On Sampling Simple Paths in Planar Graphs According to Their Lengths

Sandro Montanari[1] and Paolo Penna[2]

[1] Department of Computer Science, ETH Zurich, Switzerland
sandro.montanari@inf.ethz.ch
[2] LIAFA, Université Paris Diderot, France
penna@liafa.univ-paris-diderot.fr

Abstract. We consider the problem of sampling simple paths between two given vertices in a planar graph and propose a natural Markov chain exploring such paths by means of "local" modifications. This chain can be tuned so that the probability of sampling a path depends on its *length* (for instance, output shorter paths with higher probability than longer ones). We show that this chain is always ergodic and thus it converges to the desired sampling distribution for any planar graph. While this chain is not rapidly mixing in general, we prove that a simple restricted variant is. The restricted chain samples paths on a 2D lattice which are monotone in the vertical direction. To the best of our knowledge, this is the first example of a rapidly mixing Markov chain for sampling simple paths with a probability that depends on their lengths.

1 Introduction

Sampling (or generating) a "random" object from a large set of combinatorial objects is a fundamental problem arising in Statistical Physics, Mathematics, and Computer Science. Because the number of such objects is typically huge (i.e., exponential in the size of the input), direct enumeration is unfeasible. An efficient sampling procedure is thus an important tool for studying statistical properties of "typical" instances. For many problems, the tasks of uniformly sampling and counting the number of objects are equivalent [13] and, in most cases, $\#P$-hard. This includes counting simple paths in graphs [24], also when restricting to planar graphs [20]. For these cases, sampling is instead consider according to distributions that are "close" to the desired one [23]. A most relevant technique for the design of this kind of sampling procedures is the *Markov chain Monte Carlo* method [3].

We consider the task of sampling simple paths between two given vertices in planar graphs according to a fixed probability distribution using Markov chain Monte Carlo. Since in several applications it is natural to ask for paths of some fixed length, we thus consider the *weighted* version of the sampling problem in which the distribution depends on the length of the paths.

The Markov chain Monte Carlo method involves the design of a Markov chain whose states are the objects we wish to sample, and whose stationary distribution is the probability we want to use to sample them. The sampling procedure is then a "random walk" on the chain for a fixed number of steps, until we are certain that the probability of sampling one object is (approximately) the stationary distribution of the chain. A most critical part of this method is proving that the Markov chain is *rapidly mixing*, i.e., the number of steps needed to reach the stationary distribution is polynomially bounded by the size of the input.

While rapidly mixing Markov chains are known for several hard problems, like graph coloring [6–8, 11], knapsack [19], perfect matchings [12], independent sets [2, 9, 16], there is essentially no positive result for the case of simple st-paths on general graphs. The only chain proposed for this setting is by Roberts and Kroese [22], but it is however *not* rapidly mixing.

Most of the positive results consider restricted paths over a lattice structure. In the simplest instance of these restrictions we have paths using only downward and rightward edges of the two-dimensional grid. This case has been analyzed [15] also for *multiple* source-destination paths, where the chain can "get stuck" because all paths must be disjoint and thus some non-local moves are introduced. If we further impose the paths to stay above the main diagonal of a square grid, the so-called *staircase walks*, the number of such paths is given by the famous Catalan numbers [5]. Martin and Randall [17] consider a Markov chain for sampling such paths where the weight of a path is the number of times it hits the diagonal. In the sampler by Greenberg *et al.* [10], the weight of a path is instead the number of faces below it (the authors also consider the more general case of higher dimensional lattices). Finally, Randall and Sinclair [21] consider the case where only one end of the path is fixed, and provide an efficient sampler for all such paths of a given length in the *infinite* d-dimensional lattice.

1.1 Our Results

We study a natural Markov chain in which a current st-path in a given (undirected, unweighted) planar graph is modified according to a simple local rerouting operation (see Section 3). Roughly speaking, rerouting operations resulting in longer paths are "accepted" only with small probability, while those resulting in shorter paths are always accepted. We show that the chain always converges to the Gibbs distribution on the paths weighted according to their lengths. In other words, the probability of sampling a specific path x depends only on its length $\ell(x)$ and it is of the form

$$\pi(x) \propto \lambda^{\ell(x)}, \tag{1}$$

where $\lambda > 0$ is a parameter that can be used to "tune" the chain. Setting $\lambda = 1$ yields a uniform sampler over all st-paths, while smaller/larger values provide samples biased towards shorter/longer paths.

Despite this chain being *not* rapidly mixing in general planar graphs, we obtain an efficient sampler with Gibbs distribution (1) for the following setting, depicted in Figure 1. The paths from s to t are monotone in the vertical direction and the graph is any sub-grid of the 2D lattice without internal holes. The

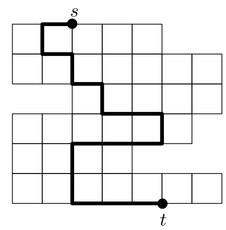

Fig. 1: An example of a vertical-monotone path in a sub-grid.

new Markov chain is a restriction of the original one maintaining the "vertical-monotonicity" of the paths (Section 5). Our main technical contribution is a rigorous proof that this chain is rapidly mixing for all $\lambda \in (0, 1]$. In the proof we combine the technique of path coupling without contraction [1] with the idea of modifying the chain by making some transitions "more lazy". Note that in this restricted setting, an efficient sampler can also be obtained using dynamic programming (see Appendix B for the details). However, our main interest is in the analysis of the mixing time of the proposed Markov chain, which is a variant of a well-known "mountain/valley" chain [17].

We show that our results are tight in the following sense (Section 4). First, the original "unrestricted" chain is *not* rapidly mixing for $\lambda = 1$ in some planar graphs. This is true even for sub-graphs of the 2D lattice, and thus for the chain sampling all paths in a grid, without restricting to vertical-monotone ones. Both for the restricted and the unrestricted chains, we show that the mixing time is exponential in the number of vertices for *every* $\lambda > 1$. The latter result is in part expected because determining if a planar graph has an Hamiltonian path is NP-hard (and for sufficiently large λ a sampler can be used to solve this problem). However, our negative results on the Markov chain are stronger in the sense that the chain remains slowly mixing even for very simple graphs where this problem can be easily solved. Thus, these results give a certain indication of the limitation of "local" chains. As for the case $\lambda = 1$, the existence of an efficient (uniform) sampler for planar graphs remains an interesting open problem.

2 Preliminaries

Planar graphs. Given a planar graph $G = (V, E)$, we use n and m to denote respectively the number of vertices $|V|$ and of undirected edges $|E|$. By planarity, the vertices in V can be drawn as points in the plane in a way such that the edges in E are non-crossing curves; we denote such a drawing as a *plane embedding* of G. In a plane embedding, any maximal region of the plane enclosed by edges of E is called a *face*; the infinite region not enclosed by any edge is called the *outer face*. We use f to denote the overall number of faces, including the outer face. According to Euler's formula, the number of faces satisfies $f = m - n + 2$. Note also that in any planar graph $f \in O(n)$. Namely, $f \leq 2n - 4$ with equality achieved by triangulated graphs, i.e., planar graphs in which every face is a triangle.

A *path* is a sequence of vertices $x = (v_1, \dots, v_l)$ such that $(v_i, v_{i+1}) \in E$, for all $i = 1, \dots, l-1$; we denote the number of edges along x as $|x| = l - 1$. In a *simple path* no vertex appears more than once. For any two vertices s and t, an *st-path* is a simple path starting at s and ending at t. Without loss of generality, we assume the graph to be 2-connected (we can otherwise easily reduce to this case).

Markov chains and mixing time. We consider Markov chains $\mathcal{M}$ whose state space Ω is finite. In our application, Ω is the set of all simple st-paths in a given planar graph. The transition matrix $P \in \mathbb{R}^{|\Omega| \times |\Omega|}$ defines the transitions of $\mathcal{M}$. That is, $P(x, y)$ is the probability that the chain moves from state x to state y in one step. Thus, $P^t(x, y)$ is the probability of moving in t steps from state x to state y, where P^t is the t^{th} power of matrix P.

A Markov chain as above is *irreducible* if, for all $x, y \in \Omega$, there exists a $t \in \mathbb{N}$ such that $P^t(x, y) > 0$. In other words, every state can be reached with non-zero probability regardless of the starting state. A Markov chain is *aperiodic* if, for all $x \in \Omega$, $\gcd\{t \in \mathbb{N} \mid P^t(x, x) > 0\} = 1$. It is well known [14] that an irreducible and aperiodic Markov chain converges to its unique *stationary distribution* π. That is, there exists a unique vector $\pi \in \mathbb{R}^{|\Omega|}$ such that $\pi P = \pi$ and, for all $x, y \in \Omega$, it holds

$$\lim_{t \to \infty} P^t(x, y) = \pi(y).$$

An aperiodic and irreducible Markov chain is called *ergodic*.

The *mixing time* of a Markov chain is the time needed for the distribution $P^t(x, \cdot)$ to get "sufficiently close" to the stationary distribution for any starting state x. Formally, the *mixing time* is defined as

$$t_{mix}(\epsilon) := \min_{t \in \mathbb{N}} \max_{x \in \Omega} \{||P^t(x, \cdot) - \pi||_{TV} \leq \epsilon\}, \tag{2}$$

where $||P^t(x, \cdot) - \pi||_{TV} = \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|$ is the total variation distance. It is common [14] to define $t_{mix} := t_{mix}(1/4)$ since $t_{mix}(\epsilon) \leq \lceil \log_2(1/\epsilon) \rceil t_{mix}$. A Markov chain is *rapidly mixing* if $t_{mix}(\epsilon)$ is bounded from above by a polynomial in $\log(|\Omega|)$ and in $\log(1/\epsilon)$.

A rapidly mixing Markov chain can be used to efficiently sample elements from Ω with probability arbitrarily close to π. Simply simulate a random walk on the chain from an arbitrary initial state x for $t = t_{mix}(\epsilon)$ time steps and return the state of the chain at time t. According to (2), the probability $P^t(x, y)$ of the returned state y is approximately $\pi(y)$.

3 A Markov Chain for Planar Graphs

We now define a Markov chain $\mathcal{M}_{paths}$ whose state space Ω is the set of all simple st-paths of a given planar graph. The transitions of $\mathcal{M}_{paths}$ are defined by rerouting a current path along one of its adjacent faces (see Figure 2).

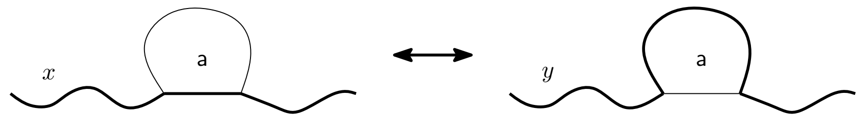

Fig. 2: Rerouting of x along face a.

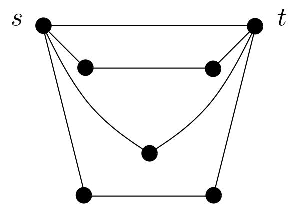

Fig. 3: A graph admitting multiple embeddings with different sets of faces.

Definition 1. *Let x be an st-path and* a *be a face adjacent to at least one edge of x. We say that x* can be rerouted *along* a *if the edges in x and* a *form a single sub-path of x of length at least one, and the path y obtained by replacing all edges common to x and* a *with the edges in* a *that do not belong to x is simple. In this case, the* rerouting operation *consists of replacing x with y.*

Note that we forbid rerouting operations that reduce the length of the current path by introducing a cycle and short-cutting it afterwards. The reason is that these operations are not "reversible" and would prevent us from using well-established methods to determine the stationary distribution.[3]

Markov chain $\mathcal{M}_{paths}$. Given a parameter $\lambda > 0$, the transition from the current state x to the next one are defined according to the following rule:

1. With probability $\frac{1}{2}$ do nothing. Otherwise,
2. Select a face a uniformly at random. If x cannot be rerouted along a then do nothing. Otherwise,
3. Move to the path y obtained by rerouting x along a with probability

$$A(x, y) := \min\left\{1, \frac{\lambda^{|y|}}{\lambda^{|x|}}\right\},$$

   and do nothing with remaining probability $1 - A(x, y)$.

For $\lambda < 1$, rerouting operations increasing the length of the current path by ℓ are accepted with probability λ^{ℓ}, while those reducing it are always accepted. The converse happens for $\lambda > 1$.

[3] The analysis of non-reversible Markov chains is in general rather difficult and it is considered an interesting problem also for simple chains [4].

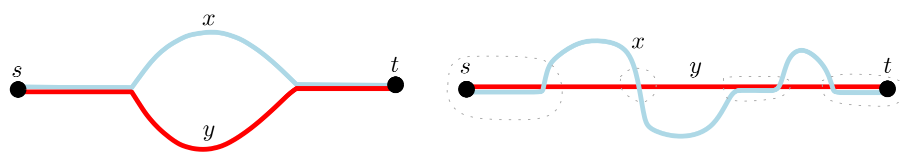

Fig. 4: Paths with $\Delta_{xy} = 2$ (left) and $\Delta_{xy} = 4$ (right).

Remark 2. We stress that a planar graph can have different embeddings, and the transitions of $\mathcal{M}_{paths}$ depend on the particular given one. Figure 3 shows an example of a graph admitting an embedding in which there is a face with three vertices. In the next section we show that the chain is ergodic for any embedding.

Remark 3. For the sake of simplicity we forbid rerouting along the outer face. Doing so would not change the presented results significantly except in making the proofs less readable.

4 Analysis of $\mathcal{M}_{paths}$

Ergodicity. For all states $x \in \Omega$ it holds that $P(x, x) \geq 1/2$, and thus $\mathcal{M}_{paths}$ is aperiodic. To show ergodicity, it then suffices to prove that any two states $x, y \in \Omega$ are connected by a path with non-zero probability. Before stating the main theorem, we first introduce the following notion of "distance" between st-paths.

Definition 4. *Given two st-paths x and y, a* maximal sub-path common to x and y *is an ordered sequence of vertices appearing in both paths and not contained in a longer sequence with the same property. We let Δ_{xy} denote the number of maximal sub-paths common to x and y.*

Note that the definition also allows "degenerate" sub-paths of just one vertex. For instance, if x and y have only the starting and the ending vertices in common, then $\Delta_{xy} = 2$. Figure 4 shows examples of paths with different values of Δ_{xy}.

Theorem 5. *For any plane embedding of a given graph and any pair of vertices s and t, the Markov chain $\mathcal{M}_{paths}$ is ergodic with diameter at most $2n^2$.*

Proof (Sketch). Any two paths x and y such that $\Delta_{xy} = 2$ are connected by a sequence of at most f paths. Moreover, for $\Delta_{xy} > 2$, there is an intermediate path x' such that $\Delta_{xx'} = 2$ and $\Delta_{x'y} < \Delta_{xy}$. Since $\Delta_{xy} < n$ and $f < 2n$ in planar graphs, the theorem follows. □

Stationary distribution. To characterize the stationary distribution of $\mathcal{M}_{paths}$, we show that, for some probability distribution π, it holds

$$\pi(x) \cdot P(x, y) = \pi(y) \cdot P(y, x), \qquad \text{for all } x, y \in \Omega. \tag{3}$$

It is well-known that π is then the stationary distribution of $\mathcal{M}_{paths}$; this property is known as the *detailed balance condition* [14, Proposition 1.19].

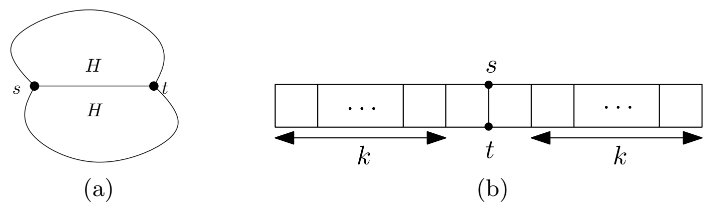

Fig. 5: Graphs with exponential bottleneck ratio.

Theorem 6. *For any planar graph and any two vertices s and t, the stationary distribution of the Markov chain $\mathcal{M}_{paths}$ is*

$$\pi(x) = \frac{\lambda^{|x|}}{Z(\lambda)}, \qquad \text{where } Z(\lambda) = \sum_{z \in \Omega} \lambda^{|z|}. \tag{4}$$

The parameter λ can be used to tune the stationary distribution of $\mathcal{M}_{paths}$. For example, by setting $\lambda = 1$ the stationary distribution is the uniform distribution over all simple st-paths.

Mixing time. We conclude this section by providing some negative results concerning the mixing time of $\mathcal{M}_{paths}$.

Theorem 7. *There exist planar graphs G and vertices s, t such that $\mathcal{M}_{paths}$ with $\lambda = 1$ is not rapidly mixing.*

Proof. We apply the well-known bottleneck theorem [14] which says that, for any subset of states $R \subset \Omega$ such that $\pi(R) \leq 1/2$, the following bound on the mixing time of the chain holds:

$$t_{mix} \geq \frac{\pi(R)}{4Q(R, \bar{R})} \qquad \text{where } Q(R, \bar{R}) = \sum_{x \in R, y \in \bar{R}} \pi(x)P(x, y). \tag{5}$$

Let G be the planar graph obtained by combining two copies of a planar graph H as follows. The two copies share only a single edge connecting s and t, and all other vertices and edges in H are duplicated (see Figure 5a). The set R consists of the subset of st-paths of G that use only edges in one of the two copies, say the upper one. Note that R contains the common single-edge path $x^* = (s, t)$, that this is the only path in R with transitions to some $y \in \bar{R}$, and that there are at most two transitions from x^* to some other state (each edge is adjacent to at most two faces). Therefore

$$Q(R, \bar{R}) = \sum_{x \in R, y \in \bar{R}} \pi(x)P(x, y) \leq 2\pi(x^*) = 2/|\Omega|.$$

In order to apply the bottleneck theorem we need $\pi(R) \leq 1/2$. This can be easily achieved by adding two more paths to the bottom copy of H (recall that

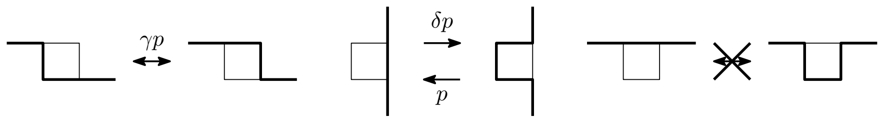

Fig. 6: The transitions of $\mathcal{M}_{mon}$; p is the probability of selecting a face u.a.r.

R consists of all paths in the upper copy of H). We then get

$$t_{mix} \geq \frac{\pi(R)}{4Q(R,\bar{R})} \geq \frac{|R|/|\Omega|}{8/|\Omega|} = |R|/8.$$

□

Corollary 8. *There exists an infinite family of planar graphs such that the mixing time of $\mathcal{M}_{paths}$ satisfies $t_{mix} \in \Omega(\lambda^{n/2})$ for all $\lambda > 1$.*

Proof (Sketch). The claim follows by considering the bottleneck ratio of the graph in Figure 5b. □

We note that Theorem 7 holds also for graphs that have a very simple structure, like two square grids sharing only edge (s, t), and for outerplanar graphs.

5 A Rapidly Mixing Chain for Vertical-Monotone Paths

We now present the rapidly mixing Markov chain $\mathcal{M}_{mon}$, which is a natural modification of $\mathcal{M}_{paths}$ for the case where the graph is a sub-graph of the two-dimensional lattice (grid) with no holes. That is, every face is either a cell of the grid or it is the outer face. The chain $\mathcal{M}_{mon}$ samples paths that are *vertical-monotone*, that is, that are only monotone in the vertical direction (if we follow the path from s to t, it never goes up). We thus assume that s lies above or at the same y-coordinate of t. Though it is straightforward to generate such paths *uniformly* at random, our goal is a weighted sampler with probability biased towards shorter paths according to the parameter λ, i.e., with distribution of the form (1).

Markov chain $\mathcal{M}_{mon}$. The chain is a modification of $\mathcal{M}_{paths}$ in which some transitions are disallowed and others are "more lazy" (see Figure 6). Specifically, the chain does not allow to replace an horizontal edge with three edges, and transitions swapping two consecutive edges of a face are only performed with probability

$$\gamma := \frac{1+\delta}{2} \quad \text{where } \delta = \lambda^2 \text{ and } \lambda \in (0, 1].$$

This choice of γ will be useful for the analysis of the mixing time. Note that we restrict to the case $\lambda \leq 1$, because the lower bound for $\lambda > 1$ of Corollary 8 holds also for $\mathcal{M}_{mon}$. Note further that $\mathcal{M}_{mon}$ is ergodic with diameter $\leq 2f$, and its stationary distribution is the same as for $\mathcal{M}_{paths}$.

5.1 Mixing Time of $\mathcal{M}_{mon}$

To bound the mixing time we use the method of *path coupling without contraction* [1]. A *path coupling* for a chain $\mathcal{M}$ can be specified by providing distributions

$$\mathbb{P}_{x,y}[X = x', Y = y'], \qquad \text{for all } x, y \in \Omega \text{ such that } P(x,y) > 0, \tag{6}$$

satisfying, for all $x, y \in \Omega$ such that $P(x,y) > 0$,

$$\mathbb{P}_{x,y}[X = x'] = P(x, x') \qquad \text{for all } x' \in \Omega, \tag{7}$$

$$\mathbb{P}_{x,y}[Y = y'] = P(y, y') \qquad \text{for all } y' \in \Omega. \tag{8}$$

We use ρ to denote the shortest-path distance in the Markov chain, i.e., $\rho(x,y)$ is the minimum number of transitions to go from x to y.

Lemma 9 (Theorem 2 in [1]). *Suppose we have a path coupling for a Markov chain $\mathcal{M}$ such that, for all x, y with $P(x,y) > 0$, it holds*

$$\mathbb{E}_{x,y}[\rho(X,Y)] \leq 1. \tag{9}$$

Then, the Markov chain $\mathcal{M}^*$ with transition matrix $P^* = (P + p_{min} I)/(1 + p_{min})$ has mixing time $t^*_{mix} \in O\left(D^2/p_{min}\right)$, where p_{min} and D denote respectively the smallest non-zero transition probability and the diameter of $\mathcal{M}$.

Note that $\mathcal{M}^*$ is the chain with transition probabilities

$$P^*(x,y) := \begin{cases} \frac{P(x,x) + p_{min}}{1 + p_{min}} & \text{if } y = x, \\ \frac{P(x,y)}{1 + p_{min}} & \text{otherwise.} \end{cases}$$

Therefore $\mathcal{M}^*$ and $\mathcal{M}$ have the same stationary distribution. This suggests naturally to run the chain $\mathcal{M}^*_{mon}$ for efficiently sample vertical-monotone paths.

Path coupling for $\mathcal{M}_{mon}$. For the sake of clarity, for every face a we define the following shorthand:

$$p_{\mathsf{a}}(x) := P(x, x \oplus \mathsf{a}),$$

where $x \oplus \mathsf{a}$ denotes the path obtained by rerouting x along a. We define a path coupling by specifying, for every pair (x,y) such that x and y differ in one face d, the probabilities in (6) to move to a pair (x', y'):

$$(x,y) \mapsto (x \oplus \mathsf{d}, y) \qquad \text{with probability } p_{\mathsf{d}}(x), \tag{10}$$

$$(x,y) \mapsto (x, y \oplus \mathsf{d}) \qquad \text{with probability } p_{\mathsf{d}}(y), \tag{11}$$

and for every other face $\mathsf{a} \neq \mathsf{d}$

$$(x,y) \mapsto (x \oplus \mathsf{a}, y \oplus \mathsf{a}) \qquad \text{with probability } \min\{p_{\mathsf{a}}(x), p_{\mathsf{a}}(y)\}, \tag{12}$$

$$(x,y) \mapsto (x \oplus \mathsf{a}, y) \qquad \text{with probability } \max\{0, p_{\mathsf{a}}(x) - p_{\mathsf{a}}(y)\}, \tag{13}$$

$$(x,y) \mapsto (x, y \oplus \mathsf{a}) \qquad \text{with probability } \max\{0, p_{\mathsf{a}}(y) - p_{\mathsf{a}}(x)\}. \tag{14}$$

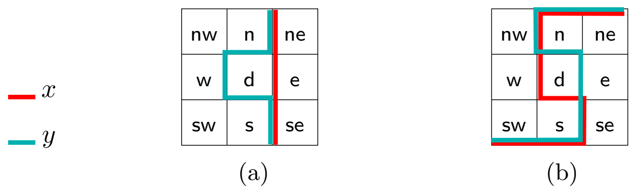

Fig. 7: Analysis of path coupling for grids (main idea).

Finally, with all remaining probability

$$(x, y) \mapsto (x, y). \tag{15}$$

One can easily check that this is indeed a path coupling, that is, (7)-(8) are satisfied. The difficulty is in proving the condition necessary to apply Lemma 9.

Lemma 10. *The path coupling defined above satisfies condition* (9)*.*

Proof (Sketch). In the coupling, the distance between two paths x and y can either increase by 1 or decrease by 1. Since the initial distance is 1, we can thus write the expected distance after one coupling step as

$$\mathbb{E}_{x,y}[\rho(X, Y)] = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 = 1 + p_2 - p_0,$$

where p_0 is the probability that the distance decreases, p_2 is the probability that it increases, and $p_1 = 1 - p_0 - p_2$.

The transitions reducing the distance are those corresponding to (10)-(11), that always happen with probability

$$p_0 = p_{\mathsf{d}}(x) + p_{\mathsf{d}}(y) = p(1 + \delta) = 2p\gamma,$$

where $p = \frac{1}{2f}$ is the probability to pick a face uniformly at random.

The distance increases instead to 2 if, for instance, the coupling uses a face for rerouting y while x stays the same. Since the coupling attempts to reroute both paths whenever possible (12), the probability that the distance becomes 2 is due to (13) and (14) only. We thus have to consider only faces for which the probability of rerouting is *different* for the two paths, that is, $p_{\mathsf{a}}(x) \neq p_{\mathsf{a}}(y)$. We illustrate the proof only for the cases in Figure 7; the remaining ones can be found in Appendix A.4.

For the case of Figure 7a, the transitions increasing the distance correspond to the four faces around d (namely, $\mathsf{n}, \mathsf{s}, \mathsf{e}, \mathsf{w}$). We thus get that

$$p_2 = p\delta + p\delta + p(\gamma - \delta) + p(\gamma - \delta) = 2p\gamma,$$

and conclude then that (9) holds for this case.

In the scenario of Figure 7b we can apply the same analysis. Here it is crucial to observe that faces ne and sw do not contribute to p_2 because in $\mathcal{M}_{mon}$ these transitions are not allowed. This is one of the cases where the monotonicity of the paths plays a crucial role. We thus get that

$$p_2 = p(1-\gamma) + p(1-\gamma) + p\delta + p\delta = 2\gamma p.$$

Therefore, (9) holds also for this case. □

Since the diameter of $\mathcal{M}_{mon}$ is $O(n)$ and the minimum non-zero transition probability is $p_{min} = \frac{\lambda^2}{2f}$, we can establish the following bound on the mixing time.

Theorem 11. *The mixing time of* $\mathcal{M}^*_{mon}$ *is* $O\left(n^3/\lambda^2\right)$ *for every* $\lambda \in (0, 1]$.

6 Conclusion

We have studied a natural Markov chain $\mathcal{M}_{paths}$ for sampling *st*-paths in any given planar graph. We have shown that this chain is always ergodic and its stationary distribution is the Gibbs distribution on the paths weighted according to the number of edges. The chain is, in general, not rapidly mixing, but it might be possible that, in some graphs, modifications of this "basic" chain yield a rapidly mixing one. We have shown that this is indeed the case when restricting the sampling to vertical-monotone paths on sub-graphs of the 2D lattice. Another possible direction might be to introduce *non-local* transitions. In this case, the chain should probably be designed "ad-hoc" for a specific class of graphs. It would also be interesting to find graph classes for which $\mathcal{M}_{paths}$ is rapidly mixing. We conjecture that this might for example be the case for regular square grids.

An interesting related question is how to *count* the number of paths of a certain length. Note that the case of vertical-monotone paths provides an excellent example to also show the limitations of our local chain. Indeed, in this case the dynamic programming algorithm for counting all paths (even non simple) in general graphs would work here. Moreover, the procedure can be easily adapted to sample paths with uniform distribution among those with a fixed length. This procedure in the end leads to an exact Gibbs sampler (1) for all values of $\lambda > 0$ (see Appendix B for the details).

Acknowledgments We are grateful to Francesco Pasquale for comments on an earlier version of this work. We also wish to thank the anonymous referees for pointing out relations to the class of outerplanar graphs, and the use of dynamic programming for the vertical-monotone paths. Part of this work has been done while the second author was at ETH Zurich. This work is supported by the EU FP7/2007-2013 (DG CONNECT.H5-Smart Cities and Sustainability), under grant agreement no. 288094 (project eCOMPASS), and by the French ANR Project DISPLEXITY.

References

1. Bordewich, M., Dyer, M.E.: Path coupling without contraction. Journal of Discrete Algorithms 5(2), 280–292 (2007)

2. Bordewich, M., Dyer, M.E., Karpinski, M.: Path coupling using stopping times and counting independent sets and colorings in hypergraphs. Random Struct. Algorithms 32(3), 375–399 (2008)
3. Bubley, R.: Randomized algorithms: approximation, generation, and counting. Springer (2011)
4. Diaconis, P., Holmes, S., Neal, R.M.: Analysis of a nonreversible Markov chain sampler. Annals of Applied Probability pp. 726–752 (2000)
5. Došlić, T.: Seven (lattice) paths to log-convexity. Acta applicandae mathematicae 110(3), 1373–1392 (2010)
6. Dyer, M., Flaxman, A.D., Frieze, A.M., Vigoda, E.: Randomly coloring sparse random graphs with fewer colors than the maximum degree. Random Struct. Algorithms 29(4), 450–465 (2006)
7. Dyer, M., Frieze, A., Hayes, T.P., Vigoda, E.: Randomly coloring constant degree graphs. Random Struct. Algorithms 43(2), 181–200 (2013)
8. Dyer, M.E., Frieze, A.M.: Randomly coloring random graphs. Random Struct. Algorithms 36(3), 251–272 (2010)
9. Dyer, M.E., Greenhill, C.S.: On markov chains for independent sets. J. Algorithms 35(1), 17–49 (2000)
10. Greenberg, S., Pascoe, A., Randall, D.: Sampling biased lattice configurations using exponential metrics. In: SODA. pp. 76–85 (2009)
11. Hayes, T.P., Vera, J.C., Vigoda, E.: Randomly coloring planar graphs with fewer colors than the maximum degree. Random Struct. Algorithms (2014)
12. Jerrum, M., Sinclair, A., Vigoda, E.: A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. J. ACM 51(4), 671–697 (2004)
13. Jerrum, M., Valiant, L.G., Vazirani, V.V.: Random generation of combinatorial structures from a uniform distribution. Theor. Comput. Sci. 43, 169–188 (1986)
14. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov chains and mixing times. American Mathematical Soc. (2009)
15. Luby, M., Randall, D., Sinclair, A.: Markov chain algorithms for planar lattice structures. SIAM J. Comput. 31(1), 167–192 (2001)
16. Luby, M., Vigoda, E.: Approximately counting up to four. In: STOC. pp. 682–687 (1997)
17. Martin, R.A., Randall, D.: Sampling adsorbing staircase walks using a new Markov chain decomposition method. In: FOCS. pp. 492–502 (2000)
18. Mihalák, M., Šrámek, R., Widmayer, P.: Approximately counting approximately-shortest paths in directed acyclic graphs. Theory of Computing Systems pp. 1–15 (2014)
19. Morris, B., Sinclair, A.: Random walks on truncated cubes and sampling 0-1 knapsack solutions. SIAM J. Comput. 34(1), 195–226 (2004)
20. Provan, J.S.: The complexity of reliability computations in planar and acyclic graphs. SIAM J. Comput. 15(3), 694–702 (1986)
21. Randall, D., Sinclair, A.: Self-testing algorithms for self-avoiding walks. Journal of Mathematical Physics 41(3), 1570–1584 (2000)
22. Roberts, B., Kroese, D.P.: Estimating the number of s-t paths in a graph. J. Graph Algorithms Appl. 11(1), 195–214 (2007)
23. Sinclair, A., Jerrum, M.: Approximate counting, uniform generation and rapidly mixing Markov chains. Inf. Comput. 82(1), 93–133 (1989)
24. Valiant, L.G.: The complexity of enumeration and reliability problems. SIAM J. Comput. 8(3), 410–421 (1979)

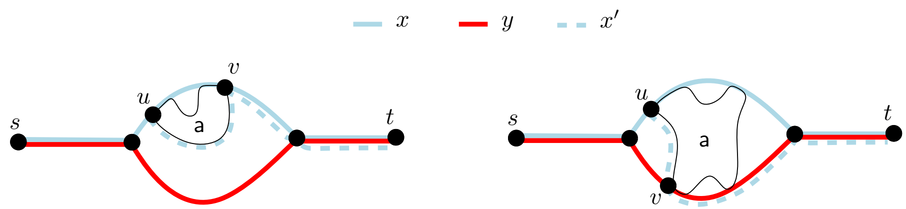

Fig. 8: The case distinction of the proof of Lemma 12.

A Postponed Proofs

A.1 Theorem 5

Lemma 12. *For any plane embedding of a graph G the following holds. If x and y are two st-paths such that $\Delta_{xy} = 2$, then they are connected in $\mathcal{M}_{paths}$ by a path of length at most f, where f is the number of faces of G.*

Proof. Consider the subgraph of G induced by the edges of x and y, and the drawing of this subgraph on the plane. Since $\Delta_{xy} = 2$, the plane is divided in exactly two regions, one of which is bounded by the edges of x and y and the other one is unbounded. The bounded region contains δ_{xy} faces of G, with $1 \leq \delta_{xy} \leq f$. We prove by induction on δ_{xy} that in $\mathcal{M}_{paths}$ there is a path between x and y of length at most δ_{xy}.

The base case for $\delta_{xy} = 1$ is trivial, because y can be obtained by rerouting x along exactly one face, the one contained in the bounded region. The path in $\mathcal{M}_{paths}$ is then the transition from x to y.

We now assume the claim to hold for every $\delta_{xy} \leq k-1$ and prove it for $\delta_{xy} = k$. Consider a face a contained in the bounded region having at least one edge which belongs to x but not to y. We distinguish among the following cases, depicted by Figure 8:

1. *Face* a *does not contain any vertex from* y. We define a path x' as follows. Let u and v be the first and last vertex of a that we encounter when moving from s to t along x, respectively. Let z be the sub-path of a connecting u to v and not containing any intermediate vertex from x. Then x' is the concatenation of the sub-path of x going from s to u, the path z, and the sub-path of x going from v to t. Note that by construction $\Delta_{xx'} = 2$ since x and x' are identical from s to u and from v to t, and they have no common vertices between u and v. Similarly, $\Delta_{yx'} = 2$ since z has no vertices from y.
   We next observe that $\delta_{xx'} + \delta_{x'y} = \delta_{xy}$, because the sub-path z partitions the faces in the bounded region determined by x and y. Note also that the partition is proper because the original bounded region contains $\delta_{xy} \geq 2$ faces, and therefore $\delta_{xx'} \leq k-1$ and $\delta_{x'y} \leq k-1$. We can thus apply the inductive hypothesis and obtain a path in $\mathcal{M}_{paths}$ from x to x' of length $\delta_{xx'}$, and a path in $\mathcal{M}_{paths}$ from x' to y of length $\delta_{x'y}$. The concatenation of the two paths yields a path in $\mathcal{M}_{paths}$ from x to y of length $\delta_{xx'} + \delta_{x'y} = \delta_{xy}$.

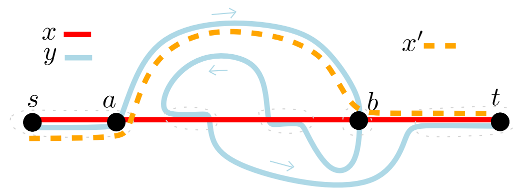

Fig. 9: The idea of the proof of Lemma 13.

2. *Face* a *contains some vertex from y.* Let u and v be the first vertex of a that we encounter by following respectively x and y along the non-common sub-paths. Let z be the sub-path of a going from u to v without intermediate vertices from x and y. Then x' is the concatenation of the sub-path of x going from s to u, the path z, and the sub-path of y going from v to t. Note that by construction $\Delta_{xx'} = 2$ since x and x' are identical from s to u and from t' to t, and they have no common vertices between u and t'. Similarly, $\Delta_{yx'} = 2$ since y and x' are identical from s to s' and from v to t, and they have no common vertices between s' and v. The same argument of the previous case implies $\delta_{xx'} + \delta_{x'y} = \delta_{xy}$ and $\delta_{xx'} \leq k-1$ and $\delta_{x'y} \leq k-1$. Thus, by inductive hypothesis, we obtain a path in $\mathcal{M}_{paths}$ from x to y of length $\delta_{xx'} + \delta_{x'y} = \delta_{xy}$.

The lemma thus follows from the fact that $\delta_{xy} \leq f$. □

One might be tempted to define the distance between two paths as the number of faces that are inside the bounded region formed by the two. However, things are slightly more complex since it is possible to have "nested" regions. The next lemma allows us to reduce to the simpler case of a single bounded region.

Lemma 13. *For any graph and any two st-paths x and y such that $\Delta_{xy} > 2$, there exists an st-path x' such that $\Delta_{xx'} = 2$ and $\Delta_{x'y} = \Delta_{xy} - 1$.*

Proof. We construct an st-path x' as in Figure 9. Starting from s, let a be the last vertex for which x and y are identical. From a, we follow y until we reach a vertex b belonging to both x and y. From b we follow again x until t. Observe that the resulting path x' is simple since in between a and b there are no vertices belonging to x. By construction, x and x' are identical from s to a and from b to t, while they differ between a and b. Therefore, it holds that $\Delta_{xx'} = 2$.

We now show that $\Delta_{x'y} \leq \Delta_{xy} - 1$. Consider the set of Δ_{xy} maximal sub-paths common to x and y. One of such sub-paths contains a, while another distinct one contains b. If we start at b and follow x towards t we encounter at most $\Delta_{xy} - 2$ such sub-paths. Therefore, the number of maximal sub-paths common to x' and y between b and t is at most $\Delta_{xy} - 2$. Moreover, observe that x' and y are identical between s and b and therefore here we have only one maximal sub-path common to x' and y. In total, there are at most $1 + \Delta_{xy} - 2$ maximal sub-paths common to x' and y, that is, $\Delta_{x'y} \leq \Delta_{xy} - 1$. □

Proof (of Theorem 5). By Lemma 12 and 13, there exists a path in $\mathcal{M}_{paths}$ from $x \in \Omega$ to $y \in \Omega$ of length at most $\Delta_{xy} \cdot f$. Since in any planar graph $f \leq 2n - 4$ and $\Delta_{xy} < n$, the distance between x and y is smaller than $2n^2$. □

A.2 Theorem 6

Proof. We show that the distribution π in (4) satisfies the detailed balance condition (3). Given two paths $x, y \in \Omega$ with $P(x, y) > 0$, assume without loss of generality that $A(y, x) = 1$ and observe that

$$P(x, y) = \frac{1}{2f} \cdot \frac{\lambda^{|y|}}{\lambda^{|x|}} \qquad \text{and} \qquad P(y, x) = \frac{1}{2f}.$$

Therefore

$$\pi(x) \cdot P(x, y) = \frac{\lambda^{|x|}}{Z(\lambda)} \cdot \frac{1}{2f} \cdot \frac{\lambda^{|y|}}{\lambda^{|x|}} = \pi(y) \cdot P(y, x),$$

□

A.3 Corollary 8

Proof. Consider the graph in Figure 5b consisting of a single row of $2k$ faces and s and t located in the middle as shown. Observe that

$$Z(\lambda) = \lambda + 2(\lambda^3 + \lambda^5 + \cdots + \lambda^{2k+1}).$$

We consider the bottleneck ratio of the set R of all st-paths on the right part of the graph and of length larger than 1. Thus

$$\pi(R) = \frac{\lambda^3 + \lambda^5 + \cdots + \lambda^{2k+1}}{Z(\lambda)} \leq 1/2.$$

Note that the only st-path in R which has a non-zero transition probability to some $y^* \in \bar{R}$ is the path x^* of length 3, and the only one such $y^* \in \bar{R}$ is the path (s, t) of length 1. Therefore, we have

$$Q(R, \bar{R}) = \pi(x^*)P(x^*, y^*) = \frac{\lambda^3}{Z(\lambda)} \frac{1}{2f} \frac{\lambda}{\lambda^3}.$$

Thus, the bottleneck theorem implies (we also use that $f \geq 2$ in our graphs)

$$t_{mix} \geq \frac{\pi(R)}{4Q(R, \bar{R})} = \frac{\lambda^3 + \lambda^5 + \cdots + \lambda^{2k+1}}{4\lambda/2f} > \lambda^{2k}.$$

Since the graph has $n = 2(2k + 1) = 4k + 2$ vertices, we get $t_{mix} \in \Omega(\lambda^{n/2})$. □

nw	n	ne
w	d	e
sw	s	se

Fig. 10: The names of the faces around a fixed face d.

A.4 Lemma 10

Let $x, y \in \Omega$ be any pair such that $\rho(x, y) = 1$. Note that the transitions defining our path coupling correspond to the following values for $\rho(X, Y)$:

$$\begin{aligned} \rho(X,Y) =&0 \qquad\qquad \text{for (10) and (11)} \\ \rho(X,Y) =&2 \qquad\qquad \text{for (13) and (14)} \\ \rho(X,Y) =&1 \qquad\qquad \text{for (12) and (15)} \end{aligned}$$

Therefore

$$\mathbb{E}_{x,y}[\rho(X,Y)] = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 = 1 + p_2 - p_0,$$

where

$$\begin{aligned} p_0 &:= \mathbb{P}_{x,y}[\rho(X,Y) = 0] = p_{\mathsf{d}}(x) + p_{\mathsf{d}}(y) \\ p_2 &:= \mathbb{P}_{x,y}[\rho(X,Y) = 2] = \sum_{\mathsf{a} \neq \mathsf{d}} |p_{\mathsf{a}}(x) - p_{\mathsf{a}}(y)| \\ p_1 &:= \mathbb{P}_{x,y}[\rho(X,Y) = 1] = 1 - p_2 - p_0 \end{aligned}$$

Our goal is then to prove that $p_2 \leq p_0$ for all $x, y \in \Omega$ with $\rho(x, y) = 1$. In order to bound p_2 it is enough to consider only the faces adjacent with the face d in which x and y differ. We name them according to Figure 10. The following claim says that we can ignore the “corner faces”. At this point it is crucial the monotonicity of the paths.

Lemma 14. *If* $\mathsf{a} \in \{\mathsf{nw}, \mathsf{ne}, \mathsf{sw}, \mathsf{se}\}$ *then* $|p_{\mathsf{a}}(x) - p_{\mathsf{a}}(y)| = 0$.

Proof. Consider the case $\mathsf{a} = \mathsf{nw}$ (the other cases are similar) and the intersection of this face with the subpath common to both x and y, depicted in Figure 11. Observe that, if vertex d is in both path x and path y, then $p_{\mathsf{nw}}(x) = p_{\mathsf{nw}}(y)$. Now consider the case where d belongs only to one of the two paths, say x, and the edges of nw common to both x and y.

If (a, c) is common (Figure 11a), the monotonicity of the paths implies that also (c, d) is common (Figure 11b), contradicting the hypothesis that d is not on y. Thus, we have that only (a, b) can be common (Figure 11c), but then the paths cannot be rerouted along nw and thus $p_{\mathsf{nw}}(x) = p_{\mathsf{nw}}(y) = 0$. □

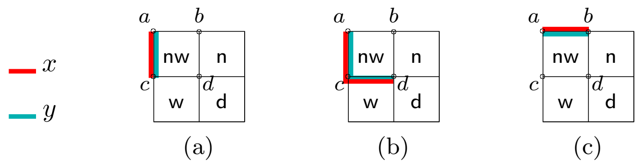

Fig. 11: The cases in the proof of Lemma 14.

Proof (of Lemma 10). By Lemma 14, our goal is to show that

$$p_2 = \sum_{\mathsf{a}\in\{\mathsf{n},\mathsf{s},\mathsf{e},\mathsf{w}\}} |p_{\mathsf{a}}(x) - p_{\mathsf{a}}(y)| \leq (1+\delta)p = 2\gamma p = p_0.$$

Dividing all terms by p, this is equivalent to show the inequality

$$\overline{p}_{\mathsf{n}} + \overline{p}_{\mathsf{s}} + \overline{p}_{\mathsf{e}} + \overline{p}_{\mathsf{w}} \leq 1 + \delta = 2\gamma,$$

where

$$\overline{p}_{\mathsf{a}} := \frac{|p_{\mathsf{a}}(x) - p_{\mathsf{a}}(y)|}{p} \qquad \text{for every face } \mathsf{a}.$$

We distinguish several cases which we group according to Figures 12 and 13. For Figure 12, it holds

$$\begin{array}{lllll} \text{Figure 12a:} & \overline{p}_{\mathsf{n}} = \gamma - \delta & \overline{p}_{\mathsf{s}} = 0 & \overline{p}_{\mathsf{e}} = \gamma & \overline{p}_{\mathsf{w}} = \delta \\ \text{Figure 12b:} & \overline{p}_{\mathsf{n}} = \gamma - \delta & \overline{p}_{\mathsf{s}} = \gamma - \delta & \overline{p}_{\mathsf{e}} = \gamma & \overline{p}_{\mathsf{w}} = \delta \\ \text{Figure 12c:} & \overline{p}_{\mathsf{n}} = 0 & \overline{p}_{\mathsf{s}} = \gamma - \delta & \overline{p}_{\mathsf{e}} = \delta & \overline{p}_{\mathsf{w}} = \gamma. \end{array}$$

The remaining cases are similar to the previous ones as they differ in only one face. For instance, the only difference between the case in Figure 12d and the one in Figure 12a is for face n. However, since $\gamma = \frac{1+\delta}{2}$, in the case of Figure 12d we have $\overline{p}_n = 1 - \gamma = \gamma - \delta$, that is, the same as in the case in Figure 12a.

For the cases of Figure 13, it holds

$$\begin{array}{lllll} \text{Figure 13a:} & \overline{p}_{\mathsf{n}} = 0 & \overline{p}_{\mathsf{s}} = 0 & \overline{p}_{\mathsf{e}} = 1 & \overline{p}_{\mathsf{w}} = \delta \\ \text{Figure 13b:} & \overline{p}_{\mathsf{n}} = \gamma - \delta & \overline{p}_{\mathsf{s}} = 0 & \overline{p}_{\mathsf{e}} = \gamma & \overline{p}_{\mathsf{w}} = \delta \\ \text{Figure 13c:} & \overline{p}_{\mathsf{n}} = 0 & \overline{p}_{\mathsf{s}} = \gamma - \delta & \overline{p}_{\mathsf{e}} = \gamma & \overline{p}_{\mathsf{w}} = \delta \\ \text{Figure 13d:} & \overline{p}_{\mathsf{n}} = \gamma - \delta & \overline{p}_{\mathsf{s}} = \gamma - \delta & \overline{p}_{\mathsf{e}} = \delta & \overline{p}_{\mathsf{w}} = \delta. \end{array}$$

The remaining cases follow from the identity $1 - \gamma = \gamma - \delta$.

To conclude the proof, we note that the case where d is not surrounded by four faces, that is, d is on the border of the grid, is even more favorable than those above, because some faces among $\mathsf{n}, \mathsf{s}, \mathsf{e}, \mathsf{w}$ will not be present. ☐

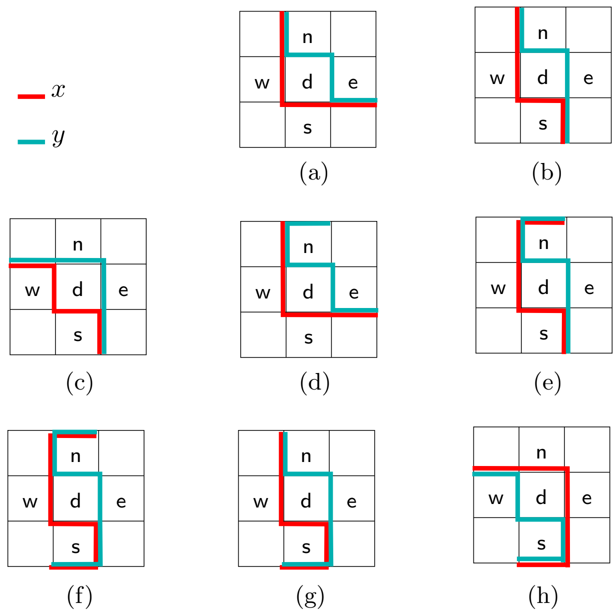

Fig. 12: The cases where x and y have the same length.

B Dynamic Programming and Exact Sampling for Vertical Monotone Paths

For the sake of completeness, in this section we describe a standard dynamic programming approach for counting paths. In general graphs, this method counts non-simple paths, but in our application the algorithm can be easily modified in order to avoid non-simple paths (see the end of this section for the details).

Suppose we have a polynomial-time procedure that returns a path selected uniformly from the subset of all paths of length ℓ, and suppose also we can compute the number m_ℓ of st-paths of length ℓ. We can use this procedure to sample paths according to the Gibbs distribution (1) as follows:

1. Select a length k with probability
$$s_k := \frac{m_k \lambda^k}{Z(\lambda)},$$
where $Z(\lambda) = \sum_{z \in \Omega} \lambda^{|z|} = \sum_\ell m_\ell \cdot \lambda^\ell$.
2. Select a path x uniformly at random in the set of all st-paths of length exactly k.

To see that we are indeed selecting paths according to (1), observe that the probability that a path x is selected is $s_k \cdot (1/m_k) = \lambda^k / Z(\lambda)$.

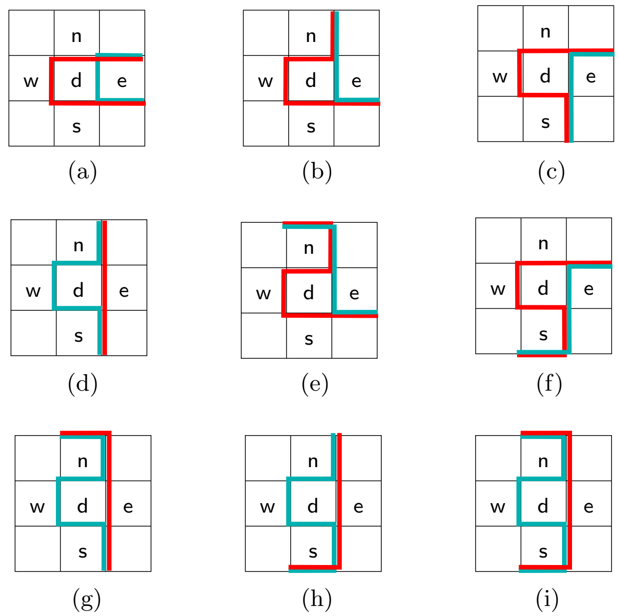

Fig. 13: The cases where the length of x and y is different.

We next describe how the standard dynamic programming for counting all paths (not necessarily simple) of length up to some L can be used to sample st-paths uniformly among those of a fixed length ℓ. For vertical-monotone paths, this algorithm can be easily modified to avoid non-simple paths. Since the graph is unweighted, we only consider lengths up to n. We compute a table $T(u, k)$ for all vertices u and for all lengths k such that $T(u, k)$ is the number of paths of length k from s to u. Initially T is identically to 0 except for $T(s, 0) = 1$.

Given the values for $k = 0, \ldots, \ell - 1$ we compute $T(u, \ell)$ as follows. For every u, consider all neighbors $v \in N_u$ and let

$$T(u, \ell) = \sum_{v \in N_u} T(v, \ell - 1). \tag{16}$$

In our application, the neighbors of u are the vertices on the grid that are placed below, on left, and on the right of u.

In order to conclude the proof, we need to show how this procedure can be used to pick a path x uniformly among all st-paths of a given length ℓ. Consider the entry $T(t, \ell)$ which tells us the number of such paths. According to (16) this value is the sum of the number of paths of length $\ell - 1$ to the neighbors of u: If v is a neighbor of u and there are $T(v, \ell - 1)$ paths from s to v, Starting from $u = t$, we pick a neighbor v with probability

$$\eta_v := T(v, \ell - 1)/T(u, \ell),$$

and repeat this step starting from the selected vertex until we reach s. The sequence of vertices that we choose is an st-path x of length ℓ. To see that x is chosen with probability $1/T(t, \ell)$ consider the sequence of vertices in $x = (x_0, x_1, \dots, x_\ell)$ where $x_0 = s$ and $x_\ell = t$. Observe that the probability of picking this sequence is the probability of picking $x_{\ell-1}, x_{\ell-2}, \dots, x_0$,

$$\begin{aligned}\eta_{x_{\ell-1}}\eta_{x_{\ell-2}}\cdots\eta_{x_0} =& \frac{T(x_{\ell-1}, \ell-1)}{T(x_\ell, \ell)} \cdot \frac{T(x_{\ell-2}, \ell-2)}{T(x_{\ell-1}, \ell-1)} \cdots \frac{T(x_0, 0)}{T(x_{\ell-1}, \ell-1)} \\ =& \frac{T(s, 0)}{T(t, \ell)} = \frac{1}{T(t, \ell)}.\end{aligned}$$

To conclude the analysis we provide a simple upper bound on the running time. First, the table has size $n \times n$ and it can be computed in time $O(n^2)$, since each node has only a constant number of neighbors. For the same reason, the procedure for selecting a random st-path of length ℓ takes $O(\ell)$ steps, and thus the overall running time can be bounded from above by $O(n^2)$.

Finally, we note how to ensure that the dynamic programming counts only simple paths in our vertical-monotone restriction. First, in the table we store three values corresponding to the paths that arrive to a node u from above, from left and from right (say T_{above}, T_{left} and T_{right}) similarly to the above computation. Once this table is computed, the selection of a random path x proceeds similarly to the method described above, except that we carry the information about the last step. For example, if $x_{\ell-1}$ is on the left of x_ℓ then we consider the number of st-paths to $x_{\ell-1}$ which arrive only from above or from the left of this node, $T'(x_{\ell-1}, \ell-1) := T_{above}(x_{\ell-1}, \ell-1) + T_{left}(x_{\ell-1}, \ell-1)$. Then, at the next iteration, when we select a neighbor $x_{\ell-2}$ of $x_{\ell-1}$ we consider the analogous quantity and the probability of selecting a particular $x_{\ell-2}$ is the ratio $T'(x_{\ell-2,\ell-2}))/T'(x_{\ell-1}, \ell-1)$.