will prove that, if $k, \ell \to \infty$ in such a way that $\ell = o(2^k)$, then

$$\mathrm{Var}(X_{k,\ell}) \sim \mathbb{E}(X_{k,\ell}^2) \sim \frac{3}{2}\, 2^{(k+1)\ell},$$

so that the relative variance satisfies

$$\mathrm{Var}\left(\frac{X_{k,\ell}}{\mathbb{E}(X_{k,\ell})}\right) \sim \frac{3}{2}\left(\frac{2^{k+1}}{(k+1)^2}\right)^{\ell}.$$

This result thus extends the very short list of examples where the variance of an importance sampler can be rigorously established [5]. Moreover, the average length of a (uniform) walk of $\mathcal{P}_{k,\ell}$ is shown to be approximately $k\ell/3$, so that the variance is again exponential in the length, as expected for other similar samplers.

The paper is organized as follows. In Section 2, we obtain an explicit expression for the generating function of the numbers $\mathbb{E}(X_{k,\ell}^2)$ (the counterpart of (1.2)). In Section 3, we derive from this expression the above asymptotic result. In Section 4, we go back to Knuth's sampler and prove that there exist two positive constants λ and β such that

$$\mathbb{E}(X_k)^{1/k^2} \to \lambda \quad \text{and} \quad \mathbb{E}(X_k^2)^{1/k^2} \to \beta.$$

The former result has actually been known since 1978 [1]. Since a variance is non-negative, $\beta \geqslant \lambda^2$. Upper (and lower) bounds on λ have been obtained in [7], based on the determination of the numbers $\mathbb{E}(X_k) = |\mathcal{W}_k|$ for small values of k, and of related numbers counting other configurations of self-avoiding walks. As shown in Section 4, a similar study, performed for the numbers $\mathbb{E}(X_k^2)$, might suffice to prove that $\beta > \lambda^2$, so that the variance would again be exponential in k^2 (which is known to be the average length of a uniform SAW crossing the $k \times k$-square [16]). We conclude with a few remarks and questions on the importance sampling of self-avoiding walks not confined to a box.

2. Exact results for NES-walks

In this section, we first describe the probability $p(w)$ of obtaining the walk w in terms of the geometry of w (Section 2.1). This description reduces the determination of the numbers $\mathbb{E}(X_{k,\ell}^2)$ to the enumeration of NES-walks according to several parameters, which we perform in Section 2.2.

2.1. The probability $p(w)$

Let w_0 be a walk of $\mathcal{P}_{k,\ell}$, written as a sequence of N, E and S steps. Let w be the prefix of w_0 that precedes the last E step. That is, $w_0 = w\mathsf{E}\mathsf{N}\cdots\mathsf{N}$. By convention, w_0 starts at height 0.

Lemma 2.1. *The probability $p(w_0)$ of obtaining w_0 via the importance sampling algorithm satisfies*

$$\frac{1}{p(w_0)} = 2 \cdot 3^{h(w)} 2^{h_c(w)} 2^{v(w)} 1^{v_c(w)},$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

where

- $h(w)$ *is the number of horizontal steps of* w *that lie neither at height* 0 *nor at height* k,
- $h_c(w)$ *is the number of horizontal contacts of* w, *that is, horizontal steps that lie at height* 0 *or* k,
- $v(w)$ *is the number of vertical steps of* w *that end neither at height* 0 *nor at height* k,
- $v_c(w)$ *is the number of vertical contacts of* w, *that is, vertical steps that end at height* 0 *or* k.

Proof. Assume the walk w_0 has length n and ends with exactly j vertical steps. The probability of the first step is $1/2$, and the probability of each of the j final steps is 1. Let s_i denote the ith step. Hence w consists of the steps $s_1, \dots, s_{n-j-1}$. For $1 \leqslant i < n-j$, the probability of s_{i+1} depends on the direction and position of s_i:

- if s_i is horizontal, but not a contact, then the probability of s_{i+1} is $1/3$,
- if s_i is a horizontal contact, then the probability of s_{i+1} is $1/2$,
- if s_i is vertical, but not a contact, then the probability of s_{i+1} is $1/2$,
- if s_i is a vertical contact, then the probability of s_{i+1} is 1.

The lemma follows. □

2.2. Enumeration of NES-walks in a strip of fixed height

Recall the expression (1.3) of the numbers $\mathbb{E}(X_{k,\ell})$. For k (the height of the rectangle) fixed, the generating function of the numbers $\mathbb{E}(X_{k,\ell})^2$ is *rational*:

$$\sum_{\ell \geqslant 1} \mathbb{E}(X_{k,\ell})^2 x^\ell = \sum_{\ell \geqslant 1} (k+1)^{2\ell} x^\ell = \frac{(k+1)^2 x}{1-(k+1)^2 x}. \tag{2.1}$$

We will determine the variance of $X_{k,\ell}$ by describing the generating function of the numbers $\mathbb{E}(X_{k,\ell}^2)$, which is also rational when k is fixed.

Proposition 2.2. *For any fixed height* k, *the generating function* $M_k(x)$ *of the numbers* $\mathbb{E}(X_{k,\ell}^2)$ *is a rational series,*

$$M_k(x) := \sum_{\ell \geqslant 1} \mathbb{E}(X_{k,\ell}^2) x^\ell = 2x \frac{N_k}{G_k},$$

where N_k *and* G_k *are polynomials in* x *satisfying the same recurrence relation,*

$$N_k = (5+9x)N_{k-2} - 4N_{k-4}$$

(*and similarly for* G_k), *with initial conditions*

$$\begin{aligned} N_1 &= 2, & G_1 &= 1-4x, \\ N_2 &= 5+3x, & G_2 &= 1-9x-6x^2, \\ N_3 &= 11+9x, & G_3 &= 1-19x-18x^2, \\ N_4 &= 23+54x+27x^2, & G_4 &= 1-36x-99x^2-54x^3. \end{aligned}$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

Example. For $k = 2$,

$$\sum_{\ell \geqslant 1} \mathbb{E}(X_{2,\ell}^2) x^\ell = 2x \frac{5+3x}{1-9x-6x^2} = 10x + 96x^2 + O(x^3).$$

Figure 3 allows us to check that the coefficient of x^2 is correct:

$$96 = 4+8+8+12+12+8+12+16+16.$$

The generating function of the variances is

$$\sum_{\ell \geqslant 1} \mathrm{Var}(X_{2,\ell}) x^\ell = 2x \frac{5+3x}{1-9x-6x^2} - \frac{9x}{1-9x}.$$

Observe that the radius of the first fraction is smaller than the radius of the second fraction. As $\ell \to \infty$,

$$\mathbb{E}(X_{2,\ell}^2) \sim \mu^\ell$$

(up to a multiplicative constant) with $\mu = (9+\sqrt{105})/2 \simeq 9.62$, while $\mathbb{E}(X_{2,\ell})^2 = 9^\ell$.

We now want to prove Proposition 2.2. Recall that

$$\mathbb{E}(X_{k,\ell}^2) = \sum_{w_0 \in \mathcal{P}_{k,\ell}} \frac{1}{p(w_0)}.$$

The expression of $p(w_0)$ given in Lemma 2.1 leads us to study a purely enumerative problem. For k fixed, let $\mathcal{T}_k$ be the set of NES-walks w that start at height 0 and are confined to the strip $0 \leqslant y \leqslant k$. We wish to count these walks by the parameters $h(w)$, $h_c(w)$, $v(w)$ and $v_c(w)$. So, let

$$\sum_{w \in \mathcal{T}_k} x^{h(w)} y^{v(w)} a^{h_c(w)} b^{v_c(w)}$$

be the associated generating function. This series is easily seen to be rational (it can be determined using a transfer-matrix approach [22, Section 4.7], or equivalently a finite-state automaton [12, p. 362]), and there are several ways to determine it. We present here what we believe to be the most direct one. It relies on a recursive description of the walks of $\mathcal{T}_k$, where we add at each time an E step and a sequence of vertical steps. This approach requires us to take into account an additional parameter, namely the height $f(w)$ of the final point of the walk w. Hence our series finally involves five variables:

$$T_k(s) \equiv T_k(x,y,a,b;s) = \sum_{w \in \mathcal{T}_k} x^{h(w)} y^{v(w)} a^{h_c(w)} b^{v_c(w)} s^{f(w)}.$$

We will denote by $\tilde{\mathcal{T}}_k$ the subset of $\mathcal{T}_k$ formed of walks that do not end at height 0 or k, and by $\tilde{T}_k(s) \equiv \tilde{T}_k(x,y,a,b;s)$ the corresponding generating function. Accordingly,

$$T_k(s) = \sum_{i=0}^{k} T_{k,i} s^i = T_{k,0} + \tilde{T}_k(s) + s^k T_{k,k},$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

where $T_{k,i}$ is the series in x, y, a and b counting walks of $\mathcal{T}_k$ ending at height i. By Lemma 2.1,

$$\begin{aligned} M_k(x) = \sum_{\ell \geqslant 1} \mathbb{E}(X_{k,\ell}^2) x^\ell &= \sum_{\ell \geqslant 1} \sum_{w_0 \in \mathcal{P}_{k,\ell}} \frac{1}{p(w_0)} x^\ell \\ &= 2 \sum_{w \in \mathcal{T}_k} 3^{h(w)} 2^{h_c(w)} 2^{v(w)} x^{1+h(w)+h_c(w)} \\ &= 2x\, T_k(3x, 2, 2x, 1; 1). \end{aligned} \tag{2.2}$$

Remark. The series $T_{k,k}$ has already been determined in the case $x = y = a = b = t$, using the same approach as here [3, Proposition 3]. The derivation is more involved here because we keep track of four parameters in the enumeration, and because we are interested in $T_k(1)$ rather than $T_{k,k}$.

Lemma 2.3. *The series $\tilde{T}_k(s)$, $T_{k,0}$ and $T_{k,k}$ satisfy the following system of equations:*

$$\begin{aligned} \left(1 - \frac{x}{1-ys} - \frac{xy\bar{s}}{1-y\bar{s}}\right) \tilde{T}_k(s) &= \frac{ys - (ys)^k}{1-ys} - \frac{x(ys)^k}{1-ys} \tilde{T}_k(1/y) \\ &\quad - \frac{x}{1-y\bar{s}} \tilde{T}_k(y) + aT_{k,0} \frac{ys-(ys)^k}{1-ys} + aT_{k,k} \frac{ys^{k-1} - y^k}{1-y\bar{s}}, \\ T_{k,0} &= 1 + bx\bar{y}\tilde{T}_k(y) + aT_{k,0} + aby^{k-1} T_{k,k}, \\ T_{k,k} &= by^{k-1} + bxy^{k-1} \tilde{T}_k(1/y) + aby^{k-1} T_{k,0} + aT_{k,k}, \end{aligned}$$

with $\bar{s} = 1/s$ and $\bar{y} = 1/y$.

Proof. We construct the walks of $\tilde{\mathcal{T}}_k$ recursively, by adding at each time a horizontal step followed by a sequence of vertical steps.

We partition the set $\tilde{\mathcal{T}}_k$ into three disjoint subsets, illustrated in Figure 4.

- The first subset consists of walks with no E step. These walks consist of i North steps, with $1 \leqslant i < k$. Their generating function is

$$\sum_{i=1}^{k-1} (ys)^i = \frac{ys - (ys)^k}{1-ys}.$$

- The second subset consists of walks in which the last E is followed by a (possibly empty) sequence of N steps. We let i denote the height of the last E step, and distinguish the cases $i = 0$ and $0 < i < k$. The generating function of this subset of $\tilde{\mathcal{T}}_k$ reads

$$\begin{aligned} &aT_{k,0} \sum_{j=1}^{k-1} (sy)^j + x \sum_{i=1}^{k-1} \left(T_{k,i} s^i \sum_{j=0}^{k-i-1} (ys)^j \right) \\ &= aT_{k,0} \frac{ys - (ys)^k}{1-ys} + x \sum_{i=1}^{k-1} \left(T_{k,i} s^i \frac{1-(ys)^{k-i}}{1-ys} \right) \\ &= aT_{k,0} \frac{ys - (ys)^k}{1-ys} + \frac{x}{1-ys} \left(\tilde{T}_k(s) - (ys)^k \tilde{T}_k(1/y) \right). \end{aligned}$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

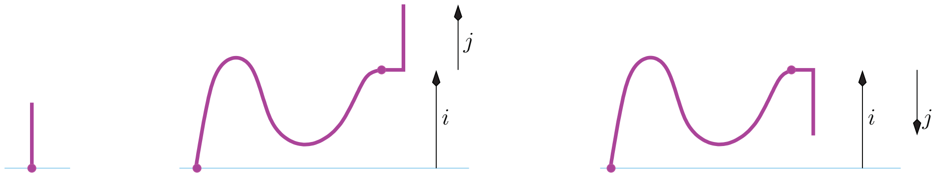


Figure 4. (Colour online) Recursive construction of bounded NES-walks.

- The third subset consists of walks in which the last E step is followed by a non-empty sequence of S steps. We let i denote the height of the last E step, and distinguish the cases $i = k$ and $0 < i < k$. The generating function of this subset of $\tilde{T}_k$ reads

$$\begin{aligned} as^k T_{k,k} \sum_{j=1}^{k-1} (y\bar{s})^j &+ x \sum_{i=1}^{k-1} \left(T_{k,i} s^i \sum_{j=1}^{i-1} (y\bar{s})^j \right) \\ &= a T_{k,k} \frac{ys^{k-1} - y^k}{1 - y\bar{s}} + x \sum_{i=1}^{k-1} \left(T_{k,i} s^i \frac{y\bar{s} - (y\bar{s})^i}{1 - y\bar{s}} \right) \\ &= a T_{k,k} \frac{ys^{k-1} - y^k}{1 - y\bar{s}} + \frac{x}{1 - y\bar{s}} \big(y\bar{s} \tilde{T}_k(s) - \tilde{T}_k(y) \big). \end{aligned}$$

Adding the three contributions gives the series $\tilde{T}_k(s)$ and establishes the first equation of the lemma.

The equations for $T_{k,0}$ and $T_{k,k}$ are obtained in a similar fashion. □

We now solve the functional equations of Lemma 2.3. The key tool is the *kernel method* (see, *e.g.*, [4, 8, 20]).

Proposition 2.4. *Let $k \geqslant 1$. The series $T_k(1) \equiv T_k(x, y, a, b; 1)$ counting NES-walks confined to a strip of height k is*

$$T_k(1) = \frac{N_k}{G_k},$$

where N_k and G_k are polynomials in x, y, a and b satisfying the same recurrence relation,

$$N_k = (1 - x + y^2(1 + x))N_{k-2} - y^2 N_{k-4}$$

(and similarly for G_k), with initial conditions

$$\begin{aligned} N_{-1} &= (1 - x - xy)(b - y)/y^2, & G_0 &= (x-1)ab/y - (x+1)(a-1), \\ N_0 &= (b - xb + xy)/y, & G_1 &= 1 - a - ab, \\ N_1 &= 1 + b, & G_2 &= (1-x)(1-a) - (x+1)yab, \\ N_2 &= 1 - x + y + by(1+x), & G_3 &= (1 - x - xy)(1-a) - yab(x + y + xy). \end{aligned}$$

Equivalently,

$$T_k(1) = \frac{1}{P(\bar{S})S^k + P(S)} \left(Q(\bar{S})S^k + Q(S) + (1 - y^2) \frac{S^k - S}{S - 1} \right), \tag{2.3}$$

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

where S is the unique formal power series[2] *in x and y satisfying*

$$S + \frac{1}{S} = (1+x)y + (1-x)\bar{y}, \tag{2.4}$$

with $\bar{y} = 1/y$, $\bar{S} = 1/S$,

$$P(s) = 1 - a + aby - s(ab + y - ay) \quad \textit{and} \quad Q(s) = 1 - by + (b - y)s. \tag{2.5}$$

The reason why we give the expressions of N_{-1} and N_0, rather than N_3 and N_4, is that they are more compact. The same reason explains why we give G_0 rather than G_4. It is of course easy to compute N_3, N_4 and G_4, and Proposition 2.2 then follows at once, using (2.2). We hope that using the same notation N_k, G_k for the enumeration problem of Proposition 2.4 and its specialization of Proposition 2.2 will not create any confusion.

Proof. First, we use the last two equations of Lemma 2.3 to express $\tilde{T}_k(y)$ and $\tilde{T}_k(1/y)$ as linear combinations of $T_{k,0}$ and $T_{k,k}$. Then, in the first equation of the lemma, we replace $\tilde{T}_k(y)$ and $\tilde{T}_k(1/y)$ by their expressions in terms of $T_{k,0}$ and $T_{k,k}$. The left-hand side is unchanged, and the right-hand side now involves only two unknown series, namely $T_{k,0}$ and $T_{k,k}$:

$$\begin{aligned}\left(1 - \frac{x}{1-ys} - \frac{xy\bar{s}}{1-y\bar{s}}\right)\tilde{T}_k(s) \\ = \frac{ys}{1-ys} + \frac{y}{b(1-y\bar{s})} + \left(\frac{ays}{1-ys} + \frac{y(a-1)}{b(1-y\bar{s})}\right)T_{k,0} + s^k\left(\frac{y(a-1)}{b(1-ys)} + \frac{ya\bar{s}}{1-y\bar{s}}\right)T_{k,k}.\end{aligned} \tag{2.6}$$

The *kernel* of this equation is the coefficient of $\tilde{T}_k(s)$. It vanishes when $s = S$ and $s = \bar{S} := 1/S$, where S is defined in the proposition. Since $\tilde{T}_k(s)$ is a polynomial in s, and S and $\bar{S}$ are Laurent series in x and y with finitely many monomials with negative exponents, the series $\tilde{T}_k(S)$ and $\tilde{T}_k(\bar{S})$ are well-defined. Replacing s by S or $\bar{S}$ in the above equation cancels the left-hand side, and hence the right-hand side. One thus obtains two linear equations between $T_{k,0}$ and $T_{k,k}$, which involve the series S. Solving them gives expressions of $T_{k,0}$ and $T_{k,k}$ in terms of S (the expression of $T_{k,k}$ is given in (2.8) below). By setting $s = 1$ in (2.6), one then expresses $\tilde{T}_k(1)$ in terms of S, and finally $T_k(1) = T_{k,0} + T_{k,k} + \tilde{T}_k(1)$. This gives (2.3).

Observe that the expression (2.3) is unchanged if we replace S by $\bar{S} = 1/S$. In particular, it can be written as a symmetric rational function in S and $\bar{S}$ (with coefficients in $\mathbb{Q}(a, b, y)$). Since S and $\bar{S}$ are the two roots of (2.4), their symmetric functions are rational functions of x and y. This implies that $T_k(1)$ is a rational series in x, y, a and b. However, the denominator of (2.3), namely $P(\bar{S})S^k + P(S)$, is not unchanged when $S \mapsto 1/S$. But let us

[2] The other solution is $1/S$, and its expansion in x and y involves negative powers of y.

define the series N_k and G_k as follows:

$$\begin{aligned}
G_{2k} &= \frac{y^k}{1-y^2}\big(P(\bar{S})S^k + P(S)\bar{S}^k\big),\\
G_{2k+1} &= \frac{y^k}{(1-y)(1+S)}\big(P(\bar{S})S^{k+1} + P(S)\bar{S}^k\big),\\
N_{2k} &= \frac{y^k}{1-y^2}\left(Q(\bar{S})S^k + Q(S)\bar{S}^k + (1-y^2)\frac{S^k-\bar{S}^{k-1}}{S-1}\right),\\
N_{2k+1} &= \frac{y^k}{(1-y)(1+S)}\left(Q(\bar{S})S^{k+1} + Q(S)\bar{S}^k + (1-y^2)\frac{S^k-\bar{S}^k}{1-\bar{S}}\right).
\end{aligned} \tag{2.7}$$

Then it is easy to check that (2.3) can be rewritten as $T_k(1) = N_k/G_k$. Moreover, the series N_k and G_k are unchanged when $S \mapsto 1/S$, and hence, by the same argument as above, they are rational functions of x, y, a and b. More precisely, each of the sequences G_{2k}, G_{2k+1}, N_{2k} and N_{2k+1} is of the form $y^k(\alpha S^k + \beta\bar{S}^k)$, where S and $\bar{S}$ are the two roots of (2.4). Hence each sequence satisfies the recurrence relation

$$u_k = (1-x+y^2(1+x))u_{k-1} - y^2u_{k-2}.$$

One easily determines the initial values for each sequence. This yields the description of N_k and G_k given in the proposition. From this description, it is clear that N_k and G_k are polynomials, as soon as $k \geqslant 1$. □

Remarks. **(1)** *Denominators.* The series $T_{k,i}$, counting walks ending at height i, are also rational, but with a denominator that is a proper multiple of the denominator of $T_k(1) = \sum_{i=0}^k T_{k,i}$. For instance,

$$T_{k,k} = \frac{b(y^2-1)(\bar{S}-S)S^k}{(P(\bar{S})S^k - P(S))(P(\bar{S})S^k + P(S))}, \tag{2.8}$$

or, in terms of polynomials,

$$T_{k,k} = \frac{by^{k-1}}{F_k}, \tag{2.9}$$

where F_k is defined by the recurrence relation

$$F_k = (1-x+(1+x)y^2)F_{k-1} - y^2F_{k-2},$$

with the initial conditions

$$F_1 = (1-a-ab)(1-a+ab) \quad \text{and} \quad F_2 = (1-a+bya)((1-x)(1-a)-(x+1)yab).$$

It is not hard to prove that G_k, the denominator of $T_k(1)$, is a divisor of F_k. The simplification that occurs in the denominator when summing the series $T_{k,i}$ over i has recently been explained combinatorially, for slightly different walk models, by Bacher [2].

(2) *Average length.* The series $T_{k,k}(tx, t, tx, t)$ counts NES-walks crossing a strip of height k according to the length (variable t) and the width (variable x). In particular, the case $t = 1$ of (2.9) reads $T_{k,k}(x, 1, x, 1) = 1/(1-(k+1)x)$, as justified combinatorially in the Introduction. In order to determine the average length $|w|$ of a uniform NES-walk w

crossing a $k \times \ell$ rectangle, we differentiate $T_{k,k}(tx,t,tx,t)$ with respect to t, and then set $t = 1$. This gives

$$\frac{\partial}{\partial t}(T_{k,k}(tx,t,tx,t))\Big|_{t=1} = \sum_{\ell \geqslant 0} x^\ell \sum_{w \in \mathcal{P}_{k,\ell}} |w| = \frac{k}{1-x(1+k)} + \frac{x(1+k)(1+k(k+2)x/3)}{(1-x(1+k))^2},$$

so that the average length is

$$\frac{(k^2+5k+3)\ell}{3(1+k)} + \frac{k(2k+1)}{3(1+k)} = \frac{k\ell}{3} + O(k+\ell). \tag{2.10}$$

3. Asymptotic results for NES-walks

We now derive asymptotic results from the previous section. Recall that $\mathbb{E}(X_{k,\ell}) = (k+1)^\ell$, so that the generating function of the numbers $\mathbb{E}(X_{k,\ell})^2$ is given by (2.1) (for k fixed), with radius of convergence $1/(k+1)^2$. The radius of convergence of the generating function of the numbers $\mathbb{E}(X_{k,\ell}^2)$ turns out to be exponentially smaller. Our study has analogies with the study of the longest run in a binary string [12, p. 308], which also requires us to analyse (explicit) rational functions depending on an integer k.

Proposition 3.1. *Let $k \geqslant 1$. The series $M_k(x)$, given in Proposition 2.2, has a unique pole ρ_k of modulus less than $1/9$, satisfying*

$$\rho_k = \frac{1}{2^{k+1}} + \frac{9}{2\cdot 4^{k+1}} - \frac{12k-23}{2\cdot 8^{k+1}} + \frac{36k^2-54k-87/8}{16^{k+1}} + O\left(\frac{k^3}{32^k}\right). \tag{3.1}$$

As $x \to \rho_k^-$,

$$M_k(x) \sim \frac{\alpha_k}{1-x/\rho_k} \tag{3.2}$$

with

$$\alpha_k = \frac{3}{2} - \frac{9k-4}{2^{k+2}} + \frac{27k^2-48k+1}{2\cdot 4^{k+1}} - \frac{81k^3-306k^2+75k+140}{2\cdot 8^{k+1}} + O\left(\frac{k^4}{16^k}\right). \tag{3.3}$$

The second moment of $X_{k,\ell}$ satisfies, uniformly in k and ℓ,

$$\mathbb{E}(X_{k,\ell}^2) = \alpha_k \rho_k^{-\ell} + O(9^\ell k).$$

In particular, if $k,\ell \to \infty$ in such a way that $\ell = o(2^k)$, then

$$\mathrm{Var}(X_{k,\ell}) \sim \mathbb{E}(X_{k,\ell}^2) \sim \frac{3}{2}\, 2^{(k+1)\ell},$$

which is much larger than $\mathbb{E}(X_{k,\ell})^2 = (k+1)^{2\ell}$. By (2.10)*, the variance is thus exponential in the average length of a* (*uniform*) *NES-walk crossing the $k \times \ell$ rectangle.*

Proof. We proceed in four steps. We first express the series $M_k(x)$ in terms of an algebraic series S, as was done for the enumerative problem in Proposition 2.4. Then, we study the analytic properties of S. We use these properties to prove that the denominator G_k of M_k has only real roots, with one positive root ρ_k and all the other roots below

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

$-1/9$. We finally apply Cauchy's formula to extract the ℓth coefficient of $M_k(x)$, which is $\mathbb{E}(X_{k,\ell}^2)$.

Step 1: the expression of M_k. By Proposition 2.2,

$$M_k(x) = 2x\frac{N_k}{G_k},$$

where the polynomials N_k and G_k can be described either by induction, or, after performing the change of variables $x \to 3x$, $a \to 2x$, $y \to 2$ and $b \to 1$ in (2.7), by[3]

$$\begin{aligned} G_{2k} &= -\frac{2^k}{3}\left(P(1/S)S^k + P(S)S^{-k}\right), \\ G_{2k+1} &= -\frac{2^k}{1+S}\left(P(1/S)S^{k+1} + P(S)S^{-k}\right), \\ N_{2k} &= -\frac{2^k}{3}\left(Q(1/S)S^k + Q(S)S^{-k} - 3\frac{S^k - S^{-k+1}}{S-1}\right), \\ N_{2k+1} &= -\frac{2^k}{1+S}\left(Q(1/S)S^{k+1} + Q(S)S^{-k} - 3\frac{S^k - S^{-k}}{1-1/S}\right), \end{aligned} \tag{3.4}$$

where S and $1/S$ are the two power series in x satisfying

$$S + \frac{1}{S} = \frac{5+9x}{2}, \tag{3.5}$$

or equivalently,

$$x = -\frac{1}{9}(2S-1)(2/S-1). \tag{3.6}$$

The polynomials $P(s)$ and $Q(s)$ are

$$P(s) = 1 + 2x - 2s(1-x) \quad \text{and} \quad Q(s) = -1 - s,$$

so that, in view of (3.6),

$$P(S) = \frac{(2S-1)(2S^2 - 11S - 4)}{9S} \quad \text{and} \quad P(1/S) = \frac{(2-S)(2-11S-4S^2)}{9S^2}.$$

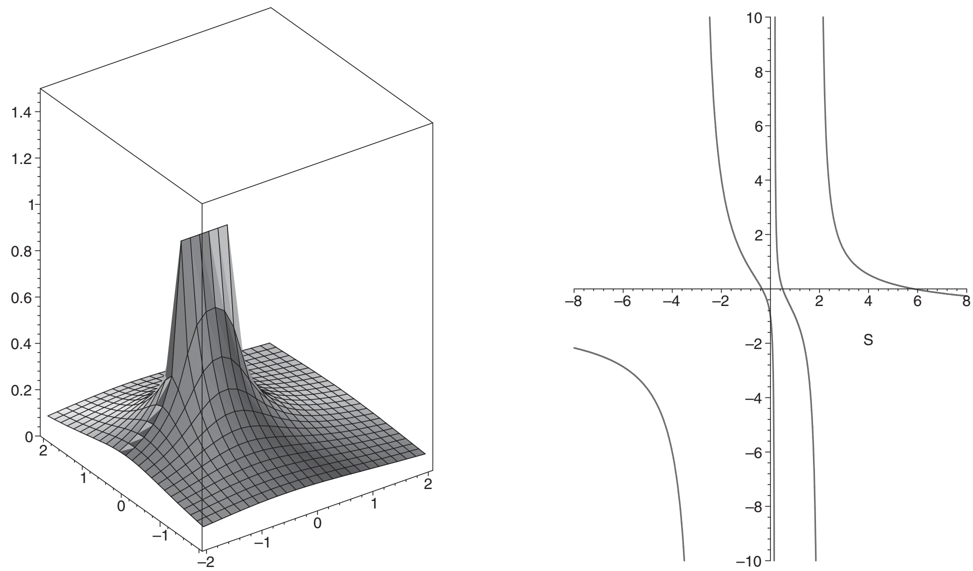
It also follows from (2.2) and (2.3) that

$$M_k(x) = \frac{2x}{P(\bar{S})S^k + P(S)}\left(Q(\bar{S})S^k + Q(S) - 3\,\frac{S^k - S}{S-1}\right). \tag{3.7}$$

Step 2: the series $S(x)$. From now on we let S denote the root of (3.5) that has constant term $1/2$:

$$S = \frac{5 + 9x - 3\sqrt{(1+x)(1+9x)}}{4}. \tag{3.8}$$

[3] From now on, we carefully avoid the notation $\bar{S} := 1/S$, since we will soon be doing complex analysis.



(a) (b)

Figure 5. (a) Plot of the modulus of S, showing the cut on the interval $[-1,-1/9]$. (b) The function $R(s)$.

Lemma 3.2. *The series S has radius of convergence $1/9$, and admits an analytic continuation, still denoted by S, in $\mathbb{C}\setminus[-1,-1/9]$. In this domain, S never vanishes, and its modulus is less than* 1.

Proof. The existence of an analytic continuation follows from basic complex analysis. If $x = u+iv$, the imaginary part of the discriminant $(1+x)(1+9x)$ reads $2v(5+9u)$. Using the principal determination of the square root, the analytic continuation of S is given by (3.8) when $\Re(x) \geqslant -5/9$, and otherwise by

$$S = \frac{5+9x+3\sqrt{(1+x)(1+9x)}}{4}.$$

A plot of the modulus of S is shown in Figure 5(a). □

Step 3: the roots of G_k

Lemma 3.3. *For $k \geqslant 1$, the denominator G_k of the series M_k has only real roots. It has a unique positive zero ρ_k, which, as $k \to \infty$, admits the expansion* (3.1)*. The other zeros are smaller than $-1/9$. As x approaches ρ_k, the series M_k behaves likes $\alpha_k/(1-x/\rho_k)$, where α_k admits the expansion* (3.3).

We could use Rouché's theorem to prove that, for any $\varepsilon > 0$, the polynomial G_k has only one root of modulus less than $1/9-\varepsilon$ for k large enough, but the above statement is more precise.

Proof. Since the case $k=1$ is trivial ($G_1 = 1-4x$), we focus on the case $k \geqslant 2$. By Proposition 2.2, the denominator G_k has degree $\lceil \frac{k+1}{2} \rceil$. The expressions (3.4) are symmetric in S and $1/S$, and thus hold for any determination of S, and thus for any $x \in \mathbb{C}$, including in the cut $[-1,-1/9]$. They show that $G_k(x)=0$ if and only if $S \neq -1$ and

$$S^{k-1} = -\frac{(2S-1)(2S^2-11S-4)}{(2-S)(2-11S-4S^2)}.$$

Conversely, if $s \in \mathbb{C} \setminus \{-1\}$ is a root of

$$s^{k-1} = -\frac{(2s-1)(2s^2-11s-4)}{(2-s)(2-11s-4s^2)} := R(s), \tag{3.9}$$

then

$$x := -\frac{1}{9}(2s-1)(2/s-1) \tag{3.10}$$

is a root of G_k. Observe that in this case, $1/s$ is also a root of (3.9), and gives rise to the same root x of G_k. Conversely, if two distinct roots s_0 and s_1 of (3.9) give rise to the same root of G_k, then $s_1 = 1/s_0$.

It is easy to relate the positions of s and x in the complex plane. By writing $s = u+iv$, one finds that x is real if and only if s is real or has modulus 1. If $s = \mathrm{e}^{\mathrm{i}\theta}$, then $x = (4\cos\theta - 5)/9$ lies in $[-1,-1/9]$. If s is real and negative, then $x \leqslant -1$, and the equality holds if and only if $s=-1$. If s is real and positive, then $x \geqslant -1/9$, and $x>0$ if and only if $s \notin [1/2, 2]$.

Since we want to prove that G_k has only real roots, let us study the roots of (3.9), distinct from -1, that are real or have modulus 1. We will prove that (3.9) has

- two pairs $\{s, 1/s\}$ of real zeros distinct from -1, one positive outside $[1/2,2]$, and one negative,
- $\lceil \frac{k-3}{2} \rceil$ pairs of zeros distinct from -1 on the unit circle.

Consequently, G_k has two real zeros outside the interval $[-1,-1/9]$, one positive, one less than -1, and $\lceil \frac{k-3}{2} \rceil$ zeros in $[-1,-1/9]$. In particular, all its roots are real.

Real roots of (3.9). An elementary study of the function $R(s)$, for $s \in \mathbb{R}$, reveals that it consists of four decreasing branches, shown in Figure 5(b), with vertical asymptotes at

$$s = -\frac{11+3\sqrt{17}}{8} \simeq -2.9, \quad s = \frac{-11+3\sqrt{17}}{8} \simeq 0.17, \quad \text{and} \quad s=2.$$

The branches intersect the s-axis at the reciprocals of these three values (and in particular at 1/2). Thus in $\mathbb{R}^+$, the equation $s^{k-1} = R(s)$ has two roots, one below 1/2 and the other beyond 2, which are necessarily the reciprocal of each other. The smallest of these increases to 1/2 as k increases: thus the corresponding value of x decreases to 0 as k increases. We denote this root of G_k by ρ_k.

If k is even, the equation $s^{k-1} = R(s)$ also has two roots in $\mathbb{R}^-$. If $k \geqslant 3$ is odd, the curve $s \mapsto s^{k-1}$ intersects the second branch of $R(s)$ at $s=-1$, but also somewhere between $s=0$ and $s=-0.508\ldots$ (which is the root obtained for $k=3$). The latter intersection point gives rise to a root of G_k smaller than -1.

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

The rest of the argument will show that all other roots of G_k lie in $[-1,-1/9]$.

Roots of (3.9) of modulus 1. We first observe that, if s has modulus 1, then the same holds for $R(s)$. More precisely, if $s=\mathrm{e}^{\mathrm{i}\theta}$, then $R(s)=\mathrm{e}^{\mathrm{i}\phi}$ with

$$\cos\phi = -\frac{56+321\cos\theta-336\cos^2\theta+128\cos^3\theta}{(5-4\cos\theta)(157+44\cos\theta-32\cos^2\theta)},$$
$$\sin\phi = -27\,\frac{(29-16\cos\theta)\sin\theta}{(5-4\cos\theta)(157+44\cos\theta-32\cos^2\theta)}.$$

Plots of $\cos\phi$ and $\sin\phi$ as a function of θ are shown in Figure 6. For $s=\mathrm{e}^{\mathrm{i}\theta}$, (3.9) is equivalent to $\cos((k-1)\theta)=\cos\phi$ and $\sin((k-1)\theta)=\sin\phi$. Given that $1/s=\mathrm{e}^{-\mathrm{i}\theta}$, we can focus on solutions such that $\theta\in[0,\pi]$. The oscillations of $\cos((k-1)\theta)$ in this interval imply that the equation $\cos((k-1)\theta)=\cos\phi$ admits at least one solution in each interval

$$\left(\frac{m-1}{k-1}\pi,\frac{m}{k-1}\pi\right],$$

for $1\leqslant m\leqslant k-1$. For each solution, $\sin((k-1)\theta)=\pm\sin\phi$, and the plot of $\sin\phi$ in Figure 6 shows that $\sin((k-1)\theta)=\sin\phi$ if and only if $\sin((k-1)\theta)\leqslant 0$, that is, if m is even. We finally note that, when k is odd, one solution is $\theta=\pi$, giving $s=-1$, which we want to exclude.

This discussion shows that (3.9) has at least $\lceil\frac{k-3}{2}\rceil$ solutions $s\neq -1$ with $\mathrm{Im}(s)>0$ on the unit circle. They give rise to as many roots of G_k in the interval $[-1,-1/9]$. With the two real roots of G_k found previously outside this interval, this gives a total of $\lceil\frac{k+1}{2}\rceil$ roots, which coincides with the degree of G_k. Hence G_k has only real roots, with one positive root ρ_k, and the others smaller than $-1/9$.

In remains to obtain an expansion of ρ_k as k grows. We first work out an expansion of the solution of (3.9) found around 1/2 (by bootstrapping in (3.9)):

$$s=\frac{1}{2}-\frac{3}{2^{k+2}}-\frac{3}{4^{k+2}}+\frac{36k+27}{4.8^{k+1}}-\frac{27\cdot 32k^2-9\cdot 16k-717}{16^{k+2}}+O\left(\frac{k^3}{32^k}\right).$$

This translates into the expansion of ρ_k using (3.10). The singular behaviour of M_k is then derived from (3.7). □

Step 4: conclusion. By Lemma 3.3 and Cauchy's formula,

$$[x^\ell]\left(M_k(x)-\frac{\alpha_k}{1-x/\rho_k}\right)=\frac{1}{2\mathrm{i}\pi}\int_{\mathcal{C}}\left(M_k(x)-\frac{\alpha_k}{1-x/\rho_k}\right)\frac{\mathrm{d}x}{x^{\ell+1}}, \tag{3.11}$$

where $\mathcal{C}$ is the circle of radius 1/9 centred at the origin. We will prove that there exists a constant C such that, for all k and $x\in\mathcal{C}$,

$$\left|M_k(x)-\frac{\alpha_k}{1-x/\rho_k}\right|\leqslant Ck,$$

so that (3.11) implies

$$[x^\ell]M_k(x)=\mathbb{E}(X_{k,\ell}^2)=\alpha_k\rho_k^{-\ell}+O(9^\ell k),$$

as stated in Proposition 3.1.

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

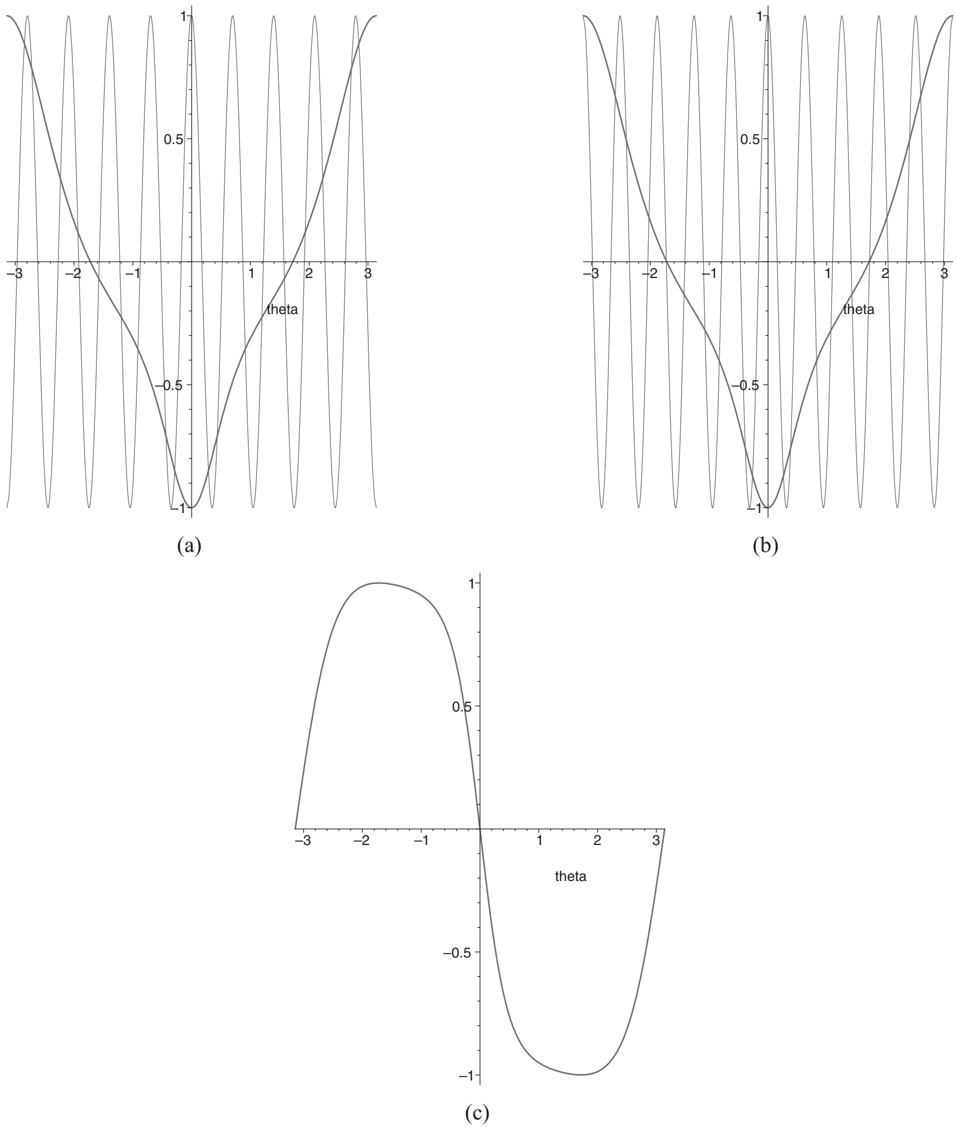


Figure 6. Plots of $\cos\phi$ (thick curve) and $\cos((k-1)\theta)$ against θ, for $\theta \in [-\pi, \pi]$, when (a) $k = 10$ and (b) $k = 11$. (c) Plot of $\sin\phi$.

It follows from (3.1) and (3.3) that $\frac{\alpha_k}{1-x/\rho_k}$ is bounded uniformly in k and $x \in \mathcal{C}$, so we only need to prove that $M_k(x) = O(k)$, uniformly in $x \in \mathcal{C}$. By Lemma 3.2, for any $x \in \mathcal{C} \setminus \{-1/9\}$, $|S(x)| < 1$. Moreover, $S(x) \to 1$ as $x \to -1/9$. Recall that $M_k(x) = 2xN_k/G_k$. By (3.4), $N_k = 2^{k/2}O(k)$, so it suffices to prove that $G_k/2^{k/2}$ is bounded away from 0, uniformly in k and $x \in \mathcal{C}$. Since $1+S$ and S^k are uniformly bounded, and $P(1/S)$ is bounded away from 0, this is equivalent to

$$\inf_{k,x\in\mathcal{C}} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0$$

(of course, S stands for $S(x)$). By the proof of Lemma 3.3,

$$S^k + \frac{P(S)}{P(1/S)}$$

does not vanish on $\mathcal{C}$. Hence it suffices to prove that

$$\liminf_k \inf_{x\in\mathcal{C}} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0.$$

Let us write $x = -e^{\pm i\theta}/9$, with $\theta \in [0, \pi]$. Then, as $\theta \to 0$,

$$S(x) = 1 - \frac{1}{2}(1 \mp i)\sqrt{\theta} + O(\theta), \tag{3.12}$$

$$|S(x)| = 1 - \frac{1}{2}\sqrt{\theta} + O(\theta), \tag{3.13}$$

$$\frac{P(S)}{P(1/S)} = 1 - \frac{20}{13}(1 \mp i)\sqrt{\theta} + O(\theta). \tag{3.14}$$

We then split interval $[0, \pi]$, to which θ belongs, into three parts.

(1) When $\sqrt{\theta} \leqslant \pi/(2k)$, there holds, uniformly in θ,

$$S(x)^k = \exp\bigl(-k(1 \mp i)\sqrt{\theta}/2\bigr) + O(1/k).$$

In particular,

$$\begin{aligned}\Re(S^k) &= \exp(-k\sqrt{\theta}/2)\cos(k\sqrt{\theta}/2) + O(1/k)\\ &\geqslant \exp(-\pi/4)/\sqrt{2} + O(1/k).\end{aligned}$$

Moreover,

$$\Re\left(\frac{P(S)}{P(1/S)}\right) = 1 + O(\sqrt{\theta}) = 1 + O(1/k),$$

uniformly in θ. Hence

$$\Re\left(S^k + \frac{P(S)}{P(1/S)}\right) = 1 + \exp(-\pi/4)/\sqrt{2} + O(1/k),$$

and

$$\liminf_k \inf_{\sqrt{\theta}\leqslant\pi/(2k)} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0.$$

(2) Let $\varepsilon > 0$ be such that, for $\sqrt{\theta} < \varepsilon$,

$$|S(x)| \leqslant 1 - \sqrt{\theta}/4 \quad \text{and} \quad \left| \frac{P(S)}{P(1/S)} \right| > 0.9.$$

Such an ε exists in view of (3.13) and (3.14). For $\pi/(2k) \leqslant \sqrt{\theta} \leqslant \varepsilon$,

$$|S|^k \leqslant (1 - \sqrt{\theta}/4)^k \leqslant (1 - \pi/(8k))^k = \exp(-\pi/8) + O(1/k),$$

so that

$$\left| S^k + \frac{P(S)}{P(1/S)} \right| \geqslant 0.9 - \exp(-\pi/8) + O(1/k) \geqslant 0.2 + O(1/k).$$

Hence

$$\liminf_k \inf_{\pi/(2k)\leqslant\sqrt{\theta}\leqslant\varepsilon} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0.$$

(3) Finally, when $\sqrt{\theta} \geqslant \varepsilon$, then $|S| < 1$ is bounded away from 1, uniformly in θ. Thus, if

$$\liminf_k \inf_{\sqrt{\theta}\geqslant\varepsilon} \left| S^k + \frac{P(S)}{P(1/S)} \right| = 0,$$

there would exist an $x \in \mathcal{C}$ such that $P(S(x)) = 0$. But this only happens when $x = 0$ or $x = (-1 \pm \sqrt{17})/4$, and none of these values lies on the circle $\mathcal{C}$.

This concludes the proof of Proposition 3.1.

4. Back to Knuth's algorithm

Let us go back to Knuth's original algorithm, described at the beginning of the paper. Recall that $\mathbb{E}(X_k)$ is the number of SAWs crossing a square of side k, and that $\mathbb{E}(X_k^2)$ is the sum of the reciprocals of the probabilities of these walks.

Proposition 4.1. *There exist two positive constants λ and β such that*

$$\mathbb{E}(X_k)^{1/k^2} \to \lambda \quad \textit{and} \quad \mathbb{E}(X_k^2)^{1/k^2} \to \beta.$$

Of course, $\beta \geqslant \lambda^2$. Moreover, writing $c(k) = \mathbb{E}(X_k)$ and $d(k) = \mathbb{E}(X_k^2)$, we have

$$\lambda = \sup_k \ c(k)^{1/(k+1)^2} \tag{4.1}$$

and

$$\beta = \sup_k \ (\sqrt{2}\, d(k))^{1/(k+1)^2}. \tag{4.2}$$

As discussed at the end of the Introduction, there is a hope of combining (4.2) and known upper bounds on λ to prove that $\beta > \lambda^2$, in which case one could conclude that the relative variance of X_k grows as κ^{k^2}, with $\kappa = \beta/\lambda^2 > 1$.

Proof. As can be expected, these results follow from a super-multiplicativity argument. The existence of λ was established for the first time in [1], and (4.1) (which allows us to produce lower bounds on λ) appears in [7]. We repeat the argument, because it applies almost verbatim to the numbers $d(k)$.

Define $\lambda := \limsup_k c(k)^{1/k^2}$. Then λ is finite, because there are only a quadratic number of edges in the $k \times k$ square, and a walk is determined by the set of its edges. Let $\varepsilon > 0$. We will prove that

$$\liminf c(K)^{1/K^2} \geqslant \lambda - \varepsilon, \tag{4.3}$$

which implies that λ is actually the limit of $c(k)^{1/k^2}$.

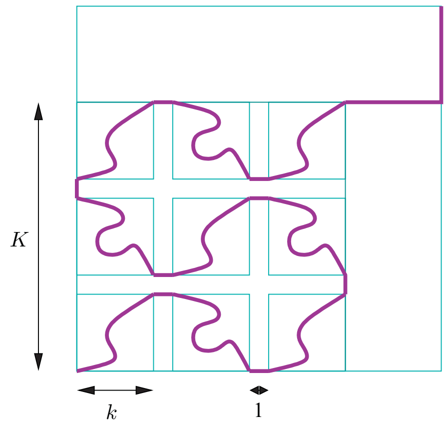


Figure 7. (Colour online) Super-multiplicativity for SAWs crossing a square.

Let $k > 0$ be such that $c(k)^{1/(k+1)^2} > \lambda - \varepsilon$. Let $K \geqslant k$, and let n be maximal so that

$$(k+1)(2n+1) - 1 \leqslant K.$$

This implies in particular that $K < (k+1)(2n+3)$. In the $K \times K$ square, insert $(2n+1)^2$ smaller squares of side k, as shown in Figure 7. In each smaller square, choose a SAW that crosses it, and build from this collection of short walks a long walk crossing the larger square, as shown in the figure. This construction implies

$$c(K) \geqslant c(k)^{(2n+1)^2}.$$

Thus

$$c(K)^{1/K^2} \geqslant c(K)^{1/((k+1)^2(2n+3)^2)} \geqslant \left(c(k)^{1/(k+1)^2}\right)^{(2n+1)^2/(2n+3)^2} \geqslant (\lambda - \varepsilon)^{(2n+1)^2/(2n+3)^2}.$$

Taking the lim inf on K boils down to taking the lim inf on n, and gives (4.3). The bound (4.1) also follows from the above inequalities.

Let us now consider the numbers $d(k)$. Again, $\beta := \limsup d(k)^{1/k^2}$ is finite, because

$$d(k) = \sum_{w \in \mathcal{W}_k} \frac{1}{p(w)} \leqslant \sum_{w \in \mathcal{W}_k} 3^{|w|} \leqslant 3^{O(k^2)} c(k),$$

where $|w|$ denotes the length of w. Now return to Figure 7. Let $w_1, w_2, \ldots$ denote the short walks, and by w the long one. It is clear for the sampling algorithm that

$$\frac{1}{p(w)} \geqslant \prod_{i=1}^{(2n+1)^2} \frac{1}{p(w_i)}.$$

It follows that

$$d(K) \geqslant d(k)^{(2n+1)^2},$$

from which one can prove, as above, that $\beta = \lim d(k)^{1/k^2}$. The above bound on $d(K)$ can actually be improved: in every row of $(2n+1)$ small squares, except maybe the top one,

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

Table 1.

		North and East	North, East and South	All four steps
Confined to $k \times k$	number	$4^k/\sqrt{k}$	$(k+1)^k$	λ^{k^2}
	relative variance	$\sqrt{k}$ [5]	$2^{k(k+1)}/(k+1)^k$ (Prop. 3.1)	κ^{k^2} ? (Prop. 4.1)
	average length	k	k^2	k^2 [16]
Unconfined, n steps	number	2^n	$(1+\sqrt{2})^n$	$(2.64\ldots)^n$ [17]
	relative variance	0	$(6/(1+\sqrt{2}))^n$	α^n (predicted by [6])

Summary of the variance predictions and results

n of the horizontal steps added between the small squares have probability 1/2 or 1/3. Hence

$$d(K) \geqslant 2^{2n^2} d(k)^{(2n+1)^2},$$

and the lower bound (4.2) now follows. □

5. Final comments

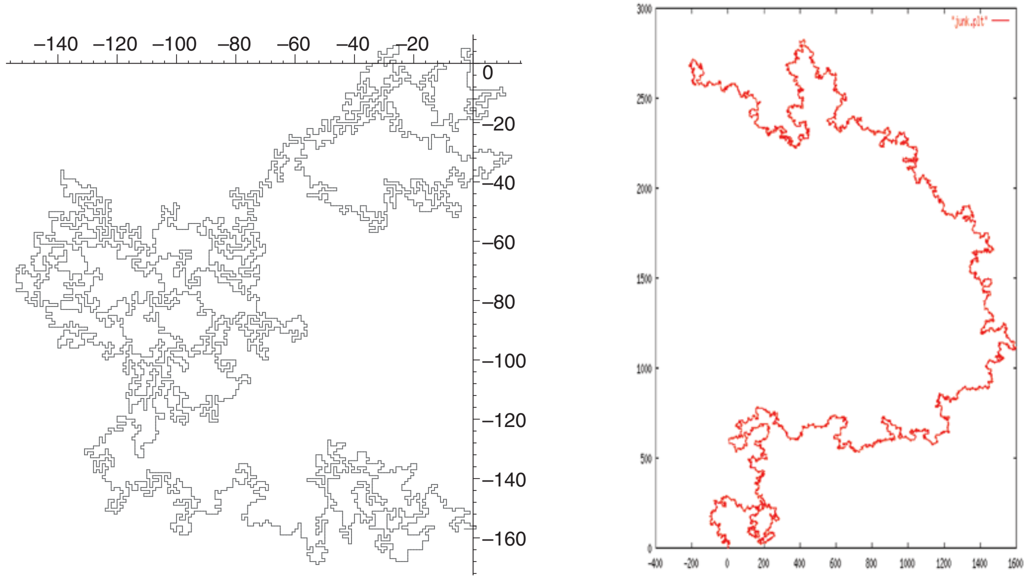
5.1. Unconfined walks

Knuth designed his algorithm to sample SAWs crossing a square of side k, but other authors have used similar ideas to sample unconfined SAWs of fixed length n. For example, the classical Rosenbluth algorithm [21] generates general SAWs step by step, by taking at each time, uniformly at random, one of the steps that preserve self-avoidance. If at some point no such step is available, and the walk has not reached length n, the algorithm restarts from scratch. This rejection step is avoided if one only samples *untrapped* self-avoiding walks, that is, walks that can be extended into a SAW of infinite length.[4] (We describe in Section 5.3 a simple procedure that detects whether a new step traps the walk, which is also useful for implementing Knuth's algorithm.) A recent numerical study, using a refinement of the above algorithm, suggests that the asymptotic properties of untrapped walks are similar to those of general SAWs, in terms of number and end-to-end distance [9].

For these algorithms, the quality of the cardinality estimator is still related to the variance of the random variable X_n equal to the reciprocal of the probability of the generated walk. As already mentioned, the variance of the Rosenbluth estimator is predicted to be exponential in n [6]. We do not know of any similar study for untrapped walks. It is easy to determine the variance of X_n for the Rosenbluth algorithm restricted to directed or partially directed walks. Our results are summarized in Table 1. In particular, for partially directed walks the relative variance is found to be exponential in n.

[4] This notion of untrapped walks differs from that in [19], where a walk is said to be untrapped as soon as it can be extended by one step.

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press



(a) (b)

Figure 8. (Colour online) (a) A random untrapped SAW of length 5000 obtained via importance sampling, and (b) a quasi-uniform SAW of length 20 000.

5.2. Kinetic distributions

It is also interesting to study the asymptotic properties of SAWs chosen according to the non-uniform (but very natural) 'kinetic' distribution that results from importance sampling. These properties may be different from those observed in the uniform case. For instance, one can expect the average end-to-end distance of kinetic unconfined SAWs to be smaller than $n^{3/4}$, because 'compact' walks in which few steps are eligible at each time have a higher probability than more spread-out walks. In fact, the kinetic end-to-end distance is conjectured [18] to grow like $n^{2/3}$. (For unconfined partially directed walks, however, the end-to-end distance is easily shown to be linear, both for the uniform and the kinetic model.) Figure 8 shows a random (untrapped) SAW that we generated by importance sampling and a (quasi-)uniform SAW generated using a pivot algorithm [13].

5.3. When does a walk get trapped?

One important feature of Knuth's algorithm, and of its adaptation to untrapped SAWs discussed in Section 5.1, is that one never appends a step that would trap the walk. Since Knuth does not explain in his paper how he detects trapping, let us describe the method we used. One obvious case of trapping in Knuth's algorithm is when the walk reaches the boundary of the square, and moves towards the origin. In all other trapping situations, the walk would have been trapped as well in the unconfined setting, so we focus on the trapping of unconfined walks.

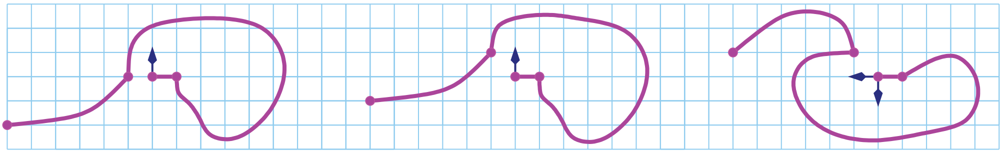


Figure 9. (Colour online) How a walk gets trapped.

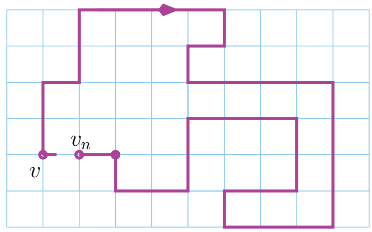


Figure 10. (Colour online) The winding number between v and v_n is -2π.

Let w be an untrapped SAW of length n, ending at vertex $v_n = (i, j)$, and, say, with a W step. There are, up to obvious symmetries, exactly three situations when adding a new step to w creates a trapped walk:

- the vertex $v = (i-1, j)$ belongs to w, one appends an N step to w and the portion of w going from v to v_n has *winding number* -2π,
- the vertex $v = (i-1, j+1)$ belongs to w, one appends an N step to w and the portion of w going from v to v_n has winding number -2π,
- the vertex $v = (i-1, j+1)$ belongs to w, one appends a W or S step to w and the portion of w going from v to v_n has winding number 2π.

These three cases are depicted in Figure 9. When computing the winding number, we add a half-edge pointing from the East to v (Figure 10). The winding number is then the difference between the number of left turns and the number of right turns, multiplied by $\pi/2$.

Acknowledgements

I am very grateful to Persi Diaconis, who first told me about this question, and then sent helpful suggestions on this manuscript and pointed to interesting related papers. I also thank the referee of the first version of this paper for his very thorough report and helpful references.

References

[1] Abbott, H. L. and Hanson, D. (1978) A lattice path problem. *Ars Combin.* **6** 163–178.
[2] Bacher, A. Generalized Dyck paths of bounded height. arXiv:1303.2724.
[3] Bacher, A. and Bousquet-Mélou, M. (2011) Weakly directed self-avoiding walks. *J. Combin. Theory Ser. A* **118** 2365–2391.
[4] Banderier, C., Bousquet-Mélou, M., Denise, A., Flajolet, P., Gardy, D. and Gouyou-Beauchamps, D. (2002) Generating functions for generating trees. *Discrete Math.* **246** 29–55.

https://doi.org/10.1017/S0963548313000643 Published online by Cambridge University Press

[5] Bassetti, F. and Diaconis, P. (2006) Examples comparing importance sampling and the Metropolis algorithm. *Illinois J. Math.* **50** 67–91.
[6] Batoulis, J. and Kremer, K. (1988) Statistical properties of biased sampling methods for long polymer chains. *J. Phys. A* **21** 127–146.
[7] Bousquet-Mélou, M., Guttmann, A. J. and Jensen, I. (2005) Self-avoiding walks crossing a square. *J. Phys. A* **38** 9159–9181.
[8] Bousquet-Mélou, M. and Petkovšek, M. (2000) Linear recurrences with constant coefficients: the multivariate case. *Discrete Math.* **225** 51–75.
[9] Chan, Y.-B. and Rechnitzer, A. (2012) A Monte Carlo study of non-trapped self-avoiding walks. *J. Phys. A* **45** 405004.
[10] Chen, Y., Diaconis, P., Holmes, S. P. and Liu, J. S. (2005) Sequential Monte Carlo methods for statistical analysis of tables. *J. Amer. Statist. Assoc.* **100** 109–120.
[11] Blitzstein, J. and Diaconis, P. (2010) A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** (4) 489–522.
[12] Flajolet, P. and Sedgewick, R. (2009) *Analytic Combinatorics*, Cambridge University Press.
[13] Kennedy, T. (2002) A faster implementation of the pivot algorithm for self-avoiding walks. *J. Statist. Phys.* **106** 407–429.
[14] Knuth, D. E. (1976) Mathematics and computer science: coping with finiteness. *Science* **194** 1235–1242.
[15] Knuth, D. E. (1996) *Selected papers on computer science*, Vol. 59 of *CSLI Lecture Notes*, CSLI Publications.
[16] Madras, N. (1995) Critical behaviour of self-avoiding walks that cross a square. *J. Phys. A* **28** 1535–1547.
[17] Madras, N. and Slade, G. (1993) *The Self-Avoiding Walk*, Probability and its Applications, Birkhäuser.
[18] Majid, I., Jan, N., Coniglio, A. and Stanley, H. E. (1984) Kinetic growth walk: A new model for linear polymers. *Phys. Rev. Lett.* **52** 1257–1260.
[19] Owczarek, A. L. and Prellberg, T. (2008) Scaling of the atmosphere of self-avoiding walks. *J. Phys. A* **41** 375004.
[20] Prodinger, H. (2003/04) The kernel method: A collection of examples. *Sém. Lothar. Combin.* **50** B50f.
[21] Rosenbluth, M. N. and Rosenbluth, A. W. (1955) Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.* **23** 356–359.
[22] Stanley, R. P. (1997) *Enumerative Combinatorics I*, Vol. 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press.