The Use of Eigenvalues for Finding Equilibrium Probabilities of Certain Markovian Two-Dimensional Queueing Problems

Winfried K. Grassmann

Department of Computer Science, University of Saskatchewan, 57 Campus Drive, Saskatoon, Saskatchewan S7N 5A9, Canada

grassman@cs.usask.ca

A number of papers have appeared recently using eigenvalues for solving steady-state queueing problems. In this paper, we analyze Markovian systems with two state variables, the level X_1 and the phase X_2, $X_1 \geq 0$, $0 \leq X_2 \leq N$. Except for some boundary levels, the rates of the events are independent of the level, and no event can change X_1, X_2, or $X_1 + X_2$ by more than 1. In this case, the eigenvectors are essentially Sturm sequences, which implies that all eigenvalues are real. The properties of the Sturm sequences allow us to design an extension of the binary search to find all eigenvalues. As it turns out, once the interval containing an eigenvalue is narrowed down sufficiently, it is preferable to use Newton's method. A computational-complexity study indicates that the resulting algorithm should be significantly faster than matrix-iterative methods. Two numerical examples are discussed involving servers with breakdowns, and in both cases, our method yields highly accurate results. Tentative reasons why this is to be expected are provided.

(*Queueing; Eigenvalues; Unreliable Servers; Sturm Sequences*)

---

1 1. Introduction

A number of authors have recently used the solutions of generalized eigenvalue problems to solve queueing models (see, among others, Adan and Resing 1999, Bertsimas 1990, Grassmann and Drekic 2000, Haverkort and Ost 1997, Mitrani and Chakka 1995) and they have found that this method often outperforms matrix-analytic methods. Here, we provide a very efficient approach using eigenvalues for finding the equilibrium probabilities of certain queueing systems with two integer state variables, the *level* X_1 and the *phase* X_2. The level can assume any non-negative value, whereas the phase is restricted to be between 0 and N,

where N is a given finite number. The rates of all events depend only on the level and the phase, that is, the system is a two-dimensional Markov chain. We also require that no event can change X_1, X_2, or X_1+X_2 by more than 1. We will call this process a (BD)(BD) process. Moreover, in our model, the rates repeat in the sense that for almost all levels, the rates of the events do not depend on the present level. This process will be called a (BDR)(BD) process, where the R stands for "repeating." The problem is to find the equilibrium probabilities for (BDR)(BD) models. Of course, equilibrium probabilities exist if and only if the process is recurrent. We assume this to be the case.

Many (BDR)(BD) processes have been analyzed in the literature. Daigle and Lucantoni (1991), for instance, analyzed a queue in a randomly changing environment, with the level representing the queue, and the phase the environment. They assume that the phase cannot change by more than 1 at any time. A number of researchers have analyzed sequential queues with blocking (Grassmann and Drekic 2000; Konheim and Reiser 1976, 1978). In this case, the level is the length of the first queue, and the phase is the length of the second queue. There are numerous papers investigating systems with two customer types, in particular Green (1985) and Stanford and Grassmann (1993, 2000), and these systems also fit into the present framework. Another problem with two state variables is the unreliable-server problem (Mitrany and Avi-Itzhak 1968) where servers are subject to breakdowns and repair. There, the level is the number of customers in the system, and the phase is the number of operational servers. This model will be used as our prime example.

In (BDR)(BD) processes, the equilibrium equations can be divided into *boundary equations* and *repeating equations*. As the name implies, the repeating equations repeat in some sense, and it turns out that they have a solution of the form $\underline{d}x^n$, where the vector $\underline{d} \neq 0$ is called the *eigenvector*, and the scalar x is called the *eigenvalue*. The reason for these names is that $\underline{d}$ and x can be found by solving a so-called generalized eigenvalue problem, sometimes also called a matrix pencil. There are several eigensolutions $(\underline{d}^{(k)}, x_k)$, and to solve the boundary equations, one must combine these solutions.

In an earlier paper (Grassmann 2002), we analyzed models where X_i, $i = 1, 2$ cannot change by more than one, which is a generalization of the (BDR)(BD) process. One result of the paper is that if $X_1 + X_2$ cannot increase by more than 1, then all eigenvalues are real and non-negative. This result greatly generalizes the work of Daigle and Lucantoni (1991), Grassmann and Drekic (2000), and Konheim and Reiser (1976, 1978). These papers show that in their models, all eigenvalues are indeed real. Incidentally, Grassmann (2002)

also shows that the eigenvalues needed to determine transient probabilities of birth-death processes are real, a result found by more complicated methods independently by Karlin and McGregor (1965) and Ledermann and Reuter (1954). The simplifications and generalization described in Grassmann (2002) were possible by use of Sturm sequences for our theoretical investigations. In this paper, we also use Sturm sequences, but this time for numerical rather than theoretical purposes: we develop an extension of the binary search to find all eigenvalues with a predetermined precision. As soon as a particular eigenvalue can be shown to be the only eigenvalue within a small range, we switch to Newton's method.

There are authors who claim that (BDR)(BD) processes are best solved by using matrix-analytic methods. The problem with matrix-analytic methods is that they do not preserve the sparsity of the matrices involved, and this can make a huge difference. Because matrix-analytic methods generally involve matrix multiplications with matrices of dimension N, one has a time complexity of $O(N^3)$, no matter how fast the algorithm converges. By using eigenvalues, on the other hand, one can exploit the fact that the matrices involved are tridiagonal, which means that the computational effort is reduced to $O(N^2)$. We do not deny that eigenvalues also have their problems, but these seem to be manageable for the problems under investigation.

2 2. Solutions Involving Eigenvalues

(BD)(BD) processes possess block-structured transition matrices, with block $Q_{i,j}$ indicating the transitions from level i to level j. The entries inside the blocks indicate the changes of the phases. In the (BDR)(BD) process, there is by definition a value n_c such that for $j \geq n_c$, $Q_{i,j} = Q_{j-i}$. In other words, except for some boundary levels, all rows of blocks are identical if properly shifted. It turns out to be convenient to combine all levels $i < n_c$ into one new level, level -1. Levels n_c, $n_c + 1$, $n_c + 2$, ... are then renumbered 0, 1, 2, 3, ... If this is done, the following transition matrix results:

$$Q^* = \begin{bmatrix} Q_{-1,-1} & Q_{-1,0} & 0 & 0 & \dots \\ Q_{0,-1} & Q_0 & Q_1 & 0 & \ddots \\ 0 & Q_{-1} & Q_0 & Q_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{1}$$

Here, $Q_{-1,-1}$ is an $(N_{-1}+1) \times (N_{-1}+1)$ matrix, $Q_{-1,0}$ is an $(N_{-1}+1) \times (N+1)$ matrix, and $Q_{0,-1}$ is an $(N+1) \times (N_{-1}+1)$ matrix. All other matrices are square with dimension

$N+1$. The problem is to find the equilibrium vector $\underline{\pi}$, the vector satisfying $\underline{\pi}Q = 0$. If $\underline{\pi}_i$ is the vector of the equilibrium probabilities of level i, all equilibrium equations can readily be expressed in block form. Here, we only need the first block equation:

$$0 = \underline{\pi}_{-1}Q_{-1,-1} + \underline{\pi}_0 Q_{0,-1}. \tag{2}$$

The level numbered -1 here causes difficulties as indicated by Neuts (1981). We avoid these difficulties by embedding the Markov chain at levels 0, 1, ... According to Kemeni et al. (1966), the resulting transition matrix becomes

$$Q = \begin{bmatrix} Q_{0,0} & Q_1 & 0 & \dots & \\ Q_{-1} & Q_0 & Q_1 & 0 & \ddots \\ 0 & Q_{-1} & Q_0 & Q_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{3}$$

where

$$Q_{0,0} = Q_0 + Q_{0,-1}(-Q_{-1,-1})^{-1}Q_{-1,0}.$$

Except for norming, the matrices Q^* and Q have the same equilibrium probabilities for levels 0,1,2, ... For Markov chains with matrix Q, we obtain the following equilibrium equations:

$$0 = \underline{\pi}_0 Q_{0,0} + \underline{\pi}_1 Q_{-1} \tag{4}$$

$$0 = \underline{\pi}_{n-1}Q_1 + \underline{\pi}_n Q_0 + \underline{\pi}_{n+1}Q_{-1}, \ n \geq 1. \tag{5}$$

Equation (4) will be referred to as the *boundary equation*, and equation (5) as the *interior equation*. The interior equation always has solutions of the form (see e.g. Grassmann and Drekic 2000)

$$\hat{\underline{\pi}}_n = \underline{d}x^n, \ n \geq 0, \ |x| < 1, \ \underline{d} \neq 0, \tag{6}$$

where $\underline{d} = [d_0, d_1, \dots, d_N]$ is an *eigenvector* and x is an *eigenvalue*. Since $\underline{\pi}_n$ must converge to zero as n goes to infinity, only eigenvalues inside the unit circle are acceptable. It was shown by Gail et al. (2000) that if the process is recurrent, and all eigenvalues are distinct, there are $N+1$ distinct solutions of the form given in (6); we denote them by $(\underline{d}^{(k)}, x_k)$, $k = 0, 1, \dots, N$. Obviously, any linear combination of these solutions also forms a solution, and we tentatively set

$$\underline{\pi}_n = \sum_{k=0}^{N} c_k \underline{d}^{(k)} x_k^n$$

where c_k, $k = 0, 1, 2, \ldots, N$ are constants to be determined from the initial conditions and the fact that the sum of all probabilities is one. If $c = [c_0, c_1, \ldots, c_N]$, $\Lambda = \mathrm{diag}(x_i)$, and $D = [\underline{d}^{(0)}, \underline{d}^{(1)}, \ldots, \underline{d}^{(N)}]^T$, we get

$$\underline{\pi}_n = c\Lambda^n D, \ n \geq 0. \tag{7}$$

This equation holds only if all eigenvalues are distinct. However, except for the eigenvalues $x = 0$, this will always be true in our model. To find c, we replace $\underline{\pi}_n$, $n = 0, 1$ in the boundary equation (4) by (7) to obtain:

$$0 = cDQ_{0,0} + c\Lambda DQ_{-1}. \tag{8}$$

Suppose we know D and Λ, and we want to find c. Since equation (8) is homogeneous, it allows us to find c only up to a factor. Hence, we also need the fact that the sum of all probabilities is 1. We return to the original problem given by (1) and solve (2) for $\underline{\pi}_{-1}$ to obtain

$$\begin{aligned} 1 &= \underline{\pi}_{-1}e_{-1} + \sum_{n=0}^{\infty} \underline{\pi}_n e \\ &= cDQ_{0,-1}(-Q_{-1,-1})^{-1}e_{-1} + \sum_{n=0}^{\infty} c\Lambda^n De \\ &= c(DQ_{0,-1}(-Q_{-1,-1})^{-1}e_{-1} + \mathrm{diag}(1/(1-x_i))De). \end{aligned} \tag{9}$$

Here, e (e_{-1}) is a column vector of $N+1$ ($N_{-1}+1$) 1's. If x_0 is the largest eigenvalue inside the unit circle, then c_0 should be greater 0. We find a solution c^* of (8) with $c_0^* = 1$. The vector c is then $c^* c_0$, and we can replace c in (9) by $c^* c_0$ and solve for c_0.

In the interior equation, we replace $\underline{\pi}_n$ by $\underline{d}x^n$ which yields

$$0 = \underline{d}x^{n-1}Q_1 + \underline{d}x^n Q_0 + \underline{d}x^{n+1}Q_{-1} = \underline{d}Q(x),$$

where

$$Q(x) = Q_1 + Q_0 x + Q_{-1}x^2.$$

This equation must be solved for $\underline{d}$ and x.

3 3. Expressions for Eigenvalues and Eigenvectors

Expanding $0 = \underline{d}Q(x)$ yields, if $q_{i,j}(x) = Q(x)_{i,j}$:

$$0 = d_0 q_{0,0}(x) + d_1 q_{1,0}(x) \tag{10}$$

$$0 = d_{i-1}q_{i-1,i}(x) + d_i q_{i,i}(x) + d_{i+1}q_{i+1,i}(x), \quad i = 1, 2, \ldots, N-1 \tag{11}$$

$$0 = d_{N-1}q_{N-1,N}(x) + d_N q_{N,N}(x). \tag{12}$$

We convert (12) to (11) by letting $q_{N+1,N}(x)$ be an arbitrary polynomial with non-negative coefficients and add $d_{N+1}q_{N+1,N}(x)$ to (12). The problem then becomes to find an x such that $d_{N+1} = d_{N+1}(x) = 0$ to satisfy (12). We now have

Theorem 1 *If no $q_{i+1,i}(x)$ is zero, all non-trivial eigenvectors satisfy $d_0 \neq 0$.*

Proof: $d_0 = 0$ implies $d_1 = 0$ unless $q_{1,0}(x) = 0$. The remainder of the proof follows by complete induction based on (11).□

Hence, if $q_{i+1,i}(x) \neq 0$ for all i, we can set $d_0 = 1$. We now introduce functions $d_i(x)$ satisfying $d_i(x) = d_i$ whenever x satisfies $d_{N+1}(x) = 0$. To find the $d_i(x)$, we set $d_0(x) = d_0 = 1$, replace all d_i by $d_i(x)$ in the above system of equations and solve for $d_i(x)$, $i = 1, 2, \ldots, N+1$. This yields

$$d_1(x) = -q_{0,0}(x)/q_{1,0}(x) \tag{13}$$

$$d_{i+1}(x) = \frac{-1}{q_{i+1,i}(x)}(d_i(x)q_{i,i}(x) + d_{i-1}(x)q_{i-1,i}(x)), \quad i = 1, 2, \ldots, N. \tag{14}$$

To deal with the case that some $q_{i+1,i}(x)$ are zero, we use determinants. We define

$$G_i(x) = d_i(x) \prod_{j=0}^{i-1} (-q_{j+1,j}(x)), \quad i = 1, 2, \ldots, N+1. \tag{15}$$

As shown by Wilkinson (1965), $G_{N+1}(x) = \det Q(x)$.

If $q_{i+1,i}(x) = 0$ for some i, then $Q(x)$ can be partitioned as follows

$$Q(x) = \begin{bmatrix} Q_1(x) & M(x) \\ 0 & Q_2(x) \end{bmatrix}$$

where all entries of $M(x)$ are zero except for the lower right corner, which is equal to $q_{i,i+1}(x)$. It is known (Perlis 1952) that in this case, $\det Q(x) = \det Q_1(x) \det Q_2(x)$. The same method applies if $q_{i-1,i}(x) = 0$. By applying this method repeatedly, one can find matrices $Q_1(x)$, $Q_2(x)$, ..., $Q_K(x)$ such that none of the $Q_i(x)$ satisfy $q_{i+1,i}(x) = 0$ or $q_{i-1,i}(x) = 0$. Clearly, x is an eigenvalue if and only if there is least one i satisfying $\det Q_i(x) = 0$.

We are in particular interested in the case where the only $q_{i-1,i}(x)$ and $q_{i+1,i}(x)$ that vanish vanish identically. In this case, the problem can be partitioned into subproblems $d^{[i]}Q_i(x) = 0$, and each subproblem can be solved individually, using the methods to be developed here to solve $dQ(x) = 0$.

4 4. Location of the Eigenvalues

To find the (standard) eigenvalues in symmetric matrices, one typically uses Sturm sequences (see e.g. Wilkinson 1965). As it turns out, many of the techniques developed there can be applied here as well. Essentially, the sequence $\{d_i(x),\ i = 0, 1, \ldots N+1\}$ is a Sturm sequence within a given interval if for any fixed x within this interval, $d_0(x) = 1$ and $d_i(x) = 0$, $i = 1, 2, \ldots, N$ implies $d_{i-1}(x)d_{i+1}(x) < 0$.

There are no events that simultaneously increase X_1 and X_2, and $(Q_1)_{i,i+1}$ is therefore 0 for all i. This implies that $q_{i,i+1}(0) = 0$. For this case, Corollary 1 of Grassmann (2002) applies, indicating that if x is an eigenvalue inside the unit circle, $0 \leq x < 1$. Let us first deal with the case $x = 0$. Since $q_{i-1,i}(0) = 0$, $\det Q(0) = q_{0,0}(0)\ q_{1,1}(0) \cdots q_{N,N}(0)$, and it follows that $x = 0$ is an eigenvalue if $q_{i,i}(0)$ vanishes for any i. Obviously, $q_{i,i}(0) = 0$ iff $(Q_1)_{i,i} = 0$. It is therefore very easy to verify whether or not $\det Q(0) = 0$. Moreover, it is not difficult to see that if $(Q_1)_{i,i}$ is zero for r different values of i, then the multiplicity of the eigenvalue at zero is r. Since in recurrent Markov chains, there are exactly $N+1$ eigenvalues inside the unit circle, we only have $N+1-r$ positive eigenvalues that still have to be determined. How to deal with eigenvalues $x = 0$ will be developed in another paper. Here, we only mention that all these eigenvalues can be removed by replacing n_c by $n_c + r$ when forming level -1, and ignoring these eigenvalues otherwise. The reader may explore this by considering the appropriate Jordan chain (see Gohberg at al. 1982, Proposition 1.9). Hence, we assume $r = 0$ in the sequel, even though the extension to the case $r > 0$ is straightforward.

Consider now the case $0 < x < 1$. Since all off-diagonal elements of Q are non-negative, $q_{i-1,i}(x)$ and $q_{i+1,i}(x)$ must be non-negative. In fact, for any given $x > 0$, $q_{i-1,i}(x)$ and $q_{i+1,i}(x)$ must either vanish identically, or they must be strictly positive. The function $q_{i+1,i}(x)$ is identically zero if and only if $(Q_1)_{i+1,i}$, $(Q_0)_{i+1,i}$, and $(Q_{-1})_{i+1,i}$ are all zero, which means that there is no way to go from phase $i+1$ to phase i without going through some boundary states. This case does arise in priority queues (Drekic and Grassmann 2002). In cases like this, $Q(x)$ must be partitioned into submatrices $Q_1(x)$, $Q_2(x)$, ..., $Q_K(x)$, and the problems $d^{[i]}Q_i(x) = 0$ must be solved separately. We will not deal with this further, except for noting that most of our methods can easily be adapted to deal with this case. The reader may want to verify this.

For every given $x > 0$, and $q_{i\pm1,i}(x) \neq 0$, the sequence $\{d_i(x),\ i = 0, 1, \ldots, N+1\}$ is a Sturm sequence. First note that $d_0(x) = 1$ does not vanish. However, $d_1(x)$, could potentially

vanish, but for this x, $d_0(x)d_2(x) < 0$ because of (14). This implies that there is no positive value of x such that $d_1(x)$ and $d_2(x)$ are both zero. The remainder of the proof follows by complete induction, and our assertion that the $d_i(x)$ describe a Sturm sequence is proven. The proof also indicates that there is no $x > 0$, such that $d_i(x)$ and $d_{i+1}(x)$ both vanish.

Basic to Sturm sequences are sign variations (see e.g. Turnbull 1952). If $d_{i+1}(x) \neq 0$, then a sign variation occurs if $d_i(x)d_{i+1}(x) < 0$. A sign variation also occurs if $d_i(x) = 0$, but $d_{i+1}(x) \neq 0$, $i \neq N$. The number of sign variations will be denoted by $n(x)$, that is

$$n(x) = \#\{d_i(x)d_{i+1}(x) < 0,\ 0 \leq i < N\} + \#\{d_i(x) = 0,\ 0 \leq i < N\}.$$

According to Grassmann (2002), we have the following important result:

Theorem 2 *Suppose the (BDR)(BD) process under consideration is recurrent, there is no eigenvalue at 0, and all phases are communicating without going through a boundary state. If* $n(x-)-n(x) = 1$, *then* $d_{N+1}(x) = 0$. *Moreover, for any* y, $0 \leq n(y-)-n(y) \leq 1$. *Finally,* $n(1-) = 0$ *and* $n(0+) = N + 1$.

It follows that there are at least $|n(x_1) - n(x_2)|$ eigenvalues between x_1 and x_2, and at least $n(0+) - n(1-) = N + 1$ eigenvalues between $0+$ and $1-$. Since there are exactly $N + 1$ eigenvalues within the unit circle, all eigenvalues are accounted for. This result will now be exploited to develop a divide-and-conquer algorithm.

5 5. A Divide-and-Conquer Algorithm to Determine the Eigenvalues

This section describes an extension of the binary search algorithm to find all eigenvalues satisfying $0 < x < 1$. As the reader may know, in binary search, one recursively divides the present search interval into two parts, and discards any interval that contains no zero. Here, we also recursively divide the search interval into two parts, and we discard any interval not containing an eigenvalue. Hence, any interval $(x_1, x_2]$ is discarded if $n(x_1) = n(x_2)$. By applying this method recursively, one will eventually end up with $N + 1$ intervals, each one containing exactly one eigenvalue, provided, of course, that all eigenvalues are distinct. At this point, the procedure becomes a binary search. This leads to the following algorithm, called getx(x1,nx1,x2,nx2), where x1= x_1 is the start of the interval, x2= x_2 is the end, nx1 is set to $n(x_1)$ and nx2 is set to $n(x_2)$:

```
procedure getx(x1,nx1, x2, nx2)
if (nx1=nx2) return
x := (x1+x2)/2
if (x2−x1 ≤ ϵ)
    then if (nx1 =nx2+1) x[nx2] := x and return
    else report multiple eigenvalues and return
nx := n(x)
getx(x1,nx1, x, nx)
getx(x, nx, x2, nx2)
return
```

The square brackets are used for the indices of arrays, and ":=" is the assignment operator. To initiate the algorithm, we use getx(0,Np1,1,0), where Np1= $N+1$. The algorithm can be improved by switching to Newton's method as soon as an interval containing only one eigenvalue is found. A failsafe method similar to the one given by Press et al. (1986) can be used for this purpose.

We now study the computational complexity of the algorithm above under certain simplifying assumptions. Clearly, the values of x can only be obtained at a certain precision, which we take to be $\epsilon = 2^{-\alpha}$. This implies that if two eigenvalues are separated by less than $2^{-\alpha}$, they must be considered as multiple eigenvalues. We note, however, that eigenvalues are typically well spaced. We also assume that $N+1$ is a power of 2, say $N+1 = 2^m$ for some integer m. We say that we are at depth k when there are 2^k intervals, each of length 2^{-k}. Hence, at depth 0, there is one interval of length 1, at depth 1, there are two intervals of length $2^{-1} = 0.5$, and so on. Let k^* be the depth where, for the first time, all eigenvalues are in different intervals. $N+1$ eigenvalues cannot be separated unless there are $N+1$ intervals, and since there are $N+1 = 2^m$ eigenvalues, $k^* \geq m$. Since the eigenvalues are separated by at least $2^{-\alpha}$, $k^* \leq \alpha$. Observe that at lower depths, the intervals rejected are longer, which means that the case where $k^* = m$ is the worst case. We therefore concentrate on this case. To reach depth m requires $2^m - 1 = N$ function evaluations, and it yields 2^m intervals, each of length 2^{-m}. From this point on, we need for each eigenvalue $\alpha - m$ function evaluations to obtain the eigenvalue with accuracy $2^{-\alpha}$. Since $m = \log_2(N+1)$, this yields $(N+1)(\alpha - \log_2(N+1)) + \log_2(N+1)$ function evaluations in total. Each function evaluation has a complexity of $O(N)$ as the reader may verify. The overall complexity to find

all eigenvalues and eigenvectors is therefore $O(N^2)$. This compares with $O(N^3)$ for finding the matrix R in matrix-analytic methods. Incidentally, if there are no multiple eigenvalues, one has $R = D^{-1}\Lambda D$, that is, one can easily find R once all eigenvalues and eigenvectors are given, but the reverse is not true.

6 6. Derivatives

To find the derivatives needed by Newton's method, differentiate (10) and (11) with respect to x:

$$\begin{aligned} 0 &= d_0(x)q'_{0,0}(x) + d_1(x)q'_{1,0}(x) + d'_0(x)q_{0,0}(x) + d'_1(x)q_{1,0}(x) \\ 0 &= d_{i-1}(x)q'_{i-1,i}(x) + d_i(x)q'_{i,i}(x) + d_{i+1}(x)q'_{i+1,i}(x) \\ &+ d'_{i-1}(x)q_{i-1,i}(x) + d'_i(x)q_{i,i}(x) + d'_{i+1}(x)q_{i+1,i}(x), \quad i = 1, 2, \ldots, N. \end{aligned}$$

We solve these equations for $d'_{i+1}(x)$ to get, noting also that $d'_0(x) = 0$:

$$\begin{aligned} d'_1(x) &= \frac{-1}{q_{1,0}(x)}(d_0(x)q'_{0,0}(x) + d_1(x)q'_{1,0}(x)) \\ d'_{i+1}(x) &= \frac{-1}{q'_{i+1,i}(x)}(d_{i-1}(x)q'_{i-1,i}(x) + d_i(x)q'_{i,i}(x) + d_{i+1}(x)q_{i+1,i}(x)' \\ &\quad + d'_{i-1}(x)q_{i-1,i}(x) + d'_i(x)q_{i,i}(x)), \ i = 1, 2, \ldots, N. \end{aligned}$$

The derivatives of d_{N+1} can also be used to decide whether or not the process in question is recurrent. In fact, we have

Theorem 3 *If $x = 1$, $d_{N+1}(x) = 0$, one has*

$$\begin{aligned} d'_{N+1}(1) \le 0 &\quad \text{if the process is positively recurrent} \\ d'_{N+1}(1) = 0 &\quad \text{if the process is null recurrent} \\ d'_{N+1}(1) \ge 0 &\quad \text{if the process is non recurrent} \end{aligned}$$

To prove the theorem, note that $d_{N+1}(1)$ is zero because $Q(1) = Q_{-1} + Q_0 + Q_1$ has row-sums of 0, that is, $e = [1, 1, \ldots, 1]^T$ is an eigenvector corresponding to the eigenvalue 1. If the process is recurrent, then x_0 is, by definition, the point where the number of sign variations decreases from 1 to 0. If there are no sign variations, $d_0(x)$ and $d_{N+1}(x)$ must have the same

sign, and since $d_0(x) = 1$, they both must be positive. Hence, for $x_0 < x < 1$, $d_{N+1}(x) > 0$. It follows that

$$d'_{N+1}(1) = \lim_{h \to 0}(d_{N+1}(1) - d_{N+1}(1-h))/h \leq 0.$$

In Grassmann (2002), we introduced the *mirrored process*, which is obtained by interchanging Q_1 and Q_{-1}, and we showed that the eigenvalues of the mirrored process are the reciprocals of the original process. (For the treatment of the case where an eigenvalue is zero, see Grassmann (2002)). It can be shown that if the original process is positively recurrent, the mirrored process is non-recurrent, and vice versa. If the original process is null recurrent, so is the mirrored process. Suppose now that original process is non-recurrent, which implies that the mirrored process is positively recurrent, and the entity corresponding to $d'_{N+1}(x)$ in the mirrored process must therefore be negative. Taking reciprocals, however, makes this derivative positive. In the null-recurrent process, $x_0 = 1$, and since 1 is always an eigenvalue, there is a double zero at 1, that is, the derivative is 0.□

We note that in all recurrent processes we encountered, $d'_{N+1}(1) \neq 0$. We also were unable to construct a recurrent process with $d'_{N+1}(1) = 0$.

Generally, eigenvalues that are close together pose problems, but if x_0 is close to 1, and the system is recurrent, no problem will arise because any eigenvalue $x = 1$ does not form part of the matrix D and the corresponding eigensolution therefore has no detrimental effect on the precision of the results.

7 7. An Example

To demonstrate our methods, consider the unreliable-server problem (Mitrany and Avi-Itzhak 1968). In this model, there are in total N servers, and these servers can fail, which means they cannot serve customers until they are repaired. The model is affected by four events, namely arrivals, departures, breakdowns, and repairs. The arrival rate is λ, the service rate μ, the failure rate is δ, and the repair rate is β. All rates depend only on the present state, that is, the system is Markovian.

In this model, let X_1 be the number of customers in the system, and let X_2 be the number of operational servers. Since there are N servers, $0 \leq X_2 \leq N$. It is convenient to introduce the abbreviations $\text{diag}_j(z_m; m)$ for a matrix with entries z_m on row m and column $m+j$, $m = 0, 1, \ldots, N$, and zeros elsewhere. Hence, $\text{diag}_1(z_m; m)$ represents a superdiagonal matrix, $\text{diag}_{-1}(z_m; m)$ a subdiagonal matrix, and $\text{diag}_0(z_m; m) = \text{diag}(z_m; m)$ a diagonal

Table 1: The Eigenvectors for $N = 3$

1	4.86546	7.8909	4.265873
1	−0.46576	−3.20734	−1.78063
1	−4.98481	4.08844	5.68804
1	−9.86546	32.44243	−35.5622

matrix. If there is no ambiguity, we will write $\text{diag}_j(z_m)$. Of course, events that increase the phase by j will lead to matrices of the form $\text{diag}_j(\gamma_m)$, where γ_m is the rate of the event while $X_2 = m$. Since the phase cannot change by more than 1, j in $\text{diag}_j(\gamma_m)$ ranges from -1 to 1. We have

$$\begin{aligned} Q_{n,n+1} &= \text{diag}(\lambda) \\ Q_{n,n-1} &= \text{diag}(\mu \min\{m,n\}; m) \\ Q_{n,n} &= \text{diag}_{-1}(\delta \min\{m,n\}; m) + \text{diag}_1(\beta(N-m); m) \\ &\quad -\text{diag}(\lambda + (\mu+\delta)\min\{m,n\} + \beta(N-m); m). \end{aligned}$$

Clearly, $Q_{n,n+1}$ is independent of n for all n. Since for $n \geq m$, $\min\{m,n\} = m$, the matrices $Q_{n,n+1}$ and $Q_{n,n-1}$ are independent of n if $n \geq N \geq m$. Hence

$$\begin{aligned} Q_1 &= \text{diag}(\lambda) \\ Q_{-1} &= \text{diag}(\mu m) \\ Q_0 &= \text{diag}_{-1}(\delta m) + \text{diag}_1(\beta(N-m)) - \text{diag}(\lambda + (\mu+\delta)m + \beta(N-m)). \end{aligned}$$

We applied our method for two different problems, one with $N = 3$ servers, and one with $N = 10$ servers. The first problem is the one with 3 servers, and $\lambda = 10$, $\mu = 6$, $\beta = 2$, $\delta = 1$, and we found the following eigenvalues

$$x_0 = 0.898106,\ x_1 = 0.607321,\ x_2 = 0.476535,\ x_3 = 0.386613.$$

Note that the eigenvalues are well spaced, and they seem to lie on a parabola. They were calculated by the method described above, using VBA with double precision, and the result was accepted as soon as the absolute amount of $d_{N+1}(x)$ was less than $\epsilon = 1E{-}5$. The matrix D obtained is given in Table 1. The vector c^* is as follows:

$$1, -0.98591, 0.465741, -0.07414.$$

Table 2: The Eigenvalues and the c_i* for $N = 10$

i	x_i	c_i^*
0	0.844949	1
1	0.717121	−3.04156
2	0.64498	4.872541
3	0.588603	−4.63062
4	0.541146	2.901963
5	0.500000	−1.26452
6	0.46383	0.39129
7	0.431798	−0.08547
8	0.403299	0.012668
9	0.37785	−0.00115
10	0.355051	$4.91E{-}05$

Table 3: The Eigenvectors for $N = 10$

1	14.494	94.545	365.44	927.00	1612.4	1947.6	1613.1	876.8	28 2.44	40.940
1	8.1660	26.189	34.465	−16.371	−132.37	−224.33	−206.67	−112.72	−34.367	−4.5403
1	3.4869	−1.9503	−24.657	−34.593	13.640	89.961	112.46	70.275	22.843	3.0894
1	−0.9681	−10.242	−0.9117	40.185	41.977	−35.777	−100.05	−80.661	−29.904	−4.3606
1	−5.4379	0.0189	32.615	−1.5319	−91.805	−47.852	96.418	130.21	59.963	9.9168
1	−10.000	30.000	0.0000	−120.00	48.000	240.00	0.0000	−240.00	−160.00	−32.000
1	−14.678	81.462	−190.11	55.706	467.59	−391.48	−590.99	329.40	505.89	135.06
1	−19.476	156.21	−646.41	1354.6	−856.05	−1649.1	2405.9	863.48	−1623.4	−703.53
1	−24.386	255.89	−1493.6	5192.0	−10423	9569.9	2716.5	−11717	3 064.7	4330.1
1	−29.396	381.91	−2870.1	13682	−42493	84053	−95384	38353	3148 6	−30456
1	−34.494	535.45	−4925.4	29732	−123076	353792	−697373	902093	−691502	238533

To find c, these values would have to be divided by $1/c_0 = 273.92$, but since we want to compare the c_i, $i \neq 0$ with c_0, we will not do this here. Note that c_3^* is small compared to $c_0^* = 1$.

We also calculated the values x_i, $i = 0, 1, \ldots, 10$ and D for the case $\lambda = 30$, $\mu = 6$, $\beta = 2$, $\delta = 1$, $N = 10$. The results are given in Tables 2 and 3. To find c_i from c_i^*, one has to divide by 369632.4. The matrix D is given in Table 3.

One striking fact is that there is an eigenvalue that has exactly the value 0.5. Since the corresponding vector is

$$1, -10,\ 30,\ 0, -120,\ 48,\ 240,\ 0,\ -240, -167,\ 32$$

this is not likely to be a coincidence.

We did two accuracy checks. In the first check, we checked how well the repeating equation (5) was satisfied. If the right-hand sides are accurate, they are zero. Hence, their

calculated values can be interpreted as residuals. For the first problem, we obtained the following residuals for $n = 1, 2, 3$, and 4:

$n = 1$	0	$-5.2E{-}17$	0	$9.45E{-}09$
$n = 2$	$-3.2E{-}18$	$2.59E{-}17$	0	$5.36E{-}09$
$n = 3$	0	0	$-1E{-}16$	$3.09E{-}09$
$n = 4$	$-6.5E{-}18$	$-3.9E{-}17$	$7.78E{-}17$	$1.81E{-}09$

Note that the first three columns are extremely accurate, far more accurate than might be expected, considering that ϵ was set to $1E{-}5$. Even the last is fairly accurate. The results for the second problem are similar. The highest residual for $X_2 < N = 10$ is $5.9E{-}17$, and the highest residual in the last column is $2.24E{-}11$.

As a second accuracy check, we evaluated the rate matrix R using

$$R = D^{-1}\Lambda D.$$

It is known that R must satisfy

$$0 = Q_1 + RQ_0 + R^2Q_{-1}. \tag{16}$$

We calculated the right-hand side of this equation, using the numerical values we obtained for R. These residuals should be close to zero. In the first problems, the residual for $X_2 < N = 3$ are all less than $3.55E{-}15$. The last column is slightly worse, with errors of $-1.3E{-}06$, $-9.5E{-}08$, $1.17E{-}07$, and $1.97E{-}07$. For the second problem, the largest residual is $1E{-}9$ for the columns corresponding to $X_2 < N$, and $1.94E{-}7$ for the ones corresponding to $X_2 = N$.

There is a reason why the column corresponding to $X_2 = N$ has worse performance than the other ones. Really, if we use equations (13) and (14), then we should get zero for $X_2 < N$, no matter if the value chosen for x is correct or wrong. The value of x only influences the residual of the last equilibrium equation. Setting the precision to $\epsilon = 1E{-}05$ therefore affects only the column with $X_2 = N$.

Since each eigenvalue is associated with a specific number of sign changes, the eigenvectors can be expected to be rather dissimilar, a fact that is confirmed by the data given in Tables 1 and 3. These tables also indicate that the smaller the eigenvalues, the greater the variations among the entries of the eigenvectors. These variations could potentially cause numerical instability, but this is not what was observed here. A tentative explanation goes as follows. Since probabilities must be between 0 and 1, c_i must be small if an eigenvector contains large

elements unless it is possible to subtract a multiple of another eigenvector or eigenvectors such that the large elements cancel. However, since the eigenvectors are very different from one another, it is not likely that all large elements of the eigenvector can simultaneously be reduced to low values, which explains the surprisingly high accuracy observed in our results. This is yet another consequence of Sturm sequences.

The fact that the equilibrium equations hold with great accuracy does not imply that the eigenvectors are accurate. The reason is that the eigenvectors are calculated recursively, which means that the rounding errors accumulate. However, this does not affect the residuals. This begs the question as to what is more important, accurate eigenvectors or small residuals. It seems, however, that most algorithms aim to minimize residuals.

8 8. Conclusions

In this paper, we discussed how to use eigenvalues and eigenvectors to find equilibrium probabilities of Markovian problems with two integer state variables X_1 and X_2, $X_1 \geq 0$, $0 \leq X_2 \leq N$, with the restriction that X_1, X_2, and $X_1 + X_2$ cannot change by more than 1. It was shown that in this case, eigenvalue solutions are efficient, and that the results are accurate. There are a number of easy extensions to our model. First, we only used the fact that $X_1 + X_2$ cannot increase by more than 1. That means that the theory also applies if both X_1 and X_2 are allowed to decrease simultaneously. Actually, as indicated in Grassmann (2002), our main tool, the Sturm sequences, are even applicable in a much wider context. However, for the sake of simplicity, we have not discussed these additional models, but using Grassmann (2002) and the results of this paper, the reader should have no difficulty obtaining results for many additional problems. In fact, it is extremely difficult to find problems where $Q(x)$ is tridiagonal and has complex eigenvalues. When writing Grassmann (2002), we spent weeks searching for such examples, yet we could find only a single one, and this was constructed by elaborate means.

If we step beyond the confines of the (BDR)(BD) process, it seems that eigenvalue solutions are advantageous for cases with large, but sparse matrices Q_i, $i = -1, 0, 1$. The number of sign changes can certainly be used in these larger problems as well, even when the eigenvectors are no longer Sturm sequences. It is easy to show that in the interval $(x_1, x_2]$, there is at least one real eigenvalue if $n(x_1) - n(x_2)$ is odd. This allows one to adopt the algorithm discussed in Section 5 to find all real eigenvalues when $Q(x)$ is an (upper or lower)

Hessenberg matrix. In addition to that, we are presently investigating how to use $n(x)$ to locate complex eigenvalues.

The theory of eigenvalues can also make important contributions to matrix-analytic methods. One issue to be explored is the role of eigenvalues $x = 0$. If there are such eigenvalues, then the rows of R are no longer independent. This fact can potentially be used to make matrix-analytic solutions more efficient. This issue is presently being investigated. Hence, there are many open problems still waiting for solutions.

Acknowledgements

The author thanks the referees for their careful reading of the paper. The research was supported, in part, by grant 8112 of NSERC.

References

Adan I.J.B.F., J.A.C Resing. 1999. A class of Markov processes on a semi-infinite strip. B. Plateau, W.J. Stewart, M. Silva, eds. *3rd International Meeting on the Numerical Solution of Markov Chains*. 41–57.

Bertsimas, D. 1990. An analytic approach to a general class of G/G/s queueing systems. *Operations Research* **38** 139–155.

Daigle, J.N., D.M. Lucantoni. 1991. Queueing systems having phase-dependent arrival and service rates. W.J. Stewart, ed. *First International Conference on the Numerical Solution of Markov Chains*. Marcel Dekker, New York. 179–215.

Drekic, S., W.K. Grassmann. 2002. An eigenvalue approach to analyzing a finite source priority queueing model. *Annals of Operations Research*. To appear.

Gail, H.R., S.L. Hantler, B.A. Taylor. 2000. Use of characteristic roots for solving infinite state Markov chains. W.K. Grassmann, ed. *Computational Probability*, International Series in Operations Research and Management Science, chapter 7. Kluwer Academic Publishers, Boston, MA. 205–254.

Gohberg,I., P. Lancaster, L. Rodman. 1982. *Matrix Polynomials*. Academic Press, New York.

Grassmann, W.K. 2002. Real eigenvalues of certain tridiagonal matrix polynomials, with

queueing applications. *Journal of Linear Algebra and Its Applications* **342** 93–106.

Grassmann, W.K., S. Drekic. 2000. An analytical solution for a tandem queue with blocking. *Queueing Systems* **36** 221–235.

Green, L. 1985. A queueing system with general-use and limited-use servers. *Operations Research* **33** 168–182.

Haverkort, B.R., A. Ost. 1997. Steady-state analysis of infinite stochastic Petri nets: a comparison between the spectral expansion and the matrix-geometric method. *Proc. 7th International Workshop on Petri Nets and Performance Models*, Saint Malo, France. IEEE Computer Society Press. 36–45.

Karlin, S., J. McGregor. 1965. Ehrenfest urn models. *Journal of Appl. Prob.* **2** 352–376.

Kemeni, J.G., J.L. Snell, A.W. Knapp. 1966. *Denumerable Markov Chains.* Van Nostrand, Princeton, NJ.

Konheim, A.G., M. Reiser. 1976. A queueing model with finite waiting room and blocking. *Journal of the ACM* **23** 328–341.

Konheim, A.G., M. Reiser. 1978. Finite capacity queueing systems with applications in computer modeling. *SIAM J. Computing* **7** 210–229.

Ledermann, W., G.E. Reuter. 1954. Spectral theory for the differential equations of simple birth and death processes. *Phil. Trans. Roy. Soc. London, Ser. A* **246** 321–369.

Mitrani, I., R. Chakka. 1995. Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method. *Performance Evaluation* **23** 241–260.

Mitrany, I., B. Avi-Itzhak. 1968. A many-server queue with service interruption. *Operations Research* **16** 628–638.

Neuts, M.F. 1981. *Matrix-Geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, MD.

Perlis, S. 1952. *Theory of Matrices.* Addison-Wesley, Cambridge, MA.

Press, W.H., B.P. Flannery, S.A. Teukolsky, W.T. Vetterling. 1986. *Numerical Recipes.* Cambridge University Press, Cambridge, UK.

Stanford, D.A., W.K. Grassmann. 1993. The bilingual server system: a queueing model featuring fully and partially qualified servers. *INFOR* **31** 261–277.

Stanford, D.A., W.K. Grassmann. 2000. Bilingual server call centres. D.R. McDonald, S.R.E. Turner, eds. *Fields Institute Communications* **28**. American Mathematical Society, Providence, RI. 31–47.

Turnbull, H.W. 1952. *Theory of Equations*, 5th ed. Oliver and Boyd, Edingurgh, Scotland.

Wilkinson, J.S. 1965. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, UK.

Accepted by Edward P.C. Kao; received September 2001; revised June 2002; accepted September 2002.