On Quantum Algorithms for Efficient Solutions of General Classes of Structured Markov Processes

Vasileios Kalantzis, Mark S. Squillante, Shashanka Ubaru
Mathematical Sciences, IBM Research
Thomas J. Watson Research Center
Yorktown Heights, NY 10598, USA

vkal@ibm.com, mss@us.ibm.com, Shashanka.Ubaru@ibm.com

1. INTRODUCTION

Multidimensional Markov processes arise in many aspects of the mathematical performance analysis, modeling and optimization of computer systems and networks. Within this context, general classes of structured Markov processes are of particular importance in both theory and practice. Our primary focus in this paper is on the general class of *M/G/1-type* processes, noting the important duality between M/G/1-type, *G/M/1-type*, and *quasi-birth-and-death* processes [1].

Consider a discrete-time Markov process $\{X(n)\,;\, n \in \mathbb{Z}_+ := \mathbb{Z}_{\geq 0}\}$ on the state space $\widehat{\Omega} = \{(i,j) : i \in \mathbb{Z}_+, j \in [M]\}$ with transition probability matrix $\mathbf{P}$, where $[M] := \{1, \ldots, M\}$ and M can be finite or infinite. The $\mathbf{P}$ matrix of the M/G/1-type process has the block Toeplitz-like structure

$$\mathbf{P} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \cdots \\ \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots \\ \mathbf{0} & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{-1} & \mathbf{A}_0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}, \text{ where } \mathbf{B}_i,\ i \in \mathbb{Z}_+,$$

and $\mathbf{A}_i$, $i \in \mathbb{Z}_{\geq -1}$, are nonnegative matrices in $\mathbb{R}^{M \times M}$ s.t. $\sum_{i=-1}^{\infty} \mathbf{A}_i$ and $\sum_{i=0}^{\infty} \mathbf{B}_i$ are stochastic [3]. Markov processes embedded at service completion epochs for the M/G/1 queue have transition probability matrices of the form of $\mathbf{P}$.

Our objective is to obtain the stationary distribution of the M/G/1-type process $\{X(n)\,;\, n \in \mathbb{Z}_+\}$. Define $\boldsymbol{\pi} := (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, $\boldsymbol{\pi}_i := (\widehat{\pi}(i,1), \widehat{\pi}(i,2), \ldots, \widehat{\pi}(i,M))$, $\widehat{\pi}(i,j) := \lim_{t \to \infty} \mathbb{P}[X(t) = (i,j)]$, $\forall (i,j) \in \widehat{\Omega}$. The limiting probability vector $\boldsymbol{\pi}$ is the stationary distribution for the stochastic process. Assuming this process to be irreducible and ergodic, its invariant probability vector $\boldsymbol{\pi}$ exists and is uniquely determined as the solution of $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ and $\boldsymbol{\pi}\mathbf{1} = 1$, where $\mathbf{1} = (1, \ldots, 1)^\top$. Various functions of the stationary distribution $\boldsymbol{\pi}$ can be used to obtain performance measures and quantities of interest including those associated with *levels* $\mathcal{L}(i) := \{(i,j) : j \in [M]\}$ and *phases* $\mathcal{P}(j) := \{(i,j) : i \in \mathbb{Z}_+\}$, such as measures of queue length and sojourn time.

One form of mathematical analysis to determine the stationary distribution of M/G/1-type processes concerns computational approaches based on numerical analysis and numerical methods. The best-known algorithms of this approach consist of variants of cyclic reduction (CR), originally proposed by Buzbee, Golub and Nielsen in 1970. Motivated by queueing-theoretic problems, Bini and Meini developed various extensions of CR for solving M/G/1-type processes [1]. This CR-based computational approach represents the most efficient solutions for computing the stationary distribution $\boldsymbol{\pi}$ of M/G/1-type processes on classical digital computers.

Even with the computational benefits of CR, the time required to compute the stationary distribution can still be prohibitive for large stochastic performance models. Quantum computers offer the potential of achieving significant advantages for certain computational problems, with the possibility of delivering polynomial-to-exponential speedups over the best solutions on classical computers. However, despite this great potential, several important quantum algorithms have been shown to provide modest or no benefits over the best classical algorithms [2]. Moreover, to our knowledge, there are no quantum algorithms for computing the stationary distribution of M/G/1-type processes and no quantum algorithms that realize the potential significant speedup.

Herein, with a primary mathematical focus, we derive the first quantum algorithms for computing the stationary distribution of M/G/1-type processes, and we derive a mathematical analysis of the computational properties of our quantum algorithms and related theoretical results. This includes establishing the potential for a polynomial-to-exponential speedup over the best classical algorithms in various settings of both theoretical and practical importance. Due to space restrictions, we refer the reader to [2] for many of the technical details on M/G/1-type processes, quantum computing, derivations of CR methods and related theoretical results, proofs of our theoretical results, and additional references.

2. ALGORITHMIC SOLUTIONS

Classical CR Algorithms. Define $\theta(\ell) := \min\{n \geq 0 : X(n) \in \mathcal{L}(\ell)\}$ and $\mathbf{G}_{jk}^{(i)} := \mathbb{P}[\theta(\ell) < \infty, X(\theta(\ell)) = (\ell,k) \,|\, X(0) = (i+\ell, j)]$, whose jk-th element denotes the probability that, starting from state $(i+\ell, j) \in \mathcal{L}(i+\ell)$ at time 0, the process enters $\mathcal{L}(\ell)$ for the first time in finite time with (ℓ,k) being the first state visited in $\mathcal{L}(\ell)$. The solution of the stationary distribution $\boldsymbol{\pi}$ is then determined in terms of the matrix $\mathbf{G}_{\min}$, which is the minimal nonnegative solution of the matrix equation $\mathbf{X} = \sum_{i=-1}^{\infty} \mathbf{A}_i \mathbf{X}^{i+1}$.

The classical CR approach basically consists of rewriting the equation $\mathbf{X} = \sum_{i=-1}^{\infty} \mathbf{A}_i \mathbf{X}^{i+1}$ in the matrix form $\mathbf{H}[\mathbf{G}^{(1)}\, \mathbf{G}^{(2)}\, \mathbf{G}^{(3)}\, \cdots]^\top = \mathbf{b} := [\mathbf{A}_{-1}\, \mathbf{0}\, \mathbf{0}\, \cdots]^\top$, and applying variants of CR to solve this rewritten matrix form. A single CR iteration consists of applying an even-odd permutation to the block rows and block columns of the matrix $\mathbf{H}$ followed by one iteration of block Gaussian elimination. This transformation maintains the unchanged structure of the system,

MAMA 2024 Venice, Italy.

and thus CR is applied recursively which yields a sequence of block Hessenberg, block Toeplitz-like infinite systems that provably converges quadratically to a limit system whose solution can be explicitly evaluated. The matrix of the infinite system at each iteration is fully characterized by its first and second block rows whose respective block entries define two formal matrix power series with an explicit functional relation that expresses in functional form the CR iteration.

The computational complexity of the classical CR algorithm is given by $O(M^3 d^C_{\max} + M^2 d^C_{\max} \log d^C_{\max})$, where $d^C_{\max}$ is the maximum numerical degree of the matrix power series generated by CR and M is the order of the block matrices of $\mathbf{P}$. This computational complexity can be prohibitive for large values of M and $d_{\max}$. We next present our quantum algorithms to address the computational bottlenecks of the classical CR algorithm on digital computers.

Quantum CR Algorithms. Focusing on the primary computational bottlenecks of classical CR, we note that the CR iterations involve Toeplitz linear systems and Toeplitz matrix products. Given a Toeplitz matrix $\mathbf{T}^{[N]}$ with entries $t_{ij} = t_{i-j}$, the entries can be viewed as Fourier coefficients of a certain 2π-periodic strictly positive continuous real-valued function f defined on $[0, 2\pi]$ as $t_k = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda) e^{-ik\lambda} d\lambda$, $k \in \{0, \pm 1, \ldots, \pm(n-1)\}$. With this generating function f of the sequence $\mathbf{T}^{[N]}$, we can construct a sequence of associated circulant matrices $\mathbf{C}^{[N]}$ that converge super linearly to these Toeplitz matrices $\mathbf{T}^{[N]}$ as N increases. In particular, we define a circulant matrix $\mathbf{C}^{[N]}$ with top row $(c_0, c_1, \ldots, c_{N-1})$ where $c_k = \frac{1}{N}\sum_{j=0}^{N-1} f(2\pi j/N) e^{2\pi i j k/N}$. The eigenvalues of $\mathbf{C}^{[N]}$ are simply $f(2\pi j/N)$, $j \in \{0, 1, \ldots, N-1\}$, and the corresponding eigenvectors are the columns of the Fourier matrix Φ. Our quantum approach therefore consists of efficient matrix operations on the matrices $\mathbf{C}^{[N]}$, for which we have the inverse $(\mathbf{C}^{[N]})^{-1} = \Phi^\dagger \Lambda^{-1} \Phi$, where Λ^{-1} is a diagonal matrix with eigenvalues $1/f(2\pi j/N)$.

Our quantum CR Alg. Q.1 executes Alg. Q.2 on a quantum computer with Toeplitz matrix inputs $\mathbf{T}_1 = (\mathbf{I} - \mathbf{U}^{(n)}_{11})$, $\mathbf{T}_2 = -\mathbf{U}^{(n)}_{21}$ having generating functions f_1, f_2. We encode the N columns of $\mathbf{T}_3 = \mathbf{U}^{(n)}_{12}$ as an initial state $|\psi_0\rangle$ and operate on each column independently and in quantum parallel. Then, we approximately compute $\mathbf{T}_2\mathbf{T}_1^{-1}|\psi_0\rangle = \Phi^\dagger \Lambda(f_2)\Lambda^{-1}(f_1)\Phi|\psi_0\rangle$, where $\Lambda(f_1)$, $\Lambda(f_2)$ are diagonal matrices corresponding to the eigenvalues of $\mathbf{T}_1$, $\mathbf{T}_2$. With oracle access to f_1, f_2 by exploiting appropriate oracle quantum circuit methods, and given the Fourier coefficients $\{t_k\}$, we compute these generating functions using efficient quantum circuits with the corresponding Fourier expansions. Next, we apply QFT to the initial state to obtain $|\psi_0'\rangle = \Phi|\psi_0\rangle$. To encode the eigenvalues of $\mathbf{T}_1$, we exploit oracle access to the values of f_1 to map $|\psi_0'\rangle = \sum_{j=0}^N b_j|j\rangle \rightarrow \sum_{j=0}^N b_j|j\rangle|f_1(2\pi j/N)\rangle$. Then, to invert the eigenvalues, we add a qubit, using controlled-rotations to invert the phases (eigenvalues), and use uncompute to obtain an approximation to the state $|\psi_1\rangle = \Lambda^{-1}(f_1)|\psi_0'\rangle$. We now encode the eigenvalues of $\mathbf{T}_2$ exploiting the oracle access to the values of f_2 and map $|\psi_1'\rangle = \sum_{j=0}^N \widetilde{b}_j|j\rangle \rightarrow \sum_{j=0}^N \widetilde{b}_j|j\rangle|f_2(2\pi j/N)\rangle$, and then apply iQFT to obtain a state proportional to $|\psi_2\rangle = \Phi^\dagger\Lambda(f_2)\Lambda^{-1}(f_1)\Phi|\psi_0\rangle$. Upon computing the state $|\psi^*\rangle$, Alg. Q.1 checks the solution accuracy by computing $\|\mathbf{1} - (\mathbf{A}_{-1} + \widehat{\mathbf{Z}}_0)\mathbf{1}\|_\infty$ on the quantum computer. Let $|a_{-1}\rangle = \mathrm{vec}(\mathbf{A}_{-1})$ and $\mathrm{mask}(|\psi^*\rangle) = \mathrm{vec}(\widehat{\mathbf{Z}}_0)$, where $\mathrm{mask}(|\psi^*\rangle)$ considers only the amplitudes corresponding to $\widehat{\mathbf{Z}}_0$ in $|\psi^*\rangle$. We load the columns of $\mathbf{A}_{-1}$ to form the state $|a_{-1}z_0\rangle = |a_{-1}\rangle + \mathrm{mask}(|\psi^*\rangle)$, and compute $\|\mathbf{1} - (\mathbf{A}_{-1} + \widehat{\mathbf{Z}}_0)\mathbf{1}\|_\infty$ from $|a_{-1}z_0\rangle$. If $\|\mathbf{1} - (\mathbf{A}_{-1} + \widehat{\mathbf{Z}}_0)\mathbf{1}\|_\infty > \epsilon$, we repeat the inner loop; else, we output $\mathbf{J} = (\mathbf{I} - \widehat{\mathbf{Z}}_0)^{-1}\mathbf{A}_{-1}$, again loading $|a_{-1}\rangle = \mathrm{vec}(\mathbf{A}_{-1})$ and using the above Toeplitz solution approach to form the state $|\widetilde{j}\rangle = \mathrm{vec}(\mathbf{J})$.

Algorithm Q.1 Quantum CR for M/G/1-type Processes

Input: Positive integer d, $M \times M$ block matrices $\mathbf{A}_i$, $i \in \{-1, 0, 1, \ldots, d-1\}$, defining the block Toeplitz, block Hessenberg matrix $\mathbf{H}$, and error tolerance $\epsilon > 0$.

Output: An approximation $\mathbf{J}$ to the matrix $\mathbf{G}_{\min}$.

1. Set $n = 0$, consider quantum circuits for matrices $\mathbf{A}^{(0)}_{-1} = \mathbf{A}_{-1}$, $\mathbf{A}^{(0)}_0 = \mathbf{A}_0$, ..., $\mathbf{A}^{(0)}_{d-1} = \mathbf{A}_{d-1}$, and $\widehat{\mathbf{A}}^{(0)}_0 = \mathbf{I} - \mathbf{A}_0$, $\widehat{\mathbf{A}}^{(0)}_1 = -\mathbf{A}_1$, ..., $\widehat{\mathbf{A}}^{(0)}_{d-1} = -\mathbf{A}_{d-1}$.

2. Call Alg. Q.2 with inputs $\mathbf{T}_1 = (\mathbf{I} - \mathbf{U}^{(n)}_{11}), \mathbf{T}_2 = -\mathbf{U}^{(n)}_{21}, \mathbf{T}_3 = \mathbf{U}^{(n)}_{12}$ and $\mathbf{T}_4 = (\mathbf{I} - \mathbf{U}^{(n)}_{22})$ to obtain a quantum output state $|\psi^*\rangle$ which is a vectorization of matrices $\mathbf{Z}_{-1}, \mathbf{Z}_0, \ldots, \mathbf{Z}_{d'-2}, \widehat{\mathbf{Z}}_0, \widehat{\mathbf{Z}}_1, \ldots, \widehat{\mathbf{Z}}_{d'-1}$.

3. Let $|a_{-1}\rangle = \mathrm{vec}(\mathbf{A}_{-1})$ and $\mathrm{mask}(|\psi^*\rangle) = \mathrm{vec}(\widehat{\mathbf{Z}}_0)$. Load columns of $\mathbf{A}_{-1}$ s.t. we form the state $|a_{-1}z_0\rangle = |a_{-1}\rangle + \mathrm{mask}(|\psi^*\rangle)$. Compute $\|\mathbf{1} - (\mathbf{A}_{-1} + \widehat{\mathbf{Z}}_0)\mathbf{1}\|_\infty$ from $|a_{-1}z_0\rangle$. If $\|\mathbf{1} - (\mathbf{A}_{-1} + \widehat{\mathbf{Z}}_0)\mathbf{1}\|_\infty > \epsilon$, set $n = n + 1$, set $\mathbf{A}^{(n)}_i = \mathbf{Z}_i$, $i \in \{-1, \ldots, d'-2\}$, set $\widehat{\mathbf{A}}^{(n)}_i = \widehat{\mathbf{Z}}_i$, $i \in \{0, \ldots, d'-1\}$, and repeat Steps **2** and **3**.

4. Load $|a_{-1}\rangle = \mathrm{vec}(\mathbf{A}_{-1})$ and use the Toeplitz solution approach in Alg. Q.2 to form the state $|j\rangle = \mathrm{vec}(\mathbf{J})$ with $\mathbf{J} = (\mathbf{I} - \widehat{\mathbf{Z}}_0)^{-1}\mathbf{A}_{-1}$.

Mathematical Analysis. After n iterations of our quantum CR, we have the general system in the matrix form $\mathbf{H}^{(n)}\mathbf{x}^{(n)}_- = \mathbf{b}$ with $\mathbf{x}^{(n)}_- = [\mathbf{G}_{\min}\mathbf{G}^{2^n+1}_{\min}\mathbf{G}^{2\cdot 2^n+1}_{\min}\mathbf{G}^{3\cdot 2^n+1}_{\min}\cdots]^\top$. For each iteration n, consider the computation of the block Hessenberg matrix $\mathbf{H}^{(n)}$ w.r.t. the matrix power series $\varphi^{(n)}(z) := \sum_{i=-1}^\infty z^{i+1}\mathbf{A}^{(n)}_i$, $\widehat{\varphi}^{(n)}(z) := \sum_{i=0}^\infty z^i \widehat{\mathbf{A}}^{(n)}_i$ defining the first two block rows of $\mathbf{H}^{(n)}$, whose computed approximations are $\vartheta^{(n)}(z)$, $\widehat{\vartheta}^{(n)}(z)$ with corresponding approximation errors $\mathbf{R}^{(n)}(z) := \vartheta^{(n)}(z) - \varphi^{(n)}(z)$, $\widehat{\mathbf{R}}^{(n)}(z) := \widehat{\vartheta}^{(n)}(z) - \widehat{\varphi}^{(n)}(z)$. Let $\mathbf{V}^{(n)}(z) = \varphi^{(n)}_{even}(z)(\mathbf{I} - \varphi^{(n)}_{odd}(z))^{-1}$, $\mathbf{W}^{(n)}(z) = (\mathbf{I} - \varphi^{(n)}_{odd}(z))^{-1}\varphi^{(n)}_{even}(z)$, $\widehat{\mathbf{W}}^{(n)}(z) = (\mathbf{I} - \varphi^{(n)}_{odd}(z))^{-1}\widehat{\varphi}^{(n)}_{even}(z)$; let $\mathcal{T}$ denote the quantum CR $n+1$ iteration transformation $(\varphi^{(n+1)}(z), \widehat{\varphi}^{(n+1)}(z)) = \mathcal{T}(\varphi^{(n)}(z), \widehat{\varphi}^{(n)}(z))$; let $\doteq$ and $\dot{\leq}$ respectively denote equality and inequality up to higher-order error terms; and let $\|\mathbf{S}(z)\|_* = \|\sum_{i=0}^\infty |\mathbf{S}_i|\|_\infty$ denote the max norm where $\mathbf{S}(z) = \sum_{i=0}^\infty z^i \mathbf{S}_i$ is a matrix power series in the Wiener algebra $\mathcal{W}$. We derive in [2] a mathematical analysis of the computational errors and computational complexity of our quantum CR. Here we summarize the theoretical results of our analysis in the following sequence of theorems, starting with a first-order analysis of the approximation errors.

THEOREM 1. *We have the first-order equalities* $\mathbf{R}^{(n+1)}(z) \doteq z\mathbf{R}^{(n)}_{odd}(z) + \mathbf{R}^{(n)}_{even}(z)\mathbf{W}^{(n)}(z) + \mathbf{V}^{(n)}(z)\mathbf{R}^{(n)}_{even}(z) + \mathbf{V}^{(n)}(z) \times \mathbf{R}^{(n)}_{odd}(z)\mathbf{W}^{(n)}(z)$ *and* $\widehat{\mathbf{R}}^{(n+1)}(z) \doteq \widehat{\mathbf{R}}^{(n)}_{odd}(z) + \mathbf{R}^{(n)}_{even}(z)\widehat{\mathbf{W}}^{(n)}(z) + \mathbf{V}^{(n)}(z)\widehat{\mathbf{R}}^{(n)}_{even}(z) + \mathbf{V}^{(n)}(z)\mathbf{R}^{(n)}_{odd}(z)\widehat{\mathbf{W}}^{(n)}(z)$, *and the corresponding first-order upper bounds* $\|\mathbf{R}^{(n+1)}(z)\|_* \dot{\leq} 2\|\mathbf{R}^{(n)}(z)\|_* \times (1 + \|\mathbf{V}^{(n)}(1)\|_\infty)$ *and* $\|\widehat{\mathbf{R}}^{(n+1)}(z)\|_* \dot{\leq} \|\widehat{\mathbf{R}}^{(n)}(z)\|_*(1 + \|\mathbf{V}^{(n)}(1)\|_\infty) + \|\mathbf{R}^{(n)}(z)\|_*(1 + \|\mathbf{V}^{(n)}(1)\|_\infty)$.

Exploiting Theorem 1, we establish an upper bound on the

Algorithm Q.2 Single Iteration n of Quantum CR

Input: Four $N \times N$ Toeplitz matrices $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4$.
Output: A quantum state $|\psi^*\rangle \approx \text{vec}(\mathbf{T}_2\mathbf{T}_1^{-1}\mathbf{T}_3 + \mathbf{T}_4)$.
1. Prepare initial state $|\psi_0\rangle$ which is vectorization of $\mathbf{T}_3$.
/ Compute the Toeplitz matrix inverse $\mathbf{T}_1^{-1}\mathbf{T}_3$ as $|\psi_1\rangle = (\mathbf{I}\otimes\mathbf{T}_1^{-1})|\psi_0\rangle$, assuming oracle access to the generating function f_1 of the Toeplitz matrix $\mathbf{T}_1$ */*
2. Compute $|\psi_0'\rangle = QFT(|\psi_0\rangle)$.
3. Suppose $|\psi_0'\rangle = \sum_{j=0}^{N-1} b_j |j\rangle$. Then, using the oracle for f_1, for each column of $\mathbf{T}_3$, compute in parallel on the parallel quantum processors $\sum_{j=0}^{N-1} b_j |j\rangle |f_1(2\pi j/N)\rangle$.
4. Add a qubit and perform a controlled-rotation on $|f_1(2\pi j/N)\rangle$ to obtain $\sum_{j=0}^{N-1} b_j |j\rangle |f_1(2\pi j/N)\rangle \left(\sqrt{1 - \frac{m^2}{f_1^2(2\pi j/N)}}|0\rangle + \frac{m}{f_1(2\pi j/N)}|1\rangle\right)$, where m is a constant s.t. $m \leq \min_j |\lambda_j|$ and λ_j are the eigenvalues of $\mathbf{C}(f_1)$.
5. Uncompute the second qubit and use amplitude amplification on the last register to obtain $|1\rangle$, and thus with high probability we attain $|\psi_1\rangle = \sqrt{\frac{1}{\sum_j m^2|b_j^2|/|f_1(2\pi j/N)|^2}} \sum_{j=0}^{N-1} b_j \frac{m}{f_1(2\pi j/N)}|j\rangle$, which is proportional to $\Lambda^{-1}|\psi_0'\rangle = \sum_{j=0}^{N-1} \frac{b_j}{f_1(2\pi j/N)}|j\rangle$ up to normalization.
/ Compute the Toeplitz matrix-vector product $|\psi_2\rangle = (\mathbf{I}\otimes\mathbf{T}_2)|\psi_1\rangle$, assuming oracle access to the generating function f_2 of the Toeplitz matrix $\mathbf{T}_2$ */*
6. Suppose $|\psi_1'\rangle = \sum_{j=0}^{N-1} \widetilde{b}_j |j\rangle$. Then, using the oracle for f_2, for each column of $\mathbf{T}_3$, compute in parallel on the quantum processors $|\psi_2'\rangle = \sum_{j=0}^{N-1} \widetilde{b}_j |j\rangle |f_2(2\pi j/N)\rangle$.
7. Compute $|\psi_2\rangle = iQFT(|\psi_2'\rangle)$.
8. Load vectorization of $\mathbf{T}_4$ as a quantum state $|\psi_3\rangle$ s.t. the amplitudes are added to the current state $|\psi_2\rangle$, i.e., prepare $|\psi^*\rangle = |\psi_2\rangle + |\psi_3\rangle$.

global errors of our quantum CR Alg. Q.1 up to iteration n.

THEOREM 2. *Suppose $\|\mathcal{E}_L^{(n)}(z)\|_* \leq \upsilon$ and $\|\widehat{\mathcal{E}}_L^{(n)}(z)\|_* \leq \upsilon$ for some $\upsilon > 0$, and $2(1 + \|\mathbf{V}^{(n)}(1)\|_*) = \gamma_n \leq \gamma$ for some $\gamma > 1$. Then, for the quantum CR Alg. Q.1, we have upper bounds on the error $\|\mathbf{E}^{(n+1)}(z)\|_* \dot{\leq} \upsilon(\gamma^{n+1} - 1)/(\gamma - 1)$ and $\|\widehat{\mathbf{E}}^{(n+1)}(z)\|_* \dot{\leq} \upsilon(\gamma^{n+1} - 1)/(\gamma - 1)$.*

The above error analysis can be provably improved by extending Alg. Q.1 w.r.t. a shifting technique that removes the root $\lambda = 1$ of the function $\chi(z) = z\mathbf{I} - \varphi(z)$. Consider the function $\widetilde{\chi}(z) = \chi(z)(\mathbf{I} - z^{-1}\mathbf{Q})^{-1}$ where $\mathbf{Q} = \mathbf{1}\mathbf{u}^\top$ for any vector $\mathbf{u}$ s.t. $\mathbf{u}^\top\mathbf{1} = 1$ and $\widetilde{\chi}(z)$ has the same roots as $\chi(z)$ except for $z = 1$ replaced by the root $z = 0$. The shifting technique then applies CR to $\widetilde{\chi}(z)$ w.r.t. the corresponding sequences of matrix power series $\widetilde{\chi}^{(n)}(z)$, $\widehat{\widetilde{\chi}}^{(n)}(z)$ in place of $\chi^{(n)}(z)$, $\widehat{\chi}^{(n)}(z)$. Our resulting quantum CR Alg. Q.3 calls Alg. Q.2 exactly as Alg. Q.1 but with inputs $\widetilde{\chi}^{(n)}(z)$ and $\widehat{\widetilde{\chi}}^{(n)}(z)$ [2]. Exploiting Theorem 1, we establish an improved upper bound on the global errors of Alg. Q.3 up to iteration n.

THEOREM 3. *Suppose $\|\mathcal{E}_L^{(n)}(z)\|_* \leq \upsilon$ and $\|\widehat{\mathcal{E}}_L^{(n)}(z)\|_* \leq \upsilon$ for some $\upsilon > 0$. Then, for the quantum CR Alg. Q.3 (see [2]) and for $\gamma_n = (1 + \theta\sigma^{2^n})^2$, we have $\|\mathbf{E}^{(n+1)}(z)\|_* \dot{\leq} \upsilon(1 + ne^{2\theta\sigma^2/(1-\sigma^2)})$ and $\|\widehat{\mathbf{E}}^{(n+1)}(z)\|_* \dot{\leq} \upsilon(1 + ne^{2\theta\sigma^2/(1-\sigma^2)})$.*

Finally, we derive a mathematical analysis of the computational complexity of our quantum CR Algs. Q.1 and Q.3, which includes the time complexity of Alg. Q.2 called by both to handle each CR iteration. Let τ_{oracle} be the time complexity to use the oracle for the generating function f and prepare the state $|f(2\pi j/N)\rangle$, and let $\tau_{readout}$ be the time complexity to read out the results from the quantum computer. Let $\mu = f_{1,\max}/f_{1,\min}$ for the generating function f_1 of Alg. Q.2, let $d_{\max}^Q$ and $d_{\max}^C$ respectively be the maximum numerical degrees of the matrix power series generated by our quantum CR and the corresponding classical CR, and let $N^Q = d_{\max}^Q \cdot M$. We then establish the following result.

THEOREM 4. *The overall computational complexity of the quantum CR Alg. Q.1 and Alg. Q.3 is given by $O(\mu(\tau_{load} + \log^2 N^Q + \tau_{oracle}) + \tau_{readout})$ with $d_{\max}^Q = O(d_{\max}^C)$. Supposing μ, τ_{oracle} to be $O(\text{poly}\log N^Q)$, we then have that the computational complexity $O(\mu(\log^2 N^Q + \tau_{oracle}))$ of Alg. Q.1 and Alg. Q.3 represents an* exponential speedup *of the computation phase over the computational complexity of $O(M^3 d_{\max}^C + M^2 d_{\max}^C \log d_{\max}^C)$ for the corresponding classical CR algorithm. Further supposing $\tau_{load}, \tau_{readout}$ to be $O(\text{poly}\log N^Q)$ or subexponential in $(\log N^Q)$, we then additionally have that the overall computational complexity $O(\mu(\tau_{load} + \log^2 N^Q + \tau_{oracle}) + \tau_{readout})$ of Alg. Q.1 and Alg. Q.3 represents a* polynomial-to-exponential speedup *over the computational complexity for the corresponding classical CR algorithm.*

REMARK 5. *Given the definition of $\mu = f_{1,\max}/f_{1,\min}$ for the generating function f_1 associated with Alg. Q.2 and given the properties of the Toeplitz matrix $\mathbf{T}_1$ (e.g., nonnegative, stochastic) together with typical ratios $f_{1,\max}/f_{1,\min}$ found in practice for M/G/1-type processes, we typically have μ to be $O(\text{poly}\log N^Q)$ for sufficiently large N^Q. In particular, for large N^Q, we have $\mu \approx \kappa$ where κ is the condition number of $\mathbf{T}_1$. By exploiting appropriate oracle quantum circuit methods (see [2]) together with the properties of the Toeplitz matrices $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ input to Alg. Q.2 and the corresponding generating functions f_1, f_2, we can have τ_{oracle} to be $O(\text{poly}\log N^Q)$ for sufficiently large N^Q. In such cases of theoretical and practical importance, Theorem 4 establishes that our quantum algorithms provide the potential for an exponential speedup of the computation phase in the decision-space on a quantum computer over that of the best-known classical algorithms, which is our primary focus from a mathematical perspective.*

Similarly, by exploiting appropriate block encoding quantum circuit methods (see [2]) together with properties of the Toeplitz matrices $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4$ input to Alg. Q.2 to support efficient data loading, we can additionally have τ_{load} to be $O(\text{poly}\log N^Q)$ or subexponential in $(\log N^Q)$ for sufficiently large N^Q; and by exploiting properties of the result matrix $\mathbf{G}_{\min}$ in cases where such matrices are sufficiently sparse to support efficient readout, we can additionally have $\tau_{readout}$ to be $O(\text{poly}\log N^Q)$ or subexponential in $(\log N^Q)$ for sufficiently large N^Q. In such cases of theoretical and practical importance, Theorem 4 establishes that our quantum algorithms provide the potential for an overall polynomial-to-exponential speedup over the best-known classical algorithms.

3. REFERENCES

[1] D. Bini, G. Latouche, B. Meini. *Numerical Methods for Structured Markov Chains*. Oxford University Press, 2005.
[2] V. Kalantzis, M. Squillante, S. Ubaru. On quantum algorithms for efficient solutions of general classes of structured Markov processes. *arXiv preprint arXiv:2404.17959*, 2024.
[3] M. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, 1989.