A number of important theorems arising in connection with Gaussian elimination are derived, using semi-regenerative analysis. The implications of these theorems to find steady-state solutions of Markov chains are analysed. The results obtained in this way are then applied to quasi birth-death processes.
PROBABILISTIC APPROACH TO COMPUTATIONAL ALGORITHMS FOR FINDING STATIONARY DISTRIBUTIONS OF MARKOV CHAINS

by

MICHAEL I. TAKSAR*
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306

and

WINFRIED K. GRASSMANN
UNIVERSITY OF SASKATCHEWAN
SASKATOON, SASKATCHEWAN

FSU TECHNICAL REPORT NO. M751
AFOSR TECHNICAL REPORT NO. 86-203

OCTOBER, 1986

*Research supported by the Air Force Office of Scientific Research under Grant Number F49620-85-C-0007.

Michael J. Taksar and Winfried K. Grassmann

Abstract

A number of important theorems arising in connection with Gaussian elimination are derived, using semi-regenerative analysis. The implications of these theorems to find steady-state solutions of Markov chains are analysed. The results obtained in this way are then applied to quasi birth-death processes.

Introduction

Originally, the determination of stationary distributions in Markov chains was done, using completely algebraic arguments. Unfortunately, algebraic arguments often do not take advantage of the particular structure of stochastic matrices which form the coefficients of the equilibrium equations. Recently, a new trend emerged which looks into the probabilistic interpretation of the algorithms for solving equilibrium equations. The objective of this paper is to further this trend by deriving computational algorithms from purely probabilistic arguments. The arguments employed in this paper are based on the semi-regenerative structure of Markov chains. This approach gives better insights into formal manipulations of equilibrium equations and provides probabilistic interpretations of the coefficients obtained at each step of such manipulations. This allows one to make connections and draw conclusions that would not be obvious otherwise. In particular, we derive a number of theorems leading to a better understanding of Gaussian elimination.
a semi-regenerative processes. These theorems allow one to derive relationships between certain matrix-geometric solutions and Gaussian elimination Neuts [1981], Gaver et al. [1984].

From a Markov chain \( \{Y_n, n=0,1,\ldots\} \) one can obtain a semi-regenerative process by recording the state of the chain only while it visits points of a given subset \( D \) of the state space of \( Y_n \). Specifically, let \( T_n, n = 1,2,\ldots \), be the time of the \( n \)th visit to \( D \) and let \( X_n \) be the position of the chain at \( T_n \). The process \( \{(X_n,T_n), n=1,2,\ldots\} \) is then a Markov renewal process (see Cinlar [1975] Chapter 10), \( \{X_n, n=1,2,\ldots\} \) is a semi-regenerative process and the \( T_n \) are the semi-regenerative epochs of the Markov renewal process. Thus, if we consider

\[
A_n = \{ Y_{T_n-1}, Y_{T_n-1+1}, \ldots, Y_{T_n-1} \}
\]

then the conditional distribution of the sequence \( \{A_nA_{n+1},\ldots\} \), given the past of the process up to \( T_{n-1} \), depends only on \( X_{n-1} \), and all \( A_n \) are conditionally independent, given \( (X_n,T_n) \). The analysis of the behaviour of \( Y_n \) from one semi-regenerative epoch to another produces the main relation between steady-state probabilities that is used for developing the algorithm. For simplicity, we assume that the Markov chain \( \{Y_n, n=0,1,\ldots\} \) is irreducible and aperiodic, and that the state space of \( Y_n \) is \( E=(0,1,2,\ldots) \). Such a Markov chain reaches steady state (see Cinlar [1975], Chapter 8), a fact that is expressed by the following relation

\[
Y_n \rightarrow Y.
\]
Alternatively, one can write

\[
P_i(Y_n = j) \rightarrow p_j, j = 0,1,\ldots,
\]

Where \( p_j \) is the distribution of \( Y_n \), and \( P_i(.) = P( . \mid Y_0 = i) \). The notation \( E_i \) must be understood in a similar way.
Let 

\[ D = \{0,1,2,...,d-1\} \]

and

\[ T = \min\{ m > 0 : Y_m \in D \}. \]  

(1)

Let \( 1_i(Y_n) \) be the indicator function for \( Y_n = i \), and let

\[ v_{ij}^{(d)} = \mathbb{E}_i\{ \sum_{m=0}^{T} 1_j(Y_m) \} = \mathbb{E}_i\{m : m < T \text{ and } Y_m = j \}. \]  

(2)

The symbol \( \# \) denotes the cardinality of a set. The expression, \( \#\{m : m > T \text{ and } Y_m = j\} \), in particular, is the number of times \( m \) meets the condition in question, and the quantity \( v_{ij}^{(d)} \) becomes the expected number of visits to state \( j \) prior to the exit from \( E-D \), given one starts from the point \( i \). The \( v_{ij}^{(d)} \) are related to the \( p_j \) according to the following theorem:

**Theorem 1:** Let \( (p_0,p_1,...) \) be the steady-state distribution of the Markov chain \( Y_n \). Then

\[ p_j = \sum_{i=1}^{d-1} v_{ij}^{(d)} p_i. \]  

(3)

**Proof:** Let

\[ T_1 = T \]

\[ T_{n+1} = \min\{ m > T_n : Y_m \in D \} \]

\[ X_n = Y_{T_n}. \]  

(4)

From these definitions, it follows that \( T_n \) is a stopping time, and that \( X_n \) can only assume the values \( 0,1,...,d-1 \). As a consequence, the strong Markov property applies, which means that \( X_n \) is a Markov chain with the state space \( D = \{0,1,...,d-1\} \). This chain is irreducible because the original chain \( Y_n \) is
irreducible. Let \( \nu_i, i = 0, 1, \ldots, d-1 \) be the unique invariant distribution for \( Y_n \).

Let \( m(i) = E[T] \) be the times between regeneration, and let

\[
K_n(i, j) = P(T > n) \{ X_n = j \}
\]

According to Cinlar [1975], Chapter 10, Theorem (6.12)

\[
\lim_{n \to \infty} P(T = j) = \frac{\sum_{k \in D} K_n(k, j)}{\sum_{k \in D} m(k)} \sum_{i=0}^{d-1} \nu_i \left( Y_m \right) > m \quad (5)
\]

We must mention Cinlar assumes that the times between the Markov-renewal have a continuous, aperiodic distribution. Since we are dealing with a discrete-time Markov renewal process, the distributions in question have in fact a periodicity of 1. However, the proof for the discrete case is basically identical to its continuous counterpart. The sum of the \( K_n(k, j) \) in (5) can be found as

\[
\sum_{m=0}^{d} P_m(Y_m = j, T > m) = \sum_{i=0}^{d-1} \sum_{k \in D} C_i \left( 1_{Y_m} \right) 1_{T > m} = v_{ij}^{(d)} \quad (6)
\]

Since \( P_m(Y_m = j) \) converges to \( p_j \), one finds from (5) and (6)

\[
p_j = \sum_{i=0}^{d} \sum_{k=0}^{d} \frac{\nu_i}{m(k)} v_{ij}^{(d)} \quad (7)
\]

If \( i, j \in D \), \( v_{ij}^{(d)} \) is 1 if \( i = j \), and zero otherwise. In this case, (6) becomes

\[
p_i = \sum_{k=0}^{d} \frac{\nu_i}{m(k)} v_{ik}^{(d)} \quad (8)
\]

a relation that is also well known from semi-Markov processes. From (7) and (8), one easily obtains (3), which proves the theorem. (3) can also be proven in different ways (Miller [1984], Grassmann et al. [1985]), but then, the connec-
tion to semi-regenerative theory is lost.

Later on, we need to relate the $v_{ij}^{(k)}$ for different values of $i,j,k$. Such relationships can be obtained from the following theorem.

Theorem 2: Let $v_{ij}^{(k)}$ be given by (2), and let $p_{ij}$ be the transition probabilities of the Markov chain $Y_n$. Then for $k > d$

$$v_{ij}^{(d)} = v_{ij}^{(k)} + \sum_{\nu=d}^{k-1} v_{ij}^{(\nu)} v_{\nu j}^{(k)}$$

(9)

Proof: Let $D$ consist of the first $d$ points form $0$ to $d-1$, and let $T$ be defined by (1). Similarly, let $K$ consist of the first $k$ points from $0$ to $k-1$, and let

$$\tau = \min\{m > 0: Y_m \in K\}$$

(10)

Since the first visit to $K$ occurs prior to the first visit to $D$, $\tau < T$, and one has

$$v_{ij}^{(d)} = E_i(\#\{m > 0: Y_m = j, m < T\})$$

$$= E_i(\#\{m: Y_m = j, 0 < m < \tau\}) + E_i(\#\{m: Y_m = j, \tau < m < T\})$$

$$= v_{ij}^{(k)} + E_i(\#\{m: Y_m = j, \tau < m < T\})$$

(11)

We now consider the successive visits to $K-D$. Thus, let $\xi(n)$ the $n$th visit to K-D, that is,

$$\xi(1) = \min\{m > 0, Y_m \in K-D\}$$

$$\xi(n) = \min\{m > \xi(n-1), Y_n \in K-D\}.$$

The $\xi(n)$ are obviously stopping times, and $\xi(n) \to \infty$ as $n \to \infty$. Moreover,

$$\tau = \min\{m > 0: Y_m \in K\} < \min\{m > 0: Y_m \in K-D\} = \xi(1).$$

(11)

In (10), the equality holds if the first visit to $K$ is also a visit to $K-D$. In a similar way, one finds from (1) and (11), provided $K-D$

$$\tau < T.$$ 

(12)

(12) holds as an equality if the first visit to $D$ coincides with the first visit to $K$. In this case, the second term of (11) is zero. If, on the other hand, $\tau < T$,
the first visit to \( K \) precedes the first visit to \( D \), that is, the first visit to \( K \) is a visit to \( K-D \). Hence,

\[ \tau = \xi (1) < T, \]

and one finds under this condition

\[
E_i\left[ \#(m; Y_m = j, \tau < m < T) \right] \\
= E_i[ \sum_{m=\xi(1)}^{\xi(\nu+1)-1} 1_j(Y_m) 1_m < T ] \\
= E_i[ \sum_{\nu=1}^{\xi(\nu+1)-1} \sum_{n=\xi(\nu)}^{\xi(\nu)-1} 1_j(Y_m) 1_m < T 1_n(Y_m) ] \\
= E_i[ \sum_{n=d}^{k-1} \sum_{m=\xi(\nu)}^{\xi(\nu)-1} 1_j(Y_m) 1_m < T 1_n(Y_m) ] \]  

(13)

We now condition on \( Y_\xi(\nu) = u \). Because of the strong Markov property for the stopping time on \( \xi(\nu) \), one has

\[
E_i \left[ \sum_{m=\xi(\nu)}^{\xi(\nu+1)-1} 1_j(Y_m) 1_m < T 1_n(Y_\xi(\nu)) \mid Y_\xi(\nu) = u \right] \\
= E_i[ \sum_{m=0}^{\xi(1)-1} 1_j(Y_m) 1_m < T 1_n(Y_0) \mid Y_0 = u ] \\
= E_i[ \sum_{m=0}^{\tau-1} 1_j(Y_m) 1_n = u ] = 1_n = u v_{nj}(k) \]  

(14)

We now use (14) to write (13) as
\[ \begin{align*}
&j-1 = \ell(v+1)-1 \\
&\mathbb{E}_i[\sum_{n=d}^{j-1} \sum_{\nu=1}^{\ell(\nu)-1} \sum_{m=\ell(\nu)}^{l_\nu} (Y_{\ell(\nu)}) 1_{\ell(\nu)<T} v_{nj}(k) ] \\
\end{align*} \]

\[ \begin{align*}
&k-1 = \\
&\mathbb{E}_i[\sum_{n=d}^{k-1} \sum_{m=0}^{l_n(T)} 1_{n(Y_m)} v_{nj}(k)] = \\
&\sum_{n=d}^{k-1} \sum_{m=0}^{l_n(T)} v_{nj}(k) 1_{n(Y_m)} ] \\
&\sum_{n=d}^{k-1} \sum_{m=0}^{l_n(T)} v_{nj}(k) v_{in}(d). \\
\end{align*} \]

In summary

\[ \mathbb{E}_i[\{m: Y_m=j, \tau m<T\}] = \sum_{n=d}^{\infty} v_{in}(d) v_{nj}(k) \] (15)

If we compare (15) with (11), we obtain (3).

Corollary: For \( i < d \)

\[ v_{ij}(d) = v_{ij}(d+1) + v_{id}(d) v_{dj}(d+1). \] (16)

A similar result was obtained in Grassmann et al [1983].

**Applications**

The theorems above are particularly useful if the states can be partitioned into groups of states \( k_0, k_1, k_2, \ldots \), called levels. Level 0 consists of state 0 only, and level \( n, n=0,1,2,\ldots \) consists of the states \( k_{n-1} \) to \( k_{n-1} \). For consistency, \( k_0 \) is defined as 0. If \( d \) in equation (3) is replaced by \( k_d \), one has
We now define the following vectors and matrices:

\[ q_n = \begin{bmatrix} p_{j1} & \cdots & p_{jK_d} \end{bmatrix} \]

and

\[ U_{nd} = \begin{bmatrix} v_{11} & \cdots & v_{1K_d} \\
                       & \ddots & \vdots \\
                       &          & v_{K_d1} \end{bmatrix} \]

Using these symbols, (17) becomes

\[ q_d = \sum_{n=0}^{d-1} q_n U_{nd} \quad (19) \]

In many applications, one has the following: For all \( d \) within a certain range, say \( d > c \), level \( d \) cannot be reached from level \( n \), \( n < d - 1 \). In other words, level \( d - 1 \) cannot be skipped. Such levels will be called non-skippable. If \( d \) is non-skippable, \( U_{d-2} \), \( U_{d-3} \), \( \ldots \) are all zero, because all \( v_{ij}^{(kd)} \) vanish. In this case, (19) becomes

\[ q_d = q_{d-1} U_{d-1,d} \quad (20) \]

Suppose now that for \( n \) within the range extending from a certain lower limit \( e \) up to \( e^* \), all levels have the following properties:

1) Level \( d \) is non-skippable

2) \( k_{d+1} = k_d + a \), where \( a \) is a constant

3) \( [p_i j_{t+1}, i\in K_d] = [p_i j_{t+1}, i\in K_g], j\geq0, g\geq d. \)

It is easy to verify that under these conditions, \( U_{d-1,d} = U_{d-1,g} = U, g \geq d. \)

From (20), one concludes that under these conditions, \( q_d \) is matrix-geometric (Wallace [1969], Neuts [1981], Gaver et al. [1984]), that is

\[ q_{d+r} = q_d U^r. \quad (21) \]
Thus, using the $v_{ij}^{(d)}$ allows one to derive a number of results. Moreover, the $v_{ij}^{(d)}$ can be found in many cases of practical importance. Grassmann et al (1985) showed how to find the $v_{ij}^{(j)}$, using a modification of Gaussian elimination. Once the $v_{ij}^{(j)}$ are determined, one can obtain the $v_{ij}^{(d)}$, $d<j$, either using (9) or (16). Moreover, the $v_{ij}^{(d)}$ can be grouped into matrices according to equation (18), which allows one to set up interesting relations for quasi birth–death processes, in particular equations (20) and (21).

References

END

12-87

DTIC