Total Variation Discrepancy of Deterministic Random Walks for Ergodic Markov Chains

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

Abstract

Motivated by a derandomization of Markov chain Monte Carlo (MCMC), this paper investigates *deterministic random walks*, which is a deterministic process analogous to a random walk. While there are some progress on the analysis of the vertex-wise discrepancy (i.e., L_∞ discrepancy), little is known about the *total variation discrepancy* (i.e., L_1 discrepancy), which plays a significant role in the analysis of an FPRAS based on MCMC. This paper investigates upper bounds of the L_1 discrepancy between the expected number of tokens in a Markov chain and the number of tokens in its corresponding deterministic random walk. First, we give a simple but nontrivial upper bound $\mathrm{O}(mt^*)$ of the L_1 discrepancy for any ergodic Markov chains, where m is the number of edges of the transition diagram and t^* is the mixing time of the Markov chain. Then, we give a better upper bound $\mathrm{O}(m\sqrt{t^* \log t^*})$ for *non-oblivious* deterministic random walks, if the corresponding Markov chain is ergodic and lazy. We also present some lower bounds.

Key words: Rotor router model, Propp machine, load balancing, Markov chain Monte Carlo (MCMC), mixing time

1 Introduction

Background Markov chain Monte Carlo (MCMC) is a powerful technique of designing randomized approximation algorithms for #P-hard problems. Jerrum et al. [21] showed the equivalence in the sense of the polynomial time computation between *almost* uniform generation and randomized approximate counting for self-reducible problems. A number of fully polynomial-time randomized approximation schemes (FPRAS) based on their technique have been developed for #P-hard problems, such as the volume of a convex body [14, 25, 11], integral of a log-concave function [25], partition function of the Ising model [19], and counting bipartite matchings [20]. When designing an FPRAS based on the technique, it is important that the *total variation distance* of the approximate distribution from the target distribution is sufficiently small, and hence analyses of the mixing times of Markov chains are central issues in a series of works on MCMC for FPRAS to guarantee a small total variation distance. See also Section 2.1 for the terminology of Markov chains.

In contrast, not many results are known about *deterministic* approximation algorithms for #P-hard problems. A remarkable progress is the correlation decay technique, independently devised by Weitz [31] and Bandyopadhyay and Gamarnik [5], and there are several recent developments on the technique. For counting 0-1 knapsack solutions, Gopalan et al. [16], and Stefankovic et al. [29] gave deterministic approximation algorithms (see also [17]). Ando and Kijima [2] gave an FPTAS based on approximate convolutions for computing the volume of a 0-1 knapsack polytope. A direct derandomization of MCMC algorithms is not known yet, but it holds a potential for a general scheme of designing deterministic approximation algorithms for #P-hard problems. *Deterministic random walks* [10, 9, 13, 8, 23, 22, 27] may be used as a substitute for Markov chains, for the purpose.

Deterministic random walk Deterministic random walk is a deterministic process analogous to a (multiple) random walk[1]. A configuration $\chi^{(t)} \in \mathbb{Z}_{\geq 0}^V$ of M tokens distributed over a (finite) vertex set V is deterministically updated from time t to $t+1$ by routers equipped on vertices. The router on a vertex $u \in V$ deterministically serves tokens on u to neighboring vertex v with a ratio (about) $P_{uv} \in [0,1]$ such that $\sum_{v \in V} P_{uv} = 1$, i.e., $P = (P_{uv}) \in \mathbb{R}^{V \times V}$ is a transition matrix (when V is finite). See Section 2.2 for the detailed description of the model with which this paper is concerned. Note that the expected configuration $\mu^{(t)} \in \mathbb{R}_{\geq 0}^V$ of M tokens in a multiple random walk at time t is given by $\mu^{(t)} = \chi^{(0)} P^t$ on the assumption that $\chi^{(0)} = \mu^{(0)}$.

Cooper and Spencer [10] investigated the rotor-router model, which is a deterministic random walk corresponding to a simple random walk, and showed for the d-dimensional (infinite) integer lattice that the maximum vertex-wise discrepancy $\|\chi^{(t)} - \mu^{(t)}\|_\infty$ is upper bounded by a constant c_d, which depends only on d but is independent of the total number of tokens.

*Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
{takeharu.shiraga,yamauchi,kijima,mak}@inf.kyushu-u.ac.jp

[1] "multiple random walk" means random walks of many tokens.

Later, it is shown that $c_1 \simeq 2.29$ [9] and c_2 is about 7.29 or 7.83 depending on the routers [13]. On the other hand, Cooper et al. [8] gave an example of a rotor-router on the infinite k-regular tree, such that its vertex-wise discrepancy gets $\Omega(\sqrt{kt})$ for an arbitrarily fixed t.

Motivated by general transition matrices, Kijima et al. [23] investigated a rotor-router model on finite multidigraphs, and gave a bound $\mathrm{O}(n|\mathcal{A}|)$ of the vertex-wise discrepancy when P is rational, ergodic and reversible, where $n = |V|$ and $\mathcal{A}$ denotes the set of multiple edges. For an arbitrary rational transition matrix P, Kajino et al. [22] gave an upper bound using the second largest eigenvalue λ^* of P and some other parameters of P. To deal with irrational transition probabilities, Shiraga et al. [27] presented a generalized notion of the rotor-router model, which they call the *functional router model*. They gave a bound $\mathrm{O}((\pi_{\max}/\pi_{\min})t^*\Delta)$ of the vertex-wise discrepancy for a specific functional router model (namely, *SRT-router model*) when P is ergodic and reversible, where t^* denotes the mixing time of P and $\pi_{\max}$ (resp. $\pi_{\min}$) is the maximum (resp. minimum) element of the stationary distribution vector π of P. Using [27], Shiraga et al. [28] discussed the time complexity of a simulation, in which they are concerned with an *oblivious* version, meaning that the states of routers are reset in each step while the deterministic random walk above mentioned carries over the states of routers to the next step.

Similar, or essentially the same concepts have been independently developed in various contexts, such as load-balancing, information spreading and self-organization. Rabani et al. [26] investigated the diffusive model for load balancing, which is an oblivious version of deterministic random walk, and showed for the model that the vertex-wise discrepancy is $\mathrm{O}\left(\Delta \log(n)/(1-\lambda^*)\right)$ when P is symmetric and ergodic, where Δ is the maximum degree of the transition diagram of P. Friedrich et al. [15] proposed the BED algorithm for load balancing, which uses some extra information in the previous time, and they gave $\mathrm{O}(d^{1.5})$ for hypercube and $\mathrm{O}(1)$ for constant-dimensional tori. Akbari and Berenbrink [1] discussed the relation between the BED algorithm and the rotor-router model, and gave the same bounds for a rotor-router model. Berenbrink et al. [6] investigated about cumulatively fair balancers algorithms, which includes the rotor-router model, and gave an upper bound $\mathrm{O}(d\min(\sqrt{\log(n)/(1-\lambda^*)}, \sqrt{n}))$ for a lazy version of simple random walks on d-regular graphs.

As a closely related topic, the behavior of the rotor-router model with a single token has also been investigated. Holroyd and Propp [18] investigated the frequency $\nu^{(t)} \in \mathbb{Z}_{\geq 0}^V$ of visits of the token in t steps, and showed that $\|\nu^{(t)}/t - \pi\|_\infty$ is $\mathrm{O}(mn/t)$. Preceding [18], Yanovski et al. [32] showed that the rotor-router model with a single token always stabilizes to a traversal of an Eulerian cycle after $2mD$ steps at most, where D denotes the diameter of the graph. This result implies that the (edge) cover time of the rotor-router model with a single token is $\mathrm{O}(mD)$ for any graph. Bampas et al. [4] gave examples of which the stabilization time gets $\Omega(mD)$. Similar analyses for the rotor-router model with many tokens have been developed, recently. Dereniowski et al. [12] investigated the cover time of the rotor-router model with M tokens, and gave an upper $\mathrm{O}(mD/\log M)$ and an example of $\Omega(mD/M)$ as a lower bound. Chalopin et al. [7] gave an upper bound of its stabilization time is $\mathrm{O}(m^4D^2 + mD\log M)$, while they also showed that the period of a cyclic stabilized states can get as large as $2^{\Omega(\sqrt{n})}$.

Our results As we stated before, the total variation distance between the target distribution and approximate samples is significant in the analysis of MCMC algorithms. While there are several works on deterministic random walks concerning the vertex-wise discrepancy $\|\chi^{(t)} - \mu^{(t)}\|_\infty$ such as [26, 23, 22, 27, 6], little is known about the total variation discrepancy $\|\chi^{(t)} - \mu^{(t)}\|_1$. This paper investigates the total variation discrepancy to develop a new analysis technique aiming at derandomizing MCMC.

To begin with, we give a simple but nontrivial upper bound for any ergodic finite Markov chains, precisely we show $\|\chi^{(t)} - \mu^{(t)}\|_1 = \mathrm{O}(mt^*)$ where t^* is the mixing time of P and m is the number of edges of the transition diagram of P. In fact, the analyses are almost the same for both the non-oblivious model, including the rotor-router model [10, 23, 22, 6], and the oblivious model like [26, 28] in which the states of routers are reset in each step, and we in Section 3 deal with the oblivious model. We also give a lower bound for the oblivious model presenting an example such that $\|\chi^{(t)} - \mu^{(t)}\|_1 = \Omega(nt^*)$, which implies that the mixing time t^* is negligible in the L_1 discrepancy for the oblivious model, in general.

Then, we in Section 4 give a better upper bound for non-oblivious deterministic random walk, precisely we show $\|\chi^{(t)} - \mu^{(t)}\|_1 = \mathrm{O}(m\sqrt{t^*\log t^*})$ when P is ergodic and lazy. Notice that the upper bound does not require reversibility. The analysis technique is a modification of Berenbrink et al. [6], in which they investigated a lazy version of simple random walks on d-regular graphs. In fact, we also remark that the analysis technique by [6] for the vertex-wise discrepancy is extended to general graphs, precisely we show that $\|\chi^{(t)} - \mu^{(t)}\|_\infty = \mathrm{O}(\Delta\sqrt{t^*\log t^*})$ when P is ergodic, lazy, symmetric. We also present some lower bounds of L_1

Conditions on P	L_∞-discrepancy		L_1-discrepancy	
E. R. symmetric	$\mathrm{O}\left(\frac{\Delta\log(n)}{1-\lambda^*}\right)$	[26]	$\mathrm{O}\left(\frac{\Delta n\log(n)}{1-\lambda^*}\right)$	
E. R. L. rational	$\mathrm{O}(n\lvert\mathcal{A}\rvert)$	[23]	$\mathrm{O}(n^2\lvert\mathcal{A}\rvert)$	
any rational	$\mathrm{O}\left(\frac{\alpha^* n\lvert\mathcal{A}\rvert}{(1-\lambda^*)^\beta}\right)$	[22]	$\mathrm{O}\left(\frac{\alpha^* n^2\lvert\mathcal{A}\rvert}{(1-\lambda^*)^\beta}\right)$	
E. R.	$\mathrm{O}\left(\frac{\pi_{\max}}{\pi_{\min}}t^*\Delta\right)$	[27]	$\mathrm{O}\left(\frac{\pi_{\max}}{\pi_{\min}}t^*\Delta n\right)$	
E. R. L. simple r.w. d-regular	$\mathrm{O}\left(d\min\left(\sqrt{\frac{\log(n)}{1-\lambda^*}},\sqrt{n}\right)\right)$	[6]	$\mathrm{O}\left(m\min\left(\sqrt{\frac{\log(n)}{1-\lambda^*}},\sqrt{n}\right)\right)$	
E.			$\mathrm{O}(mt^*)$	Thm. 3.1
E. L.			$\mathrm{O}(m\sqrt{t^*\log t^*})$	Thm. 4.1
E. R. L. symmetric	$\mathrm{O}(\Delta\sqrt{t^*\log t^*})$	Thm. 4.2		

E.: ergodic, R.: reversible, L.: lazy

Table 1: Summary of known results on $\|\chi^{(t)}-\mu^{(t)}\|_\infty$ for finite graphs, and this work.

discrepancy for non-oblivious models.

Table 1 shows a summary of known results [26, 23, 22, 27, 6] on $\|\chi^{(t)}-\mu^{(t)}\|_\infty$, and the results by this work. The column of "L_1 discrepancy" shows the upper bounds of $\|\chi^{(t)}-\mu^{(t)}\|_1$ implied by the previous results [26, 23, 22, 27, 6], in comparison with upper bounds obtained by this paper.

2 Preliminaries

2.1 Random walk / Markov chain

As a preliminary step, we introduce some terminology of Markov chains (cf. [24]). Let $V=\{1,\ldots,n\}$ be a finite set, and let $P\in\mathbb{R}_{\geq 0}^{n\times n}$ be a transition matrix on V, which satisfies $\sum_{v\in V}P_{u,v}=1$ for any $u\in V$, where $P_{u,v}$ denotes the (u,v) entry of P ($P^t_{u,v}$ denotes (u,v) entry of P^t, as well). Let $\mathcal{G}=(V,\mathcal{E})$ be the transition digram of P, meaning that $\mathcal{E}=\{(u,v)\in V\times V\mid P_{u,v}>0\}$. Let $\mathcal{N}^+(v)$ and $\mathcal{N}^-(v)$ respectively denote the out-neighborhood and the in-neighborhood[2] of $v\in V$ on $\mathcal{G}$. For convenience, let $m=|\mathcal{E}|$, $\delta^+(v)=|\mathcal{N}^+(v)|$ and $\delta^-(v)=|\mathcal{N}^-(v)|$.

A finite Markov chain is called *ergodic* if P is *irreducible*[3] and *aperiodic*[4]. It is well known that any ergodic P has a unique *stationary distribution* $\pi\in\mathbb{R}^n_{>0}$ (i.e., $\pi P=\pi$), and the limit distribution is π (i.e., $\lim_{t\to\infty}\xi P^t=\pi$ for any probability distribution $\xi\in\mathbb{R}^n_{\geq 0}$ on V). Let ξ and ζ be probability distributions on V, then the *total variation distance* $\mathcal{D}_{\rm tv}$ between ξ and ζ is defined by

$$\mathcal{D}_{\rm tv}(\xi,\zeta) \overset{\text{def.}}{=} \max_{A\subseteq V}\left|\sum_{v\in A}(\xi_v-\zeta_v)\right| = \frac{1}{2}\|\xi-\zeta\|_1=\frac{1}{2}\sum_{v\in V}|\xi_v-\zeta_v|. \tag{2.1}$$

The *mixing time* of P is defined by[5]

$$\tau(\varepsilon)\overset{\text{def.}}{=}\max_{v\in V}\min\left\{t\in\mathbb{Z}_{\geq 0}\mid \mathcal{D}_{\rm tv}(P^t_{v,\cdot},\pi)\leq\varepsilon\right\} \tag{2.2}$$

for $\varepsilon>0$, and let

$$t^*\overset{\text{def.}}{=}\tau(1/4), \tag{2.3}$$

which is often used as an important characterization of P.

Let $\mu^{(0)}=(\mu^{(0)}_1,\ldots,\mu^{(0)}_n)\in\mathbb{Z}^n_{\geq 0}$ denote an initial configuration of M tokens over V. Suppose that each token randomly moves according to P. Let $\mu^{(t)}$ denote the *expected* configuration of tokens at time $t\in\mathbb{Z}_{\geq 0}$ in a Markov chain: then $\mu^{(t)}=\mu^{(0)}P^t$ holds. By the definition of mixing time, $(1/2)\|\mu^{(t)}/M-\pi\|_1\leq\varepsilon$ holds for any $t\geq\tau(\varepsilon)$ if P is ergodic.

2.2 Deterministic random walk: framework

A *deterministic random walk* is a deterministic process imitating $\mu^{(t)}$. Let $\chi^{(0)}=\mu^{(0)}$ and $\chi^{(t)}\in\mathbb{Z}^n_{\geq 0}$ denote the configuration of tokens at time $t\in\mathbb{Z}_{\geq 0}$

[2] $\mathcal{N}^+(v)=\{u\in V\mid P_{v,u}>0\}$ and $\mathcal{N}^-(v)=\{u\in V\mid P_{u,v}>0\}$.

[3] P is irreducible if $\forall u,v\in V,\exists t>0, P^t_{u,v}>0$ if and only if transition diagram of P is connected.

[4] P is aperiodic if $\forall v\in V,\mathrm{GCD}\{t\in\mathbb{Z}_{>0}\mid P^t_{v,v}>0\}=1$.

[5] $P^t_{v,\cdot}$ denotes the v-th row vector of P^t.

in a deterministic random walk. An update in a deterministic random walk is defined by $Z_{v,u}^{(t)}$ denoting the number of tokens moving from v to u at time t, where $Z_{v,u}^{(t)}$ must satisfy the condition that

$$\sum_{u\in\mathcal{N}^+(v)} Z_{v,u}^{(t)} = \chi_v^{(t)} \tag{2.4}$$

for any $v \in V$. Then, $\chi^{(t+1)}$ is defined by

$$\chi_u^{(t+1)} \stackrel{\text{def.}}{=} \sum_{v\in\mathcal{N}^-(u)} Z_{v,u}^{(t)} \tag{2.5}$$

for any $u \in V$. We will explain some specific deterministic random walks in Sections 3.1 and 4.1 by giving precise definitions of $Z_{v,u}^{(t)}$. We are interested in the question of whether $\chi^{(t)}$ approximates $\mu^{(t)}$ well in terms of the total variation discrepancy, i.e., the question is how large $\max_{A\subseteq V} |\chi_A^{(t)} - \mu_A^{(t)}| = (1/2)\|\chi^{(t)} - \mu^{(t)}\|_1$ ever gets.

At the end of this section, we introduce two notations which we will use in the paper. For any $\xi \in \mathbb{R}^V$ and $S \subseteq V$, let ξ_S denote $\sum_{v\in S}\xi_v$. For example, $\mu_S^{(t)} = \sum_{v\in S}\mu_v^{(t)}$ and $P_{u,S} = \sum_{v\in S} P_{u,v}$. For any $\xi \in \mathbb{R}^n$, $P \in \mathbb{R}^{n\times n}$ and $u \in V$, let $(\xi P)_u$ denotes the u-th element of the vector ξP, i.e., $(\xi P)_u = \sum_{v\in V}\xi_v P_{v,u}$.

3 Upper and lower bounds for oblivious model

This section is concerned with an oblivious version of deterministic random walk, which is closely related to the models in [26, 28].

3.1 Oblivious model

Given a transition matrix P and a configuration $\chi^{(t)}$ of tokens, we define $Z_{v,u}^{(t)}$ as follows. Assume that an arbitrary ordering $u_1, \ldots, u_{\delta^+(v)}$ on $\mathcal{N}^+(v)$ is prescribed for each $v \in V$. Then, let

$$Z_{v,u_i}^{(t)} = \begin{cases} \left\lfloor \chi_v^{(t)} P_{v,u_i} \right\rfloor + 1 & (i \le i^*) \\ \left\lfloor \chi_v^{(t)} P_{v,u_i} \right\rfloor & (\text{otherwise}) \end{cases} \tag{3.6}$$

where $i^* \stackrel{\text{def.}}{=} \chi_v^{(t)} - \sum_{i=1}^{\delta^+(v)} \lfloor \chi_v^{(t)} P_{v,u_i} \rfloor$ denotes the number of "surplus" tokens. It is easy to check that the condition (2.4) holds for any $v \in V$ and $t \in \mathbb{Z}_{\ge 0}$. Then, the configuration $\chi^{(t+1)}$ is updated according to (2.5), recursively. The following observation is easy from the definition (3.6) of $Z_{v,u}^{(t)}$.

Observation 3.1. *For any oblivious model, $|Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u}| \le 1$ holds for any $u, v \in V$ and $t \in \mathbb{Z}_{\ge 0}$.*

3.2 Upper bound

In this section, we give an upper bound of the total variation discrepancy.

Theorem 3.1. *Suppose $P \in \mathbb{R}_{\ge 0}^{n\times n}$ is ergodic. Then, for any oblivious model,*

$$\max_{S\subseteq V} \left|\chi_S^{(T)} - \mu_S^{(T)}\right| \;\le\; \frac{3}{2} m t^* = \mathrm{O}(m t^*)$$

holds for any $T \in \mathbb{Z}_{\ge 0}$.

Notice that the bound is independent of the number of tokens M, meaning that the total variation distance $\mathcal{D}_{\mathrm{tv}}(\chi^{(T)}/M, \mu^{(T)}/M)$ approaches asymptotically to zero as increasing the number of tokens M. We also remark that Theorem 3.1 only assumes that P is ergodic.

Proof. Let $\phi^{(t)} = \chi^{(t)} - \chi^{(t-1)}P$, for convenience. By (2.5) and Observation 3.1,

$$\begin{aligned} |\phi_u^{(t+1)}| &= \left|\left(\chi^{(t+1)} - \chi^{(t)}P\right)_u\right| \\ &= \left|\sum_{v\in\mathcal{N}^-(u)} (Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u})\right| \\ &\le \sum_{v\in\mathcal{N}^-(u)} \left|Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u}\right| \\ &\le \delta^-(u) \end{aligned} \tag{3.7}$$

holds for any $u \in V$ and $t \in \mathbb{Z}_{\ge 0}$. Next, we see that

$$\begin{aligned} &\sum_{t=0}^{T-1} \phi^{(T-t)} P^t \\ &= \sum_{t=0}^{T-1} \left(\chi^{(T-t)} P^t - \chi^{(T-t-1)} P^{t+1}\right) \\ &= \chi^{(T)} P^0 - \chi^{(0)} P^T = \chi^{(T)} - \mu^{(T)} \end{aligned} \tag{3.8}$$

holds, since $\mu^{(T)} = \chi^{(0)} P^T$ holds by the assumption that $\mu^{(0)} = \chi^{(0)}$. By (3.8),

$$\begin{aligned} \chi_S^{(T)} - \mu_S^{(T)} &= \left(\sum_{t=0}^{T-1} \phi^{(T-t)} P^t\right)_S \\ &= \sum_{t=0}^{T-1} \sum_{u\in V} \phi_u^{(T-t)} P_{u,S}^t \\ &= \sum_{t=0}^{T-1} \sum_{u\in V} \phi_u^{(T-t)} \left(P_{u,S}^t - \pi_S\right) \end{aligned} \tag{3.9}$$

holds, where the last equality follows from the fact that

$$\begin{aligned}\sum_{u\in V}\phi_u^{(t)} &= \sum_{u\in V}\left(\chi^{(t)}-\chi^{(t-1)}P\right)_u\\ &= \sum_{u\in V}\chi_u^{(t)}-\sum_{u\in V}\sum_{v\in V}\chi_v^{(t-1)}P_{v,u}\\ &= M-M=0\end{aligned}$$

holds for any $t\geq 1$. By (3.9), we obtain that

$$\begin{aligned}&\left|\chi_S^{(T)}-\mu_S^{(T)}\right|\\ &\quad\leq \left|\sum_{t=0}^{\alpha t^*-1}\sum_{u\in V}\phi_u^{(T-t)}P_{u,S}^t\right|\\ &\qquad+\left|\sum_{t=\alpha t^*}^{T-1}\sum_{u\in V}\phi_u^{(T-t)}\left(P_{u,S}^t-\pi_S\right)\right|\end{aligned}\tag{3.10}$$

for every positive integer α. Now, we give upper bounds of each term of (3.10). For the first term of (3.10), it is easy to see that

$$\begin{aligned}&\left|\sum_{t=0}^{\alpha t^*-1}\sum_{u\in V}\phi_u^{(T-t)}P_{u,S}^t\right|\\ &\quad\leq \sum_{t=0}^{\alpha t^*-1}|P_{u,S}^t|\sum_{u\in V}|\phi_u^{(T-t)}|\\ &\quad\leq \sum_{t=0}^{\alpha t^*-1}\sum_{u\in V}\delta^-(u)=m\alpha t^*\end{aligned}\tag{3.11}$$

holds by (3.7). To bound the second term of (3.10), we use the following lemma (See Appendix A for the proof).

LEMMA 3.1. *[27] Suppose $P\in\mathbb{R}_{\geq 0}^{n\times n}$ is* ergodic. *Then*

$$\sum_{t=\alpha t^*}^{\infty}\mathcal{D}_{\rm tv}\left(P_{u,\cdot}^t,\pi\right)\leq\frac{t^*}{2^\alpha}$$

holds for any $u\in V$ and for any $\alpha\in\mathbb{Z}_{>0}$. ■

By Lemma 3.1, we obtain that

$$\begin{aligned}&\left|\sum_{t=\alpha t^*}^{T-1}\sum_{u\in V}\phi_u^{(T-t)}\left(P_{u,S}^t-\pi_S\right)\right|\\ &\quad\leq \sum_{t=\alpha t^*}^{T-1}\sum_{u\in V}|\phi_u^{(T-t)}|\left|P_{u,S}^t-\pi_S\right|\\ &\quad\leq \frac{t^*}{2^\alpha}\sum_{u\in V}\max_{0\leq t\leq T}|\phi_u^{(T-t)}|\leq\frac{mt^*}{2^\alpha},\end{aligned}\tag{3.12}$$

where the last inequality follows from (3.7). Now, we obtain the claim from (3.10)–(3.12) by letting $\alpha=1$. □

3.3 Lower bound

We give the following lower bound for an oblivious model. This proposition implies that the factor t^* in total variation discrepancy for an oblivious model is negligible, in general.

PROPOSITION 3.1. *There exists an oblivious model such that*

$$\max_{S\subseteq V}\left|\chi_S^{(T)}-\mu_S^{(T)}\right|=\Omega(nt^*)$$

holds for any time T after mixing.

Proof. Let $V=\{0,\ldots,n-1\}$, and let a transition matrix P be defined by $P_{u,u}=(k-1)/k$ for any $u\in V$, and $P_{u,v}=1/k(n-1)$ for any $u,v\in V$ such that $u\neq v$, i.e., P denotes a simple random walk on K_n with a self-loop probability $(k-1)/k$ for any vertex. For this P, it is not difficult to check that

$$\tau(\varepsilon)\leq k\ln\varepsilon^{-1}\tag{3.13}$$

holds for any $0<\varepsilon<1$ (see Appendix A). Now, we give a corresponding oblivious deterministic random walk. Let us assume that the prescribed ordering for each $v\in V$ starts with v itself (recall the definition of an oblivious deterministic random walk in Section 3.1). Let

$$\chi_u^{(0)}=\begin{cases}k & (u\in A)\\ 0 & (u\in B),\end{cases}$$

where $A=\{0,\ldots,\lceil n/2\rceil-1\}$ and $B=\{\lceil n/2\rceil,\ldots,n-1\}$ ($M=k\lceil n/2\rceil$). Then, the initial configuration is stable, i.e., $\chi^{(t)}=\chi^{(0)}$, since each $v\in A$ serves $\lfloor k\cdot\frac{k-1}{k}\rfloor+1=k$ tokens to itself (notice that the "surplus" token stays at v according to the prescribed ordering). Thus it is easy to see that

$$\begin{aligned}\max_{S\subseteq V}|\chi_S^{(t)}-\mu_S^{(t)}| &\geq |\chi_A^{(t)}-\mu_A^{(t)}|\geq k\left\lceil\frac{n}{2}\right\rceil-\left(\frac{kn}{4}+\varepsilon\right)\\ &\geq \frac{kn}{4}-\varepsilon\geq\frac{n\tau(\varepsilon)}{4\ln\varepsilon^{-1}}-\varepsilon=\Omega(nt^*)\end{aligned}$$

holds for any $t\geq\tau(\varepsilon)$. We obtain the claim. □

Remark that our example on a complete graph, implies only $\Omega(\sqrt{m}t^*)$ lower bound. The gap between the upper and lower bounds remains as open.

4 Upper and lower bounds for non-oblivious model

Observation 3.1 for oblivious models only claims that $|Z_{v,u}^{(t)}-\chi_v^{(t)}P_{v,u}|\leq 1$ holds for any $t\in\mathbb{Z}_{\geq 0}$. In this section, we introduce the *SRT-router model* (cf, [27]), which satisfies $|\sum_{s=0}^{t}(Z_{v,u}^{(s)} - \chi_v^{(s)} P_{v,u})| < 1$ for any $t \in \mathbb{Z}_{\geq 0}$, and we obtain an improved bound when the Markov chain is *lazy*[6].

4.1 The SRT-router model

The SRT-router model, based on the *shortest remaining time* (SRT) rule [3, 30, 27], is a generalized version of the rotor-router model. In the model, we define an *SRT-router* $\sigma_v \colon \mathbb{Z}_{\geq 0} \to \mathcal{N}^+(v)$ on each $v \in V$ for a given P. Roughly speaking, $\sigma_v(i)$ denotes the destination of the $(i+1)$-st launched token at v. Given $\sigma_v(0), \ldots, \sigma_v(i-1)$, inductively $\sigma_v(i)$ is defined as follows. First, let

$$T_i(v) = \left\{u \in \mathcal{N}^+(v) \mid \tfrac{|\{j \in [0,i) | \sigma_v(j) = u\}|}{i+1} - P_{v,u} < 0\right\},$$

where $[z, z') \stackrel{\text{def.}}{=} \{z, z+1, \ldots, z'-1\}$ (and we remark $[z, z) = \emptyset$). Then, let $\sigma_v(i)$ be $u^* \in T_i(v)$ minimizing the value

$$\frac{|\{j \in [0,i) \mid \sigma_v(j) = u\}| + 1}{P_{v,u}}$$

over choices $u \in T_i(v)$. If there are two or more such $u \in T_i(v)$, then let u^* be the minimum in them in an arbitrary prescribed order. The ordering $\sigma_v(0), \sigma_v(1), \ldots$ is known as the *shortest remaining time* (SRT) rule (see e.g., [3, 30, 27]).

In an SRT-router model, there are $\chi_v^{(t)}$ tokens on a vertex v at time t, and each vertex v serves tokens on v to the neighboring vertices one by one according to $\sigma_v(i)$, like a rotor-router. For example, if there are a tokens on v at time $t = 0$, then $|\{j \in [0, a) \mid \sigma_v(j) = u\}|$ tokens move to each $u \in \mathcal{N}^+(v)$, and if there are b tokens on v at $t = 1$, then $|\{j \in [a, a+b) \mid \sigma_v(j) = u\}|$ tokens move to each $u \in \mathcal{N}^+(v)$, and so on.

Formally, the model is defined by

$$Z_{v,u}^{(t)} = \left|\left\{j \in \left[\textstyle\sum_{s=0}^{t-1}\chi_v^{(s)}, \sum_{s=0}^{t}\chi_v^{(s)}\right) \mid \sigma_v(j) = u\right\}\right|. \tag{4.14}$$

It is clear that the definition (4.14) satisfies (2.4). Then, the configuration of tokens is recursively defined by (2.5).

The following proposition is due to Angel et al. [3] and Tijdeman [30].

Proposition 4.1. [30, 3] *For any SRT-router model,*

$$\Big||\{j \in [0, z) \mid \sigma_v(j) = u\}| - z \cdot P_{v,u}\Big| < 1$$

holds for any $v, u \in V$ and for any integer $z > 0$.

[6] P is lazy if $P_{u,u} \geq 1/2$ holds for each $u \in V$.

Proposition 4.1 suggests that $|Z_{v,u}^{(t)} - \chi_v^{(t)} P_{v,u}|$ is small enough. In fact, Proposition 4.1 and (4.14) say a stronger fact that

$$\begin{aligned}
&\left|\sum_{t=a}^{b}\left(Z_{v,u}^{(t)} - \chi_v^{(t)}P_{v,u}\right)\right| \\
&= \left|\sum_{t=a}^{b}\left|\left\{j \in \left[\textstyle\sum_{s=0}^{t-1}\chi_v^{(s)}, \sum_{s=0}^{t}\chi_v^{(s)}\right) \mid \sigma_v(j) = u\right\}\right| - \sum_{t=a}^{b}\chi_v^{(t)}P_{v,u}\right| \\
&= \left|\left|\left\{j \in \left[\textstyle\sum_{s=0}^{a-1}\chi_v^{(s)}, \sum_{s=0}^{b}\chi_v^{(s)}\right) \mid \sigma_v(j) = u\right\}\right| - \sum_{t=a}^{b}\chi_v^{(t)}P_{v,u}\right| \\
&\leq \max_{\substack{z,z' \in \mathbb{Z}_{\geq 0} \\ \text{s.t. } z' > z}}\Big||\{j \in [z, z') \mid \sigma_v(j) = u\}| - (z'-z)P_{v,u}\Big| \\
&< 2
\end{aligned} \tag{4.15}$$

holds for any $a, b \in \mathbb{Z}_{\geq 0}$ s.t. $a \leq b$. We will use (4.15) in our analysis, in Section 4.2.

4.2 Better upper bound for the SRT-router model

Now, we show for any ergodic and lazy P the following theorem, modifying the technique in [6].

Theorem 4.1. *Suppose $P \in \mathbb{R}_{\geq 0}^{n \times n}$ is ergodic and lazy. Then for any SRT model,*

$$\begin{aligned}
\max_{S \subseteq V}\left|\chi_S^{(T)} - \mu_S^{(T)}\right| &\leq 2m\Big(24\sqrt{t^*\lceil\lg t^*\rceil + 1} - 9\Big) \\
&= \mathrm{O}\left(m\sqrt{t^* \log t^*}\right)
\end{aligned}$$

holds for any $T \in \mathbb{Z}_{\geq 0}$.

Proof. Let $\phi^{(t)} = \chi^{(t)} - \chi^{(t-1)}P$, for convenience. The major difference between an oblivious model and an SRT-router model is that

$$\begin{aligned}
\left|\sum_{t=a}^{b}\phi_u^{(t+1)}\right| &= \left|\sum_{t=a}^{b}\sum_{v \in \mathcal{N}^-(u)}(Z_{v,u}^{(t)} - \chi_v^{(t)}P_{v,u})\right| \\
&\leq \sum_{v \in \mathcal{N}^-(u)}\left|\sum_{t=a}^{b}(Z_{v,u}^{(t)} - \chi_v^{(t)}P_{v,u})\right| \\
&\leq 2\delta^-(u)
\end{aligned} \tag{4.16}$$

holds for any $u \in V$ and $b \geq a$ in an SRT-router model since (4.15) holds. By (3.9), we obtain that

$$\begin{aligned} (4.17)\quad & \left|\chi_S^{(T)} - \mu_S^{(T)}\right| \\ & \leq \left|\sum_{t=0}^{\alpha t^*+2}\sum_{u\in V}\phi_u^{(T-t)}P_{u,S}^t\right| \\ & \quad + \left|\sum_{t=\alpha t^*+3}^{T-1}\sum_{u\in V}\phi_u^{(T-t)}\left(P_{u,S}^t - \pi_S\right)\right| \end{aligned}$$

for any positive integer α. Now, we give a better upper bound of the first term of (4.17). As a preliminary, we remark the following proposition and lemma (see Appendix A for proofs).

Proposition 4.2. (summation by parts) *Let $F_t = \sum_{i=0}^{t} f_i$. Then,*

$$\sum_{t=0}^{T} f_t g_t = F_T g_T + \sum_{t=0}^{T-1} F_t(g_t - g_{t+1})$$

holds for any $T \in \mathbb{Z}_{\geq 0}$ and for any f_i, g_i $(0 \leq i \leq T)$. ■

Lemma 4.1. *Suppose that $P \in \mathbb{R}_{\geq 0}^{n\times n}$ is* ergodic *and* lazy. *Then,*

$$\sum_{t=0}^{T}\mathcal{D}_{\rm tv}\left(P_{u,\cdot}^t, P_{u,\cdot}^{t+1}\right) \leq 24\sqrt{T} - 11$$

holds for any $u \in V$ and for any $T \in \mathbb{Z}_{>0}$. ■

Using Proposition 4.2, (4.16), and Lemma 4.1, we obtain

$$\begin{aligned} (4.18)\quad & \left|\sum_{t=0}^{\alpha t^*+2}\phi_u^{(T-t)}P_{u,S}^t\right| \\ & = \left|\left(\sum_{i=0}^{\alpha t^*+2}\phi_u^{(T-i)}\right)P_{u,S}^{\alpha t^*+2} \right. \\ & \quad \left. + \sum_{t=0}^{\alpha t^*+1}\left(\sum_{i=0}^{t}\phi_u^{(T-i)}\right)\left(P_{u,S}^t - P_{u,S}^{t+1}\right)\right| \\ & \leq \left|\sum_{i=0}^{\alpha t^*+2}\phi_u^{(T-i)}\right|P_{u,S}^{\alpha t^*+2} \\ & \quad + \sum_{t=0}^{\alpha t^*+1}\left|\sum_{i=0}^{t}\phi_u^{(T-i)}\right|\left|P_{u,S}^t - P_{u,S}^{t+1}\right| \\ & \leq 2\delta^-(u) + 2\delta^-(u)\cdot\left(24\sqrt{\alpha t^*+1} - 11\right) \\ & = 2\delta^-(u)\left(24\sqrt{\alpha t^*+1} - 10\right) \end{aligned}$$

for any $u \in V$, where α is an arbitrary positive integer. Finally, (4.17), (4.18), (3.12) and (4.16) imply that

$$\begin{aligned} & \left|\chi_S^{(T)} - \mu_S^{(T)}\right| \\ & \leq \sum_{u\in V}\left|\sum_{t=0}^{\alpha t^*+2}\phi_u^{(T-t)}P_{u,S}^t\right| + \sum_{u\in V}\left|\sum_{t=\alpha t^*+3}^{T-1}\phi_u^{(T-t)}P_{u,S}^t\right| \\ & \leq 2m\left(24\sqrt{\alpha t^*+1} - 10\right) + \frac{t^*}{2^\alpha}\sum_{u\in V}\max_{0\leq t\leq T}|\phi_u^{(T-t)}| \\ & \leq 2m\left(24\sqrt{t^*\lceil\lg t^*\rceil + 1} - 10\right) + 2m\cdot\frac{t^*}{2^{\lceil\lg t^*\rceil}} \\ & \leq 2m\left(24\sqrt{t^*\lceil\lg t^*\rceil + 1} - 9\right) \end{aligned}$$

where the last inequality is obtained by letting $\alpha = \lceil\lg t^*\rceil$. We obtain the claim. □

4.3 Lower bounds

This section discusses a lower bound of the total variation discrepancy. First, we observe the following proposition, which is caused by the integral gap between $\chi^{(T)} \in \mathbb{Z}^V$ and $\mu^{(T)} \in \mathbb{R}^V$.

Proposition 4.3. *Suppose that P is ergodic and its stationary distribution is uniform. Then, for any $\chi^{(T)} \in \mathbb{Z}_{\geq 0}^n$ with an appropriate number of tokens M,*

$$\max_{S\subseteq V}\left|\chi_S^{(T)} - \mu_S^{(T)}\right| = \Omega(n)$$

holds for any time T after mixing.

Proof. Let $M = (k-(1/2))n$ be the number of tokens for an arbitrary positive integer k. Note that $\widetilde{\mu}_v^{(t)} = \mu_v^{(t)}/M$ converges to $1/n$ for any $v \in V$ since the stationary distribution is uniform. Precisely, for any $S \subseteq V$ and $T \geq \tau(1/(8k))$,

$$(4.19)\qquad \frac{|S|}{n} - \frac{1}{8k} \leq \sum_{v\in S}\widetilde{\mu}_v^{(T)} \leq \frac{|S|}{n} + \frac{1}{8k}$$

holds by the definition (2.2) of the mixing time $\tau(\varepsilon)$.

Let T be an arbitrary time, and let $S = \{v \in V \mid \chi_v^{(T)} \geq k\}$. First, we consider the case that $|S| \geq n/2$. Then, we see that $\sum_{v\in S}\chi_v^{(T)} \geq k|S|$ holds. At the same time

$$\begin{aligned} \sum_{v\in S}\mu_v^{(T)} &= \sum_{v\in S}M\widetilde{\mu}_v^{(T)} \leq \left(k - \frac{1}{2}\right)n\cdot\left(\frac{|S|}{n} + \frac{1}{8k}\right) \\ &\leq \left(k - \frac{1}{2}\right)|S| + \frac{n}{8} \end{aligned}$$

holds. Thus

$$\sum_{v\in S}\left(\chi_v^{(T)}-\mu_v^{(T)}\right) \geq k|S|-\left(\left(k-\frac{1}{2}\right)|S|+\frac{n}{8}\right) = \frac{1}{2}|S|-\frac{n}{8}\geq\frac{n}{8}$$

where the last inequality follows $|S|\geq n/2$. We obtain the claim in the case. Next, we consider the other case, meaning that $|S|<n/2$. Then we see that $\sum_{v\in\overline{S}}\chi_v^{(T)}\leq(k-1)|\overline{S}|$ since $\chi_v^{(T)}<k$ for any $v\in\overline{S}$. At that time,

$$\begin{aligned}\sum_{v\in\overline{S}}\mu_v^{(T)} &= \sum_{v\in\overline{S}}M\widetilde{\mu}_v^{(T)}\\ &\geq \left(k-\frac{1}{2}\right)n\cdot\left(\frac{|\overline{S}|}{n}-\frac{1}{8k}\right)\\ &\geq \left(k-\frac{1}{2}\right)|\overline{S}|-\frac{n}{8}\end{aligned}$$

holds. Thus

$$\begin{aligned}&\sum_{v\in\overline{S}}\left(\mu_v^{(T)}-\chi_v^{(T)}\right)\\ &\geq \left(\left(k-\frac{1}{2}\right)|\overline{S}|+\frac{n}{8}\right)-(k-1)|\overline{S}|\\ &= \frac{1}{2}|\overline{S}|-\frac{n}{8}\geq\frac{n}{8}\end{aligned}$$

where the last inequality follows $|\overline{S}|\geq n/2$. We obtain the claim. $\square$

We also give a better lower bound for a specific sequence of SRT-router models.

PROPOSITION 4.4. *There exist an example of SRT model such that*

$$\max_{S\subseteq V}\left|\chi_S^{(T)}-\mu_S^{(T)}\right|\geq\frac{n^2}{8}=\Omega(m)$$

holds for any $T>0$.

Proof. We consider a random walk on a complete graph $K_{2n'}$, i.e., let $V=\{0,1,\ldots,2n'-1\}$ ($n'\in\mathbb{Z}_{>0}$) and $P_{u,v}=1/(2n')$ for any $u,v\in V$. Let $A=\{0,1,\ldots,n'-1\}$, $B=\{n',n'+1\ldots,2n'-1\}$ and let

$$\chi_u^{(0)}=\begin{cases}(2k+1)n' & (u\in A)\\ 0 & (u\in B),\end{cases}$$

for an arbitrary $k\in\mathbb{Z}_{\geq 0}$. Note that $M=\|\chi^{(0)}\|_1=(2k+1)(n')^2$. Since this P mixes in a single step, $\mu_A^{(t)}=\mu_B^{(t)}=(2k+1)(n')^2/2$ holds for any $t>0$. We define the SRT-router $\sigma_u(i)$ as

$$\sigma_u(i \bmod 2n')=i$$

for any $u\in V$. Then, it is not difficult to check that $\chi_A^{(t)}=(k+1)(n')^2$ and $\chi_B^{(t)}=k(n')^2$ when t is odd, as well as that $\chi_A^{(t)}=k(n')^2$ and $\chi_B^{(t)}=(k+1)(n')^2$ when $t>0$ is even. Thus,

$$\max_{S\subseteq V}|\chi_S^{(t)}-\mu_S^{(t)}|\geq|\chi_A^{(t)}-\mu_A^{(t)}|=\frac{(n')^2}{2}=\frac{n^2}{8}$$

holds for any $t>0$. We obtain the claim. $\square$

4.4 Vertex-wise discrepancy

This section presents an upper bound on the *single vertex discrepancy* $\|\chi^{(T)}-\mu^{(T)}\|_\infty$, which is an extended version of [6] to ergodic, reversible and lazy Markov chains, in general.

THEOREM 4.2. *Suppose* $P\in\mathbb{R}_{\geq 0}^{n\times n}$ *is ergodic, reversible*[7]*, and lazy. Then for any SRT-router model,*

$$\begin{aligned}\left|\chi_w^{(T)}-\mu_w^{(T)}\right| &\leq \frac{2\pi_w}{\pi_{\min}}\Delta\left(48\sqrt{t^*\lceil\lg t^*\rceil+1}-19\right)\\ &= \mathrm{O}\left(\frac{\pi_{\max}}{\pi_{\min}}\Delta\sqrt{t^*\log t^*}\right)\end{aligned}$$

holds for any $w\in V$ *and for any* $T\in\mathbb{Z}_{\geq 0}$*, where* $\Delta=\max_{u\in V}|\mathcal{N}^+(u)|(=\max_{u\in V}|\mathcal{N}^-(u)|)$, $\pi_{\max}=\max_{u\in V}\pi_u$ *and* $\pi_{\min}=\min_{u\in V}\pi_u$.

Proof. By a combination of (4.17), (4.18), (3.12), and (4.16), we obtain that

$$\begin{aligned}\left|\chi_w^{(T)}-\mu_w^{(T)}\right| &\leq 2\Delta\sum_{u\in V}|P_{u,w}^{\alpha t^*+2}|\\ &\quad+2\Delta\sum_{t=0}^{\alpha t^*+2}\sum_{u\in V}\left|P_{u,w}^{t}-P_{u,w}^{t+1}\right|\\ &\quad+2\Delta\sum_{t=\alpha t^*+3}^{T-1}\sum_{u\in V}\left|P_{u,w}^{t}-\pi_w\right|\end{aligned}\tag{4.20}$$

holds, where α is an arbitrary positive integer. The condition that P is reversible, i.e., $\pi_uP_{u,w}^t=\pi_wP_{w,u}^t$ holds for any $u,v\in V$, implies that

$$\begin{aligned}\sum_{u\in V}P_{u,w}^t &= \sum_{u\in V}\frac{\pi_w}{\pi_u}P_{w,u}^t\\ &\leq \frac{\pi_w}{\pi_{\min}}\sum_{u\in V}P_{w,u}^t=\frac{\pi_w}{\pi_{\min}}\end{aligned}\tag{4.21}$$

[7] P is reversible if the *detailed balance equation* $\pi_vP_{v,u}=\pi_uP_{u,v}$ holds for any $u,v\in V$. Notice that a reversible ergodic P is symmetric if its stationary distribution is uniform, and vice versa.

holds. Lemma 4.1 implies that

$$\begin{aligned}(4.22)\qquad \sum_{t=0}^{\alpha t^*+1}\sum_{u\in V}\left|P^t_{u,w}-P^{t+1}_{u,w}\right| &= \sum_{t=0}^{\alpha t^*+1}\sum_{u\in V}\left|\frac{\pi_w}{\pi_u}\left(P^t_{w,u}-P^{t+1}_{w,u}\right)\right| \\ &\leq \frac{\pi_w}{\pi_{\min}}\sum_{t=0}^{\alpha t^*+1}\sum_{u\in V}\left|P^t_{w,u}-P^{t+1}_{w,u}\right| \\ &= \frac{\pi_w}{\pi_{\min}}\sum_{t=0}^{\alpha t^*+1}\|P^t_{w,u}-P^{t+1}_{w,u}\|_1 \\ &= \frac{2\pi_w}{\pi_{\min}}\sum_{t=0}^{\alpha t^*+1}\mathcal{D}_{\rm tv}\left(P^t_{w,\cdot},P^{t+1}_{w,\cdot}\right) \\ &\leq \frac{2\pi_w}{\pi_{\min}}\left(24\sqrt{\alpha t^*+1}-11\right)\end{aligned}$$

holds, as well as Lemma 3.1 implies that

$$\begin{aligned}(4.23)\quad \sum_{t=\alpha t^*+3}^{T-1}\sum_{u\in V}\left|P^t_{u,w}-\pi_w\right| &= \sum_{t=\alpha t^*+3}^{T-1}\sum_{u\in V}\left|\frac{\pi_w}{\pi_u}\left(P^t_{w,u}-\pi_u\right)\right| \\ &\leq \frac{\pi_w}{\pi_{\min}}\sum_{t=\alpha t^*+3}^{T-1}\sum_{u\in V}\left|P^t_{w,u}-\pi_u\right| \\ &= \frac{2\pi_w}{\pi_{\min}}\sum_{t=\alpha t^*+3}^{T-1}\mathcal{D}_{\rm tv}\left(P^t_{w,\cdot},\pi\right)\leq\frac{2\pi_w}{\pi_{\min}}\frac{t^*}{2^\alpha}\end{aligned}$$

holds. Thus, a combination (4.20)–(4.23) implies that

$$\begin{aligned}\left|\chi^{(T)}_w-\mu^{(T)}_w\right| &\leq 2\Delta\frac{\pi_w}{\pi_{\min}}+2\Delta\frac{2\pi_w}{\pi_{\min}}\left(24\sqrt{\alpha t^*+1}-11\right) \\ &\quad +2\Delta\frac{2\pi_w}{\pi_{\min}}\frac{t^*}{2^\alpha} \\ &\leq \frac{2\pi_w}{\pi_{\min}}\Delta\left(48\sqrt{t^*\lceil\lg t^*\rceil+1}-19\right)\end{aligned}$$

holds where the last inequality follows by letting $\alpha = \lceil\lg t^*\rceil$. We obtain the claim. □

5 Concluding Remarks

In this paper, we gave two upper bounds of the *total variation discrepancy*, one is $\|\chi^{(t)}-\mu^{(t)}\|_1 = \mathrm{O}(mt^*)$ for any ergodic Markov chains and the other is $\|\chi^{(t)}-\mu^{(t)}\|_1 = \mathrm{O}(m\sqrt{t^*\log t^*})$ for any lazy and ergodic Markov chains. We also showed some lower bounds. The gap between upper and lower bounds is a future work. Development of a deterministic approximation algorithm based on deterministic random walks for #P-hard problems is a challenge.

6 Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments, especially for the proof of Proposition 3.1. This work is supported by JSPS KAKENHI Grant Number 15J03840, 15K15938, 25700002, 15H02666.

References

[1] H. Akbari and P. Berenbrink, Parallel rotor walks on finite graphs and applications in discrete load balancing, Proc. SPAA 2013, 186–195.

[2] E. Ando and S. Kijima, An FPTAS for the volume computation of 0-1 knapsack polytopes based on approximate convolution integral, Lecture Notes in Computer Science, **8889** (ISAAC 2014), 376–386.

[3] O. Angel, A.E. Holroyd, J. Martin, and J. Propp, Discrete low discrepancy sequences, arXiv:0910.1077, 2009.

[4] E. Bampas, L. Gasieniec, N. Hanusse, D. Ilcinkas, R. Klasing, and A. Kosowski, Euler tour lock-in problem in the rotor-router model, Proc. DISC 2009, 423-435

[5] A. Bandyopadhyay and D. Gamarnik, Counting without sampling: asymptotics of the log-partition function for certain statistical physics models, Random Structures & Algorithms, **33** (2008), 452–479.

[6] P. Berenbrink, R. Klasing, A. Kosowski, F. Mallmann-Trenn, and P. Uznanski, Improved analysis of deterministic load-balancing schemes, Proc. PODC 2015, 301–310.

[7] J. Chalopin, S. Das, P. Gawrychowski, A. Kosowski, A. Labourel and P. Uznanski, Limit behavior of the multi-agent rotor-router system, Proc. DISC 2015, 123–139.

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

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

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

[11] B. Cousins, S. Vempala: Bypassing KLS: Gaussian cooling and an $\mathrm{O}^*(n^3)$ volume algorithm, Proc. STOC 2015, 539–548.

[12] D. Dereniowski, A. Kosowski, D. Pajak, and P. Uznanski, Bounds on the cover time of parallel rotor walks, LIPICS, **25** (STACS 2014), 263–275.

[13] B. Doerr and T. Friedrich, Deterministic random walks on the two-dimensional grid, Combinatorics, Probability and Computing, **18** (2009), 123–144.

[14] M. Dyer, A. Frieze, R. Kannan, A random polynomial-time algorithm for approximating the volume of convex bodies, Journal of the ACM, **38** (1991), 1–17.

[15] T. Friedrich, M. Gairing, and T. Sauerwald, Quasirandom load balancing, SIAM Journal on Computing, **41** (2012), 747–771.
[16] P. Gopalan, A. Klivans, and R. Meka, Polynomial-time approximation schemes for knapsack and related counting problems using branching programs, arXiv:1008.3187v1, 2010.
[17] P. Gopalan, A. Klivans, R. Meka, D. Stefankovic, S. Vempala, E. Vigoda, An FPTAS for #knapsack and related counting problems, Proc. of FOCS 2011, 817–826, 2011.
[18] 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.
[19] M. R. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, SIAM Journal on Computing, **22** (1993), 1087–1116.
[20] M. R. 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.
[21] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, Random generation of combinatorial structures from a uniform distribution. Theoretical Computer Science, **32** (1986), 169–188.
[22] H. Kajino, S. Kijima, and K. Makino, Discrepancy analysis of deterministic random walks on finite irreducible digraphs, discussion paper, 2014.
[23] S. Kijima, K. Koga, and K. Makino, Deterministic random walks on finite graphs, Random Structures & Algorithms, **46** (2015), 739–761.
[24] D. A. Levine, Y. Peres, and E. L. Wilmer, Markov Chain and Mixing Times, The American Mathematical Society, 2008.
[25] L. Lovasz and S. Vempala, Fast algorithms for logconcave functions: sampling, rounding, integration and optimization, Proc. FOCS 2006, 57–68.
[26] Y. Rabani, A. Sinclair, and R. Wanka, Local divergence of Markov chains and analysis of iterative load balancing schemes, Proc. FOCS 1998, 694–705.
[27] T. Shiraga, Y. Yamauchi, S. Kijima, and M. Yamashita, Deterministic random walks for rapidly mixing chains, arXiv:1311.3749, 2013.
[28] T. Shiraga, Y. Yamauchi, S. Kijima, and M. Yamashita, L_∞-discrepancy analysis of polynomial-time deterministic samplers emulating rapidly mixing chains, Lecture Notes in Computer Science, **8591** (COCOON 2014), 25–36.
[29] 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.
[30] R. Tijdeman, The chairman assignment problem, Discrete Mathematics. **32** (1980), 323–330.
[31] D. Weitz, Counting independent sets up to the tree threshold, Proc. STOC 2006, 140–149.
[32] V. Yanovski, I.A. Wagner, and A.M. Bruckstein, A distributed ant algorithm for efficiently patrolling a network, Algorithmica, **37** (2003), 165–186.

A Supplemental proofs

A.1 Proof of Lemma 3.1

For convenience, let $h(t) = \max_{u \in V} \mathcal{D}_{\rm tv}(P^t_{u,\cdot}, \pi)$. We use the following proposition to obtain Lemma 3.1.

PROPOSITION A.1. *[27] (Proposition 2.1) For any integers ℓ ($\ell \geq 1$) and k ($0 \leq k < t^*$),*

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

holds for any $u \in V$. ■

Proof. [Proof of Lemma 3.1] By Proposition A.1,

$$\begin{aligned}\sum_{t=\alpha t^*}^{\infty} \mathcal{D}_{\rm tv}(P^t_{u,\cdot}, \pi) &\leq \sum_{t=\alpha t^*}^{\infty} h(t) = \sum_{\ell=\alpha}^{\infty} \sum_{k=0}^{t^*-1} h(\ell t^* + k) \\ &\leq \sum_{\ell=\alpha}^{\infty} \sum_{k=0}^{t^*-1} \frac{1}{2^{\ell+1}} = t^* \cdot \frac{1/2^{\alpha+1}}{1-(1/2)} \\ &= \frac{t^*}{2^\alpha}\end{aligned}$$

holds. We obtain the claim. □

A.2 Supplemental proof of Proposition 3.1

We give a proof of (3.13) for Proposition 3.1.

PROPOSITION A.2. *Let*

$$P_{u,v} = \begin{cases} \frac{k-1}{k} & (\text{if } v = u) \\ \frac{1}{k(n-1)} & (\text{otherwise}). \end{cases}$$

Then

$$\tau(\varepsilon) \leq k \ln \varepsilon^{-1}.$$

Proof. For this P, it is not difficult to check

$$P^t_{u,v} = \begin{cases} \frac{1}{n} + \frac{n-1}{n}\left(1 - \frac{n}{k(n-1)}\right)^t & (\text{if } v = u) \\ \frac{1}{n} - \frac{1}{n}\left(1 - \frac{n}{k(n-1)}\right)^t & (\text{otherwise}) \end{cases}$$

holds for any $t \geq 0$. Thus,

$$\begin{aligned}(\text{A.1})\ \mathcal{D}_{\rm tv}(P^t_{u,\cdot}, \pi) &= \frac{1}{2}\left(|P^t_{u,u} - \pi_u| + \sum_{v \in V, v \neq u} |P^t_{u,v} - \pi_v|\right) \\ &= \frac{n-1}{n}\left(1 - \frac{n}{k(n-1)}\right)^t\end{aligned}$$

holds for any $t \geq 0$. By (A.1),

$$\begin{aligned}
\tau(\varepsilon) &= \left\lceil \frac{\ln \varepsilon^{-1} - \ln \frac{n}{n-1}}{\ln\left(1 - \frac{n}{k(n-1)}\right)^{-1}} \right\rceil \\
&\leq \ln \varepsilon^{-1} \cdot \frac{k(n-1)}{n} \\
&\leq k \ln \varepsilon^{-1}
\end{aligned}$$

holds, where we used the fact that $\log(1-x)^{-1} \geq x$ holds for any x $(0 < x < 1)$. We obtain the claim. □

A.3 Proof of Proposition 4.2

Proof. Let $F_t = \sum_{i=0}^{t} f_i$. Then, $f_t = F_t - F_{t-1}$ holds.

$$\begin{aligned}
\sum_{t=0}^{T} f_t g_t &= f_0 g_0 + \sum_{t=1}^{T} f_t g_t = f_0 g_0 + \sum_{t=1}^{T} (F_t - F_{t-1}) g_t \\
&= f_0 g_0 + \sum_{t=1}^{T} F_t g_t - \sum_{t=1}^{T} F_{t-1} g_t \\
&= \sum_{t=0}^{T} F_t g_t - \sum_{t=0}^{T-1} F_t g_{t+1} \\
&= F_T g_T + \sum_{t=0}^{T-1} F_t (g_t - g_{t+1}).
\end{aligned}$$

□

A.4 Proof of Lemma 4.1

To bound $\mathcal{D}_{\mathrm{tv}}\left(P_{u,\cdot}^{t}, P_{u,\cdot}^{t+1}\right)$, we use the following proposition.

Proposition A.3. *[24](p.69, Proposition 5.6) Suppose $P \in \mathbb{R}_{\geq 0}^{n \times n}$ is ergodic and lazy. Then*

$$\mathcal{D}_{\mathrm{tv}}(P_{u,\cdot}^{t}, P_{u,\cdot}^{t+1}) \leq \frac{12}{\sqrt{t}}$$

holds for any $u \in V$ and for any $t > 0$. ■

Proof. [Proof of Lemma 4.1] By Proposition A.3,

$$\begin{aligned}
\sum_{t=0}^{T} \mathcal{D}_{\mathrm{tv}}\left(P_{u,\cdot}^{t}, P_{u,\cdot}^{t+1}\right) &\leq 1 + \sum_{t=1}^{T} \mathcal{D}_{\mathrm{tv}}\left(P_{u,\cdot}^{t}, P_{u,\cdot}^{t+1}\right) \\
&\leq 1 + \sum_{t=1}^{T} \frac{12}{\sqrt{t}} \\
&\leq 1 + 12\left(2\sqrt{T} - 1\right) \\
&= 24\sqrt{T} - 11
\end{aligned}$$

holds, and we obtain the claim. We remark that we use the integral-comparison fact $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq 2\sqrt{T} - 1$. □