Journal of Applied Mathematics and Stochastic Analysis, **11**:3 (1998), 391-396.

COVARIANCE AND RELAXATION TIME IN FINITE MARKOV CHAINS

JULIAN KEILSON
University of Rochester
William E. Simon Graduate School of Business Administration
Rochester, NY 14627 USA

(Received February, 1998; Revised March, 1998)

The relaxation time T_{REL} of a finite ergodic Markov chain in continuous time, i.e., the time to reach ergodicity from some initial state distribution, is loosely given in the literature in terms of the eigenvalues λ_j of the infinitesimal generator $\underline{\underline{Q}}$. One uses $T_{REL} = \theta^{-1}$ where $\theta = \min_{\lambda_j \neq 0} \{\text{Re}\lambda_j[-\underline{\underline{Q}}]\}$. This paper establishes for the relaxation time θ^{-1} the theoretical solidity of the time reversible case. It does so by examining the structure of the quadratic distance $d(t)$ to ergodicity. It is shown that, for any function $f(j)$ defined for states j, the correlation function $\rho_f(\tau)$ has the bound $|\rho_f(\tau)| \leq \exp[-\theta|\tau|]$ and that this inequality is tight. The argument is almost entirely in the real domain.

Key words: Finite Markov Chains, Covariance, Relaxation Time.
AMS subject classifications: 60J27.

1. Introduction

Let $J(t)$ be any ergodic Finite Markov Chain in continuous time with generator $\underline{\underline{Q}}$. A single underscore will be used to denote vectors and a double underscore will be used for matrices. Let $\underline{p}^T(t) = \underline{p}^T(0)e^{t\underline{\underline{Q}}}$ be the state probability vector so that

$$\lim_{t\to\infty} \underline{p}^T(t) = \underline{e}^T = (e_n)_1^K > \underline{0}^T; \quad \underline{e}^T\underline{\underline{Q}} = \underline{0}^T.$$

We are interested in the relaxation time of $J(t)$. For time-reversible chains where all eigenvalues of $\underline{\underline{Q}}$ are real the relaxation time is well understood (cf. Keilson [1]). For more general chains with real eigenvalues and eigenvalues occurring in complex conjugate pairs, all eigenvalues $\lambda_j[\underline{\underline{Q}}]$ other than zero have $\text{Re}\lambda_j[\underline{\underline{Q}}] < 0$ (see appendix). Let $\theta = \min_{\lambda_j \neq 0}\{\text{Re}\lambda_j[-\underline{\underline{Q}}]\}$. The value $T_{REL} = \theta^{-1}$ is employed loosely for the relaxation time in the literature. This paper establishes for the relaxation time

θ^{-1} the theoretical solidity of the time reversible case.

Let $\underline{\underline{e}}_D = \text{diag}(e_n)$. Recall that $\sqrt{\underline{x}^T \underline{\underline{U}}\, \underline{x}}$ is a vector norm when $\underline{\underline{U}}$ is positive definite. The scalar function

$$d(t) = \sqrt{(\underline{p}^T(t) - \underline{e}^T)\underline{\underline{e}}_D^{-1}(\underline{p}^T(t) - \underline{e}^T)} \tag{1}$$

is then a vector norm and a distance to ergodicity.

It has been shown by D.G. Kendall [3] that the distance $d(t)$ is monotone decreasing for time reversible chains. It has also been shown by Keilson and Vasicek [2] that this distance decreases to zero for *all* ergodic chains. An independent proof will be given in this paper.

2. The Structure of the Distance to Ergodicity

The structure of $d(t)$ for all finite ergodic chains is examined more deeply here. This structure is used to establish the relaxation time θ^{-1} entirely in the real domain without any reference to complex eigenvalues until the end. We use the following notation:

Definitions:

2a) $\underline{\underline{e}}_D = \text{diag}(e_n)$ 2b) $\underline{\nu}^T(t) = (\underline{p}^T(t) - \underline{e}^T)\underline{\underline{e}}_D^{-1/2}$

2c) $w(t) = \underline{\nu}^T(t)\underline{\nu}(t)$ 2d) $d(t) = \sqrt{w(t)}$

2e) $\underline{\underline{Q}}^R = \underline{\underline{e}}_D^{-1}\underline{\underline{Q}}^T\underline{\underline{e}}_D$ 2f) $\underline{\underline{Q}}^{\#} = \frac{1}{2}\left[\underline{\underline{Q}} + \underline{\underline{Q}}^R\right]$

2g) $\underline{\underline{B}} = \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}\,\underline{\underline{e}}_D^{-1/2}$ 2h) $\underline{\underline{C}} = \frac{1}{2}\left[\underline{\underline{B}} + \underline{\underline{B}}^T\right] = \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{\underline{e}}_D^{-1/2}$

The superscript R refers to the reverse chain. Note that $\underline{\underline{Q}}$, $\underline{\underline{Q}}^R$, and $\underline{\underline{Q}}^{\#}$ all generate chains which have the same ergodic vector $\underline{e}^T$.

Theorem A: *For any finite ergodic Markov chain* $J(t)$:

(*a*) $\underline{\underline{Q}}^{\#}$ *is the Q-matrix of an ergodic chain;*

(*b*) $\underline{\underline{C}} = \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{\underline{e}}_D^{-1/2}$ *is symmetric and negative-semidefinite with eigenvalues* $\lambda_1 = 0,\ \lambda_j < 0,\ j \neq 1$.

(*c*) *The stationary chain generated by* $\underline{\underline{Q}}^{\#}$ *is time-reversible.*

Proof: $\underline{\underline{Q}}$ and $\underline{\underline{Q}}^R = \underline{\underline{e}}_D^{-1}\underline{\underline{Q}}^T\underline{\underline{e}}_D$ have the zero structure needed to be an ergodic Q matrix as does $\underline{\underline{Q}}^{\#} = \frac{1}{2}\left[\underline{\underline{Q}} + \underline{\underline{Q}}^R\right]$. The matrices $\underline{\underline{Q}}$, $\underline{\underline{Q}}^R$ and $\underline{\underline{Q}}^{\#}$ all have row sum zero. Also $2\underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{\underline{e}}_D^{-1/2} = \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}\,\underline{\underline{e}}_D^{-1/2} + \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^R\underline{\underline{e}}_D^{-1/2} = \underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}\,e_D^{-1/2} + \underline{\underline{e}}_D^{-1/2}\underline{\underline{Q}}^T\underline{\underline{e}}_D^{1/2}$ is symmetric. Hence $\underline{\underline{e}}_D\underline{\underline{Q}}^{\#} = \underline{\underline{e}}_D^{1/2}\left[\underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{\underline{e}}_D^{-1/2}\right]\underline{\underline{e}}_D^{1/2}$ is symmetric and $J^{\#}(t)$ governed by $\underline{\underline{Q}}^{\#}$ is time-reversible (cf. Keilson [1]). □

Theorem B: *Let* $\theta = min_{\lambda_j \neq 0}\{\lambda_j[-\underline{\underline{C}}]\} = min_{\lambda_j \neq 0}\{\lambda_j[-\underline{\underline{Q}}^{\#}]\}$. $T^{\#}_{REL} = \theta^{-1}$ *is then the relaxation time of the ergodic time reversible chain* $J^{\#}(t)$. *For any finite ergodic Markov chain, it has been shown in* [4] *that*

$$\frac{d(t)}{d(0)} = \sqrt{\frac{w(t)}{w(0)}} \le e^{-\theta t}. \tag{3}$$

The rate θ will be called the *global decay rate* of $d(t) = \sqrt{w(t)}$. The proof is given here.

Proof: From the definitions, one has at once $\frac{d}{dt}\underline{\nu}^T(t) = \underline{\nu}^T(t)\underline{\underline{B}}$ and $\frac{d}{dt}\underline{\nu}(t) = \underline{\underline{B}}^T\underline{\nu}(t)$. Hence $\frac{d}{dt}w(t) = \frac{d}{dt}[\underline{\nu}^T(t)\underline{\nu}(t)] = \underline{\nu}^T(t)\underline{\underline{B}}\,\underline{\nu}(t) + \underline{\nu}^T(t)\underline{\underline{B}}^T\underline{\nu}(t)$. This implies

$$\frac{d}{dt}w(t) = 2\left[\underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)\right]. \tag{4}$$

Since $\underline{x}^T\underline{\underline{C}}\,\underline{x} < 0$, for all real $\underline{x} \ne \underline{0}$, $\frac{d}{dt}w(t) \le 0$ and $w(t)$ decreases in t. The matrix $\underline{\underline{C}}$ has principal left eigenvector $\underline{e}_C^T = (e_n^{1/2})$ corresponding to eigenvalue 0 and $\underline{\nu}^T(t)\underline{e}_C = (\underline{p}^T t) - \underline{e}^T)\underline{1} = 0$. Thus $\underline{\nu}^T(t)$ is orthogonal to the principal rank one eigenspace of $\underline{\underline{C}}$. When $\underline{p}^T(0) \ne \underline{e}^T$, $\underline{\nu}^T(t)$ moves in this space, and does not vanish. One has from the Rayleigh-Ritz principal

$$\frac{d}{dt}\log w(t) = 2\frac{\underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)}{\underline{\nu}^T(t)\underline{\nu}(t)} \le -2\min_{\lambda_j > 0}\{\lambda_j[-\underline{\underline{C}}]\} = -2\theta.$$

If one integrates from 0 to t, Theorem B follows. □

Convexity Lemma: *The function* $w(t)$ *is convex and*

$$\frac{d^2}{dt^2}w(t) = 4\underline{\nu}^T(t)\underline{\underline{C}}^2\underline{\nu}(t) \ge 0. \tag{5}$$

Proof: From $\frac{d}{dt}w(t) = \frac{d}{dt}[\underline{\nu}^T(t)\underline{\nu}(t)] = 2\underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)$ one has

$$\frac{d^2}{dt^2}w(t) = \frac{d}{dt}[\underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)] = 2\underline{\nu}^T(t)[\underline{\underline{B}}\,\underline{\underline{C}} + \underline{\underline{C}}\,\underline{\underline{B}}^T]\underline{\nu}(t).$$

But

$$2[\underline{\underline{B}}\,\underline{\underline{C}} + \underline{\underline{C}}\,\underline{\underline{B}}^T] = \underline{\underline{B}}[\underline{\underline{B}} + \underline{\underline{B}}^T] + [\underline{\underline{B}} + \underline{\underline{B}}^T]\underline{\underline{B}}^T = (\underline{\underline{B}} + \underline{\underline{B}}^T)^2 + (\underline{\underline{B}}\,\underline{\underline{B}}^T - \underline{\underline{B}}^T\underline{\underline{B}})$$

and $(\underline{\underline{B}}\,\underline{\underline{B}}^T - \underline{\underline{B}}^T\underline{\underline{B}})$ is antisymmetric. The lemma then follows. □

A stronger result is available.

Theorem C: *For any finite ergodic Markov chain, with* $d(0) \ne 0$,

(*a*) $w(t)$ *is convex and decreasing in* t;

(*b*) $\log w(t)$ *is convex in* t, *i.e.* $\frac{w'(t)}{w(t)}$ *increases with* t;

(*c*) $\frac{w'(t)}{w(t)} \le -2\theta$. *This equality is tight, i.e., an initial state vector can be found for which* $\frac{w'(t)}{w(t)} = -2\theta$ *for all* t.

Proof: From the proof of Theorem B, $w'(t) < 0$. We must show that $[\log w(t)]'' = \frac{w(t)w''(t) - [w'(t)]^2}{w^2} \ge 0$. Calculation gives

$$w(t)w''(t) - [w'(t)]^2 = 4\left\{[\underline{\nu}^T(t)\underline{\nu}(t)][\underline{\nu}^T(t)\underline{\underline{C}}^2\underline{\nu}(t)] - [\underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)]^2\right\}$$

where $\underline{\underline{C}} = \underline{e}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{e}_D^{-1/2}$. Moreover since $\underline{\underline{C}}$ is symmetric, the Schwartz inequality gives

$$\left| \underline{\nu}^T(t)\underline{\underline{C}}\,\underline{\nu}(t)\right|^2 \leq \left| \underline{\nu}^T(t)\underline{\underline{C}}\right|^2[\underline{\nu}^T(t)\underline{\nu}(t)] = \left[\underline{\nu}^T(t)\underline{\underline{C}}^2\underline{\nu}(t)\right][\underline{\nu}^T(t)\underline{\nu}(t)]$$

and $w(t)w''(t) - [w'(t)]^2 \geq 0$. This proves log-convexity.

For $\underline{p}^T(0) \neq \underline{e}^T$, the Schwartz inequality is strict and the convexity of $\log\frac{w(t)}{w(0)}$ is strict unless $\underline{\nu}^T(t)\underline{\underline{C}} = K\underline{\nu}^T$ for some constant K. This can only happen when $\underline{\nu}^T(t)$ is an eigenvector of $\underline{\underline{C}}$, i.e., when $(\underline{p}^T(t) - \underline{e}^T)\underline{\underline{e}}_D^{-1/2}\underline{\underline{C}} = \lambda_j[\underline{\underline{C}}](\underline{p}^T(t) - \underline{e}^T)\underline{\underline{e}}_D^{-1/2}$.

We next show that the inequality (3) in Theorem B is tight in that, for any ergodic chain one can always find a $\underline{p}^T(0)$ for which $\frac{w(t)}{w(0)} = \exp[-2\theta]t$. If $\frac{w'(0)}{w(0)} = -2\theta$, knowledge that $\frac{w'(t)}{w(t)}$ is increasing and $\frac{w'(t)}{w(t)} \leq -2\theta$ implies that $\frac{w'(t)}{w(t)} = -2\theta$ for all t. Let $\underline{u}_1^T$ be any real orthonormal eigenvector of $\underline{\underline{C}}$ for the eigenvalue $-\theta$, and α be small and real. If $\underline{\nu}^T(0) = (\underline{p}^T(0) - \underline{e}^T)\underline{\underline{e}}_D^{-1/2} = \alpha\underline{u}_1^T$, so that $\underline{\nu}^T(0)\underline{\underline{C}} = -\alpha\theta\underline{u}_1^T$ then $\frac{w'(0)}{w(0)} = 2\,\frac{\underline{\nu}^T(0)\underline{\underline{C}}\,\underline{\nu}(0)}{\underline{\nu}^T(0)\underline{\nu}(0)} = -2\theta$ as needed. If one chooses $\underline{p}^T(0) = \underline{e}^T + \alpha\underline{u}_1^T\underline{\underline{e}}_D^{1/2}$, one will have $\underline{p}^T(0) > \underline{0}^T$ for α sufficiently small. Moreover one will also have $\underline{p}^T(0)\underline{1} = 1$. For

$$(\underline{p}^T(0) - \underline{e}^T)\underline{1} = \alpha\underline{u}_1^T\underline{\underline{e}}_D^{1/2}\underline{1} \text{ and } -\theta\underline{u}_1^T\underline{\underline{e}}_D^{1/2}\underline{1} = \underline{u}_1^T\underline{\underline{C}}\,\underline{\underline{e}}_D^{1/2}\underline{1} = \underline{u}_1^T\underline{\underline{e}}_D^{1/2}\underline{\underline{Q}}^{\#}\underline{1} = 0. \quad \square$$

Time-reversible case: When $J(t)$ governed by $\underline{\underline{Q}}$ is time reversible, a special case of the above, $\underline{\underline{Q}} = \underline{\underline{Q}}^R = \underline{\underline{Q}}^{\#}$ and θ is just the reciprocal of the relaxation time described in [1]. For this time reversible case, for any λ_j and real eigenvector $\underline{u}_j^T$ of $\underline{\underline{Q}}$, one can find initial vectors $\underline{p}^T(0)$ for which $w(t)$ will have purely exponential decay at rate $2\,|\,\lambda_j\,|$ faster than 2θ. The global decay rate is still θ.

3. The Covariance Function

Let $J(t)$ be any finite ergodic Markov chain which is stationary. Let $f(j)$ be any real function of state j. The covariance function is $R_f(\tau) = \mathrm{cov}[f(J(t), f(J(t+\tau)]$ and (cf. [1]) $R_f(\tau) = \underline{f}^T\underline{\underline{e}}_D[\underline{\underline{p}}(\tau) - \underline{1}\,\underline{e}^T]\underline{f}$ for $\tau > 0$.

The correlation function is $\rho_f(\tau) = \frac{R_f(\,|\,\tau\,|\,)}{R_f(0)}$.

Theorem D: *For any finite ergodic Markov chain $J(t)$ which is stationary, the correlation function satisfies*

$$|\,\rho_f(\tau)\,| \leq exp[-\theta\,|\,\tau\,|\,]$$

and the inequality is tight.

Proof: Without loss of generality, we may assume that $f_n > 0$ since a positive constant may always be added to f_n without altering the covariance. Let $\underline{\underline{p}}(\tau) =$

$[p_{mn}(\tau)]$ be the transition probability matrix of $J(t)$. For $\underline{\nu}^T(\tau) = \frac{\underline{f}^T \underline{\underline{e}}_D}{\underline{f}^T \underline{\underline{e}}_D \underline{1}}[\underline{p}(\tau) - \underline{1}\,\underline{e}^T]\,\underline{\underline{e}}_D^{-1/2}$, algebra gives

$$R_f(\tau) = \underline{f}^T \underline{\underline{e}}_D[\underline{p}(\tau) - \underline{1}\,\underline{e}^T]\underline{\underline{e}}_D^{-1/2}\underline{\underline{e}}_D^{-1/2}[\underline{\underline{e}}_D \underline{f} - (\underline{f}^T \underline{\underline{e}}_D \underline{1})\underline{e}]$$

$$= (\underline{f}^T \underline{\underline{e}}_D \underline{1})^2[\underline{\nu}^T(\tau)\underline{\nu}(0)].$$

From the Schwartz inequality,

$$\rho_f^2(\tau) = \frac{R_f^2(\tau)}{R_f^2(0)} = \frac{[\underline{\nu}^T(\tau)\underline{\nu}(0)]^2}{[\underline{\nu}^T(0)\underline{\nu}(0)]^2} \le \frac{|\underline{\nu}(\tau)|^2}{|\underline{\nu}(0)|^2} = \frac{w(\tau)}{w(0)} \le \exp[-2\theta\tau].$$

The tightness of the inequality follows from the tightness in Theorem C. This proves the theorem. □

4. The Relaxation Time

In [1] the relaxation time was defined by $T_{REL} = \sup_f \int_0^\infty \rho_f(\tau)d\tau$. This was motivated by the similarity of $\rho_f(\tau)$ to a survival function. One then has at once from Theorem D, $T_{REL} = \theta^{-1}$.

One must finally relate the decay rate θ to the eigenvalues of $\underline{\underline{Q}}$. Suppose that there are eigenvalues of $\underline{\underline{Q}}$ with negative real part $-\zeta$ and that other eigenvalues have a more negative real part. Consider $w(t) = \underline{\nu}^T(t)\underline{\nu}(t) = (\underline{p}^T(t) - \underline{e}^T)\underline{\underline{e}}_D^{-1}(\underline{p}(t) - \underline{e})$. $e^{2\zeta t}w(t) = [e^{\zeta t}(\underline{p}^T(t) - \underline{e}^T)]\underline{\underline{e}}_D^{-1}[e^{\zeta t}(\underline{p}(t) - \underline{e})] = [e^{\zeta t}d(t)]^2$ and that this is log-convex in t, all ζ. For $\underline{p}^T(t) - \underline{e}^T \neq 0$, $\limsup_{t\to\infty} e^{2\zeta t}w(t) = 0$, when $\zeta < -\theta^*$ and $\limsup_{t\to\infty} e^{2\zeta t}w(t) = \infty$, when $\zeta > = \theta^*$. From the tightness in Theorem C, we must then identify θ^* with θ.

A calculation has been carried out using the symbolic and numerical power of Maple for the chain $J(t)$ starting in any state and generator

$$\underline{\underline{Q}} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The graph given by Maple is found to be log-convex as predicted. The symbolic expression for $w(t)$ has the asymptotic decay rate predicted.

Acknowledgement

The author wishes to thank P. Gundelpudi and A. Roy for their interest and help.

Appendix

Lemma: *If* $\lambda_j[\underline{\underline{Q}}]$ *is the eigenvalue of an ergodic finite Markov chain in continuous time with infinitesmal generator* $\underline{\underline{Q}}$*, then apart from the principal eigenvalue at* 0*, all eigenvalues of* $\underline{\underline{Q}}$ *have a strictly negative real part.*

Proof: We may use uniformization [1] to write $\underline{\underline{Q}} = -\nu[\underline{\underline{I}} - \underline{\underline{a}}_\nu]$ where ν is any positive rate exceeding the largest exit rate from a state and $\underline{\underline{a}}_\nu$ is a stochastic ergodic (irreducible and aperiodic) matrix. Then $\lambda_j[\underline{\underline{Q}}] = -\nu(1 - \lambda_j[\underline{\underline{a}}_\nu])$ and $|\lambda_j[\underline{\underline{a}}_\nu]| < 1$ for other than $\lambda_1[\underline{\underline{a}}_\nu] = 1$. The lemma then follows.

Remark: One can have a stochastic matrix which is ergodic and has purely imaginary eigenvalues.

An example with eigenvalues 1, $-\frac{1}{2}$, $\frac{1}{2}i$, $-\frac{1}{2}i$ is

$$\begin{bmatrix} \frac{1}{8} & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{5}{8} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}.$$

References

[1] Keilson, J., *Markov Chain Models - Rarity and Exponentiality*, Springer-Verlag, Applied Mathematical Sciences Series **28** 1979.

[2] Keilson, J. and Vasicek, O.A., Monotone measures of ergodicity for Markov chains, Special Issue for Ryszard Syski, *J. of Appl. Math. and Stoch. Anal.*, **11**:3 (1998), 281-286.

[3] Kendall, D.G., *Unitary Dilations of Markov Transition Operators and the Corresponding Integral Representations for Transition Probability Matrices*, (ed. by U. Grenander), Probability and Statistics, Almqvist and Wiksell, Stockholm 1959.

[4] Roy, Abhijit, *Transient Analysis of Queueing Systems*, Ph.D. Thesis, The University of Rochester (1996). This is available from UMI, Ann Arbor, MI.