World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:7, No:1, 2013

On the Hierarchical Ergodicity Coefficient

Yilun Shang

Abstract—In this paper, we deal with the fundamental concepts and properties of ergodicity coefficients in a hierarchical sense by making use of partition. Moreover, we establish a hierarchial Hajnal's inequality improving some previous results.

Keywords—Stochastic matrix, ergodicity coefficient, partition.

I. Introduction

THROUGHOUT the paper all matrices are of fixed size $n \times n$. A stochastic matrix P is a square matrix with non-negative entries and unit row sums. For an n-dimensional column vector x, we say it is stochastic if x has non-negative entries and $x^T\mathbf{1} = 1$, where $\mathbf{1}$ is the n-dimensional column vector with all entries equal to 1. Ergodicity coefficient [5], [15] is a continuous scalar function $\mu(\cdot)$ on the set of stochastic matrices P (P is regarded as a point in $\mathbb{R}^{n^2}$) that satisfies $0 \le \mu(P) \le 1$. An ergodicity coefficient is called proper if

$$\mu(P) = 0 \iff P = \mathbf{1}x^T,$$

for some stochastic vector x (i.e., all rows of P are identical).

Ergodicity coefficient is used to measure the convergence rate of infinite products of stochastic matrices especially in the context of inhomogeneous Markov chains [17], [18]. Roughly speaking, a Markov chain is said to be ergodic if the associated matrix products converge to a stochastic matrix whose rows are all identical, that is, a rank-one matrix. For relevant backgrounds on ergodicity and some classical stochastic matrices such as Markov matrices and scrambling matrices, we refer to [7], [8], [16] and references therein.

For a stochastic matrix P, denote p_{ij} the entry of P on the ith row and jth column. Some common ergodicity coefficients are defined as follows:

$$\tau(P) = 1 - \min_{i,j} \textstyle\sum_k \min\{p_{ik}, p_{jk}\},$$

$$\alpha(P) = \max_k \max_{i,j} |p_{ik} - p_{jk}|,$$

$$\beta(P) = 1 - \textstyle\sum_k \min_i p_{ik}.$$

τ is called the Markov-Dobrushin ergodicity coefficient [4], [9]. It is easy to check [16] that they are all proper, and furthermore, the following relation holds for every stochastic matrix P

$$\alpha(P) \le \tau(P) \le \beta(P).$$

Other variants of ergodicity coefficient can be found in e.g. [1], [2], [3], [8].

Y. Shang is with the Institute for Cyber Security, University of Texas, San Antonio, TX, 78249 USA (e-mail: shylmath@hotmail.com).

Let $\|\cdot\|$ be a vector norm in $\mathbb{R}^n$. Define the Hajnal diameter of a stochastic matrix P with row vectors $P_1, P_2, \cdots, P_n$ as

$$\Delta(P) = \max_{i,j} \|P_i - P_j\|.$$

Clearly, $\Delta(P) = 0$ if and only if all rows of P are identical. As a useful tool in studying the ergodicity of Markov chains, the well known Hajnal's inequality (see the comment below), in its general form, is established as follows.

Theorem 1. [20] *For any two stochastic matrices P and Q,*

$$\Delta(PQ) \le \tau(P)\Delta(Q).$$

Note that when we take the L_1 norm $\|\cdot\|_1$,

$$\Delta(P) = 2\tau(P).$$

The resulting inequality is sharper than Hajnal's in [5]. Historically, the classical form of the inequality was due to Markov [9], and was rediscovered by others after [5] was published. Some authors also call it Hajnal's inequality (see [14], [16] for historical development).

In this paper, motivated by applications in search of a cluster consensus between a set of agents [6], [19], we aim to refine ergodicity coefficients as well as Hajnal's inequality by means of partition (in a hierarchical nature). A partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$ of the set $[n] = \{1, 2, \cdots, n\}$ is a sequence of subsets of $[n]$ such that $\cup_{s=1}^K \mathcal{S}_s = [n]$ and $\mathcal{S}_s \cap \mathcal{S}_t = \emptyset$ for $s \ne t$. As we will see below the partition $\mathcal{S}$ allows us to extend some known results on ergodicity elegantly.

We should mention that, in the literature, the application of partitions to ergodicity coefficients has been made in some papers (c.f. [10], [11], [12]). However, the coefficients studied here have totally different definitions and the results are different (see the discussion below).

II. Hierarchical ergodicity coefficient

By virtue of the partition $\mathcal{S}$, in this section we extend the concepts of ergodicity coefficient mentioned above to those in the hierarchical sense.

For a given partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$, a *hierarchical ergodicity coefficient*, denoted $\mu_{\mathcal{S}}(\cdot)$, is a sort of ergodicity coefficient defined for stochastic matrices P. We say $\mu_{\mathcal{S}}$ is *hierarchial proper* if

$$\mu_{\mathcal{S}}(P) = 0 \iff P = \sum_{s=1}^K \mathbf{1}_{\mathcal{S}_s} x_s^T,$$

where, for $s = 1, \cdots, K$, $\mathbf{1}_{\mathcal{S}_s}$ is the sum of ith n-dimensional coordinate vectors $e_i = (0, \cdots, 0, \overset{i\text{th}}{1}, 0, \cdots, 0)^T$ over all $i \in \mathcal{S}_s$, and x_s is a stochastic vector.

A hierarchical proper ergodicity coefficient is equal to zero if the rows corresponding to each subset $\mathcal{S}_s$ for $s = 1, \cdots, K$ of the stochastic matrix are identical. As such, the rank of the stochastic matrix equals K. Clearly, we reproduce the concept of proper ergodicity coefficient when $K = 1$. In what follows, we instantiate the above general $\mu_{\mathcal{S}}$ in some special cases. Specifically, the hierarchical counterparts for ergodicity coefficients τ, α and β can be formulated as

$$\tau_{\mathcal{S}}(P) = 1 - \min_{1\le s\le K} \min_{i,j\in\mathcal{S}_s} \sum_k \min\{p_{ik}, p_{jk}\},$$

$$\alpha_{\mathcal{S}}(P) = \max_k \max_{1\le s\le K} \max_{i,j\in\mathcal{S}_s} |p_{ik} - p_{jk}|,$$

$$\beta_{\mathcal{S}}(P) = 1 - \min_{1\le s\le K} \sum_k \min_{i\in\mathcal{S}_s} p_{ik}.$$

It is direct to check that all these ergodicity coefficients are hierarchical proper. The definition of $\tau_{\mathcal{S}}$ first appears in [11] geared towards the study of uniform weak $\mathcal{S}$-ergodicity of Markov chains. A sub-multiplicative property $\tau_{\mathcal{S}}(PQ) \le \tau_{\mathcal{S}}(P)\tau_{\mathcal{S}}(Q)$ was established. In [10], an analogous generalized Markov-Dobrushin ergodicity coefficient was introduced as

$$\bar{\tau}_{\mathcal{S}}(P) = 1 - \min_{1\le s,t\le K} \min_{\substack{i\in\mathcal{S}_s \\ j\in\mathcal{S}_t, s\ne t}} \sum_k \min\{p_{ik}, p_{jk}\}.$$

Note that $\bar{\tau}_{\mathcal{S}}$ measures the inter-group differences rather than the intra-group ones. Under a decomposability condition, for any non-unit eigenvalue λ of P, Păun [10] showed that

$$|\lambda| \le \bar{\tau}_{\mathcal{S}}(P).$$

A stochastic matrix P is said to be doubly stochastic if its transpose P^T is also stochastic. We have the following result for hierarchial ergodicity coefficient of doubly stochastic matrices.

Proposition 1. *For a given partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$, if P is a doubly stochastic matrix and $\mu_{\mathcal{S}}$ a hierarchical proper ergodicity coefficient, then*

$$\mu_{\mathcal{S}}(P) = \mu_{\mathcal{S}}(P^T) = 0 \iff P = P^T.$$

Proof. $\Longrightarrow$: By assumption, we can write

$$P = \sum_{s=1}^K \mathbf{1}_{\mathcal{S}_s} x_s^T,$$

where, for $s = 1, \cdots, K$, $\mathbf{1}_{\mathcal{S}_s}$ is the sum of ith n-dimensional coordinate vectors $e_i = (0, \cdots, 0, \overset{i\text{th}}{1}, 0, \cdots, 0)^T$ over all $i \in \mathcal{S}_s$, and x_s is a stochastic vector. Thus,

$$P^T = \sum_{s=1}^K x_s \mathbf{1}_{\mathcal{S}_s}^T.$$

Since $\mu_{\mathcal{S}}(P^T) = 0$, we obtain

$$\sum_{s=1}^K x_s \mathbf{1}_{\mathcal{S}_s}^T = \sum_{s=1}^K \mathbf{1}_{\mathcal{S}_s} y_s^T$$

for stochastic vectors y_s $(s = 1, \cdots, K)$. Let $x_s = (x_{s1}, x_{s2}, \cdots, x_{sn})^T$ and $y_s = (y_{s1},\ y_{s2},\ \cdots,\ y_{sn})^T$, for all $s = 1, \cdots, K$. Expanding both sides of the above equality, we observe directly that $y_{si} = x_{si}$ for all $s = 1, \cdots, K$ and $i = 1, \cdots, n$. It follows that $y_s = x_s$, and hence $P = P^T$.

$\Longleftarrow$: This direction follows simply from the definition of hierarchical proper ergodicity coefficient. □

We remark that if $K = 1$, Proposition 1 reduces to the known result [8], [13]

$$\mu(P) = \mu(P^T) = 0 \iff P = \frac{1}{n}\mathbf{1}\mathbf{1}^T.$$

To see how Proposition 1 works, a simple non-trivial example could be $n = 3$, $\mathcal{S} = \{\{1, 3\}, \{2\}\}$ and

$$P = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}.$$

A useful relation between the above mentioned hierarchical ergodicity coefficients $\tau_{\mathcal{S}}, \alpha_{\mathcal{S}}$ and $\beta_{\mathcal{S}}$ are stated below. Analogous and further relations in non-hierarchial sense can be found in e.g. [8], [13], [16].

Proposition 2. *For a given partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$, if P is a stochastic matrix, then the following relation holds*

$$\alpha_{\mathcal{S}}(P) \le \tau_{\mathcal{S}}(P) \le \beta_{\mathcal{S}}(P).$$

Proof. We start with the first inequality. By definition of ergodicity coefficient $\alpha_{\mathcal{S}}$, we have

$$\alpha_{\mathcal{S}}(P) = p_{mr} - p_{lr} \ge 0$$

for some indices l, m and r. Hence

$$\begin{aligned} \tau_{\mathcal{S}}(P) &= \frac{1}{2} \max_{1\le s\le K} \max_{i,j\in\mathcal{S}_s} \sum_k |p_{ik} - p_{jk}| \\ &\ge \frac{1}{2} \sum_k |p_{mk} - p_{lk}|. \end{aligned}$$

We now partition $[n]$ into two subsets: $\mathcal{P}_m$, consisting of indices k such that $p_{mk} \ge p_{lk}$, and $\mathcal{P}_l$, consisting of indices k such that $p_{mk} < p_{lk}$. Since P is a stochastic matrix which has unit row sums, we obtain

$$\sum_{k\in\mathcal{P}_l} p_{mk} = 1 - \sum_{k\in\mathcal{P}_m} p_{mk}$$

and

$$\sum_{k\in\mathcal{P}_l} p_{lk} = 1 - \sum_{k\in\mathcal{P}_m} p_{lk}.$$

Therefore,

$$\begin{aligned} \sum_k |p_{mk} - p_{lk}| &= \sum_{k\in\mathcal{P}_m} (p_{mk} - p_{lk}) + \sum_{k\in\mathcal{P}_l} (p_{lk} - p_{mk}) \\ &= 2 \sum_{k\in\mathcal{P}_m} (p_{mk} - p_{lk}). \end{aligned}$$

Recall that for any $k \in \mathcal{P}_m$, we have $p_{mk} - p_{lk} \geq 0$. Thus, we obtain

$$\begin{aligned} \tau_{\mathcal{S}}(P) &\geq \frac{1}{2}\sum_k |p_{mk} - p_{lk}| \\ &= \sum_{k \in \mathcal{P}_m} (p_{mk} - p_{lk}) \\ &= p_{mr} - p_{lr} + \sum_{\substack{k \in \mathcal{P}_m \\ k \neq r}} (p_{mk} - p_{lk}) \\ &\geq \alpha_{\mathcal{S}}(P), \end{aligned}$$

as desired.

As for the second inequality, for every $s = 1, \cdots, K$, we can choose two indices i_s and j_s such that

$$\min_{i,j \in \mathcal{S}_s} \sum_k \min\{p_{ik}, p_{jk}\} = \sum_k \min\{p_{i_s k}, p_{j_s k}\}.$$

Hence, we obtain

$$\begin{aligned} 1 - \tau_{\mathcal{S}}(P) &= \min_{1 \leq s \leq K} \sum_k \min\{p_{i_s k}, p_{j_s k}\} \\ &\geq \min_{1 \leq s \leq K} \sum_k \min_{i \in \mathcal{S}_s} p_{ik} \\ &= 1 - \beta_s(P). \end{aligned}$$

□

III. Extended Hajnal's inequality

In this section, we extend the Hajnal inequality to the hierarchical case.

To this end, we first introduce a hierarchical Hajnal diameter [6]. Given a partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$, and a stochastic matrix P, which has row vectors $P_1, P_2, \cdots, P_n$, the *hierarchial Hajnal diameter* is defined as

$$\Delta_{\mathcal{S}}(P) = \max_{1 \leq s \leq K} \max_{i,j \in \mathcal{S}_s} \|P_i - P_j\|.$$

Note that $\Delta_{\mathcal{S}}(P) = 0$ if and only if $P = \sum_{s=1}^K \mathbf{1}_{\mathcal{S}_s} x_s^T$ for some stochastic vectors $x_1, \cdots, x_K$.

Given a partition $\mathcal{S}$, a stochastic matrix P is said to be *hierarchical balanced* if, for $i \in \mathcal{S}_s$,

$$\gamma_{st} := \sum_{j \in \mathcal{S}_t} p_{ij}$$

is independent of i. In other words, γ_{st} only depends on the partition indices s and t. This property is dubbed "inter-cluster common influence" in [6].

Theorem 2. *For a given partition $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_K\}$, if P and Q are two hierarchical balanced stochastic matrices, then*

$$\Delta_{\mathcal{S}}(PQ) \leq \tau_{\mathcal{S}}(P)\Delta_{\mathcal{S}}(Q).$$

In particular, if $Q = I$ (i.e., the identity matrix), then

$$\Delta_{\mathcal{S}}(P) \leq \tau_{\mathcal{S}}(P).$$

Proof. Let $R = PQ$ and $R_1, R_2, \cdots, R_n$ denote the rows of R. Namely, $R = (R_1^T, R_2^T, \cdots, R_n^T)^T$. Moreover, suppose that $Q_1, Q_2, \cdots, Q_n$ are the rows of Q. Then we have

$$R_i = \sum_k p_{ik} Q_k,$$

for all $i = 1, \cdots, n$.

Fix $i \in \mathcal{S}_t$ and $j \in \mathcal{S}_t$ for some $1 \leq t \leq K$. We have

$$R_i = \sum_{s=1}^K \sum_{k \in \mathcal{S}_s} p_{ik} Q_k$$

and

$$R_j = \sum_{s=1}^K \sum_{k \in \mathcal{S}_s} p_{jk} Q_k.$$

Define a set Y as follows

$$Y = \{y = (y_1, y_2, \cdots, y_K) : y_s \in \mathcal{S}_s, s = 1, \cdots, K\}.$$

In other words, y is an index vector with the sth element belonging to $\mathcal{S}_s$, for $s = 1, \cdots, K$. For every $y \in Y$, we define a convex combination of $Q_1, \cdots, Q_n$ as

$$\begin{aligned} C_y &= \sum_{s=1}^K \left(\sum_{\substack{k \in \mathcal{S}_s \\ k \neq y_s}} \min\{p_{ik}, p_{jk}\} Q_k \right. \\ &\quad \left. + \left(\gamma_{st} - \sum_{\substack{k \in \mathcal{S}_s \\ k \neq y_s}} \min\{p_{ik}, p_{jk}\} \right) Q_{y_s} \right). \end{aligned}$$

Since R_i and R_j are in the convex hull of $\{C_y : y \in Y\}$, we obtain

$$\begin{aligned} \|R_i - R_j\| &\leq \max_{y, y' \in Y} \|C_y - C_{y'}\| \\ &\leq \max_{y, y' \in Y} \sum_{s=1}^K \left(\gamma_{st} - \sum_{k \in \mathcal{S}_s} \min\{p_{ik}, p_{jk}\} \right) \\ &\quad \cdot \|Q_{y_s} - Q_{y'_s}\| \\ &\leq \tau_{\mathcal{S}}(P)\Delta_{\mathcal{S}}(Q). \end{aligned}$$

By the arbitrariness of i, j and t, we finally arrive at

$$\Delta_{\mathcal{S}}(R) \leq \tau_{\mathcal{S}}(P)\Delta_{\mathcal{S}}(Q)$$

as desired.

When $Q = I$, it is easy to see that $\Delta_{\mathcal{S}}(Q) = 1$ and therefore $\Delta_{\mathcal{S}}(P) \leq \tau_{\mathcal{S}}(P)$. □

We remark that the stronger inequality

$$\Delta_{\mathcal{S}}(PQ) \leq \alpha_{\mathcal{S}}(P)\Delta_{\mathcal{S}}(Q)$$

does not hold. To see this, set $K = 1$. Using the l_1 norm for Δ, it reads $\tau(PQ) \leq \alpha(P)\tau(Q)$. Taking $Q = P$, and choosing P periodic of period 2, we obtain $\tau(P) = \tau(P^2) = 1$. If $\alpha(P) < 1$ the result of the above implies $1 < 1$, a contradiction.

Acknowledgment

The author expresses his sincere gratitude to the anonymous referee for careful reading of the manuscript and useful comments that helped to improve presentation of results and accentuate important details.

REFERENCES

[1] M. Akelbek and S. Kirkland, Coefficients of ergodicity and the scrambling index. *Linear Algebra Appl.*, 430(2009) 1111–1130.

[2] M. Artzrouni, The local coefficient of ergodicity of a nonnegative matrix. *SIAM J. Matrix Anal. Appl.*, 25(2003) 507–516.

[3] M. Artzrouni and O. Gavart, Nonlinear matrix iterative processes and generalized coefficient of ergodicity. *SIAM Matrix Anal. Appl.*, 21(2000) 1343–1353.

[4] R. L. Dobrushin, Central limit theorem for nonstationary Markov chains. I. *Theory Probab. Appl.*, 1(1956) 65–79.

[5] J. Hajnal, Weak ergodicity in non-homogeneous Markov chains. *Proc. Camb. Phil. Soc.*, 54(1958) 233–246.

[6] Y. Han, W. Lu, and T. Chen, Cluster consensus in discrete-time networks of multi-agents with adapted inputs. To appear in *IEEE Trans. Neural Netw. Learn. Syst.*

[7] D. J. Hartfiel, *Nonhomogeneous Matrix Products*. World Scientific, New Jersey, 2002.

[8] I. C. F. Ipsen and T. M. Selee, Ergodicity coefficients defined by vector norms. *SIAM J. Matrix Anal. Appl.*, 32(2011) 153–200.

[9] A. A. Markov, Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.* 15(1906) 135–156.

[10] U. Păun, A class of ergodicity coefficients, and applications. *Math. Rep. (Bucur.)*, 4(2002) 225–232.

[11] U. Păun, New classes of ergodicity coefficients, and applications. *Math. Rep. (Bucur.)*, 6(2004) 141–158.

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

[13] A. Paz, Ergodic theorems for infinite probabilistic tables. *Ann. Math. Statist.*, 41(1970) 539–550.

[14] E. Seneta, On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Camb. Phil. Soc.*, 74(1973) 507–513.

[15] E. Seneta, Explicit forms for ergodicity coefficients and spectrum localization. *Linear Algebra Appl.*, 60(1984) 187–197.

[16] E. Seneta, *Non-negative Matrices and Markov Chains*. Springer-Verlag, New York, 2006.

[17] Y. Shang, Exponential random geometric graph process models for mobile wireless networks. *Proc. of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, Zhangjiajie, 2009, 56–61.

[18] Y. Shang, Multi-agent coordination in directed moving neighborhood random networks. *Chin. Phys. B*, 19(2010) 070201.

[19] Y. Shang, L^1 group consensus of multi-agent systems with stochastic inputs under directed interaction topology. *Int. J. Control*, 86(2013) 1–8.

[20] J. Shen, A geometric approach to ergodic non-homogeneous Markov chains. In: (Eds. T.-X. He) *Wavelet Analysis and Multiresolution Methods*, Lecture Notes in Pure and Applied Mathematics, 212(2000) 341–366.

Yilun Shang received his Ph.D. in 2010 at Shanghai Jiao Tong University, performed his postdoctoral studies at University of Texas at San Antonio from 2010 to 2013, and will work as a researcher at Singapore University of Technology and Design from 2013. He is an invited reviewer of Mathematical Reviews and Zentralblatt MATH. His main research field is structure and dynamics of complex networks.