BRIDGING ETAQA AND RAMASWAMI'S FORMULA FOR THE SOLUTION OF M/G/1-TYPE PROCESSES *

ANDREAS STATHOPOULOS†, ALMA RISKA‡, ZHILI HUA†, AND EVGENIA SMIRNI†

Abstract.
For some time, the method of Ramaswami has been the established way to analyze M/G/1-type processes. The ETAQA method, proposed previously in [15], has offered a more efficient alternative for the exact computation of a general class of metrics for M/G/1-type processes. However, the stability of ETAQA and its relation to Ramaswami's method were not well understood. In this paper, we derive a new formulation that improves the numerical stability and computational performance of ETAQA. As with ETAQA, the resulting methodology, *new*ETAQA, solves a homogeneous system of equations to obtain the aggregate probability of a finite set of classes of states from the state space. In contrast to ETAQA, *new*ETAQA constructs its matrix $\mathbf{X}$ in a way similar to the method of Ramaswami, decoupling the computation of the probabilities of the first two initial classes of states from the computation of the aggregate probability. Because direct methods are used to solve this system, the decoupling implies an often significant speedup over ETAQA. In addition, we show that the matrix $\mathbf{X}$ is an M-matrix, and under certain conditions, $\mathbf{X}$ is also diagonally dominant and thus can be factored stably. More importantly, we show that the *new*ETAQA method is just an efficient way to implement Ramaswami's method. We also discuss alternative normalization conditions for Ramaswami's method. Our numerical experiments demonstrate the stability of our method for both stiff and well behaved processes, and for both low and high system utilizations.

Keywords: M/G/1-type processes; matrix-analytic methods; diagonally dominant M-matrices; numerical stability

1. Introduction. Matrix analytic techniques, pioneered by Marcel Neuts [12, 13], provide a framework that is widely used for the exact analysis of a general and frequently encountered class of queuing models. From these models we focus on those that capture the behavior of M/G/1 queues. The class of models that can be analyzed using M/G/1-type Markov chains are the important classes of MAP/G/1 queues and its generalization of BMAP/G/1 queues, where the arrival process is the Markovian arrival process and the batch Markovian arrival process, respectively, [13, 8]. BMAP/G/1 queues are often used as the modeling tool of choice when examining the performance of modern computer or communication systems.

Various analytic methodologies have been developed for the solution of M/G/1-type Markov processes [13, 14, 10, 11, 4, 15], with most notable the Ramaswami's algorithm which provides a numerically stable method for the calculation of the steady-state probabilities [14]. Traditional matrix-analytic algorithms are based on stochastic complementation [16] and compute the stationary probability vector of the Markov process with a recursive function. The key in the matrix-analytic solution is the computation of an auxiliary matrix called $\mathbf{G}$, on which the recursion for the computation of the stationary probability vectors is based. Iterative procedures are used for determining $\mathbf{G}$ [10].

In [15], an alternative methodology for the solution of M/G/1-type Markov processes has been proposed. Given that the state space of an M/G/1-type processes is partitioned into the boundary states $\mathcal{S}^{(0)} = \{s_1^{(0)}, \ldots, s_m^{(0)}\}$ and the sets of states $\mathcal{S}^{(j)} = \{s_1^{(j)}, \ldots, s_n^{(j)}\}$, for $j \geq 1$, this methodology, called ETAQA, differs from the

*Work supported by the National Science Foundation (ITR/AP-0112727, CCR-0098278, ACI-0090221, and ITR-0428330), and Seagate Research.

†Department of Computer Science, College of William and Mary, Williamsburg, Virginia 23187-8795, (andreas/zlhua/esmirni@cs.wm.edu).

‡Seagate Research, 1251 Waterfront Place, Pittsburgh, PA 15222, (alma.riska@seagate.com).

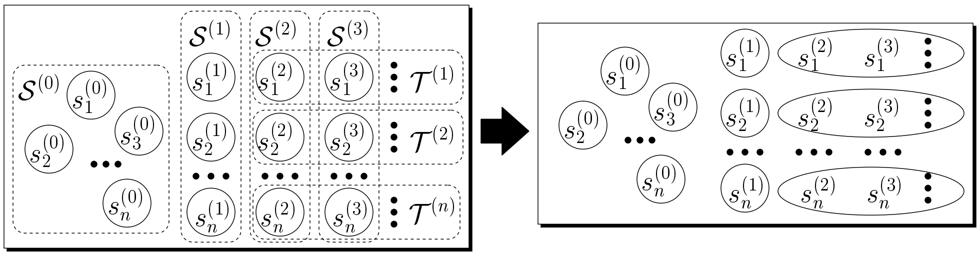

FIG. 1.1. *Aggregation of an infinite $\mathcal{S}$ into a finite number of classes of states.*

classic methods by not providing a recursive computation of the stationary probability vectors, i.e., $\boldsymbol{\pi}^{(j)}$ in $\mathcal{S}^{(j)}$ for $j \geq 1$ is *not* computed as a function of $\boldsymbol{\pi}^{(i)}$ for $i < j$. Instead, ETAQA directly computes $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and the aggregate vector $\boldsymbol{\pi}^{(*)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$. Essentially, ETAQA recasts the problem into that of solving a finite linear system of $m + 2n$ unknowns, where m is the number of states in the boundary portion of the process and n is the number of states in each of the repetitive "levels" of the state space (see Figure 1.1), and obtains the exact solution. The computation of the aggregate probability distribution that ETAQA provides is exact. ETAQA results in significantly more time-efficient solutions than the traditional methods for the M/G/1-type processes as shown both by our detailed big-O complexity analysis and by experimental comparisons against the most efficient methods [15], i.e., the fast FFT implementation of Ramamswami's formula [9]. We stress that ETAQA offers a more efficient alternative for a certain class of exact analyses of M/G/1-type processes (e.g., ETAQA can compute the average queue length and any of its higher moments exactly, but it does not compute the entire probability distribution of the process). Nonetheless, the matrix built by ETAQA does not have a structure that is easily classifiable as numerically stable, i.e., yielding a small, bounded growth factor in the factorization process.

This paper presents *new*ETAQA, an alternative methodology that bridges Ramaswami's method and ETAQA, by first obtaining the probabilities $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$, and second, encapsulating the infinite recursive formula of Ramaswami in a solution of an $n \times n$ system for the aggregate vector of probabilities, $\boldsymbol{\pi}^{(*)}$. Because of the direct solution of the linear systems, an $O(n^3)$ process, this decoupling results in a significant computational speedup over ETAQA.

More importantly, the $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ are solutions of the same infinitesimal generator matrix as in Ramaswami's method, and thus their computation is numerically stable. Computing the rest of the matrix in *new*ETAQA is also stable as it involves only additions and multiplications of positive numbers. Furthermore, the matrix yielding $\boldsymbol{\pi}^{(*)}$ is an M-matrix, which is also diagonally dominant when the system is "strongly" recurrent. In this case, the computation of $\boldsymbol{\pi}^{(*)}$ is also unconditionally numerically stable. Our experiments show that even for stiff processes (process with extreme variations between their rates), and under extremely high utilization, *new*ETAQA's numerical behavior is comparable to (and often better than) that of Ramaswami, when the latter one is iterated up to machine precision.

Finally, the new formulation in *newETAQA* allows us also to clearly show that it is theoretically equivalent to the most popular implementation of Ramaswami's method that uses an a-priori normalization of the vector $\boldsymbol{\pi}^{(0)}$ to its exact, final norm. This exact norm is the result of an inversion of the same M-matrix that we use to solve for the aggregate vector $\boldsymbol{\pi}^{(*)}$. The difference is that this implementation of

Ramaswami's method ignores this information, going on to rebuilt it iteratively. We also discuss a more efficient, and potentially more stable normalization alternative for Ramaswami's method, which to our knowledge has not been pointed out before.

In Section 2 we summarize Ramaswami's formula and the two normalization alternatives for its implementation as well as the ETAQA method. *new*ETAQA is presented in Section 3, followed by a discussion on its time complexity, its numerical stability, and its relation to Ramaswami's method. Section 5 presents experiments that demonstrate the numerical stability of the method under various conditions. We summarize our contributions in Section 6.

2. Solutions of M/G/1-type processes. In this paper, we assume continuous time Markov chains, or CTMCs, hence we refer to the infinitesimal generator $\mathbf{Q}$, but our discussion applies just as well to discrete time Markov chains, or DTMCs. For the case of M/G/1-type processes, the repetitive structure of the Markov chains allows for the following state space partitioning. The state space $\mathcal{S}$ is partitioned into the boundary states $\mathcal{S}^{(0)} = \{s_1^{(0)}, \ldots, s_m^{(0)}\}$ and the sets of states $\mathcal{S}^{(j)} = \{s_1^{(j)}, \ldots, s_n^{(j)}\}$, for $j \geq 1$, while $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(j)}$, are the stationary probability vectors for states in $\mathcal{S}^{(0)}$ and $\mathcal{S}^{(j)}$, for $j \geq 1$. The infinitesimal generator $\mathbf{Q}_{M/G/1}$ is block-partitioned as:

$$\mathbf{Q}_{M/G/1} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \widehat{\mathbf{F}}^{(2)} & \widehat{\mathbf{F}}^{(3)} & \widehat{\mathbf{F}}^{(4)} & \cdots \\ \widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2.1}$$

We use the letters "L", "F", and "B" according to whether they describe "local", 'forward", and "backward" transition rates, respectively, in relation to a set of states $\mathcal{S}^{(j)}$ for $j \geq 1$, and a "$\widehat{\ }$" for matrices related to $\mathcal{S}^{(0)}$.

For the solution of M/G/1-type processes, several algorithms exist [4, 2, 10, 13]. These algorithms first compute the matrix $\mathbf{G}$ as the solution of the matrix equation:

$$\mathbf{B} + \mathbf{L} \cdot \mathbf{G} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)} \cdot \mathbf{G}^{j+1} = \mathbf{0}. \tag{2.2}$$

The matrix $\mathbf{G}$, which is stochastic if the process is recurrent and irreducible, has an important probabilistic interpretation: an entry (k, l) in $\mathbf{G}$ expresses the conditional probability of the process first entering $\mathcal{S}^{(j-1)}$ through state l, given that it starts from state k of $\mathcal{S}^{(j)}$ [13, page 81][1]. The $\mathbf{G}$ matrix is obtained by solving Eq.(2.2) iteratively. However, recent advances show that the computation of $\mathbf{G}$ is more efficient when displacement structures are used based on the representation of M/G/1-type processes by means of QBD processes [10, 2, 1, 7]. The most efficient algorithm for the computation of $\mathbf{G}$ is the cyclic reduction algorithm [2].

2.1. Ramaswami's recursive formula. The calculation of the stationary probability vector is based on the recursive Ramaswami's formula [14], which defines the following recursive relation among stationary probability vectors $\boldsymbol{\pi}^{(j)}$ for $j \geq 1$:

[1]The probabilistic interpretation of $\mathbf{G}$ is the same for both DTMCs and CTMCs. The interpretation in [13, page 81] is consistent with the discussion in [7, page 142], where CTMCs are taken into consideration.

$$\boldsymbol{\pi}^{(j)} = -\left(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(j)} + \sum_{k=1}^{j-1}\boldsymbol{\pi}^{(k)}\mathbf{S}^{(j-k)}\right)\mathbf{S}^{(0)^{-1}} \quad \forall j \geq 1, \tag{2.3}$$

where $\widehat{\mathbf{S}}^{(j)}$ and $\mathbf{S}^{(j)}$ are defined as follows (letting $\mathbf{F}^{(0)} \equiv \mathbf{L}$):

$$\widehat{\mathbf{S}}^{(j)} = \sum_{l=j}^{\infty}\widehat{\mathbf{F}}^{(l)} \cdot \mathbf{G}^{l-j}, \quad j \geq 1, \qquad \mathbf{S}^{(j)} = \sum_{l=j}^{\infty}\mathbf{F}^{(l)} \cdot \mathbf{G}^{l-j}, \quad j \geq 0. \tag{2.4}$$

Given the above definition of $\boldsymbol{\pi}^{(j)}$ and any normalization condition, a unique vector $\tilde{\boldsymbol{\pi}}^{(0)}$ can be obtained by solving the following homogeneous system of m linear equations:

$$\tilde{\boldsymbol{\pi}}^{(0)}\left(\widehat{\mathbf{L}}^{(0)} - \widehat{\mathbf{S}}^{(1)} \cdot \mathbf{S}^{(0)^{-1}} \cdot \widehat{\mathbf{B}}\right) = \mathbf{0}. \tag{2.5}$$

Faster implementations of Ramaswami's formula are also known [9]. Once $\boldsymbol{\pi}^{(0)}$ is known using Eq.(2.5), the stationary probability vector is computed using matrix-generating functions associated with triangular Toeplitz matrices, that can be computed efficiently using fast Fourier transforms (FFTs). FFTs, however, can become the source of numerical instability. Moreover, the iterative nature of the algorithm remains, requiring computation of all the intermediate states.

2.2. Normalization choices in Ramaswami's method. When solving equation (2.5), the exact size of $\boldsymbol{\pi}^{(0)}$ is not known, so any normalization condition, say $\tilde{\boldsymbol{\pi}}^{(0)}\mathbf{1}^T = 1$, would suffice to compute the relative ratios of the probabilities inside $\boldsymbol{\pi}^{(0)}$. Once $\tilde{\boldsymbol{\pi}}^{(0)}$ is known, we can use recurrence (2.3) to compute $\tilde{\boldsymbol{\pi}}^{(j)}$ for $j \geq 1$, stopping at $j = j_0$, when $\tilde{\boldsymbol{\pi}}^{(j_0)}\mathbf{1}^T$ has reached machine precision. At this point, we can normalize all the probabilities to sum to 1, i.e., $\boldsymbol{\pi}^{(j)} = \tilde{\boldsymbol{\pi}}^{(j)}/(\sum_k \tilde{\boldsymbol{\pi}}^{(k)}\mathbf{1}^T)$. We call this an a posteriori normalization approach.

The stopping criterion presumes that levels beyond j_0 will be of smaller norm, and so their contribution is negligible, if measurable at all. Although, the convergence of $\tilde{\boldsymbol{\pi}}^{(j)}\mathbf{1}^T$ in Ramaswami's formula may not be strictly monotonic in the first few iterations, this is not expected to cause premature stopping of the iteration because recurrency of the chain asserts that the larger probability mass (far larger than machine precision) is in the earlier states. After the first few iterations, the sum of probabilities in each level j will start decreasing monotonically at every step. This eventual monotonicity can also be described analytically for the practical case of any banded infinitesimal generator $\mathbf{Q}_{M/G/1}$, where the matrix geometric approach obtains $\boldsymbol{\pi}^{(j)} = \boldsymbol{\pi}^{(j-1)}\mathbf{R}$, where $\mathbf{R}$ is a matrix with spectral radius less than 1 [13]. Finally, note that because we have assumed $\tilde{\boldsymbol{\pi}}^{(0)}\mathbf{1}^T = 1$, all subsequent probabilities $\tilde{\boldsymbol{\pi}}^{(j)}$ will be of larger magnitude than their corresponding $\boldsymbol{\pi}^{(j)}$. Thus, converging to machine precision for the $\tilde{\boldsymbol{\pi}}^{(j)}$ will, in fact, compute more accurately the $\boldsymbol{\pi}^{(j)}$.

A different approach to implementing Ramaswami's recurrence, is to precompute the exact size (normalization condition) for $\boldsymbol{\pi}^{(0)}$ [9, 3]. The condition involves replacing one of the equations in the singular Schur complement in (2.5) with the following equation:

$$\boldsymbol{\pi}^{(0)}1^T - \boldsymbol{\pi}^{(0)}\left(\sum_{j=1}^{\infty}\widehat{\mathbf{S}}^{(j)}\right) \cdot \left(\sum_{j=0}^{\infty}\mathbf{S}^{(j)}\right)^{-1}\mathbf{1}^T = 1. \tag{2.6}$$

This condition obviates the a posteriori normalization step for all computed probabilities, because $\boldsymbol{\pi}^{(0)}$ and therefore any subsequent probability is computed to their exact norm. Unfortunately, the above condition involves the computation and the inversion of the sum of all $S^{(j)}$, which requires several $O(n^3)$ operations. More importantly, this approach is essentially the second step of *new*ETAQA as described in the following sections. When this step is completed, *new*ETAQA has the $\boldsymbol{\pi}^{(*)}$, with which a host of measures can be obtained easily. Unlike *new*ETAQA, however, this version of Ramaswami's method ignores that $\boldsymbol{\pi}^{(*)}$ is available, and generates the whole stationary probability vector.

Ramaswami's method, involves a solution of the Schur complement equation in (2.5). Because its deriving matrix is an infinitesimal generator, the Schur complement is also known to have a growth factor of 1. In other words its factorization can be carried out in a numerically stable way. In addition, the iterative computation of $\boldsymbol{\pi}^{(j)}$ entails additions and multiplications with only positive numbers which can be performed in a numerically stable way if a summation algorithm similar to Kahan's is utilized [5, 13, 14]. In that case, the summation can be carried out accurately even with a series of numbers that are smaller than machine precision. The numerical properties of solving the system in (2.6) for the second approach are discussed in the following sections.

2.3. ETAQA for M/G/1-type processes. In [15] a new method for the solution of M/G/1-type processes has been defined. This technique, called ETAQA, computes only $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$ and the aggregated probability vector $\boldsymbol{\pi}^{(*)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$. This approach is exact and very efficient with respect to both its time and space complexity (see [15] for a detailed discussion).

The first step toward the solution of an M/G/1-type process is the computation of matrix $\mathbf{G}$. We assume that $\mathbf{G}$ is available, i.e., it has been computed using an efficient iterative method, e.g., the cyclic reduction algorithm [2], or that it is explicitly obtained. Then the ETAQA theorem for the solution of M/G/1-type processes can be formulated as follows (see [15]).

THEOREM 2.1 (ETAQA). *Assume an ergodic CTMC with infinitesimal generator* $\mathbf{Q}_{M/G/1}$ *having the structure shown in Eq.(2.1) and stationary probability vector* $\boldsymbol{\pi} = [\boldsymbol{\pi}^{(0)}, \ \boldsymbol{\pi}^{(1)}, \ \boldsymbol{\pi}^{(2)}, \ \ldots \ ]$. *Let also* $\boldsymbol{\pi}^{(*)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$. *Then the homogeneous system of linear equations*

$$\mathbf{x} \cdot \mathbf{X} = \mathbf{0}, \tag{2.7}$$

where $\mathbf{X} \in \mathbb{R}^{(m+2n)\times(m+2n)}$ *is defined as follows*

$$\mathbf{X} = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} - \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} & \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} \\ \widehat{\mathbf{B}} & \mathbf{L} - \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\ \mathbf{0} & \mathbf{B} - \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \end{array} \right], \tag{2.8}$$

admits a unique solution $\mathbf{x} = [\boldsymbol{\pi}^{(0)}, \ \boldsymbol{\pi}^{(1)}, \ \boldsymbol{\pi}^{(*)}]$, *with normalization* $\mathbf{x}\mathbf{1}^T = 1$.

Computing $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\boldsymbol{\pi}^{(*)}$ only does not preclude the computation of a class of stationary measures of interest. More specifically, ETAQA allows for the computation of measures of interest that can be expressed as the expected reward rate

$$r = \sum_{j=0}^{\infty} \sum_{i \in \mathcal{S}^{(j)}} \boldsymbol{\rho}_i^{(j)} \boldsymbol{\pi}_i^{(j)},$$

where $\boldsymbol{\rho}_i^{(j)}$ is the *reward rate* of state $s_i^{(j)}$.[2] If the reward rate of state $s_i^{(j)}$, for $j \geq 2$ and $i = 1, \ldots, n$, is a polynomial of degree k in j with arbitrary coefficients $\mathbf{a}_i^{[0]}, \mathbf{a}_i^{[1]}, \ldots, \mathbf{a}_i^{[k]}$, then the computation of these measures requires the solution of $k+1$ sets of linear equations, appropriately defined. For more details on the construction of this set of linear equations and its computational cost, we direct the interested reader to [15].

3. *new*ETAQA: Improved Aggregate Solution for M/G/1-type Processes. Although ETAQA provides computational savings in comparison to Ramaswami's formula [15], it gives rise to matrix $\mathbf{X}$ in eq.(2.8), which although it has row sums equal to zero, it is not an infinitesimal generator. Blocks $\mathbf{X}_{1,2}, \mathbf{X}_{2,2}, \mathbf{X}_{3,2}$ may include both positive and negative elements. The effect is that when factorizing the matrix $\mathbf{X}$ as discussed in classic numerical literature for Markov chains ([17]), we cannot theoretically bound the growth factor in the factorization. It is common knowledge in numerical analysis that numerical instability because of excessive growth factors is not encountered in practice, when pivoting strategies are employed. It is primarily the condition number of the matrix that determines the level of accuracy. Even infinitesimal generators (with growth factor of 1), can have huge condition numbers that limit their accuracy. Still, a more numerically favorable structure for the matrix $\mathbf{X}$ would be desirable. Additionally, notice that the block $\mathbf{X}_{3,2}$ couples the computation of $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ to that of $\boldsymbol{\pi}^{(*)}$ which is not required in Ramaswami's formula, beyond the normalization step.

In this section, we reformulate ETAQA's derivation to utilize the first step of Ramaswami's formula. Then, $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ are obviously computed numerically stably. Moreover, this gives rise to a matrix $\mathbf{X}_{new}$ which entails only additions, and thus can be computed stably. We show that the new matrix is an M-matrix, and that under certain conditions, it is a diagonally dominant M-matrix, which in turn suggests that the factorization of $\mathbf{X}_{new}$ yields a growth factor of 1, and a numerically stable solution.

Given the infinitesimal generator $\mathbf{Q}_{M/G/1}$ defined as in Eq.(2.1), we first rewrite the matrix equality $\boldsymbol{\pi} \cdot \mathbf{Q}_{M/G/1} = \mathbf{0}$ as:

$$
\left\{
\begin{array}{ll}
\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{L}} + \boldsymbol{\pi}^{(1)} \cdot \widehat{\mathbf{B}} & = \mathbf{0} \\
\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(1)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{B} & = \mathbf{0} \\
\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(2)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{F}^{(1)} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(3)} \cdot \mathbf{B} & = \mathbf{0} \\
\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(3)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{F}^{(2)} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{F}^{(1)} + \boldsymbol{\pi}^{(3)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(4)} \cdot \mathbf{B} & = \mathbf{0} \\
\vdots &
\end{array}
\right. \tag{3.1}
$$

Assuming that $\mathbf{G}$ from equation (2.2) is available, then the following theorem can be formulated.

THEOREM 3.1 (*new*ETAQA). *Assume an ergodic CTMC with infinitesimal generator $\mathbf{Q}_{M/G/1}$ having the structure shown in Eq.(2.1), with stationary probability vector $\boldsymbol{\pi} = [\boldsymbol{\pi}^{(0)}, \ \boldsymbol{\pi}^{(1)}, \ \boldsymbol{\pi}^{(2)}, \ ...]$. Let also $\boldsymbol{\pi}^{(*)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$. Then the homogeneous system of linear equations*

$$
\mathbf{x} \cdot \mathbf{X}_{new} = \mathbf{0}, \tag{3.2}
$$

[2] For example, to compute the expected queue length in steady state, where $\mathcal{S}^{(j)}$ represents the system states with j customers in the queue, we let $\boldsymbol{\rho}_i^{(j)} = j$. To compute the second moment of the queue length, we let $\boldsymbol{\rho}_i^{(j)} = j^2$.

where $\mathbf{X}_{new} \in \mathbb{R}^{(m+2n)\times(m+2n)}$ *is defined as follows*

$$\mathbf{X}_{new} = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} + \widehat{\mathbf{S}}^{(2)} \cdot \mathbf{G} & \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} \\ \widehat{\mathbf{B}} & \mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G} & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\ \mathbf{0} & \mathbf{0} & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \end{array}\right], \tag{3.3}$$

admits a unique solution $\mathbf{x} = [\boldsymbol{\pi}^{(0)},\ \boldsymbol{\pi}^{(1)},\ \boldsymbol{\pi}^{(*)}]$, *with normalization* $\mathbf{x} \cdot \mathbf{1}^T = 1$.

Proof. We first show that $[\boldsymbol{\pi}^{(0)},\ \boldsymbol{\pi}^{(1)},\ \boldsymbol{\pi}^{(*)}]$ is a solution of Eq.(3.2) by verifying that it satisfies three matrix equations corresponding to the three blocks of columns we used to define $\mathbf{X}_{new}$. The normalization equation is obviously satisfied:

$$\boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(*)} \cdot \mathbf{1}^T = 1. \tag{3.4}$$

(i) The first set of m equations is the first line in Eq.(3.1):

$$\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{L}} + \boldsymbol{\pi}^{(1)} \cdot \widehat{\mathbf{B}} = \mathbf{0}. \tag{3.5}$$

(ii) The second set of n equations is derived by starting with the second line in Eq.(3.1):

$$\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(1)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{B} = \mathbf{0}.$$

Because $\boldsymbol{\pi}^{(2)}$ is not part of our solution (it is part of $\boldsymbol{\pi}^{(*)}$), we express $\boldsymbol{\pi}^{(2)}$ in terms of $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ using Ramaswami's formula:

$$\boldsymbol{\pi}^{(2)} = -(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(2)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(1)})(\mathbf{S}^{(0)})^{-1}$$

Substituting this for $\boldsymbol{\pi}^{(2)}$ in the second line in Eq.(3.1) we obtain:

$$\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(1)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L} - (\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(2)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(1)})(\mathbf{S}^{(0)})^{-1} \cdot \mathbf{B} = \mathbf{0}. \tag{3.6}$$

We can further simplify the above by expressing $\mathbf{B}$ using the definition of the $\mathbf{G}$ matrix:

$$\mathbf{B} + \mathbf{L}\mathbf{G} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}\mathbf{G}^{j+1} = \mathbf{0} \ \Leftrightarrow$$

$$\mathbf{B} = -(\mathbf{L}\mathbf{G} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}\mathbf{G}^{j+1}) = -(\mathbf{L} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}\mathbf{G}^{j}) \cdot \mathbf{G} \ = -\mathbf{S}^{(0)} \cdot \mathbf{G} \tag{3.7}$$

Substituting this expression for $\mathbf{B}$ in equation (3.6) we obtain:

$$\begin{aligned} &\boldsymbol{\pi}^{(0)} \cdot \widehat{\mathbf{F}}^{(1)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L} + (\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(2)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(1)}) \cdot \mathbf{G} = \\ &\boldsymbol{\pi}^{(0)} \cdot (\widehat{\mathbf{F}}^{(1)} + \widehat{\mathbf{S}}^{(2)} \cdot \mathbf{G}) + \boldsymbol{\pi}^{(1)} \cdot (\mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G}) = \mathbf{0}. \end{aligned} \tag{3.8}$$

(iii) Another set of n equations is obtained by summing all the remaining lines in Eq.(3.1):

$$\boldsymbol{\pi}^{(0)} \cdot \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \boldsymbol{\pi}^{(1)} \cdot \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)} \cdot \left(\mathbf{L} + \sum_{i=1}^{\infty} \mathbf{F}^{(i)}\right) + \sum_{j=3}^{\infty} \boldsymbol{\pi}^{(j)} \cdot \mathbf{B} = \mathbf{0}$$

Since $\sum_{j=3}^{\infty} \boldsymbol{\pi}^{(j)} \cdot \mathbf{B}$ can be expressed as a function of $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\boldsymbol{\pi}^{(*)}$ only, the above equation can be rewritten as:

$$\boldsymbol{\pi}^{(0)} \cdot \left(\sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} \right) + \boldsymbol{\pi}^{(1)} \cdot \left(\sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \right) +$$
$$\boldsymbol{\pi}^{(*)} \cdot \left(\sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \right) = \mathbf{0}. \tag{3.9}$$

In steps (i) through (iii) we showed that the vector $[\boldsymbol{\pi}^{(0)},\ \boldsymbol{\pi}^{(1)},\ \boldsymbol{\pi}^{(*)}]$ satisfies Eqs. (3.5), (3.8), (3.9), and the normalization (3.4), hence it is a non trivial solution of Eq.(3.2). Now we have to show that this solution is unique, or equivalently, that the rank of $\mathbf{X}_{new}$ is $m + 2n - 1$. Because the last n columns are the same as in the ETAQA matrix, and these have previously been proved linearly independent, it suffices to show that all except one of the first $m + n$ columns of $\mathbf{X}_{new}$ are linearly independent. Note that the sub-matrix

$$\tilde{\mathbf{X}}_{11} = \left[\begin{array}{c|c} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} + \widehat{\mathbf{S}}^{(2)} \cdot \mathbf{G} \\ \widehat{\mathbf{B}} & \mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G} \end{array} \right] \tag{3.10}$$

of $\mathbf{X}_{new}$ is an infinitesimal generator that results when applying stochastic complementation to $\mathbf{Q}_{M/G/1}$ so that it only contains states in $\mathcal{S}^{(0)}$ and $\mathcal{S}^{(1)}$. Thus, its rank is $m+n-1$ which does not change if we extend it with two zero submatrices to yield the first $m+n$ columns of $\mathbf{X}_{new}$, and this completes the proof. ☐

Because QBD processes are a special case of M/G/1-type processes, the aggregate solution that we propose for M/G/1-type processes holds also for QBD processes.

COROLLARY 3.2 (*new*ETAQA for QBDs). *The matrix* $\mathbf{X}_{new}$ *for QBD processes takes the following form:*

$$\mathbf{X}_{new} = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \mathbf{F} & \mathbf{0} \\ \widehat{\mathbf{B}} & \mathbf{L} + \mathbf{F} \cdot \mathbf{G} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} & \mathbf{L} + \mathbf{F} + \mathbf{F} \cdot \mathbf{G} \end{array} \right] \tag{3.11}$$

Proof. The proof that the matrix $\mathbf{X}$ in Eq.(3.11) admits a unique solution of the form $[\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(*)}]$ follows the same steps as the proof of Theorem 2.1. ☐

In some cases a larger number of probability vectors may be needed, i.e., if one needs the exact vectors of $\boldsymbol{\pi}^{(i)}$, $2 \leq i < k$, where k is a predefined level. *new*ETAQA can be extended to these cases, as Theorem 3.1 can be restated as follows.

COROLLARY 3.3 (*new*ETAQA parameterized). *Assume an ergodic CTMC with infinitesimal generator* $\mathbf{Q}_{M/G/1}$ *having the structure shown in Eq.(2.1), with stationary probability vector* $\boldsymbol{\pi} = [\boldsymbol{\pi}^{(0)},\ \boldsymbol{\pi}^{(1)},\ \boldsymbol{\pi}^{(2)},\ ...]$. *Let also* $\boldsymbol{\pi}^{(k,*)} = \sum_{i=k}^{\infty} \boldsymbol{\pi}^{(i)}$. *Then the homogeneous system of linear equations*

$$\mathbf{x}_k \cdot \mathbf{X}_{new}(k) = \mathbf{0} \quad ,$$

where $\mathbf{X}_{new}(k) \in \mathbb{R}^{(m+k\cdot n)\times(m+k\cdot n)}$ *is defined by*

$$\mathbf{X}_{new}(k) = \tag{3.12}$$

$$\left[\begin{array}{c|c|c|c|c|c|c}
\widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \widehat{\mathbf{F}}^{(2)} & \cdots & \widehat{\mathbf{F}}^{(k-2)} & \widehat{\mathbf{F}}^{(k-1)} + \widehat{\mathbf{S}}^{(k)} \cdot \mathbf{G} & \sum_{i=k}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=k+1}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} \\
\widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots & \mathbf{F}^{(k-3)} & \mathbf{F}^{(k-2)} + \mathbf{S}^{(k-1)} \cdot \mathbf{G} & \sum_{i=k-1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=k}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\
\mathbf{0} & \mathbf{B} & \mathbf{L} & \cdots & \mathbf{F}^{(k-4)} & \mathbf{F}^{(k-3)} + \mathbf{S}^{(k-2)} \cdot \mathbf{G} & \sum_{i=k-2}^{\infty} \mathbf{F}^{(i)} + \sum_{i=k-1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L} & \mathbf{F}^{(1)} + \mathbf{S}^{(2)} \cdot \mathbf{G} & \sum_{i=2}^{\infty} \mathbf{F}^{(i)} + \sum_{i=3}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} & \mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G} & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & 0 & \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G}
\end{array}\right],$$

admits a unique solution $\mathbf{x}_k = \left[\ \boldsymbol{\pi}^{(0)}\ \boldsymbol{\pi}^{(1)}\ \ldots\ \boldsymbol{\pi}^{(k-1)}\ \boldsymbol{\pi}^{(k,*)}\ \right]$, *with normalization condition* $\mathbf{x}_k \cdot \mathbf{1}^T = 1$.

Proof. The proof is trivial as it follows the same steps as the proof of Theorem 2.1. ∎

We remark briefly that computing the matrix $\mathbf{X}_{new}(k)$ requires no additional costs than computing the matrix $\mathbf{X}_{new}$, as all matrix multiplications and sums are computed in the process. Naturally, the factorization of $\mathbf{X}_{new}(k)$ is more expensive because of the larger dimension $m + kn$. Still, the block Hessenberg structure of $\mathbf{X}_{new}(k)$ implies a factorization cost that grows as $O(k^2n^3)$. Although it is interesting to identify the level k at which the $\mathbf{X}_{new}(k)$ approach starts to become more expensive than Ramaswami's method, we defer this question for future research as it distracts from the focus of this paper.

4. Benefits of new ETAQA and connection to Ramaswami. There are obvious computational advantages from the new decoupled structure of $\mathbf{X}_{new}$,

$$\mathbf{X}_{new} = \left[\begin{array}{c|c} \tilde{\mathbf{X}}_{11} & \mathbf{X}_{12} \\ \mathbf{0} & \mathbf{X}_{33} \end{array}\right].$$

Instead of factorizing an $m + 2n$ matrix with $O(2/3(m + 2n)^3)$ computational cost, we solve the system in two phases. We first compute $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ as a non trivial solution of the infinitesimal generator $\tilde{\mathbf{X}}_{11}$ in (3.10). This involves $O(2/3(m + n)^3)$ operations and it is as numerically stable as Ramaswami's or any other algorithm. At the second step, we use $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ to compute $\boldsymbol{\pi}^{(*)}$ from the Schur complement as follows:

$$\boldsymbol{\pi}^{(*)} = -[\boldsymbol{\pi}^{(0)}\boldsymbol{\pi}^{(1)}] \cdot \mathbf{X}_{12} \cdot \mathbf{X}_{33}^{-1}. \tag{4.1}$$

This involves an additional $O(2/3n^3)$ computational cost. Overall, the inversion process alone in *new*ETAQA is 3 times faster than the one in ETAQA if $m = n$, and it is 4 times faster if $m = 1$. The total computational costs for each method should also include the time to build $\mathbf{X}$ and $\mathbf{X}_{new}$.

The following corollary is central not only to estimating the computational costs, but also for establishing the connection with Ramaswami's method.

Corollary 4.1.

$$\mathbf{X}_{new} = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \widehat{\mathbf{S}}^{(1)} & \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} \\ \widehat{\mathbf{B}} & \mathbf{S}^{(0)} & \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{0} & \sum_{i=0}^{\infty} \mathbf{S}^{(i)} \end{array} \right],$$

$$\mathbf{X} = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \sum_{i=1}^{\infty} \widehat{\mathbf{F}}^{(i)} - \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} & \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} \\ \widehat{\mathbf{B}} & \sum_{i=0}^{\infty} \mathbf{F}^{(i)} - \sum_{i=1}^{\infty} \mathbf{S}^{(i)} & \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{B} - \sum_{i=0}^{\infty} \mathbf{F}^{(i)} - \sum_{i=0}^{\infty} \mathbf{S}^{(i)} & \sum_{i=0}^{\infty} \mathbf{S}^{(i)} \end{array} \right].$$

Proof. First, we show the equality for each subblock of $\mathbf{X}_{new}$.

$$\begin{aligned}
\mathbf{X}_{new(12)} &= \widehat{\mathbf{F}}^{(1)} + \widehat{\mathbf{S}}^{(2)} \cdot \mathbf{G} = \widehat{\mathbf{F}}^{(1)} \cdot \mathbf{G}^{1-1} + (\sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} \cdot \mathbf{G}^{i-2}) \cdot \mathbf{G} \\
&= \sum_{i=1}^{\infty} \widehat{\mathbf{F}}^{(i)} \cdot \mathbf{G}^{i-1} = \widehat{\mathbf{S}}^{(1)} \\
\mathbf{X}_{new(22)} &= \mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G} = \mathbf{F}^{(0)} \cdot \mathbf{G}^{0} + (\sum_{i=1}^{\infty} \mathbf{F}^{(i)} \cdot \mathbf{G}^{i-1}) \cdot \mathbf{G} = \sum_{i=0}^{\infty} \mathbf{F}^{(i)} \cdot \mathbf{G}^{i} = \mathbf{S}^{(0)} \\
\mathbf{X}_{new(13)} &= \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} = \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + \sum_{i=3}^{\infty} \sum_{l=i}^{\infty} \widehat{\mathbf{F}}^{(l)} \cdot \mathbf{G}^{l-i} \cdot \mathbf{G} \\
&= \sum_{i=3}^{\infty} \left(\widehat{\mathbf{F}}^{(i-1)} + \sum_{l=i}^{\infty} \widehat{\mathbf{F}}^{(l)} \cdot \mathbf{G}^{l-i+1} \right) = \sum_{i=3}^{\infty} \sum_{l=i-1}^{\infty} \widehat{\mathbf{F}}^{(l)} \cdot \mathbf{G}^{l-i+1} \\
&= \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i-1)} = \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} \\
\mathbf{X}_{new(23)} &= \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} = \sum_{i=2}^{\infty} \left(\mathbf{F}^{(i-1)} + \sum_{l=i}^{\infty} \mathbf{F}^{(l)} \cdot \mathbf{G}^{l-i+1} \right) \\
&= \sum_{i=2}^{\infty} \mathbf{S}^{(i-1)} = \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\
\mathbf{X}_{new(33)} &= \mathbf{X}_{new(23)} + \mathbf{L} + \mathbf{S}^{(1)} \cdot \mathbf{G} = \sum_{i=1}^{\infty} \mathbf{S}^{(i)} + \mathbf{S}^{(0)} = \sum_{i=0}^{\infty} \mathbf{S}^{(i)}.
\end{aligned}$$

Using the above expressions we obtain the following for the old ETAQA matrix:

$$\begin{aligned}
\mathbf{X}_{(12)} &= \widehat{\mathbf{F}}^{(1)} - \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} = \widehat{\mathbf{F}}^{(1)} + \sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} - \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} \\
\mathbf{X}_{(22)} &= \mathbf{L} - \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} = \mathbf{F}^{(0)} + \sum_{i=1}^{\infty} \mathbf{F}^{(i)} - \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\
\mathbf{X}_{(32)} &= \mathbf{B} - \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} = \mathbf{B} - (\sum_{i=0}^{\infty} \mathbf{S}^{(i)} - \mathbf{L} - \sum_{i=1}^{\infty} \mathbf{F}^{(i)}).
\end{aligned}$$

□

We start examining the consequences of the Corollary by noting first that the third block column of the old ETAQA matrix is added to its second block column. Therefore, we could formulate the old ETAQA method to use the following equivalent matrix, which can be computed without subtractions except for the (2,3) block:

$$\mathbf{X}' = \left[\begin{array}{c|c|c} \widehat{\mathbf{L}} & \sum_{i=1}^{\infty} \widehat{\mathbf{F}}^{(i)} & \sum_{i=2}^{\infty} \widehat{\mathbf{S}}^{(i)} \\ \widehat{\mathbf{B}} & \sum_{i=0}^{\infty} \mathbf{F}^{(i)} & \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{B} - \sum_{i=0}^{\infty} \mathbf{F}^{(i)} & \sum_{i=0}^{\infty} \mathbf{S}^{(i)} \end{array}\right]. \tag{4.2}$$

Finally, we could formulate $\mathbf{X}_{new}(k)$ of Eq. (3.12) in a similar way:

$$\mathbf{X}_{new}(k) = \left[\begin{array}{c|c|c|c|c|c|c} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \widehat{\mathbf{F}}^{(2)} & \cdots & \widehat{\mathbf{F}}^{(k-2)} & \widehat{\mathbf{S}}^{(k-1)} & \sum_{i=k}^{\infty} \widehat{\mathbf{S}}^{(i)} \\ \widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots & \mathbf{F}^{(k-3)} & \mathbf{S}^{(k-2)} & \sum_{i=k-1}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \cdots & \mathbf{F}^{(k-4)} & \mathbf{S}^{(k-3)} & \sum_{i=k-2}^{\infty} \mathbf{S}^{(i)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L} & \mathbf{S}^{(1)} & \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} & \mathbf{S}^{(0)} & \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & 0 & \sum_{i=0}^{\infty} \mathbf{S}^{(i)} \end{array}\right]. \tag{4.3}$$

4.1. Computational costs of the methods. Following [17], we note that in order to solve a homogeneous linear system of equations stably, we perform an LU decomposition of the matrix and replace the last diagonal element of U (U(n,n) should be zero) with machine precision. Then, we set the last element of the zero right hand side to 1 and proceed with the solution. Later, we can enforce the required normalization on the solution vector.

Before we solve for the probability vector, all methods require the computation of all $\mathbf{S}^{(i)}, \widehat{\mathbf{S}}^{(i)}$ for $i = 0, \ldots, L_m$, where L_m is a maximum bandwidth of the infinitesimal generator. The algorithm to compute those is simply: $\mathbf{S}^{(L_m)} = \mathbf{F}^{(L_m)}$, and $\mathbf{S}^{(i)} = \mathbf{F}^{(i)} + \mathbf{S}^{(i+1)} \cdot \mathbf{G}$, $i = L_m - 1, \ldots, 0$. Therefore, the total computational time is $O(2L_m n^3)$. The $\widehat{\mathbf{S}}^{(i)}$ are computed similarly for a total cost of $O(2(L_m - 1)2mn^2)$. Thus, the setup time is

$$T_{setup} = O\left(2L_m n^3 + 4(L_m - 1)mn^2\right).$$

Ramaswami's method. Based on the first alternative that we described, we need to invert the matrix $\mathbf{S}^{(0)}$ (for a cost $O(4/3n^3)$), setup the Schur complement of equation (2.5) (cost $O(2mn^2 + 2m^2n)$), and solve it for $\boldsymbol{\pi}^{(0)}$ (cost $O(2/3m^3)$). After that, every probability vector $\boldsymbol{\pi}^{(i)}$ is obtained through the recurrence (2.3) for a cost of $O(2mn + (2L_m + 1)n^2)$. Assuming that Nit number of iterations are needed, the total cost of Ramaswami's method, including T_{setup} is:

$$T_{Ram} = O\left((2L_m + 4/3)n^3 + (4L_m - 2)mn^2 + 2m^2n + 2/3m^3 + Nit \cdot \left(2mn + (2L_m + 1)n^2\right)\right).$$

ETAQA. Using the alternative matrix in equation (4.2) for the old ETAQA method, we observe that the setup time covers all expensive operations (i.e., matrix multiplications) for building the matrix. What is still needed is the summations in the subblocks. Note that the sub blocks in line 2, ((2,2), (2,3)) can be obtained as a byproduct of the subblocks line 3. Overall, however, these summations should be negligible compared with the setup and the inversions in the program. Thus, the total time of old ETAQA should be:

$$T_{old} = O\left(2L_m n^3 + 4(L_m - 1)mn^2 + 2/3(m+2n)^3\right).$$

***new*ETAQA.** The computation of the matrix $\mathbf{X}_{new}$ is obviously slightly faster than $\mathbf{X}$, because it does not require any additions to obtain the second block column. The difference, however, is negligible. The rest of the three steps are: the solution of the $\tilde{X}_{11}$ (cost $O(2/3(m+n)^3)$, the multiplication with $\mathbf{X}_{12}$ (cost $O(mn+n^2)$), and finally the solution with $\mathbf{X}_{33}$ (cost $O(2/3)n^3$).

$$T_{new} = O\left((2L_m + 2/3)n^3 + 4(L_m - 1)mn^2 + 2/3(m+n)^3\right).$$

The table in Figure 4.1 shows how these times simplify in the extreme case of $m = 1$ and in the typical $m = n$. Note that T_{new} is always smaller than T_{old}, and it is always smaller than Ramaswami's method for small values of m, regardless of the number of iterations of the latter. When $m = n$, T_{new} is comparable to T_{Ram} when the iterations and L_m are both small. In practice, however, Ramaswami's method requires a large number of steps to converge, especially as the utilization of the system grows. The graph shown in Figure 4.1, shows that the number of iterations grows exponentially with the system utilization $(1-\boldsymbol{\pi}^{(0)})$, and that it is relatively insensitive to the size of the case.

4.2. Numerical stability of *new*ETAQA. The probabilities $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ are obtained in the same way as with Ramaswami's method (by Schur complementation), and therefore their computation is stable. To obtain $\boldsymbol{\pi}^{(*)}$, we need to factorize $\mathbf{X}_{33}$. We show that this is usually a very stable process.

LEMMA 4.2. *(see 2.5.3.12 [6]) A matrix* $\mathbf{X}$ *is an M-matrix if and only if there is a positive vector* $\mathbf{y} \in \Re^n$ *with* $\mathbf{y} \cdot \mathbf{X} > 0$.

THEOREM 4.3. *The matrix* $\mathbf{X}_{33}$ *is a negative M-matrix, i.e.,* $-\mathbf{X}_{33}$ *is an M-matrix.*

Proof. We show for a negative M-matrix, because $\mathbf{X}_{33} = \sum \mathbf{F}^{(i)} + \mathbf{L} + \sum \mathbf{S}^{(i)}\mathbf{G}$ has negative diagonals and positive off diagonals. Based on the Lemma above, we must find a positive vector y, with $y \cdot \mathbf{X}_{33} < 0$. If the chain is recurrent, there exists the probability vector $\boldsymbol{\pi}^{(*)}$, which is obviously positive. From the Schur complement derivation of $\boldsymbol{\pi}^{(*)}$ in equation (4.1), and the definition of the matrix $\mathbf{X}_{new}$ we have that $\mathbf{X}_{12}$ has positive elements, and therefore $\boldsymbol{\pi}^{(*)} \cdot \mathbf{X}_{33} = -[\boldsymbol{\pi}^{(0)}\boldsymbol{\pi}^{(1)}] \cdot \mathbf{X}_{12} < 0$. □

The M-matrix property of $\mathbf{X}_{33}$ is a highly desirable property but to guarantee a bounded growth factor in factorizations, the matrix also has to be diagonally dominant. The following describes a sufficient condition.

LEMMA 4.4. *The matrix* $\mathbf{X}_{33}$ *is diagonally dominant if* $\sum_{i=1}^{\infty} i\mathbf{F}^{(i)}\mathbf{1}^T < \mathbf{B}\mathbf{1}^T$.

Proof. From the definition of $\mathbf{X}_{new}$ we have:

$$\mathbf{X}_{33} = \sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty}\sum_{l=i}^{\infty} \mathbf{F}^{(l)}\mathbf{G}^{l-i+1}.$$

	T_{Ram}	T_{old}	T_{new}
$m=1$	$(2L_m+4/3)n^3 + Nit \cdot (2L_m+1)n^2$	$(2L_m+16/3)n^3$	$(2L_m+4/3)n^3$
$m=n$	$(6L_m+2)n^3 + Nit \cdot (2L_m+3)n^2$	$(6L_m+14)n^3$	$(6L_m+6)n^3$

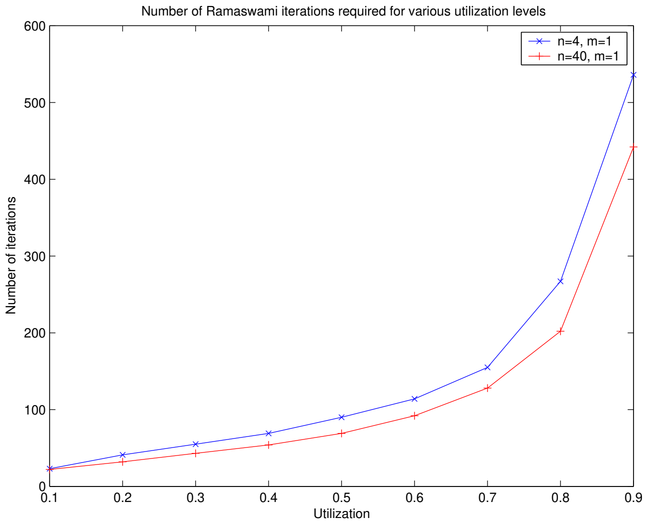

FIG. 4.1. *The table compares of computational costs for two typical values of m and for each of the three methods: Ramaswami, old ETAQA, and* new*ETAQA. The graph below shows for two cases $m=1, n=4$ and $m=1, n=40$ that the dominant cost in Ramaswami's method is the iterative process, more so as the utilization increases.*

Let $\mathbf{e}_j$ be the j-th unit vector. $\mathbf{F}^{(i)}, \mathbf{B}$ and $\mathbf{L}$ are subblocks of the infinitesimal generator, and thus satisfy:

$$\mathbf{L}_{jj} = -\sum_{k \neq j} \mathbf{L}_{j,k} - \mathbf{e}_j \sum_{i=1}^{\infty} \mathbf{F}^{(i)} \mathbf{1}^T - \mathbf{e}_j \mathbf{B} \mathbf{1}^T. \tag{4.4}$$

For any row j, we denote by d_j the absolute value of the diagonal element of $\mathbf{X}_{33}$. Then,

$$\begin{aligned} d_j &= \left| \sum_{i=1}^{\infty} \mathbf{F}_{jj}^{(i)} + \mathbf{L}_{jj} + \left(\sum_{i=1}^{\infty} \sum_{l=i}^{\infty} \mathbf{F}^{(l)} \mathbf{G}^{l-i+1} \right)_{(j,j)} \right| \\ &\equiv |f_j + \mathbf{L}_{jj} + h_j|. \end{aligned} \tag{4.5}$$

In the above, the scalars f_j and h_j are used to denote the summations of $\mathbf{F}_{jj}^{(i)}$ and the (j,j) diagonal element of the double summation of matrices, correspondingly. Let us now denote by Δ_j the sum of all off-diagonal (all positive) elements $\mathbf{X}_{33}(jk), k \neq j$. Being careful not to add the diagonals in the row sums, and using equations (4.4), (4.5), and $\mathbf{G}\mathbf{1}^T = \mathbf{1}^T$, we have:

$$\Delta_j = \sum_{k \neq j} \mathbf{L}_{j,k} + \mathbf{e}_j \sum_{i=1}^{\infty} \mathbf{F}^{(i)} \mathbf{1}^T - f_j + \mathbf{e}_j \sum_{i=1}^{\infty} \sum_{l=i}^{\infty} \mathbf{F}^{(l)} \mathbf{1}^T - h_j$$

$$
\begin{aligned}
&= \sum_{k \neq j} \mathbf{L}_{j,k} + \mathbf{e}_j \sum_{i=1}^{\infty} \mathbf{F}^{(i)} \mathbf{1}^T + \mathbf{e}_j \sum_{i=1}^{\infty} i \mathbf{F}^{(i)} \mathbf{1}^T - f_j - h_j \\
&< \sum_{k \neq j} \mathbf{L}_{j,k} + \mathbf{e}_j \sum_{i=1}^{\infty} \mathbf{F}^{(i)} \mathbf{1}^T + \mathbf{e}_j \mathbf{B} \mathbf{1}^T - f_j - h_j \\
&= -\mathbf{L}_{jj} - f_j - h_j = | - \mathbf{L}_{jj} - f_j - h_j | \qquad\qquad (\text{since } \Delta_j > 0) \\
&= d_j .
\end{aligned}
$$

□

The assumption of the above Lemma has an interesting interpretation from the point of view of recurrency of the Markov chain. It is well known, [7, 16], that for the chain to be recurrent, the following condition has to be satisfied:

$$
\tilde{\boldsymbol{\pi}} \sum_{i=1}^{\infty} i \mathbf{F}^{(i)} \mathbf{1}^T < \tilde{\boldsymbol{\pi}} \mathbf{B} \mathbf{1}^T ,
$$

where $\tilde{\boldsymbol{\pi}}$ is the solution of the infinitesimal generator $(\mathbf{B} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{F}^{(i)})$. Obviously, our assumption in the Lemma is stronger, and is not satisfied by all recurrent chains. Yet, it characterizes the chains that maintain diagonal dominance. First, observe that for "strongly recurrent" chains, i.e., chains where $\boldsymbol{\pi}^{(*)} \mathbf{1}^T$ is very small, the above condition is more likely to be satisfied. Even if it is not, the effect on the relative error in the solution is small. More interestingly, observe that our condition requires that each row on the left side of the inequality is smaller than each row on the right. The recurrency condition above, however, only requires that the weighted average of the rows on the left is less than the average of the rows on the right. Therefore, for Markov chains whose elements of $\mathbf{F}^{(i)}$ and $\mathbf{B}$ do not vary greatly along rows and from each other, the two conditions are expected to be close. Moreover, high variability of the elements between rows is also a cause of high condition numbers for $\mathbf{X}_{33}$ and the generator as well, showing how this property captures the stability behavior of such chains.

4.3. Connection of *new*ETAQA to Ramaswami. We have frequently pointed out that the computation of the probability vectors $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ are identical for both *new*ETAQA and Ramaswami's method. To compute $\boldsymbol{\pi}^{(*)}$, Ramaswami's method has the choice between an a posteriori normalization, driving the iterations to machine precision (assuming $\boldsymbol{\pi}^{(0)} \mathbf{1}^T = 1$ for equation (2.5)), or solving the additional equation (2.6) to compute the exact norm of $\boldsymbol{\pi}^{(0)}$. It should be clear by now, that the latter condition is identical to *new*ETAQA, as the additional equation involves the inverse of $\mathbf{X}_{33}$! Thus, using Ramaswami's method with equation (2.6) to compute $\boldsymbol{\pi}^{(*)}$ is completely wasteful, as it does not realize that $\boldsymbol{\pi}^{(*)}$ is implicitly available as the result of the normalization condition. Of course, other $\boldsymbol{\pi}^{(j)}$ may be required. Following Ramaswami's formula with the a posteriori normalization should be much more efficient, although still far slower than *new*ETAQA.

An immediate consequence of the above observations is the equivalent numerical behavior of *new*ETAQA and Ramaswami's method with equation (2.6). There seems to be a discrepancy with the literature, therefore, because Ramaswami's method is reported as unconditionally numerically stable. The practical difference from *new*ETAQA is that the inversion of $\mathbf{X}_{33}$ in equation (2.6) is not used to obtain the $\boldsymbol{\pi}^{(*)}$, but only to normalize $\boldsymbol{\pi}^{(0)}$. An error in the inversion would imply an error in the size of $\boldsymbol{\pi}^{(0)} \mathbf{1}^T$, which may in turn signal the recursive formula to stop too early, without having computed the accurate $\boldsymbol{\pi}^{(*)}$. In that case, the numerical behavior of Ramaswami's method is identical to *new*ETAQA. The numerical advantage of Ramaswami's method is that the summation can be continued beyond machine precision, thus correcting some of the numerical errors in the norm estimation. Note that because the probability vector has to be renormalized at the end, there is no predetermined target value for the norm, and the stopping criterion becomes unclear.

5. Experiments on numerical stability. We provide experiments that confirm our theoretical results and observations about both conditioning and stability of the problem. To be able to control both utilization (which essentially characterizes the drift of the process) and conditioning, as well as to describe a tractable chain, we have used a test case with only two $\mathbf{F}$ off diagonal identical blocks, as described below:

$$\mathbf{F} = \mathbf{F}^{(1)} = \mathbf{F}^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 200 & 200 & 200 & 200 \\ 100 & 100 & 100 & 100 \end{bmatrix}.$$

The rest of the parameters are defined below:

$$\begin{array}{l} \mathbf{B} = \mathbf{z}^T \cdot \mathbf{y}, \text{with various vectors } \mathbf{z} \text{ and} \mathbf{y} = [.25, .25, .25, .25] \\ \mathbf{L} = -diag(\sum_{rows} \mathbf{B} + 2 * \sum_{rows} \mathbf{F}) \\ \widehat{\mathbf{F}} = \mathbf{F}(1,:), \widehat{\mathbf{B}} = \sum_{rows} \mathbf{B}, \widehat{\mathbf{L}} = -2 * \sum_{rows} \mathbf{F}(1,:) \end{array}.$$

The above choice simplifies the computation of $\mathbf{G} = \mathbf{z}^T \cdot \mathbf{1}$.

Our first set of experiments is run in Matlab to establish the relative effect of utilization and the stiffness of the coefficients in the conditioning of $\mathbf{X}_{33}$. Table 5.1 shows the vectors $\mathbf{z}$ we have used and the coefficient of variation (CV) of their elements. Our goal was not only to increase the condition number but also to force $\mathbf{X}_{33}$ to lose diagonal dominance. Figure 5.1 plots the condition number of $\mathbf{X}_{33}$ against various levels of system utilization $(1 - \boldsymbol{\pi}^{(0)})$. Each curve corresponds to a different vector $\mathbf{z}$, displaying its CV. To achieve various utilization levels without affecting the CV, we change $\mathbf{z}$ by multiplying all its elements by a constant factor (.9) until we reach the recurrency threshold.

We first observe that the condition number grows both with the stiffness of the coefficients and with the system utilization, but the CV has by far the dominant role. The dependency on the utilization seems linear for most of the utilization spectrum, and it is not until utilization approaches 0.96 (i.e., a barely recurrent chain) that the condition number rises rapidly. Therefore, *new*ETAQA is not expected to have numerical problems even for chains that are barely recurrent. For those cases, Ramaswami's method would require vastly more time to converge as Figure 4.1 shows.

In our second set of experiments we measure the actual numerical error incurred by each of the three methods, ETAQA, *new*ETAQA and Ramaswami. To do so, we need a very accurate solution vector as our reference. As analytical solutions are not known, even for the simple test case described above, we ran our tests in Maple, which can simulate arbitrary floating point precision in software. In the experiments, we first run *new*ETAQA, requiring Maple to compute the probability vector with all floating point arithmetic happening in software with 100 digits in the mantissa. Then, we switch to a much smaller number of digits in the mantissa, and we run the

CV	**z**
1.0	[750, 742, 750, 730]
4.26	[1e5, 1e2, 5e3, 600.44]
5.16	[1e10, 1e4, 3e3, 340.44]
6.24	[1e14, 3e6, 1.1e3, 64]

TABLE 5.1

The choices of the vector $\mathbf{z}$ *in creating smooth or highly stiff coefficients in* $\mathbf{B}$.

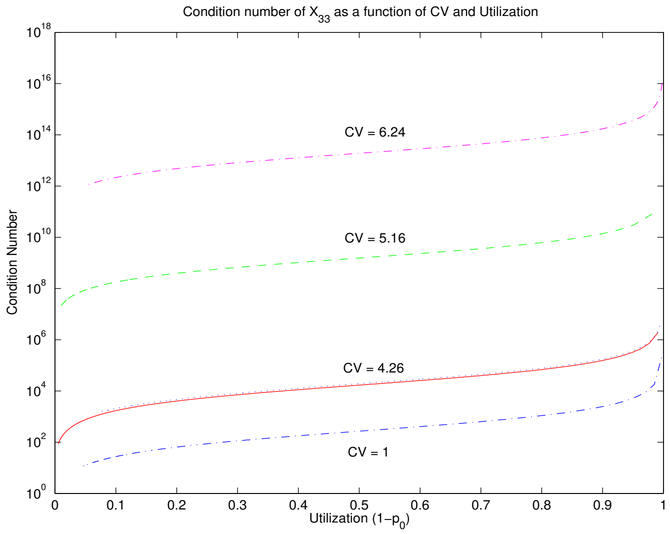

FIG. 5.1. *The norm-2 condition number of* $\mathbf{X}_{33}$ *as a function of system utilization* $(1 - \boldsymbol{\pi}^{(0)})$ *for four different cases with increasing coefficient of variation (CV) in the matrix elements. The stiffness caused by high CVs has a larger impact on the condition number.*

three methods above, reporting the 2-norm of the difference of their solution from the accurate solution. We report results for 6, 8, 10, 12, 14, and 16 digits in the mantissa, effectively simulating machines with various numerical precisions.

Figure 5.2 shows results from two highly ill conditioned problems, which are constructed as in the previous set of experiments, the first with $\mathbf{z} =$[1e14,1e5,1e3,12.71], and the second with $\mathbf{z} =$[1e14,1e5,1e3,19]. Despite their high conditioning, their utilizations are very different; the first 0.988, almost non-recurrent, and the second 0.50. The ETAQA matrix and method is implemented in the original way. The *new*ETAQA is implemented in the way described in this paper. For Ramaswami's method we use the a priori normalization of (2.6). For the first test case, Ramaswami's code is iterated until $\boldsymbol{\pi}^{(j)}\mathbf{1}^T < 10^{-(\text{Digits used})}$, while for the second case, $\boldsymbol{\pi}^{(j)}\mathbf{1}^T < 10^{-(\text{Digits used}+1)}$.

The first observation is that none of the methods achieves accuracy equal to the (virtual) machine precision, with *new*ETAQA being the closest for both cases. The condition number of $\mathbf{X}_{33}$ is obviously the limiting factor. The original ETAQA is performing surprisingly well too. Ramaswami's method, despite several hundreds of thousands of iterations, did not achieve a better accuracy. This is exactly the behavior we predicted in the previous section. In fact, in the second test Ramaswami's method has come slightly closer to the precision of *new*ETAQA, as we allowed it to iterate

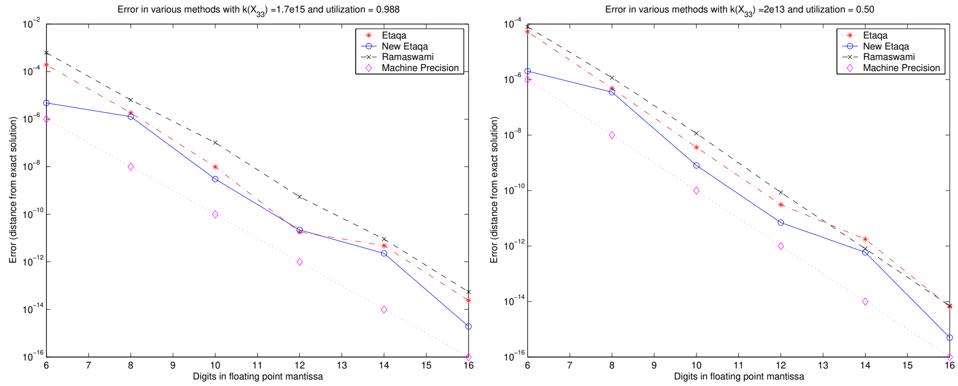

FIG. 5.2. *Numerical accuracy of the methods of Ramaswami, ETAQA and* new*ETAQA for two ill conditioned cases. We plot the difference from the accurate solution* $\|\mathbf{x} - \mathbf{x}_{accurate}\|_2$ *for each method, under various machine precisions. The methods are implemented in Maple, which allows for software simulated arbitrary machine precision, by setting the digits in the mantissa of floating point numbers. 100 digits are used for the accurate solution.*

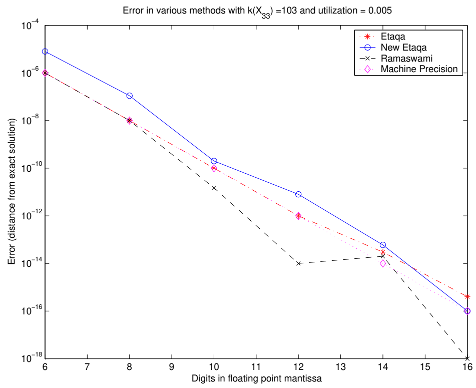

FIG. 5.3. *Numerical accuracy of the methods of Ramaswami, ETAQA and* new*ETAQA for a well conditioned case. We plot the difference from the accurate solution* $\|\mathbf{x} - \mathbf{x}_{accurate}\|_2$ *for each method, under various machine precisions. The methods are implemented in Maple, which allows for software simulated arbitrary machine precision, by setting the digits in the mantissa of floating point numbers. 100 digits are used for the accurate solution.*

longer, crossing the machine precision boundary by one digit.

For the third experiment, we modified a $\mathbf{z}$ from Table 5.1 so that it yields a very low condition number and an extremely low utilization. We run the same experiments as before, only now Ramaswami's code is allowed to converge to two digits lower than the machine precision, i.e., $10^{-(\text{Digits used}+2)}$. Figure 5.3 shows the results. The difference between ETAQA and *new*ETAQA is negligible, and they are both within the machine precision noise. Ramaswami's method, however, is able to achieve, in certain cases, accuracy about two orders of magnitude better than the machine precision. This is because we allowed it to continue computing probabilities for two more orders of magnitude lower, and also because the low utilization implies that just a few iterations are sufficient. This is in full agreement with our theoretical results.

6. Conclusions. We have provided a restructuring of the ETAQA matrix for the solution of the M/G/1-type processes, that leads to significant computational savings for the direct computation of the aggregate probability vector $\boldsymbol{\pi}^{(*)}$, as explained via Big-O notation. More importantly, we provide a theoretical explanation of the proximity of the *new*ETAQA method to Ramaswami's recursive formula. The computation of the $\boldsymbol{\pi}^{(*)}$, that ETAQA is based on, is also implicitly provided (but not explicitly used) by Ramaswami's formula.

Additionally, the computations in this new ETAQA formulation are numerically stable as the resulting $\mathbf{X}_{new}$ matrix is an M-matrix, which under conditions of strong recurrency, is also diagonally dominant. Regardless of diagonal dominance, the numerical behavior is theoretically equivalent to that of Ramaswami's method that uses a typical normalization condition. The advantage of Ramaswami's method is that it could continue the iteration beyond machine precision, albeit at additional computational expense and an unclear stopping criterion. These theoretical connections have thus provided new intuition on Ramaswami's numerical behavior, and have suggested an alternative implementation that is faster and potentially more stable. Experimentation using arbitrary floating point precision is in strong agreement with the above theoretical results.

Acknowledgments. We thank B. Meini and V. Ramaswami for insightful discussions.

REFERENCES

[1] D. A. Bini and B. Meini. Using displacement structure for solving non-skip-free M/G/1 type Markov chains. In A. S. Alfa and S. R. Chakravarthy, editors, *Advances in Matrix Analytic Methods for Stochastic Models*, pages 17–37, NJ, 1998. Notable Publications Inc.

[2] D. A. Bini, B. Meini, and V. Ramaswami. Analyzing M/G/1 paradigms through QBDs: the role of the block structure in computing the matrix G. In G. Latouche and P. Taylor, editors, *Advances in Matrix Analytic Methods for Stochastic Models*, pages 73–86, NJ, 2000. Notable Publications Inc.

[3] G. Ciardo and E. Smirni. ETAQA: An efficient technique for the analysis of QBD-processes by aggregation. *Performance Evaluation*, 36-37:71–93, 1999.

[4] W. K. Grassman and D. A. Stanford. Matrix analytic methods. In W. K. Grassman, editor, *Computational Probability*, pages 153–204, Boston, MA, 2000. Kluwer Academic Publishers.

[5] Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM, Philadelphia, 1996.

[6] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1991.

[7] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* SIAM, Philadelphia PA, 1999. ASA-SIAM Series on Statistics and Applied Probability.

[8] D. M. Lucantoni. The BMAP/G/1 queue: A tutorial. In L. Donatiello and R. Nelson, editors, *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, pages 330–358. Springer-Verlag, 1993.

[9] B. Meini. An improved FFT-based version of Ramaswami's formula. *Comm. Statist. Stochastic Models*, 13:223–238, 1997.

[10] B. Meini. Solving M/G/1-type Markov chains: Recent advances and applications. *Comm. Statist. Stochastic Models*, 14(1 - 2):479–496, 1998.

[11] N.C. Oguz N. Akar and K. Sohraby. Matrix-geometric solution in finite and infinite m/g/1 type markov chains: A unifying generalized state-space approach. *Journal of Selected Areas in Communications*, 16(5), 1998.

[12] M. F. Neuts. *Matrix-geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, MD, 1981.

[13] M. F. Neuts. *Structured Stochastic Matrices of M/G/1-type and their Applications.* Marcel Dekker, New York, NY, 1989.

[14] V. Ramaswami. A stable recursion for the steady state vector in Markov chains of M/G/1-type. *Commun. Statist. Stochastic Models*, 4:183–189, 1988.

[15] A. Riska and E. Smirni. Exact aggregate solutions for M/G/1-type Markov processes. In *Proceedings of ACM SIGMETRICS Conference*, pages 86–96, Marina del Rey, CA, June 2002.

[16] A. Riska and E. Smirni. M/G/1-type Markov processes: A tutorial. In M. C. Calzarossa and S. Tucci, editors, *Performance Evaluation of Complex Computer Systems: Techniques and Tools, LNCS 2459*, pages 36–63. Springer-Verlag, 2002.

[17] William J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.