Coefficients of Ergodicity for Stochastically Monotone Markov Chains

G. Ch. Pflug; W. Schachermayer

Journal of Applied Probability, Vol. 29, No. 4. (Dec., 1992), pp. 850-860.

Stable URL:
http://links.jstor.org/sici?sici=0021-9002%28199212%2929%3A4%3C850%3ACOEFSM%3E2.0.CO%3B2-%23

Journal of Applied Probability is currently published by Applied Probability Trust.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/apt.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Wed Aug 22 02:58:31 2007

J. Appl. Prob. **29**, 850–860 (1992)
Printed in Israel
© *Applied Probability Trust* 1992

COEFFICIENTS OF ERGODICITY FOR STOCHASTICALLY MONOTONE MARKOV CHAINS

G. CH. PFLUG* AND
W. SCHACHERMAYER,* *University of Vienna*

Abstract

In this paper we show that to each distance d defined on the finite state space S of a strongly ergodic Markov chain there corresponds a coefficient ρ_d of ergodicity based on the Wasserstein metric. For a class of stochastically monotone transition matrices P, the infimum over all such coefficients is given by the spectral radius of $P - R$, where $R = \lim_k P^k$ and is attained. This result has a probabilistic interpretation of a control of the speed of convergence of $(P^k)_{k=1}^{\infty}$ by the metric d and is linked to the second eigenvalue of P.

ERGODIC COEFFICIENTS; WASSERSTEIN METRIC; BIRTH AND DEATH MATRIX

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60J10

1. The main results

Let $S = \{1, 2, \cdots, N\}$ be the finite state space of a Markov chain with transition matrix $P = (p_{ij})$. We recall that P is called strongly ergodic, if there is a matrix R with identical rows such that

$$R = \lim_{k \to \infty} P^k \tag{1}$$

(see [4], [5]). It is an important problem to assign a numerical value to the speed of convergence in (1). A tool for doing so is given by the following concept.

1.1. *Definition.* A coefficient of ergodicity is a scalar function ρ defined on the set of stochastic matrices satisfying
(i) $0 \leqq \rho(P)$,
(ii) $\rho(P) = 0$ if and only if P has identical rows,
(iii) for all stochastic matrices P_1, P_2, $\rho(P_1P_2) \leqq \rho(P_1)\rho(P_2)$.

1.2. *Example.* Typical examples of coefficients of ergodicity are Dobrushin's coefficient [2]:

Received 14 December 1990; revision received 10 October 1991.

* Postal address: Universität Wien, Institut für Statistik und Informatik, Universitätstrasse 5/9, A-1010 Wien, Austria.

$$\rho_0(P) := \frac{1}{2} \sup_{i,j} \sum_k |p_{ik} - p_{jk}| = 1 - \inf_{i,j} \sum_k \min(p_{ik}, p_{jk}),$$

and Birkhoff's coefficient [10]:

$$\rho_{\mathrm{B}}(P) := \sup_{\substack{x,y>0 \\ x \neq \lambda y}} \frac{d_B(x'P, y'P)}{d_B(x', y')},$$

where

$$d_{\mathrm{B}}(x', y') = \max_{i,j} \ln \left[\frac{x_i y_j}{x_j y_i} \right] \qquad x_i, y_i \neq 0.$$

It is known that $\rho_{\mathrm{B}} \geqq \rho_0$ ([10], p. 110) and that P is strongly ergodic if and only if there is a $M > 0$ such that $\rho_0(P^M) < 1$ (see [5]). There are however examples of strongly ergodic matrices with $\rho_0(P) = \rho_{\mathrm{B}}(P) = 1$, for example the 'birth and death' matrices considered below.

In the present paper we study the coefficients of ergodicity induced by a metric d on S via the Wasserstein distance, which extends d to $\mathscr{P}(S)$, the set of all probability measures on S (for reference see [3]).

1.3. *Definition.* For $\mu, \nu \in \mathscr{P}(S)$, the Wasserstein distance between μ and ν is defined as

$$d^W(\mu, \nu) = \inf\{\mathbb{E}(d(X, Y)) : X \text{ and } Y \text{ are } S\text{-valued random variables with distributions } \mu \text{ and } \nu \text{ respectively}\}. \tag{2}$$

By identifying the points i of S with the corresponding Dirac measures δ_i one readily observes that the Wasserstein distance d^W on $\mathscr{P}(S)$ indeed extends the original metric d on S.

Let us now introduce the central concept of this paper.

1.4. *Definition.* Let P be a stochastic matrix on the finite state space S. For a metric d on S define

$$\rho_d(P) = \sup_{i,j \in S} \frac{d^W(\delta_i P, \delta_j P)}{d(i,j)}. \tag{3}$$

We call $\rho_d(P)$ the d-coefficient of P, because it is easily verified that ρ_d satisfies the requirements of Definition 1.1. Indeed, in order to establish 1.1 (iii) it suffices to observe that

$$\rho_d(P) = \sup_{\mu, \nu \in \mathscr{P}(S)} \frac{d^W(\mu P, \nu P)}{d^W(\mu, \nu)}. \tag{4}$$

Note in passing that, in general, $\rho_d(P)$ may be greater than 1. However, there always is a metric which gives $\rho_d(P) \leqq 1$, as shown by the subsequent example.

1.5. *Example*. Let d_0 be the discrete metric on S, $d_0(i,j)=\delta_{i,j}$. As is easily seen, the corresponding Wasserstein extension is $d_0(\mu,\nu)=(1/2)\,\|\mu-\nu\|_1$ and hence

$$\rho_{d_0}(P)=(1/2)\sup_{i,j}\sum_k |p_{ik}-p_{jk}| = \rho_0(P)\leqq 1.$$

Thus Dobrushin's coefficient defined in 1.2 appears as a special case of Definition 1.4.

We obtain the following characterisation of strong ergodicity (the proof is given in 3.4 below).

1.6. *Theorem*. The stochastic matrix $P=(p_{ij})$ on the finite state space S is strongly ergodic iff there is a metric d on S such that $\rho_d(P)<1$.

The central question treated in this paper is to investigate the infimum $\kappa(P)$ of $\rho_d(P)$ for a strongly ergodic transition matrix P, when d runs through the metrics on S. We show (Proposition 3.1) that a lower bound for $\kappa(P)$ is given by the absolute value $|\lambda_2|$ of the second eigenvalue of P.

In general $\kappa(P)$ is bigger than $|\lambda_2|$, as will be shown in Example 3.2. However, for an important subclass of stochastic matrices $\kappa(P)$ is attained by some $\rho_d(P)$ and equals the second largest eigenvalue of P. In order to describe this subclass, some definitions are needed.

1.7. *Definition*. A metric d on S is called *one-dimensional* if d is inherited from the euclidean distance of $\mathbb{R}$ via an order-preserving injection of S into $\mathbb{R}$; in other words if for every $1\leqq i<j\leqq N$,

$$d(i,j)=\sum_{k=i}^{j-1} d(k,k+1).$$

A row vector $v=(v_i)_{i=1}^N$ satisfying

$$v_{i+1}-v_i=d(i,i+1), \qquad 1\leqq i<N$$

will be called a *primitive* of the one-dimensional metric d.

The Wasserstein metric pertaining to a one-dimensional distance may easily be computed according to 1.8 (or 1.11).

1.8. *Lemma*. Let d be a one-dimensional metric on S and $\mu=(\mu_1,\cdots,\mu_N)$, $\nu=(\nu_1,\cdots,\nu_N)\in\mathscr{P}(S)$. Let $F=(F_1,\cdots,F_N)$, $G=(G_1,\cdots,G_N)$ be the distribution functions of μ and ν respectively,

$$F_i=\sum_{j=1}^i \mu_j, \qquad G_i=\sum_{j=1}^i \nu_j,$$

and $v=(v_i)_{i=1}^N$ a primitive of d. Then

$$d^W(\mu,\nu)=\sum_{i=1}^{N-1} |F_i-G_i|(v_{i+1}-v_i). \tag{5}$$

This lemma is just a special case of the following well-known result, a proof of which may be found in [11].

1.9. *Lemma.* Let μ, ν be two probability measures with finite expectations on $\mathbb{R}$ endowed with the euclidean metric d_1. Then

$$d_1^W(\mu, \nu) = \int_{-\infty}^{+\infty} |F(u) - G(u)|\, du \tag{6}$$

where F (or G) is the distribution function of μ (or ν).

1.10. *Definition.* On the set $\mathscr{P}(S)$ we define a partial order by saying $\mu \geqq \nu$ if the distribution function of ν dominates that of μ, i.e.

$$\mu \geqq \nu \Leftrightarrow \sum_{j=i}^{N} \mu_j \geqq \sum_{j=i}^{N} \nu_j, \qquad \text{for } 1 \leqq i \leqq N.$$

A stochastic matrix P on S is called *stochastically monotone* [1] if $\mu \geqq \nu \Rightarrow \mu P \geqq \nu P$.

1.11. *Remark.* If μ, ν are measures on S with $\mu \geqq \nu$ and d is a one-dimensional metric on S then Lemma 1.8 reduces to a particularly easy formula, namely

$$d^W(\mu, \nu) = \langle \mu - \nu, \upsilon \rangle \tag{7}$$

where υ is a primitive of d and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^N$. Indeed, we may argue by 'partial integration':

$$\begin{aligned} d^W(\mu, \nu) &= \sum_{i=1}^{N-1} (G_i - F_i)(\upsilon_{i+1} - \upsilon_i) \\ &= \sum_{i=1}^{N-1} \left(\sum_{j=1}^{i} (\nu_j - \mu_j) \right) (\upsilon_{i+1} - \upsilon_i) \\ &= \sum_{j=1}^{N-1} \left((\nu_j - \mu_j) \sum_{i=j}^{N-1} (\upsilon_{i+1} - \upsilon_i) \right) \\ &= \sum_{j=1}^{N} (\mu_j - \nu_j)(\upsilon_j - \upsilon_N) \\ &= \langle \mu - \nu, \upsilon \rangle. \end{aligned}$$

1.12. *Definition.* A stochastic matrix P on the state space S is called *lumpable* [7] if there is a non-trivial partition of the state space $S = \bigcup_{k=1}^{K} A_k$ such that

$$i \mapsto \sum_{j \in A_k} p_{ij} \text{ is constant within each partition set.}$$

We say that P is *lumpable into intervals* if there is a partition as above such that all A_k are intervals of S.

The notions of lumpability and lumpability into intervals do not coincide, as may be seen by considering

$$P = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

Taking the partition $\{1, 3\}, \{2\}$ one observes that P is lumpable; it is just as easy to verify that P is not lumpable into intervals.

Now we can formulate the central result of this paper.

1.13. *Main theorem.* Let P be a strongly ergodic stochastically monotone matrix on the finite state space $S = \{1, \cdots, N\}$ which is not lumpable into intervals.

Then there is a simple eigenvalue $\lambda_2 \in]0, 1[$, which satisfies $\lambda_2 \geqq |\lambda_i|$ for all eigenvalues λ_i of P except for the simple eigenvalue $\lambda_1 = 1$. The corresponding right eigenvector $v = (v_i)_{i=1}^N$ is strictly increasing.

Defining the one-dimensional metric d on S by $d(i, j) = |v_i - v_j|$ we have $\rho_d(P) = \lambda_2$ and this metric is unique in the following sense: for every one-dimensional metric d_1 on S such that $\rho_{d_1}(P) \leqq \lambda_2$, d_1 coincides with d up to a scalar factor.

The proof of this theorem is given in Section 4. Whereas Theorem 1.13 discusses the role of the *right* eigenvector v associated with λ_2, the following theorem investigates the role of the *left* eigenvector (the proof is given in 4.2 below).

1.14. *Theorem.* Let $\nu = (\nu_1, \cdots, \nu_N)$ be the left eigenvector pertaining to λ_2, normalized by $\langle \nu, v \rangle = \Sigma_{i=1}^N \nu_i v_i = 1$. If λ_2 is the only eigenvalue of P of absolute value equal to λ_2, then, for every $i, j \in S$, $\lim_{n \to \infty} (\delta_i - \delta_j) P^n / \lambda_2^n = d(i, j)\nu$ where the convergence takes place with respect to the norm $\|\cdot\|_1$.

2. Birth and death matrices as basic examples

Let P be a stochastic matrix describing a general birth and death process on $S = \{1, \cdots, N\}$, i.e.

$$P = \begin{bmatrix} r_1 & p_1 & & & \\ q_2 & r_2 & p_2 & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & & q_{N-1} & r_{N-1} & p_{N-1} \\ & & & q_N & r_N \end{bmatrix}$$

such that, letting $q_2 = p_N = 0$, for every $1 \leqq i \leqq N$ we have $q_i + r_i + p_i = 1$. Note that P is stochastically monotone iff for $1 \leqq i \leqq N - 1$, $q_i + r_i \geqq q_{i+1}$. We claim that P is not lumpable into intervals iff for every $2 \leqq i \leqq N - 1$, q_i and p_i are different from 0. Indeed suppose that $q_i = 0$ for some $2 \leqq i \leqq N - 1$ (the case of $p_i = 0$ is similar): then the partition of S into $\{1\}, \{2\}, \cdots, \{i - 1\}, \{i, i + 1, \cdots, N\}$ satisfies the requirements of Definition 1.12. The proof of the reverse implication is left to the reader.

As a special case, consider the matrix

$$P_a = \begin{bmatrix} a & 1-a & 0 & & & \\ a & 0 & 1-a & & \mathbf{0} & \\ & \ddots & & \ddots & \ddots & \\ \mathbf{0} & & a & 0 & 1-a & \\ & & & a & 1-a & \end{bmatrix}$$

for $0 < a < 1$. Clearly P_a is strongly ergodic and the stationary distribution is given by $((1-a)/a)^n)_{n=1}^{N}$ (up to the normalizing factor $\Sigma_{n=1}^{N}((1-a)/a)^n$). If $N \geqq 4$ then $\rho_0(P) = 1$ and in fact $\rho_0(P^k) = 1$ for $k \leqq N-3$. Hence Dobrushin's coefficient of ergodicity gives non-trivial information only for high powers of P.

On the other hand consider, for $a \neq 1/2$, the metric

$$d(i,j) = \left| \left(\frac{a}{1-a}\right)^{i/2} - \left(\frac{a}{1-a}\right)^{j/2} \right| . \tag{8}$$

An easy direct computation gives

$$d^W(\delta_i P_a, \delta_j P_a) \leqq 2\sqrt{a(1-a)}\, d(i,j) \tag{9}$$

hence

$$\rho_d(P_a) \leqq 2\sqrt{a(1-a)} \tag{10}$$

which — at least in the case $a \neq 1/2$ — gives non-trivial information about P_a itself and not only about some high power of P_a. Considering $i, j \in S$, (9) tells us that by applying P_a the Wasserstein distance of the probability measures $\delta_i P_a$ and $\delta_j P_a$ is smaller than the original distance by a factor of at least $\rho_d(P_a)$.

Note that equality holds in (9) iff neither i nor j equals 1 or N. In particular, equality holds in (10) for $N \geqq 4$.

The question arises whether this metric is 'optimal'. For the application of the Main Theorem 1.13 we have to find the eigenvalues of P_a, which are $\lambda_1 = 1$ and for $2 \leqq j \leqq N$,

$$\lambda_j = 2\sqrt{a(1-a)} \cos\left(\frac{(j-1)\pi}{N}\right) \tag{11}$$

(see [6], Chapter 4.4). By considering the corresponding right eigenvectors one finds that the 'optimal' metric is not (8), but

$$D(i,j) = \sum_{k=i}^{j-1} \left(\frac{a}{1-a}\right)^{k/2} \sin\left(\frac{k\pi}{N}\right).$$

For this metric

$$\rho_D(P_a) = \lambda_2 = 2\sqrt{a(1-a)} \cos\left(\frac{\pi}{N}\right) < 1$$

for all $0 < a < 1$.

3. Proof of Theorem 1.6

The proof requires some preliminary results.

3.1. *Proposition.* Let P be a stochastic matrix on the finite state space $S = \{1, \cdots, N\}$ and d a metric on S. Then $\rho_d(P) \geqq |\lambda_2|$ where $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ are the eigenvalues of P ordered such that $1 = \lambda_1 \geqq |\lambda_2| \geqq \cdots \geqq |\lambda_N|$.

Proof. Denote by H the subspace of $l^1(S)$ orthogonal to the constant function $(1, \cdots, 1)$ on S. Note that H is invariant under P by left multiplication and that the restriction of P to H has the eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$.

We first show the proposition for the case of the discrete metric $d = d_0$ considered in Example 1.5. In this case $d_0(\mu, \nu) = \| \mu - \nu \|_1$ and one easily observes that $\rho_{d_0}(P)$ is just the norm of the restriction of P to H with respect to the norm $\| \cdot \|$ on H. The assertion of the proposition thus reduces to the well-known fact that the operator of norm of $P|_H$ is greater than or equal to the spectral radius $|\lambda_2|$ of $P|_H$. In fact this argument shows that, for $n \in \mathbb{N}$, $\rho_{d_0}(P^n) \geqq |\lambda_2|^n$.

Now we pass to the general case. For an arbitrary metric d on S find $c_1, c_2 > 0$ such that, for $i, j \in S$,

$$c_1 d_0(i,j) \leqq d(i,j) \leqq c_2^{-1} d_0(i,j).$$

Hence for every $n \in \mathbb{N}$

$$\begin{aligned}(\rho_d(P))^n &\geqq \rho_d(P^n)\\ &\geqq c_1 c_2 \rho_{d_0}(P^n)\\ &\geqq c_1 c_2 |\lambda_2|^n\end{aligned}$$

which by taking the nth root immediately gives the assertion.

The absolute value $|\lambda_2|$ of the second eigenvalue of a strongly ergodic transition matrix therefore is a lower bound for $\rho_d(P)$. In the next example we shall see that in general $\kappa(P) = \inf\{\rho_d(P) : d \text{ a metric on } S\}$ is strictly bigger than $|\lambda_2|$.

3.2. *Example.* Let $0 < c < 1$ and consider the strongly ergodic transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ c & 1-c & 0 \end{pmatrix}.$$

The eigenvalues are $\lambda_1 = 1$ and $\lambda_{2,3} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} - c}$. In particular, for $c = 1/4$, $\lambda_2 = \lambda_3 = -1/2$. Note that $(\delta_2 - \delta_3)P = \delta_3 - c\delta_1 - (1-c)\delta_2$ whence for any metric d on S

$$d^W(\delta_2 P, \delta_3 P) = cd(1,3) + (1-c)d(2,3),$$

which implies that

$$\rho_d(P) \geqq d^W(\delta_2 P, \delta_3 P)/d(2,3) > 1 - c.$$

For $c = 1/4$ this gives $1 - c = 3/4$ which is strictly bigger than $|\lambda_2| = 1/2$.

3.3. *Lemma*. Let P be a transition matrix and suppose that there is a non-empty subset A of the set of all (non-ordered) pairs $\{i,j\}$, $i,j \in S$, $i \neq j$, with the following property. For every $\{i,j\} \in A$ and $k \in \operatorname{supp}(\delta_i P)$, $l \in \operatorname{supp}(\delta_j P)$, we have $\{k,l\} \in A$.

Then P fails to be strongly ergodic.

Proof. The assumption in fact implies that if $\mu, \nu \in \mathscr{P}(S)$ are such that for $i \in \operatorname{supp}(\mu)$, $j \in \operatorname{supp}(\nu)$ we have $\{i,j\} \in A$ then the same property holds true for the measures μP and νP.

Indeed, write

$$\mu = \sum a_p \delta_{i_p}, \qquad \nu = \sum_q b_q \delta_{j_q}$$

where $a_p > 0$, $b_q > 0$ and $\Sigma_p a_p = \Sigma_q b_q = 1$. Each pair $\{i_p, j_q\}$ belongs to A, hence by hypothesis for each $k \in \operatorname{supp}(\mu P)$ and $l \in \operatorname{supp}(\nu P)$ the pair $\{k,l\}$ belongs to A, as $k \in \operatorname{supp}(\delta_{i_p} P)$ and $l \in \operatorname{supp}(\delta_{i_q} P)$ for some p and q.

Hence for $\{i,j\} \in A$ it follows inductively that for every $m \in \mathbb{N}$ the measures $\delta_i P^m$ and $\delta_j P^m$ have the property defined above. In particular they have disjoint supports and therefore

$$d_0(\delta_i P^m, \delta_j P^m) = \tfrac{1}{2} \| (\delta_i - \delta_j) P^m \|_1 = 1 \qquad \text{for } m \in \mathbb{N},$$

whence $\rho_0(P^m) = 1$ for every $m \in \mathbb{N}$ and P fails to be strongly regular by the remark in 1.2 above.

3.4. *Proof of Theorem* 1. We proceed inductively. Consider the discrete metric d_0 and let A_0 be the set of all pairs $\{i,j\}$, $i \neq j$, $i,j \in S$. For each such pair $\{i,j\}$ we have

$$d_0^W(\delta_i P, \delta_j P) \leqq d_0(i,j). \tag{12}$$

Let A_1 be the set of all pairs $\{i,j\}$ such that equality holds true in (12). By the above lemma, A_1 is a proper subset of A_0. Define

$$d_1(i,j) = \begin{cases} 1, & \text{if } \{i,j\} \in A_1 \\ 1 - \varepsilon_1, & \text{if } \{i,j\} \notin A_1 \end{cases}$$

where $1/2 \geqq \varepsilon_1 > 0$ is chosen small enough such that

$$d_1^W(\delta_i P, \delta_j P) \leqq d_1(i,j) \tag{13}$$

still holds true with strict inequality for all $\{i,j\} \notin A_1$. Note that for $\{i,j\} \in A_1$ the above inequality also holds true (possibly with equality) as for those $\{i,j\}$ we have only decreased the left-hand side compared to (12). If $A_1 = \varnothing$ we stop the construction. Otherwise, suppose $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \neq \varnothing$ and $d_1, \cdots, d_n$ have been defined such that $d_n(i,j) = 1$ for $\{i,j\} \in A_n$, $d_n(i,j) < 1$ for $\{i,j\} \notin A_n$ and such that

$$d_n^W(\delta_i P, \delta_j P) \leqq d_n(i,j) \tag{14}$$

for every $\{i,j\} \in A_0$ with strict inequality holding at least for $\{i,j\} \notin A_n$.

Let A_{n+1} be the set of all $\{i,j\}$ such that equality holds true in (14). By the above lemma A_{n+1} is strictly contained in A_n; otherwise the set $A = A_n = A_{n+1}$ would satisfy the requirements of the lemma in contradiction to the strong ergodicity of P. Define

$$d_{n+1}(i,j) = \begin{cases} 1, & \text{if } \{i,j\} \in A_{n+1} \\ 1 - \varepsilon_{n+1}, & \text{if } \{i,j\} \in A_n \setminus A_{n+1} \\ d_n(i,j), & \text{otherwise} \end{cases}$$

where $1/2 \geqq \varepsilon_{n+1} > 0$ is chosen small enough that

$$d_{n+1}^W(\delta_i P, \delta_j P) \leqq d_{n+1}(i,j) \tag{15}$$

still holds true with strict inequality for every $\{i,j\} \notin A_{n+1}$. This completes the inductive step. After a finite number of times we arrive at the case $A_{n+1} = \varnothing$ which implies that in (15) strict inequality holds true for all pairs $\{i,j\}$. Hence d_{n+1}, which is a metric on S in view of $\varepsilon_n \leqq 1/2$, verifies $\rho_{d_{n+1}}(P) < 1$.

Thus Theorem 1.6 is proved.

4. Proof of the main theorems

4.1. *Proof of Theorem* 1.13. Let

$$s_{ij} = \sum_{k=1}^{j} p_{ik}, \qquad j = 1, \cdots, N$$

$$s_{i0} = 0,$$

$$t_{ij} = s_{ij} - s_{i+1,j}, \qquad i = 1, \cdots, N-1.$$

Let T be the $[N-1] \times [N-1]$ matrix

$$T = (t_{ij})_{\substack{i=1,\cdots,N-1, \\ j=1,\cdots,N-1}}.$$

For any vector $v = (v_1, \cdots, v_N)$ let v^- be the $[N-1]$-dimensional vector $v^- = (v_2 - v_1, \cdots, v_N - v_{N-1})$. We claim that the following assertion is true: each eigenvalue $\lambda \neq 1$ of P with right eigenvector v is also an eigenvalue of T with right eigenvector v^-. Let λ be such an eigenvector. Then

$$\lambda v_i = \sum_{j=1}^{N} p_{ij} v_j = \sum_{j=1}^{N} (s_{ij} - s_{ij-1}) v_j$$

$$= \sum_{j=1}^{N-1} s_{ij}(v_j - v_{j+1}) + v_N.$$

Therefore, for $i = 1, \cdots, N-1$,

$$\lambda v_i^- = \lambda(v_{i+1} - v_i) = \sum_{j=1}^{N-1} (s_{ij} - s_{i+1,j})(v_{j+1} - v_j) = \sum_{j=1}^{N-1} t_{ij} v_j^-.$$

Conversely let $\lambda \neq 1$ be an eigenvalue of T with corresponding right eigenvector w. We prove now λ is an eigenvalue of P and there is a right eigenvector v, such that $w = v^{-}$.

Let v_N be the solution of

$$\lambda v_N = v_N - \sum_{j=1}^{N-1} s_{N_j} w_j$$

and define

$$v_j = v_N - \sum_{k=j}^{N-1} w_k, \qquad j = 1, \cdots, N-1.$$

Since

$$s_{ij} = s_{N_j} + \sum_{k=i}^{N-1} t_{kj}, \qquad i = 1, \cdots, N-1$$

we get

$$\begin{aligned}\sum_{j=1}^{N} p_{ij} v_j &= v_N - \sum_{j=1}^{N} (s_{ij} - s_{ij-1}) \sum_{k=j}^{N-1} w_k \\ &= v_N - \sum_{j=1}^{N-1} s_{ij} w_j = v_N - \sum_{j=1}^{N-1} \left(s_{N_j} + \sum_{k=i}^{N-1} t_{kj} \right) w_j \\ &= v_N - \sum_{j=1}^{N-1} s_{N_j} w_j - \lambda \sum_{k=i}^{N-1} w_k = \lambda v_N - \lambda \sum_{k=i}^{N-1} w_k = \lambda v_i,\end{aligned}$$

which proves the assertion above.

In order to apply the well-known Perron–Frobenius theorem on positive matrices, we have to check that T is a positive, irreducible matrix.

The fact that

$$t_{ij} = \sum_{k=1}^{j} p_{ik} - \sum_{k=1}^{j} p_{i+1,k} \geqq 0$$

is a reformulation of the fact that P is stochastically monotone. Suppose that T is reducible. Then there is a set of indices $A \neq \emptyset$, $A^c \neq \emptyset$ such that

$$t_{ij} = \sum_{k=1}^{j} p_{ik} - \sum_{k=1}^{j} p_{i+1,k} = 0$$

for $i \in A, j \in A^c$. Let B be the set $B = \{i : i \in A \text{ or } i - 1 \in A\}$. We may represent the set B as a finite union of disjoint intervals $B = \bigcup \{i : i_k \leqq i \leqq l_k\}$. Then $i_k \in A$, $i_k - 1 \notin A$ and $l_k \notin A$. Let the partition sets be all intervals of B and all the remaining singletons. We show that P is lumpable into these intervals.

For every k

$$\sum_{j=i_k}^{l_k} p_{ij} = \sum_{j=1}^{l_k} p_{ij} - \sum_{j=1}^{i_k - 1} p_{ij}.$$

If i is the same partition as i', then

$$\sum_{j=1}^{l} p_{ij} = \sum_{j=1}^{l} p_{i+1,j} = \cdots = \sum_{j=1}^{l} p_{i'j}$$

for all $l \notin A$. Since $l_k \notin A$ and $i_k - 1 \notin A$, we see that $\Sigma_{j=i_k}^{l_k} p_{ij}$ is constant in every interval and this contradicts the assumption that P is not lumpable into intervals.

The Perron–Frobenius theorem implies that T has a positive maximal eigenvalue λ_2, which has multiplicity 1 and a positive eigenvector w. As just proved, λ_2 is the second largest eigenvalue of P and there is a right eigenvector v of P with $w = v^-$. Consequently, v is strictly increasing.

4.2. *Proof of Theorem* 1.14. Consider the restriction of the operator $\lambda_2^{-1}P$ to the hyperplane H of $l^1(S)$ orthogonal to the constant function $(1, 1, \cdots, 1)$ on S.

As in the proof of 3.1 observe that this operator has a simple eigenvalue equal to 1 and all other eigenvalues of absolute value strictly less than 1. Hence $\lambda_2^{-n}P^n \,|_H$ converges to the projection onto the eigenspace corresponding to the eigenvalue λ_2 of P. As v is the left and v the right eigenvector pertaining to λ_2, normalized by $\langle v, v\rangle = 1$ we obtain

$$\lim_{n \mapsto \infty} (S_i - S_j)P^n/\lambda_2^n = \langle \delta_i - \delta_j, v \rangle v$$

$$= d(i, j)v$$

the last equality following from the definition of d.

4.3. *Remark.* It is not true in general that in the setting of the main theorem λ_2 is the only eigenvalue of P of absolute value equal to λ_2. For the matrix P_a of the basic example we get for instance $\lambda_N = -\lambda_2$ (see (11)).

References

[1] Daley, D. J. (1968) Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitsth.* **10**, 305–317.

[2] Dobrushin, R. L. (1956) Central limit theorems for non-stationary Markov chains I. *Theory Prob. Appl.* **1**, 65–80.

[3] Dudley, R. M. (1976) *Probability and Metrics.* Aarhus University Lecture Notes **45**, 1–126.

[4] Isaacson, D. and Luecke, G. R. (1978) Strongly ergodic Markov chains and rates of convergence using spectral conditions. *Stoch. Proc. Appl.* **7**, 113–121.

[5] Isaacson, D. and Madsen, R. (1976) *Markov Chains.* Wiley, New York.

[6] Karlin, S. (1966) *A First Course in Stochastic Processes.* Academic Press, New York.

[7] Kemeny, J. G. and Snell, J. L. (1976) *Finite Markov Chains.* Springer-Verlag, New York.

[8] Nummelin, E. and Tweedie, R. L. (1978) Geometric ergodicity and R-positivity for general Markov chains. *Ann. Prob.* **6**, 404–420.

[9] Rachev, S. T. (1984) The Monge–Kantorovich mass transformance problem and its stochastic applications. *Theory Prob. Appl.* **29**, 647–676.

[10] Seneta, E. (1981) *Non-negative Matrices and Markov Chains*, 2nd edn. Springer-Verlag, Berlin.

[11] Vallender, S. S. (1973) Calculation of the Wasserstein-distance between probability distributions on the line. *Theory Prob. Appl.* **18**, 784–786.