Deterministic Random Walks for Rapidly Mixing Chains

Takeharu Shiraga* Yukiko Yamauchi* Shuji Kijima* Masafumi Yamashita*

November 18, 2013

Abstract

The rotor-router model, also known as the Propp machine, is a deterministic process analogous to a random walk on a graph. Instead of distributing tokens to randomly chosen neighbors, the rotor-router deterministically serves the neighbors in a fixed order. This paper is concerned with a generalized model, functional-router model. While the rotor-router is an analogy with a random walk consisting of only rational transition probabilities using parallel edges, the functional-router imitates a random walk containing irrational transition probabilities. For the functional-router model, we investigate the discrepancy on a single vertex between the number of tokens in the functional-router model and the expected number of tokens in a random walk. We present an upper bound of the discrepancy in terms of the mixing rate of the corresponding transition matrix. Using the result, we give a polynomial time deterministic sampler for particular #P-complete problems, such as 0-1 knapsack solutions, linear extensions, matchings, etc., for which rapidly mixing chains are known; Our deterministic algorithm provides samples from a “distribution” with a point-wise distance at most $\varepsilon$ from the target distribution, in time polynomial in the input size and $\varepsilon^{-1}$.

Key words: rotor-router model, Markov chain Monte Carlo, mixing time, #P-complete.

1 Introduction

The rotor-router model, also known as the Propp machine, is a deterministic process analogous to a random walk on a graph [41, 8, 30]. Instead of distributing tokens to randomly chosen neighbors, the rotor-router model deterministically serves the neighbors in a fixed order by associating to each vertex a “rotor-router” pointing to one of its neighbors. Doerr et al. [6, 10] first called the rotor-router model deterministic random walk, meaning a “derandomized, hence deterministic, version of a random walk.”

Single vertex discrepancy for multiple-walk. Cooper and Spencer [8] are concerned with the model of multiple tokens (multiple-walk) on $\mathbb{Z}^n$, and investigated the discrepancy on a single vertex: they gave a bound that $|\chi_v^{(t)} - \mu_v^{(t)}| \leq c_n$ where $\chi_v^{(t)}$ (resp. $\mu_v^{(t)}$) denotes the number (resp. expected number) of tokens on vertex $v \in \mathbb{Z}^n$ in a rotor-router model (resp. corresponding random walk) at time $t$ on the condition that $\mu_v^{(0)} = \chi_v^{(0)}$ for any $v$, and $c_n$ is a constant depending only on $n$ but independent of the total number of tokens in the system. Cooper et al. [6] showed $c_1 \simeq 2.29$, and Doerr and Friedrich [10] showed that $c_2$ is about 7.29 or 7.83 depending on the routing rules. On the other hand, Cooper et al. [5] gave an example of $|\chi_v^{(t)} - \mu_v^{(t)}| = \Omega(\sqrt{kt})$ on infinite $k$-regular trees, the example implies that the discrepancy can get infinitely large as increasing the total number of tokens.

Motivated by a derandomization of Markov chains, Kijima et al. [30] are concerned with multiple-walks on general finite multidigraphs $(V, A)$, and gave a bound $|\chi_v^{(t)} - \mu_v^{(t)}| = O(|V||A|)$ in case that

*Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
shiraga@tcslab.csce.kyushu-u.ac.jp, {yamauchi, kijima, mak}@inf.kyushu-u.ac.jp
corresponding Markov chain is ergodic, reversible and lazy. They also gave some examples of $|\chi_v(t) - \mu_v(t)| = \Omega(|\mathcal{A}|)$. Kajino et al. [29] sophisticated the approach by [30], and gave a bound in terms of the second largest eigenvalue and eigenvectors of the corresponding Markov chain, for an arbitrary irreducible finite Markov chain, which may not be lazy, reversible nor aperiodic.

For some specific finite graphs, namely hypercubes and tori, some bounds in terms of logarithm of the size of transition diagram are known. For $n$-dimensional hypercube, Kijima et al. [30] gave a bound $O(n^3)$, and Kajino et al. [29] improved the bound to $O(n^2)$. Recently, Akbari and Berenbrink [1] gave a bound $O(n^{1.5})$, using results by Friedrich et al. [17]. Akbari and Berenbrink [1] also gave a bound $O(1)$ for constant dimensional tori. Those analyses highly depends on the structures of the specific graphs, and it is difficult to extend the technique to other combinatorial graphs. Kijima et al. [30] gave rise to a question for constant dimensional tori. Those analyses highly depend on the structures of the specific graphs, and it is difficult to extend the technique to other combinatorial graphs. Kijima et al. [30] gave rise to a question.

Our Results. This paper is concerned with a parallel walk version of the functional-router model, a generalization of the rotor-router model. While the rotor-router is an analogy with a random walk with rational transition probabilities, the functional-router imitates a random walk containing irrational transition probabilities by routing-functions defined on vertices. In the functional-router model, a configuration of $M$ tokens over a finite set $V = \{1, \ldots, N\}$ is deterministically updated; let $\chi(t) = (\chi_1(t), \ldots, \chi_N(t)) \in \mathbb{Z}_N$ denote the configuration at time $t = 0, 1, 2, \ldots$, i.e., $\sum_{v \in V} \chi_v(t) = M$. For comparison, let $\mu(t) = \chi(0)$, and let $\mu(t) = \mu(0)P^t$, then $\mu(t) \in \mathbb{R}^N$ denotes the expected configuration of $M$ tokens independently according to $P$ for $t$ steps. A main contribution of the paper is to show that $|\chi_v(t) - \mu_v(t)| \leq 3(\pi_{\text{max}}/\pi_{\text{min}})t^* \Delta$ holds for any $v \in V$ at any time $t$ in case that the corresponding transition matrix $P$ is ergodic and reversible, where
\( \pi_{\text{max}} \) and \( \pi_{\text{min}} \) are maximum/minimum values of \( \pi \) respectively, \( t^* \) is the mixing rate of the corresponding Markov chain, and \( \Delta \) is the maximum degree of the transition diagram.

This result suggests a polynomial-time deterministic algorithm for uniform sampling of some \#P-complete problems, such as knapsack solutions, linear extensions, matchings, q-colorings etc., for which rapidly mixing chains exist. Thus, our result affirmatively answers the question by Kijima et al. [30]. Setting the number of tokens \( M \geq 3\varepsilon^{-1}t^*\Delta \) for an arbitrary \( \varepsilon (0 < \varepsilon < 1) \), our algorithm provides \( M \) samples with a “distribution” \( \tilde{\chi}^{(t)} := \chi^{(t)}/M \), of which the point-wise distance \( \|\tilde{\chi}^{(t)} - \pi\|_\infty \) is at most \( \varepsilon \) from the uniform distribution \( \pi \) over the target set. For instance, our algorithm runs in \( O^*(n^{1.1} \varepsilon^{-1}) \) time for \( n \)-dimensional 0-1 knapsack solutions, in \( O^*(n^3 \varepsilon^{-1}) \) time for linear extensions of \( n \) elements poset, in \( O^*(m^4 n^4 \varepsilon^{-1}) \) time for all matchings in a graph with \( n \) vertices and \( m \) edges, in \( O^*(c_q n^4 \varepsilon^{-1}) \) time for all \( q \)-colorings in a graph with \( n \) vertices, where \( c_q \) is a constant depending on \( q \) and the max. degree of the input graph. O* notation ignores \( \text{poly}(\log(\varepsilon^{-1}), \log m, \log n) \) factors. Note that those orders of magnitude are not optimized, for simplicity of the main arguments. Unfortunately, these running times are the best possible in terms of \( \varepsilon^{-1} \) for any deterministic sampler, because of the integrality gap of the number of tokens.

An example of a random walk containing irrational transition probabilities is the \( \beta \)-random walk devised by Ikeda et al. [24], which achieves an \( O(N^2) \) hitting time and an \( O(N^2 \log N) \) cover time for \textit{any graphs}. Another example should be the Markov chain Monte Carlo (MCMC), such as Gibbs samplers for the Ising model (cf. [44, 37, 39]), reversible Markov chains for queeining networks (cf. [31]), etc.

Recently, derandomization of randomized algorithms is an interesting and challenging topic. Gopalan, Klivans, and Meka [19], and Stefankovic, Vempala, and Vigoda [45] gave deterministic approximation algorithms for counting knapsack solutions (cf. [20]), where randomized approximation algorithms based on MCMC [38] or based on rejection sampling via dynamic programming [16] had been known for the problem. Cooper, Ilcinkas, Klasing, and Kosowski [7] were concerned with exploration of an anonymous graph using derandomization of random walks. This paper is motivated by a development of a general scheme for a derandomization of randomized algorithms based on Markov chains, such as MCMC, and the functional-router model proposed by the paper is hopefully a step for this goal.

**Organization** This paper is organized as follows. In Section 2, we briefly reviews MCMC, as a preliminary of analysis. In Section 3, we describe functional-router model, and explain our main result. In Section 4, we prove the main theorem. In Section 5, we show some constructions of routing function, and give upper bounds for the models. In Section 6, we present a deterministic sampling algorithm, and show examples of polynomial-time uniform samplers of \#P-complete problems, namely for knapsack solutions, linear extensions, matchings, and \( q \)-colorings.

## 2 Preliminaries: Markov Chain Monte Carlo

As a preliminary step of our deterministic sampling, this section briefly reviews the Markov chain Monte Carlo (MCMC). See e.g., [44, 37, 39] for detail of MCMC.

Let \( V \) be a finite set, and suppose that we wish to sample from \( V \) with a probability proportional to a given positive vector \( f = (f_1, \ldots, f_N) \in \mathbb{R}_{\geq 0}^N \); for example, we are concerned with uniform sampling of 0-1 knapsack solutions in Section 6.3 where \( V \) denotes the set of 0-1 knapsack solutions and \( f_v \) is 1 for each \( v \in V \). The idea of a Markov chain Monte Carlo (MCMC) is to sample from a limit distribution of a Markov chain which is equal to the target distribution \( f/\|f\|_1 \) where \( \|f\|_1 = \sum_{v \in V} f_v \) is the normalizing constant.

Let \( P \in \mathbb{R}_{\geq 0}^{N \times N} \) be a transition matrix of a Markov chain with the state space \( V \), where \( P_{u,v} \) denotes the transition probability from \( u \) to \( v \) \((u, v \in V)\). A transition matrix \( P \) is irreducible if \( P_{u,v} > 0 \) for any \( u \) and \( v \) in \( V \), and is aperiodic if \( \text{GCD}\{t \in \mathbb{Z}_{>0} \mid P^t_{x,x} > 0\} = 1 \) holds for any \( x \in V \), where \( P^t_{u,v} \) denotes the
entry of $P^t$, the $t$-th power of $P$. An irreducible and aperiodic transition matrix is called \textit{ergodic}. It is well-known for an ergodic $P$, there is a unique \textit{stationary distribution} $\pi \in \mathbb{R}_+^N$, i.e., $\pi P = \pi$, and the limit distribution is $\pi$, i.e., $\xi P^\infty = \pi$ for any probability distribution $\xi \in \mathbb{R}_+^N$ on $V$.

An ergodic Markov chain defined by a transition matrix $P \in \mathbb{R}_{\geq 0}^{N \times N}$ is \textit{reversible} if the \textit{detailed balance equation}

$$f_u P_{u,v} = f_v P_{v,u}$$

holds for any $u, v \in V$. When $P$ satisfies the detailed balance equation, it is not difficult to see that $fP = f$ holds, meaning that $f/\|f\|_1$ is the limit distribution (see e.g., [37]). Let $\xi$ and $\zeta$ be a distribution on $V$, then the \textit{total variation distance} $D_{tv}$ between $\xi$ and $\zeta$ is defined by

$$D_{tv}(\xi, \zeta) \overset{\text{def.}}{=} \max_{A \subseteq V} \sum_{v \in A} (\xi_v - \zeta_v) = \frac{1}{2} \|\xi - \zeta\|_1.$$  

(2)

Note that $D_{tv}(\xi, \zeta) \leq 1$, since $\|\xi\|_1$ and $\|\zeta\|_1$ are equal to one, respectively. The \textit{mixing time} of a Markov chain is defined by

$$\tau(\varepsilon) \overset{\text{def.}}{=} \max_{v \in V} \min \{t \in \mathbb{Z}_{\geq 0} \mid D_{tv}(P_{v,\cdot}^t, \pi) \leq \varepsilon\}$$

for any $\varepsilon > 0$, where $P_{v,\cdot}^t$ denotes the $v$-th row vector of $P^t$; i.e., $P_{v,\cdot}^t$ denotes the distribution of a Markov chain at time $t$ starting from the initial state $v \in V$. In other words, the distribution $P_{v,\cdot}^t$ of the Markov chain after $\tau(\varepsilon)$ transition satisfies $D_{tv}(P_{v,\cdot}^t, \pi) \leq \varepsilon$, meaning that we obtain an approximate sample from the target distribution.

For convenience, let $h(t) \overset{\text{def.}}{=} \max_{v \in V} D_{tv}(P_{v,\cdot}^t, \pi)$ for $t \geq 0$, then it is well-known that $h$ satisfies a kind of \textit{submultiplicativity}. We will use the following proposition in the analysis of our algorithm in Section 4. See [37] or Appendix A for the proof.

\begin{proposition}
For any integers $\ell (\ell \geq 1)$ and $k (0 \leq k < \tau(\gamma))$,

$$h(\ell \cdot \tau(\gamma) + k) \leq \frac{1}{2} (2\gamma)^\ell$$

holds for any $\gamma (0 < \gamma < 1/2)$.
\end{proposition}

Thus, $t^* \overset{\text{def.}}{=} \tau(1/4)$, called \textit{mixing rate}, is often used as a characterization of $P$.

\section{Model and Main result}

A \textit{functional-router model} is a deterministic process analogous to a multiple random walk. Roughly speaking, tokens on $u$ moves to a neighboring vertex $v$ with probability $P_{u,v}$ in a multiple random walk, whereas a router defined on each vertex $u$ deterministically serves tokens to $v$ at a rate of $P_{u,v}$ in a functional-router model.

To get the idea, let us start with explaining the rotor-router model (see e.g., [8] [30]), which corresponds to a simple random walk on a graph, in Section 3.1.
3.1 Rotor-router model

Let $G = (V,E)$ be a simple undirected graph where $V = \{1, \ldots, N\}$. Let $\mathcal{N}(v)$ denote the neighborhood of $v \in V$. For convenience, let $\delta(v) = |\mathcal{N}(v)|$. Let $\chi^{(0)} \in \mathbb{Z}^N_{\geq 0}$ be an initial configuration of tokens, and let $\chi^{(t)} \in \mathbb{Z}^N_{\geq 0}$ denote the configuration of tokens at time $t \in \mathbb{Z}_{\geq 0}$ in the rotor-router model. A configuration $\chi^{(t)}$ is updated by rotor-routers on vertices, as follows. Without loss of generality, we may assume that an ordering $u_0, \ldots, u_{\delta(v)-1}$ is defined on $\mathcal{N}(v)$ for each $v \in V$. Then, a rotor-router $\sigma_v: \mathbb{Z}_{\geq 0} \to \mathcal{N}(v)$ on $v \in V$ is defined by

$$\sigma_v(j) \overset{\text{def.}}{=} u_i \mod \delta(v)$$

for $j \in \mathbb{Z}_{\geq 0}$. Let

$$Z_{v,u}^{(t)} \overset{\text{def.}}{=} \left| \left\{ j \in \{0,\ldots, \chi_v^{(t)} - 1 \} \mid \sigma_v \left( j + \sum_{s=0}^{t-1} \chi_v^{(s)} \right) = u \right\} \right|$$

for $v,u \in V$, where $Z_{v,u}^{(t)}$ denotes the number of tokens served from $v$ to $u$ in the update. Then, $\chi^{(t+1)}$ is defined by

$$\chi_u^{(t+1)} \overset{\text{def.}}{=} \sum_{v \in V} Z_{v,u}^{(t)}$$

for each $u \in V$.

It is not difficult to see that

$$\frac{\left| \left\{ j \in \{0,\ldots, z - 1 \} \mid \sigma_v(j) = u \right\} \right|}{z} \overset{z \to \infty}{\to} \frac{1}{\delta(v)}$$

holds, which implies that the “outflow ratio” $\sum_{s=0}^{\infty} Z_{v,u}^{(s)}/\sum_{s=0}^{\infty} \chi_v^{(s)}$ of tokens at $v$ to $u$ approaches asymptotically to $1/\delta(v)$ as $t$ increasing. Thus, the rotor-router hopefully approximates a distribution of tokens by a random walk.

3.2 Functional-router model

Let $P \in \mathbb{R}^{N \times N}_{\geq 0}$ be a transition matrix of a Markov chain with a state space $V \overset{\text{def.}}{=} \{1, \ldots, N\}$, where $P_{u,v}$ denotes the transition probability from $u$ to $v$ ($u,v \in V$). Note that $P_{u,v}$ may be irrational. In this paper, we assume that $P$ is ergodic and reversible (see Section 2). Let $\mu^{(0)} = (\mu_1^{(0)}, \ldots, \mu_N^{(0)}) \in \mathbb{Z}^N_{\geq 0}$ denote an initial configuration of $M$ tokens over $V$, and let $\mu^{(t)} \in \mathbb{R}^N_{\geq 0}$ denote the expected configuration of tokens independently according to $P$ at time $t \in \mathbb{Z}_{\geq 0}$, i.e., $\|\mu^{(t)}\|_1 = M$ and $\mu^{(t)} = \mu^{(0)} P^t$.

Let $G = (V,E)$ be the transition digram of $P$, meaning that $E = \{(u,v) \in V^2 \mid P_{u,v} > 0\}$. Note that $E$ may contain self-loop edges, and also note that $|E| \leq N^2$ holds. Let $\mathcal{N}(v)$ denote the (out-)neighborhood of $v \in V$, i.e., $\mathcal{N}(v) = \{ u \in V \mid P_{v,u} > 0 \}$, and let $\delta(v) = |\mathcal{N}(v)|$. Note that $v \in \mathcal{N}(v)$ if $P_{v,v} > 0$.

Let $\chi^{(0)} = \mu^{(0)}$, and let $\chi^{(t)} \in \mathbb{Z}^N_{\geq 0}$ denote the configuration of tokens at time $t \in \mathbb{Z}_{\geq 0}$ in the functional-router model. A configuration $\chi^{(t)}$ is updated by functional-routers $\sigma_v: \mathbb{Z}_{\geq 0} \to \mathcal{N}(v)$ defined on each $v \in V$ to imitate $P_{v,v}$. To be precise, let

$$I_{v,u}[z,z'] \overset{\text{def.}}{=} \left| \left\{ j \in \{z,\ldots, z' - 1 \} \mid \sigma_v(j) = u \right\} \right|$$

1 In Section 2.2, we are concerned with the model on multidigraphs.
2 e.g., $P_{u,v} = \sqrt{\pi}/10$, $\exp(-10)$, $\sin(\pi/6)$, etc. are allowed.
3 Since $P$ is reversible, $u \in \mathcal{N}(v)$ if and only if $v \in \mathcal{N}(u)$, and then we abuse $\mathcal{N}(v)$ for in-neighborhood of $v \in V$. 

5
Figure 1: In this example, \( V = \{1, 2\} \), \( \chi(0) = (7, 0), \chi(1) = (4, 3), \chi(2) = (5, 2) \). \( \mathcal{I}_{1,2}[4, 10] = |\{j \in \{4, \ldots, 9\} | \sigma_1(j) = 2\}| = 2 \).

for \( v, u \in V \) and for any \( z, z' \in \mathbb{Z}_{\geq 0} \) satisfying \( z < z' \), for convenience. Then, \( \sigma_v \) is designed such as to minimize

\[
\frac{\mathcal{I}_{v,u}[0, z]}{z} - P_{v,u}
\]

for \( z \in \mathbb{Z}_{\geq 0} \). Some specific routing-functions are given in Sections 5.1, 5.2, and 5.3. Let

\[
Z_{v,u}^{(t)} = \mathcal{I}_{v,u} \left[ \sum_{s=0}^{t-1} \chi_{v}^{(s)}, \sum_{s=0}^{t} \chi_{v}^{(s)} \right]
\]

for \( v, u \in V \), where \( Z_{v,u}^{(t)} \) denotes the number of tokens served from \( v \) to \( u \) in the update. Then, \( \chi^{(t+1)} \) is defined by

\[
\chi_u^{(t+1)} \overset{\text{def.}}{=} \sum_{v \in V} Z_{v,u}^{(t)}
\]

for each \( u \in V \).

We in Section 5 show examples of functional-routers, in which the “outflow ratio” \( \sum_{s=0}^{t} Z_{v,u}^{(s)} / \sum_{s=0}^{t} \chi_{v}^{(s)} \) from \( v \) to \( u \) approaches asymptotically to \( P_{v,u} \) as \( t \) increases, meaning that the functional-router hopefully approximate a distribution of tokens by a random walk.

### 3.3 Main results

Our goal is to estimate the discrepancy \( |\chi_w^{(T)} - \mu_w^{(T)}| \) for \( w \in V \) and \( T \geq 0 \) for the functional router model described in Section 3.2. Let

\[
\Psi_{\sigma} = \max_{v \in V, u \in \mathcal{N}(v), t \geq 0} \left| Z_{v,u}^{(t)} - \chi_{v}^{(t)} P_{v,u} \right|
\]

and then, the following is our main theorem.

**Theorem 3.1.** Let \( P \in \mathbb{R}^{N \times N}_{\geq 0} \) be a transition matrix of a reversible and ergodic Markov chain with a state space \( V \), and let \( \pi \) be the stationary distribution of \( P \). Then, the configurations \( \chi^{(T)} \) and \( \mu^{(T)} \) of tokens in a functional-routing model and in its corresponding random walk satisfy

\[
|\chi_w^{(T)} - \mu_w^{(T)}| \leq \Psi_{\sigma} \frac{2(1 - \gamma)}{1 - 2\gamma} \tau(\gamma) \frac{\pi_w}{\pi_{\min}} \Delta
\]

for any \( w \in V, T \geq 0 \) and \( \gamma (0 < \gamma < 1/2) \).
We remark that
\[ \Psi_\sigma \leq \max_{v \in V, u \in \N(v), z, z' \in \Z_{\geq 0} \text{ s.t. } z' > z} \left| I_{v,u}[z, z'] - (z' - z) P_{v,u} \right| \]  \tag{9}
holds, since
\[ Z^{(t)}_{v,u} - \chi^{(t)}_v P_{v,u} = I_{v,u} \left[ \sum_{s=0}^{t-1} \chi^{(s)}_v, \sum_{s=0}^{t} \chi^{(s)}_v \right] - \left( \sum_{s=0}^{t} \chi^{(s)}_v - \sum_{s=0}^{t-1} \chi^{(s)}_v \right) P_{v,u} \]
holds by the definition. We give upper bounds of \( \Psi_\sigma \) for some specific routing-functions. See sections 5.1, 5.2, and 5.3 considering (9).

4 Analysis of the Point-wise Distance

This section proves Theorem 5.1. Our proof technique in this section is similar to previous results [8, 30]. To begin with, we establish the following key lemma.

Lemma 4.1. Let \( P \in \R^{N \times N}_{\geq 0} \) be a transition matrix of a reversible and ergodic Markov chain with a state space \( V \), and let \( \pi \) be the stationary distribution of \( P \). Then, the configurations \( \chi^{(T)} \) and \( \mu^{(T)} \) of tokens in the algorithm and in corresponding random walk satisfy
\[ \chi^{(T)}_w - \mu^{(T)}_w = \sum_{t=0}^{T-1} \sum_{u \in V} \sum_{v \in \N(u)} \left( Z^{(t)}_{v,u} - \chi^{(t)}_v P_{v,u} \right) \left( P^{T-t-1}_{u,w} \right) - \pi_w \]
holds for any \( w \in V \) and for any \( T \geq 0 \).

Proof. Remark that
\[ \chi^{(T)}_w - \mu^{(T)}_w = \left( \chi^{(T)} - \mu^{(0)} P^T \right)_w = \left( \chi^{(T)} P^0 - \chi^{(0)} P^T \right)_w \]  \tag{10}
holds where the last equality follows the assumption \( \chi^{(0)} = \mu^{(0)} \). It is not difficult to see that
\[ \chi^{(T)} P^0 - \chi^{(0)} P^T = \left( \chi^{(T)} P^0 - \chi^{(T-1)} P^1 \right) + \left( \chi^{(T-1)} P^1 - \chi^{(T-2)} P^2 \right) + \cdots \]
\[ + \left( \chi^{(2)} P^{T-2} - \chi^{(1)} P^{T-1} \right) + \left( \chi^{(1)} P^{T-1} - \chi^{(0)} P^T \right) \]
\[ = \sum_{t=0}^{T-1} \left( \chi^{(t+1)} P^{T-t-1} - \chi^{(t)} P^{T-t} \right) \]
holds, thus we have
\[ 10 = \sum_{t=0}^{T-1} \left( \left( \chi^{(t+1)} P^{T-t-1} \right)_w - \left( \chi^{(t)} P^{T-t} \right)_w \right) \]
\[ = \sum_{t=0}^{T-1} \left( \sum_{u \in V} \chi^{(t+1)}_u P^{T-t-1}_{u,w} - \sum_{u \in V} \left( \chi^{(t)}_u P_u \right) P^{T-t-1}_{u,w} \right) \]
\[ = \sum_{t=0}^{T-1} \sum_{u \in V} \left( \chi^{(t+1)}_u - \left( \chi^{(t)}_u P_u \right) \right) P^{T-t-1}_{u,w}. \]  \tag{11}
While \( \sum_{u \in V} \left( \chi_u^{(t+1)} - (\chi^{(t)} P)_u \right) P_{u,w}^{T-t-1} \) in (11) may not be 0 in general, remark that
\[
\sum_{u \in V} \left( \chi_u^{(t+1)} - (\chi^{(t)} P)_u \right) = \sum_{u \in V} \chi_u^{(t+1)} - \sum_{u \in V} \sum_{v \in V} \chi_v^{(t)} P_{v,u} \\
= \sum_{u \in V} \chi_u^{(t+1)} - \sum_{v \in V} \chi_v^{(t)} \sum_{u \in V} P_{v,u} \\
= M - M = 0
\]
holds. for any \( t \geq 0 \). Hence
\[
(11) = \sum_{t=0}^{T-1} \sum_{u \in V} \left( \chi_u^{(t+1)} - (\chi^{(t)} P)_u \right) P_{u,w}^{T-t-1} - \sum_{t=0}^{T-1} \sum_{u \in V} \left( \chi_u^{(t+1)} - (\chi^{(t)} P)_u \right) \pi_w \\
= \sum_{t=0}^{T-1} \sum_{u \in V} \left( \chi_u^{(t+1)} - (\chi^{(t)} P)_u \right) (P_{u,w}^{T-t-1} - \pi_w) \tag{12}
\]
holds. Since \( P \) is reversible, \( Z_{v,u}^{(t)} = 0 \) for any \( v \notin \mathcal{N}(u) \) and \( \chi_u^{(t+1)} = \sum_{v \in V} Z_{v,u}^{(t)} = \sum_{v \in \mathcal{N}(u)} Z_{v,u}^{(t)} \) holds by definition (7). Thus,
\[
(12) = \sum_{t=0}^{T-1} \sum_{u \in V} \left( \sum_{v \in \mathcal{N}(u)} Z_{v,u}^{(t)} - \sum_{v \in \mathcal{N}(u)} \chi_v^{(t)} P_{v,u} \right) (P_{u,w}^{T-t-1} - \pi_w) \\
= \sum_{t=0}^{T-1} \sum_{u \in V} \sum_{v \in V} \left( Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u} \right) (P_{u,w}^{T-t-1} - \pi_w)
\]
holds, and we obtain the claim. \( \square \)

Now, we are concerned with reversible Markov chains, and show the theorem 3.1.

**Proof of Theorem 3.1.** By Lemma 4.1 and (8), we obtain that
\[
\left| \chi_u^{(T)} - \mu_u^{(T)} \right| \leq \sum_{t=0}^{T-1} \sum_{u \in V} \sum_{v \in \mathcal{N}(u)} \left| Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u} \right| P_{u,w}^{T-t-1} - \pi_w \\
\leq \Psi_\sigma \sum_{t=0}^{T-1} \sum_{u \in V} \sum_{v \in \mathcal{N}(u)} \left| P_{u,w}^{T-t-1} - \pi_w \right| \\
= \Psi_\sigma \sum_{t=0}^{T-1} \sum_{u \in V} \delta(u) \left| P_{u,w}^t - \pi_w \right| \tag{13}
\]
holds. Since \( P \) is reversible, \( P_{u,w}^t = \frac{\pi_w}{\pi_u} P_{w,u}^t \) holds for any \( w \) and \( u \) in \( V \) (see Proposition A.5 in Appendix A). Thus
\[
(13) = \Psi_\sigma \sum_{t=0}^{T-1} \sum_{u \in V} \delta(u) \left| P_{u,w}^t - \pi_u \right| \\
\leq \Psi_\sigma \Delta \frac{\pi_w}{\pi_{\min}} \sum_{t=0}^{T-1} \sum_{u \in V} \left| P_{u,w}^t - \pi_u \right| \\
= 2\Psi_\sigma \Delta \frac{\pi_w}{\pi_{\min}} \sum_{t=0}^{T-1} D_{tv} (P_{w,u}^t, \pi) \tag{14}
\]
where the last equality follows the fact that \( \sum_{u \in V} |P_{w,u}^{t} - \pi_u| = 2 \mathcal{D}_{tv} (P_{w,u}^{t}, \pi) \), by the definition of the total variation distance \([2]\). By Proposition \([2.1]\) we obtain the following.

**Lemma 4.2.** For any \( v \in V \) and for any \( T > 0 \),

\[
\sum_{t=0}^{T-1} \mathcal{D}_{tv} (P_{w,v}^{t}, \pi) \leq \frac{1 - \gamma}{1 - 2\gamma} \tau(\gamma)
\]

holds for any \( \gamma \) \((0 < \gamma < 1/2)\).

**Proof.** Let \( h(t) = \max_{w \in V} \mathcal{D}_{tv} (P_{w,v}^{t}, \pi) \), for convenience. Then, \( h(t) \) is at most 1 for any \( t \geq 0 \), by the definition of the total variation distance \([2]\). By Proposition \([2.1]\)

\[
\sum_{t=0}^{T-1} \mathcal{D}_{tv} (P_{w,v}^{t}, \pi) = \sum_{t=0}^{T-1} h(t) = \sum_{\ell=0}^{\tau(\gamma)-1} \sum_{k=0}^{\tau(\gamma)-1} h(\ell \cdot \tau(\gamma) + k)
\]

\[
= \sum_{k=0}^{\tau(\gamma)-1} h(k) + \sum_{\ell=1}^{\tau(\gamma)-1} \sum_{k=0}^{\tau(\gamma)-1} h(\ell \cdot \tau(\gamma) + k) \leq \sum_{k=0}^{\tau(\gamma)-1} 1 + \sum_{\ell=1}^{\tau(\gamma)-1} \sum_{k=0}^{\tau(\gamma)-1} \frac{1}{2} (2\gamma)^{\ell}
\]

\[
= \tau(\gamma) + \sum_{\ell=1}^{\tau(\gamma)} \frac{1}{2} (2\gamma)^{\ell} = \tau(\gamma) + \frac{\gamma}{1 - 2\gamma} \tau(\gamma) = \frac{1 - \gamma}{1 - 2\gamma} \tau(\gamma)
\]

holds, and we obtain the claim. \(\square\)

Now we obtain Theorem \([3.1]\) from \([14]\) and Lemma \([4.2]\).

### 5 Various Model

In section \([4]\) we gave a bound of \(|\chi^{(t)}(v) - \mu^{(t)}(v)|\) using \(\Psi_\sigma\), in general. This section shows some routing functions, namely greedy routing in Section \([5.1]\) rotor-router on multigraph in Section \([5.2]\) and the functional-router based on the van der Corput sequence in Section \([5.3]\) and gives upper bounds on \(|\chi^{(T)}(v) - \mu^{(T)}(v)|\) for them. Greedy routing (Section \([5.1]\)), originally given by Holroyd and Propp by the name of stack-walk, and the functional-router based on the van der Corput sequence (Section \([5.3]\)) are to be used for irrational transition probabilities, while the rotor-router on multidigraph (Section \([5.2]\)) was discussed in \([30, 29]\) to deal with rational transition probabilities.

#### 5.1 Greedy routing

This section introduces greedy routing, which is given by Angel et al. \([2]\) and Tijdeman \([46]\). The functional-router \(\sigma_v(i) \ (i \in \mathbb{Z}_{\geq 0})\) on \(v \in V\) of the greedy-routing is defined, as follows. For \(v \in V\) and \(i \in \mathbb{Z}_{\geq 0}\), let

\[
L_i(v, u) = (i + 1)P_{v,u} - I_{v,u}[0, i]
\]

for each \(u \in \mathcal{N}(v)\), and let

\[
T_i(v) = \{u \in \mathcal{N}(v) \mid L_i(v, u) > 0\}
\]

Then, let \(\sigma_v(i)\) be \(u^* \in T_i(v)\) minimizing the value

\[
\frac{1 - L_i(v, u)}{P_{v,u}}
\]
in all \( u \in T_i(v) \). If there are two or more such \( u \in T_i(v) \), then let \( u^* \) be the minimum in them in a (prescribed) order.

Since \( \sigma_v(i) \in T_i(v) \), we can see that \( L_i(v, u) = (i + 1)P_{v,u} - \mathcal{I}_{v,u}[0, i + 1] > -1 \) holds for any \( u, v \) and \( i \), by an induction on \( i \in \mathbb{Z}_{\geq 0} \). The following theorem is due to Angel et al. \cite{2} and Tijdeman \cite{46}.

**Theorem 5.1.** \cite{46,2} For any transition matrix \( P \),

\[
\mathcal{I}_{v,u}[0, z] - z \cdot P_{v,u} \leq 1
\]

holds for any \( v, u \in V \) and any \( z \in \mathbb{Z}_{>0} \).

Theorem 5.1 is firstly given by Tijdeman \cite{46}, where he gave a slightly better bound \( \mathcal{I}_{v,u}[0, z] - z \cdot P_{v,u} \leq 1 - (2(\delta(v) - 1))^{-1} \), in fact. Angel et al. \cite{2} rediscovered Theorem 5.1 in the context of deterministic random walk (see also \cite{21}), where they also showed a similar statement holds even when the corresponding probability is time-inhomogeneous.

Theorem 5.1 and (9) imply that

\[
\Psi_\sigma \leq \max_{v \in V, u \in \mathcal{N}(v), z, z' \in \mathbb{Z}_{\geq 0} s.t. z' > z} \mathcal{I}_{v,u}[z, z'] - (z' - z)P_{v,u} \leq 2
\]

holds for the greedy-routing model. Thus, we immediately obtain the following theorem by Theorem 3.1 and (15).

**Theorem 5.2.** Let \( P \in \mathbb{R}^{N \times N}_{\geq 0} \) be a transition matrix of a reversible and ergodic Markov chain with a state space \( V \), and let \( \pi \) be the stationary distribution of \( P \). Then, the configurations \( \chi^{(T)} \) and \( \mu^{(T)} \) of tokens respectively in a greedy-routing model and in its corresponding random walk satisfy

\[
\left| \chi^{(T)}_{w} - \mu^{(T)}_{w} \right| \leq \frac{6\pi \cdot t^* \cdot \Delta}{\pi_{\min}}
\]

for any \( w \in V \) and \( T \geq 0 \).

### 5.2 Rotor-router on multidigraph

The rotor-router model described in Section 3.2 can be generally considered on digraphs with parallel edges (i.e., multidigraphs). Kijima et al. \cite{30} and Kajino et al. \cite{29} are concerned with the rotor-router model on finite multidigraphs. Suppose that \( P \) is a transition matrix with rational entries. For each \( v \in V \), let \( \delta(v) \in \mathbb{Z}_{\geq 0} \) be a common denominator (or the least common denominator) of \( P_{v,u} \) for all \( u \in \mathcal{N}(v) \), meaning that \( \delta(v) \cdot P_{v,u} \) is integer for each \( u \in \mathcal{N}(v) \). We define a rotor-router \( \sigma_v(0), \sigma_v(1), \ldots, \sigma_v(\delta(v) - 1) \) arbitrarily satisfying that

\[
\left| \{ j \in [0, \ldots, \delta(v)) \mid \sigma_v(j) = u \} \right| = \delta(v) \cdot P_{v,u}
\]

for any \( v \in V \) and \( u \in \mathcal{N}(v) \). Then, \( \sigma_v(i) \) is defined by

\[
\sigma_v(i) = \sigma_v(i \ mod \ \delta(v)) \left( \equiv \sigma_v \left( i - \delta(v) \cdot \left\lfloor \frac{i}{\delta(v)} \right\rfloor \right) \right). \tag{16}
\]

For the rotor router on a multidigraph, it is not difficult to observe the following.

**Observation 5.3.** \( \Psi_\sigma = \max_v \delta(v) \) holds for the rotor-router model on a multidigraph.

By Theorem 5.1 and the above observation, we obtain the following theorem.
**Theorem 5.4.** Let $P \in \mathbb{Q}_{\geq 0}^{N \times N}$ be a transition matrix of a reversible and ergodic Markov chain with a state space $V$, and let $\pi$ be the stationary distribution of $P$. Then, the configurations $\chi^{(T)}$ and $\mu^{(T)}$ of tokens respectively in a rotor-router model and in its corresponding random walk satisfy

$$|\chi^{(T)}_w - \mu^{(T)}_w| \leq \frac{3\pi_m}{\pi_{\min}} \max_v \delta(v) \cdot t^* \Delta$$

for any $w \in V$ and $T \geq 0$.

### 5.3 Routing based on the van der Corput sequence

This section gives a router $\sigma$ based on the *van der Corput sequence* [48, 40], which is a well-known low-discrepancy sequence.

The van der Corput sequence $\psi: \mathbb{Z}_{\geq 0} \to [0, 1)$ is defined as follows. Suppose $i \in \mathbb{Z}_{\geq 0}$ is represented in binary as $i = \sum_{j=0}^{\lfloor \log_2 i \rfloor} \beta_j(i) \cdot 2^j$ using $\beta_j(i) \in \{0, 1\}$ ($j \in \{0, 1, \ldots, \lfloor \log_2 i \rfloor\}$). Then, we define

$$\psi(i) \overset{\text{def.}}{=} \sum_{j=0}^{\lfloor \log_2 i \rfloor} \beta_j(i) \cdot 2^{-(j+1)}. \quad (17)$$

and $\psi(0) \overset{\text{def.}}{=} 0$. For example, $\psi(1) = 1 \times 1/2 = 1/2$, $\psi(2) = 0 \times 1/2 + 1 \times 1/4 = 1/4$, $\psi(3) = 1 \times 1/2 + 1 \times 1/4 = 3/4$, $\psi(4) = 0 \times 1/2 + 0 \times 1/4 + 1 \times 1/8 = 1/8$, $\psi(5) = 1 \times 1/2 + 0 \times 1/4 + 1 \times 1/8 = 5/8$, $\psi(6) = 0 \times 1/2 + 1 \times 1/4 + 1 \times 1/8 = 3/8$, and so on. Clearly, $\psi(i) \in [0, 1)$ holds for any (finite) $i \in \mathbb{Z}_{\geq 0}$.

Now, given $i \in \mathbb{Z}_{\geq 0}$, we define $\sigma_v(i)$ as follows. Without loss of generality, we may assume that an ordering $u_1, \ldots, u_{\delta(v)}$ is defined on $\mathcal{N}(v)$ for $v \in V$. Then, we define the routing function $\sigma_v: \mathbb{Z}_{\geq 0} \to \mathcal{N}(v)$ on $v \in V$ such that $\sigma_v(i) = u_k \in \mathcal{N}(v)$ satisfies that

$$\sum_{j=1}^{k-1} P_{v,u_j} \leq \psi(i) < \sum_{j=0}^{k} P_{v,u_j}$$

for $k \in \{1, \ldots, \delta(v)\}$, where $\sum_{j=1}^{0} P_{v,u_j} = 0$, for convenience.

For the van der Corput sequence, the following theorem, due to van der Corput [48], is known.

**Theorem 5.5.** [48] For any transition matrix $P$,

$$|\mathcal{I}_{v,u}[0, z) - z \cdot P_{v,u}| \leq \log(z + 1)$$

holds for any $v, u \in V$ and any $z \in \mathbb{Z}_{\geq 0}$.

More sophisticated bounds are found in [40]. Carefully examining Theorem 5.5, we obtain the following lemma. See Appendix [B] for the proof.

**Lemma 5.6.** For any transition matrix $P$,

$$|\mathcal{I}_{v,u}[z, z') - (z' - z)P_{v,u}| \leq 2 \log(z' - z + 1)$$

holds for any $v, u \in V$, and for any $z, z' \in \mathbb{Z}_{\geq 0}$ satisfying $z' > z$.

Lemma 5.6 suggests the following lemma.

**Lemma 5.7.** $\Psi_\sigma \leq 2 \log(M + 1)$ holds for the van der Corput sequence.

By Theorem 5.5 and Lemma 5.7, we obtain the following.
Theorem 5.8. Let $P \in \mathbb{R}^{N \times N}_{\geq 0}$ be a transition matrix of a reversible and ergodic Markov chain with a state space $V$, and let $\pi$ be the stationary distribution of $P$. Then, the configurations $\chi^{(T)}$ and $\mu^{(T)}$ of tokens respectively in a functional-routing model based on the van der Corput sequence and in its corresponding random walk satisfy
\[
|\chi_{w}^{(T)} - \mu_{w}^{(T)}| \leq \frac{66\pi_w}{\pi_{\text{min}}} \cdot \log(M + 1) \cdot t^* \Delta
\]
for any $w \in V$ and $T \geq 0$.

Though the bound depends on $\log M$, $|\chi^{(t)}_v / M - \mu^{(t)}_v / M| = O(\log(M)/M)$ holds in terms of $M$, meaning that the discrepancy approaches asymptotically to zero as increasing the number of tokens $M$.

6 Deterministic Sampling

This section presents deterministic sampling algorithm based on a (version of) functional routing model, as an application. We explain our algorithm in Section 6.1 and show our main theorem in Section 6.2. Note that our following algorithm is oblivious; while the rotor-router model or greedy-routing model memorizes the configurations of tokens and routers, our algorithm memorizes the configuration of tokens only. It makes the description of the algorithm simple, compared with other deterministic random walks.

We show detailed description of deterministic sampling algorithms for particular applications, such as 0-1 knapsack solutions (Section 6.3), linear extensions (Section 6.4), matchings (Section 6.5), and $q$-coloring (Section 6.6), where we also discuss the computational complexities of our algorithm for the applications. Note that a similar (or essentially the same) bounds also hold on the greedy-routing model.

6.1 Sampling algorithm

The algorithm is essentially the greedy-routing model described in Section 5.1 but routing-functions are reset at (the beginning of) each time. Let $\chi^{(0)} = \mu^{(0)}$, and let $\chi^{(t)} \in \mathbb{Z}^N_{\geq 0}$ denote the configuration of tokens at time $t \in \mathbb{Z}_{\geq 0}$ in our algorithm. A configuration $\chi^{(t)}$ is updated, imitating $P_{v,u}$, as follows. Without loss of generality, we may assume that an arbitrary ordering $u_1, \ldots, u_{\delta(v)}$ is defined on $\mathcal{N}(v)$ for each $v \in V$.

Then, we define the number of tokens $Z_{v,u_i}^{(t)}$ sent from $v$ to $u_i$ during the time interval from $t$ to $t + 1$ by
\[
Z_{v,u_i}^{(t)} = \begin{cases} 
\chi_v^{(t)} P_{v,u_i} + 1 & (i \leq i^*) \\
\chi_v^{(t)} P_{v,u_i} & (\text{otherwise})
\end{cases}
\] (18)

where
\[
i^* = \chi_v^{(t)} - \sum_{i=1}^{\delta(v)} \left\lfloor \chi_v^{(t)} P_{v,u_i} \right\rfloor
\]
denoting the number of “surplus” tokens. Then, $\chi^{(t+1)}$ is defined by
\[
\chi_u^{(t+1)} \overset{\text{def}}{=} \sum_{v \in V} Z_{v,u}^{(t)}
\] (19)
for each $u \in V$. It is not difficult to see the following observation.

Observation 6.1. For the above algorithm,
\[
|Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u}| \leq 1
\]
holds for any $u, v, \in V$ and $t \geq 0$. 

12
6.2 Upper bound of the point-wise distance

Let $\tilde{\mu}^{(t)} = \mu^{(t)}/M$, for simplicity, then clearly $\tilde{\mu}^{(\infty)} = \pi$ holds, since $P$ is ergodic (see Section 3.1). By the definition of the mixing time, $D_{tv}(\tilde{\mu}^{(\tau(\varepsilon))}, \pi) \leq \varepsilon$ holds where $\tau(\varepsilon)$ denotes the mixing time of $P$, meaning that $\tilde{\mu}$ approximates the target distribution $\pi$ well. Thus, we hope our deterministic sampler that the “distribution” $\tilde{\chi}^{(T)} = \chi^{(T)}/M$ approximates the target distribution $\pi$ well. We define a point-wise distance $D_{pw}(\xi, \zeta)$ between $\xi \in \mathbb{R}^N_{\geq 0}$ and $\zeta \in \mathbb{R}^N_{\geq 0}$ satisfying $\|\xi\|_1 = \|\zeta\|_1 = 1$ by

$$D_{pw}(\xi, \zeta) \overset{\text{def.}}{=} \max_{v \in V} |\xi_v - \zeta_v| = \|\xi - \zeta\|_{\infty}. \quad (20)$$

**Theorem 6.2.** Let $P \in \mathbb{R}^{N \times N}_{\geq 0}$ be a reversible transition matrix with a stationary distribution $\pi$, then

$$D_{pw}\left(\tilde{\chi}^{(T)}, \tilde{\mu}^{(T)}\right) \leq \frac{\pi_{\max}}{\pi_{\min}} \cdot \frac{3t^* \Delta}{M}$$

holds for any $T \geq 0$, where $\pi_{\max} = \max\{\pi_v \mid v \in V\}$ and $\pi_{\min} = \min\{\pi_v \mid v \in V\}$.

**Proof.** We can apply Theorem 3.1 for algorithm described in section 6.1 since (19) holds. Note that $\Psi_{\sigma} = 1$ holds by observation 6.1, we have

$$\left|\chi_w^{(T)} - \mu_w^{(T)}\right| \leq \frac{2(1 - \gamma)}{1 - 2\gamma} \tau(\gamma) \frac{\pi_w}{\pi_{\min}} \Delta$$

holds. Then, taking $\gamma = 1/4$ and divide by $M$ both sides, we obtain the claim. \hfill \Box

In a special case that the stationary distribution is uniform, we obtain the following by Theorem 6.2.

**Corollary 6.3.** Let $P \in \mathbb{R}^{N \times N}_{\geq 0}$ be an ergodic and reversible transition matrix with a uniform stationary distribution $\pi$. Set $M \geq 6\varepsilon^{-1}t^* \Delta$, then the “distribution” $\tilde{\chi}^{(T)}$ of the deterministic sampler after $T \geq \tau(\varepsilon/2)$ steps satisfies that $D_{pw}\left(\tilde{\chi}^{(T)}, \pi\right) \leq \varepsilon$.

In the following section, we show some examples of polynomial-time deterministic samplers for uniform sampling of combinatorial objects, whose counting is known to be #P-complete.

6.3 01 knapsack solutions

Given $a \in \mathbb{R}^n_{\geq 0}$ and $b \in \mathbb{R}_{\geq 0}$, the set of the 0-1 knapsack solutions is defined by $\Omega_{\text{Kna}} = \{x \in \{0, 1\}^n \mid \sum_{i=1}^n a_i x_i \leq b\}$. We define a transition matrix $P_{\text{Kna}} \in \mathbb{R}^{[\Omega_{\text{Kna}} \times [\Omega_{\text{Kna}}]}$ by

$$P_{\text{Kna}}(x, y) = \begin{cases} 1/2n & \text{(if } y \in N_{\text{Kna}}(x)) \\ 1 - |N_{\text{Kna}}(x)|/2n & \text{(if } y = x) \\ 0 & \text{(otherwise)} \end{cases} \quad (21)$$

for $x, y \in \Omega_{\text{Kna}}$, where $N_{\text{Kna}}(x) = \{y \in \Omega_{\text{Kna}} \mid \|x - y\|_1 = 1\}$. Note that stationary distribution of $P_{\text{Kna}}$ is uniform distribution since $P_{\text{Kna}}$ is symmetric. The following theorem is due to Morris and Sinclair [38].

**Theorem 6.4.** [38] The mixing time $\tau(\gamma)$ of $P_{\text{Kna}}$ is $O(n^{2+\alpha} \log \gamma^{-1})$ for any $\alpha > 0$ and for any $\gamma > 0$.

For the Markov chain defined by $P_{\text{Kna}}$, our deterministic sampler is described as follows. Note that the following implementation is not optimized the time and space complexity, for simplicity of the arguments.
Algorithm 1.  
Step 0. Set $W_0^i := 0$ for each $i = 1, \ldots, M$. $W^T[i]$ stores a solution in $\Omega_{Kna}$, where token $i$ is.
Step 1. For $(t = 0$ to $T - 1$){
    (a). Set list $S_x^{(t)} := \{i \in \{1, \ldots, M\} \mid W^T[i] = x\}$ for each $x \in \Omega_{Kna}$ as long as $S_x^{(t)} \neq \emptyset$.
    (b). Serve tokens in $S_x^{(t)}$ to neighboring vertices according to (13) for each $x \in \Omega_{Kna}$ satisfying that $S_x^{(t)} \neq \emptyset$, and set $W^{t+1}[i]$ be the solution in $\Omega_{Kna}$ at which token $i$ arrived.
}
Step 2. Output $W^T[i]$ for each $i = 1, \ldots, M$.

**Theorem 6.5.** For an arbitrary $\varepsilon$ $(0 < \varepsilon < 1)$, set $M := c_1 n^{\frac{1}{2} + \alpha - \varepsilon}$ and $T := c_2 n^{\frac{2}{3} + \alpha} \log \varepsilon^{-1}$ with appropriate constants $c_1, c_2$ and $\alpha$, then Algorithm 1 outputs $M$ samples over $\Omega_{Kna}$ satisfying that

\[
D_{\text{pw}} \left( \chi^{(T)}, \pi \right) \leq \varepsilon
\]

where $\pi$ is the uniform distribution over $\Omega_{Kna}$. The running time of Algorithm 1 is

\[
O(TM \log (M) n \log a, \log b) = O^*(n^{1+2\alpha} \varepsilon^{-1})
\]

where $O^*$ ignores poly log term.

**Proof.** We check Algorithm 1 for each Step. Step 0 sets all $M$ tokens on $0 \in \Omega_{Kna}$, which takes $O(Mn)$ time. Step 1(a) constructs the configuration $\chi^{(t)}$ of $M$ tokens over $\Omega_{Kna}$. Note that the number of lists is at most $M$, since Step 1(a) constructs a list only for $v \in \Omega_{Kna}$ where at least one token exists. Step 1(a) takes $O(M \log (M) n)$ time, by heapifying $W^T[i]$ $(i = 1, \ldots, M)$ with the lexicographical order on $\Omega_{Kna}$. Step 1(b) updates configuration according to our deterministic sampling algorithm in Section 6.1. It takes $O(n \log (a, \log b))$ time to find all feasible solutions neighboring to $x$. Once the algorithm finds all feasible solutions neighboring to $x$, then it is easy to let every token of $\chi^{(t)}$ to go to the neighboring vertex according to (18), in $O(n^{\alpha})$ time, like the router-router. Since we repeat Step 1 $T$ times, then we obtain the time complexity $O(TM \log (M) n \log (a, \log b))$.

Now, (22) is clear from Corollary 6.3 since Algorithm 1 is an implementation of the deterministic sampler in Section 6.1.

**6.4 Linear extensions of a poset**

Let $S = \{1, 2, \ldots, n\}$, and $Q = (S, \leq)$ be a partial order. A linear extension of $Q$ is a total order $X = (S, \preceq)$ which respects $Q$, i.e., for all $i, j \in S$, $i \leq j$ implies $i \preceq j$. Let $\Omega_{\text{Lin}}$ denote the set of all linear extensions of $Q$. We define a relationship $X \sim_p X' (p \in \{1, \ldots, n\})$ for a pair of linear extensions $X$ and $X' \in \Omega_{\text{Lin}}$ satisfying that $x_p = x'_p, x_p+1 = x'_p$, and $x_i = x'_i$ for all $i \neq p, p + 1$, i.e.,

\[
X = (x_1, x_2, \ldots, x_{p-1}, x_p, x_{p+1}, x_{p+2}, \ldots, x_n)
\]

\[
X' = (x_1, x_2, \ldots, x_{p-1}, x_p, x_{p+1}, x_{p+2}, \ldots, x_n)
\]

holds. Then, we define a transition matrix $P_{\text{Lin}} \in \mathbb{R}^{\Omega_{\text{Lin}} \times \Omega_{\text{Lin}}}$ by

\[
P_{\text{Lin}}(X, X') = \begin{cases} 
F(p)/2 & \text{(if } X' \sim_p X) \\
1 - \sum_{I \in \mathcal{N}_{\text{Lin}}(X)} P_{\text{Lin}}(X, I) & \text{(if } X' = X) \\
0 & \text{(otherwise)}
\end{cases}
\]

for $X, X' \in \Omega_{\text{Lin}}$, where $\mathcal{N}_{\text{Lin}}(X) = \{Y \in \Omega_{\text{Lin}} \mid X \sim_p Y (p \in \{1, \ldots, n-1\})\}$ and $F(p) = \frac{p(n-p)}{2(n^3 - n)}$. Note that $P_{\text{Lin}}$ is ergodic and reversible, and its stationary distribution is uniform on $\Omega_{\text{Lin}}$ [4]. The following theorem is due to Bubley and Dyer [4].
Theorem 6.6. For \( P_{\text{Lin}} \),
\[
\tau(\gamma) \leq \left\lceil \frac{1}{6} (n^3 - n) \ln \frac{n^2}{4\gamma} \right\rceil
\]
holds for any \( \gamma > 0 \).

For the Markov chain defined by \( P_{\text{Lin}} \), our deterministic sampler is described as follows.

Algorithm 2.

Step 0. Set \( W^0[i] := X' \) for each \( i = 1, \ldots, M \). % \( X' \in \Omega_{\text{Lin}} \) is a linear extension.

Step 1. For \( (t = 0 \text{ to } T - 1) \{ \)

(a). Set list \( S^{(t)}_X := \{ i \in \{1, \ldots, M\} \mid W^t[i] = X \} \) for each \( X \in \Omega_{\text{Lin}} \) as long as \( S^{(t)}_X \neq \emptyset \).

(b). Serve tokens in \( S^{(t)}_X \) to neighboring vertices according to (18) for each \( X \in \Omega_{\text{Lin}} \) satisfying that \( S^{(t)}_X \neq \emptyset \), and set \( W^{t+1}[i] \) be the solution in \( \Omega_{\text{Lin}} \) at which token \( i \) arrived.

\}

Step 2. Output \( W^T[i] \) for each \( i = 1, \ldots, M \).

Theorem 6.7. For an arbitrary \( \varepsilon (0 < \varepsilon < 1) \), set \( M := 6n \left\lceil \frac{1}{3} (n^3 - n) \ln n \right\rceil \varepsilon^{-1} \) and \( T := \left\lceil \frac{1}{6} (n^3 - n) \ln \frac{n^2}{2\varepsilon} \right\rceil \)
then Algorithm 2 outputs \( M \) sample over \( \Omega_{\text{Lin}} \) satisfying that
\[
\mathcal{D}_{pw} \left( \chi^{(T)}, \pi \right) \leq \varepsilon
\]  \hspace{1cm} (24)
where \( \pi \) is the uniform distribution over \( \Omega_{\text{Lin}} \). The running time of Algorithm 2 is
\[
O(TnM \log(M)) = O^*(n^8 \varepsilon^{-1})
\]
where \( O^* \) ignores poly log term.

Proof. We check Algorithm 2 for each Step. Step 0 sets all \( M \) tokens on \( X' \in \Omega_{\text{Lin}} \), which takes \( O(Mn + n^2) \) time, since it takes \( O(n^2) \) time to pick a linear extension \( X' \) by topological sorting. Step 1(a) takes \( O(nM \log(M)) \) time, by heapifying \( W^t[i] (i = 1, \ldots, M) \) with the lexicographic order on \( \Omega_{\text{Lin}} \). Step 1(b) takes \( O(n \log n) \) time to find all feasible solutions neighboring to \( X \). Since we repeat Step 1 \( T \) times, then we obtain the time complexity \( O(TnM \log(M)) \).

Now, (24) is clear from Corollary 6.3 since Algorithm 2 is an implementation of the deterministic sampler in Section 6.1. \( \square \)

6.5 Matchings in a graph

Counting all matchings in a graph, related to the Hosoya index \( [22] \), is known to be \#P-complete \( [47] \). Jerrum and Sinclair \( [27] \) gave a rapidly mixing chain. This Section is concerned with sampling of all matchings in a graph. Remark that counting all perfect matchings in a bipartite graph, related to the permanent, is also well-known \#P-complete problem, and Jerrum, Sinclair, and Vigoda \( [28] \) gave a celebrated FPRAS based on an MCMC method using annealing. To apply our algorithm to sampling perfect matchings, we need some assumptions on the input graph (see e.g., \( [44, 27, 28] \)).

Let \( H = (U, F) \) be an undirected graph, where \( |U| = n \) and \( |F| = m \). A matching in \( H \) is a subset \( M \subseteq F \) such that no edges in \( M \) share an endpoint. Let \( N_C(M) = \{ e = \{ u, v \} \mid e \notin M \} \)

\footnote{In fact, to obtain this order of magnitude, we need a (simple) data structure, which is constructed in \( O(n^3) \) (or \( O(n^2 \log n) \) is possible) time at Step 0. However, here we omit the detail, for simplicity of the argument.}
\( \mathcal{M} \), both \( u \) and \( v \) are matched in \( \mathcal{M} \) and let \( \mathcal{N}_{\text{Mat}}(\mathcal{M}) = \{ e \mid e \notin \mathcal{N}_{\mathcal{C}}(\mathcal{M}) \} \). Then, for \( e = \{ u, v \} \in \mathcal{N}_{\text{Mat}}(\mathcal{M}) \), we define \( \mathcal{M}(e) \) by

\[
\mathcal{M}(e) = \begin{cases} 
\mathcal{M} - e & \text{(if } e \in \mathcal{M} \text{)} \\
\mathcal{M} + e & \text{(if } u \text{ and } v \text{ are unmatched in } \mathcal{M} \text{)} \\
\mathcal{M} + e - e' & \text{(if exactly one of } u \text{ and } v \text{ is matched in } \mathcal{M}, \text{ and } e' \text{ is the matching edge)}.
\end{cases}
\]

Let \( \Omega_{\text{Mat}} \) denote the set of all possible matchings of \( H \). The we define the transition matrix \( P_{\text{Mat}} \in \mathbb{R}^{|\Omega_{\text{Mat}}| \times |\Omega_{\text{Mat}}|} \) by

\[
P_{\text{Mat}}(\mathcal{M}, \mathcal{M}') = \begin{cases} 
1/2m & (\text{if } \mathcal{M}' = \mathcal{M}(e)) \\
1 - |\mathcal{N}_{\text{Mat}}(\mathcal{M})|/2m & (\text{if } \mathcal{M}' = \mathcal{M}) \\
0 & (\text{otherwise})
\end{cases}
\]

for any \( \mathcal{M}, \mathcal{M}' \in \Omega_{\text{Mat}} \). Note that \( P_{\text{Mat}} \) is ergodic and reversible, and its stationary distribution is uniform on \( \Omega_{\text{Mat}} \) \([27]\). The following theorem is due to Jerrum and Sinclair \([27]\).

**Theorem 6.8.** \([27]\) For \( P_{\text{Mat}} \),

\[
\tau(\gamma) \leq 4mn(n \ln n + \ln \gamma^{-1})
\]

holds for any \( \gamma > 0 \).

For the Markov chain defined by \( P_{\text{Mat}} \), our deterministic sampler is described as follows.

**Algorithm 3.**

Step 0. Set \( W^0[i] := \emptyset \) for each \( i = 1, \ldots, M \).

Step 1. For \( t = 0 \) to \( T - 1 \)\{

(a). Set \( S_M^{(t)} := \{ i \in \{ 1, \ldots, M \} \mid W^t[i] = \mathcal{M} \} \) for each \( \mathcal{M} \in \Omega_{\text{Mat}} \) as long as \( S_M^{(t)} \neq \emptyset \).

(b). Serve tokens in \( S_M^{(t)} \) to neighboring vertices according to \((18)\) for each \( \mathcal{M} \in \Omega_{\text{Mat}} \) satisfying that \( S_M^{(t)} \neq \emptyset \), and set \( W^{t + 1}[i] \) be the solution in \( \Omega_{\text{Mat}} \) at which token \( i \) arrived.

\}

Step 2. Output \( W^T[i] \) for each \( i = 1, \ldots, M \).

**Theorem 6.9.** For an arbitrary \( \varepsilon \ (0 < \varepsilon < 1) \), set \( M := 24(m + 1)mn(n \ln n + \ln 4)\varepsilon^{-1} \) and \( T := 4mn(n \ln n + \ln(\varepsilon/2)^{-1}) \), then Algorithm 3 outputs a sample \( \mathcal{M} \) over \( \Omega_{\text{Mat}} \) satisfying that

\[
\mathcal{D}_{pw} \left( \hat{\chi}(T), \pi \right) \leq \varepsilon
\]

where \( \pi \) is the uniform distribution over \( \Omega_{\text{Mat}} \). The running time of Algorithm 3 is

\[
O(TmM \log(M)) = O^*(m^4n^4\varepsilon^{-1})
\]

where \( O^* \) ignores poly log term.

**Proof.** We check Algorithm 3 for each Step. Step 0 sets all \( M \) tokens on \( \emptyset \in \Omega_{\text{Mat}} \), which takes \( O(Mm) \) time. Step 1(a) takes \( O(mM \log(M)) \) time, by heapifying \( W^t[i] \ (i = 1, \ldots, M) \) with the lexicographic order on \( \Omega_{\text{Mat}} \). Step 1(b) takes \( O(m^2) \) time to find all feasible solutions neighboring to \( \mathcal{M} \). Since we repeat Step 1 \( T \) times, then we obtain the time complexity \( O(TmM \log(M)) \).

Now, \((25)\) is clear from Corollary 6.3 since Algorithm 3 is an implementation of the deterministic sampler in Section 6.1.

\[ \square \]
6.6 \( q \)-coloring of a graph

Let \( H = (U, F) \) be an undirected graph, where \( |U| = n \) and \( |F| = m \), and let \( Q = \{1, \ldots, q\} \) be a set of colors. The set of proper \( q \)-coloring of \( H \) is defined by \( \Omega_{\text{Col}} = \{c \in Q^n \mid c_v \neq c_w \ (\forall (v, w) \in F)\} \).

We define the relationship \( c \sim_v c' (v \in V) \) for a pair of coloring \( c, c' \in \Omega_{\text{Col}} \) satisfying that \( c_v \neq c'_v \) and \( c_u = c'_u \) for all \( u \neq v \). Let \( Q(c_v) = Q \setminus c(N(v)) \), where \( c(A) = \{c_u \mid u \in A\} \) for a set \( A \subseteq U \), and we define the transition matrix \( P_{\text{Col}} \in \mathbb{R}^{\Omega_{\text{Col}} \times \Omega_{\text{Col}}} \) by

\[
P_{\text{Col}}(c, c') = \begin{cases} 1/(n \cdot |Q(c_v)|) & \text{(if } c' \sim_v c) \\ 1 - \sum_{i \in N_{\text{Col}}(c)} P_{\text{Col}}(c, i) & \text{(if } c' = c) \\ 0 & \text{(otherwise)} \end{cases}
\]

for any \( c, c' \in \Omega_{\text{Col}} \), where \( N_{\text{Col}}(c) = \{c' \in \Omega_{\text{Col}} \mid c' \sim_v c (v \in U)\} \). This chain is known as the Glauber dynamics for proper \( q \)-colorings. Note that \( P_{\text{Col}} \) is ergodic and reversible, and its stationary distribution is uniform on \( \Omega_{\text{Col}} \) [25]. The following theorem is due to Jerrum [25].

**Theorem 6.10.** [25] For \( P_{\text{Col}} \),

\[
\tau(\gamma) \leq \left\lceil \left( \frac{q - \Delta_H}{q - 2\Delta_H} \right) n \ln \frac{n}{\gamma} \right\rceil
\]

holds for any \( \gamma > 0 \) in case that \( q > 2\Delta_H \), where \( \Delta_H \) denotes the maximum degree of \( H \).

For the Markov chain defined by \( P_{\text{Col}} \), our deterministic sampler is described as follows.

**Algorithm 4.**

Step 0. Set \( W^0[i] := c' \) for each \( i = 1, \ldots, M \). % \( c' \) is some coloring of \( \Omega_{\text{Col}} \).

Step 1. For \( (t = 0) \to (T - 1) \) {

(a) Set list \( S_C^{(t)} := \{i \in \{1, \ldots, M\} \mid W^t[i] = c\} \) for each \( c \in \Omega_{\text{Col}} \) as long as \( S_C^{(t)} \neq \emptyset \).

(b) Serve tokens in \( S_C^{(t)} \) to neighboring vertices according to (18) for each \( c \in \Omega_{\text{Col}} \) satisfying that \( S_C^{(t)} \neq \emptyset \), and set \( W^{t+1}[i] \) to be the solution in \( \Omega_{\text{Col}} \) at which token \( i \) arrived.

}

Step 2. **Output** \( W^T[i] \) for each \( i = 1, \ldots, M \).

Note that maximum degree of transition diagram of \( P_{\text{Col}} \) is at most \( nq \), we obtain the following theorem.

**Theorem 6.11.** For an arbitrary \( \varepsilon \) (\( 0 < \varepsilon < 1 \)), set \( M := 6nq\varepsilon^{-1} \left\lceil \left( \frac{q - \Delta_H}{q - 2\Delta_H} \right) n \ln 4n \right\rceil \) and \( T := \left\lceil \left( \frac{q - \Delta_H}{q - 2\Delta_H} \right) n \ln \frac{2n}{\varepsilon} \right\rceil \), then Algorithm 4 outputs \( M \) sample over \( \Omega_{\text{Col}} \) satisfying that

\[
\mathcal{D}_{pw} \left( \chi^{(T)}, \pi \right) \leq \varepsilon
\]

where \( \pi \) is the uniform distribution over \( \Omega_{\text{Col}} \). The running time of Algorithm 4 is

\[
O(TnM \log(M)) = O^* \left( \left( \frac{q - \Delta_H}{q - 2\Delta_H} \right)^2 qn^4 \varepsilon^{-1} \right)
\]

where \( O^* \) ignores poly log term.

**Proof.** We check Algorithm 4 for each Step. Step 0 sets all \( M \) tokens on \( c' \in \Omega_{\text{Col}} \), which takes \( O(n\Delta_H) \) time. Step 1(a) takes \( O(nM \log(M)) \) time, by heapifying \( W^t[i] \) \( (i = 1, \ldots, M) \) with the lexicographic order on \( \Omega_{\text{Col}} \). Step 1(b) takes \( O(nq\Delta_H) \) time to find all feasible solutions neighboring to \( c \). Since we repeat Step 1 \( T \) times, then we obtain the time complexity \( O(TnM \log(M)) \).

Now, (26) is clear from Corollary 6.3 since Algorithm 4 is an implementation of the deterministic sampler in Section 6.1.

---

17
7 Concluding Remarks

This paper is concerned with the functional-router model, that is a generalization of the rotor-router model, and gave an upper bound of $|\chi_v^{(t)} - \mu_v^{(t)}|$ for the general function-router model in case of the corresponding Markov chain is reversible. We also better bounds for several specific models, such as greedy-routing, rotor-router model on multidigraphs, and the functional-routing model based on the van der Corput sequence. We then proposed a deterministic sampling algorithm, and gave an upper bound of the point-wise distance. Using the deterministic sampling algorithm, we obtain polynomial-time deterministic algorithms for uniform sampling of knapsack solutions, linear extensions, matchings, $q$-coloring, etc. A bound of the point-wise distance independent of $\pi_{\max}/\pi_{\min}$ is a future work. Development of deterministic approximation algorithms based on the deterministic sampler for #P-hard problems is a challenge.

Acknowledgment

The authors would like to thank an anonymous reviewer of the previous version of the paper for letting us know the stack-walk by Holroyd and Propp.

References

[1] H. Akbari and Petra Berenbrink, Parallel rotor walks on finite graphs and applications in discrete load balancing, Proceedings of the 25th ACM symposium on Parallelism in algorithms and architectures (SPAA 2013), 186–195.

[2] O. Angel, A.E. Holroyd, J. Martin, and J. Propp, Discrete low discrepancy sequences, arXiv:0910.1077.

[3] S. Angelopoulos, B. Doerr, A. Huber, and K. Panagiotou, Tight bounds for quasirandom rumor spreading, The Electronic Journal of Combinatorics, 16 (2009), #R102.

[4] R. Bubley and M. Dyer, Faster random generation of linear extensions, Discrete Mathematics, 201 (1999), 81–88.

[5] J. Cooper, B. Doerr, T. Friedrich, and J. Spencer, Deterministic random walks on regular trees, Random Structures & Algorithms, 37 (2010), 353–366.

[6] J. Cooper, B. Doerr, J. Spencer, and G. Tardos, Deterministic random walks on the integers, European Journal of Combinatorics, 28 (2007), 2072–2090.

[7] C. Cooper, D. Ilcinkas, R. Klasing, and A. Kosowski, Derandomizing random walks in undirected graphs using locally fair exploration strategies, Distributed Computing, 24 (2011), 91–99.

[8] J. Cooper and J. Spencer, Simulating a random walk with constant error, Combinatorics, Probability and Computing, 15 (2006), 815–822.

[9] B. Doerr, Introducing quasirandomness to computer science, Efficient Algorithms, volume 5760 of Lecture Notes in Computer Science, Springer Verlag, (2009), 99–111.

[10] B. Doerr and T. Friedrich, Deterministic random walks on the two-dimensional grid, Combinatorics, Probability and Computing, 18 (2009), 123–144.
[11] B. Doerr, T. Friedrich, M. Künemann, and T. Sauerwald, Quasirandom rumor spreading: An experimental analysis, ACM Journal of Experimental Algorithmics, 16 (2011), 3.3:1–3.3:13.

[12] B. Doerr, T. Friedrich, and T. Sauerwald, Quasirandom rumor spreading. Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2008), 773–781.

[13] B. Doerr, T. Friedrich, and T. Sauerwald, Quasirandom rumor spreading: Expanders, push vs. pull, and robustness, Proceedings of the 36th International Colloquium on Automata, Languages and Programming, (ICALP 2009), 366–377.

[14] B. Doerr, T. Friedrich, and T. Sauerwald, Quasirandom rumor spreading on expanders, Electronic Notes in Discrete Mathematics, 34 (2009), 243–247.

[15] B. Doerr, A. Huber, and A. Levavi, Strong robustness of randomized rumor spreading protocols, Proceedings of the 20th International Symposium on Algorithms and Computation (ISAAC 2009), 812–821.

[16] M. Dyer, Approximate counting by dynamic programming, Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC 2003), 693–699.

[17] T. Friedrich, M. Gairing, and T. Sauerwald, Quasirandom load balancing, SIAM Journal on Computing, 41 (2012), 747–771.

[18] T. Friedrich and T. Sauerwald, The cover time of deterministic random walks, The Electronic Journal of Combinatorics, 17 (2010), R167.

[19] P. Gopalan, A. Klivans, and R. Meka, Polynomial-time approximation schemes for knapsack and related counting problems using branching programs, arXiv:1008.3187v1, 2010.

[20] P. Gopalan, A. Klivans, R. Meka, D. Stefankovic, S. Vempala, and E. Vigoda, An FPTAS for #knapsack and related counting problems, Proceedings of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS 2011), 817–826.

[21] A.E. Holroyd and J. Propp, Rotor walks and Markov chains, M. Lladser, R.S. Maier, M. Mishna, A. Rechnitzer, (eds.), Algorithmic Probability and Combinatorics, The American Mathematical Society, 2010, 105–126.

[22] H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, Bulletin of the Chemical Society of Japan, 44 (1971), 2332–2339.

[23] A. Huber and N. Fountoulakis, Quasirandom broadcasting on the complete graph is as fast as randomized broadcasting, Electronic Notes in Discrete Mathematics, 34 (2009), 553–559.

[24] S. Ikeda, I. Kubo, M. Yamashita, The hitting and cover times of random walks on finite graphs using local degree information, Theoretical Computer Science, 410 (2009), 94–100.

[25] M. Jerrum, A very simple algorithm for estimating the number of k-colorings of a low-degree graph, Random Structures Algorithms 7(1995), 157–165.

[26] D. Jerison, L. Levine, S. Sheffield, Logarithmic fluctuations for internal DLA, Journal of the American Mathematical Society, 25 (2012), 271–301.
[27] M. Jerrum and A. Sinclair, Approximation algorithms for NP-hard problems, D.S. Hochbaum ed., The Markov chain Monte Carlo method: an approach to approximate counting and integration, PWS Publishing, 1996.

[28] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, Journal of the ACM, 51 (2004), 671–697.

[29] H. Kajino, S. Kijima, and K. Makino, Discrepancy analysis of deterministic random walks on finite irreducible graphs, discussion paper.

[30] S. Kijima, K. Koga, and K. Makino, Deterministic random walks on finite graphs, Proceedings of Analytic Algorithmics and Combinatorics (ANALCO 2012), 16–25.

[31] S. Kijima and T. Matsui, Approximation algorithm and perfect sampler for closed Jackson networks with single servers, SIAM Journal on Computing, 38 (2008), 1484–1503.

[32] M. Kleber, Goldbug variations, The Mathematical Intelligencer, 27 (2005), 55–63.

[33] G.F. Lawler, M. Bramson, and D. Griffeath, Internal diffusion limited aggregation, The Annals of Probability, 20 (1992), 2117–2140.

[34] L. Levine and Y. Peres, The rotor-router shape is spherical, The Mathematical Intelligencer, 27 (2005), 9–11.

[35] L. Levine and Y. Peres, Spherical asymptotics for the rotor-router model in \( \mathbb{Z}^d \), Indiana University Mathematics Journal, 57 (2008), 431–450.

[36] L. Levine and Y. Peres, Strong spherical asymptotics for rotor-router aggregation and the divisible sandpile, Potential Analysis, 30 (2009), 1–27.

[37] D.A. Levine, Y. Peres, and E.L. Wilmer, Markov Chain and Mixing Times, American Mathematical Society, 2008.

[38] B. Morris and A. Sinclair, Random walks on truncated cubes and sampling 0-1 knapsack solutions, SIAM Journal on Computing, 34 (2004), 195–226.

[39] R. Montenegro and P. Tetali, Mathematical Aspects of Mixing Times in Markov Chains, NOW Publishers, 2006.

[40] H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, Bull. Amer. Math. Soc., 84(1978), 957–1042

[41] V. Priezzhev, D. Dhar, A. Dhar, and S. Krishnamurthy, Eulerian walkers as a model of self-organized criticality, Physical Review Letters, 77 (1996), 5079–5082.

[42] J. Propp and D. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Structures Algorithms, 9 (1996), 223–252.

[43] T. Shiraga, Y. Yamauchi, S. Kijima, and M. Yamashita, Deterministic random walks for irrational transition probabilities, IPSJ SIG Technical Reports, 2012-AL-142(2), 2012. (in Japanese).

[44] A. Sinclair, Algorithms for Random Generation & Counting, A Markov chain approach, Birkhäuser, 1993.
[45] D. Stefankovic, S. Vempala, and E. Vigoda, A deterministic polynomial-time approximation scheme for counting knapsack solutions, SIAM Journal on Computing, 41 (2012), 356–366.

[46] R. Tijdeman, The chairman assignment problem, Discrete Math. 32 (1980), 323–330.

[47] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM Journal on Computing, 8 (1979), 410–421.

[48] J. G. van der Corput, Verteilungsfunktionen, Proc. Akad. Amsterdam, 38 (1935), 813–821, 1058–1066.

A Fundamental Properties of Markov Chain and Mixing Time

A.1 Proof of Proposition 2.1

In this Section, we show Proposition 2.1 (see e.g., [39, 37]).

Proposition 2.1 For any integers \(\ell (\ell \geq 1)\) and \(k (0 \leq k < \tau (\gamma))\),
\[h (\ell \cdot \tau (\gamma) + k) \leq \frac{1}{2} (2\gamma)^\ell\]
holds for any \(0 < \gamma < 1/2\).

To begin with, we define
\[\bar{h}(t) \overset{\text{def}}{=} \max_{v,w \in V} D_{tv}(P^t_v, P^t_w).\] (27)

Then, we show the following.

Lemma A.1. Let \(\xi, \zeta \in \mathbb{R}^{|V|}\) be probability distributions. Then,
\[D_{tv}(\xi P^t, \zeta P^t) \leq \bar{h}(t)\]
holds for any \(\xi, \zeta \in \mathbb{R}^{|V|}\) and for any \(t \geq 0\).

Proof. First, we obtain that
\[D_{tv}(\xi P^t - \zeta P^t) = \frac{1}{2} \left\| \sum_{v \in V} \xi_v P^t_v - \sum_{w \in V} \zeta_w P^t_w \right\|_1\]
\[= \frac{1}{2} \left\| \sum_{v \in V} \xi_v P^t_v - \sum_{w \in V} \zeta_w P^t_w \right\|_1\]
\[= \frac{1}{2} \left\| \sum_{v \in V} \sum_{w \in V} \xi_v \zeta_w (P^t_v - P^t_w) \right\|_1.\] (28)

Second equality follows the fact that \(\sum_{u \in V} \xi_u = \sum_{u \in V} \zeta_u = 1\), since \(\xi\) and \(\zeta\) are probability distributions. Thus,
\[\frac{1}{2} \sum_{v \in V} \sum_{w \in V} \xi_v \zeta_w \left\| P^t_v - P^t_w \right\|_1\]
\[\leq \frac{1}{2} \max_{v,w \in V} \left\| P^t_v - P^t_w \right\|_1 \sum_{v \in V} \sum_{w \in V} \xi_v \zeta_w = \bar{h}(t)\]
holds, and we obtain the claim. \(\square\)
Lemma A.2.

\[ h(t) \leq \bar{h}(t) \leq 2h(t) \]

holds for any \( t \geq 0 \).

**Proof.** Let \( e_v \in \mathbb{R}^{|V|} \) denote the \( v \)-th unit vector. By Lemma A.1,

\[ D_{tv}(P^t_{v,\cdot}, \pi) = D_{tv}(e_v P^t_{\cdot \cdot}, \pi P^t_{\cdot \cdot}) \leq \bar{h}(t) \]

holds for any \( v \in V \), and we obtain \( h(t) \leq \bar{h}(t) \).

By the definition of the total variation distance,

\[ D_{tv}(P^t_{v,\cdot}, P^t_{w,\cdot}) = \frac{1}{2} \sum_{u \in V} |P^t_{v,u} - \pi_u + \pi_u - P^t_{w,u}| \leq \frac{1}{2} \sum_{u \in V} |P^t_{v,u} - \pi_u| + \frac{1}{2} \sum_{u \in V} |\pi_u - P^t_{w,u}| \leq 2h(t) \]

holds for any \( v, w \in V \). We obtain \( \bar{h}(t) \leq 2h(t) \). \( \square \)

**Lemma A.3.** Suppose a vector \( \xi \in \mathbb{R}^{|V|} \) satisfies \( \sum_{i \in V} \xi_i = 0 \), then

\[ \|\xi P^t\|_1 \leq \|\xi\|_1 \bar{h}(t) \]

holds for any \( t \geq 0 \).

**Proof.** Let \( \xi^+, \xi^- \in \mathbb{R}^{|V|} \) be defined by \( \xi^+_i = \max\{\xi_i, 0\} \) and \( \xi^-_i = \max\{-\xi_i, 0\} \). Then,

\[ \xi^+_i - \xi^-_i = \max\{\xi_i, 0\} - \max\{-\xi_i, 0\} = \xi_i \quad (29) \]

holds, thus we obtain

\[ \xi = \xi^+ - \xi^- \quad (30) \]

Since \( \sum_{i \in V} \xi_i = 0 \),

\[ \sum_{i \in V} \xi^+_i = \sum_{i \in V} \xi^-_i \quad (31) \]

holds by (29), and we obtain the first inequality.

Now, we show the second inequality. By the definition of \( \xi^+ \) and \( \xi^- \),

\[ \xi^+_i + \xi^-_i = \max\{\xi_i, 0\} + \max\{-\xi_i, 0\} = |\xi_i| \]

holds. Hence

\[ \sum_{i \in V} \xi^+_i + \sum_{i \in V} \xi^-_i = \sum_{i \in V} |\xi_i| \quad (32) \]

holds. Thus, by (31) and (32),

\[ \sum_{i \in V} \xi^+_i = \sum_{i \in V} \xi^-_i = \frac{1}{2} \sum_{i \in V} |\xi_i| = \frac{1}{2} \|\xi\|_1 \quad (33) \]
holds, hence $\frac{\xi^+}{\frac{1}{2}\|\xi\|_1}$ and $\frac{\xi^-}{\frac{1}{2}\|\xi\|_1}$ are probabilistic distribution. Finally, by Lemma A.1 and (30),

$$\|\xi P^t\|_1 = \|\xi^+ P^t - \xi^- P^t\|_1 = \frac{1}{2}\|\xi\|_1 \cdot \left\| \frac{\xi^+}{\frac{1}{2}\|\xi\|_1} P^t - \frac{\xi^-}{\frac{1}{2}\|\xi\|_1} P^t \right\|_1 \leq \|\xi\|_1 \bar{h}(t)$$

holds, and we obtain the claim.

**Lemma A.4.**

$$h(s + t) \leq h(s)\bar{h}(t), \quad \text{and} \quad \bar{h}(s + t) \leq \bar{h}(s)\bar{h}(t)$$

hold for any $s, t \geq 0$.

**Proof.** By Lemma A.3

$$\frac{1}{2} \|e_v P^{s+t} - \pi\|_1 = \frac{1}{2} \|(e_v P^s - \pi) P^t\|_1 \leq \frac{1}{2} \|e_v P^s - \pi\|_1 \bar{h}(t)$$

holds for any $v \in V$, and we get $h(s + t) \leq h(s)\bar{h}(t)$. Similarly,

$$\frac{1}{2} \|e_v P^{s+t} - e_w P^{s+t}\|_1 = \frac{1}{2} \|(e_v P^s - e_w P^s) P^t\|_1 \leq \frac{1}{2} \|e_v P^s - e_w P^s\|_1 \bar{h}(t)$$

holds for any $v, w \in V$, and we get $\bar{h}(s + t) \leq \bar{h}(s)\bar{h}(t)$. \qed

**Proof of Proposition 2.1** Using Lemma A.4

$$h(\ell \cdot \tau(\gamma) + k) \leq h(\ell \cdot \tau(\gamma))\bar{h}(k) \leq h(\ell \cdot \tau(\gamma)) \leq h(\tau(\gamma)) \cdot \bar{h}(\ell \cdot \tau(\gamma)) = h(\tau(\gamma)) \cdot (\bar{h} \bar{h}(\tau(\gamma)))^{\ell-1} \quad (34)$$

holds. By Lemma A.2

$$(34) \leq h(\tau(\gamma)) \cdot (2h(\tau(\gamma)))^{\ell-1} \leq \gamma \cdot (2\gamma)^{\ell-1} \leq \frac{1}{2}(2\gamma)^\ell$$

holds, and we obtain the claim. \qed

**A.2 Supplementary proof of Theorem 3.1**

We show the following proposition, appearing in the proof of Theorem 3.1.

**Proposition A.5.** If $P$ is reversible, then

$$\pi_u P^t_{u,v} = \pi_v P^t_{v,u}$$

holds for any $u, v \in V$ and for any $t \geq 1$. 

---

23
We obtain the claim.

\[ \pi_u P_{u,v}^{t+1} = \pi_u \sum_{w \in V} P_{w,u}^t P_{w,v} = \sum_{w \in V} \pi_w P_{w,u}^t P_{w,v} = \sum_{w \in V} P_{w,u}^t \pi_w P_{w,v} = \pi_v \sum_{w \in V} P_{w,u}^t P_{w,v} = \pi_v P_{v,u}^{t+1}. \]

We obtain the claim.

\[ \square \]

### B Supplemental Proofs in Section 5.3

This section presents a proof of Theorem 5.5 and Lemma 5.6. To begin with, we remark two lemmas concerning the function \( \psi \) defined by (17).

#### Lemma B.1.

For any \( i \in \mathbb{Z}_{\geq 0} \) and any \( k \in \mathbb{Z}_{\geq 0} \),

\[ \psi(i) = \left( i \mod 2^k \right) + \psi \left( \left\lfloor \frac{i}{2^k} \right\rfloor \right) \cdot \frac{1}{2^k} \]

holds.

**Proof.** In case of \( i < 2^k \), the claim is easy since \( \psi(i \mod 2^k) = \psi(i) \) and \( \psi(\lfloor i/2^k \rfloor) = \psi(0) = 0 \) holds. Suppose that \( i \geq 2^k \), and that \( i \) is represented in binary as \( i = \sum_{j=0}^{[\log_2 i]} \beta_j(i) \cdot 2^j \) using \( \beta_j(i) \in \{0, 1\} \) \((j \in \{0, 1, \ldots, [\log_2 i]\})\). Then,

\[ i \mod 2^k = \left( \sum_{j=0}^{[\log_2 i]} \beta_j(i) \cdot 2^j \right) \mod 2^k = \sum_{j=0}^{k-1} \beta_j(i) \cdot 2^j, \]  

and

\[ \left\lfloor \frac{i}{2^k} \right\rfloor = \sum_{j=k}^{[\log_2 i]} \beta_j(i) \cdot 2^{j-k} \]

hold, respectively. Let \( l = j - k \) and let \( b_l = \beta_{l+k}(i) \) \((l = 0, 1, \ldots, [\log_2 i] - k)\), for convenience, then

\[ \psi \left( i \mod 2^k \right) + \psi \left( \left\lfloor \frac{i}{2^k} \right\rfloor \right) \cdot \frac{1}{2^k} = \left( \sum_{j=0}^{k-1} \beta_j(i) 2^j \right) + \psi \left( \sum_{j=k}^{N} \beta_j(i) 2^{j-k} \right) \cdot \frac{1}{2^k} = \sum_{j=0}^{k-1} \beta_j(i) 2^{-j+1} + \left( \sum_{l=0}^{[\log_2 i]-k} b_l \cdot 2^{-(j+1)} \right) \cdot \frac{1}{2^k} = \sum_{j=0}^{k-1} \beta_j(i) 2^{-j+1} + \frac{[\log_2 i]-k}{2^k} \sum_{l=0}^{[\log_2 i]} b_l \cdot 2^{-(j+1)} = \sum_{j=0}^{[\log_2 i]} \beta_j(i) 2^{-j+1} = \psi(i) \]
and, we obtain the claim.

**Lemma B.2.** For any $k \in \mathbb{Z}_{\geq 0}$ and $\alpha \in \{0, 1, \ldots, 2^k - 1\}$,

$$\psi(2^k + \alpha) = \frac{1}{2^{k+1}} + \psi(\alpha)$$

holds.

**Proof.** By Lemma B.1,

$$\psi(2^k + \alpha) = \psi\left( (2^k + \alpha) \mod 2^k \right) + \psi\left( \left\lfloor \frac{2^k + \alpha}{2^k} \right\rfloor \right) \cdot \frac{1}{2^k}$$

$$= \psi(\alpha) + \psi\left( \frac{1}{2^k} \right)$$

$$= \frac{1}{2^{k+1}} + \psi(\alpha)$$

holds, where the last equality follows $\psi(1) = 1/2$ by the definition.

Now, we define

$$\Phi[z, z'] \overset{\text{def}}{=} \{ \psi(i) \mid i \in \{z, z + 1, \ldots, z' - 1\} \}$$

(35)

for $z, z' \in \mathbb{Z}_{\geq 0}$ satisfying $z < z'$. For convenience, we define $\Phi[z, z] = \emptyset$. It is not difficult to see that

$$\Phi[0, 2^k] = \left\{ \frac{i}{2^k} \mid i \in \{0, 1, \ldots, 2^k - 1\} \right\}$$

(36)

holds for any $k \in \mathbb{Z}_{\geq 0}$. In general, we can show the following lemma, using Lemmas B.1 and B.2.

**Lemma B.3.** For any $z \in \mathbb{Z}_{\geq 0}$ and for any $k \in \mathbb{Z}_{\geq 0}$

$$\left| \Phi[z, z + 2^k] \cap \left[ \frac{i}{2^k}, \frac{i + 1}{2^k} \right) \right| = 1$$

holds for any $i \in \{0, 1, \ldots, 2^k - 1\}$.

**Proof.** By Lemma B.1,

$$\Phi[z, z + 2^k] = \left\{ \psi(z + i) \mid i \in \{0, 1, \ldots, 2^k - 1\} \right\}$$

$$= \left\{ \psi \left( (z + i) \mod 2^k \right) + \psi\left( \left\lfloor \frac{z + i}{2^k} \right\rfloor \right) \mid i \in \{0, 1, \ldots, 2^k - 1\} \right\}$$

holds. Since $0 \leq \psi(z') < 1$ holds for any $z' \in \mathbb{Z}_{\geq 0}$,

$$\psi \left( (z + i) \mod 2^k \right) + \psi\left( \left\lfloor \frac{z + i}{2^k} \right\rfloor \right)$$

$$\in \left[ \psi \left( (z + i) \mod 2^k \right) + \psi\left( (z + i) \mod 2^k \right) + \frac{1}{2^k} \right]$$

(37)
holds for each $i \in \{0, 1, \ldots, 2^k - 1\}$. The observation \(36\) implies that
\[
\left\{ \psi \left( \left( z + i \right) \mod 2^k \right) \mid i \in \{0, 1, \ldots, 2^k - 1\} \right\}
= \left\{ \frac{j}{2^k} \mid j \in \{0, 1, \ldots, 2^k - 1\} \right\}
\tag{38}
\]
holds. Notice that \(38\) implies $\psi \left( \left( z + i \right) \mod 2^k \right) \neq \psi \left( \left( z + i' \right) \mod 2^k \right)$ for any distinct $i, i' \in \{0, 1, \ldots, 2^k - 1\}$. Now, the claim is clear by \(37\) and \(38\). □

Note that Lemma B.3 implies that
\[
\Phi(z, z + 2^k) = \left\{ \frac{i}{2^k} + \frac{\theta(i)}{2^k} \mid i \in \{0, 1, \ldots, 2^k - 1\} \right\}
\tag{39}
\]
using appropriate $\theta(i) \in [0, 1)$ for $i = 0, 1, \ldots, 2^k - 1$.

**Lemma B.4.** Let $z_0, k \in \mathbb{Z}_{\geq 0}$, and let $x, y \in [0, 1)$ satisfy $x < y$. Then,
\[
\left| \Phi(z_0, z_0 + 2^k) \cap [x, y] \right| - 2^k \cdot (y - x) < 2
\]
holds.

**Proof.** Lemma B.3 and Equation 39 implies that there exists $s, t \in \{0, 1, \ldots, 2^k - 1\}$ such that $s \leq t$ and
\[
\frac{s}{2^k} + \frac{\theta(s)}{2^k} \leq x < \frac{s + 1}{2^k} + \frac{\theta(s + 1)}{2^k}
\tag{40}
\]
\[
\frac{t}{2^k} + \frac{\theta(t)}{2^k} \leq y < \frac{t + 1}{2^k} + \frac{\theta(t + 1)}{2^k}
\tag{41}
\]
where we assume
\[
-\frac{1}{2^k} + \frac{\theta(-1)}{2^k} = 0 \quad \text{and} \quad \frac{2^k}{2^k} + \frac{\theta(2^k)}{2^k} = 1
\]
for convenience. Then, it is clear that
\[
\Phi(z_0, z_0 + 2^k) \cap [x, y] = \left\{ \frac{i}{2^k} + \frac{\theta(i)}{2^k} \mid i \in \{s, s + 1, \ldots, t - 1\} \right\}
\]
where $s = t$ means that $\Phi(z_0, z_0 + 2^k) \cap [x, y] = \emptyset$. By (40) and (41), we obtain
\[
2^k \cdot x - \theta(s + 1) - 1 < s \leq 2^k \cdot x - \theta(s),
\]
\[
2^k \cdot y - \theta(t + 1) - 1 < t \leq 2^k \cdot y - \theta(t)
\]
respectively, and hence we obtain that
\[
2^k \cdot (y - x) - \theta(t + 1) - 1 + \theta(s) < t - s < 2^k \cdot (y - x) - \theta(t) + \theta(s + 1) + 1.
\]
Since $|\Phi(z_0, z_0 + 2^k) \cap [x, y]| = t - s$ and $0 \leq \theta(i) < 1$ ($i \in \{0, 1, \ldots, 2^k - 1\}$),
\[
2^k \cdot (y - x) - 2 < \left| \Phi(z_0, z_0 + 2^k) \cap [x, y] \right| < 2^k \cdot (y - x) + 2
\]
holds, and we obtain the claim. □
Using Lemma B.4, we obtain the following.

**Lemma B.5.** Let \( z_0 \in \mathbb{Z}_{\geq 0}, z \in \mathbb{Z}_{> 0}, \) and let \( x, y \in [0, 1) \) satisfy \( x < y. \) Then,

\[
\left| \frac{\Phi[z_0, z_0 + z]}{z} \cap [x, y) \right| - (y - x) \right| < 2 \left| \frac{\lg z}{z} \right| + 2
\]

holds.

**Proof.** For simplicity, let

\[
\Phi^* [z_0, z_0 + \ell] \quad \text{def.} \quad \Phi(z_0, z_0 + \ell) \cap [x, y)
\]

for any \( \ell \in \mathbb{Z}_{> 0}. \) Then, notice that

\[
|\Phi^* [z_0, z_0 + z]| = |\Phi^* [z_0, z_0 + z']| + |\Phi^* [z_0 + z', z_0 + z)|
\]

holds for any \( z' \in \mathbb{Z}_{> 0} \) satisfying \( z' < z. \) Now, suppose \( z \) is represented as \( z = \sum_{j=0}^{\lfloor \lg z \rfloor} \beta_j(z) \cdot 2^j \) in binary, where \( \beta_j(z) \in \{0, 1\}. \) Using Lemma B.4, we obtain that

\[
|\Phi^* [z_0, z_0 + z]| = \Phi^* \left[ z_0, z_0 + \sum_{j=0}^{\lfloor \lg z \rfloor} \beta_j(z) \cdot 2^j \right]
\]

\[
= \Phi^* \left[ z_0, z_0 + \beta_0(z) \cdot 2^0 \right] + \sum_{k=0}^{\lfloor \lg z \rfloor - 1} \Phi^* \left[ z_0 + \sum_{j=0}^{k} \beta_j(z) \cdot 2^j, z_0 + \sum_{j=0}^{k+1} \beta_j(z) \cdot 2^j \right]
\]

\[
< \beta_0(z) \cdot 2^0 \cdot (y - x) + 2 + \sum_{k=0}^{\lfloor \lg z \rfloor - 1} \left( \beta_{k+1}(z) \cdot 2^{k+1} \cdot (y - x) + 2 \right)
\]

\[
= \sum_{k=0}^{\lfloor \lg z \rfloor} \left( \beta_k(z) \cdot 2^k \cdot (y - x) + 2 \right)
\]

\[
= 2(\lfloor \lg z \rfloor + 1) + (y - x) \sum_{k=0}^{\lfloor \lg z \rfloor} \beta_k(z) \cdot 2^k
\]

\[
= 2(\lfloor \lg z \rfloor + 1) + z \cdot (y - x).
\]

In a similar way, we also have

\[
|\Phi^* [z_0, z_0 + z]| > 2(\lfloor \lg z \rfloor - 1) + z \cdot (y - x),
\]

and we obtain the claim. \( \Box \)

By Lemma B.5, it is not difficult to see Theorem 5.5 and Lemma 5.6 holds.