2005 International Conference on Analysis of Algorithms DMTCS proc. **AD**, 2005, 371–382

Rapidly mixing chain and perfect sampler for logarithmic separable concave distributions on simplex†

Shuji Kijima[1] and Tomomi Matsui[1]

[1]*Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan.*

In this paper, we are concerned with random sampling of an n dimensional integral point on an $(n-1)$ dimensional simplex according to a multivariate discrete distribution. We employ sampling via Markov chain and propose two "hit-and-run" chains, one is for approximate sampling and the other is for perfect sampling. We introduce an idea of *alternating inequalities* and show that a *logarithmic separable concave* function satisfies the alternating inequalities. If a probability function satisfies alternating inequalities, then our chain for approximate sampling mixes in $\mathrm{O}(n^2 \ln(K\varepsilon^{-1}))$, namely $(1/2)n(n-1)\ln(K\varepsilon^{-1})$, where K is the side length of the simplex and ε $(0 < \varepsilon < 1)$ is an error rate. On the same condition, we design another chain and a perfect sampler based on monotone CFTP (Coupling from the Past). We discuss a condition that the expected number of total transitions of the chain in the perfect sampler is bounded by $\mathrm{O}(n^3 \ln(Kn))$.

Keywords: Markov chain, Mixing time, Path coupling, Coupling from the past, Log-concave function.

1 Introduction

A sampling from a *log-concave* distribution has many interesting applications [2, 1, 12], and approximate samplers via Markov chain from log-concave distributions on a convex body have been studied [9, 10, 14]. Frieze and Kannan proved that a "ball walk" (chain) rapidly mixes, by using log-Sobolev inequalities [10]. Lovász and Vempala proved with considering the conductance that both of "ball walk" and "hit-and-run walk" (chain) mix in $\mathrm{O}(n^4 \mathrm{poly}(\ln n, \ln \varepsilon^{-1}))$ and that the amortized mixing times of them are bounded by $\mathrm{O}(n^3 \mathrm{poly}(\ln n, \ln \varepsilon^{-1}))$ [14].

Randall and Winkler discussed approximate uniform sampling via Markov chain on a simplex [19]. They treated both continuous and discrete state spaces, and showed that mixing time of a type of "hit-and-run" chain is $\Theta(n^3 \ln(n\varepsilon^{-1}))$. They proved the upper bound by using a technique of warm start and two phase coupling, and the lower bound by using a probabilistic second-moment argument.

We are concerned with random sampling an n dimensional integral point on a simplex from a *logarithmic separable concave* probability function. We show that a type of "hit-and-run" chain mixes in $\mathrm{O}(n^2 \ln(K\varepsilon^{-1}))$ where K is the side length of a simplex, with using path coupling technique [3]. There are several straightforward applications, just as restricted to a logarithmic separable concave probability function on a simplex. One is a computation of the normalizing constant of the product form solution for a closed queueing network [4, 11]. Another is MCMC based exact test for independence in medical statistics, in bioinformatics, and so on [16, 15].

We also show that another "hit-and-run" chain provides a perfect sampler based on monotone CFTP (Coupling from the Past). An ordinary sampler via Markov chain is an approximate sampler, whereas Propp and Wilson proposed monotone CFTP algorithm which realizes a perfect (exact) sampling from stationary distribution. One of the great advantages of perfect sampling is that we never need to determine the error rate ε when to use. Another is that a perfect sampler becomes faster than any approximate sampler based on a Markov chain when we need a sample according to highly accurate distribution. There are some other algorithms for perfect sampling via Markov chain, e.g., Wilson's read once algorithm [20] and Fill's interruptible algorithm [7, 8]. In this paper, we just review monotone CFTP algorithm, while we can employ others in a similar way.

†Supported by Superrobust Computation Project of the 21st Century COE Program "Information Science and Technology Strategic Core."

2 Preliminaries and Results

2.1 Logarithmic Separable Concave Function

We denote the set of real numbers (non-negative, positive real numbers) by $\mathbb{R}$ ($\mathbb{R}_+$, $\mathbb{R}_{++}$), and the set of integers (non-negative, positive integers) by $\mathbb{Z}$ ($\mathbb{Z}_+$, $\mathbb{Z}_{++}$), respectively. Let us consider the set of non-negative integral points in an $n-1$ dimensional simplex

$$\Xi \stackrel{\text{def.}}{=} \left\{x = (x_1, x_2, \ldots, x_n) \in \mathbb{Z}_+^n \mid \sum_{i=1}^n x_i = K\right\}.$$

Given a vector of single-variable positive functions $\boldsymbol{f} = (f_1, f_2, \ldots, f_n)$ where $f_i : \mathbb{Z}_+ \to \mathbb{R}_{++}$ for each i, we concern with sampling on Ξ according to distribution π defined by

$$\pi(x) = \frac{1}{C}\prod_{i=1}^n f_i(x_i) \quad \text{for } x = (x_1, x_2, \ldots, x_n) \in \Xi, \tag{1}$$

where $C \stackrel{\text{def.}}{=} \sum_{x\in\Xi}\prod_{i=1}^n f_i(x_i)$. A function f_i is called *log-concave* if and only if $\forall x \in \mathbb{Z}_+$, $\ln f_i(x+1) \geq (1/2)\left(\ln f_i(x) + \ln f_i(x+2)\right)$. The function π of (1) is said as *logarithmic separable concave* if each f_i is a log-concave function. In the following, we give two examples of practical applications in our setting that is logarithmic separable concave function on a simplex.

Example 1 ([11]) In queueing network theory, closed Jackson network is a basic and significant model. It is well known that a closed Jackson network has a product form solution. Given a vector of parameters $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}_{++}^n$ and a vector $(s_1, s_2, \ldots, s_n) \in \mathbb{Z}_+^n$ of number of servers on nodes, the product form solution π_{J} on Ξ is defined by

$$\pi_{\mathrm{J}}(x) \stackrel{\text{def.}}{=} \frac{1}{C_{\mathrm{J}}}\prod_{i=1}^n \frac{\alpha_i^{x_i}}{\min\{x_i, s_i\}!\,\min\{x_i, s_i\}^{x_i - \min\{x_i, s_i\}}} \text{ for } x \in \Xi,$$

where C_{J} is normalizing constant. No weakly (thus genuinly) polynomial-time algorithm is known for computing C_{J} and we can design a polynomial-time randomized approximation scheme by using a polynomial time sampler of π_{J}.

Example 2 ([16, 15]) Discretized Dirichlet Distribution often appears in statistical methods in the bioinformatics. Given a vector of parameters (Dirichlet parameters) $(u_1, u_2, \ldots, u_n) \in \mathrm{R}_+^n$, Discretized Dirichlet Distribution π_{D} on $\Xi_{++} \stackrel{\text{def.}}{=} \{x = (x_1, \ldots, x_n) \in \mathbb{Z}_{++} \mid \sum_{i=1}^n x_i = K\}$ $(K \geq n)$ is defined by

$$\pi_{\mathrm{D}}(x) \stackrel{\text{def.}}{=} \frac{\prod_{i=1}^n (x_i/K)^{u_i - 1}}{\sum_{x\in\Xi_{++}}\prod_{i=1}^n (x_i/K)^{u_i-1}} \text{ for } x \in \Xi_{++}.$$

When every parameter satisfies $u_i \geq 1$, $\pi_{\mathrm{D}}(x)$ is a logarithmic separable concave function. There is rejection sampling for (continuous) Dirichlet distribution and it is know that rejection sampling is inefficient when Dirichlet parameters are very small.

2.2 Main results

In this subsection, we describe two theorems, which are our main results.

Given a pair of probability distributions ν_1 and ν_2 on a finite state space Ω, the *total variation distance* between ν_1 and ν_2 is defined by $\mathrm{d_{TV}}(\nu_1, \nu_2) \stackrel{\text{def.}}{=} \frac{1}{2}\sum_{x\in\Omega}|\nu_1(x) - \nu_2(x)|$. The *mixing time* of an ergodic Markov chain is defined by $\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{x\in\Omega}\{\min\{t \mid \forall s \geq t,\ \mathrm{d_{TV}}(P_x^s, \pi) \leq \varepsilon\}\}$ $(0 < \varepsilon < 1)$ where π is the stationary distribution and P_x^s is the probability distribution of the chain at time period $s \geq 0$ with initial state x (at time period 0).

Theorem 2.1 *For any logarithmic separable distribution π on Ξ, there is an ergodic Markov chain whose stationary distribution is π and which has the mixing time $\tau(\varepsilon)$ $(0 < \varepsilon < 1)$ satisfying*

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2}\ln(K\varepsilon^{-1}).$$

We will give an actual Markov chain and show Theorem 2.1 in Section 3.

Suppose that we have an ergodic Markov chain with a finite state space Ω where the transition rule $X \mapsto X'$ can be described by a deterministic function $\phi : \Omega \times [0,1) \to \Omega$, called *update function*, and a random number Λ uniformly distributed over $[0,1)$ satisfying $X' = \phi(X, \Lambda)$. We introduce a partial order "$\succeq$" on the state space Ω. A transition rule expressed by a deterministic update function ϕ is called *monotone* (with respect to $(\Omega, \succeq)$) if $\forall \lambda \in [0,1)$, $\forall x, \forall y \in \Omega$, $x \succeq y \Rightarrow \phi(x,\lambda) \succeq \phi(y,\lambda)$. We also say that a chain is *monotone* if the chain is defined by a *monotone* update function. We say that a monotone chain has a unique pair of maximum and minimum when there exists a unique pair $(x_{\mathrm{U}}, x_{\mathrm{L}}) \in \Omega^2$ satisfying $\forall x \in \Omega$, $x_{\mathrm{U}} \succeq x \succeq x_{\mathrm{L}}$.

Theorem 2.2 *For any logarithmic separable distribution π on Ξ, there is an ergodic monotone Markov chain whose stationary distribution is π and which has a unique pair of maximum and minimum. Thus it directly gives a perfect sampler.*

In Section 4, we will propose another chain, briefly review the monotone CFTP algorithm, and prove Theorem 2.2.

3 Approximate Sampler

Now we propose a new Markov chain $\mathcal{M}_{\mathrm{A}}$ with state space Ξ. A transition of $\mathcal{M}_{\mathrm{A}}$ from a current state $X \in \Xi$ to a next state X' is defined as follows. First, we chose a pair of distinct indices $\{j_1, j_2\}$ uniformly at random. Next, put $k = X_{j_1} + X_{j_2}$, and chose $l \in \{0, 1, \ldots, k\}$ with probability

$$\frac{f_{j_1}(l) f_{j_2}(k-l)}{\sum_{s=0}^{k} f_{j_1}(s) f_{j_2}(k-s)} \qquad \left(\equiv \frac{f_{j_1}(l) f_{j_2}(k-l) \prod_{j \notin \{j_1, j_2\}} f_j(X_j)}{\sum_{s=0}^{k} f_{j_1}(s) f_{j_2}(k-s) \prod_{j \notin \{j_1, j_2\}} f_j(X_j)} \right)$$

then set

$$X'_i = \begin{cases} l & (\text{for } i = j_1), \\ k - l & (\text{for } i = j_2), \\ X_i & (\text{otherwise}). \end{cases}$$

Since $f_i(x)$ is a positive function, the Markov chain $\mathcal{M}_{\mathrm{A}}$ is irreducible and aperiodic, so ergodic, hence has a unique stationary distribution. Also, $\mathcal{M}_{\mathrm{A}}$ satisfies detailed balance equation, thus the stationary distribution is π defined by (1).

Let $g_{ij}^k : \mathbb{Z} \to \mathbb{R}_+$ be the cumulative distribution function defined by

$$g_{ij}^k(l) \stackrel{\text{def.}}{=} \frac{\sum_{s=0}^{l} f_i(s) f_j(k-s)}{\sum_{s=0}^{k} f_i(s) f_j(k-s)} \quad \text{for } l \in \{0, 1, \ldots, k\}.$$

We also define $g_{ij}^k(-1) \stackrel{\text{def.}}{=} 0$, for convenience. Then we can simulate the Markov chain $\mathcal{M}_{\mathrm{A}}$ with the function g_{ij}^k as follows. First, choose a pair $\{i, j\}$ of indices with the probability $2/(n(n-1))$. Next, put $k = X_i + X_j$, generate an uniformly random real number $\Lambda \in [0,1)$, choose l satisfying $g_{ij}^k(l-1) \leq \Lambda \leq g_{ij}^k(l)$, and set $X'_i = l$ and $X'_j = k - l$.

In the rest of this section, we estimate the mixing time of our chain $\mathcal{M}_{\mathrm{A}}$. We introduce *alternating inequalities* defined by

$$\frac{\sum_{s=0}^{l} f_i(s) f_j(k+1-s)}{\sum_{s=0}^{k+1} f_i(s) f_j(k+1-s)} \leq \frac{\sum_{s=0}^{l} f_i(s) f_j(k-s)}{\sum_{s=0}^{k} f_i(s) f_j(k-s)} \leq \frac{\sum_{s=0}^{l+1} f_i(s) f_j(k+1-s)}{\sum_{s=0}^{k+1} f_i(s) f_j(k+1-s)} \quad (l \in \{0, 1, \ldots, k\}), \quad (2)$$

where (i, j) is a pair of indices and k is an arbitrarily positive integer in $\{0, 1, \ldots, K\}$ (See Figure 1). The inequalities (2) are equivalent to that

$$g_{ij}^{k+1}(l) \leq g_{ij}^k(l) \leq g_{ij}^{k+1}(l+1) \quad (l \in \{0, \ldots, k\}). \tag{3}$$

We show that a logarithmic separable concave function satisfies alternating inequalities.

Lemma 3.1 *If f_i is a log-concave function for each $i \in \{1, 2, \ldots, n\}$, then alternating inequalities (2) hold for any $l \in \{0, 1, \ldots, k\}$, for any pair of distinct indices $\{i, j\}$ ($i, j \in \{1, 2, \ldots, n\}$) and for any $k \in \mathbb{Z}_+$.*

0	$f_i(0)f_j(k)/A$	$f_i(1)f_j(k-1)/A$	$\cdots$	$f_i(k)f_j(0)/A$	1

0	$f_i(0)f_j(k+1)/A'$	$f_i(1)f_j(k)/A'$	$f_i(2)f_j(k-1)/A'$	$\cdots$	$f_i(k+1)f_j(0)/A'$	1

Fig. 1: A figure of alternating inequalities for a pair of indices (i, j) and a non-negative integer k. In the figure, $A \stackrel{\text{def.}}{=} \sum_{s=0}^{k} f_i(s)f_j(k-s)$ and $A' \stackrel{\text{def.}}{=} \sum_{s=0}^{k+1} f_i(s)f_j(k+1-s)$ are normalizing constants.

Proof: When $k = 0$, it is obvious. First, we show that for any fixed $k \in \mathbb{Z}_{++}$, alternating inequalities (2) hold for any $l \in \{0, 1, \ldots, k\}$, if and only if

$$\begin{aligned}
&\left(\sum_{s=0}^{l} f_i(s)f_j(k+1-s)\right)\left(\sum_{s'=l+1}^{k} f_i(s')f_j(k-s')\right) \\
&\quad \leq \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s'=l+1}^{k+1} f_i(s')f_j(k+1-s')\right), \qquad (4)\\
\text{and}\quad &\left(\sum_{s=0}^{l} f_i(k+1-s)f_j(s)\right)\left(\sum_{s'=l+1}^{k} f_i(k-s')f_j(s')\right) \\
&\quad \leq \left(\sum_{s=0}^{l} f_i(k-s)f_j(s)\right)\left(\sum_{s'=l+1}^{k+1} f_i(k+1-s')f_j(s')\right), \qquad (5)
\end{aligned}$$

hold for any $l \in \{0, 1, \ldots, k-1\}$. We obtain (4) by transforming the former inequality of (2) for each $l \in \{0, 1, \ldots, k-1\}$ as follows,

$$\begin{aligned}
&\frac{\sum_{s=0}^{l} f_i(s)f_j(k+1-s)}{\sum_{s=0}^{k+1} f_i(s)f_j(k+1-s)} \leq \frac{\sum_{s=0}^{l} f_i(s)f_j(k-s)}{\sum_{s=0}^{k} f_i(s)f_j(k-s)} \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=0}^{k} f_i(s)f_j(k-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=0}^{k+1} f_i(s)f_j(k+1-s)\right) \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=0}^{l} f_i(s)f_j(k-s) + \sum_{s=l+1}^{k} f_i(s)f_j(k-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=0}^{l} f_i(s)f_j(k+1-s) + \sum_{s=l+1}^{k+1} f_i(s)f_j(k+1-s)\right) \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=l+1}^{k} f_i(s)f_j(k-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=l+1}^{k+1} f_i(s)f_j(k+1-s)\right).
\end{aligned}$$

In a similar way, we obtain (5) by transforming the latter inequality of (2) for each $l \in \{0, 1, \ldots, k-1\}$ as follows,

$$\begin{aligned}
&\frac{\sum_{s=0}^{l} f_i(s)f_j(k-s)}{\sum_{s=0}^{k} f_i(s)f_j(k-s)} \leq \frac{\sum_{s=0}^{l+1} f_i(s)f_j(k+1-s)}{\sum_{s=0}^{k+1} f_i(s)f_j(k+1-s)} \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=0}^{k+1} f_i(s)f_j(k+1-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l+1} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=0}^{k} f_i(s)f_j(k-s)\right) \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=0}^{l+1} f_i(s)f_j(k+1-s) + \sum_{s=l+2}^{k+1} f_i(s)f_j(k+1-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l+1} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=0}^{l} f_i(s)f_j(k-s) + \sum_{s=l+1}^{k} f_i(s)f_j(k-s)\right) \\
&\Leftrightarrow \left(\sum_{s=0}^{l} f_i(s)f_j(k-s)\right)\left(\sum_{s=l+2}^{k+1} f_i(s)f_j(k+1-s)\right) \\
&\qquad \leq \left(\sum_{s=0}^{l+1} f_i(s)f_j(k+1-s)\right)\left(\sum_{s=l+1}^{k} f_i(s)f_j(k-s)\right) \\
&\Leftrightarrow \left(\sum_{s'=k-l}^{k} f_i(k-s')f_j(s')\right)\left(\sum_{s''=0}^{k-l-1} f_i(k+1-s'')f_j(s'')\right) \\
&\qquad \leq \left(\sum_{s''=k-l}^{k+1} f_i(k+1-s'')f_j(s'')\right)\left(\sum_{s'=0}^{k-l-1} f_i(k-s')f_j(s')\right) \\
&\Leftrightarrow \left(\sum_{s'=l'+1}^{k} f_i(k-s')f_j(s')\right)\left(\sum_{s''=0}^{l'} f_i(k+1-s'')f_j(s'')\right)
\end{aligned}$$

$$\leq \left(\sum_{s''=l'+1}^{k+1} f_i(k+1-s'')f_j(s'')\right)\left(\sum_{s'=0}^{l'} f_i(k-s')f_j(s')\right),$$

where $l' = k - l - 1 \in \{0, 1, \dots, k-1\}$.

Next, we show the inequality (4) hold for any $k \in \mathbb{Z}_{++}$. With considering the expansion, it is enough to show that

$$f_i(s)f_j(k+1-s)f_i(s')f_j(k-s') \leq f_i(s)f_j(k-s)f_i(s')f_j(k+1-s') \tag{6}$$

$\forall s, \forall s' \in \{0, 1, \dots, k\}$ satisfying $0 \leq s < s' \leq k$. From the hypothesis of Lemma, f_j is a log-concave function for any index $j \in \{1, 2, \dots, n\}$, thus with considering $(k - s') < (k - s' + 1) \leq (k - s) < (k - s + 1)$, the inequality

$$\ln f_j(k-s') + \ln f_j(k-s+1) \leq \ln f_j(k-s'+1) + \ln f_j(k-s)$$

holds. Since $f_i(x) > 0$ for any $x \in \mathbb{Z}_+$, the inequality (6) hold $\forall s, \forall s' \in \{0, 1, \dots, k\}$ satisfying $0 \leq s < s' \leq k$. For the inequality (5), we obtain the claim in the same way as (4) by interchanging i and j. □

Theorem 3.2 *If $\boldsymbol{f} = (f_1, f_2, \dots, f_n)$ satisfies alternating inequalities for any pair of distinct indices $\{i, j\}$ $(i, j \in \{1, 2, \dots, n\})$ and for any $k \in \{0, 1, \dots, K\}$, then Markov chain $\mathcal{M}_{\mathrm{A}}$ mixes rapidly and the mixing time $\tau(\varepsilon)$ for $0 < \varepsilon < 1$ satisfies*

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(K\varepsilon^{-1}).$$

Proof: Let $G = (\Xi, \mathcal{E})$ be an undirected simple graph with vertex set Ξ and edge set $\mathcal{E}$ defined as follows. A pair of vertices $\{x, y\}$ is an edge of G if and only if $(1/2)\sum_{i=1}^{n} |x_i - y_i| = 1$. Clearly, the graph G is connected. We define the length $l_{\mathrm{A}}(e)$ of an edge $e \in \mathcal{E}$ by $l_{\mathrm{A}}(e) \overset{\text{def.}}{=} 1$. For each pair $(x, y) \in \Xi^2$, we define the distance $d_{\mathrm{A}}(x, y)$ be the length of the shortest path between x and y on G with l_{A}. Clearly, the diameter of G, defined by $\max_{x,y\in\Xi}\{d_{\mathrm{A}}(x, y)\}$, is bounded by K.

We define a joint process $(X, Y) \mapsto (X', Y')$ for any pair $\{X, Y\} \in \mathcal{E}$. Pick a pair of distinct indices $\{i_1, i_2\}$ uniformly at random. Then put $k_X = X_{i_1} + X_{i_2}$ and $k_Y = Y_{i_1} + Y_{i_2}$, generate a uniform random number $\Lambda \in [0, 1)$, chose $l_X \in \{0, 1, \dots, k_X\}$ and $l_Y \in \{0, 1, \dots, k_Y\}$ which satisfy $g_{i_1 i_2}^{k_X}(l_X - 1) \leq \Lambda < g_{i_1 i_2}^{k_X}(l_X)$ and $g_{i_1 i_2}^{k_Y}(l_Y - 1) \leq \Lambda < g_{i_1 i_2}^{k_Y}(l_Y)$, and set $X'_{i_1} = l_X$, $X'_{i_2} = k_X - l_X$, $Y'_{i_1} = l_Y$ and $Y'_{i_2} = k_Y - l_Y$.

Now we show that

$$\mathrm{E}[d_{\mathrm{A}}(X', Y')] \leq \beta d_{\mathrm{A}}(X, Y), \text{ and } \beta = 1 - \frac{2}{n(n-1)}, \text{ for any pair } \{X, Y\} \in \mathcal{E}.$$

Suppose that $X, Y \in \mathcal{E}$ satisfies $|X_j - Y_j| = 1$ for $j \in \{j_1, j_2\}$, and $|X_j - X_j| = 0$ for $j \notin \{j_1, j_2\}$.

Case 1: When the neither of indices j_1 nor j_2 are chosen, i.e., $\{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset$. Put $k = X_{i_1} + X_{i_2}$. It is easy to see that $\Pr[X'_{i_1} = l] = \Pr[Y'_{i_1} = l]$ for any $l \in \{0, \dots, k\}$ since $Y_{i_1} + Y_{i_2} = k$. We set $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$. Then $d_{\mathrm{A}}(X', Y') = d_{\mathrm{A}}(X, Y)$ holds.

Case 2: When the both of indices j_1 and j_2 are chosen, i.e., $\{i_1, i_2\} = \{j_1, j_2\}$. In the same way as Case 1, we can set $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$. Then $d_{\mathrm{A}}(X', Y') = 0$ holds.

Case 3: When exactly one of j_1 and j_2 is chosen, i.e., $|\{i_1, i_2\} \cap \{j_1, j_2\}| = 1$. Without loss of generality, we can assume that $i_1 = j_1$ and that $X_{i_1} + 1 = Y_{i_1}$. Put $k = X_{i_1} + X_{i_2}$, then $Y_{i_1} + Y_{i_2} = k + 1$ obviously. Now, we consider the joint process as a random number $\Lambda \in [0, 1)$ is given. Assume that $l \in \{0, 1, \dots, k\}$ satisfies $g_{i_1 i_2}^{k}(l - 1) \leq \Lambda < g_{i_1 i_2}^{k}(l)$, then alternating inequalities (3) imply that $g_{i_1 i_2}^{k+1}(l - 1) \leq \Lambda < g_{i_1 i_2}^{k+1}(l + 1)$. Therefore, if $X'_{i_1} = l$ then Y'_{i_1} should be in $\{l, l+1\}$ by the joint process. Thus we always obtain that $[X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} + 1 = Y'_{i_2}]$ or $[X'_{i_1} + 1 = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}]$. Hence $d_{\mathrm{A}}(X', Y') = d_{\mathrm{A}}(X, Y)$ holds.

With considering that Case 2 occurs with probability $2/(n(n-1))$, we obtain that

$$\mathrm{E}[d_{\mathrm{A}}(X', Y')] \leq \left(1 - \frac{2}{n(n-1)}\right) d_{\mathrm{A}}(X, Y).$$

Since the diameter of G is bounded by K, Theorem 3.3 (Path Coupling Theorem), which we describe following this proof, implies that the mixing time $\tau(\varepsilon)$ satisfies

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(K\varepsilon^{-1}).$$

□

The following Path Coupling Theorem proposed by Bubbly and Dyer is a useful technique for bounding the mixing time.

Theorem 3.3 (Path Coupling [3]) *Let $\mathcal{M}$ be a finite ergodic Markov chain with a finite state space Ω. Let $H = (\Omega, \mathcal{E})$ be a connected undirected graph with vertex set Ω and edge set $\mathcal{E} \subseteq \binom{\Omega}{2}$. Let $l : \mathcal{E} \to \mathbb{R}_{++}$ be a positive length defined on the edge set. For any pair of vertices $\{x, y\}$ of H, the distance between x and y, denoted by $d(x, y)$ and/or d(y,x), is the length of a shortest path between x and y, where the length of a path is the sum of the lengths of edges in the path. Suppose that there exists a joint process $(X, Y) \mapsto (X', Y')$ with respect to $\mathcal{M}$ whose marginals are faithful copies of $\mathcal{M}$ and*

$$\exists \beta,\ 0 < \beta < 1,\ \forall \{X, Y\} \in \mathcal{E},\ \mathrm{E}\left[d(X', Y')\right] \leq \beta \cdot d(X, Y).$$

Then the mixing time $\tau(\varepsilon)$ of Markov chain $\mathcal{M}$ satisfies $\tau(\varepsilon) \leq (1-\beta)^{-1} \ln(\varepsilon^{-1} D/d)$, where $d \stackrel{\text{def.}}{=} \min\{d(x, y) \mid \forall x, y \in \Omega\}$ and $D \stackrel{\text{def.}}{=} \max\{d(x, y) \mid \forall x, \forall y \in \Omega\}$.

The above theorem differs from the original theorem in [3] since the integrality of the edge length is not assumed. We drop the integrality and introduced the minimum distance d. This modification is not essential and we can show Theorem 3.3 similarly. We do not require this modification in the proof of Theorem 3.2, but we will use the modified version in the proof of Theorem 4.3 described later.

Outline of proof: Let π be the stationary distribution of $\mathcal{M}$, and let P_x^t for $x \in \Omega$ and $t \in \mathbb{Z}_{++}$ be the distribution of X^t which is a random variable of $\mathcal{M}$ at time t with initial state x. Let $(Y^0, Y^1, \ldots,)$ be a random process of $\mathcal{M}$ where Y^0 is a random variable according to π. Clearly, Y^t is according to π for any t. Let A^* be a subset of Ω satisfying

$$\sum_{s \in A^*} \left(P_x^t(s) - \pi(s)\right) = \max_{A \subseteq \Omega} \left\{ \sum_{s \in A} \left(P_x^t(s) - \pi(s)\right) \right\}.$$

Then we can show that

$$\begin{aligned}
\mathrm{d_{TV}}\left(P_x^t, \pi\right) &= \max_{A \subseteq \Omega} \left\{ \sum_{s \in A} \left(P_x^t(s) - \pi(s)\right) \right\} = \sum_{s \in A^*} \left(\Pr\left[X^t = s\right] - \Pr\left[Y^t = s\right]\right) \\
&\leq \sum_{s \in A^*} \Pr\left[X^t = s, Y^t \neq s\right] \leq \sum_{s \in \Omega} \Pr\left[X^t = s, Y^t \neq s\right] = \Pr\left[X^t \neq Y^t\right] = \sum_{x \neq y} \Pr\left[X^t = x, Y^t = y\right] \\
&\leq \sum_{x,y} \frac{d(x,y)}{d} \Pr\left[X^t = x, Y^t = y\right] = \frac{\mathrm{E}\left[d(X^t, Y^t)\right]}{d} \qquad (7)
\end{aligned}$$

hold. For any distinct pair $X^t, Y^t \in \Omega$ at any time t, there exists a sequence of elements $Z_1, Z_2, \ldots, Z_r$ ($Z_i \in \Omega$), corresponding to a shortest path between X and Y, such as $Z_1 = X$, $Z_r = Y$ and $d(X, Y) = \sum_{i=1}^{r-1} d(Z_i, Z_{i+1})$. Let Z_i' ($i \in \{1, 2, \ldots, r\}$) denote the next state of Z_i after a transition, then

$$d(X^{t+1}, Y^{t+1}) \leq \textstyle\sum_{i=1}^{r-1} d(Z_i', Z_{i+1}')$$

holds for any transition, since the distance between states is defined by the length of a shortest path. From the hypothesis of Theorem,

$$\begin{aligned}
\mathrm{E}\left[d(X^{t+1}, Y^{t+1})\right] &\leq \mathrm{E}\left[\textstyle\sum_{i=1}^{r-1} d(Z_i', Z_{i+1}')\right] = \textstyle\sum_{i=1}^{r-1} \mathrm{E}\left[d(Z_i', Z_{i+1}')\right] \\
&\leq \beta \cdot d(Z_i, Z_{i+1}) = \beta \cdot d(X^t, Y^t) \qquad (8)
\end{aligned}$$

hold. If we set $t = -(1/\ln\beta)\cdot\ln(D/d\varepsilon)$, then t satisfies that $t \le (1-\beta)^{-1}\ln(D/d\varepsilon) = \tau(\varepsilon)$. Thus recursively applying the result of (7) and (8), we can show that

$$\mathrm{d_{TV}}(P_x^t,\pi) \le \frac{\mathrm{E}[d(X^t,Y^t)]}{d} \le \beta\cdot\frac{d(X^{t-1},Y^{t-1})}{d} \le \beta^t\cdot\frac{d(X^0,Y^0)}{d} \le \mathrm{e}^{\ln\frac{D}{d\varepsilon}}\cdot\frac{D}{d} = \varepsilon$$

hold without respect to the initial state x. □

With the above discussions, we obtain Theorem 2.1.

4 Perfect Sampler

4.1 Monotone CFTP

Suppose that we have an ergodic monotone Markov chain $\mathcal{M}$ with partially ordered finite state space $(\Omega,\succeq)$. The transition $X \mapsto X'$ of $\mathcal{M}$ is defined by a monotone update function $\phi : \Omega\times[0,1)\to\Omega$, i.e., ϕ satisfies that $\forall\lambda\in[0,1), \forall x,\forall y\in\Omega,\ x\succeq y \Rightarrow \phi(x,\lambda)\succeq\phi(y,\lambda)$. Also suppose that there exists a unique pair of states $(x_{\mathrm{U}},x_{\mathrm{L}})$ in partially ordered set $(\Omega,\succeq)$, satisfying $x_{\mathrm{U}}\succeq x\succeq x_{\mathrm{L}}$, $\forall x\in\Omega$. The result of transitions of the chain from the time t_1 to t_2 ($t_1 < t_2$) with a sequence of random numbers $\boldsymbol{\lambda} = (\lambda[t_1],\lambda[t_1+1],\ldots,\lambda[t_2-1])\in[0,1)^{t_2-t_1}$ is denoted by $\Phi_{t_1}^{t_2}(x,\boldsymbol{\lambda}) : \Omega\times[0,1)^{t_2-t_1}\to\Omega$ where $\Phi_{t_1}^{t_2}(x,\boldsymbol{\lambda}) \stackrel{\text{def.}}{=} \phi(\phi(\cdots(\phi(x,\lambda[t_1]),\ldots,\lambda[t_2-2]),\lambda[t_2-1])$. Then, a standard monotone Coupling From The Past algorithm is expressed as follows.

Algorithm 1 (Monotone CFTP Algorithm [18])

Step 1 Set the starting time period $T := -1$ to go back, and set $\boldsymbol{\lambda}$ be the empty sequence.

Step 2 Generate random real numbers $\lambda[T],\lambda[T+1],\ldots,\lambda[\lceil T/2\rceil-1]\in[0,1)$, and insert them to the head of $\boldsymbol{\lambda}$ in order, i.e., put $\boldsymbol{\lambda} := (\lambda[T],\lambda[T+1],\ldots,\lambda[-1])$.

Step 3 Start two chains from x_{U} and x_{L} at time period T, and run each chain to time period 0 according to the update function ϕ with the sequence of numbers in $\boldsymbol{\lambda}$. (Here we note that both chains use the common sequence λ.)

Step 4 [Coalescence check]

(a) If $\Phi_T^0(x_{\mathrm{U}},\boldsymbol{\lambda}) = \Phi_T^0(x_{\mathrm{L}},\boldsymbol{\lambda})$, then return the common state $\Phi_T^0(x_{\mathrm{U}},\boldsymbol{\lambda}) = \Phi_T^0(x_{\mathrm{L}},\boldsymbol{\lambda})$ and stop.

(b) Else, update the starting time period $T := 2T$, and go to Step 2.

Theorem 4.1 (Monotone CFTP [18]) *Suppose that a Markov chain defined by an update function ϕ is monotone with respect to a partially ordered set of states $(\Omega,\succeq)$, and $\exists(x_{\mathrm{U}},x_{\mathrm{L}})\in\Omega^2$, $\forall x\in\Omega$, $x_{\mathrm{U}}\succeq x\succeq x_{\mathrm{L}}$. Then the monotone CFTP algorithm (Algorithm 1) terminates with probability 1, and obtained value is a realization of a random variable exactly distributed according to the stationary distribution.* □

4.2 Monotone Markov Chain

We propose another new Markov chain $\mathcal{M}_{\mathrm{P}}$. The transition rule of $\mathcal{M}_{\mathrm{P}}$ is defined by the following update function $\phi : \Xi\times[1,n)\to\Xi$. For a current state $X\in\Xi$, the next state $X' = \phi(X,\lambda)\in\Xi$ with respect to a random number $\lambda\in[1,n)$ is defined by

$$X'_i = \begin{cases} l & (\text{for } i = \lfloor\lambda\rfloor), \\ k-l & (\text{for } i = \lfloor\lambda\rfloor+1), \\ X_i & (\text{otherwise}), \end{cases}$$

where $k = X_{\lfloor\lambda\rfloor} + X_{\lfloor\lambda\rfloor+1}$ and $l\in\{0,1,\ldots,k\}$ satisfies

$$g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l-1) < \lambda - \lfloor\lambda\rfloor \le g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l).$$

Our chain $\mathcal{M}_{\mathrm{P}}$ is a modification of $\mathcal{M}_{\mathrm{A}}$, obtained by restricting to choose only a consecutive pair of indices. Clearly, $\mathcal{M}_{\mathrm{P}}$ is ergodic. The chain has a unique stationary distribution π defined by (1).

In the following, we show the monotonicity of $\mathcal{M}_{\mathrm{P}}$. Here we introduce a partial order "$\succeq$" on Ξ. For any state $x\in\Xi$, we introduce *cumulative sum vector* $c_x = (c_x(0),c_x(1),\ldots,c_x(n))\in\mathbb{Z}_+^{n+1}$ defined by

$$c_x(i) \stackrel{\text{def.}}{=} \begin{cases} 0 & (\text{for } i = 0), \\ \sum_{j=1}^{i} x_j & (\text{for } i\in\{1,2,\ldots,n\}). \end{cases}$$

For any pair of states $x, y \in \Xi$, we say $x \succeq y$ if and only if $c_x \geq c_y$. Next, we define two special states $x_{\mathrm{U}}, x_{\mathrm{L}} \in \Xi$ by $x_{\mathrm{U}} \stackrel{\text{def.}}{=} (K, 0, \cdots, 0)$ and $x_{\mathrm{L}} \stackrel{\text{def.}}{=} (0, \ldots, 0, K)$. Then we can see easily that $\forall x \in \Xi$, $x_{\mathrm{U}} \succeq x \succeq x_{\mathrm{L}}$.

Theorem 4.2 *If $\boldsymbol{f}$ satisfies alternating inequalities (2) for any consecutive pair of indices $(j, j+1)$ ($j \in \{1, \ldots, n-1\}$) and for any $k \in \{0, 1, \ldots, K\}$, then Markov chain $\mathcal{M}_{\mathrm{P}}$ is monotone on the partially ordered set $(\Xi, \succeq)$, i.e., $\forall \lambda \in [1, n)$, $\forall X, \forall Y \in \Omega$, $X \succeq Y \Rightarrow \phi(X, \lambda) \succeq \phi(Y, \lambda)$.*

Proof: We say that a state $X \in \Omega$ *covers* $Y \in \Omega$ (at j), denoted by $X \cdot\!\!\succ Y$ (or $X \cdot\!\!\succ_j Y$), when

$$X_i - Y_i = \begin{cases} +1 & (\text{for } i = j), \\ -1 & (\text{for } i = j+1), \\ 0 & (\text{otherwise}). \end{cases}$$

For any pair of states X, Y satisfying $X \succeq Y$, it is easy to see that there exists a sequence of states $Z_1, Z_2, \ldots, Z_r$ with appropriate length satisfying $X = Z_1 \cdot\!\!\succ Z_2 \cdot\!\!\succ \cdots \cdot\!\!\succ Z_r = Y$. Then it is enough to show that if a pair of states $X, Y \in \Omega$ satisfies $X \cdot\!\!\succ Y$, then $\forall \lambda \in [1, n)$, $\phi(X, \lambda) \succeq \phi(Y, \lambda)$, to obtain that $\phi(X, \lambda) = \phi(Z_1, \lambda) \succeq \phi(Z_2, \lambda) \succeq \cdots \succeq \phi(Z_r, \lambda) = \phi(Y, \lambda)$.

In the following, we show that if a pair of states $X, Y \in \Omega$ satisfies $X \cdot\!\!\succ_j Y$, then $\forall \lambda \in [1, n)$, $\phi(X, \lambda) \succeq \phi(Y, \lambda)$. We denote $\phi(X, \lambda)$ by X' and $\phi(Y, \lambda)$ by Y' for simplicity. For any index $i \neq \lfloor \lambda \rfloor$, it is easy to see that $c_X(i) = c_{X'}(i)$ and $c_Y(i) = c_{Y'}(i)$, and so $c_{X'}(i) - c_{Y'}(i) = c_X(i) - c_Y(i) \geq 0$ since $X \succeq Y$. In the following, we show that $c_{X'}(\lfloor \lambda \rfloor) \geq c_{Y'}(\lfloor \lambda \rfloor)$.

Case 1: If $\lfloor \lambda \rfloor \neq j-1$ and $\lfloor \lambda \rfloor \neq j+1$. Put $k = X_{\lfloor \lambda \rfloor} + X_{\lfloor \lambda \rfloor + 1}$, then it is easy to see that $Y_{\lfloor \lambda \rfloor} + Y_{\lfloor \lambda \rfloor + 1} = k$. Accordingly $X'_{\lfloor \lambda \rfloor} = Y'_{\lfloor \lambda \rfloor} = l$ where l satisfies

$$g^k_{\lfloor \lambda \rfloor (\lfloor \lambda \rfloor + 1)}(l-1) \leq \lambda - \lfloor \lambda \rfloor < g^k_{\lfloor \lambda \rfloor (\lfloor \lambda \rfloor + 1)}(l),$$

hence $c_{X'}(\lfloor \lambda \rfloor) = c_{Y'}(\lfloor \lambda \rfloor)$ holds.

Case 2: Consider the case that $\lfloor \lambda \rfloor = j-1$. Put $k+1 = X_{j-1} + X_j$, then $Y_{j-1} + Y_j = k$, since $X \cdot\!\!\succ_j Y$. From the definition of cumulative sum vector,

$$\begin{aligned} & c_{X'}(\lfloor \lambda \rfloor) - c_{Y'}(\lfloor \lambda \rfloor) = c_{X'}(j-1) - c_{Y'}(j-1) \\ = \;& c_{X'}(j-2) + X'_{j-1} - c_{Y'}(j-2) - Y'_{j-1} = c_X(j-2) + X'_{j-1} - c_Y(j-2) - Y'_{j-1} \\ = \;& X'_{j-1} - Y'_{j-1}. \end{aligned}$$

Thus, it is enough to show that $X'_{j-1} \geq Y'_{j-1}$. Now suppose that $l \in \{0, 1, \ldots, k\}$ satisfies $g^k_{(j-1)j}(l-1) \leq \lambda - \lfloor \lambda \rfloor < g^k_{(j-1)j}(l)$ for λ. Then $g^{k+1}_{(j-1)j}(l-1) \leq \lambda - \lfloor \lambda \rfloor < g^{k+1}_{(j-1)j}(l+1)$ holds, since alternating inequalities (3) imply that $g^{k+1}_{(j-1)j}(l-1) \leq g^k_{(j-1)j}(l-1) < g^{k+1}_{(j-1)j}(l) \leq g^{k+1}_{(j-1)j}(l+1)$. Thus we have that if $Y'_{j-1} = l$ then $X'_{j-1} = l$ or $l+1$. In other words,

$$\begin{pmatrix} X'_{j-1} \\ Y'_{j-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} k \\ k \end{pmatrix}, \begin{pmatrix} k+1 \\ k \end{pmatrix} \right\}$$

and $X'_{j-1} \geq Y'_{j-1}$ hold. Accordingly, we have that $c_{X'}(\lfloor \lambda \rfloor) \geq c_{Y'}(\lfloor \lambda \rfloor)$.

Case 3: Consider the case that $\lfloor \lambda \rfloor = j+1$. We can show $c_{X'}(\lfloor \lambda \rfloor) \geq c_{Y'}(\lfloor \lambda \rfloor)$ in a similar way to Case 2. □

Since $\mathcal{M}_{\mathrm{P}}$ is a monotone chain, we can design a perfect sampler based on monotone CFTP [18], which we briefly introduced in the previous subsection.

With the above discussions, we obtain Theorem 2.2.

4.3 Expected Running Time

Here, we discuss on an expected running time of our perfect sampler. First, we introduce the following condition.

Condition 1 *Indices of $\boldsymbol{f} = (f_1, f_2, \ldots, f_n)$ is arranged to satisfy that*

$$\sum_{l=0}^{k} \left(\frac{\sum_{s=0}^{l} f_i(s) f_{i+1}(k-s)}{\sum_{s=0}^{k} f_i(s) f_{i+1}(k-s)} - \frac{\sum_{s=0}^{l} f_i(s) f_{i+1}(k+1-s)}{\sum_{s=0}^{k+1} f_i(s) f_{i+1}(k+1-s)} \right) \equiv \sum_{l=0}^{k} \left(g^k_{i(i+1)}(l) - g^{k+1}_{i(i+1)}(l) \right) \geq \frac{1}{2}$$

for any $k \in \{0, 1, \ldots, K\}$.

Our goal in this subsection is to show the following theorem.

Theorem 4.3 *If $\boldsymbol{f}$ satisfies alternating inequalities (2) and Condition 1, then the expected number of whole transitions required in the monotone CFTP algorithm is upper bounded by* $\mathrm{O}(n^3 \ln(Kn))$.

First, we show the following lemma.

Lemma 4.4 *If $\boldsymbol{f}$ satisfies alternating inequalities (2) and Condition 1, Markov chain $\mathcal{M}_{\mathrm{P}}$ mixes rapidly and the mixing time $\tau(\varepsilon)$ of $\mathcal{M}_{\mathrm{P}}$ satisfies that for* $0 < \varepsilon < 1$,

$$\tau(\varepsilon) \leq n(n-1)^2 \ln\left(Kn/(2\varepsilon)\right).$$

Proof: Let $G = (\Xi, \mathcal{E})$ be the graph defined in the proof of Theorem 3.2 in Section 3. For each edge $e = \{X, Y\} \in \mathcal{E}$, there exists a unique pair of indices $j_1, j_2 \in \{1, 2, \ldots, n\}$ called the *supporting pair* of e, satisfying

$$|X_i - Y_i| = \begin{cases} 1 & (i = j_1, j_2), \\ 0 & (\text{otherwise}). \end{cases}$$

We define the length $l(e)$ of an edge $e = \{X, Y\} \in \mathcal{E}$ by $l(e) \overset{\text{def.}}{=} (1/(n-1)) \sum_{i=1}^{j^*-1}(n-i)$ where $j^* = \max\{j_1, j_2\} \geq 2$ and $\{j_1, j_2\}$ is the supporting pair of e. Note that $1 \leq \min_{e \in \mathcal{E}} l(e) \leq \max_{e \in \mathcal{E}} l(e) \leq n/2$. For each pair $X, Y \in \Xi$, we define the distance $d(X, Y)$ be the length of a shortest path between X and Y on G. Clearly, the diameter of G, i.e., $\max_{(X,Y) \in \Xi^2} d(X, Y)$, is bounded by $Kn/2$, since $d(X, Y) \leq (n/2) \sum_{i=1}^{n} (1/2)|X_i - Y_i| \leq (n/2)K$ for any $(X, Y) \in \Xi^2$. The definition of edge length implies that for any edge $\{X, Y\} \in \mathcal{E}$, $d(X, Y) = l(\{X, Y\})$.

We define a joint process $(X, Y) \mapsto (X', Y')$ as $(X, Y) \mapsto (\phi(X, \Lambda), \phi(Y, \Lambda))$ with uniform real random number $\Lambda \in [1, n)$ and the update function ϕ defined in the previous subsection. Now we show that

$$\mathrm{E}[d(X', Y')] \leq \beta \cdot d(X, Y) \text{ where } \beta = 1 - \frac{1}{n(n-1)^2}, \tag{9}$$

for any pair $\{X, Y\} \in \mathcal{E}$. In the following, we denote the supporting pair of $\{X, Y\}$ by $\{j_1, j_2\}$. Without loss of generality, we can assume that $j_1 < j_2$, and $X_{j_2} + 1 = Y_{j_2}$.

Case 1: When $\lfloor \Lambda \rfloor = j_2 - 1$, we will show that

$$\mathrm{E}[d(X', Y') \mid \lfloor \Lambda \rfloor = j_2 - 1] \leq d(X, Y) - \frac{1}{2} \cdot \frac{n - j_2 + 1}{n - 1}.$$

In case of $j_1 = j_2 - 1$, $X' = Y'$ holds with conditional probability 1. Hence $d(X', Y') = 0$. In the following, we consider the case $j_1 < j_2 - 1$. Put $k = X_{j_2-1} + X_{j_2}$, then $Y_{j_2-1} + Y_{j_2} = k+1$ holds since $X_{j_2} + 1 = Y_{j_2}$. Now suppose that $l \in \{0, 1, \ldots, k\}$ satisfies $g^{k}_{(j_2-1)j_2}(l-1) \leq \Lambda - \lfloor \Lambda \rfloor < g^{k}_{(j_2-1)j_2}(l)$. Then $g^{k+1}_{(j_2-1)j_2}(l-1) \leq \Lambda - \lfloor \Lambda \rfloor < g^{k+1}_{(j_2-1)j_2}(l+1)$ holds, since alternating inequalities (3) imply that $g^{k+1}_{(j_2-1)j_2}(l-1) \leq g^{k}_{(j_2-1)j_2}(l-1) < g^{k+1}_{(j_2-1)j_2}(l) \leq g^{k+1}_{(j_2-1)j_2}(l+1)$. Thus we have that if $Y'_{j_2-1} = l$ then $X'_{j_2-1} = l$ or $l+1$. In other words,

$$\begin{pmatrix} X'_{j_2-1} \\ Y'_{j_2-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ldots, \begin{pmatrix} k \\ k \end{pmatrix}, \begin{pmatrix} k \\ k+1 \end{pmatrix} \right\}.$$

holds. If $X'_{j_2-1} = Y'_{j_2-1}$, the supporting pair of $\{X', Y'\}$ is $\{j_1, j_2\}$ and so $d(X', Y') = d(X, Y)$. If $X'_{j_2-1} \neq Y'_{j_2-1}$, the supporting pair of $\{X', Y'\}$ is $\{j_1, j_2 - 1\}$ and so $d(X', Y') = d(X, Y) - (n - j_2 + 1)/(n-1)$. With considering that

$$\Pr[X'_{j_2-1} \neq Y'_{j_2-1} \mid \lfloor \Lambda \rfloor = j_2 - 1] = \textstyle\sum_{l=0}^{k'} \left(g^{k'}_{(j_2-1),j_2}(l) - g^{k'+1}_{(j_2-1),j_2}(l) \right),$$

Condition 1 implies that

$$\begin{aligned} \Pr\left[X'_{j_2-1} \neq Y'_{j_2-1} \mid \lfloor \Lambda \rfloor = j_2 - 1\right] &\geq 1/2, \\ \Pr\left[X'_{j_2-1} = Y'_{j_2-1} \mid \lfloor \Lambda \rfloor = j_2 - 1\right] &\leq 1/2. \end{aligned}$$

Thus we obtain

$$\begin{aligned}\mathrm{E}\left[d(X',Y') \mid \lfloor\Lambda\rfloor = j_2 - 1\right] &\leq \frac{1}{2}d(X,Y) + \frac{1}{2}\left(d(X,Y) - \frac{n-j_2+1}{n-1}\right) \\ &= d(X,Y) - \frac{1}{2}\cdot\frac{n-j_2+1}{n-1}.\end{aligned}$$

Case 2: When $\lfloor\Lambda\rfloor = j_2$, we can show that

$$\mathrm{E}[d(X',Y')|\lfloor\Lambda\rfloor = j_2] \leq d(X,Y) + \frac{1}{2}\cdot\frac{n-j_2}{n-1}$$

in a similar way to Case 1.

Case 3: When $\lfloor\Lambda\rfloor \neq j_2 - 1$ and $\lfloor\Lambda\rfloor \neq j_2$, it is easy to see that the supporting pair $\{j_1', j_2'\}$ of $\{X', Y'\}$ satisfies $j_2 = \max\{j_1', j_2'\}$. Thus $d(X,Y) = d(X',Y')$.

The probability of appearance of Case 1 is equal to $1/(n-1)$, and that of Case 2 is less than or equal to $1/(n-1)$. From the above, we have that

$$\begin{aligned}\mathrm{E}[d(X',Y')] &\leq d(X,Y) - \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2+1}{n-1} + \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2}{n-1} = d(X,Y) - \frac{1}{2(n-1)^2} \\ &\leq \left(1 - \frac{1}{2(n-1)^2}\cdot\frac{1}{\max_{\{X,Y\}\in\mathcal{E}}\{d(X,Y)\}}\right)d(X,Y) = \left(1 - \frac{1}{n(n-1)^2}\right)d(X,Y).\end{aligned}$$

Since the diameter of G is bounded by $Kn/2$, Theorem 3.3 implies that the mixing time $\tau(\varepsilon)$ for $0 < \varepsilon < 1$ satisfies

$$\tau(\varepsilon) \leq n(n-1)^2 \ln\left(Kn/(2\varepsilon)\right).$$

$\square$

Next we estimate the expectation of *coalescence time* $T_* \in \mathbb{Z}_+$ of $\mathcal{M}_{\mathrm{P}}$ defined by $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi^0_{-t}(x, \mathbf{\Lambda})\}$. Note that T_* is a random variable.

Lemma 4.5 *If* $\boldsymbol{f}$ *satisfies alternating inequalities (2) and Condition 1, the coalescence time* T_* *of* $\mathcal{M}_{\mathrm{P}}$ *satisfies* $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln Kn)$.

Proof: Let $G = (\Xi, \mathcal{E})$ be the undirected graph and $d(X,Y)$, $\forall X, \forall Y \in \Xi$, be the metric on G, both of which are used in the proof of Lemma 4.4. We define $D \stackrel{\text{def.}}{=} d(x_{\mathrm{U}}, x_{\mathrm{L}})$ and $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1 + \ln D)$. By using the inequality (9) obtained in the proof of Lemma 4.4, we have

$$\begin{aligned}\Pr\left[T_* > \tau_0\right] &= \Pr\left[\Phi^0_{-\tau_0}(x_{\mathrm{U}}, \mathbf{\Lambda}) \neq \Phi^0_{-\tau_0}(x_{\mathrm{L}}, \mathbf{\Lambda})\right] = \Pr\left[\Phi^{\tau_0}_0(x_{\mathrm{U}}, \mathbf{\Lambda}) \neq \Phi^{\tau_0}_0(x_{\mathrm{L}}, \mathbf{\Lambda})\right] \\ &\leq \textstyle\sum_{(X,Y)\in\Xi^2} d(X,Y) \Pr\left[X = \Phi^{\tau_0}_0(x_{\mathrm{U}}, \mathbf{\Lambda}), Y = \Phi^{\tau_0}_0(x_{\mathrm{L}}, \mathbf{\Lambda})\right] \\ &= \mathrm{E}\left[d\left(\Phi^{\tau_0}_0(x_{\mathrm{U}}, \mathbf{\Lambda}), \Phi^{\tau_0}_0(x_{\mathrm{L}}, \mathbf{\Lambda})\right)\right] \leq \left(1 - \frac{1}{n(n-1)^2}\right)^{\tau_0} d(x_{\mathrm{U}}, x_{\mathrm{L}}) \\ &= \left(1 - \frac{1}{n(n-1)^2}\right)^{n(n-1)^2(1+\ln D)} D \leq \mathrm{e}^{-1}\mathrm{e}^{-\ln D} D = \frac{1}{\mathrm{e}}.\end{aligned}$$

By the *submultiplicativity* of coalescence time (see [18] e.g.), for any $k \in \mathbb{Z}_+$, $\Pr[T_* > k\tau_0] \leq (\Pr[T_* > \tau_0])^k \leq (1/\mathrm{e})^k$. Thus

$$\begin{aligned}\mathrm{E}[T_*] &= \textstyle\sum_{t=0}^{\infty} t \Pr[T_* = t] \leq \tau_0 + \tau_0 \Pr[T_* > \tau_0] + \tau_0 \Pr[T_* > 2\tau_0] + \cdots \\ &\leq \tau_0 + \tau_0/\mathrm{e} + \tau_0/\mathrm{e}^2 + \cdots = \tau_0/(1 - 1/\mathrm{e}) \leq 2\tau_0.\end{aligned}$$

Clearly $D \leq Kn$, then we obtain the result that $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln Kn)$. $\square$

Proof of Theorem 4.3 Let T_* be the coalescence time of our chain. Note that T_* is a random variable. Put $m = \lceil \log_2 T_* \rceil$. Algorithm 1 terminates when we set the starting time period $T = -2^m$ at $(m+1)$st iteration. Then the total number of simulated transitions is bounded by $2(2^0 + 2^1 + 2^2 + \cdots + 2^K) < 2 \cdot 2 \cdot 2^m \leq 8T_*$, since we need to execute two chains from both x_{U} and x_{L}. Thus the expectation of the total number of transitions of $\mathcal{M}_{\mathrm{P}}$ required in the monotone CFTP algorithm is bounded by $\mathrm{O}(\mathrm{E}[8T_*]) = \mathrm{O}(n^3 \ln Kn)$. $\square$

Note that we can obtain Condition 1 for closed Jackson networks with single servers model, that is a special case of Example 1 such that $s_i = 1$ ($i \in \{1, 2, \ldots, n\}$), by arranging parameters $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ in non-increasing order such as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$. For Discretized Dirichlet Distribution (Example 2), we can obtain Condition 1 by arranging parameters in non-increasing order such as $u_1 \geq u_2 \geq \cdots \geq u_n$ [15].

5 Concluding Remarks

In this paper, we concern with random sampling of an integral point on $n-1$ dimensional simplex of side length K, according to a multivariate discrete distribution. We introduce an idea of *alternating inequalities*, and we propose two hit-and-run chains, one is rapidly mixing chain, the other is monotone chain. We show that a logarithmic separable concave function satisfies the alternating inequalities. Here we note that there exist functions which satisfy alternating inequalities, though they are not log-concave, for example, Discretized Dirichlet Distribution with parameters less than 1 [16, 15].

One of future works is an extension of our results to general log-concave functions on a simplex. It should gives an efficient algorithm for universal portfolios [12]. Another is an extension to general convex bodies, especially to base polytope of a polymatroid. A set of two-rowed contingency tables for given marginal sums is a simple example of base polytope of a polymatroid. For two-rowed contingency tables, there is an approximately uniform sampler by Dyer and Greenhill [6] and a perfect uniform sampler by our another work [13]. Matsui, et al. [17], including common author of this paper, gave rapidly mixing chain for sampling two-rowed tables from hyper geometric distribution, which is also a logarithmic-separable concave probability function. For tables with any constant number of rows, Cryan, et al. [5] showed that a heat-bath chain mixes rapidly.

References

[1] M. Y. An, Log-concave Probability Distributions: Theory and Statistical Testing, Papers 96-01, Centre for Labour Market and Social Research, Danmark, 1996. (available from http://ideas.repec.org/s/fth/clmsre.html)

[2] M. Bagnoli and T. Bergstrom, Log-Concave Probability and its Applications, papers 89-23, Michigan-Center for Research on Economic & Social Theory, 1989. (available from http://ideas.repec.org/s/fth/michet.html)

[3] R. Bubley and M. Dyer, Path coupling: A Technique for Proving Rapid Mixing in Markov Chains, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS 1997), 223–231.

[4] W. Chen and C. A. O'Cinneide, Towards a Polynomial-Time Randomized Algorithm for Closed Product-Form Networks, ACM Transactions on Modeling and Computer Simulation, **8** (1998), 227–253.

[5] M. Cryan, M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin, Rapidly Mixing Markov Chains for Sampling Contingency Tables with Constant Number of Rows, Proceedings of the 43rd Annual Symposium on Foundations of Computer Science (FOCS 2002), 711–720.

[6] M. Dyer and C. Greenhill, Polynomial-Time Counting and Sampling of Two-Rowed Contingency Tables, Theoretical Computer Sciences, **246** (2000), 265–278.

[7] J. Fill, An Interruptible Algorithm for Perfect Sampling via Markov Chains, The Annals of Applied Probability, **8** (1998), 131–162.

[8] J. Fill, M. Machida, D. Murdoch, and J. Rosenthal, Extension of Fill's Perfect Rejection Sampling Algorithm to General Chains, Random Structures and Algorithms, **17** (2000), 290–316.

[9] A. Frieze, R. Kannan, and N. Polson, Sampling from Log-Concave Distribution, The Annals of Applied Probability, **4** (1994), 812–837.

[10] A. Frieze and R. Kannan, Log-Sobolev Inequalities and Sampling from Log-Concave Distributions, The Annals of Applied Probability, **9** (1999), 14–26.

[11] J. R. Jackson, Networks of Waiting Lines, The Journal of Operations Research Society of America, **5** (1957), 518–521.

[12] A. Kalai and S. Vempala, Efficient Algorithms for Universal Portfolios, Journal of Machine Learning Research, **3** (2002), 423–440.

[13] S. Kijima and T. Matsui, Polynomial Time Perfect Sampling Algorithm for Two-rowed Contingency Tables, Random Structures Algorithms, to appear.

[14] L. Lovász and S. Vempala, Logconcave Functions: Geometry and Efficient Sampling Algorithms, Proceedings of the 44th Symposium on Foundations of Computer Science (FOCS 2003), 640–649.

[15] T. Matsui and S. Kijima, Polynomial Time Perfect Sampler for Discretized Dirichlet Distribution, METR 2003-17, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[16] T. Matsui, M. Motoki, and N. Kamatani, Polynomial Time Approximate Sampler for Discretized Dirichlet Distribution, 14th International Symposium on Algorithms and Computation (ISAAC 2003), LNCS, **2906** (2003) 676–685.

[17] T. Matsui, Y. Matsui, and Y. Ono, Random Generation of $2 \times 2 \times \cdots \times 2 \times J$ Contingency Tables, Theoretical Computer Science, **326** (2004), 117–135.

[18] J. Propp and D. Wilson, Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics, Random Structures Algorithms, **9** (1996), 223–252.

[19] D. Randall and P. Winkler, Mixing Points on an Interval, Proceedings of the Second Workshop on Analytic Algorithms and Combinatorics (ANALCO 2005), http://www.siam.org/meetings/analco05/program.htm.

[20] D. Wilson, How to Couple from the Past Using a Read-Once Source of Randomness, Random Structures Algorithms, **16** (2000), 85–113.