Random sampling of lattice paths with constraints, via transportation

Lucas Gerin[1,2]

[1]UNIVERSITÉ PARIS-OUEST,MODAL'X-*bât. G, 200 avenue de la République,F-92000 Nanterre*
[2]ANR GAMMA, `http://www-apr.lip6.fr/anrGamma`

We build and analyze in this paper Markov chains for the random sampling of some one-dimensional lattice paths with constraints, for various constraints. These chains are easy to implement, and sample an "almost" uniform path of length n in $n^{3+\varepsilon}$ steps. This bound makes use of a certain *contraction property* of the Markov chain, and is proved with an approach inspired by optimal transport.

Keywords: lattice paths, random sampling, MCMC, discrete Ricci curvature

1 Lattice Paths with Constraints

Lattice paths arise in several areas in probability and combinatorics, either in their own interest (as realizations of random walks, or because of their interesting combinatorial properties: see [1] for the latter) or because of fruitful bijections with various families of trees, tilings, words. The problem we discuss here is to efficiently sample uniform (or *almost* uniform) paths in a family of paths with constraints.

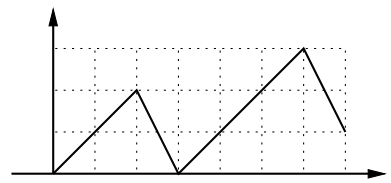
There are several reasons for which one may want to generate uniform samples of lattice paths: to make and try conjectures on the behaviour of a large "typical" path, test algorithms running on paths (or words, trees,...). In view of random sampling, it is often very efficient to make use of the combinatorial structure of the family of paths under study. In some cases, this yields linear-time (in the length of the path) *ad-hoc* algorithms [2, 6]. However, the nature of the constraints makes sometimes impossible such an approach, and there is a need for robust algorithms that work in lack of combinatorial knowledge.

Luby,Randall and Sinclair [11] design a Markov chain that generate sets of non-intersecting lattice paths. This was motivated by a classical (and simple, see illustrations in [4, 14]) correspondence between dimer configurations on an hexagon, rhombae tilings of this hexagon and families of non-intersecting lattice paths. As the first step for the analysis of this chain, Wilson [14] introduces a peak/valley Markov chain (see details below) over some simple lattice paths and obtain sharp bounds for its mixing time. We present in this paper a variant of this Markov chain, which is valid for various constraints and whose analysis is simple. It generates an "almost" uniform path of length n in $n^{3+\varepsilon}$ steps, this bound makes use of a certain *contraction property* of the chain.

Appart from the algorithmic aspect, the peak/valley process seems to have a physical relevancy as a simplified model for the evolution of *quasicrystals* (see a discussion on a related process in the introduction of [4]). In particular, the mixing time of this Markov seems to have some importance.

1365–8050 © 2010 Discrete Mathematics and Theoretical Computer Science (DMTCS), Nancy, France

Notations



We fix three integers $n, a, b > 0$, and consider the paths of length n, with steps $+a/-b$, that is, the words of n letters taken in the alphabet $\{a, -b\}$. Such a word $s = (s_1, s_2, \ldots, s_n)$ is identified to the path $S = (S_1, \ldots, S_n) := (s_1, s_1 + s_2, \ldots, s_1 + s_2 + \cdots + s_n)$. On the right, one sees the lattice path $S = (1, 2, 0, 1, 2, 3, 1)$ associated with the word $s = (1, 1, -2, 1, 1, 1, -2)$.

To illustrate the methods and the results, we focus on some particular sub-families $\mathcal{A}_n \subset \{a, -b\}^n$:

1. Discrete *meanders*, denoted by $\mathcal{M}_n$, which are simply the non-negative paths: $S \in \mathcal{M}_n$ if for any $i \leq n$ we have $S_i \geq 0$. This example is mainly illustrative because the combinatorial properties of meanders make it possible to perform exact sampling very efficiently (an algorithm running in $\mathcal{O}(n^{1+\varepsilon})$ steps is given in [2], an order that we cannot get in the present paper).

2. Paths with *walls*. A path with a wall of height h between r and s is a path such that $S_i \geq h$ for any $r \leq i \leq s$ (see Fig. 1 for an example). These are denoted by $\mathcal{W}_n = \mathcal{W}_n(h, r, s)$, they appear in statistical mechanics as toy models for the analysis of random interfaces and polymers (see examples in [7]).

3. *Excursions*, denoted by $\mathcal{E}_n$, which are non-negative paths such that $S_n = 0$. In the case $a = b = 1$, these correspond to well-parenthesed words and are usually called Dyck words. In the general case, Duchon [6] proposes a rejection algorithm which generates excursions in linear time.

4. *Culminating paths* of size n, denoted further by $\mathcal{C}_n$, which are non-negative paths whose maximum is attained at the last step: for any i we have $0 \leq S_i \leq S_n$. They have been introduced in [2], motivated in particular by the analysis of some algorithms in bioinformatics.

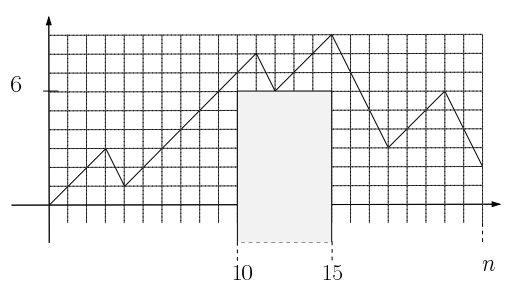


Fig. 1: A path of steps $+1/-2$, with a wall of height $h = 6$ between $i = 10$ and $j = 15$.

2 Sampling with Markov chains

We will consider Markov chains in a family $\mathcal{A}_n$, where all the probability transitions are symmetric. For a modern introduction to Markov chains, we refer to [8]. Hence we are given a transition matrix $(p_{i,j})$ of

size $|\mathcal{A}_n| \times |\mathcal{A}_n|$ with

$$p_{i,j} = p_{j,i} \text{ whenever } i \neq j,$$
$$p_{i,i} = 1 - \sum_{j \neq i} p_{i,j}.$$

Lemma 1 *If such a Markov chain is irreducible, then it admits as unique stationary distribution the uniform distribution $\pi = \pi(\mathcal{A}_n)$ on $\mathcal{A}_n$.*

Proof: The equality $\pi(i)p_{i,j} = \pi(j)p_{j,i}$ holds for any two vertices i, j. This shows that the probability distribution π is reversible for $(p_{i,j})$, and hence stationary. It is unique if the chain is irreducible. □

This lemma already provides us with a scheme for sampling an almost uniform path in $\mathcal{A}_n$, without knowing much about $\mathcal{A}_n$. To do so, we define a "flip" operator on paths, this is an operator

$$\begin{array}{rccc} \phi: & \mathcal{A}_n \times \{1, \ldots, n\} \times \{\downarrow, \uparrow\} \times \{+, -\} & \to & \mathcal{A}_n \\ & (\mathbf{S}, i, \varepsilon, \delta) & \mapsto & \phi(\mathbf{S}, i, \varepsilon, \delta). \end{array}$$

When $i \in \{1, 2, \ldots, n-1\}$ the path $\phi(\mathbf{S}, i, \uparrow, \delta)$ is defined as follows : if $(s_i, s_{i+1}) = (-b, a) =$ then these two steps are changed into $(a, -b) =$. The $n-2$ other steps remain unchanged. If $(s_i, s_{i+1}) \neq (-b, a)$ then $\phi(\mathbf{S}, i, \uparrow, \delta) = \mathbf{S}$. Note that in the case $i \in \{1, 2, \ldots, n-1\}$ the value of ϕ does not depend on δ.

For the case $i = n$, if $\delta = +$, we define $\phi(\mathbf{S}, n, \varepsilon, \delta)$ as before as if there would be a $+a$ as the end if the path. For instance, in the case where $S_n = -b$, the path $\phi(\mathbf{S}, n, \uparrow, +)$, the n-th step is turned into a.

The path $\phi(\mathbf{S}, i, \downarrow, \delta)$ is defined equally: if $i < n$ and $(s_i, s_{i+1}) =$, it turns into . When $\delta = -$, one flips as if there would be a $-b$ at the end of the path.

For culminating paths, we have to take another definition of $\phi(\mathbf{S}, n, \uparrow, \delta), \phi(\mathbf{S}, n, \downarrow, \delta)$, see Section 2.1.

We are also given a probability distribution $\mathbf{p} = (p_i)_{1 \leq i \leq n}$, and we assume that $p_i > 0$ for each i. We will consider a particular sequence $\mathbf{p}$ later on, but at this point we can take the uniform distribution in $\{1, \ldots, n\}$. We describe the algorithm below in Algorithm 1.

In words, this algorithm performs the Markov chain in $\mathcal{A}_n$ with transition matrix $P = (P_{\mathbf{R},\mathbf{S}})_{\mathbf{R},\mathbf{S} \in \mathcal{A}_n}$ defined as follows:

$$\begin{cases} P_{\mathbf{R},\mathbf{S}} & = p_i/2, \text{ if } \mathbf{S} \neq \mathbf{R} \text{ and } \mathbf{S} = \phi(\mathbf{R}, i, \varepsilon, \delta) \text{ for some } i, \varepsilon, \delta \\ P_{\mathbf{R},\mathbf{S}} & = 0 \text{ otherwise,} \\ P_{\mathbf{R},\mathbf{R}} & = 1 - \sum_{\mathbf{S} \neq \mathbf{R}} P_{\mathbf{R},\mathbf{S}}. \end{cases}$$

Proposition 2 *Denote by $S(t)$ the random path obtained after the t-th run of the loop in Algorithm 1. When $t \to \infty$, the sequence $S(t)$ converges in law to the uniform distribution in $\mathcal{A}_n$. Moreover, the execution of Algorithm 1 until time T is linear in T.*

Algorithm 1 Approximate sampling of a path in $\mathcal{A}_n$

initialize $\mathbf{S} = (+a, +a, +a, \dots, +a)$
$I_1, I_2, \cdots \leftarrow$ i.i.d. r.v. with law $\mathbf{p}$
$\varepsilon_1, \varepsilon_2, \cdots \leftarrow$ i.i.d. uniform r.v. in $\{\uparrow, \downarrow\}$
$\delta_1, \delta_2, \cdots \leftarrow$ i.i.d. uniform r.v. in $\{+, -\}$
for $t = 1$ to T **do**
  if $\phi(\mathbf{S}, I_t, \varepsilon_t, \delta_t)$ is in $\mathcal{A}_n$ **then**
    $\mathbf{S} \leftarrow \phi(\mathbf{S}, I_t, \varepsilon_t, \delta_t)$
  end if
end for

Proof: For the first claim, we have to check that the chain is aperiodic and irreducible. Aperiodicity comes from the (many) loops. Irreducibility will follow from Lemma 4. For the second claim, notice that the time needed for the test "$\phi(\mathbf{S}, I_t, \varepsilon_t,)$ is in $\mathcal{A}_n$" can be considered as constant, since for the families $\mathcal{M}_n$ and $\mathcal{E}_n$ we only have to compare $0, S_i$ while for the family $\mathcal{W}_n$ we only have to compare S_i with the height of the wall at i. For the case of the culminating paths, see below in Section 2.1. □

We now choose the distribution (p_i). Instead of $p_i = 1/n$, we will use the probability distribution defined by

$$p_i := i(2n - i)\kappa_0 + a \quad (\text{ for } i = 1, \dots, n), \tag{1}$$

where

$$\kappa_0 = \frac{3}{2n^2(n+1)}$$
$$a = 1/4n^3.$$

We let the reader check that $(p_i)_{i \le n}$ is indeed a probability distribution. The reason for which we use this particular distribution will appear in the proof of Proposition 5. We will then need the following relation: for each $1 \le i \le n - 1$,

$$p_i - p_{i-1}/2 - p_{i+1}/2 = \kappa_0. \tag{2}$$

For Algorithm 1 to be efficient, we need to know how $S(T)$ is close in law to π. This question is related to the spectral properties of the matrix P. In particular, the speed of convergence is governed by the spectral gap (*i.e.* $1 - \lambda$, where λ is the largest of the modulus of the eigenvalues different from one, see [10] for example), but this quantity is not known in general. Some geometrical methods [5] allow to bound from below $1 - \lambda$, but they assume a precise knowledge of the structure of the graph defined by the chain P. It seems that such results do not apply here.

Instead, we will study the metric properties of the chain P with respect to a natural distance on $\mathcal{A}_n$, and show that it satisfies a certain *contraction property*.

2.1 The variant of Algorithm 1 for culminating paths

Unchanged, our Markov chain P cannot generate culminating paths since the path $\mathbf{S} = (a, a, \dots, a)$ would then be isolated: it has no peak/valley and $\phi(\mathbf{S}, n, \downarrow, -) = (a, a, \dots, -b)$ which is not culminating.

Thus we propose a slight modification for the family $\mathcal{C}_n$. We only change the definition of $\phi(\mathbf{S}, i, \varepsilon, \delta)$ when $i = n$ (it won't depend on δ). Since the maximum is reached at n, the $\lceil b/a \rceil + 1$ last steps are necessarily

$$(a, a, \ldots, a) \text{ or } (-b, a, \ldots, a).$$

We thus define $\phi(\mathbf{S}, n, \uparrow, \delta)$ as the path obtained by changing the $\lceil b/a \rceil + 1$ last steps into $(a, a, \ldots, a)$ (regardless of their initial values in $\mathbf{S}$) and $\phi(\mathbf{S}, n, \downarrow, \delta)$ as the path obtained by changing the $\lceil b/a \rceil + 1$ last steps into $(-b, a, \ldots, a)$.

Notice that despite this change the execution time of each loop of Algorithm 1 is still a $\mathcal{O}(1)$:

- If $I_t < n$, the time needed for the test "$\phi(\mathbf{S}, I_t, \varepsilon_t, \delta_t)$ is in $\mathcal{A}_n$" can be considered as constant, since we only have to compare $0, S_i, S_n$.

- If $I_t = n$, the new value S_n is compared with the maximum of S, which can be done in $\mathcal{O}(n)$. Fortunately, this occurs with probability $p_n = \mathcal{O}(1/n)$, so that the time-complexity of each loop is, on average, a $\mathcal{O}(1)$.

3 Error estimates with contraction

Going back to a more general setting, we consider a Markov chain in a finite set V, endowed with a metric d. For a vertice $x \in V$ and a transition matrix P, we denote by $P\delta_x$ (resp. $P^t\delta_x$) the law of the Markov chain associated with P at time 1 (resp. t), when starting from x. For $x, y \in V$, the main assumption made on P is that there is a coupling between $P\delta_x, P\delta_y$ (that is, a random variable (X_1, Y_1) with $X_1 \overset{\text{law}}{=} P\delta_x, Y_1 \overset{\text{law}}{=} P\delta_y$) such that

$$\mathbb{E}\left[d(X_1, Y_1)\right] \leq (1 - \kappa) d(x, y), \tag{3}$$

for some $\kappa > 0$, which is called the *Ricci curvature* of the chain, by analogy with the Ricci curvature in differential geometry[(i)]. If the inequality holds, then it implies ([10],p.189) that P admits a unique stationary measure π and that, for any x,

$$\| P^t\delta_x - \pi \|_{\mathrm{TV}} \leq (1 - \kappa)^t \mathrm{diam}(V), \tag{4}$$

where $\mathrm{diam}(V)$ is the diameter of the graph with vertices V induced by the Markov chain. The notation $\| \cdot \|_{\mathrm{TV}}$ stands, as usual, for the *Total Variation* distance over the probability distributions on V defined by

$$\| \mu_1 - \mu_2 \|_{\mathrm{TV}} := \sup_{A \subset V} |\mu_1(A) - \mu_2(A)| .$$

Hence, a positive Ricci curvature gives the exponential convergence to the stationary measure, with an exact (*i.e.* (4) is non-asymptotic in t) bound. In many situations, a smart choice for the coupling between X_1, X_2 gives a sharp rate of convergence in eq. (4) (see some striking examples in [12]).

[(i)] The Ricci curvature is actually the largest positive number such that (3) holds, for all the couplings of $P\delta_x, P\delta_y$; here we should rather say that Ricci curvature is larger than κ.

3.1 Metric properties of P

To apply the Ricci curvature machinery, we endow each $\mathcal{A}_n$ with the L^1-distance

$$d_1(S, S') = \frac{1}{a+b} \sum_{i=0}^{n} |S_i - S'_i|.$$

(Notice that $|S_i - S'_i|$ is always a multiple of $a + b$.) For our purpose, it is fundamental that this metric space is *geodesic*.

Definition 3 *A Markov chain P in a finite set V is said to be* geodesic *with respect to the distance d on V if for any two $x, y \in V$ with $d(x, y) = k$, there exist $k + 1$ vertices $x_0 = x, x_1, \ldots, x_k = y$ in V such that for each i*

- $d(x_i, x_{i+1}) = 1$ *;*
- *x_i and x_{i+1} are neighbours in the Markov chain P (*i.e. $P(x_i, x_{i+1}) > 0$*).*

This implies in particular that P is irreducible and that the diameter of P is smaller than $\max_{x,y} d(x, y)$.

Lemma 4 *For each family $\mathcal{C}_n$,$\mathcal{W}_n$,$\mathcal{E}_n$,$\mathcal{M}_n$ the Markov chain of Algorithm 1 is geodesic with respect to d_1.*

Proof of Lemma 4: The proof goes by induction on k. We fix $S \neq T$ (and denote by s, t the corresponding words) ; we want to decrease $d_1(S, T)$ by one, by applying the operator ϕ with proper i, ε. We denote by $i_0 \in \{1, \ldots, n\}$ the first index for which $S \neq T$. For instance we have $T_{i_0} = S_{i_0} + a + b$. Let j be the position of the left-most peak in T in $\{i_0 + 1, i_0 + 2, \ldots, n\}$, if such a peak exists. Then $S' := \phi(\mathbf{T}, j, \downarrow, \delta)$ is also in $\mathcal{A}_n$: it is immediate for the families $\mathcal{M}_n, \mathcal{W}_n, \mathcal{C}_n, \mathcal{E}_n$. We have $d_1(S, S') = k - 1$.

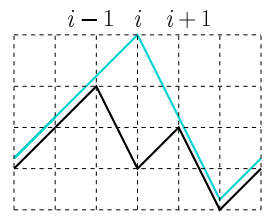
If there is no peak in T after i_0, then $(t_{i_0+1}, t_{i_0+2}, \ldots, t_n) = (a, a, \ldots, a)$. Hence we try to increase the final steps of S by one. To do so, we choose $S' := \phi(\mathbf{S}, n, \uparrow, \delta)$ if $S \neq \phi(\mathbf{S}, n, \uparrow, \delta)$, or $S' = \phi(\mathbf{S}, j, \uparrow, \delta)$ where j is the position of the right-most $-b$ otherwise (we choose the right-most one to ensure that $\phi(\mathbf{S}, j, \uparrow, \delta)$ remains culminating in the case where $\mathcal{A}_n = \mathcal{C}_n$.).

□

For meanders, excursions and walls, we will show that the Ricci curvature of P with respect to the distance d_1 is (at least) of order $1/n^3$.

Proposition 5 *For the three families $\mathcal{M}_n, \mathcal{E}_n, \mathcal{W}_n$, the Ricci curvature of the associated Markov chain, with weights (p_i) defined as in* (1)*, is larger than κ_0.*

Proof of Proposition 5:



Fix $\mathbf{S}, \mathbf{T}$ in $\mathcal{A}_n \in \{\mathcal{M}_n, \mathcal{E}_n, \mathcal{W}_n\}$, we first assume that $\mathbf{S}, \mathbf{T}$ are neighbours, for instance $\mathbf{T} = \phi(\mathbf{S}, i, \uparrow,)$ for some i. Let $(\mathbf{S}^1, \mathbf{S}^2)$ be the random variable in $\mathcal{A}_n \times \mathcal{A}_n$ whose law is defined by

$$(\mathbf{S}^1, \mathbf{S}^2) \overset{\text{(law)}}{=} (\phi(\mathbf{S}, \mathcal{I}, E), \phi(\mathbf{T}, \mathcal{I}, E)),$$

where $\mathcal{I}$ is a r.v. taking values in $\{1, \ldots, n\}$ with distribution $\mathbf{p}$ and E is uniform in $\{\uparrow, \downarrow\}$. In other words, we run one loop of Algorithm 1 simultaneously on both paths.

We want to show that $\mathbf{S}^1, \mathbf{S}^2$ are, on average, closer than $\mathbf{S}, \mathbf{T}$. Different cases may occur, depending on $\mathcal{I}$ and on the index i where $\mathbf{S}, \mathbf{T}$ differ.

Case 1. $i = 1, 2, \ldots, n-2$**.**

Case 1a. $\mathcal{I} = i$**.** This occurs with probability p_i and, no matter the value of E, we have $\mathbf{S}^1 = \mathbf{S}^2$.

Case 1b. $\mathcal{I} = i-1$ **or** $i+1$**.** We consider the case $i-1$. Since $\mathbf{S}$ and $\mathbf{T}$ coincide everywhere but in i, we necessarily have one of these two cases:

- there is a peak in $\mathbf{S}$ at $i-1$ and neither a peak nor a valley in $\mathbf{T}$ at $i-1$ (as in the figure on the right) ;
- there is a valley in $\mathbf{T}$ at $i-1$ and neither a peak nor a valley in $\mathbf{S}$ at $i-1$.

In the first case for instance, then we may have $d_1(\mathbf{S}^1, \mathbf{S}^2) = 2$ if $E = \downarrow$, while the distance remains unchanged if $E = \uparrow$. The case $\mathcal{I} = i+1$ is identical. This shows that with a probability smaller than $p_{i-1}/2 + p_{i+1}/2$ we have $d_1(\mathbf{S}^1, \mathbf{S}^2) = 2$.

Case 1c. $\mathcal{I} \neq i-1, i, i+1$ **and** $\mathcal{I} \neq n$**.** In this case, $\mathbf{S}$ and $\mathbf{T}$ are possibly modified in $\mathcal{I}$, but if there is a modification it occurs in both paths. It is immediate since for the families $\mathcal{M}_n$,$\mathcal{W}_n$ and $\mathcal{E}_n$ since the constraints are local.

Case 2. $i = n-1$**.** In this case, it is easy to check that, because of our definition of $\phi(\mathbf{S}, n, \varepsilon, \delta)$, we have

$$\mathbb{E}\left[d_1(\mathbf{S}^1, \mathbf{S}^2)\right] \leq 1 - p_{n-1} + p_{n-2}/2 + p_n/2 = 1 - \kappa_0.$$

Case 3. $i = n$**.** We have

$$\mathbb{E}\left[d_1(\mathbf{S}^1, \mathbf{S}^2)\right] \leq 1 + p_{n-1}/2 - p_n/2 = 1 - \kappa_0.$$

Thus, we have proven that when $\mathbf{S}, \mathbf{T}$ only differ at i

$$\begin{aligned}\mathbb{E}\left[d_1(\mathbf{S}^1, \mathbf{S}^2)\right] &\leq 2 \times (p_{i-1}/2 + p_{i+1}/2) + 0 \times p_i + 1 \times (1 - p_i - p_{i-1}/2 - p_{i+1}/2) \\ &\leq (1 - \kappa_0) \times 1 = (1 - \kappa_0) d_1(S, T).\end{aligned} \tag{5}$$

What makes Ricci curvature very useful is that if this inequality holds for pairs of neighbours then it holds for any pair, as noticed in [3]. Indeed, take $k+1$ paths $S_0 = S, S_1, \ldots, S_k = T$ as in Lemma 4 and apply the triangular inequality for d_1:

$$\begin{aligned}\mathbb{E}\left[d_1(\phi(S, \mathcal{I}, E), \phi(T, \mathcal{I}, E))\right] &\leq \sum_{i=0}^{k-1} \mathbb{E}\left[d_1(\phi(S_i, \mathcal{I}, E), \phi(S_{i+1}, \mathcal{I}, E))\right] \\ &\leq (1 - \kappa_0) k = (1 - \kappa_0) d_1(S, T).\end{aligned}$$

□

Remark 6 *It is easy to exhibit some S, T such that ineq.* (5) *is in fact an equality. In the case where $p_i = 1/n$, this equality reads $\mathbb{E}\left[d_1(\mathbf{S}^1, \mathbf{S}^2)\right] = d_1(S, T)$, and we cannot obtain a positive Ricci curvature (though this does not prove that there is not another coupling or another distance for which we could get a $\kappa > 0$ in the case $p_i = 1/n$.).*

We recall that for each family $\mathcal{A}_n$, $\mathrm{diam}(\mathcal{A}_n) = \max d_1(\mathbf{S}, \mathbf{T}) \leq n(n+1)/2$. Hence, combining Proposition 5with Eq. (4) gives our main result:

Theorem 7 *For meanders, excursions and path with walls, Algorithm 1 returns an almost uniform sample of π, as soon as $T \gg n^3$. Precisely, for any itinialization of Algorithm 1,*

$$\| \mathbf{S}(T) - \pi \|_{\mathrm{TV}} \leq \mathrm{diam}(\mathcal{A}_n)(1-\kappa)^T \leq \frac{n(n+1)}{2} \exp\left(-\frac{3}{2n^2(n+1)}T\right).$$

Another formulation of this result is that the mixing time of the associated Markov chain, defined as usual by

$$t_{\mathrm{mix}} := \left\{\inf\ t \geq 0\ ;\ \sup_{v \in V} \| P^t \delta_v - \pi \|_{\mathrm{TV}} \leq e^{-1}\right\} \tag{6}$$

(e^{-1} is here by convention), is smaller than $n^2(n+1)\log n$. For culminating paths, the argument of Case 1c fails and (5) does not hold, we are not able to prove such a result as Theorem 7. However, it seems empirically that the mixing time is also of order $n^3 \log n$ (with a constant strongly dependent on a, b). A way to prove this could be the following observation: take $(\mathbf{S}^0, \mathbf{T}^0) = (\mathbf{S}, \mathbf{T})$ two any culminating paths, and define

$$(\mathbf{S}^{t+1}, \mathbf{T}^{t+1}) = (\phi(\mathbf{S}^t, I_t, \varepsilon_t, \delta_t), \phi(\mathbf{T}^t, I_t, \varepsilon_t, \delta_t)),$$

where $I_t, \varepsilon_t, \delta_t$ are those in Algorithm 1. The sequence $(\| \mathbf{S}^t - \mathbf{T}^t \|_\infty)_t$ is decreasing throughout the process. Unfortunately we cannot get a satisfactory bound for the time needed for this quantity to decrease by one.

3.2 Related works

Bounding mixing times via a contraction property over the transportation metric is quite a standard technique, the main ideas dating back to Dobrushin (1950's). A modern introduction is made in [10]. For geodesic spaces, this technique has been developped in [3] under the name *path coupling*.

As mentioned in the introduction, the Markov chain P on lattice paths with uniform weights $p_i = 1/n$ has in fact already been introduced for paths starting and ending at zero (sometimes called *bridges*) in [11], and its mixing time has been estimated in [14]. Wilson also proves a mixing time of order $n^3 \log n$, by showing that (3) holds with a different distance (namely, a kind of Fourier transform of the heights of the paths)[ii]. This is the concavity of this Fourier transform which gives a good mixing time, exactly as the concavity of our p_i's speeds up the convergence of our chain.

Wilson's method is developped only for bridges in [14] and it is not completely straightforward to use it when the endpoints are not fixed. For instance, take $n = 7$ and $a = b = 1$, and consider the paths $+++--++$ and $---+++-$. There are more "bad moves" (moves that take away these paths) than "good moves".

[ii] Notice that a, b do not have the same meaning in Wilson's paper: a (resp. b) stands for the number of positive (resp. negative) steps.

4 *Coupling From The Past* with P

Propp-Wilson's Coupling From The Past (CFTP) [13] is a very general procedure for the exact sampling of the stationary distribution of a Markov chain. It is efficient if the chain is monotonous with respect to a certain order relation $\preceq$ on the set V of vertices, with two extremal points denoted $\hat{0}, \hat{1}$ (*i.e.* such that $\hat{0} \preceq x \preceq \hat{1}$ for any vertex x). This is the case here for each family $\mathcal{C}_n$,$\mathcal{W}_n$,$\mathcal{E}_n$,$\mathcal{M}_n$, with the partial order

$$\mathbf{S} \preceq \mathbf{T} \text{ iff } S_i \leq T_i \text{ for any } i.$$

For the family $\mathcal{M}_{10}$ with $a = 1, b = -2$ for instance, we have

$$\hat{0} = \hat{0}_{\text{meanders}} = (1, 1, -2, 1, 1, -2, 1, 1, -2, 1),$$
$$\hat{1} = \hat{1}_{\text{meanders}} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1).$$

We describe CFTP, with our notations, in Algorithm 2.

Algorithm 2 CFTP: Exact sampling of a path in $\mathcal{A}_n$

$\mathbf{S} \leftarrow \hat{0}$, $\mathbf{T} \leftarrow \hat{1}$
$\ldots, I_{-2}, I_{-1} \leftarrow$ i.i.d. r.v. with law $\mathbf{p}$
$\ldots, \varepsilon_{-2}, \varepsilon_{-1} \leftarrow$ i.i.d. uniform r.v. in $\{\uparrow, \downarrow\}$
$\ldots, \delta_{-2}, \delta_{-1} \leftarrow$ i.i.d. uniform r.v. in $\{+, -\}$
$\tau = 1$
repeat
  $\mathbf{S} \leftarrow \hat{0}$, $\mathbf{T} \leftarrow \hat{1}$
  for $t = -\tau$ to 0 **do**
    if $\phi(\mathbf{S}, I_t, \varepsilon_t,)$ is in $\mathcal{A}_n$ **then** $\mathbf{S} \leftarrow \phi(\mathbf{S}, I_t, \varepsilon_t, \delta_t)$
    if $\phi(\mathbf{T}, I_t, \varepsilon_t,)$ is in $\mathcal{A}_n$ **then** $\mathbf{T} \leftarrow \phi(\mathbf{T}, I_t, \varepsilon_t, \delta_t)$
  end for
  $\tau \leftarrow 2\tau$
until $\mathbf{S} = \mathbf{T}$

We refer to ([8],Chap.10) for a very clear introduction to CFTP, and we only outline here the reasons why this indeed gives an exact sampling of the stationary distribution.

- The output of the algorithm (if it ever ends!) is the state of the chain P that has been running "since time $-\infty$", and thus has reached stationnarity.

- The exit condition $\mathbf{S} = \mathbf{T}$ ensures that it is not worth running the chain from T steps earlier, since the trajectory of any lattice path $\hat{0} \preceq \mathbf{R} \preceq \hat{1}$ is "sandwiched" between those of $\hat{0}, \hat{1}$, and therefore ends at the same value.

Proposition 8 *Algorithm 2 ends with probability* 1 *and returns an exact sample of the uniform distribution over* $\mathcal{A}_n$*. For the families* $\mathcal{W}_n$,$\mathcal{E}_n$,$\mathcal{M}_n$*, this takes on average* $\mathcal{O}(n^3(\log n)^2)$ *time units.*

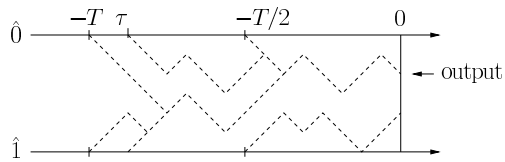


Fig. 2: A sketchy representation of CFTP : trajectories starting from $\hat{0}, \hat{1}$ at time $-T/2$ don't meet before time zero, while those starting at time $-T$ do.

Let us mention that in the case where the mixing time is not rigorously known, Algorithm 2 (when it ends) outputs an exact uniform sample and therefore is of main practical interest compared to MCMC.

Proof of Proposition 8: It is shown in [13] that Algorithm 2 returns an exact sampling in $\mathcal{O}(t_{\text{mix}} \log H)$ runs of the chain, where t_{mix} is defined in (6) and H is the length of the longest chain of states between $\hat{0}$ and $\hat{1}$. It is a consequence of the proof of Lemma 4 that $H = \mathcal{O}(n^2)$. We have seen that $t_{\text{mix}} = \mathcal{O}(n^3 \log n)$. (Recall that each test in Algorithm 2 takes, on average, $\mathcal{O}(1)$ time units.) $\square$

We recall that CFTP has a major drawback compared to MCMC. For the algorithm to be correct, we have to reuse the same random variables $I_t, \varepsilon_t, \delta_t$, so that space-complexity is in fact linear in $n^3(\log n)^2$. This may become an issue when n is large.

5 Concluding remarks and simulations

1. In Fig.3, we show simulations of the three kinds of paths, for $a = 1, b = 2, n = 600$. We observe that the final height of the culminating path is very low (about 30), it would be interesting to use our algorithm to investigate the behaviour of this height when $n \to \infty$; this question was left open in [2].

2. One may wonder to what extent this work applies to other families $\mathcal{A}_n$ of paths. The main assumption is that the family of paths should be a geodesic space w.r.t. distance d_1. This is true for example if the following condition on $\mathcal{A}_n$ is fulfilled:

$$(R, T \in \mathcal{A}_n \text{ and } R \preceq S \preceq T) \Rightarrow S \in \mathcal{A}_n.$$

Notice however that this is quite a strong requirement, and it is not verified for culminating paths for instance.

3. A motivation to sample random paths is to make and test guesses for some functionals of these paths, taken on average over $\mathcal{A}_n$. Consider a function $f : \mathcal{A}_n \to \mathbb{R}$, we want an approximate value of $\pi(f) := \text{card}(\mathcal{A}_n)^{-1} \sum_{s \in \mathcal{A}_n} f(s)$, if the exact value is out of reach by calculation. We estimate this quantity by

$$\hat{\pi}(f) := \frac{1}{T} \sum_{t=1}^{T} f\left(\mathbf{S}(t)\right), \tag{7}$$

(recall that $S(t)$ is the value of the chain at time t). For Algorithm 1 to be efficient in practice, we have to bound

$$\mathbb{P}\left(|\pi(f) - \hat{\pi}(f)| > r\right), \tag{8}$$

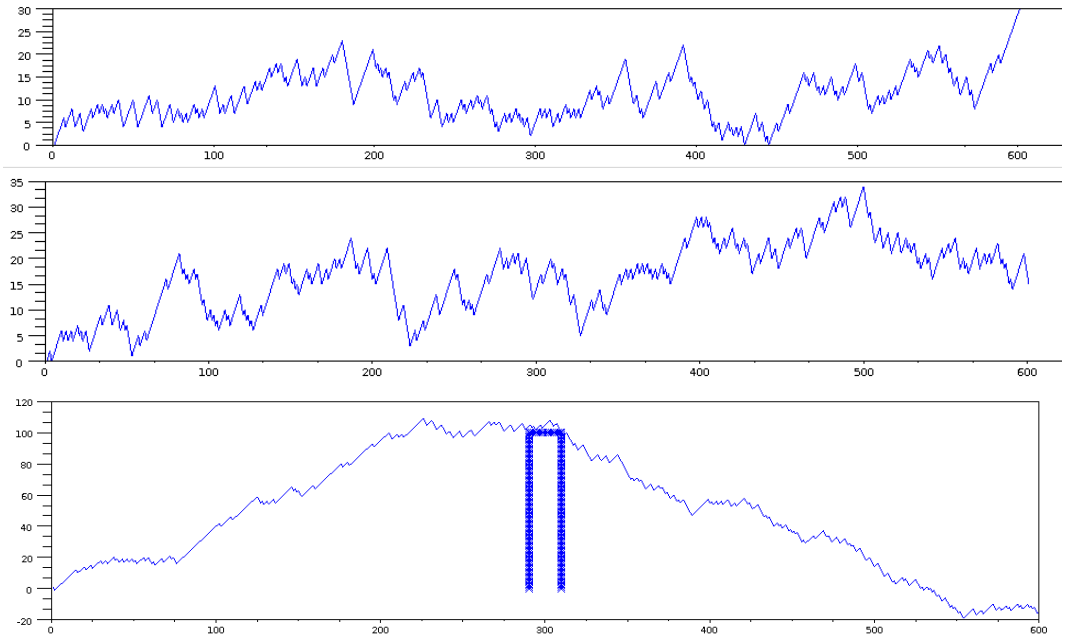


Fig. 3: (Almost) uniform paths of length 600, with $a = 1, b = 2$. From top to bottom: a culminating path, a meander, a path with wall (shown by an arch).

for any fixed $r > 0$, by a non-asymptotic (in T) quantity. This can be done with ([9], Th.4-5), in which one can find concentration inequalities for (8). The sharpness of these inequalities depends on κ and on the geometrical structure of $\mathcal{A}_n$.

Aknowledgements. Many thanks to Frédérique Bassino and the other members of ANR GAMMA for the support ; I also would like to thank Élie Ruderman for the English corrections. A referee raised a serious error in the first version of this paper, I am grateful to them.

References

[1] C. Banderier and Ph. Flajolet. Basic Analytic Combinatorics of Directed Lattice Paths, *Theoretical Computer Science* **281** (1):37-80 (2002).

[2] M.Bousquet-Mélou and Y. Ponty. Culminating paths, *Discrete Math and Theoretical Computer Science* **10** (2):125-152 (2008).

[3] R.Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. Proceedings of the *38th Annual Symposium on Foundations of Computer Science*, p.223-231 (1997).

[4] N.Destainville. Flip dynamics in octagonal rhombus tiling sets, *Physical Review Letters* **88** (2002).

[5] P.Diaconis and D.W.Stroock. Geometric Bounds for Eigenvalues of Markov Chains, *Annals of Applied Probability* **1** (1):36-61 (1991).

[6] Ph.Duchon. On the enumeration and generation of generalized Dyck words. *Discrete Mathematics* **225** (1-3)121–135 (2000).

[7] M.E.Fisher. Walks, Walls, Wetting, and Melting. *Journal of Statistical Physics* **34** (5):667-729 (1984).

[8] O.Häggström. *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society (2002).

[9] A.Joulin and Y.Ollivier. Curvature, concentration, and error estimates for Markov chain Monte Carlo. To appear in *Annals of Probability*, `arXiv:0904.1312` (2009).

[10] D.A.Levin, Y.Peres, and E.L.Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society (2009).

[11] M.Luby, D.Randall, A.Sinclair. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing*, **31** (1):167–192 (2001).

[12] Y.Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis* **256** (3):810-864 (2009).

[13] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9** (1):223-252, (1996).

[14] D.B.Wilson. Mixing times of Lozenge tiling and card shuffling Markov chains. *Annals of Applied Probability* **14** (1):274–325 (2004).