COVARIANCE AND RELAXATION TIME INFINITE 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 \( \lambda_j \) of the infinitesimal generator \( Q \). One uses \( T_{REL} = \theta^{-1} \) where \( \theta = \min_{\lambda_j \neq 0} \{\text{Re}\lambda_j[-Q]\} \). This paper establishes for the relaxation time \( \theta^{-1} \) 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 \( \vert \rho_f(\tau) \vert \leq \exp(-\theta \vert \tau \vert) \) 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 \( Q \). A single underscore will be used to denote vectors and a double underscore will be used for matrices. Let \( p^T(t) = p^T(0)e^{tQ} \) be the state probability vector so that

\[
\lim_{t \to \infty} p^T(t) = \xi^T = (e_n)_1^K > \xi^T Q = \xi^T Q.
\]

We are interested in the relaxation time of \( J(t) \). For time-reversible chains where all eigenvalues of \( Q \) 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 \( \lambda_j[Q] \) other than zero have \( \text{Re}\lambda_j[Q] < 0 \) (see appendix). Let \( \theta = \min_{\lambda_j \neq 0} \{\text{Re}\lambda_j[-Q]\} \). The value \( T_{REL} = \theta^{-1} \) is employed loosely for the relaxation time in the literature. This paper establishes for the relaxation time
Let $e_D = \text{diag}(e_n)$. Recall that $\sqrt{e^TUe}$ is a vector norm when $U$ is positive definite. The scalar function

$$d(t) = \sqrt{(\nu^T(t) - e^T)e_D^{-1}(\nu^T(t) - e^T)}$$

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 $\theta^{-1}$ entirely in the real domain without any reference to complex eigenvalues until the end. We use the following notation:

**Definitions:**

2a) $e_D = \text{diag}(e_n)$

2b) $\nu^T(t) = (p^T(t) - e^T)e_D^{-1/2}$

2c) $w(t) = \nu^T(t)\nu(t)$

2d) $d(t) = \sqrt{w(t)}$

2e) $Q^R = e_D^{-1/2}Q^Te_D^{-1/2}$

2f) $Q^# = \frac{1}{2} \left[ Q + Q^R \right]$

2g) $B = e_D^{1/2}Q^#e_D^{-1/2}$

2h) $C = \frac{1}{2} \left[ B + B^T \right] = e_D^{1/2}Q^#e_D^{-1/2}$

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

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

(a) $Q^#$ is the $Q$-matrix of an ergodic chain;

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

(c) The stationary chain generated by $Q^#$ is time-reversible.

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

**Theorem B:** Let $\theta = \min \lambda_j \neq 0 \{ \lambda_j[-C] \} = \min \lambda_j \neq 0 \{ \lambda_j[-Q^#] \}$. $T^#_{\text{REL}} = \theta^{-1}$ 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
\[ \frac{d(t)}{d(0)} = \sqrt{\frac{w(t)}{w(0)}} \leq e^{-\theta t}. \] (3)

The rate \( \theta \) will be called the global decay rate of \( d(t) = w(t) \). The proof is given here.

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

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

Since \( x^TCx < 0 \), for all real \( x \neq 0 \), \( \frac{d}{dt}w(t) \leq 0 \) and \( w(t) \) decreases in \( t \). The matrix \( C \) has principal left eigenvector \( \epsilon_C = (e_1^{1/2}) \) corresponding to eigenvalue 0 and \( \nu^T(t)\epsilon_C = (p^T(t) - e^T)1 = 0 \). Thus \( \nu^T(t) \) is orthogonal to the principal rank one eigenspace of \( C \). When \( p^T(0) \neq e^T \), \( \nu^T(t) \) moves in this space, and does not vanish.

One has from the Rayleigh-Ritz principal

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

If one integrates from 0 to \( t \), Theorem B follows.

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

\[ \frac{d^2}{dt^2}w(t) = 4\nu^T(t)C^2\nu(t) \geq 0. \] (5)

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

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

But

\[ 2[B + C + C^T] = B[B + B^T] + [B + B^T]B^T = (B + B^T)^2 + (B B^T - B^T B) \]

and \( (B B^T - B^T B) \) is antisymmetric. The lemma then follows.

A stronger result is available.

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

(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) \( \frac{w'(t)}{w(t)} \leq -2\theta. \) 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\[ [\log w(t)]'' = \frac{w(t)w''(t) - [w'(t)]^2}{w^2} \geq 0. \] Calculation gives

\[ w(t)w''(t) - [w'(t)]^2 = 4\left[ [\nu^T(t)\nu(t)][\nu^T(t)\nu(t)] - [\nu^T(t)C\nu(t)]^2 \right] \]

where \( C = e_0^{1/2}Qe_D^{1/2}e_D^{1/2} \). Moreover since \( C \) is symmetric, the Schwartz inequality
gives
\[ |\nu^T(t)\nu(t)|^2 \leq |\nu^T(t)\nu(t)|^2 [\nu^T(t)\nu(t)] = [\nu^T(t)\nu(t)][\nu^T(t)\nu(t)] \]
and \( w(t)w''(t) - [w'(t)]^2 \geq 0 \). This proves log-convexity.

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

We next show that the inequality (3) in Theorem B is tight in that, for any ergodic chain one can always find a \( p^T(0) \) for which \( w(t) < -20 \) implies that \( w'(t) - 20 \) for all \( t \).

Let \( u_1^T \) be any real orthonormal eigenvector of \( \mathbb{C} \) for the eigenvalue \( -\theta \), and \( \alpha \) be small and real. If \( \nu^T(t) = (p^T(t) - \varepsilon^T)\mathbb{C}^{-1/2} = \alpha u_1^T \), so that \( \nu^T(t)C = -\alpha \theta u_1^T \) then
\[
\frac{w'(t)}{w(0)} = 2 \frac{\nu^T(t)\nu(t)}{\nu^T(t)\nu(t)} = -2\theta \text{ as needed.}
\]
If one chooses \( p^T(0) = \varepsilon^T + \alpha u_1^T \mathbb{C}^{-1/2} \), one will have \( p^T(0) > u_1^T \) for \( \alpha \) sufficiently small. Moreover one will also have \( p^T(0) \| = 1 \). For
\[
(p^T(0) - \varepsilon^T)\| = \alpha u_1^T \mathbb{C}^{-1/2} \text{ and } -\theta u_1^T \mathbb{C}^{-1/2} \| = u_1^T C^1/2 \| = u_1^T \mathbb{C}^{-1/2} \mathbb{C}^1/2 \| = u_1^T \mathbb{C}^1/2 \mathbb{C}^1/2 \| = 0. \]

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

### 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 \( R_f(\tau) = \text{cov}[f(J(t)), f(J(t+\tau))] \) and (cf. [1]) \( R_f(\tau) = \int_\tau^T \int_\tau^T R_f(|\tau|) \).

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 \( p(\tau) = \)
be the transition probability matrix of $J(t)$. For $
u^T(\tau) = \frac{\int_T \mathcal{D} \nu(\tau)}{\int_{\mathcal{D}} \mathcal{D} \nu(\tau)}$, algebra gives

$$R_f(\tau) = \frac{\int_T \mathcal{D} \nu(\tau) - 1}{\mathcal{D} \nu(\tau)} \frac{1}{2} \prod (f) \mathcal{D} \nu(\tau) \nu(0).$$

From the Schwartz inequality,

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

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 $\theta$ to the eigenvalues of $Q$. Suppose that there are eigenvalues of $Q$ with negative real part $-\zeta$ and that other eigenvalues have a more negative real part. Consider $w(t) = \nu^T(t) \nu(t) = (\mathcal{D} \nu(\tau) - \mathcal{D} \nu(0)) \mathcal{D} \nu(\tau) - \mathcal{D} \nu(0)$. $e^{2\zeta t}w(t) = [e^{\zeta t}(\mathcal{D} \nu(\tau) - \mathcal{D} \nu(0))] \mathcal{D} \nu(\tau) = [e^{\zeta t}(\mathcal{D} \nu(\tau) - \mathcal{D} \nu(0))] \mathcal{D} \nu(\tau)$ and that this is log-convex in $t$, all $\zeta$. For $\mathcal{D} \nu(\tau) - \mathcal{D} \nu(0) \neq 0$, $\limsup_{t \to \infty} e^{2\zeta t}w(t) = 0$, 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 $\theta^*$ with $\theta$.

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

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

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

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

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

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

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

References


Submit your manuscripts at http://www.hindawi.com