COMPUTING MOMENTS OF FIRST PASSAGE TIMES
TO A SUBSET OF STATES IN MARKOV CHAINS

TUĞRUL DAYAR* AND NAIL AKAR†

Abstract. This paper presents a relatively efficient and accurate method to compute the moments of first passage times to a subset of states in finite ergodic Markov chains. With the proposed method, the moment computation problem is reduced to the solution of a linear system of equations with the right-hand side governed by a novel recurrence for computing the higher order moments. We propose to use a form of the Grassmann-Taksar-Heyman (GTH) algorithm to solve these linear equations. Due to the form of the linear systems involved, the proposed method does not suffer from the drawbacks associated with GTH in a row-wise sparse implementation.

Key words. Markov chains, unsafe states, first passage times, mean, variance, moments, GTH algorithm

AMS subject classifications. 60J10, 60J27, 65F05, 65F10

1. Introduction. Let the state space of a Markov Chain (MC) be partitioned into the subsets of *safe* and *unsafe* states. The question "How much time on the average will it take the system to reach an unsafe state if it starts in a safe state and what is its variance?" is to be answered when reliability is of concern in the system that is under investigation. Such reliability measures need to be known, for instance, in repairable computer systems to forecast the time to an unavailable system [22, pp. 328–330], or in railway systems [16, p. 12] where the passing of a train to a prohibited track on the red light is considered to be unsafe.

Determining the mean and variance of first passage times (or hitting times) in finite ergodic MCs [10, 11, 12] has centered around computing the stationary vector of the MC at hand and the fundamental matrix [14] (or the group generalized inverse [17]). In [23, p. 10], an iterative technique that does not require the computation of the stationary vector, the fundamental matrix, or the group generalized inverse, had been proposed. Recently, another technique appeared in [7] to compute these measures through Laplace transforms. Unfortunately, all of these techniques require dealing with systems in the order of the state space size. However, when the mean and variance of first passage times only to a subset of states in the MC are sought, it is unnecessary to consider systems as large as the state space size. In fact, it suffices to solve a linear system with as many equations as the number of safe states for two different right-hand sides.

In this paper, we study the more general problem of computing *all* moments of first passage times to a subset of states in finite ergodic MCs. The proposed method reduces the computation of the first k moments to the solution of k linear systems with a common coefficient matrix but different right-hand sides (one for each moment) where the right-hand sides are governed by a novel recurrence involving the universally well-known Pascal's triangle. We also use a form of the Grassmann-Taksar-Heyman (GTH) [6] algorithm for solving such linear systems. Due to the form of the linear systems involved, the proposed method does not suffer from the drawbacks associated with GTH in a row-wise sparse implementation.

The remainder of the paper is organized as follows. The second section provides an overview of the existing methods for the computation of the mean and variance of first passage times to a subset of states in MCs. The third section introduces the more efficient method which is based on the novel recurrence to compute all moments of first passage times to the subset of states. The fourth section shows how this method can be implemented relatively accurately using a form of the GTH algorithm and discusses a row-wise sparse implementation. The fifth section includes numerical experiments with MCs from the literature and the sixth section concludes the paper. The recurrence for computing the moments of first passage times in discrete-time MCs (DTMCs) is proved in the appendix.

In the following, P denotes the transition probability matrix of a finite ergodic MC of order n with steady state probability (row) vector π. Being the steady state vector, π is also the stationary vector and satisfies $\pi P = \pi$ with the normalization condition $\sum_{i=1}^{n} \pi_i = 1$. The proposed method in this paper

---

*Department of Computer Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey (tugrul@cs.bilkent.edu.tr).

†Department of Electrical and Electronics Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey (akar@ee.bilkent.edu.tr).

concerns DTMCs, but can be easily extended to continuous-time MCs (CTMCs) as we do in one of the test problems.

2. Mean and variance of first passage times to a subset of states. Without loss of generality, let us assume that P has the following block partitioning

$$(1) \qquad P = \begin{pmatrix} P_{\mathcal{S},\mathcal{S}} & P_{\mathcal{S},\mathcal{U}} \\ P_{\mathcal{U},\mathcal{S}} & P_{\mathcal{U},\mathcal{U}} \end{pmatrix},$$

where $\mathcal{S}$ and $\mathcal{U}$ are respectively the subsets of *safe* and *unsafe* states; let the number of states in these subsets be respectively $n_{\mathcal{S}}$ and $n_{\mathcal{U}}$. Otherwise, P can always be symmetrically permuted to the block form in (1). Here $P_{\mathcal{S},\mathcal{S}}$ is the square submatrix of order $n_{\mathcal{S}}$ obtained from P by deleting the rows and columns that correspond to the states in $\mathcal{U}$. Let the stationary vector be partitioned conformally as in $\pi = (\pi_{\mathcal{S}} \ \pi_{\mathcal{U}})$, where $\pi_{\mathcal{S}}$ and $\pi_{\mathcal{U}}$ are the subvectors of π associated with the states in $\mathcal{S}$ and $\mathcal{U}$, respectively.

The mean and variance of first passage times to subset $\mathcal{U}$ from states in subset $\mathcal{S}$ can be determined by first computing π using the GTH algorithm [6] and forming the aggregated matrix

$$K = \begin{pmatrix} P_{\mathcal{S},\mathcal{S}} & P_{\mathcal{S},\mathcal{U}}e \\ \frac{\pi_{\mathcal{U}}}{\|\pi_{\mathcal{U}}\|_1}P_{\mathcal{U},\mathcal{S}} & \frac{\pi_{\mathcal{U}}}{\|\pi_{\mathcal{U}}\|_1}P_{\mathcal{U},\mathcal{U}}e \end{pmatrix},$$

where e is a column vector of ones with appropriate length. Since $\pi_K = (\pi_{\mathcal{S}} \ \ \|\pi_{\mathcal{U}}\|_1)$ is the stationary vector of K, its fundamental matrix [14, p. 75] can be computed from

$$(2) \qquad Z = (I - K + e\pi_K)^{-1},$$

where I is the identity matrix of appropriate order. Now let us assume that $\mathrm{diag}(A)$ returns a column vector formed of the diagonal elements of matrix A, and $\mathrm{diag}(a)$ returns a diagonal matrix having the elements of vector a along its diagonal. Then the fundamental matrix enables us to compute the first and second moment matrices, namely $F^{(1)}$ and $F^{(2)}$, respectively, of first passage times among states in K from [14, p. 79]

$$F^{(1)} = (I - K + e\,\mathrm{diag}(Z)^T)D$$

and [14, p. 83]

$$F^{(2)} = F^{(1)}(2\,\mathrm{diag}(\mathrm{diag}(Z))D - I) + 2(ZF^{(1)} - e\,\mathrm{diag}(ZF^{(1)})^T),$$

where D is the diagonal matrix having reciprocals of the elements of π_K along its diagonal. The (i,j)th element of $F^{(1)}$, denoted by $f_{i,j}^{(1)}$, is the mean first passage time from state i to state j in K. Similarly, the (i,j)th element of $F^{(2)}$, denoted by $f_{i,j}^{(2)}$, is the second moment of the first passage time from state i to state j in K.

In terms of the entries of $F^{(1)}$ and $F^{(2)}$, the mean and variance of first passage times from state $i \in \mathcal{S}$ to $\mathcal{U}$ are respectively given by

$$(3) \qquad m_i^{(1)} = f_{i,n_{\mathcal{S}}+1}^{(1)}$$

and

$$(4) \qquad v_i = f_{i,n_{\mathcal{S}}+1}^{(2)} - (f_{i,n_{\mathcal{S}}+1}^{(1)})^2,$$

where $f_{i,n_{\mathcal{S}}+1}^{(2)}$ is the second moment of first passage time from state $i \in \mathcal{S}$ to $\mathcal{U}$ and will be denoted by $m_i^{(2)}$. We remark that the time complexity of obtaining the two measures in (3) and (4) with the described approach is $O(n^3)$ due to the necessity to compute π.

It is well known that the computation of the fundamental matrix through matrix inversion as in (2) is prone to numerical inaccuracies. This affects the accuracy of the first and second moment matrices of first passage times that are computed using the fundamental matrix. Through a series of papers [10, 11, 12], Heyman and O'Leary have identified this situation and have provided algorithms that aim at remedying

it. Their idea is based on using the GTH algorithm [6] to factorize a singular linear system involving the MC at hand and to compute the stationary vector accurately.

Since we are interested in computing the moments of first passage times to a subset of states, in the next section we present a method that does not require the stationary vector of the MC to be computed and has a time complexity that depends on $n_{\mathcal{S}}$, as opposed to n with the existing techniques. After this presentation, we show how a form of the GTH algorithm can be used to provide relatively accurate results in solving the nonsingular linear systems that arise in this context.

3. A more efficient method. Let X be the random variable associated with the first passage times to subset $\mathcal{U}$ conditioned on an initial probability vector τ in the subset of safe states, $\mathcal{S}$. Given $T = P_{\mathcal{S},\mathcal{S}}$, we have [15, p. 50]

$$E[X(X - 1) \cdots (X - k + 1)] = \tau \gamma^{(k)}, \tag{5}$$

where

$$\gamma^{(k)} = k!(I - T)^{-k} T^{k-1} e \tag{6}$$

is defined as the kth factorial moment vector of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$ for $k > 0$. This result follows from probabilistic arguments and is based on aggregating the states in $\mathcal{U}$ to a single state and constructing an absorbing MC in which all states in $\mathcal{S}$ are transient and the aggregated state is the single absorbing state (see [15, pp. 47–50]).

When $k = 1$, (5) becomes

$$E[X] = \tau(I - T)^{-1} e. \tag{7}$$

Then the mean first passage time from state $i \in \mathcal{S}$ to $\mathcal{U}$, denoted by $m_i^{(1)}$, follows from (7) by substituting e_i^T (i.e., ith row of I) for τ. More generally, the first moment vector $m^{(1)}$, whose ith element is $m_i^{(1)}$ can be obtained from

$$(I - T)m^{(1)} = e. \tag{8}$$

We remark that $(I - T)$ is a nonsingular M-matrix[1] and $(I - T)^{-1} \geq 0$ [2]. Hence, $m^{(1)} = (I - T)^{-1} e$, simply the row sums of the nonnegative matrix $(I - T)^{-1}$. This nonnegative matrix is referred to as the fundamental matrix of an absorbing MC in which the states in $\mathcal{S}$ are transient and those in $\mathcal{U}$ are ergodic [14, p. 46]. In fact, (8) is available in [14, p. 51] and has been derived following a different approach.

When $k = 2$, (5) and (6) yield

$$E[X^2] = E[X] + 2\tau(I - T)^{-2} T e. \tag{9}$$

Defining $m^{(2)}$ to be the second moment vector of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$, we obtain

$$(I - T)m^{(2)} = 2m^{(1)} - e \tag{10}$$

from (9) after some algebra and again considering all rows of I for τ. Although not exactly in the form of (10), an equivalent representation for $m^{(2)}$ is available in [14, p. 51].

Once $m^{(2)}$ is available, the variance of the first passage time from state i in $\mathcal{S}$ to $\mathcal{U}$ can be computed from

$$v_i = m_i^{(2)} - (m_i^{(1)})^2 \quad \text{for } i \in \mathcal{S}. \tag{11}$$

Equations (8) and (10) provide an alternative way (to that in section 2) of computing the first and second moment vectors of first passage times to a subset of states in finite ergodic MCs. Observe that neither a complete factorization based on P or the stationary vector of P is required to compute these measures. In conclusion, one needs to factorize $(I - P_{\mathcal{S},\mathcal{S}})$ once, compute $m^{(1)}$, and then compute $m^{(2)}$ using $m^{(1)}$, e, and the factors of $(I - P_{\mathcal{S},\mathcal{S}})$.

---

[1] Any matrix A of the form $A = sI - B$ with $s > 0$ and $B \geq 0$ for which $s \geq \rho(B)$, the spectral radius of B.

In fact, it is possible to compute (higher) moments of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$ using the recurrence

$$(12) \qquad Sm^{(i+1)} = \sum_{j=0}^{i} (-1)^{i-j} \binom{i+1}{j} m^{(j)} \quad \text{for} \quad i \geq 0 \quad \text{with} \quad m^{(0)} = e,$$

where $S = I - P_{\mathcal{S},\mathcal{S}}$. Here $\binom{i+1}{j}$ denotes the binomial coefficient in Pascal's triangle that is read "$(i+1)$ choose j" [5, p. 153]. To the best of our knowledge, the recurrence in (12) has not been given before and is proved in the appendix.

In the next section we show how the more efficient method which essentially solves linear systems of the form $Sy = c$ in (12), where $S = I - P_{\mathcal{S},\mathcal{S}}$, can be implemented relatively accurately using a form of the GTH algorithm and discuss a row-wise sparse implementation.

4. Implementation issues. In order to compute (higher) moment vectors of first passage times, we use Algorithm 1 which follows from (12) that has the matrix $S = I - P_{\mathcal{S},\mathcal{S}}$ on the left-hand side and a linear combination of e and the previous moment vectors on the right-hand side. We LU factorize S once at the outset using GTH as discussed in the next subsection. Here, L and U are respectively the unit-lower triangular factor and the upper-triangular factor of S. The very reason we use GTH is to improve the accuracy in the computed factors when S is close to being singular. Otherwise, S could very well be factorized using Gaussian elimination (GE), but in no way should be inverted and passed to the right-hand side in (12). Finally, the integer coefficients of the linear combination on the right-hand side of (12) may be computed easily by observing the alternating signs and using the well known identity [5, p. 158]

$$\binom{i+1}{j} = \binom{i}{j-1} + \binom{i}{j}$$

that binomial coefficients satisfy.

<div align="center">

ALGORITHM 1

Computing the kth moment vector of first passage times to a subset of states in DTMCs.

</div>

LU factorize S using GTH;
$a_0 = 1$; $a_1 = -1$; $m^{(0)} = e$;
Solve for $m^{(1)}$ in $LUm^{(1)} = m^{(0)}$;
For $i = 1$ up to $k - 1$ {
    $a_{i+1} = -1$;
    For $j = i$ down to 1
        $a_j = a_{j-1} - a_j$;
    $a_0 = -a_0$;
    Solve for $m^{(i+1)}$ in $LUm^{(i+1)} = \sum_{j=0}^{i} a_i m^{(i)}$;
}

4.1. Factorization using GTH. The GTH algorithm due to Grassmann-Taksar-Heyman [6] emerges from probabilistic arguments and shows how one can compute the stationary vector of a MC using only nonnegative numbers and avoiding subtraction operations. This algorithm achieves significantly greater accuracy than other algorithms in the literature [8, 20]. The entrywise relative error in the stationary vector computed using the GTH algorithm is shown to be small, independent of the structure of the matrix and the magnitudes of its elements [19].

A form of the GTH algorithm is used in the disaggregation phase of the iterative aggregation-disaggregation (IAD) algorithm geared towards nearly completely decomposable (NCD) MCs [4] (we will return to NCD MCs in the section on numerical results). Within that phase, linear systems of the form $y^T S = c^T$ (or the transposed form $S^T y = c$) need to be solved for y. To each such system one more unknown and one more equation are added so as to have a homogeneous linear system with a coefficient matrix that is a singular M-matrix. Here we follow a slightly different approach and construct a non-homogeneous, but consistent, linear system with a singular coefficient matrix.

Consider $Ax = b$, where

$$(13) \qquad A = \begin{pmatrix} I - P_{\mathcal{S},\mathcal{S}} & -P_{\mathcal{S},\mathcal{U}}e \\ w^T & -w^T e \end{pmatrix}, \; x = \begin{pmatrix} y \\ 0 \end{pmatrix}, \text{ and } b = \begin{pmatrix} c \\ v \end{pmatrix}.$$

Observe that A has zero row sums since $(I - P_{\mathcal{S},\mathcal{S}})e = e - P_{\mathcal{S},\mathcal{S}}e = P_{\mathcal{S},\mathcal{U}}e$, and A is a singular M-matrix with rank $n_{\mathcal{S}}$. The values of the vector w and the scalar v are irrelevant since in the LU factorization of the singular matrix A, the last row of the upper-triangular factor turns out to be zero.

The properties of a singular M-matrix coupled with the idea of using nonnegative arithmetic in the GTH algorithm suggests the following modification to Gaussian Elimination (GE) when upper-triangularizing A. At each step of GE, rather than computing the pivot element in the usual way, one can correct the pivot by replacing it with the negated sum of the off-diagonal elements in the unreduced part of the same row as the pivot. This follows from the fact that at each step of GE, the unreduced square submatrix that is to be updated turns out to be a singular M-matrix having zero row sums. We remark that other than providing better roundoff properties, this modification has the additional advantage of being able to be carried out completely in row-wise sparse format with delayed row updates [23]. In particular, the computation of pivots is in no way orthogonal to the access scheme of row-wise sparse format as opposed to the situation with the systems solved in IAD (see section 4 in [4]) or the original formulation of the GTH algorithm [6]. It is this modified GE which we refer to as the GTH implementation in this paper.

The GTH reduction needs to be performed for only $(n_{\mathcal{S}} - 1)$ steps. This amounts to premultiplying A by a unit lower-triangular multiplier matrix as in

$$\begin{pmatrix} L^{-1} & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} I - P_{\mathcal{S},\mathcal{S}} & -P_{\mathcal{S},\mathcal{U}}e \\ w^T & -w^T e \end{pmatrix} = \begin{pmatrix} U & -L^{-1}P_{\mathcal{S},\mathcal{U}}e \\ w^T & -w^T e \end{pmatrix}$$

so that $(I - P_{\mathcal{S},\mathcal{S}}) = LU$, where L is unit lower-triangular and U is upper-triangular. The product $-L^{-1}P_{\mathcal{S},\mathcal{U}}e$ needs to be computed along the way to enable the correction of the pivots. Then y can be obtained from $LUy = c$ through forward and backward substitutions (cf. (13)). Note that both L and U need to be stored since we will solve for (at least) two right-hand sides.

4.2. Analysis. In this subsection we discuss issues pertaining to the row-wise sparse factorization of $S = I - P_{\mathcal{S},\mathcal{S}}$ using GTH and the sparse implementation of Algorithm 1. Let nz_S, nz_L, and nz_U denote respectively the number of nonzeros in S, the strictly lower-triangular part of L (since 1's along its diagonal need not be stored), and U. Also let k denote the order of the highest moment vector to be computed.

The nonzeros in nz_L and nz_U depend on the sparsity pattern of S, and there are myriad ways discussed in the literature to reduce the fill-in that results from its sparse factorization. However, a discussion on fill reducing orderings and their implications is beyond the scope of this paper. Therefore, assuming that $\mathcal{S}$ is not reordered and a row-wise sparse format is used, one needs $(nz_L + nz_U)$ real and $(nz_L + nz_U + n_{\mathcal{S}} + 1)$ integer storage space to store and index the nonzeros in the L and U factors of S [23, pp. 80–81]. Since the sparse factorization can be performed through delayed row updates (see subsection 4.1), it suffices to have an additional real row vector of length n_s to expand the row to be updated, do the reduction, compact the reduced row, and store it. When GTH is used to carry out the factorization, a real vector of length n_s which sums the rows of S to zero (i.e., last column of A in (13)) should also be allocated. This real vector of length n_s is the only extra storage space compared to using GE.

Algorithm 1 requires k real vectors of length n_s to store the moment vectors and one real vector of length n_s to store the right-hand side in (12). This right-hand side is a linear combination with well-defined integer coefficients of the previous moment vectors, and does not pose problems from a computational point of view other than overflow as long as the integer coefficients are stored in a real vector of length $k + 1$ (see the coefficients a_i). Hence, the total storage space required by Algorithm 1 in addition to the sparse factors of S is $(k + 1)$ real vectors of length n_s and one real vector of length $(k + 1)$. Note that since the absolute values of the coefficients in (12) follow from Pascal's triangle and that this triangle is symmetric, it is possible to do by storing only half of the coefficients.

The number of floating-point arithmetic operations (flops) to perform sparse GTH on S is equal to the number of flops to perform sparse GE on S, minus $2(n_{\mathcal{S}} - 1)$ flops for those that do not get executed in GTH due to the new way of computing the pivots (note that the pivot in the first row is not updated) plus nz_U flops due to computing each pivot by summing the nonzeros in the strictly upper triangular part of U

and the element of the updated extra column in the same row as the pivot. Hence, the only disadvantage of GTH with respect to GE in this setting is the extra $(nz_U - 2n_\mathcal{S} + 2)$ flops. Assuming that nz_U is linear in $n_\mathcal{S}$ for the sparse matrix S, this extra cost in GTH is negligible since the number of flops to perform sparse GE on S will be quadratic in $n_\mathcal{S}$. In passing to the analysis of Algorithm 1, we remark that each pair of forward and backward substitutions (respectively with the matrices L and U) for a different right-hand side requires $2(nz_L + nz_U) - n_\mathcal{S}$ flops.

In the ith pass, the outer for-loop in Algorithm 1 requires the computation of $(i+1)$ coefficients (which takes i floating-point subtractions on integers), the computation of a right-hand side vector (which takes $2(i+1)n_s$ flops), and a pair of forward and backward substitutions (which take $2(nz_L + nz_U) - n_\mathcal{S}$ flops). For a total of $(k-1)$ passes, these amount to $(k-1)[2(nz_L + nz_U) + n_s(k+1) + k/2]$ flops.

In the following section we provide results of numerical experiments with two NCD MCs and a reliability problem from the literature.

5. Numerical results. In this section, we present results with three problems to illustrate the proposed method. We have implemented the proposed method in C using IEEE floating-point arithmetic with and without GTH [21]. We refer to the latter approach as the proposed method with GE. Using these methods, we have computed the moment vectors of first passage times to a subset of states in each MC. In all problems, the results of the proposed method with GTH and GE in double precision floating-point arithmetic (i.e., about sixteen decimal digits of precision) agree to more than seven decimal digits. Hence, we have taken the first seven decimal digits of these results as exact results, and have compared the results of the proposed method with GTH and GE in single precision floating-point arithmetic (i.e., about seven decimal digits of precision) against these exact results. Furthermore, the condition number of each coefficient matrix is reported so as to indicate the inherent difficulty associated with computing the moment vectors accurately. Note that 10^{38} is about the size of the largest positive number in single precision floating-point arithmetic, and any attempt to compute moment vectors having elements larger than this value will be futile. Assuming that k is the highest order moment vector that can be computed in single precision floating-point arithmetic for a given problem, in order to save space we report only the first, second, and kth moment vectors. Since k in each of the problems turns out to be less than or equal to twenty-two, the computation of binomial coefficients in single precision arithmetic does not pose problems. In other words, the largest entry in the twenty-second row of Pascal's triangle, 22 choose 11, is equal to 705,432, and can be represented in seven decimal digits of precision.

The first two problems are also considered in [11] and are examples of nearly completely decomposable (NCD) MCs [3, 18], which are irreducible stochastic matrices that can be symmetrically permuted to the block form

$$(14) \qquad P_{n \times n} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NN} \end{pmatrix} \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{matrix}$$

where nonzero elements in the off-diagonal blocks are small compared with those in the diagonal blocks [23, p. 286]. Let $P = \text{diag}(P_{11}, P_{22}, \ldots, P_{NN}) + E$. The diagonal blocks P_{ii} are square, of order n_i, with $n = \sum_{i=1}^{N} n_i$. The quantity $\|E\|_\infty$ is referred to as the degree of coupling and is taken to be a measure of the decomposability of P. Such matrices are known to be numerically challenging. The third problem is the reliability model that appears in [22].

5.1. Test Problem 1. The first problem is the 8×8 matrix due to Courtois [3] with all row sums equal to 1 and given by

$$P = \begin{pmatrix} 0.85 & 0 & 0.149 & 0.0009 & 0 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.65 & 0.249 & 0 & 0.0009 & 0.00005 & 0 & 0.00005 \\ 0.1 & 0.8 & 0.0996 & 0.0003 & 0 & 0 & 0.0001 & 0 \\ 0 & 0.0004 & 0 & 0.7 & 0.2995 & 0 & 0.0001 & 0 \\ 0.0005 & 0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0 & 0 \\ 0 & 0.00005 & 0 & 0 & 0.00005 & 0.6 & 0.2499 & 0.15 \\ 0.00003 & 0 & 0.00003 & 0.00004 & 0 & 0.1 & 0.8 & 0.0999 \\ 0 & 0.00005 & 0 & 0 & 0.00005 & 0.1999 & 0.25 & 0.55 \end{pmatrix}.$$

This matrix has the three subsets of NCD states, $\{1,2,3\}$, $\{4,5\}$, $\{6,7,8\}$, for a degree of coupling of 0.001.

Let us take $\mathcal{S} = \{1,2,3\}$, which corresponds to an NCD subset of states. For this choice of safe states, the 2-norm condition number of the coefficient matrix is $cond_2(I - P_{\mathcal{S},\mathcal{S}}) = 1.5 \times 10^3$ from Matlab. For this problem the first ten moment vectors can be computed in single precision floating-point arithmetic. The computed results are

$$
\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \end{pmatrix}_{exact} = \begin{pmatrix} 1.122227 \\ 1.122461 \\ 1.123047 \end{pmatrix} \times 10^3, \qquad
\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.122239 \\ 1.122473 \\ 1.123059 \end{pmatrix} \times 10^3,
$$

$$
\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \end{pmatrix}_{GTH} = \begin{pmatrix} 1.122227 \\ 1.122461 \\ 1.123047 \end{pmatrix} \times 10^3;
$$

$$
\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \end{pmatrix}_{exact} = \begin{pmatrix} 2.518215 \\ 2.518743 \\ 2.520059 \end{pmatrix} \times 10^6, \qquad
\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \end{pmatrix}_{GE} = \begin{pmatrix} 2.518271 \\ 2.518799 \\ 2.520115 \end{pmatrix} \times 10^6,
$$

$$
\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \end{pmatrix}_{GTH} = \begin{pmatrix} 2.518216 \\ 2.518744 \\ 2.520060 \end{pmatrix} \times 10^6;
$$

$$
\begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \end{pmatrix}_{exact} = \begin{pmatrix} 1.147324 \\ 1.147565 \\ 1.148164 \end{pmatrix} \times 10^{37}, \qquad
\begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.147451 \\ 1.147692 \\ 1.148292 \end{pmatrix} \times 10^{37},
$$

$$
\begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \end{pmatrix}_{GTH} = \begin{pmatrix} 1.147325 \\ 1.147566 \\ 1.148166 \end{pmatrix} \times 10^{37}.
$$

We remark that the first moment vector of first passage times from the chosen NCD subset to the remaining states have elements that are close to each other and are in the order 10^3. See [1, 9] for detailed information regarding this phenomenon. The relative errors in $(m_{GE}^{(1)}, m_{GE}^{(2)}, m_{GE}^{(10)})$ and $(m_{GTH}^{(1)}, m_{GTH}^{(2)}, m_{GTH}^{(10)})$ for $\mathcal{S} = \{1,2,3\}$ are respectively $(1.1 \times 10^{-5}, 2.2 \times 10^{-5}, 1.1 \times 10^{-4})$ and $(0, 4.0 \times 10^{-7}, 1.7 \times 10^{-6})$ in the infinity norm.

Now, let us consider $\mathcal{S} = \{1,4,6\}$ in which each state comes from a different NCD subset. For this choice of safe states, we have $cond_2(I - P_{\mathcal{S},\mathcal{S}}) = 2.7 \times 10^0$ from Matlab. For this problem the first twenty-two moment vectors can be computed in single precision floating-point arithmetic. The computed results are

$$
\begin{pmatrix} m_1^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \end{pmatrix}_{exact} = \begin{pmatrix} 6.687500 \\ 3.333333 \\ 2.500000 \end{pmatrix} \times 10^0, \qquad
\begin{pmatrix} m_1^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \end{pmatrix}_{GE} = \begin{pmatrix} 6.687501 \\ 3.333333 \\ 2.500000 \end{pmatrix} \times 10^0,
$$

$$
\begin{pmatrix} m_1^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \end{pmatrix}_{GTH} = \begin{pmatrix} 6.687501 \\ 3.333333 \\ 2.500000 \end{pmatrix} \times 10^0;
$$

$$
\begin{pmatrix} m_1^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \end{pmatrix}_{exact} = \begin{pmatrix} 8.261667 \\ 1.888889 \\ 1.000000 \end{pmatrix} \times 10^1, \qquad
\begin{pmatrix} m_1^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \end{pmatrix}_{GE} = \begin{pmatrix} 8.261670 \\ 1.888889 \\ 1.000000 \end{pmatrix} \times 10^1,
$$

$$
\begin{pmatrix} m_1^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \end{pmatrix}_{GTH} = \begin{pmatrix} 8.261668 \\ 1.888889 \\ 1.000000 \end{pmatrix} \times 10^1;
$$

$$
\begin{pmatrix} m_1^{(22)} \\ m_4^{(22)} \\ m_6^{(22)} \end{pmatrix}_{exact} = \begin{pmatrix} 2.814053 \times 10^{38} \\ 9.561367 \times 10^{30} \\ 3.840642 \times 10^{27} \end{pmatrix}, \qquad
\begin{pmatrix} m_1^{(22)} \\ m_4^{(22)} \\ m_6^{(22)} \end{pmatrix}_{GE} = \begin{pmatrix} 2.814063 \times 10^{38} \\ 9.561357 \times 10^{30} \\ 3.840647 \times 10^{27} \end{pmatrix},
$$

$$\begin{pmatrix} m_1^{(22)} \\ m_4^{(22)} \\ m_6^{(22)} \end{pmatrix}_{GTH} = \begin{pmatrix} 2.814057 \times 10^{38} \\ 9.561385 \times 10^{30} \\ 3.840641 \times 10^{27} \end{pmatrix}.$$

In this case, the relative errors in $(m_{GE}^{(1)}, m_{GE}^{(2)}, m_{GE}^{(22)})$ and $(m_{GTH}^{(1)}, m_{GTH}^{(2)}, m_{GTH}^{(22)})$ are respectively $(1.5 \times 10^{-7}, 3.6 \times 10^{-7}, 3.6 \times 10^{-6})$ and $(1.5 \times 10^{-7}, 1.2 \times 10^{-7}, 1.4 \times 10^{-6})$ in the infinity norm. It is not surprising to encounter higher relative errors when computing the larger means and variances of first passage times corresponding to the NCD subset $\mathcal{S} = \{1, 2, 3\}$.

5.2. Test Problem 2. The second problem we consider is inspired by the class of 10×10 matrices used in [20]. Let

$$S = \begin{pmatrix} 0.1 & 0.3 & 0.1 & 0.2 & 0.3 & \beta & 0 & 0 & 0 & 0 \\ 0.2 & 0.1 & 0.1 & 0.2 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.2 & 0.4 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.2 & 0.1 & 0.2 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0.3 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 & 0 & 0.1 & 0.2 & 0.2 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.2 & 0.1 & 0.3 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.3 & 0.2 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.2 & 0.1 & 0.3 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.7 & 0 & 0 & 0.2 \end{pmatrix}.$$

If $D = \text{diag}([1/(1+\beta), 1, 1, 1, 1, 1/(1+\beta), 1, 1, 1, 1]^T)$, then $P = DS$ is an NCD MC of order 10 with degree of coupling $\beta/(1 + \beta)$. Note that P has two subsets of NCD states, $\{1, 2, 3, 4, 5\}$ and $\{6, 7, 8, 9, 10\}$, that communicate with each other over a single state in each direction.

We let $\mathcal{S} = \{1, 2, 3, 4, 5\}$. For this choice of safe states and $\beta = 10^{-7}$, we have $cond_2(I - P_{\mathcal{S},\mathcal{S}}) = 5.3 \times 10^7$ from Matlab. For this problem only the first four moment vectors can be computed in single precision floating-point arithmetic. The computed results are

$$\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_5^{(1)} \end{pmatrix}_{exact} = \begin{pmatrix} 3.478633 \\ 3.478633 \\ 3.478633 \\ 3.478633 \\ 3.478633 \end{pmatrix} \times 10^7, \qquad \begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_5^{(1)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.484419 \\ 1.484419 \\ 1.484419 \\ 1.484419 \\ 1.484419 \end{pmatrix} \times 10^8,$$

$$\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_5^{(1)} \end{pmatrix}_{GTH} = \begin{pmatrix} 3.478632 \\ 3.478633 \\ 3.478633 \\ 3.478633 \\ 3.478633 \end{pmatrix} \times 10^7;$$

$$\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_5^{(2)} \end{pmatrix}_{exact} = \begin{pmatrix} 2.420177 \\ 2.420177 \\ 2.420177 \\ 2.420177 \\ 2.420177 \end{pmatrix} \times 10^{15}, \qquad \begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_5^{(2)} \end{pmatrix}_{GE} = \begin{pmatrix} 4.406998 \\ 4.406999 \\ 4.406998 \\ 4.406998 \\ 4.406999 \end{pmatrix} \times 10^{16},$$

$$\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_5^{(2)} \end{pmatrix}_{GTH} = \begin{pmatrix} 2.420177 \\ 2.420177 \\ 2.420177 \\ 2.420177 \\ 2.420177 \end{pmatrix} \times 10^{15};$$

$$\begin{pmatrix} m_1^{(4)} \\ m_2^{(4)} \\ m_3^{(4)} \\ m_4^{(4)} \\ m_5^{(4)} \end{pmatrix}_{exact} = \begin{pmatrix} 3.514354 \\ 3.514355 \\ 3.514355 \\ 3.514354 \\ 3.514354 \end{pmatrix} \times 10^{31}, \qquad \begin{pmatrix} m_1^{(4)} \\ m_2^{(4)} \\ m_3^{(4)} \\ m_4^{(4)} \\ m_5^{(4)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.165298 \\ 1.165298 \\ 1.165298 \\ 1.165298 \\ 1.165298 \end{pmatrix} \times 10^{34},$$

$$
\begin{pmatrix} m_1^{(4)} \\ m_2^{(4)} \\ m_3^{(4)} \\ m_4^{(4)} \\ m_5^{(4)} \end{pmatrix}_{GTH} = \begin{pmatrix} 3.514354 \\ 3.514354 \\ 3.514354 \\ 3.514354 \\ 3.514354 \end{pmatrix} \times 10^{31}.
$$

We remark that the first moment vector of first passage times from the chosen NCD subset to the remaining states have elements that are identical to each other and are of order 10^7. A similar argument is valid for the second moment vector of first passage times, though with the squared exponent. This situation can be attributed to the very strongly coupled interaction among the states in the NCD subset and to the very weak single transition to the remaining states. The relative errors in $(m_{GE}^{(1)}, m_{GE}^{(2)}, m_{GE}^{(4)})$ and $(m_{GTH}^{(1)}, m_{GTH}^{(2)}, m_{GTH}^{(4)})$ are respectively $(3.3 \times 10^0, 1.7 \times 10^1, 3.3 \times 10^2)$ and $(2.9 \times 10^{-7}, 0, 2.8 \times 10^{-7})$ in the infinity norm. Hence, the proposed method with GE does not have even a single digit of accuracy in any of the moment vectors of first passage times.

5.3. Test Problem 3. The third problem appears in [22] and is about a repairable computer system. The system has two processors and three memory units connected to a bus. All of these components can fail independently of each other and are repaired by a single repairman. The system is up if at least one processor, two memory units and the bus are operational. It is assumed that no failures occur when the system is down. We consider the first example model described in [22, pp. 328–329] having 21 states, 6 of which are those corresponding to operational states. For our purposes, these 6 states are classified as safe states and the rest as unsafe states.

In this case, the underlying problem yields a CTMC whose infinitesimal generator matrix is denoted by Q. This matrix has nonnegative off-diagonal elements and diagonal elements that are negated row sums of its off-diagonal elements; hence, $Qe = 0$. Now, let Y be the random variable associated with the first passage times to subset $\mathcal{U}$ conditioned on an initial probability vector τ in the subset of safe states, $\mathcal{S}$. Assuming the same kind of block partitioning as in (1) and $R = Q_{\mathcal{S},\mathcal{S}}$, the moments of Y (cf. (5)) are given by

$$
(15) \qquad E[Y^k] = k!\tau(-R^{-1})^k e
$$

for $k \geq 1$ [15, p. 45].

When $k = 1$, (15) becomes

$$
E[Y] = -\tau R^{-1} e.
$$

Then the first moment vector of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$ can be obtained from

$$
(16) \qquad -Rm^{(1)} = e.
$$

We remark that (16) becomes identical to (8) when one replaces $-R$ with $(I - T)$. The coefficient matrix, $-R$, is again a nonsingular M-matrix and $-R^{-1} \geq 0$. Hence, $m^{(1)} = -R^{-1}e$, simply the row sums of the nonnegative matrix $-R^{-1}$.

When $k = 2$, (15) becomes

$$
E[Y^2] = 2\tau R^{-2} e.
$$

Then the second moment vector of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$ can be obtained from

$$
(17) \qquad -Rm^{(2)} = 2m^{(1)}.
$$

Note that (17) has a different right-hand side compared to that of (10).

As for DTMCs, it is possible to compute higher moments of first passage times from states in $\mathcal{S}$ to $\mathcal{U}$ using a recurrence. However, the recurrence is much simpler in this case and follows from (15) as

$$
(18) \qquad -Rm^{(i+1)} = (i + 1)m^{(i)} \quad \text{for} \quad i \geq 0 \quad \text{with} \quad m^{(0)} = e.
$$

In conclusion, one needs to factorize $-Q_{\mathcal{S},\mathcal{S}}$ once, compute $m^{(1)}$, and then compute $m^{(2)}$ using $m^{(1)}$ and the factors of $-Q_{\mathcal{S},\mathcal{S}}$. The factorization can be performed using GTH as in section 4, but this time on the coefficient matrix

$$A = \begin{pmatrix} -Q_{\mathcal{S},\mathcal{S}} & -Q_{\mathcal{S},\mathcal{U}}e \\ w^T & -w^T e \end{pmatrix}.$$

in (13).

An algorithm similar to Algorithm 1 can be given for CTMCs. The matrix S should be replaced with R and R should be LU factorized using GTH. There should be a single for-loop consisting of a single statement in which the linear system with the right-hand side, $(i+1)$ times the ith moment vector, in (18) is solved for the $(i+1)$st moment vector through forward and backward substitutions using the L and U factors of R. The storage requirement of the algorithm for CTMCs is smaller than its counterpart for DTMCs since there is no need for a real vector of length $(k+1)$ to store coefficients and there is no need for k real vectors of length n_s to store moment vectors. Other than storage space for the L and U factors, it suffices to have two real vectors of length n_s, one for the moment vector to be computed and one for the right-hand side. The time requirement is also smaller since the computation of the right-hand side takes altogether only $(n_s + 1)$ flops. For a total of $(k-1)$ passes over the for-loop, the computation of the right-hand side and the forward and backward substitutions amount to $(k-1)[2(nz_L + nz_U) + k/2]$ flops.

In the reliability problem, $\mathcal{S} = \{1, 2, 3, 4, 6, 7\}$ and we have $cond_2(-Q_{\mathcal{S},\mathcal{S}}) = 2.7 \times 10^3$ from Matlab. For this problem the first ten moment vectors can be computed in single precision floating-point arithmetic. The computed results are

$$\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \\ m_7^{(1)} \end{pmatrix}_{exact} = \begin{pmatrix} 9.031009 \\ 8.940884 \\ 8.448356 \\ 8.751327 \\ 8.664630 \\ 7.766966 \end{pmatrix} \times 10^2, \quad \begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \\ m_7^{(1)} \end{pmatrix}_{GE} = \begin{pmatrix} 9.031031 \\ 8.940908 \\ 8.448377 \\ 8.751351 \\ 8.664653 \\ 7.766987 \end{pmatrix} \times 10^2,$$

$$\begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ m_3^{(1)} \\ m_4^{(1)} \\ m_6^{(1)} \\ m_7^{(1)} \end{pmatrix}_{GTH} = \begin{pmatrix} 9.031010 \\ 8.940884 \\ 8.448356 \\ 8.751328 \\ 8.664630 \\ 7.766965 \end{pmatrix} \times 10^2;$$

$$\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \\ m_7^{(2)} \end{pmatrix}_{exact} = \begin{pmatrix} 1.627131 \\ 1.610873 \\ 1.521222 \\ 1.576529 \\ 1.560893 \\ 1.397454 \end{pmatrix} \times 10^6, \quad \begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \\ m_7^{(2)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.627139 \\ 1.610882 \\ 1.521230 \\ 1.576538 \\ 1.560901 \\ 1.397461 \end{pmatrix} \times 10^6,$$

$$\begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ m_3^{(2)} \\ m_4^{(2)} \\ m_6^{(2)} \\ m_7^{(2)} \end{pmatrix}_{GTH} = \begin{pmatrix} 1.627131 \\ 1.610873 \\ 1.521222 \\ 1.576530 \\ 1.560893 \\ 1.397454 \end{pmatrix} \times 10^6;$$

$$\begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \\ m_4^{(10)} \\ m_6^{(10)} \\ m_7^{(10)} \end{pmatrix}_{exact} = \begin{pmatrix} 1.280432 \\ 1.267638 \\ 1.197083 \\ 1.240611 \\ 1.228306 \\ 1.099680 \end{pmatrix} \times 10^{36}, \quad \begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \\ m_4^{(10)} \\ m_6^{(10)} \\ m_7^{(10)} \end{pmatrix}_{GE} = \begin{pmatrix} 1.280466 \\ 1.267672 \\ 1.197115 \\ 1.240645 \\ 1.228339 \\ 1.099710 \end{pmatrix} \times 10^{36},$$

$$
\begin{pmatrix} m_1^{(10)} \\ m_2^{(10)} \\ m_3^{(10)} \\ m_4^{(10)} \\ m_6^{(10)} \\ m_7^{(10)} \end{pmatrix}_{GTH} = \begin{pmatrix} 1.280433 \\ 1.267639 \\ 1.197084 \\ 1.240612 \\ 1.228307 \\ 1.099681 \end{pmatrix} \times 10^{36}.
$$

The relative errors in $(m_{GE}^{(1)}, m_{GE}^{(2)}, m_{GE}^{(10)})$ and $(m_{GTH}^{(1)}, m_{GTH}^{(2)}, m_{GTH}^{(10)})$ are respectively $(2.7 \times 10^{-6}, 5.5 \times 10^{-6}, 2.7 \times 10^{-5})$ and $(1.1 \times 10^{-7}, 6.1 \times 10^{-7}, 7.8 \times 10^{-7})$ in the infinity norm. In this problem, the unit of time is hour, which indicates that on the average failure does not happen before a month (see $m^{(1)}$) and thus the system is considered reliable.

6. Conclusion. We have presented a relatively efficient and accurate method to compute the moments of first passage times to a subset of states in finite ergodic MCs. The method amounts to solving linear systems with a common coefficient matrix but different right-hand sides with the latter governed by a novel recurrence introduced in this paper. This proposed recurrence involves the Pascal's triangle and is shown to lead to a stable implementation. The efficiency and accuracy of the proposed method rest respectively on the smaller linear systems solved and the GTH implementation used. Due to the form of the linear systems involved, the method can be implemented in row-wise sparse format with delayed updates. The proposed method is expected to be useful in large MCs that have a relatively small number of safe states and pose computational problems.

Appendix. We prove (12) after we present some basic results about Stirling numbers, which are closely associated with binomial coefficients. We follow the discussion and notation in [5]. Further information concerning Stirling numbers may be obtained from [13].

LEMMA 1. *The kth factorial moment vector, $\gamma^{(k)}$, is a linear combination of the first k moment vectors $m^{(i)}$, $i \in \{1, 2, \ldots, k\}$, and is in the form*

$$
(19) \qquad \gamma^{(k)} = \sum_{i=0}^{k} (-1)^{k-i} \begin{bmatrix} k \\ i \end{bmatrix} m^{(i)} \quad for \quad k \geq 0 \quad with \quad m^{(0)} = e,
$$

where $\begin{bmatrix} k \\ i \end{bmatrix}$ denotes the Stirling number of the first kind that is read "k cycle i".

Proof. This result follows from the left-hand side of (5), our definition of $m^{(i)}$ as the ith moment vector, the definition of Stirling numbers of the first kind [5, p. 245], and the definition involving Stirling numbers of the first kind with alternating signs [5, p. 249]. □

REMARK 1. *Stirling numbers of the first kind satisfy the recurrence [5, p. 250]*

$$
\begin{bmatrix} k \\ i \end{bmatrix} = (k-1) \begin{bmatrix} k-1 \\ i \end{bmatrix} + \begin{bmatrix} k-1 \\ i-1 \end{bmatrix} \quad for \quad k > 0
$$

with

$$
\begin{bmatrix} k \\ i \end{bmatrix} = 0 \quad for \quad k < i, \qquad \begin{bmatrix} k \\ 0 \end{bmatrix} = 0 \quad for \quad k > 0, \quad and \quad \begin{bmatrix} k \\ k \end{bmatrix} = 1.
$$

The next result is intimately related with Lemma 1 and is used in the proof of (12).

LEMMA 2. *The kth moment vector, $m^{(k)}$, is a linear combination of the first k factorial moment vectors $\gamma^{(i)}$, $i \in \{1, 2, \ldots, k\}$, and is in the form*

$$
(20) \qquad m^{(k)} = \sum_{i=0}^{k} \begin{Bmatrix} k \\ i \end{Bmatrix} \gamma^{(i)} \quad for \quad k \geq 0 \quad with \quad \gamma^{(0)} = e,
$$

where $\begin{Bmatrix} k \\ i \end{Bmatrix}$ denotes the Stirling number of the second kind that is read "k subset i".

Proof. This result follows from Lemma 1, the definition of Stirling numbers of the second kind [5, p. 244], and (6.10) in [5, p. 248].   □

REMARK 2. *Stirling numbers of the second kind satisfy the recurrence [5, p. 250]*

$$\begin{Bmatrix} k \\ i \end{Bmatrix} = i \begin{Bmatrix} k-1 \\ i \end{Bmatrix} + \begin{Bmatrix} k-1 \\ i-1 \end{Bmatrix} \quad for \quad k > 0$$

with

$$\begin{Bmatrix} k \\ i \end{Bmatrix} = 0 \quad for \quad k < i, \quad \begin{Bmatrix} k \\ 0 \end{Bmatrix} = 0 \quad for \quad k > 0, \quad and \quad \begin{Bmatrix} k \\ k \end{Bmatrix} = 1.$$

The following is a fundamental result which is also used in the proof of (12).

LEMMA 3. *The factorial moment vectors can be computed from the recurrence*

$$(21) \qquad S\gamma^{(k+1)} = \sum_{i=0}^{k} (-1)^{k-i} \frac{(k+1)!}{i!} \gamma^{(i)} \quad for \quad k \geq 0 \quad with \quad \gamma^{(0)} = e.$$

Proof. We recall that S is a nonsingular M-matrix and $S^{-1} \geq 0$. Furthermore, $S^{-1}T = TS^{-1}$, where $T = I - S$. The basis $S\gamma^{(1)} = e$ is easy to derive from (6). Using the fact that S^{-1} and T commute, after some algebra on (6), we obtain

$$(22) \qquad \gamma^{(k+1)} = (k+1)S^{-1}T\gamma^{(k)} \quad for \quad k > 0.$$

The recurrence between two consecutive factorial moment vectors in (22) yields

$$(23) \qquad \gamma^{(k+1)} = (k+1)!(S^{-1}T)^k \gamma^{(1)} \quad for \quad k \geq 0.$$

Premultipliying both sides of (6) by T and using the fact that S^{-1} and T commute, we obtain

$$(24) \qquad T\gamma^{(k)} = k!(S^{-1}T)^{k-1}S^{-1}Te.$$

But,

$$(25) \qquad S^{-1}Te = \gamma^{(1)} - e$$

since $T = I - S$ and $S\gamma^{(1)} = e$. Hence, (24) becomes

$$\frac{T\gamma^{(k)}}{k!} = (S^{-1}T)^{k-1}\gamma^{(1)} - (S^{-1}T)^{k-1}e.$$

Using (23), this can be written as

$$(26) \qquad \frac{T\gamma^{(k)}}{k!} = \frac{\gamma^{(k)}}{k!} - (S^{-1}T)^{k-1}e.$$

Factoring $(S^{-1}T)^{k-1}$ as $(S^{-1}T)^{k-2}(S^{-1}T)$ in (26), using (25), and continuing in this manner, we obtain

$$(27) \qquad T\gamma^{(k)} = \sum_{i=0}^{k} (-1)^{k-i} \frac{k!}{i!} \gamma^{(i)} \quad for \quad k > 0 \quad with \quad \gamma^{(0)} = e.$$

Now, rewriting (22) as

$$S\gamma^{(k+1)} = (k+1)T\gamma^{(k)}$$

and using the result in (27) on its right-hand side, we have (21).   □

COROLLARY 1. *The $(k+1)$st moment vector can be computed from the linear system obtained by premultiplying both sides of (20) by S and then using the recurrence in (21).*

Now, we state and prove the theorem.

THEOREM 1. *Let P be the transition probability matrix of an ergodic DTMC, $\mathcal{S}$ be a subset of states in P, and $m^{(k+1)}$ denote the $(k+1)$st moment vector of first passage times from states in $\mathcal{S}$ to states outside $\mathcal{S}$ for $k \geq 0$. Then $m^{(k+1)}$ can be computed from the recurrence*

$$(28) \qquad Sm^{(k+1)} = \sum_{i=0}^{k} (-1)^{k-i} \binom{k+1}{i} m^{(i)} \quad \text{for} \quad k \geq 0 \quad \text{with} \quad m^{(0)} = e,$$

where $P_{\mathcal{S},\mathcal{S}}$ is the submatrix of P corresponding to the states in $\mathcal{S}$ and $S = I - P_{\mathcal{S},\mathcal{S}}$.

Proof. Our objective is to express S times the $(k+1)$st moment vector in terms of the first k moment vectors as in (28) (i.e., (12) in the text). To that end, following Corollary 1, we write

$$(29) \qquad Sm^{(k+1)} = \sum_{i=0}^{k} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} S\gamma^{(i+1)} \quad \text{for} \quad k \geq 0.$$

After substituting the result in (19) in the right-hand side of (21) and then substituting the obtained result in the right-hand side of (29), we get

$$(30) \qquad Sm^{(k+1)} = \sum_{i=0}^{k} \sum_{j=0}^{i} \sum_{l=0}^{j} (-1)^{i-l} \frac{(i+1)!}{j!} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} \begin{bmatrix} j \\ l \end{bmatrix} m^{(l)} \quad \text{for} \quad k \geq 0 \quad \text{with} \quad m^{(0)} = e.$$

We prove the theorem by showing that (30) reduces to (28); in other words,

$$(31) \qquad \sum_{i=0}^{k} \sum_{j=0}^{i} \sum_{l=0}^{j} (-1)^{i-l} \frac{(i+1)!}{j!} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} \begin{bmatrix} j \\ l \end{bmatrix} m^{(l)} = \sum_{i=0}^{k} (-1)^{k-i} \binom{k+1}{i} m^{(i)} \quad \text{for} \quad k \geq 0.$$

Let $Y(k)$ refer to the left-hand side of (31). Note that therein it is possible to replace the upper-limit of the third summation with k since $j < k$ and $\begin{bmatrix} j \\ l \end{bmatrix} = 0$ for $l > j$. Hence, we can write

$$Y(k) = \sum_{i=0}^{k} \sum_{j=0}^{i} \sum_{l=0}^{k} (-1)^{i-l} \frac{(i+1)!}{j} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} \begin{bmatrix} j \\ l \end{bmatrix} m^{(l)}$$

$$= \sum_{l=0}^{k} (-1)^{k-l} m^{(l)} \sum_{i=0}^{k} (-1)^{i-k} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} (i+1)! \sum_{j=0}^{i} \begin{bmatrix} j \\ l \end{bmatrix} / j! \; .$$

But

$$i! \sum_{j=0}^{i} \begin{bmatrix} j \\ l \end{bmatrix} / j! = \begin{bmatrix} i+1 \\ l+1 \end{bmatrix}$$

from identity (6.21) on p. 251 in [5]. Therefore, we have

$$Y(k) = \sum_{l=0}^{k} (-1)^{k-l} m^{(l)} \sum_{i=0}^{k} (-1)^{i-k} \begin{Bmatrix} k+1 \\ i+1 \end{Bmatrix} (i+1) \begin{bmatrix} i+1 \\ l+1 \end{bmatrix} \; .$$

After replacing the lower-limit of the second summation with l since $\begin{bmatrix} i+1 \\ l+1 \end{bmatrix} = 0$ for $i < l$ and then exchanging the variables i and l, we obtain

$$Y(k) = \sum_{i=0}^{k} (-1)^{k-i} m^{(i)} \sum_{l=i}^{k} (-1)^{l-k} (l+1) \begin{Bmatrix} k+1 \\ l+1 \end{Bmatrix} \begin{bmatrix} l+1 \\ i+1 \end{bmatrix} \; .$$

Hence, the only thing that remains is to show that

$$\sum_{l=i}^{k}(-1)^{l-k}(l+1)\begin{Bmatrix}k+1\\l+1\end{Bmatrix}\begin{bmatrix}l+1\\i+1\end{bmatrix}=\begin{pmatrix}k+1\\i\end{pmatrix}.$$

But this follows from identity (12) on p. 188 in [13].   □

REFERENCES

[1] K. E. AVRACHENKOV AND M. HAVIV, *The first Laurent series coefficients for singularly perturbed stochastic matrices*, Linear Algebra and Its Applications, 386 (2004), pp. 243–259.

[2] A. BERMAN AND R. J. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, SIAM Press, Philadelphia, Pennslyvania, 1994.

[3] P.-J COURTOIS, Decomposability: Queueing and Computer System Applications. Academic Press, New York, 1977.

[4] T. DAYAR AND W. J. STEWART, *On the effects of using the Grassmann-Taksar-Heyman method in iterative aggregation-disaggregation*, SIAM Journal on Scientific Computing, 17 (1996), pp. 287–303.

[5] R. L. GRAHAM, D. E. KNUTH, AND O. PATASHNIK, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[6] W. K. GRASSMANN, M. I. TAKSAR, AND D. P. HEYMAN, *Regenerative analysis and steady state distributions for Markov chains*, Operations Research, 33 (1985), pp. 1107–1116.

[7] P. G. HARRISON AND W. J. KNOTTENBELT, *Passage time distribution in large Markov chains*, Performance Evaluation Review, 30:1 (2002), pp. 77–85.

[8] W. J. HARROD AND R. J. PLEMMONS, *Comparison of some direct methods for computing the stationary distributions of Markov chains*, SIAM Journal on Scientific and Statistical Computing, 5 (1984), pp. 453–469.

[9] R. HASSIN AND M. HAVIV, *Mean passage times and nearly uncoupled Markov chains*, SIAM Journal on Discrete Mathematics, 5 (1992), pp. 386-397.

[10] D. P. HEYMAN, *Accurate computation of the fundamental matrix of a Markov chain*, SIAM Journal on Matrix Analysis and Applications, 16 (1995), pp. 954–963.

[11] D. P. HEYMAN AND D. P. O'LEARY, *What is fundamental for Markov chains: first passage times, fundamental matrices, and group generalized inverses*, in W. J. Stewart, ed., Computations with Markov Chains, 151-161, Kluwer Academic Publishers, Boston, Massachusetts, 1995.

[12] D. P. HEYMAN AND D. P. O'LEARY, *Overcoming instability in computing the fundamental matrix for a Markov chain*, SIAM Journal on Matrix Analysis and Applications, 19 (1998), pp. 534–540.

[13] C. JORDAN, *Calculus of Finite Differences*, 3rd edition, Chelsea Publishing Company, New York, 1979.

[14] J. G. KEMENY AND J. L. SNELL, *Finite Markov Chains*, Springer-Verlag, New York, 1983.

[15] G. LATOUCHE AND V. RAMASWAMI, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM Press, Philadelphia, Pennsylvania, 1999.

[16] I. MAREK, *Iterative aggregation/disaggregation methods for computing some characteristics of Markov chains. II. Fast convergence*, Applied Numerical Mathematics, 45 (2003), pp. 11–28.

[17] C. D. MEYER, *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Review, 17 (1975), pp. 443–464.

[18] C. D. MEYER, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Review, 31 (1989), pp. 240–272.

[19] C. A. O'CINNEIDE, *Entrywise perturbation theory and error analysis for Markov chains*, Numerische Mathematik, 65 (1993), pp. 109–120.

[20] C. C. PAIGE, P. H. STYAN, AND P. G. WACHTER, *Computation of the stationary distribution of a Markov chain*, Journal of Statistical Computation and Simulation, 4 (1975), pp. 1156–1179.

[21] Software for computing moments of first passage times to a subset of states in Markov chains. http://www.cs.bilkent.edu.tr/~tugrul/software.html

[22] E. DE SOUZA E SILVA AND H. R. GAIL, *Calculating cumulative operational time distributions of repairable computer systems*, IEEE Transactions on Computers, C-35 (1986), pp. 322–332.

[23] W. J. STEWART, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, New Jersey, 1994.