*Dedicated to Academician Marius Iosifescu*
*on the occasion of his 75th birthday*

# A HYBRID METROPOLIS-HASTINGS CHAIN

UDREA PĂUN

We consider a hybrid Metropolis-Hastings chain on a known finite state space; its design is based on the $G$ method (because this method can perform some interesting things, see Sections 1 and 2 and also [17]) and its analysis (the convergence rate) is based on the $G$ method and ergodicity coefficients. Finally, we give two special cases of the hybrid chain, namely, when the state space is $\mathbb{S}_n$, the set of permutations of order $n$ (e.g., the Mallows model is defined on this space), and when this is $\{0, 1, \ldots, h\}^n$, $h, n \geq 1$ (e.g., the Ising model is defined on $\{0, 1\}^n$).

*AMS 2010 Subject Classification:* 60J10, 60J20, 65C05, 68Q25.

*Key words:* $\Delta$-ergodic theory, $G$ method, ergodicity coefficient, Metropolis-Hastings chain, hybrid Metropolis-Hastings chain, convergence rate.

## 1. $G_{\Delta_1,\Delta_2}$ AND $\overline{G}_{\Delta_1,\Delta_2}$ IN ACTION

To design and analyze our hybrid Metropolis-Hastings chain we need to extend certain notions and results from the $\Delta$-ergodic theory in a more general framework, namely, that of nonnegative matrices, and, then, to give, at least for the $\Delta$-ergodic theory, other results. (See mainly [17] for this section; see [13–17] and references therein for the general $\Delta$-ergodic theory.)

Set

$$\mathrm{Par}(E) = \{\Delta \mid \Delta \text{ is a partition of } E\},$$

where $E$ is a nonempty set. We shall agree that the partitions do not contain the empty set.

*Definition* 1.1. Let $\Delta_1, \Delta_2 \in \mathrm{Par}(E)$. We say that $\Delta_1$ *is finer than* $\Delta_2$ if $\forall V \in \Delta_1$, $\exists W \in \Delta_2$ such that $V \subseteq W$.

Write $\Delta_1 \preceq \Delta_2$ when $\Delta_1$ is finer than $\Delta_2$.

Set

$$\langle m \rangle = \{1, 2, \ldots, m\}, \quad m \geq 1,$$

$$N_{m,n} = \{F \mid F \text{ is a nonnegative } m \times n \text{ matrix}\},$$
$$S_{m,n} = \{F \mid F \text{ is a stochastic } m \times n \text{ matrix}\},$$
$$N_n = N_{n,n} \quad \text{and} \quad S_n = S_{n,n}.$$

Let $F = (F_{ij}) \in N_{m,n}$. (The entries of a matrix $Z$ will be denoted $Z_{ij}$.) Let $\emptyset \neq U \subseteq \langle m \rangle$ and $\emptyset \neq V \subseteq \langle n \rangle$. Define the matrices

$$F_U = (F_{ij})_{i \in U, j \in \langle n \rangle}, \quad F^V = (F_{ij})_{i \in \langle m \rangle, j \in V}, \quad \text{and} \quad F_U^V = (F_{ij})_{i \in U, j \in V}.$$

*Definition* 1.2. Let $P \in N_{m,n}$. We say that $P$ is a *generalized stochastic matrix* if $\exists a \geq 0$, $\exists Q \in S_{m,n}$ such that $P = aQ$.

*Definition* 1.3 ([17]). Let $P \in N_{m,n}$. Let $\Delta \in \operatorname{Par}(\langle m \rangle)$ and $\Sigma \in \operatorname{Par}(\langle n \rangle)$. We say that $P$ is a $[\Delta]$-*stable matrix on* $\Sigma$ if $P_K^L$ is a generalized stochastic matrix, $\forall K \in \Delta$, $\forall L \in \Sigma$. In particular, a $[\Delta]$-stable matrix on $(\{i\})_{i \in \langle n \rangle}$ is called $[\Delta]$-*stable* for short $((\{i\})_{i \in \langle n \rangle} := (\{1\}, \{2\}, \ldots, \{n\}))$.

*Definition* 1.4 ([17]). Let $P \in N_{m,n}$. Let $\Delta \in \operatorname{Par}(\langle m \rangle)$ and $\Sigma \in \operatorname{Par}(\langle n \rangle)$. We say that $P$ is a $\Delta$-*stable matrix on* $\Sigma$ if $\Delta$ is the least fine partition for which $P$ is a $[\Delta]$-stable matrix on $\Sigma$. In particular, a $\Delta$-stable matrix on $(\{i\})_{i \in \langle n \rangle}$ is called $\Delta$-*stable* while a $(\langle m \rangle)$-stable matrix on $\Sigma$ is called *stable on* $\Sigma$ for short. A stable matrix on $(\{i\})_{i \in \langle n \rangle}$ is called *stable* for short.

Let $\Delta_1 \in \operatorname{Par}(\langle m \rangle)$ and $\Delta_2 \in \operatorname{Par}(\langle n \rangle)$. Define

$$G_{\Delta_1, \Delta_2} = \{P \mid P \in S_{m,n} \text{ and } P \text{ is a } [\Delta_1]\text{-stable matrix on } \Delta_2\}$$

(see [17] and, for an equivalent definition, [12]),

$$\overline{G}_{\Delta_1, \Delta_2} = \{P \mid P \in N_{m,n} \text{ and } P \text{ is a } [\Delta_1]\text{-stable matrix on } \Delta_2\},$$

and, if $m = n$,

$$G_\Delta = G_{\Delta, \Delta}$$

(see [11] for an equivalent definition) and

$$\overline{G}_\Delta = \overline{G}_{\Delta, \Delta}.$$

Let $P \in \overline{G}_{\Delta_1, \Delta_2}$. Let $K \in \Delta_1$ and $L \in \Delta_2$. Then $\exists a_{K,L} \geq 0$, $\exists Q_{K,L} \in S_{|K|,|L|}$ such that $P_K^L = a_{K,L} Q_{K,L}$. Set

$$P^{-+} = (P_{KL}^{-+})_{K \in \Delta_1, L \in \Delta_2}, \quad P_{KL}^{-+} = a_{K,L}, \quad \forall K \in \Delta_1, \ \forall L \in \Delta_2$$

(see also [17]). If confusion can arise we write $P^{-+(\Delta_1, \Delta_2)}$ instead of $P^{-+}$. In this article, when we work with the operator $(\cdot)^{-+} = (\cdot)^{-+}(\Delta_1, \Delta_2)$ we suppose, for labelling the rows and columns of matrices, that $\Delta_1$ and $\Delta_2$ are ordered sets, even if we omit to precise this. To give an example, let

$$P = \begin{pmatrix} 2 & 3 & 7 & 0 \\ 5 & 0 & 6 & 1 \\ 1 & 0 & 0 & 2 \end{pmatrix}.$$

Obviously, $P \in \overline{G}_{\Delta_1,\Delta_2}$, where $\Delta_1 = (\{1,2\},\{3\})$ and $\Delta_2 = (\{1,2\},\{3,4\})$. Further, we have

$$P^{-+} = P^{-+(\Delta_1,\Delta_2)} = \begin{pmatrix} 5 & 7 \\ 1 & 2 \end{pmatrix}.$$

($\{1,2\}$ and $\{3\}$ are the first and the second element of $\Delta_1$, respectively; on the basis of this order, the first and the second row of $P^{-+}$ are labelled $\{1,2\}$ and $\{3\}$, respectively. The columns of $P^{-+}$ are labelled similarly.)

The next result is the main one of this section; it is a generalization of Theorem 2.3 in [17].

THEOREM 1.5. *Let* $P \in \overline{G}_{\Delta_1,\Delta_2} \subseteq N_{m,n}$ *and* $Q \in \overline{G}_{\Delta_2,\Delta_3} \subseteq N_{n,p}$. *Then*
(i) $PQ \in \overline{G}_{\Delta_1,\Delta_3} \subseteq N_{m,p}$;
(ii) $(PQ)^{-+} = P^{-+}Q^{-+}$.

*Proof.* (i) Let $P \in \overline{G}_{\Delta_1,\Delta_2}$ and $Q \in \overline{G}_{\Delta_2,\Delta_3}$. Then $\forall K \in \Delta_1$, $\forall U \in \Delta_2$, $\forall L \in \Delta_3$, $\exists a_{K,U} \geq 0$, $\exists A_{K,U} \in S_{|K|,|U|}$, $\exists b_{U,L} \geq 0$, $\exists B_{U,L} \in S_{|U|,|L|}$ such that $P_K^U = a_{K,U}A_{K,U}$ and $Q_U^L = b_{U,L}B_{U,L}$.

Let $K \in \Delta_1$ and $L \in \Delta_3$. Let $i \in K$. We have

$$\sum_{l\in L}(PQ)_{il} = \sum_{l\in L}\sum_{k\in\langle n\rangle} P_{ik}Q_{kl} = \sum_{k\in\langle n\rangle} P_{ik}\sum_{l\in L} Q_{kl} = \sum_{W\in\Delta_2}\sum_{k\in W} P_{ik}\sum_{l\in L} Q_{kl} =$$

$$= \sum_{W\in\Delta_2}\sum_{k\in W} P_{ik}b_{W,L} = \sum_{W\in\Delta_2} b_{W,L}\sum_{k\in W} P_{ik} = \sum_{W\in\Delta_2} a_{K,W}b_{W,L}.$$

It follows that $\sum\limits_{l\in L}(PQ)_{il}$ only depends on constants $a_{K,W}$, $b_{W,L}$, $W \in \Delta_2$, $\forall i \in K$. Therefore, $PQ \in \overline{G}_{\Delta_1,\Delta_3}$.

(ii) See the proof of (i). $\square$

In this article, a vector $x$ is a row vector and $x'$ denotes its transpose. Set $e = e(n) = (1,1,\ldots,1) \in \mathbf{R}^n$, $\forall n \geq 1$.

The next result is a generalization of Theorem 2.10 in [17]; it is another main result of this section.

THEOREM 1.6. *Let* $P_1 \in \overline{G}_{(\langle m_1\rangle),\Delta_2} \subseteq N_{m_1,m_2}$, $P_2 \in \overline{G}_{\Delta_2,\Delta_3} \subseteq N_{m_2,m_3}$, $\ldots, P_{n-1} \in \overline{G}_{\Delta_{n-1},\Delta_n} \subseteq N_{m_{n-1},m_n}$, $P_n \in \overline{G}_{\Delta_n,(\{i\})_{i\in\langle m_{n+1}\rangle}} \subseteq N_{m_n,m_{n+1}}$. *Then*
(i) $P_1P_2\ldots P_n$ *is a stable matrix*;
(ii) $\pi = P_1^{-+}P_2^{-+}\ldots P_n^{-+}$, *where* $e'\pi := P_1P_2\ldots P_n$ (*i.e.*, $\pi = (P_1P_2\ldots P_n)_{\{i\}}$ *for a fixed* $i$).

*Proof.* (i) By Theorem 1.5(i) and induction we have $P_1P_2\dots P_n \in \overline{G}_{(\langle m_1\rangle),(\{i\})_{i\in\langle m_{n+1}\rangle}}$. Therefore, $P_1P_2\dots P_n$ is a stable matrix. (A matrix $P \in N_{q,r}$ is a stable matrix if and only if $P \in \overline{G}_{(\langle q\rangle),(\{i\})_{i\in\langle r\rangle}}$.)

(ii) By (i), $\pi = (P_1P_2\dots P_n)^{-+}$. Now, (ii) follows by Theorem 1.5(ii) and induction. □

Let $P \in N_{m,n}$. Define

$$\alpha(P) = \min_{1\le i,j\le m} \sum_{k=1}^{n} \min(P_{ik}, P_{jk})$$

(if $P \in S_{m,n}$, then $\alpha(P)$ is called the *Dobrushin ergodicity coefficient of* $P$ ([4]; see, e.g., also [6, p. 56])) and

$$\overline{\alpha}(P) = \frac{1}{2} \max_{1\le i,j\le m} \sum_{k=1}^{n} |P_{ik} - P_{jk}|.$$

*Remark* 1.7 (see, e.g., [7, p. 143]). If $P \in S_{m,n}$, then $\overline{\alpha}(P) = 1 - \alpha(P)$.

Theorem 1.6(i) can be used, e.g., to see if a finite Markov chain has a finite convergence time (see also [17]). Below we give other applications (Theorem 1.8 and Remark 1.9) – the best results of this section – of $G_{\Delta_1,\Delta_2}$ and of $\overline{G}_{\Delta_1,\Delta_2}$; they give bounds for $\alpha(P_1P_2\dots P_n)$, where $P_1, P_2, \dots, P_n$ are nonnegative matrices (an interesting case is that when these matrices are sparse large stochastic). When we study products of nonnegative matrices using $G_{\Delta_1,\Delta_2}$ and/or $\overline{G}_{\Delta_1,\Delta_2}$ we shall refer this as the *G method*.

Theorem 1.8. *Let* $P_1 \in N_{m_1,m_2}$, $P_2 \in N_{m_2,m_3}, \dots, P_n \in N_{m_n,m_{n+1}}$. *Let* $\Delta_1 = (\langle m_1\rangle)$, $\Delta_2 \in \mathrm{Par}(\langle m_2\rangle), \dots, \Delta_n \in \mathrm{Par}(\langle m_n\rangle)$, $\Delta_{n+1} = (\{i\})_{i\in\langle m_{n+1}\rangle}$. *Consider the matrices* $L_l = ((L_l)_{VW})_{V\in\Delta_l, W\in\Delta_{l+1}}$ *(*$(L_l)_{VW}$ *is the entry* $(V,W)$ *of matrix* $L_l$*)*, $U_l = ((U_l)_{VW})_{V\in\Delta_l, W\in\Delta_{l+1}}$, $l \in \langle n\rangle$, *where*

$$(L_l)_{VW} := \min_{i\in V} \sum_{j\in W} (P_l)_{ij} \text{ and } (U_l)_{VW} := \max_{i\in V} \sum_{j\in W} (P_l)_{ij},$$

$\forall l \in \langle n\rangle$, $\forall V \in \Delta_l$, $\forall W \in \Delta_{l+1}$. *Then*

$$\sum_{K\in\Delta_{n+1}} (L_1L_2\dots L_n)_{\langle m_1\rangle K} \le \alpha(P_1P_2\dots P_n) \le \sum_{K\in\Delta_{n+1}} (U_1U_2\dots U_n)_{\langle m_1\rangle K}.$$

*(Since* $L_1L_2\dots L_n$ *and* $U_1U_2\dots U_n$ *are* $1 \times |\langle m_{n+1}\rangle|$ *matrices, they can be thought as being row vectors, but above we used and below we shall use the matrix notation for entries instead of the vector one. E.g., above the matrix notation* $(L_1L_2\dots L_n)_{\langle m_1\rangle K}$ *was used instead of the vector one* $(L_1L_2\dots L_n)_K$ *because, in this article, the notation* $A_U$*, where* $A \in N_{p,q}$ and $\emptyset \ne U \subseteq \langle p\rangle$*, means something different.)*

*Proof.* Consider the matrices $E_1, F_1 \in \overline{G}_{\Delta_1,\Delta_2}$, $E_2$, $F_2 \in \overline{G}_{\Delta_2,\Delta_3}, \ldots, E_n$, $F_n \in \overline{G}_{\Delta_n,\Delta_{n+1}}$ such that $E_l^{-+} = L_l$ and $F_l^{-+} = U_l$, $\forall l \in \langle n \rangle$. Since (see Theorem 1.6)

$$(E_1 E_2 \ldots E_n)_{\{i\}} = (E_1 E_2 \ldots E_n)^{-+} = E_1^{-+} E_2^{-+} \cdots E_n^{-+} = L_1 L_2 \ldots L_n$$

and

$$(F_1 F_2 \ldots F_n)_{\{i\}} = (F_1 F_2 \ldots F_n)^{-+} = F_1^{-+} F_2^{-+} \cdots F_n^{-+} = U_1 U_2 \ldots U_n,$$

$\forall i \in \langle m_1 \rangle$, we have

$$(E_1 E_2 \ldots E_n)_{ij} = (L_1 L_2 \ldots L_n)_{\langle m_1 \rangle \{j\}}$$

and

$$(F_1 F_2 \ldots F_n)_{ij} = (U_1 U_2 \ldots U_n)_{\langle m_1 \rangle \{j\}},$$

$\forall i \in \langle m_1 \rangle$, $\forall j \in \langle m_{n+1} \rangle$.

We choose the matrices $E_l, F_l$, $l \in \langle n \rangle$, such that

$$(E_l)_{ij} \leq (P_l)_{ij} \leq (F_l)_{ij}, \quad \forall l \in \langle n \rangle,\ \forall i \in \langle m_l \rangle,\ \forall j \in \langle m_{l+1} \rangle;$$

this is possible as follows. Let $l \in \langle n \rangle$. Let $V \in \Delta_l$ and $W \in \Delta_{l+1}$. Let $i \in V$.

*Case* 1. $(P_l)_{ij} = 0$, $\forall j \in W$. No problem. (We must take $(E_l)_{ij} = 0$, $\forall j \in W$, and we can take, e.g., $(F_l)_{ij} = 0$, $\forall j \in W$.)

*Case* 2. $\exists j \in W$ such that $(P_l)_{ij} > 0$. First, we construct the entries $(E_l)_{ij}$, $j \in W$. We take $(E_l)_{ij} = 0$, $\forall j \in W$, if $(L_l)_{VW} = 0$. (Recall that $E_l^{-+} = L_l$ and $(L_l)_{VW} = 0$ imply $(E_l)_{ij} = 0$, $\forall j \in W$.) Suppose, now, that $(L_l)_{VW} > 0$. Let $(P_l)_{ij_1}, (P_l)_{ij_2}, \ldots, (P_l)_{ij_k}$ be all the nonnegative entries of $(P_l)_{\{i\}}^W$. Suppose that $(P_l)_{ij_1} \leq (P_l)_{ij_2} \leq \cdots \leq (P_l)_{ij_k}$. We know that $E_l^{-+} = L_l$ and $(L_l)_{VW} \leq \sum_{t=1}^{k} (P_l)_{ij_t}$. If $(L_l)_{VW} \leq (P_l)_{ij_k}$, we take $(E_l)_{ij_k} = (L_l)_{VW}$ and $(E_l)_{ij} = 0$, $\forall j \in W$, $j \neq j_k$. Otherwise, if $(L_l)_{VW} \leq (P_l)_{ij_{k-1}} + (P_l)_{ij_k}$, we take $(E_l)_{ij_k} = (P_l)_{ij_k}$, $(E_l)_{ij_{k-1}} = (L_l)_{VW} - (E_l)_{ij_k}$, and $(E_l)_{ij} = 0$, $\forall j \in W$, $j \neq j_{k-1}, j_k$. Otherwise, if $(L_l)_{VW} \leq (P_l)_{ij_{k-2}} + (P_l)_{ij_{k-1}} + (P_l)_{ij_k}$, we take $(E_l)_{ij_k} = (P_l)_{ij_k}$, $(E_l)_{ij_{k-1}} = (P_l)_{ij_{k-1}}$, $(E_l)_{ij_{k-2}} = (L_l)_{VW} - (E_l)_{ij_{k-1}} - (E_l)_{ij_k}$, and $(E_l)_{ij} = 0$, $\forall j \in W$, $j \neq j_{k-2}, j_{k-1}, j_k$. Etc. Second, we construct the entries $(F_l)_{ij}$, $j \in W$. Using the above notation, we take $(F_l)_{ij_k} = (P_l)_{ij_k} + \left( (U_l)_{VW} - \sum_{t=1}^{k} (P_l)_{ij_t} \right)$ and $(F_l)_{ij} = (P_l)_{ij}$, $\forall j \in W$, $j \neq j_k$.

Finally, by

$$(L_1 L_2 \ldots L_n)_{\langle m_1 \rangle \{j\}} = (E_1 E_2 \ldots E_n)_{ij} \leq (P_1 P_2 \ldots P_n)_{ij} \leq$$
$$\leq (F_1 F_2 \ldots F_n)_{ij} = (U_1 U_2 \ldots U_n)_{\langle m_1 \rangle \{j\}}, \quad \forall i \in \langle m_1 \rangle,\ \forall j \in \langle m_{n+1} \rangle,$$

we have

$$\sum_{K\in\Delta_{n+1}}(L_1L_2\dots L_n)_{\langle m_1\rangle K}\le\alpha(P_1P_2\dots P_n)\le\sum_{K\in\Delta_{n+1}}(U_1U_2\dots U_n)_{\langle m_1\rangle K}.$$

(We used the fact that $\alpha(P)\le\alpha(Q)$ if $P\le Q$ $(P,Q\in N_{p,q})$.) $\square$

Let $P\in N_{m,n}$. Let $\Delta\in\mathrm{Par}(\langle m\rangle)$ and $\Sigma\in\mathrm{Par}(\langle n\rangle)$. Define

$$P^+=(P^+_{iJ})_{i\in\langle m\rangle,\,J\in\Sigma},\quad P^+_{iJ}=\sum_{k\in J}P_{ik},\quad \forall i\in\langle m\rangle,\ \forall J\in\Sigma$$

(see also [15]). If confusion can arise we write $P^{+\Sigma}$ instead of $P^+$. In this article, when we work with the operator $(\cdot)^+=(\cdot)^+(\Sigma)$ we suppose, for labelling the columns of matrices, that $\Sigma$ is an ordered set, even if we omit to precise this.

*Remark* 1.9. (a) By Theorem 1.8 we have

$$\underline{b}=\underline{b}(P_1,P_2,\dots,P_n):=\max_{\substack{\Delta_1=(\langle m_1\rangle),\\ \Delta_2\in\mathrm{Par}(\langle m_2\rangle),\dots,\\ \Delta_n\in\mathrm{Par}(\langle m_n\rangle)\\ \Delta_{n+1}=(\{i\})_{i\in\langle m_{n+1}\rangle}}}\sum_{K\in\Delta_{n+1}}(L_1L_2\dots L_n)_{\langle m_1\rangle K}\le$$

$$\le\alpha(P_1P_2\dots P_n)\le$$

$$\le\min_{\substack{\Delta_1=(\langle m_1\rangle),\\ \Delta_2\in\mathrm{Par}(\langle m_2\rangle),\dots,\\ \Delta_n\in\mathrm{Par}(\langle m_n\rangle)\\ \Delta_{n+1}=(\{i\})_{i\in\langle m_{n+1}\rangle}}}\sum_{K\in\Delta_{n+1}}(U_1U_2\dots U_n)_{\langle m_1\rangle K}:=\overline{b}(P_1,P_2,\dots,P_n)=\overline{b}.$$

(b) If $P_1,P_2,\dots,P_n$ are stochastic matrices, then $\overline{b}\ge 1$ while if $P_1,P_2,\dots,P_n$ are substochastic matrices, then it is possible that $\overline{b}$ be smaller or equal to 1. Further, as to the products of stochastic matrices, since $\alpha(P)\le 1$, $\forall P\in S_{m,n}$, it follows that the first inequality from Theorem 1.8 (also, the first inequality from (a)) is only interesting. This inequality can be even an equation in some special cases. E.g., let

$$P=\begin{pmatrix}\frac34&\frac14&0\\ 0&\frac34&\frac14\\ 0&0&1\end{pmatrix}.$$

If we take $\Delta_1=(\langle 3\rangle)$, $\Delta_2=(\{1\},\{2,3\})$, and $\Delta_3=(\{i\})_{i\in\langle 3\rangle}=(\{1\},\{2\},\{3\})$, then

$$L_1L_2=\begin{pmatrix}0&\frac14\end{pmatrix}\begin{pmatrix}\frac34&\frac14&0\\ 0&0&\frac14\end{pmatrix}=\begin{pmatrix}0&0&\frac1{16}\end{pmatrix}.$$

By Theorem 1.8, $0+0+\frac{1}{16}=\frac{1}{16}\leq\alpha(P^2)$. On the other hand, since

$$P^2=\begin{pmatrix} \frac{9}{16} & \frac{6}{16} & \frac{1}{16} \\ 0 & \frac{9}{16} & \frac{7}{16} \\ 0 & 0 & 1 \end{pmatrix},$$

we obtain $\alpha(P^2)=\frac{1}{16}$ by direct computation. To give other examples, we can, e.g., use some examples from [17].

(c) Let $P_1\in N_{m_1,m_2}$, $P_2\in N_{m_2,m_3},\ldots,$ $P_n\in N_{m_n,m_{n+1}}$. Let $\Delta_1\in \operatorname{Par}(\langle m_1\rangle)$, $\Delta_2,\Delta_2'\in\operatorname{Par}(\langle m_2\rangle),\ldots,\Delta_n,\Delta_n'\in\operatorname{Par}(\langle m_n\rangle)$, $\Delta_{n+1}\in\operatorname{Par}(\langle m_{n+1}\rangle)$. By the proof of Theorem 1.8 there are matrices $E_1,F_1\in\overline{G}_{\Delta_1,\Delta_2}$, $E_2,F_2\in\overline{G}_{\Delta_2',\Delta_3},\ldots,E_n,F_n\in\overline{G}_{\Delta_n',\Delta_{n+1}}$ with

$$(E_l)_{KL}^{-+}=\min_{i\in K}\sum_{j\in L}(P_l)_{ij}\quad\text{and}\quad(F_l)_{KL}^{-+}=\max_{i\in K}\sum_{j\in L}(P_l)_{ij},$$

$\forall l\in\langle n\rangle,\ \forall K\in\Delta_l',\ \forall L\in\Delta_{l+1}$, where $\Delta_1':=\Delta_1$, such that

$$E_l\leq P_l\leq F_l,\quad\forall l\in\langle n\rangle.$$

It follows that

$$E_1E_2\ldots E_n\leq P_1P_2\ldots P_n\leq F_1F_2\ldots F_n$$

and, therefore,

$$\alpha(E_1E_2\ldots E_n)\leq\alpha(P_1P_2\ldots P_n)\leq\alpha(F_1F_2\ldots F_n).$$

Obviously, the conclusion of Theorem 1.8 is a special case of the latter sequence of above inequalities, namely, when $E_1E_2\ldots E_n$ and $F_1F_2\ldots F_n$ are stable matrices.

(d) By (c) we have, in particular,

$$P_1E_2\ldots E_n\leq P_1P_2\ldots P_n\leq P_1F_2\ldots F_n$$

and, therefore,

$$\alpha(P_1E_2\ldots E_n)\leq\alpha(P_1P_2\ldots P_n)\leq\alpha(P_1F_2\ldots F_n).$$

Suppose that $\Delta_2,\Delta_3,\ldots,\Delta_{n+1}$ and $E_2,F_2,E_3,F_3,\ldots,E_n,F_n$ are as in Theorem 1.8 and its proof, respectively. Further (see Theorem 1.6 and the proof of

Theorem 1.8),

$$P_1E_2\dots E_n=\begin{pmatrix}(P_1)_{\{1\}}E_2\dots E_n\\(P_1)_{\{2\}}E_2\dots E_n\\\vdots\\(P_1)_{\{m_1\}}E_2\dots E_n\end{pmatrix}=\begin{pmatrix}((P_1)_{\{1\}}E_2\dots E_n)^{-+((\{1\}),\Delta_{n+1})}\\((P_1)_{\{2\}}E_2\dots E_n)^{-+((\{2\}),\Delta_{n+1})}\\\vdots\\((P_1)_{\{m_1\}}E_2\dots E_n)^{-+((\{m_1\}),\Delta_{n+1})}\end{pmatrix}=$$

$$=\begin{pmatrix}((P_1)_{\{1\}})^{-+((\{1\}),\Delta_2)}E_2^{-+(\Delta_2,\Delta_3)}\dots E_n^{-+(\Delta_n,\Delta_{n+1})}\\((P_1)_{\{2\}})^{-+((\{2\}),\Delta_2)}E_2^{-+(\Delta_2,\Delta_3)}\dots E_n^{-+(\Delta_n,\Delta_{n+1})}\\\vdots\\((P_1)_{\{m_1\}})^{-+((\{m_1\}),\Delta_2)}E_2^{-+(\Delta_2,\Delta_3)}\dots E_n^{-+(\Delta_n,\Delta_{n+1})}\end{pmatrix}=$$

$$=\begin{pmatrix}((P_1)_{\{1\}})^{+\Delta_2}L_2\dots L_n\\((P_1)_{\{2\}})^{+\Delta_2}L_2\dots L_n\\\vdots\\((P_1)_{\{m_1\}})^{+\Delta_2}L_2\dots L_n\end{pmatrix}=P_1^{+\Delta_2}L_2\dots L_n.$$

We also have

$$P_1F_2\dots F_n=P_1^{+\Delta_2}U_2\dots U_n.$$

Finally, we obtain

$$\alpha(P_1^{+\Delta_2}L_2\dots L_n)\le\alpha(P_1P_2\dots P_n)\le\alpha(P_1^{+\Delta_2}U_2\dots U_n).$$

Obviously, the bounds $\alpha(P_1^{+\Delta_2}L_2\dots L_n)$ and $\alpha(P_1^{+\Delta_2}U_2\dots U_n)$ are better than $\alpha(E_1E_2\dots E_n)$ and $\alpha(F_1F_2\dots F_n)$ from (c), respectively.

Although below we do not give applications of Remark 1.9, this could be used to improve the convergence rate of our hybrid Metropolis-Hastings chain from the next section.

## 2. OUR HYBRID METROPOLIS-HASTINGS CHAIN

We design a hybrid Metropolis-Hastings chain on a known finite state space based on the $G$ method because this method can perform some interesting things, such as

1) to see if a finite Markov chain has a finite convergence time (see [17] and also Section 1);

2) to see if a product of stochastic (more generally, nonnegative) matrices is positive (see, e.g., Theorem 2.3 below; see also Theorem 1.6 (we can replace the stochastic (more generally, nonnegative) matrices with their incidence matrices));

3) to design positive products of stochastic matrices, the latter having certain given properties (see, e.g., this section);

4) to give bounds for the ergodicity coefficients $\alpha$ and $\overline{\alpha}$ of the products of stochastic matrices

(see also [17]) for other applications). The analysis of our hybrid Metropolis-Hastings chain is based on the $G$ method and ergodicity coefficients.

*Definition* 2.1 (see, e.g., [20, p. 80]). Let $P \in N_{m,n}$.

(a) We say that $P$ is a *row-allowable matrix* if it has at least one positive entry in each row.

(b) We say that $P$ is a *column-allowable matrix* if it has at least one positive entry in each column.

Below we also use notation from Theorem 1.8 and its proof (see also Remark 1.9). Let $S = \langle r \rangle$.

THEOREM 2.2. *Let $P_1, P_2, \ldots, P_t \in S_r$. Let $\Delta_1, \Delta_2, \ldots, \Delta_{t+1} \in \mathrm{Par}(S)$, $\Delta_1 = (S)$, $\Delta_{t+1} = (\{i\})_{i \in S}$. If $L_l$ (see Theorem* 1.8) *is a column-allowable matrix, $\forall l \in \langle t \rangle$, then $P_1 P_2 \ldots P_t > 0$.*

*Proof.* Obviously, $L_1 > 0$. It follows, by induction, that $L_1 L_2 \ldots L_t > 0$. Since (see the proof of Theorem 1.8)

$$(P_1 P_2 \ldots P_t)_{\{i\}} \geq (E_1 E_2 \ldots E_t)_{\{i\}} = L_1 L_2 \ldots L_t > 0, \quad \forall i \in S,$$

we have $P_1 P_2 \ldots P_t > 0$. $\square$

Let $P \in N_{m,n}$. Define

$$\overline{P} = (\overline{P}_{ij}) \in N_{m,n}, \quad \overline{P}_{ij} = \begin{cases} 1 & \text{if } P_{ij} > 0, \\ 0 & \text{if } P_{ij} = 0, \end{cases}$$

$\forall i \in \langle m \rangle$, $\forall j \in \langle n \rangle$. We call $\overline{P}$ the *incidence matrix of* $P$ (see, e.g., [6, p. 222]).

Let $\pi = (\pi_i)_{i \in S} = (\pi_1, \pi_2, \ldots, \pi_r)$ be a probability distribution on $S$. One way to sample approximately from $S$ is by means of the well-known Metropolis-Hastings chain ([10] and [5]). Let $Q \in S_r$ be an irreducible matrix such that $\overline{Q}$ is a symmetric matrix. Define

$$P = (P_{ij}) \in S_r, \quad P_{ij} = \begin{cases} 0 & \text{if } j \neq i \text{ and } Q_{ij} = 0, \\ Q_{ij} \min(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}) & \text{if } j \neq i \text{ and } Q_{ij} > 0, \\ 1 - \sum\limits_{k \neq i} P_{ik} & \text{if } j = i. \end{cases}$$

Since $\pi P = \pi$ ($\pi_i P_{ij} = \pi_j P_{ji}$, $\forall i, j \in S$, implies $\pi P = \pi$), we have $P^n \to e'\pi$ as $n \to \infty$. The Markov chain with transition matrix $P$ is called *Metropolis-Hastings* (see also [2–3] and [18] for some historical notes, open problems, and results on this chain).

Further, we define our hybrid Metropolis-Hastings chain.

Let $\Delta_1, \Delta_2, \ldots, \Delta_{t+1} \in \text{Par}(S)$ with $\Delta_1 = (S) \succ \Delta_2 \succ \cdots \succ \Delta_{t+1} = (\{i\})_{i\in S}$. (We set $\Delta \succ \Delta'$ if $\Delta' \preceq \Delta$ and $\Delta' \neq \Delta$, where $\Delta, \Delta' \in \text{Par}(E)$, see Section 1.) Let $Q_1, Q_2, \ldots, Q_t \in S_r$ such that

(C1) $\overline{Q}_1, \overline{Q}_2, \ldots, \overline{Q}_t$ are symmetric matrices;

(C2) $(Q_l)_K^L = 0$, $\forall l \in \langle t\rangle - \{1\}$, $\forall K, L \in \Delta_l$, $K \neq L$ (this assumption implies that $Q_2, Q_3, \ldots, Q_t$ are block diagonal matrices);

(C3) $(Q_l)_K^U$ is a row-allowable matrix, $\forall l \in \langle t\rangle$, $\forall K \in \Delta_l$, $\forall U \in \Delta_{l+1}$, $U \subseteq K$.

Although $Q_l$, $l \in \langle t\rangle$, are not irreducible matrices if $l \geq 2$, we define the matrices $P_l$, $l \in \langle t\rangle$, as in the Metropolis-Hastings case, namely,

$$P_l = ((P_l)_{ij}) \in S_r,\ (P_l)_{ij} = \begin{cases} 0 & \text{if } j \neq i \text{ and } (Q_l)_{ij} = 0, \\ (Q_l)_{ij} \min(1, \frac{\pi_j (Q_l)_{ji}}{\pi_i (Q_l)_{ij}}) & \text{if } j \neq i \text{ and } (Q_l)_{ij} > 0, \\ 1 - \sum\limits_{k\neq i} (P_l)_{ik} & \text{if } j = i, \end{cases}$$

$\forall l \in \langle t\rangle$. Set $P = P_1 P_2 \ldots P_t$.

Theorem 2.3. *Concerning $P$ above we have $\pi P = \pi$ and $P > 0$.*

*Proof.* Since $\pi_i (P_l)_{ij} = \pi_j (P_l)_{ji}$, $\forall l \in \langle t\rangle$, $\forall i, j \in S$, we have $\pi P_l = \pi$, $\forall l \in \langle t\rangle$. Further, $\pi P_l = \pi$, $\forall l \in \langle t\rangle$, implies $\pi P = \pi$. By Theorem 2.2, $P = P_1 P_2 \ldots P_t > 0$. $\square$

By Theorem 2.3, $P^n \to e'\pi$ as $n \to \infty$. $P$ determines a Markov chain; we call this chain the *hybrid Metropolis-Hastings chain*. In particular, we call this chain the *hybrid Metropolis chain* when $Q_1, Q_2, \ldots, Q_t$ are symmetric matrices.

We need the next result for the analysis of hybrid Metropolis-Hastings chain.

Theorem 2.4. (*A less or more known result.*) *Let $P \in S_r$ be an aperiodic irreducible matrix. Consider a Markov chain with transition matrix $P$ and limit probability distribution $\pi$. Let $p_n$ be the probability distribution of chain at time $n$, $\forall n \geq 0$. Then*

$$\|p_n - \pi\|_1 \leq 2\overline{\alpha}(P^n).$$

*Proof.* (The proof is less or more known.) Let $\mu$ and $\nu$ be two probability distributions on $\langle m\rangle$ and $Q \in S_{m,n}$. It is known that

$$\|\mu Q - \nu Q\|_1 \leq \|\mu - \nu\|_1 \overline{\alpha}(Q)$$

(see, e.g., [7, p. 147]). Using the above result, $p_n = p_0 P^n$, and $\pi P = \pi$, we have

$$\|p_n - \pi\|_1 = \|p_0 P^n - \pi P^n\|_1 \leq \|p_0 - \pi\|_1 \overline{\alpha}(P^n) \leq 2\overline{\alpha}(P^n). \quad \square$$

Now, bounds for $\|p_n - \pi\|_1$ of the hybrid Metropolis-Hastings chain can follow from Theorems 1.8 and 2.4 and Remark 1.9. This analysis can be realized for each hybrid chain or for certain collections of hybrid chains. Obviously, it is necessary that $t$ be as small as possible.

Suppose that $\Delta_l = \left(K_1^{(l)}, K_2^{(l)}, \ldots, K_{u_l}^{(l)}\right)$, $\forall l \in \langle t+1 \rangle$. Below we consider a case which can be analyzed more easily than the others, namely, that satisfying, moreover, the conditions:

(c1) $|K_1^{(l)}| = |K_2^{(l)}| = \cdots = |K_{u_l}^{(l)}|$, $\forall l \in \langle t+1 \rangle$ with $u_l \geq 2$;

(c2) $r = r_1 r_2 \ldots r_t$ with $r_1 r_2 \ldots r_l = |\Delta_{l+1}|$, $\forall l \in \langle t-1 \rangle$, and $r_t = |K_1^{(t)}|$ (this is compatible with $\Delta_1 \succ \Delta_2 \succ \cdots \succ \Delta_{t+1}$);

(c3) (c3.1) $Q_l$ is a symmetric matrix such that (c3.2) $(Q_l)_{ii} > 0$, $\forall i \in S$, and $(Q_l)_{i_1 j_1} = (Q_l)_{i_2 j_2}$, $\forall i_1, i_2, j_1, j_2 \in S$ with $i_1 \neq j_1, i_2 \neq j_2$, and $(Q_l)_{i_1 j_1}$, $(Q_l)_{i_2 j_2} > 0$, $\forall l \in \langle t \rangle$ ((c3.2) says, to put it otherwise, that all the positive entries of $Q_l$, excepting the entries $(Q_l)_{ii}$, $\forall i \in S$, are equal, $\forall l \in \langle t \rangle$);

(c4) $(Q_l)_K^U$ has in each row just one positive entry, $\forall l \in \langle t \rangle$, $\forall K \in \Delta_l$, $\forall U \in \Delta_{l+1}$ with $U \subseteq K$ (this is compatible with (c3.1) because $(Q_l)_V^W$ is a square matrix, $\forall l \in \langle t \rangle$, $\forall V, W \in \Delta_{l+1}$).

By (C2), (c1), (c2), and (c4),

$$|\{j \mid j \in S \text{ and } (Q_l)_{ij} > 0\}| = r_l, \quad \forall i \in S, \ \forall l \in \langle t \rangle,$$

and by (c1) and (c2),

$$|K_1^{(l+1)}| = \frac{r}{|\Delta_{l+1}|} = \frac{r_1 r_2 \ldots r_t}{r_1 r_2 \ldots r_l} = r_{l+1} r_{l+2} \cdots r_t, \quad \forall l \in \langle t-1 \rangle.$$

To give an example for which the conditions (C1)–(C3) and (c1)–(c4) hold, we consider $S = \langle 6 \rangle$ ($r = 6 = 3 \cdot 2$), $\Delta_1 = (\langle 6 \rangle)$, $\Delta_2 = (\{1,2\}, \{3,4\}, \{5,6\})$, $\Delta_3 = (\{i\})_{i \in \langle 6 \rangle} = (\{1\}, \{2\}, \ldots, \{6\})$,

$$Q_1 = \begin{pmatrix} \frac{2}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{2}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{2}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{2}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{2}{4} & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{2}{4} \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

Further, we choose the positive entries of matrices $Q_l$, $l \in \langle t \rangle$; these are not choose at random (it is interesting to bear in mind this fact). Let $l \in \langle t \rangle$. Set

$$f_l = \min_{i,j \in S, \ (Q_l)_{ij} > 0} \frac{\pi_j}{\pi_i}, \quad g_l = \frac{1}{f_l},$$

and (see (c3) again)

$$x_l = (Q_l)_{ij},$$

where $i, j \in S$ are fixed such that $i \neq j$ and $(Q_l)_{ij} > 0$. Obviously, $f_l \leq 1$, $g_l \geq 1$, and

$$g_l = \max_{i,j \in S, (Q_l)_{ij} > 0} \frac{\pi_j}{\pi_i}.$$

We have

$$(P_l)_{ii} = 1 - \sum_{j \neq i} (P_l)_{ij} \geq 1 - (r_l - 1)x_l.$$

We impose

$$1 - (r_l - 1)x_l \geq x_l f_l.$$

It follows that

$$x_l \leq \frac{1}{f_l + r_l - 1}.$$

We choose the matrix $Q_l$ such that

$$x_l = \frac{1}{f_l + r_l - 1}.$$

To see that this choice is possible, we need to prove that $(Q_l)_{ii} > 0$ (see (c3) above). Indeed,

$$(Q_l)_{ii} = 1 - \sum_{j \in S, j \neq i} (Q_l)_{ij} = 1 - (r_l - 1)x_l = 1 - \frac{r_l - 1}{f_l + r_l - 1} > 0.$$

Below we give another main result of this article.

THEOREM 2.5. *Under the above assumptions we have*
(i) $\overline{\alpha}(P) \leq 1 - r\frac{f_1}{f_1+r_1-1}\frac{f_2}{f_2+r_2-1}\cdots\frac{f_t}{f_t+r_t-1} =$
$= 1 - r\frac{1}{1+(r_1-1)g_1}\frac{1}{1+(r_2-1)g_2}\cdots\frac{1}{1+(r_t-1)g_t}$;
(ii) $\|p_n - \pi\|_1 \leq 2(1 - r\frac{f_1}{f_1+r_1-1}\frac{f_2}{f_2+r_2-1}\cdots\frac{f_t}{f_t+r_t-1})^n =$
$= 2(1 - r\frac{1}{1+(r_1-1)g_1}\frac{1}{1+(r_2-1)g_2}\cdots\frac{1}{1+(r_t-1)g_t})^n, \ \forall n \geq 0,$
*where $p_n$ is the probability distribution at time $n$ of the hybrid Metropolis chain with the transition matrix $P = P_1P_2\ldots P_t$, $\forall n \geq 0$;*

(iii) *an upper bound of the minimum number $n$ of steps such that $\|p_n - \pi\|_1 \leq \varepsilon$ is $[(\ln\frac{\varepsilon}{2})/(\ln\overline{a}(P))]$ if $0 < \varepsilon \leq 2$ (recall that $\|p_n - \pi\|_1 \leq 2$, $\forall n \geq 0$), $\overline{a}(P) > 0$, where*

$$\overline{a}(P) := 1 - r\frac{f_1}{f_1 + r_1 - 1}\frac{f_2}{f_2 + r_2 - 1}\cdots\frac{f_t}{f_t + r_t - 1}.$$

*Proof.* (i) The inequality follows from Remark 1.7 and Theorem 1.8 (we have

$$(P_l)_{ij} \geq x_l f_l = \frac{f_l}{f_l + r_l - 1}, \quad \forall i, j \in S \text{ with } (P_l)_{ij} > 0, \ \forall l \in \langle t \rangle).$$

The equation is obvious.

(ii) It follows from Theorem 2.4, (i), and the well-known inequality

$$\overline{\alpha}(T_1T_2) \leq \overline{\alpha}(T_1)\overline{\alpha}(T_2), \quad \forall T_1 \in S_{m_1,m_2},\ \forall T_2 \in S_{m_2,m_3}$$

([4]; see, e.g., also [6, p. 58] or [7, p. 145]).

(iii) We impose $2(\overline{\alpha}(P))^n \leq \varepsilon$. Then $\|p_n - \pi\|_1 \leq \varepsilon$ if

$$n \geq [(\ln \tfrac{\varepsilon}{2})/(\ln \overline{\alpha}(P))]. \quad \square$$

*Remark* 2.6. (a) Theorem 2.5(ii) gives a geometric convergence rate of the hybrid Metropolis chain. Obviously, the chain has a fast convergence if $f_l$ is close to 1, $\forall l \in \langle t\rangle$.

(b) By Theorem 2.5(ii), if $\pi = (\pi_i)_{i\in S}$ is the uniform distribution on $S = \langle r\rangle$ (in this case, $f_l = 1$, $\forall l \in \langle t\rangle$), then $\|p_1 - \pi\|_1 = 0$ (i.e., we have an exact sampling from uniform distribution in one step due to $P$ or, equivalently, in $t$ steps due to $P_1, P_2, \ldots, P_t$).

(c) The bound from Theorem 2.5(ii) could not be the best one; this is an open problem.

To speed up the hybrid Metropolis-Hastings chain, we can replace the product $P_{s+1}P_{s+2}\cdots P_t$ $(1 \leq s < t)$ from $P = P_1P_2\ldots P_sP_{s+1}\cdots P_t$ by the $\Delta_{s+1}$-stable matrix (recall that $\Delta_l = (K_1^{(l)}, K_2^{(l)}, \ldots, K_{u_l}^{(l)})$, $\forall l \in \langle t+1\rangle$)

$$P^* = P^*(s) := \begin{pmatrix} A_1^{*(s+1)} & & & \\ & A_2^{*(s+1)} & & \\ & & \ddots & \\ & & & A_{u_{s+1}}^{*(s+1)} \end{pmatrix},$$

where

$$A_z^{*(s+1)} := \frac{1}{\sum\limits_{i\in K_z^{(s+1)}} \pi_i} e'(|K_z^{(s+1)}|)(\pi_i)_{i\in K_z^{(s+1)}}, \quad \forall z \in \langle u_{s+1}\rangle$$

(recall that $e = e(n) = (1,1,\ldots,1) \in \mathbf{R}^n$ and $e'$ is its transpose). E.g., if $S = \langle 8\rangle$, $\Delta_1 = (\langle 8\rangle)$, $\Delta_2 = (\{1,2,3,4\},\{5,6,7,8\})$, $\Delta_3 = (\{1,2\},\{3,4\},\{5,6\},\{7,8\})$, $\Delta_4 = (\{i\})_{i\in\langle 8\rangle} = (\{1\},\{2\},\ldots,\{8\})$, then, for $s = 2$, $P = P_1P_2P^*$ with

$$P^* = P^*(2) = \begin{pmatrix} \frac{1}{a_1^{(3)}}W_1^{*(3)} & 0 & 0 & 0 \\ 0 & \frac{1}{a_2^{(3)}}W_2^{*(3)} & 0 & 0 \\ 0 & 0 & \frac{1}{a_3^{(3)}}W_3^{*(3)} & 0 \\ 0 & 0 & 0 & \frac{1}{a_4^{(3)}}W_4^{*(3)} \end{pmatrix},$$

where $a_1^{(3)} := \pi_1 + \pi_2$, $a_2^{(3)} := \pi_3 + \pi_4$, $a_3^{(3)} := \pi_5 + \pi_6$, $a_4^{(3)} := \pi_7 + \pi_8$, $W_1^{*(3)} := e'(2)(\pi_1, \pi_2)$, $W_2^{*(3)} := e'(2)(\pi_3, \pi_4)$, $W_3^{*(3)} := e'(2)(\pi_5, \pi_6)$, and $W_4^{*(3)} := e'(2)(\pi_7, \pi_8)$ (obviously, $A_k^{*(3)} = \frac{1}{a_k^{(3)}} W_k^{*(3)}$, $\forall k \in \langle 4 \rangle$), while, for $s = 1$, $P = P_1 P^*$ with

$$P^* = P^*(1) = \begin{pmatrix} \frac{1}{a_1^{(2)}} W_1^{*(2)} & 0 \\ 0 & \frac{1}{a_2^{(2)}} W_2^{*(2)} \end{pmatrix},$$

where $a_1^{(2)} := \pi_1 + \pi_2 + \pi_3 + \pi_4$, $a_2^{(2)} := \pi_5 + \pi_6 + \pi_7 + \pi_8$, $W_1^{*(2)} := e'(4)(\pi_1, \pi_2, \pi_3, \pi_4)$, and $W_2^{*(2)} := e'(4)(\pi_5, \pi_6, \pi_7, \pi_8)$. Moreover, setting

$$C_1 = \begin{pmatrix} \frac{1}{a_1^{(2)}} U_1^{*(2)} & 0 \\ 0 & \frac{1}{a_2^{(2)}} U_2^{*(2)} \end{pmatrix},$$

where $U_1^{*(2)} := e'(4)(\pi_1 + \pi_2, 0, \pi_3 + \pi_4, 0)$, and $U_2^{*(2)} := e'(4)(\pi_5 + \pi_6, 0, \pi_7 + \pi_8, 0)$, we have $P^*(1) = C_1 P^*(2)$. (This result can be generalized easily.) Consequently, we can work with $C_1$ and $P^*(2)$ instead of $P^*(1)$. This is an interesting thing because the matrices $C_1$ and $P^*(2)$ are sparser than $P^*(1)$.

To unify the theory of the hybrid chain with right $\Delta$-stable matrices as above and that without right $\Delta$-stable matrices as above, below we consider that $s \in \langle t \rangle$ and set $r_{t+1} = 1$,

$$P = P(s) = \begin{cases} P_1 P_2 \ldots P_t & \text{if } s = t, \\ P_1 P_2 \ldots P_s P^* & \text{if } 1 \leq s < t, \end{cases}$$

and

$$\overline{\alpha}(P) = \begin{cases} 1 - r \frac{f_1}{f_1 + r_1 - 1} \frac{f_2}{f_2 + r_2 - 1} \cdots \frac{f_t}{f_t + r_t - 1} & \text{if } P = P_1 P_2 \ldots P_t, \\ 1 - r \frac{f_1}{f_1 + r_1 - 1} \frac{f_2}{f_2 + r_2 - 1} \cdots \frac{f_s}{f_s + r_s - 1} \frac{1}{r_{s+1} r_{s+2} \cdots r_{t+1}} & \text{if } P = P_1 P_2 \ldots P_s P^*. \end{cases}$$

The two results below are generalizations of Theorems 2.3 and 2.5, respectively.

THEOREM 2.7. *Concerning $P$ above we have $\pi P = \pi$ and $P > 0$.*

*Proof.* See the proof of Theorem 2.3. $\square$

THEOREM 2.8. *Concerning $P$ above we have (recall that $|K_1^{(l+1)}| = r_{l+1} r_{l+2} \cdots r_{t+1}$, $\forall l \in \langle t \rangle$)*

(i) $\overline{\alpha}(P) \leq 1 - r \frac{f_1}{f_1 + r_1 - 1} \frac{f_2}{f_2 + r_2 - 1} \cdots \frac{f_s}{f_s + r_s - 1} \frac{1}{r_{s+1} r_{s+2} \cdots r_{t+1}} =$
$= 1 - r \frac{1}{1 + (r_1 - 1) g_1} \frac{1}{1 + (r_2 - 1) g_2} \cdots \frac{1}{1 + (r_s - 1) g_s} \frac{1}{r_{s+1} r_{s+2} \cdots r_{t+1}}$;

(ii) $\|p_n - \pi\|_1 \leq 2(1 - r \frac{f_1}{f_1 + r_1 - 1} \frac{f_2}{f_2 + r_2 - 1} \cdots \frac{f_s}{f_s + r_s - 1} \frac{1}{r_{s+1} r_{s+2} \cdots r_{t+1}})^n =$

$$= 2(1 - r\tfrac{1}{1+(r_1-1)g_1}\tfrac{1}{1+(r_2-1)g_2}\cdots\tfrac{1}{1+(r_s-1)g_s}\tfrac{1}{r_{s+1}r_{s+2}\cdots r_{t+1}})^n,\ \forall n \geq 0,$$

*where $p_n$ is the probability distribution at time $n$ of the hybrid Metropolis chain with transition matrix $P$, $\forall n \geq 0$;*

(iii) *an upper bound of the minimum number $n$ of steps such that $\|p_n - \pi\|_1 \leq \varepsilon$ is $[(\ln \frac{\varepsilon}{2})/(\ln \overline{a}(P))]$ if $0 < \varepsilon \leq 2$, $\overline{a}(P) > 0$.*

*Proof.* See the proof of Theorem 2.5. (We use the partitions $\Delta_1 = (S), \Delta_2, \ldots, \Delta_{s+1}, (\{i\})_{i\in S}$ if $1 \leq s < t$; it is easy to compute $(P^*)^{-+(\Delta_{s+1},(\{i\})_{i\in S})}$.) $\square$

*Remark* 2.9. If $1 \leq s_1 < s_2 \leq t$, then

$$1 - r\frac{1}{1+(r_1-1)g_1}\frac{1}{1+(r_2-1)g_2}\cdots\frac{1}{1+(r_{s_1}-1)g_{s_1}}\frac{1}{r_{s_1+1}r_{s_1+2}\cdots r_{t+1}} \leq$$
$$\leq 1 - r\frac{1}{1+(r_1-1)g_1}\frac{1}{1+(r_2-1)g_2}\cdots\frac{1}{1+(r_{s_2}-1)g_{s_2}}\frac{1}{r_{s_2+1}r_{s_2+2}\cdots r_{t+1}}.$$

Consequently, the convergence rate obtained for $s_1$ is better than that for $s_2$.

## 3. SPECIAL CASES

Below we consider two special cases of our hybrid Metropolis-Hastings chain; one refer to $S = \mathbb{S}_n$, where $\mathbb{S}_n$ is the set of permutations of order $n$, and the other to $S = \{0, 1, \ldots, h\}^n$, $h, n \geq 1$. (In Sections 1 and 2 we used $S = \langle r\rangle$ for simplification; $S$ can be any finite set.) These special cases can easily be implemented on a computer.

First, we consider $S = \mathbb{S}_n$. In this case, $r := |S| = n!$. Let $\mathbb{A}_n^l$ be the set of arrangements using $l$ of $n$ objects, $\forall l \in \langle n\rangle$. Set (see [17])

$$K_{(i_1,i_2,\ldots,i_l)} = \{\sigma \mid \sigma \in \mathbb{S}_n \text{ and } \sigma(s) = i_s,\ \forall s \in \langle l\rangle\}, \quad \forall l \in \langle n-1\rangle,$$
$$\Delta_1 = (\mathbb{S}_n),$$

and

$$\Delta_{l+1} = (K_{(i_1,i_2,\ldots,i_l)})_{(i_1,i_2,\ldots,i_l)\in\mathbb{A}_n^l}, \quad \forall l \in \langle n-1\rangle.$$

Obviously, we have $\Delta_n = (\{\sigma\})_{\sigma\in\mathbb{S}_n}$ and $r_1 = n$, $r_2 = n-1, \ldots, r_{n-1} = 2$. (We can also study other decompositions of $n!$, such as $r_1 = n(n-1)$, $r_2 = (n-2)(n-3), \ldots, r_k = 3\cdot 2$ if $n = 2k+1 \geq 5$ and $r_1 = n(n-1)$, $r_2 = (n-2)(n-3), \ldots, r_{k-1} = 4\cdot 3\cdot 2$ if $n = 2k \geq 6$.)

Define the matrices $Q_l = ((Q_l)_{\sigma\tau})$, $l \in \langle n-1\rangle$, by (see above of Theorem 2.5; see also [17])

$$(Q_l)_{\sigma\tau} = \begin{cases} \frac{1}{f_l+n-l} & \text{if } \tau = \sigma\circ(l,m) \text{ for some } m \in \{l+1, l+2, \ldots, n\}, \\ 0 & \text{if } \tau \neq \sigma\circ(l,m),\ \forall m \in \{l+1, l+2, \ldots, n\}, \\ 1 - \sum\limits_{\gamma\neq\sigma}(Q_l)_{\sigma\gamma} & \text{if } \tau = \sigma, \end{cases}$$

$\forall l \in \langle n-1\rangle$, $\forall \sigma, \tau \in \mathbb{S}_n$ ($(l, m)$ is a transposition). Further, we have (see Section 2; $t = n-1$, $s \in \langle n-1\rangle$, and the matrices $Q_l$, $l \in \langle n-1\rangle$, are symmetric)

$$P_l = ((P_l)_{\sigma\tau}), \quad (P_l)_{\sigma\tau} = \begin{cases} (Q_l)_{\sigma\tau} \min(1, \frac{\pi_\tau}{\pi_\sigma}) & \text{if } \tau \neq \sigma, \\ 1 - \sum\limits_{\gamma \neq \sigma} (P_l)_{\sigma\gamma} & \text{if } \tau = \sigma, \end{cases}$$

$\forall l \in \langle n-1\rangle$, $\forall \sigma, \tau \in \mathbb{S}_n$, and

$$P = \begin{cases} P_1 P_2 \dots P_{n-1} & \text{if } s = n-1, \\ P_1 P_2 \dots P_s P^* & \text{if } 1 \leq s < n-1. \end{cases}$$

Then, by Theorem 2.8(ii), we have

$$\|p_u - \pi\|_1 \leq 2 \left(1 - \frac{n!}{(n-s)!} \frac{f_1}{f_1 + n - 1} \frac{f_2}{f_2 + n - 2} \cdots \frac{f_s}{f_s + n - s}\right)^u =$$
$$= 2 \left(1 - \frac{n!}{(n-s)!} \frac{1}{1 + (n-1)g_1} \frac{1}{1 + (n-2)g_2} \cdots \frac{1}{1 + (n-s)g_s}\right)^u, \quad \forall u \geq 0.$$

If $f_1, f_2, \dots, f_{n-1} \geq f$ for some $f \in (0, 1]$, then $g_1, g_2, \dots, g_{n-1} \leq g := \frac{1}{f}$ ($f$ and $g$ depend or not on $n$ ). In this case, we have

$$\|p_u - \pi\|_1 \leq 2 \left(1 - \frac{n!}{(n-s)!} \frac{f}{f + n - s} \frac{f}{f + n - s + 1} \cdots \frac{f}{f + n - 1}\right)^u =$$
$$= 2 \left(1 - \frac{n!}{(n-s)!} \frac{1}{1 + (n-s)g} \frac{1}{1 + (n-s+1)g} \cdots \frac{1}{1 + (n-1)g}\right)^u, \quad \forall u \geq 0,$$

and, moreover, we can replace $f_l$ by $f$ in the definition of $Q_l$ above, $\forall l \in \langle n-1\rangle$ (see above of Theorem 2.5 again).

If $\pi = (\pi_\sigma)_{\sigma \in \mathbb{S}_n}$ is the uniform distribution on $\mathbb{S}_n$, we can take $f = 1$ and, with this choice, we have $\|p_1 - \pi\|_1 = 0$ (one step due to $P$ or, equivalently, $n-1$ steps due to $P_1, P_2, \dots, P_{n-1}$ (the latter is a well-known result, see, e.g., [8, pp. 139–141], [17], and [19])).

*Remark* 3.1. It is easy to obtain two bounds for

$$1 - \frac{n!}{(n-s)!} \frac{1}{1 + (n-s)g} \frac{1}{1 + (n-s+1)g} \cdots \frac{1}{1 + (n-1)g};$$

more exactly, from

$$\frac{n!}{(n-s)!} \frac{1}{1 + (n-s)g} \frac{1}{1 + (n-s+1)g} \cdots \frac{1}{1 + (n-1)g} \leq$$
$$\leq \frac{n!}{(n-s)!} \frac{1}{(n-s)g} \frac{1}{(n-s+1)g} \cdots \frac{1}{(n-1)g} = \frac{n}{n-s} \frac{1}{g^s}$$

and (since $g \geq 1$)

$$\frac{n!}{(n-s)!}\frac{1}{1+(n-s)g}\frac{1}{1+(n-s+1)g}\cdots\frac{1}{1+(n-1)g} \geq$$
$$\geq \frac{n!}{(n-s)!}\frac{1}{(n-s+1)g}\frac{1}{(n-s+2)g}\cdots\frac{1}{ng} = \frac{1}{g^s}$$

we have

$$1 - \frac{n}{n-s}\frac{1}{g^s} \leq$$
$$\leq 1 - \frac{n!}{(n-s)!}\frac{1}{1+(n-s)g}\frac{1}{1+(n-s+1)g}\cdots\frac{1}{1+(n-1)g} \leq 1 - \frac{1}{g^s}$$

THEOREM 3.2. *If $g$ and $s$ are fixed ($g, s \geq 1$), then*

$$\lim_{n\to\infty}\left(1 - \frac{n!}{(n-s)!}\frac{1}{1+(n-s)g}\frac{1}{1+(n-s+1)g}\cdots\frac{1}{1+(n-1)g}\right) = 1 - \frac{1}{g^s}.$$

*Proof.* See Remark 3.1. □

To give some numerical results (with their error estimates), we consider

$$\pi_\sigma = \frac{\theta^{d(\sigma,\sigma_0)}}{Z}, \quad \forall \sigma \in \mathbb{S}_n,$$

where $0 < \theta \leq 1$, $\sigma_0 \in \mathbb{S}_n$, $d(\sigma, \sigma_0)$ = minimum number of transpositions required to bring $\sigma$ to $\sigma_0$, and $Z = \sum_{\sigma\in\mathbb{S}_n} \theta^{d(\sigma,\sigma_0)}$ (see, e.g., [1, p. 104], or [2], or [3]). $d$ is called the *Cayley metric* and this construction is called the *Mallows model through Cayley metric* (see, e.g., [1, Chapter 6] for other examples of metrics on $\mathbb{S}_n$ and, therefore, other examples of Mallows models). In this case, we take $f = \theta$ and, therefore, $g = \frac{1}{\theta}$ because $f_1 = f_2 = \cdots = f_{n-1} = \theta$ and, therefore, $g_1 = g_2 = \cdots = g_{n-1} = \frac{1}{\theta}$ (note that $f$ and $g$ do not depend on $n$); $f_1 = f_2 = \cdots = f_{n-1} = \theta$ because

$$|d(\sigma, \sigma_0) - d(\tau, \sigma_0)| \leq 1,$$

$\forall \sigma \in \mathbb{S}_n$, $\forall \tau \in \mathbb{S}_n$ with $\tau = \sigma \circ (l, m)$ for some transposition $(l, m)$, $l \neq m$ (we use the fact that $d$ is metric and $d(\sigma, \tau) = 1$, $\forall \sigma \in \mathbb{S}_n$, $\forall \tau \in \mathbb{S}_n$ with $\tau$ as above) and

$$|d(\sigma, \sigma_0) - d(\tau, \sigma_0)| = 1$$

when $\tau$ is as above and, e.g., $\sigma = \sigma_0$. We also note that $\min_{\sigma,\tau\in\mathbb{S}_n} \frac{\pi_\sigma}{\pi_\tau} = \theta^{\frac{n(n-1)}{2}}$ and $\max_{\sigma,\tau\in\mathbb{S}_n} \frac{\pi_\sigma}{\pi_\tau} = (\frac{1}{\theta})^{\frac{n(n-1)}{2}}$.

Now, if, e.g., $\theta = \frac{1}{2}$, we have $g = 2$. For $n = 30$ ($30! \simeq 2.6525 \cdot 10^{32}$; in this case, the greatest value of $\pi_\sigma$ is $\frac{1}{Z}$ and the smallest is $\frac{1}{2^{435}Z}$) and $s = 12$ ($s = 12$ could not be the best choice), we have, e.g.,

$$\|p_u - \pi\|_1 \leq 0.086255\cdots \text{ if } u = 10000$$

(10000 steps due to $P$ or, equivalently, $13 \cdot 10000 = 130000$ steps due to $P_1, P_2, \ldots, P_{12}, P^*$) and

$$\|p_u - \pi\|_1 \leq 0.00371996\cdots \text{ if } u = 20000.$$

For $n = 40$ ($40! \simeq 8.1592 \cdot 10^{47}$) and $s = 22$, we have, e.g.,

$$\|p_u - \pi\|_1 \leq 0.0580266\cdots \text{ if } u = 10^7$$

($10^7$ steps due to $P$ or, equivalently, $23 \cdot 10^7$ steps due to $P_1, P_2, \ldots, P_{22}, P^*$) and

$$\|p_u - \pi\|_1 \leq 0.00988384\cdots \text{ if } u = 15 \cdot 10^6.$$

We have not used the result from Theorem 2.8(iii). Obviously, this can easily be made for any given error $0 < \varepsilon \leq 2$. We also note that estimating the errors is a very large gap in Markov chain Monte Carlo theory.

Second, we consider $S = \{0, 1, \ldots, h\}^n$, $h, n \geq 1$. This case is interesting to analyze because, e.g., the Ising model (with an external field or not) is defined on $S = \{0, 1\}^n$ while the Potts model and grey-scale images (see, e.g., [9, Chapter 6]) are even defined on $S = \{0, 1, \ldots, h\}^n$. Using the notation from Section 2, we have $r := |S| = (1 + h)^n$. If $n = n_1 + n_2 + \cdots + n_w$, then $r = (1 + h)^{n_1}(1 + h)^{n_2} \cdots (1 + h)^{n_w}$ and, consequently, we can take $r_i = (1 + h)^{n_i}$, $\forall i \in \langle w\rangle$. Below we only consider the case $n_1 = n_2 = \cdots = n_w := v$, $r_i = (1 + h)^v$, $\forall i \in \langle w\rangle$. In this case, $n = vw$.

Set

$$K_{(x_1,x_2,\ldots,x_{vl})} = \{(y_1, y_2, \ldots, y_n) \mid (y_1, y_2, \ldots, y_n) \in S \text{ and } y_i = x_i, \forall i \in \langle vl\rangle\},$$

$\forall l \in \langle w\rangle$, $\forall x_1, x_2, \ldots, x_{vl} \in \{0, 1, \ldots, h\}$ (obviously,

$$K_{(x_1,x_2,\ldots,x_{vw})} = K_{(x_1,x_2,\ldots,x_n)} = \{(x_1, x_2, \ldots, x_n)\}),$$

$$\Delta_1 = (S),$$

and

$$\Delta_{l+1} = (K_{(x_1,x_2,\ldots,x_{vl})})_{x_1,x_2,\ldots,x_{vl} \in \{0,1,\ldots,h\}}, \quad \forall l \in \langle w\rangle.$$

Define the symmetric matrices $Q_l$, $l \in \langle w \rangle$, by ($x = (x_1, x_2, \ldots, x_n)$, etc.)

$$(Q_l)_{xy} = \begin{cases} \frac{1}{f_l+(1+h)^v-1} & \text{if } (y_1, y_2, \ldots, y_{v(l-1)}) = (x_1, x_2, \ldots, x_{v(l-1)}) \text{ when } \\ & 1 \le v(l-1),\ (y_{vl+1}, y_{vl+2}, \ldots, y_n) = (x_{vl+1}, \\ & x_{vl+2}, \ldots, x_n) \text{ when } vl+1 \le n, \text{ and } x \ne y, \\ 0 & \text{if } (y_1, y_2, \ldots, y_{v(l-1)}) \ne (x_1, x_2, \ldots, x_{v(l-1)}) \text{ when } \\ & 1 \le v(l-1) \text{ or } (y_{vl+1}, y_{vl+2}, \ldots, y_n) \ne (x_{vl+1}, \\ & x_{vl+2}, \ldots, x_n) \text{ when } vl+1 \le n, \\ 1 - \sum\limits_{z \ne x} (Q_l)_{xz} & \text{if } x = y, \end{cases}$$

$\forall l \in \langle w \rangle$, $\forall x, y \in S$. Further, we have (see Section 2)

$$P_l = ((P_l)_{xy}),\ (P_l)_{xy} = \begin{cases} (Q_l)_{xy} \min(1, \frac{\pi_y}{\pi_x}) & \text{if } y \ne x, \\ 1 - \sum\limits_{z \ne x} (P_l)_{xz} & \text{if } y = x, \end{cases}$$

$\forall l \in \langle w \rangle$, $\forall x, y \in S$, and

$$P = \begin{cases} P_1 P_2 \ldots P_w & \text{if } s = w, \\ P_1 P_2 \ldots P_s P^* & \text{if } 1 \le s < w. \end{cases}$$

Then, by Theorem 2.8(ii), we have ($s \in \langle w \rangle$)

$$\|p_u - \pi\|_1 \le 2 \left(1 - (1+h)^{sv} \frac{f_1}{f_1 + (1+h)^v - 1} \cdots \frac{f_s}{f_s + (1+h)^v - 1}\right)^u =$$

$$= 2 \left(1 - (1+h)^{sv} \frac{1}{1 + ((1+h)^v - 1) g_1} \cdots \frac{1}{1 + ((1+h)^v - 1) g_s}\right)^u, \quad \forall u \ge 0.$$

If $f_1, f_2, \ldots, f_w \ge f$ for some $f \in (0, 1]$ and, consequently, $g_1, g_2, \ldots, g_w \le g := \frac{1}{f}$, then

$$\|p_u - \pi\|_1 \le 2 \left(1 - (1+h)^{sv} \left(\frac{f}{f + (1+h)^v - 1}\right)^s\right)^u =$$

$$= 2 \left(1 - (1+h)^{sv} \frac{1}{(1 + ((1+h)^v - 1) g)^s}\right)^u, \quad \forall u \ge 0,$$

and, moreover, we can replace $f_l$ by $f$ in the definition of $Q_l$ above, $\forall l \in \langle w \rangle$ (see above of Theorem 2.5 again).

Recall (see Remark 2.6(b)) that if $\pi = (\pi_i)_{i \in S}$ is the uniform distribution on $S = \{0, 1, \ldots, h\}^n$ (in this case, $f_1 = f_2 = \cdots = f_w = 1$), then $\|p_1 - \pi\|_1 = 0$.

To give some numerical results, we consider the Ising model (see, e.g., [9, Chapter 6]) on the $m_1 \times m_2$ grid. This grid is a graph $G = (V, E)$, where

$$V := \{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\}$$

is the vertex set and

$$E := \{(a,b) \mid a = (a_1,a_2),\ b = (b_1,b_2) \in V \text{ and } a_1 = b_1,\ a_2 - b_2 = -1 \\ \text{or } a_1 - b_1 = -1,\ a_2 = b_2\}$$

is the edge set. To define the Ising model on this grid, we must consider $S = \{0,1\}^{|V|}$; in this case, $r = 2^{|V|}$ and $n = |V| = m_1 m_2$. Let $q : \langle |V| \rangle \to V$ be a bijective function. Let

$$H(x) = \sum_{\substack{i,j \in \langle |V| \rangle \\ (q(i),q(j)) \in E}} 1[x_{q(i)} \neq x_{q(j)}], \quad \forall x = (x_{q(1)}, x_{q(2)}, \ldots, x_{q(n)}) \in S,$$

where

$$1[x_{q(i)} \neq x_{q(j)}] = \begin{cases} 1 & \text{if } x_{q(i)} \neq x_{q(j)}, \\ 0 & \text{if } x_{q(i)} = x_{q(j)}, \end{cases}$$

$\forall i, j \in \langle |V| \rangle$, $i \neq j$. Define

$$\pi_x = \pi_x(\beta) = \frac{\exp(-\beta H(x))}{Z_\beta}, \quad \forall x \in S,$$

where $\beta \in \mathbf{R}^* := \mathbf{R} - \{0\}$ is a parameter, and

$$Z_\beta = \sum_{x \in S} \exp(-\beta H(x)).$$

The probability distribution $\pi = (\pi_x)_{x \in S}$ is called the *Ising model* (*without external field*). Taking $q(1) = (1,1)$, $q(2) = (1,2), \ldots, q(m_2) = (1,m_2)$, $q(m_2+1) = (2,1)$, $q(m_2+2) = (2,2), \ldots, q(2m_2) = (2,m_2), \ldots, q((m_1-1)m_2+1) = (m_1,1)$, $q((m_1-1)m_2+2) = (m_1,2), \ldots, q(m_1 m_2) = (m_1,m_2)$, we have (recall that $n = vw = m_1 m_2$, where $v$ can be equal or not to $m_1$ or $m_2$)

$$|H(x) - H(y)| \leq 3v + 1$$

if $(x_{q(1)}, x_{q(2)}, \ldots, x_{q(v(l-1))}) = (y_{q(1)}, y_{q(2)}, \ldots, y_{q(v(l-1))})$ when $1 \leq v(l-1)$ and $(x_{q(vl+1)}, x_{q(vl+2)}, \ldots, x_{q(n)}) = (y_{q(vl+1)}, y_{q(vl+2)}, \ldots, y_{q(n)})$ when $vl + 1 \leq n$. It follows that we can take $f = \exp(-\beta(3v+1))$ and, consequently, $g = \exp(\beta(3v+1))$ if $\beta > 0$ and $f = \exp(\beta(3v+1))$ and, consequently, $g = \exp(-\beta(3v+1))$ if $\beta < 0$. Obviously, we obtain better results if we use $f_l$, $l \in \langle w \rangle$, instead of $f$ (or $g_l$, $l \in \langle w \rangle$, instead of $g$).

If we take, e.g., $m_1 = m_2 = 10$, $v = 1$, $s = w$ ( $s = w$ is the worst case), and $\beta = \frac{1}{64}$, then $n = 100$, $w = 100$, $r = 2^{100} \simeq 1.26765 \cdot 10^{30}$, and, e.g.,

$$\|p_u - \pi\|_1 \leq 0.235\ldots \text{ if } u = 50$$

(50 steps due to $P$ or, equivalently, $50 \cdot 100 = 5000$ steps due to $P_1, P_2, \ldots, P_{100}$) (we used $g = \exp(\frac{1}{16})$; e$= 2.718\ldots$ is the base of exponential function exp),

$$\|p_u - \pi\|_1 \leq 0.027\ldots \text{ if } u = 100,$$

and

$$\|p_u - \pi\|_1 \leq 0.00032\ldots \text{ if } u = 150$$

($2^{100}$ is near to 30! from the example of Mallows model, but here we have good bounds for $\|p_u - \pi\|_1$ in a small number $u$ of steps; obviously, this is easy to understand because our (general) upper bound for $\|p_u - \pi\|_1$ from Theorem 2.8(ii) depends on $u$ and the values of the parameters of the model and of Markov chain) while if we take, e.g., $m_1 = 20$, $m_2 = 10$, $v = 1$, $s = w$, and $\beta = \frac{1}{64}$, then $n = 200$, $w = 200$, $r = 2^{200} \simeq 1.60693 \cdot 10^{60}$, and, e.g.,

$$\|p_u - \pi\|_1 \leq 0.346\ldots \text{ if } u = 1000,$$
$$\|p_u - \pi\|_1 \leq 0.144\ldots \text{ if } u = 1500,$$

and

$$\|p_u - \pi\|_1 \leq 0.060\ldots \text{ if } u = 2000$$

(we note also that $2^{200}$ is much larger than 40! from the example of Mallows model, but here we also have good bounds for $\|p_u - \pi\|_1$ in a small number $u$ of steps).

We conclude saying that we believe that the homogeneous Markov chain framework is too narrow to design fast algorithms of Metropolis-Hastings type. Moreover, we believe that our hybrid Metropolis-Hastings chain works better than the Metropolis-Hastings chain, at least on $\mathbb{S}_n$, the set of permutations of order $n$, and on $\{0, 1, \ldots, h\}^n$.

## REFERENCES

[1] P. Diaconis, *Group Representation in Probability and Statistics.* Institute of Mathematical Statistics, Hayward, CA, 1988.

[2] P. Diaconis and L. Saloff-Coste, *What do we know about the Metropolis algorithm*? J. Comput. System Sci. **57** (1998), 20–36.

[3] P. Diaconis, *The Markov Monte Carlo revolution.* Bull. Amer. Math. Soc. **46** (2009), 179–205.

[4] R.L. Dobrushin, *Central limit theorem for nonstationary Markov chains,* I, II. Theory Probab. Appl. **1** (1956), 65–80, 329–383.

[5] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications.* Biometrika **57** (1970), 97–109.

[6] M. Iosifescu, *Finite Markov Processes and Their Applications.* Wiley, Chichester & Ed. Tehnică, Bucharest, 1980; corrected republication by Dover, Mineola, N.Y., 2007.

[7] D.L. Isaacson and R.W. Madsen, *Markov Chains*: *Theory and Applications.* Wiley, New York, 1976; republication by Krieger, 1985.

[8] D.E. Knuth, *The Art of Computer Programming.* Volume 2. *Seminumerical Algorithms*, 2nd Edition. Addison-Wesley, Reading, 1981.

[9] N. Madras, *Lectures on Monte Carlo Methods.* American Mathematical Society, Providence, Rhode Island, 2002.

[10] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equations of state calculations by fast computing machines*. J. Chem. Phys. **21** (1953), 1087–1092.

[11] U. Păun, *Weak and uniform weak* $\Delta$*-ergodicity for* $[\Delta]$*-groupable finite Markov chains*. Math. Rep. (Bucur.) **6**(**56**) (2004), 275–293.

[12] U. Păun, *Ergodicity coefficients of several matrices*. Math. Rep. (Bucur.) **7**(**57**) (2005), 125–148.

[13] U. Păun, *General* $\Delta$*-ergodic theory of finite Markov chains*. Math. Rep. (Bucur.) **8**(**58**) (2006), 83 –117.

[14] U. Păun, *General* $\Delta$*-ergodic theory*: $\Delta$*-stability, basis, and new results*. Math. Rep. (Bucur.) **8**(**58**) (2006), 219–238.

[15] U. Păun, *General* $\Delta$*-ergodic theory*: *an extension*. Rev. Roumaine Math. Pures Appl. **53** (2008), 209– 226.

[16] U. Păun, *General* $\Delta$*-ergodic theory, with some results on simulated annealing*. Math. Rep. (Bucur.) **13**( **63**) (2011), 171–196.

[17] U. Păun, $G_{\Delta_1,\Delta_2}$ *in action*. Rev. Roumaine Math. Pures Appl. **55** (2010), 387–406.

[18] M. Richey, *The evolution of Markov chain Monte Carlo methods*. Amer. Math. Monthly **117** (2010), 383–413.

[19] L. Saloff-Coste and J. Zúñiga, *Convergence of some time inhomogeneous Markov chains via spectral techniques*. Stochastic Process. Appl. **117** (2007), 961–979.

[20] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd Edition. Springer-Verlag, Berlin–Heidelberg–New York, 1981; revised printing, 2006.

*Received 10 August 2011*

*Romanian Academy*
*"Gheorghe Mihoc-Caius Iacob" Institute*
*of Mathematical Statistics and Applied Mathematics*
*Calea 13 Septembrie nr. 13*
*050711 Bucharest 5, Romania*
*paun@csm.ro*