Quantum Metropolis-Hastings algorithm with the target distribution calculated by quantum Monte Carlo integration

Koichi Miyamoto *

Center for Quantum Information and Quantum Biology, Osaka University, Toyonaka, Osaka 560-0043, Japan

(Received 17 March 2023; accepted 10 July 2023; published 28 July 2023)

The Markov chain Monte Carlo (MCMC) method, especially the Metropolis-Hastings (MH) algorithm, is a widely used technique for sampling from a target probability distribution P on a state space Ω and applied to various problems such as estimation of parameters in statistical models in the Bayesian approach. Quantum algorithms for MCMC have been proposed, yielding the quadratic speedup with respect to the spectral gap Δ compared to classical counterparts. In this paper, we consider the quantum version of the MH algorithm in the case that calculating P is costly because the log-likelihood L for a state $x \in \Omega$ is obtained via computing the sum of many terms, $\frac{1}{M}\sum_{i=0}^{M-1} \ell(i,x)$. We propose calculating L by quantum Monte Carlo integration and combine it with the existing method called quantum simulated annealing (QSA) to generate the quantum state that encodes P in amplitudes. We consider not only state generation but also finding a credible interval for a parameter, a common task in Bayesian inference. In the proposed method for credible interval calculation, the number of queries to the quantum circuit to compute ℓ scales on Δ, the required accuracy ϵ, and the standard deviation σ of ℓ as $\tilde{O}(\sigma/\epsilon^2\Delta^{3/2})$, in contrast to $\tilde{O}(M/\epsilon\Delta^{1/2})$ for QSA with L calculated exactly. Therefore, the proposed method is advantageous if σ scales on M sublinearly. As one such example, we consider parameter estimation in a gravitational wave experiment, where $\sigma = O(M^{1/2})$.

DOI: 10.1103/PhysRevResearch.5.033059

I. INTRODUCTION

A. Background and motivation

Following the recent rapid development of quantum computing, various quantum algorithms are studied extensively, along with their industrial and scientific applications. Among them, quantum algorithms for the Markov Chain Monte Carlo (MCMC) method are one of the prominent ones [1–7]. MCMC is a methodology for sampling from a probability distribution P on a sample space (state space) Ω by generating a Markov chain whose stationary distribution is P (see [8]). It is widely used in various situations, for example, estimation of parameters in statistical models in the Bayesian approach [9].

In spite of its usefulness, MCMC often has an issue of computational time, since many iterations of state generation may be needed for the chain to sufficiently converge to the target distribution P. Classically, the iteration number for sufficient convergence scales as $\tilde{O}(\Delta^{-1})$, where Δ is the spectral gap of the chain (see the definition in Sec. II B).

Quantum MCMC algorithms can be remedies for this: using the quantum walk [10] as a building block, they generates (an approximation of) a quantum state $|P\rangle$ that encodes P in amplitudes with $\tilde{O}(\Delta^{-1/2})$ queries to the walk operator, which indicates the quadratic quantum speedup. More concretely, the approach in Refs. [1–5] called quantum simulated annealing (QSA) takes the following strategy. We consider an initial distribution P_0 and a series of Markov chains with stationary distributions $P_1, \ldots, P_l$ such that these distributions sufficiently *overlap*; that is, the quantum states that encode them satisfy $|\langle P_i|P_{i+1}\rangle|^2 \geqslant \text{const}$. Then, starting from the state $|P_0\rangle$ that encodes P_0, we sequentially generate $|P_1\rangle, \ldots, |P_{l-1}\rangle$, and finally $|P_l\rangle$, which is close to $|P\rangle$. In particular, Reference [5] considered applying this to Bayesian inference, where P is written as $P(x) \propto P_0(x)e^{-L(x)}$ with the negative log-likelihood L, and presented the procedure to generate $|P\rangle$ making $\tilde{O}(\sqrt{\bar{L}/\Delta})$ calls to the quantum walk operator, where $\bar{L}$ is the expectation of L under the distribution P_0.

Among various types of MCMC, the Metropolis-Hastings (MH) algorithm [11,12] is particularly prominent. In this algorithm, it is supposed that the target distribution P is efficiently computable except for the normalization constant and we are given some proposal distribution T for transition among possible states, which is also efficiently computed. Then, accepting the proposed transition with some probability determined by P and T, we generate a chain, which is guaranteed to converge to P. Because of its simplicity, the MH algorithm is widely used. Some of the previous quantum algorithms for MCMC are in fact based on the MH algorithm [4,6,7].

In this paper, we consider a quantum Metropolis-Hastings algorithm in a specific but ubiquitous and important situation. That is, we focus on the case that the target distribution P is computed via summation of many terms. More specifically, we suppose that $P(x) \propto P_0(x)e^{-L(x)}$ and L is written as $L(x) = \frac{1}{M}\sum_{i=0}^{M-1} \ell(i,x)$ with a large integer M and a function ℓ, except for efficiently computable terms (see Sec. III B for the exact problem setting). In the context of Bayesian

*miyamoto.kouichi.qiqb@osaka-u.ac.jp

inference, this corresponds to the situation that the negative log-likelihood is a sum of many terms. In this case, calculation of P can be time consuming, even if ℓ is efficiently computed. Naively, we need to iterate calculations of $\ell(i, x)$ and additions M times.

As an example of such a case, we can take parameter estimation in a gravitational wave (GW) detection experiment (see the review [13] and references therein). In a GW laser interferometer such as LIGO and Virgo [14–17], a GW signal is explored in noisy detector output data by matched filtering [18–21], and, if detected, estimation of parameters in the waveform of the GW is performed in the Bayesian approach. The log-likelihood in this parameter estimation is given as a sum of contributions from various Fourier modes of the data and thus has the aforementioned form. More generally, many statistical inference problems with a large number of independent samples fall into the considered case.

As far as the author knows, previous studies on quantum MCMC algorithms have not focused on the difficulty to compute a target distribution of the above type. References [1–5] assumed the availability of the quantum circuit to generate the quantum state that encodes the transition matrix of the Markov chain. Although Refs. [6,7] broke down the operators needed in the quantum MH algorithm in more detail, they assumed the availability of the quantum circuit to determine acceptance or rejection, and did not consider the detail of target distribution calculation. When it comes to quantum algorithms for data analysis in GW experiments, although there are studies on GW detection [22–25] and the quantum MH algorithm for GW parameter estimation [26], the issue of log-likelihood computation has not been focused on. The focus in Ref. [26] is the comparison between the classical and quantum MH algorithms with respect to the time to solution (TTS), a metric for the time to obtain the maximum likelihood parameter set with high probability, by the numerical simulation, and Ref. [26] did not consider the issue of the log-likelihood computation cost.

B. Our contributions

In this paper, we consider how to speed up calculation of P using another quantum algorithm as a subroutine of the quantum MH algorithm. Concretely, we use quantum Monte Carlo integration (QMCI) [27,28]. Based on quantum amplitude estimation (QAE) [29], QMCI estimates $E[F(X)]$, the expectation of a function F of a random variable X, providing quadratic speedup compared to classical Monte Carlo integration. For example, if we have a bound σ^2 on the variance of $F(X)$, QMCI yields an estimate with accuracy ϵ, calling the quantum circuits to generate a quantum state encoding X's distribution and compute F $\tilde{O}(\sigma/\epsilon)$ times. As a special case, we can use QMCI to estimate the sum of many terms. In fact, QMCI is utilized for calculating the signal-to-noise ratio (SNR) in the quantum algorithm for GW matched filtering proposed in Ref. [23], and using QMCI in the quantum MH algorithm is a similar idea.

We note that QMCI gives an erroneous estimate and thus the Markov chain based on it has a stationary distribution different from the original one P. Fortunately, MCMC with such a perturbation has been studied [30–32], and, according to their results, we can set the accuracy in QMCI to obtain the distribution close to P.

We make a further consideration from a practical perspective. Some previous studies considered only preparing the quantum state $|P\rangle$, but what we want in real life is not the quantum state but the results of some statistical analysis on P as classical data. Then, this paper presents not only how to prepare $|P\rangle$ but also the procedure to obtain the *credible interval* of a parameter θ in a statistical model. It is an interval where θ falls with a prefixed probability in the distribution P and a quantity we often aim to find in Bayesian inference. Given a quantum circuit V_P to prepare $|P\rangle$, we can estimate the cumulative distribution function (CDF) of θ by QMCI using V_P iteratively. We then find the credible interval via binary search on the CDF. We also consider applying this method to credible interval calculation for GW parameters.

C. Main result

Table I is a summary of the complexities in state generation and credible interval calculation, which means the numbers of calls to the oracle to compute ℓ, in various methods for sufficiently small error tolerance ϵ. Here, the exact QSA method is QSA with L calculated exactly by M iterative calculations of ℓ. We see that the complexity of the proposed method is equal to that of the exact QSA method with the factor M replaced with $\sigma/\Delta_{\min}\epsilon$, where σ^2 is the variance of ℓ, ϵ is the required accuracy, and $\Delta_{\min}$ is the lower bound of spectral gaps of Markov chains related to the considered problem. This is because, following the result of Ref. [30], we estimate L by QMCI with accuracy $\Delta_{\min}\epsilon$ in order to reach a distribution close to P with accuracy ϵ, and thus its complexity becomes $\tilde{O}(\sigma/\Delta_{\min}\epsilon)$. As a result, with respect to ϵ and $\Delta_{\min}$, the complexity of the proposed method is worse than the exact QSA method and even the classical MH algorithm. Nevertheless, it may be advantageous with respect to M, if σ scales on M sublinearly. In fact, in the case of GW parameter estimation, σ can be $O(\sqrt{M})$, which means that the proposed method is quadratically faster than the exact QSA method and the classical MH algorithm with respect to M.

D. Organization

Section II is preliminary one, where we briefly explain the MH algorithm, QSA, and QMCI. In Sec. III, we present our methods for generating the state $|P\rangle$ and finding the credible interval in details. In Sec. IV, we consider the application to GW parameter estimation. Section V summarizes this paper.

II. PRELIMINARY

A. Notation

Here, we summarize some notations used in this paper.

$\mathbb{R}_+$ denotes the set of all positive real numbers. For $n \in \mathbb{N}$, we define $[n] := \{1, \ldots, n\}$ and $[n]_0 := \{0, 1, \ldots, n-1\}$.

We hereafter consider systems consisting of quantum registers (or simply registers), sets of single or multiple qubits. A ket $|\psi\rangle$ denotes a state vector of a quantum state on a register, and we sometimes put a subscript to clarify the register on which the state is generated: $|\psi\rangle_R$ is a state

TABLE I. The complexities (number of queries to the oracle to compute ℓ) in various tasks in various methods for sufficiently small ϵ [precisely speaking, ϵ satisfying Eq. (78)]. Here, ϵ represents the total variation distance between the generated state and $|P\rangle$ and the error in the cumulative distribution function for state generation and credible interval calculation, respectively (see Secs. III C and III D for details). $\Delta_{\min}$ is a lower bound of spectral gaps of some Markov chains (see Sec. III C for details). ρ is the SNR in GW matched filtering (see Sec. IV for details).

Task	Proposed method	Exact QSA	Classical MH
Generate $\|P\rangle$	$\tilde{O}\left(\frac{\sigma \bar{L}^{1/2}}{\Delta_{\min}^{3/2}\epsilon}\right)$	$\tilde{O}\left(\frac{M\bar{L}^{1/2}}{\Delta_{\min}^{1/2}}\right)$	Not applicable
Credible interval (general)	$\tilde{O}\left(\frac{\sigma \bar{L}^{1/2}}{\Delta_{\min}^{3/2}\epsilon^2}\right)$	$\tilde{O}\left(\frac{M\bar{L}^{1/2}}{\Delta_{\min}^{1/2}\epsilon}\right)$	$\tilde{O}\left(\frac{M}{\Delta\epsilon^2}\right)$
Credible interval (GW)	$\tilde{O}\left(\frac{\rho M^{1/2}\bar{L}^{1/2}}{\Delta_{\min}^{3/2}\epsilon^2}\right)$	$\tilde{O}\left(\frac{M\bar{L}^{1/2}}{\Delta_{\min}^{1/2}\epsilon}\right)$	$\tilde{O}\left(\frac{M}{\Delta\epsilon^2}\right)$

on a register R. Similarly, we sometimes put a subscript to a symbol representing an operator to indicate the register on which the operator acts. In particular, I_R denotes the identity operator on a register R.

For $x \in \mathbb{R}$, $|x\rangle$ denotes the computational basis state on a register whose bit string corresponds to a finite-precision binary representation of x. This representation includes the sign of x in some way (e.g., the two's complement method [33]). We assume that any number considered in this paper is represented with a sufficiently large number of qubits and thus neglect rounding errors. For a real vector $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $|x\rangle$ denotes a computational basis state on d-register system $|x\rangle = |x_1\rangle \cdots |x_d\rangle$.

For a vector $x = (x_1, \ldots, x_d) \in \mathbb{C}^d$, we define its k-norm as $\|x\|_k := (\sum_{i=1}^d |x_i|^k)^{1/k}$ with $k \in \mathbb{N}$ and max norm as $\|x\|_\infty := \max\{|x_1|, \ldots, |x_d|\}$. For a matrix $A \in \mathbb{C}^{m\times n}$, we define $\|A\|_k := \sup_{x\in\mathbb{C}^n, \|x\|_k=1} \|Ax\|_k$ with $k \in \mathbb{N} \cup \{\infty\}$, and denote by $\|A\|_F$ its Frobenius norm. We simply write $\|\cdot\|_2$ as $\|\cdot\|$. $\||\psi\rangle\|$ is a 2-norm of the (unnormalized) state vector $|\psi\rangle$.

If $x, y \in \mathbb{C}^d$ satisfy $\|x - y\| \leqslant \epsilon$ with some $\epsilon \in \mathbb{R}_+$, we say that x is ϵ-close to y and x is an ϵ approximation of y. If quantum states $|\psi\rangle$ and $|\phi\rangle$ on a same register satisfy $\||\psi\rangle - |\phi\rangle\| \leqslant \epsilon$ with some $\epsilon \in \mathbb{R}_+$, we say that $|\psi\rangle$ is ϵ-close to $|\phi\rangle$ and $|\psi\rangle$ is an ϵ approximation of $|\phi\rangle$.

For a nonsingular matrix A, we define its condition number as $\|A\| \cdot \|A^{-1}\|$.

Letting $(\Omega, 2^\Omega, P)$ be a probability space with a finite sample space Ω, we write $P(x) = P(\{x\})$ for $x \in \Omega$ and also call the measure P the probability distribution or distribution. We denote by $\mathbb{E}_P[\cdot]$ the expectation with respect to P.

The indicator function $\mathbf{1}_C$ takes 1 if the condition C is satisfied and 0 otherwise.

B. Metropolis-Hastings algorithm

1. Outline

We briefly summarize the MH algorithm [11,12], whose aim is sampling a random variable X that obeys some target probability distribution. For more details, see Ref. [8].

Every value X can take is called a *state*, and the set of the states is called the *state space* and hereafter denoted by Ω. In this paper, we consider the situation that Ω is a finite subset in $\mathbb{R}^d$, where $d \in \mathbb{N}$. This is because a quantum computer can only represent real numbers in finite precision using a finite number of qubits, which is the case also on a classical computer. Of course, X can take continuous values in many situations, but we assume that continuous X is well approximated in a discrete manner and errors from this are negligible, as stated in Sec. II A.

For every $x \in \Omega$, we denote by $P(x) \in (0, 1)$ the probability that X takes x in the target distribution. We assume that $P(x)$ can be written as $P(x) = p(x)/Z$, where $p(x)$ is an easily computable function and $Z := \sum_{x\in\Omega} p(x)$ is the normalization factor. Although Z is often hard to be computed, the MH algorithm works even if we do not know Z, as explained later.

In the MH algorithm, starting from some initial state x_0, we sequentially get states by making transitions over Ω as follows. For every $x \in \Omega$, we set some easy-to-sample proposal distribution $T(x, \cdot) : \Omega \to (0, 1)$, for example, the normal distribution centered at x (strictly, its discrete approximation). Letting $x_i \in \Omega$ be the ith state, we randomly choose $\tilde{x}_{i+1} \in \Omega$ with probability $T(x_i, \tilde{x}_{i+1})$ as a candidate for the next state. Then, calculating the acceptance ratio $A(x_i, \tilde{x}_{i+1})$, which is defined for $x, y \in \Omega$ as

$$A(x, y) := \min\left\{1, \frac{P(y)T(y, x)}{P(x)T(x, y)}\right\}, \tag{1}$$

we set the next state x_{i+1} to $\tilde{x}_{i+1}$ with probability $A(x_i, \tilde{x}_{i+1})$ or stay at x_i otherwise. Note that the target distribution P appears in A in the form of the ratio $P(y)/P(x)$, which means that we need to compute only $p(x)$, not Z.

As a consequence, the sequence generated by the MH algorithm is a Markov chain with a following transition matrix $W = (W_{x,y})$: it is indexed by $x, y \in \Omega$ and its (x, y) entry, which corresponds to the probability that the transition to y occurs provided that the current state is x, is

$$W_{x,y} = \begin{cases} T(x, y)A(x, y) & \text{if } x \neq y \\ 1 - \sum_{z\in\Omega\setminus\{x\}} T(x, z)A(x, z) & \text{if } x = y. \end{cases} \tag{2}$$

The convergence property of this Markov chain is affected by the *spectral gap* Δ. It is defined as

$$\Delta := 1 - |\lambda_1|, \tag{3}$$

where λ_1 is the eigenvalue of W with the second largest modulus. It is known that the eigenvalue of W with the largest modulus is 1 and it is nondegenerate (Lemmas 12.1 and 12.2 of Ref. [8]), which means $|\lambda_1| < 1$ and $0 < \Delta < 1$. To present the formal statement on the convergence rate, we introduce the total variance distance, a metric to measure the difference between two probability distributions.

Definition 1. For probability measures P and Q on a measurable space $(\Omega, \mathcal{F})$, the total variance distance is

Algorithm 1. Metropolis-Hastings algorithm.

Require:
- Initial probability distribution P_0 on Ω
- Burn-in length $n_\mathrm{b} \in \mathbb{N}$
- Number n of sample values of X we need
- Function to sample from the proposal distribution T
- Function to compute T
- Function to compute the target probability P except for the normalization constant

1: Sample the initial point $x_0 \in \Omega$ from P_0.
2: **for** $i = 0, \ldots, n_\mathrm{b} + n - 1$ **do**
3: Sample $\tilde{x}_{i+1} \in \Omega$ from the distribution $T(x_i, \cdot)$.
4: Compute $A(x_i, \tilde{x}_{i+1})$ in Eq. (1).
5: Set $x_{i+1} = \tilde{x}_{i+1}$ with probability $A(x_i, \tilde{x}_{i+1})$ or $x_{i+1} = x_i$ otherwise.
6: **end for**
7: Output $x_{n_b+1}, \ldots, x_{n_b+n}$.

defined as

$$\|P - Q\|_\mathrm{TV} := \sup_{A \in \mathcal{F}} |P(A) - Q(A)|. \tag{4}$$

Since we are now considering finite Ω, Eq. (4) becomes $\|P - Q\|_\mathrm{TV} = \max_{A \in 2^\Omega} |P(A) - Q(A)|$.

Then, we have the following theorem.

Theorem 1 (*Theorem 12.4 in Ref. [8]*). Let W be the transition matrix of a reversible irreducible Markov chain with a finite state space Ω and a stationary distribution Π. Let Δ be the spectral gap of the chain. Let $\mathcal{P}$ be a set of all probability distributions on Ω and, for $n \in \mathbb{N}$ and $\mu \in \mathcal{P}$, denote by μW^n the probability distribution after n steps of the Markov chain with the initial distribution μ. Then, for any $n \in \mathbb{N}$,

$$d(n) \leqslant \frac{(1-\Delta)^n}{2\sqrt{\Pi_\mathrm{min}}}, \tag{5}$$

where $d(n) := \sup_{\mu \in \mathcal{P}} \|\mu W^n - \Pi\|_\mathrm{TV}$ and $\Pi_\mathrm{min} := \min_{x \in \Omega} \Pi(x)$, and thus, for any $\epsilon \in \mathbb{R}_+$,

$$t_\mathrm{mix}(\epsilon) := \min\{n \in \mathbb{N} \mid d(n) \leqslant \epsilon\} \leqslant \frac{1}{\Delta} \log\left(\frac{1}{\epsilon \Pi_\mathrm{min}}\right). \tag{6}$$

Since the early part of the Markov chain has not converged yet, we usually discard it, which is called *burn-in*. Then, the procedure of the MH algorithm is summarized in Algorithm 1.

The obtained sequence can be used for, e.g., calculating expectations of random variables. On the error in this, we have the following theorem.

Theorem 2 (*Theorem 11 in Ref. [34]*). Consider the Markov chain on a finite sample space Ω generated with a transition matrix W and initial distribution P_0. Denote its stationary distribution and spectral gap by Π and Δ, respectively. Let the second largest eigenvalue of W be λ_1' and let $\Delta' := 1 - \lambda_1'$. Then, for any function $f : \Omega \to \mathbb{R}$ and $n_\mathrm{b}, n \in \mathbb{N}$,

$$e_{n,n_\mathrm{b},f}^2 := \mathbb{E}_\mathrm{MC}[(S_{n,n_\mathrm{b},f} - \mathbb{E}_\Pi[f(x)])^2] \leqslant \frac{2\|f\|_\infty^2}{n\Delta'} + \frac{4\left\|\frac{P_0}{\Pi} - 1\right\|_\infty^{1/2} \|f\|_\infty^2 (1-\Delta)^{n_\mathrm{b}}}{n^2\Delta^2} \tag{7}$$

holds. Here,

$$S_{n,n_\mathrm{b},f} := \frac{1}{n} \sum_{i=1}^{n} f\left(x_{n_\mathrm{b}+i}\right), \tag{8}$$

x_i, $i \in [n_\mathrm{b} + n]$ is the ith entry in the chain, $\mathbb{E}_\mathrm{MC}[\cdot]$ denotes the expectation with respect to the randomness of the generated chain, $\|f\|_\infty := \max_{x \in \Omega} |f(x)|$, and $\|\frac{P_0}{\Pi} - 1\|_\infty := \max_{x \in \Omega} |\frac{P_0(x)}{\Pi(x)} - 1|$.

$S_{n,n_\mathrm{b},f}$ is an estimate of $\mathbb{E}_\Pi[f(x)]$, the expectation of f based on n samples from the Markov chain with n_b burn-in samples discarded. Then the theorem implies that, to suppress the root mean square error $e_{n,n_\mathrm{b},f}$ of $S_{n,n_\mathrm{b},f}$ to $\epsilon \in (0, \|f\|_\infty)$, it is sufficient to take the burn-in length

$$n_\mathrm{b} = O\left(\frac{\log\left(\left\|\frac{P_0}{\Pi} - 1\right\|_\infty^{1/2}\right)}{\Delta}\right) \tag{9}$$

and the sample number

$$n = O\left(\max\left\{\frac{\|f\|_\infty^2}{\Delta'\epsilon^2}, \frac{\|f\|_\infty}{\Delta\epsilon}\right\}\right). \tag{10}$$

Since $\Delta_1 \geqslant \Delta$ holds, n is also upper bounded as

$$n = O\left(\frac{\|f\|_\infty^2}{\Delta\epsilon^2}\right). \tag{11}$$

If $\Delta' = \Delta$, which holds when λ_1' is the second largest eigenvalue also in modulus, that is, $\lambda_1 = \lambda_1'$, bound (10) becomes Eq. (11). Including the burn-in, the total step number is also of order (11).

2. *Perturbed acceptance ratio*

We consider the case where we can compute not the exact acceptance ratio but some approximation of it. We expect that, in such a situation, although the chain converges to some distribution different from the original target distribution, the difference is small as far as the error in the acceptance ratio is small. The bound on such a difference has been studied in previous studies [30–32], and, in this paper, we use the result of Ref. [30]. As a preparation to present it, we introduce and a concept called uniform ergodicity [35].

Definition 2. If, for a Markov chain on a finite state space, there exist $\rho \in (0, 1)$ and $C \in \mathbb{R}$ such that

$$\|\Pi_n - \Pi\| \leqslant C\rho^n \tag{12}$$

holds for any initial distribution and $n \in \mathbb{N}$, where Π_n is the distribution after n steps and Π is the stationary distribution, we say that the Markov chain is (C, ρ)-uniformly ergodic.

Apparently, for a reversible irreducible Markov chain, an upper bound of ρ is $1 - \Delta$.

Then, the theorem we will later use is the following. This is the restriction of Corollary 2.3 in Ref. [30] to the case that we can compute an approximate acceptance ratio that is deterministic and has a bounded error.

Theorem 3. Consider Algorithm 1. Suppose that the generated Markov chain $\mathcal{C}$ is (C, ρ)-uniformly ergodic. Also consider another chain $\mathcal{C}'$ the same as $\mathcal{C}$ except that the acceptance ratio A is replaced with $A' : \Omega \times \Omega \to [0, 1]$ such

that

$$\max_{x,y\in\Omega} |A'(x,y) - A(x,y)| \leqslant \epsilon \tag{13}$$

with some $\epsilon \in \mathbb{R}$. Then, the stationary distributions Π and Π' of $\mathcal{C}$ and $\mathcal{C}'$ satisfy

$$\|\Pi - \Pi'\|_{\mathrm{TV}} \leqslant \epsilon\left(\lambda + \frac{C\rho^{\lambda}}{1-\rho}\right) \tag{14}$$

with $\lambda := \lceil \frac{\log(1/C)}{\log\rho} \rceil$.

Because of Theorem 1, for a reversible irreducible Markov chain, we obtain

$$\|\Pi - \Pi'\|_{\mathrm{TV}} \leqslant \epsilon\left(\left\lceil \frac{\log(2\sqrt{\Pi_{\min}})}{\log(1-\Delta)} \right\rceil + \frac{1}{\Delta}\right), \tag{15}$$

by substituting $\rho = 1 - \Delta$ and $C = \frac{1}{2\sqrt{\Pi_{\min}}}$ in Eq. (14).

C. Quantum walk operator

Among several versions of quantum walk operators proposed so far, this paper adopts that in Ref. [6]. It is dedicated for the MH algorithm for Ising models, and will be generalized in Sec. III A. Suppose that we are now considering an Ising system with n_{sp} spins, which means that $\Omega = \{-1,1\}^{\times n_{\mathrm{sp}}}$. We use a system of two quantum registers R_{S} and R_{M} and a qubit R_{C}. R_{S} holds a string x of ± 1 with length n_{sp}, that represents the current spin configuration. R_{M} holds a bit string $z \in \{0,1\}^{\times n_{\mathrm{sp}}}$ that represents the next spin flip: if the ith bit z_i of z is 1 (respectively 0), the ith spin x_i in $x \in \{-1,1\}^{\times n_{\mathrm{sp}}}$ is changed to $-x_i$ (respectively unchanged). We denote by $x \odot z$ the spin configuration generated by x and z under this rule. Then, we consider the following operator:

$$U_{\mathrm{IS}} = R_{\mathrm{IS}} V_{\mathrm{IS}}^{\dagger} B_{\mathrm{IS}}^{\dagger} F_{\mathrm{IS}} B_{\mathrm{IS}} V_{\mathrm{IS}}. \tag{16}$$

The component operators are as follow. V_{IS} is a unitary on R_{M} that acts as

$$V_{\mathrm{IS}}|0\rangle_{R_{\mathrm{M}}} = \sum_{z\in\mathcal{M}} \sqrt{p_{\mathrm{fl}}(z)}|z\rangle_{R_{\mathrm{M}}}. \tag{17}$$

Here, $\mathcal{M} \subset \{0,1\}^{\times n_{\mathrm{sp}}}$ is the set of possible spin flips and $p_{\mathrm{fl}} : \mathcal{M} \to (0,1)$ is a probability distribution on $\mathcal{M}$, which we use as the proposal distribution of the next flip. Note that it is assumed that possible flips $\mathcal{M}$ and the probability $p_{\mathrm{fl}}(z)$ that a flip $z \in \mathcal{M}$ is proposed are independent of the current spin configuration. We can associate p_{fl} with the proposal distribution T in Algorithm 1 as

$$T(x,y) = \begin{cases} p_{\mathrm{fl}}(z) & \text{if } y = x \odot z \text{ with some } z \in \mathcal{M} \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where $x, y \in \{0,1\}^{\times n_{\mathrm{sp}}}$. B_{IS} is the rotation gate on R_{C} controlled by R_{S} and R_{M}, whose rotation angle is determined by the acceptance ratio. That is, it acts as[1]

$$\begin{aligned} &B_{\mathrm{IS}}|x\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} \\ &\quad = |x\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}} \\ &\qquad \otimes \begin{pmatrix} \sqrt{1-A(x,x\odot z)} & -\sqrt{A(x,x\odot z)} \\ \sqrt{A(x,x\odot z)} & \sqrt{1-A(x,x\odot z)} \end{pmatrix}|\phi\rangle_{R_{\mathrm{C}}} \end{aligned} \tag{19}$$

for any $x \in \Omega$, $z \in \mathcal{M}$ and state $|\phi\rangle_{R_{\mathrm{C}}}$ on R_{C}, where $A(x, x\odot z) = P(x\odot z)T(x\odot z, x)/P(x)T(x, x\odot z)$ with some target distribution P on $x \in \{-1,1\}^{\times n_{\mathrm{sp}}}$. F_{IS} is a gate to apply the spin flip under the control by R_{C}, which acts as

$$\begin{aligned} &F_{\mathrm{IS}}|x\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} \\ &\quad = \begin{cases} |x\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle_{R_{\mathrm{C}}} = |0\rangle_{R_{\mathrm{C}}} \\ |x\odot z\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}}|1\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle_{R_{\mathrm{C}}} = |1\rangle_{R_{\mathrm{C}}}. \end{cases} \end{aligned} \tag{20}$$

Finally,

$$R_{\mathrm{IS}} = 2\Lambda_{0,\mathrm{IS}} - I_{R_{\mathrm{S}}} \otimes I_{R_{\mathrm{M}}} \otimes I_{R_{\mathrm{C}}}, \tag{21}$$

where

$$\Lambda_{0,\mathrm{IS}} := I_{R_{\mathrm{S}}} \otimes |0\rangle\langle 0|_{R_{\mathrm{M}}} \otimes |0\rangle\langle 0|_{R_{\mathrm{C}}}. \tag{22}$$

Although this is different from quantum walk operators in previous studies [1–5,10,36], it has the following property on its spectrum, which is same as previous ones, and thus can be used as an alternative.

Theorem 4. Consider the Markov chain generated by Algorithm 1 with the state space $\Omega = \{-1,1\}^{\times n_{\mathrm{sp}}}$, the target distribution P, and the proposal distribution T in Eq. (18). Denote by Δ its spectral gap. Let $\mathcal{A} = \mathrm{span}\{|x\rangle_{R_{\mathrm{S}}}|0\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}}|x \in \Omega\}$ and $\mathcal{B} = V_{\mathrm{IS}}^{\dagger}B_{\mathrm{IS}}^{\dagger}F_{\mathrm{IS}}B_{\mathrm{IS}}V_{\mathrm{IS}}\mathcal{A}$. Then, on $\mathcal{A}+\mathcal{B}$, $|P\rangle$ is the unique eigenstate of U_{IS} with eigenvalue 1, and any other eigenvalue is written as $e^{i\theta}$ with $\theta \in \mathbb{R}$ such that $|\theta| \geqslant \arccos(1-\Delta)$.

We call operators that have this property the quantum walk operators for the Markov chain. Although this theorem only considers the phase gap of the quantum walk operator unlike the previous results such as Theorem 1 in Ref. [10] and Theorem 1 in Ref. [3], which consider the entire spectrum in more detail, it is sufficient for our purpose for the reason explained in Sec. II D.

D. Quantum simulated annealing

In previous studies on quantum versions of MCMC [1–5], the aim is generating a quantum state in which the target distribution P is encoded in the amplitude, that is,

$$|P\rangle := \sum_{x\in\Omega} \sqrt{P(x)}|x\rangle. \tag{23}$$

In this paper, we use the method proposed in Ref. [5]. In that paper, following Ref. [3], the author took the strategy called

[1]Note that this B_{IS} is equivalent to B in Ref. [6]. For the latter, only the action on states in the form of $|x\rangle_{R_{\mathrm{S}}}|z\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}}$ is shown, but Eq. (19) shows the operator as an entire matrix. In this definition of B_{IS}, there is an ambiguity of an overall constant factor in the second column in the matrix acting on $|\phi\rangle_{R_{\mathrm{C}}}$, and it does not affect the discussions in this paper.

QSA, which was inspired by simulated annealing. That is, assuming that P is in the form of

$$P(x) \propto P_0(x)\exp(-L(x)) \tag{24}$$

with the *prior distribution* P_0 and the *negative log-likelihood* $L(x)$, we sequentially prepare the quantum states $|P_{\beta_1}\rangle, \ldots, |P_{\beta_l}\rangle$ from the initial state $|P_0\rangle$. Here, these states encode the distributions in the form of

$$P_\beta(x) \propto P_0(x)\exp(-\beta L(x)) \tag{25}$$

with parameters called temperatures $\beta_0 = 0 < \beta_1 < \cdots < \beta_l = 1$, and largely overlap: $|\langle P_{\beta_i}|P_{\beta_{i+1}}\rangle|^2 \geqslant p$ with $p = \Theta(1)$. The method in Ref. [5] is twofold: obtain a set of appropriate values of $\{\beta_i\}$ and transform $|P_0\rangle$ to $|P_{\beta_l}\rangle$ via $|P_{\beta_1}\rangle, \ldots, |P_{\beta_{l-1}}\rangle$. Both phases are based on the following result, how to approximately construct the phase gate, which multiplies a phase factor ω to the state vector for the state $|P_{\beta_i}\rangle$ but not for orthogonal states, using the quantum walk operator for the Markov chain converging to P_{β_i}. This is summarized as the following theorem.

Theorem 5 (Corollary 2 in Ref. [3]). Consider a Markov chain $\mathcal{C}$ on a finite state space Ω with the transition matrix W, the stationary distribution P, and the spectral gap Δ. Let ω be a complex number with unit modulus and define a unitary

$$R^{\omega}_{|P\rangle} := \omega\Lambda^{\parallel}_{|P\rangle} + \Lambda^{\perp}_{|P\rangle}, \tag{26}$$

where $\Lambda^{\parallel}_{|P\rangle}$ and $\Lambda^{\perp}_{|P\rangle}$ are the projection onto the subspace spanned by a state $|P\rangle$ on a register R and that onto the orthogonal subspace, respectively, Then, for any $\delta \in (0, 1)$, we have access to a unitary operator $\tilde{R}^{\omega}_{|P\rangle,\delta}$ that has the following properties:

(i) It acts on a system of R and n_{anc} ancillary qubits, where $n_{\mathrm{anc}} = O(\log(\frac{1}{\Delta})\log(\frac{1}{\delta}))$.

(ii) It uses the controlled version of the quantum walk operator U for $\mathcal{C}$ and its inverse $O(\frac{\log(1/\delta)}{\sqrt{\Delta}})$ times.

(iii) For any state $|\Xi\rangle$ on R, $\tilde{R}^{\omega}_{|P\rangle,\delta}|\Xi\rangle|0\rangle^{\otimes n_{\mathrm{anc}}} = (R^{\omega}_{|P\rangle}|\Xi\rangle)|0\rangle^{\otimes n_{\mathrm{anc}}} + |\xi\rangle$, where $|\xi\rangle$ is an unnormalized state on the entire system with $\||\xi\rangle\| \leqslant \delta$.

Note that, as special cases, this theorem covers the original version of Corollary 2 in Ref. [3] with $\omega = \omega_{\pi/3} := e^{i\frac{\pi}{3}}$, and Theorem 6 in Ref. [36] with $\omega = -1$, which corresponds to the reflection operator with respect to $|P\rangle$.

The outline of constructing $R^{\omega}_{|P\rangle}$ with U is as follows. Given $|\psi_i\rangle$ an eigenstate of U with the corresponding eigenvalue $\lambda_i = e^{i\theta_i}$, we can compute an estimate of the phase θ_i onto another register R_{ph} by quantum phase estimation (QPE) [37,38]. Because of Theorem 4, the difference between the phase of the eigenstate $|P\rangle$, which is zero, and that of any other eigenstate is larger than $\arccos(1-\Delta) = \Omega(\sqrt{\Delta})$, and thus $|P\rangle$ can be distinguished from other eigenstates via QPE with accuracy $O(\sqrt{\Delta})$. Thus, the above QPE followed by a phase gate controlled by the register R_{ph} is the phase gate that acts only on $|P\rangle$, that is, $R^{\omega}_{|P\rangle}$. Because QPE outputs an estimate within the desired accuracy not certainly but with a finite failure probability, this implementation of $R^{\omega}_{|P\rangle}$ gives an approximate gate $\tilde{R}^{\omega}_{|P\rangle,\delta}$.

Given the phase gate with $\omega = \omega_{\pi/3}$, we can generate $|P_{\beta_{i+1}}\rangle$ from $|P_{\beta_i}\rangle$ by Grover's $\frac{\pi}{3}$-amplitude amplification [39], which is summarized as follows.

Theorem 6. Let $|\phi_1\rangle$ and $|\phi_2\rangle$ be quantum states on the same register satisfying $|\langle\phi_1|\phi_2\rangle|^2 \geqslant p$ with some $p \in (0, 1]$. For $i \in \{1, 2\}$, denote by $\Lambda^{\parallel}_{|\phi_i\rangle}$ and $\Lambda^{\perp}_{|\phi_i\rangle}$ the projection on the subspace spanned by $|\phi_i\rangle$ and that on the orthogonal subspace, respectively, and define the unitary $R^{\omega_{\pi/3}}_{|\phi_i\rangle} := \omega_{\pi/3}\Lambda^{\parallel}_{|\phi_i\rangle} + \Lambda^{\perp}_{|\phi_i\rangle}$. Define the unitaries $U_{i,m}$ recursively as follows:

$$\begin{aligned} U_{i,0} &= I, \\ U_{i,m+1} &= U_{i,m}R^{\omega_{\pi/3}}_{|\phi_i\rangle}U^{\dagger}_{i,m}R^{\omega_{\pi/3}}_{|\phi_{i+1}\rangle}U_{i,m}. \end{aligned} \tag{27}$$

Then, for any $m \in \mathbb{N}$,

$$|\langle\phi_2|U_{i,m}|\phi_1\rangle|^2 \geqslant 1 - (1-p)^{3^m} \tag{28}$$

holds. This implies that, for any $\epsilon \in \mathbb{R}_+$, we can prepare $|\widetilde{\phi_2}\rangle$ ϵ-close to $|\phi_2\rangle$ from $|\phi_1\rangle$ using the unitaries in $\{R^{\omega_{\pi/3}}_{|\phi_1\rangle}, R^{\omega_{\pi/3}}_{|\phi_2\rangle}, (R^{\omega_{\pi/3}}_{|\phi_1\rangle})^{\dagger}, (R^{\omega_{\pi/3}}_{|\phi_2\rangle})^{\dagger}\}$ $O(\frac{\log(\frac{1}{\epsilon})}{\log\frac{1}{1-p}})$ times.

Strictly speaking, we do not have the exact operator $R^{\omega_{\pi/3}}_{|P\rangle}$ but an approximate one $\tilde{R}^{\omega_{\pi/3}}_{|P\rangle,\epsilon}$. According to Ref. [3], we can transform $|P_0\rangle$ to a state $|\widetilde{P_1}\rangle$ ϵ-close to $|P_1\rangle$, via $|\widetilde{P_{\beta_1}}\rangle, \ldots, |\widetilde{P_{\beta_{l-1}}}\rangle$ the approximate states of $|P_{\beta_1}\rangle, \ldots, |P_{\beta_{l-1}}\rangle$, by $\frac{\pi}{3}$-amplitude amplifications with $\tilde{R}^{\omega_{\pi/3}}_{|P_{\beta_i}\rangle,\epsilon'}$ used instead of $R^{\omega_{\pi/3}}_{|P_{\beta_i}\rangle}$, where ϵ' is some real number set depending on l, ϵ, and p. We hereafter call this method the approximate $\frac{\pi}{3}$-amplitude amplification (A$\frac{\pi}{3}$AA) with accuracy ϵ and overlap p. Its complexity is summarized as follows.

Theorem 7 (Theorem 5 in Ref. [5], originally Theorem 2 in Ref. [3]). Consider l Markov chains $\mathcal{C}_1, \ldots, \mathcal{C}_l$ on a finite state space Ω with stationary distributions $p_1, \ldots, p_l$ and spectral gaps lower bounded by $\Delta \in (0, 1)$. Let p_0 be another probability distribution on Ω and suppose that the state $|p_0\rangle$ is given on a register R. Assume that, for some $p \in (0, 1)$, $|\langle p_i|p_{i+1}\rangle|^2 \geqslant p$ holds for any $i \in [l]_0$. Then, for any $\epsilon \in (0, 1)$, we have an access to a unitary operator U_{QSA} on the system of R and n_{anc} qubits that acts as

$$U_{\mathrm{QSA}}|0\rangle_R|0\rangle^{\otimes n_{\mathrm{anc}}} = |p_l\rangle_R|0\rangle^{\otimes n_{\mathrm{anc}}} + |\zeta\rangle, \tag{29}$$

making $O(\frac{l}{\sqrt{\Delta}}\log^2\frac{l}{p\epsilon}\log\frac{1}{p})$ queries to the controlled quantum walk operators for $\mathcal{C}_1, \ldots, \mathcal{C}_l$. Here,

$$n_{\mathrm{anc}} = O\left(\log\left(\frac{1}{\Delta}\right)\log\left(\frac{l\log\left(\frac{l}{\epsilon}\right)}{\log\left(\frac{1}{1-p}\right)}\right)\right), \tag{30}$$

and $|\zeta\rangle$ is an (unnormalized) state on the entire system with $\||\zeta\rangle\| \leqslant \epsilon$.

Besides, given the phase gate with $\omega = -1$, which is, namely, the reflection operator, we can use nondestructive amplitude estimation (NAE) [5], a modification of QAE [29], to estimate $|\langle P_{\beta_i}|P_{\beta_{i+1}}\rangle|^2$.

Theorem 8 (*Theorem 6 in Ref. [5]*). Given a quantum state $|\phi\rangle$ on a register R and two operators $R_\phi = 2|\phi\rangle\langle\phi| - I$ and $R_{\phi'} = 2|\phi'\rangle\langle\phi'| - I$, where $|\phi'\rangle$ is another state on R, for any $\epsilon, \delta \in (0, 1)$, there exists a quantum algorithm with the following properties:

(i) With probability at least $1-\delta$, it outputs an ϵ approximation of $|\langle\phi'|\phi\rangle|^2$ and a flag 1, and restores the state $|\phi\rangle$.

(ii) Otherwise, output a flag 0.

(iii) It uses R_ϕ and $R_{\phi'}$ $O(\frac{\log(1/\delta)}{\epsilon})$ times.

Algorithm 2. Quantum simulated annealing (Algorithm 1 in Ref. [5], modified).

Require:
- Access to a unitary operator O_{P_0} to generate $|P_0\rangle$:
$$O_{P_0}|0\rangle = |P_0\rangle. \tag{31}$$
- For any $\beta \in (0, 1]$, an access to a quantum walk operator U_β for a Markov chain that has P_β in Eq. (25) as the stationary distribution and the spectral gap lower bounded $\Delta_{\min}$.
- Accuracy $\epsilon \in \mathbb{R}_+$ for the final state.
- Failure probability $\eta \in (0, 1)$.

Output: Either of
(A) sequence $\beta_0 = 0 < \beta_1 < \cdots < \beta_{l-1} < \beta_l = 1$ such that $l \leqslant l_{\max}$ and $|\langle P_{\beta_i}|P_{\beta_{i+1}}\rangle|^2 \geqslant \frac{9}{10e^2}$ for any $i \in [l]_0$, and flg $= 1$
(B) flg $= 0$

1: Set $|\widetilde{P_0}\rangle = |P_0\rangle$.
2: **for** $i = 0, 1, \ldots, l_{\max} - 1$ **do**
3: Find the largest $\beta' \in (\beta_i, 1]$ such that $|\langle P_{\beta_i}|P_{\beta'}\rangle|^2 \geqslant e^{-2}$ by binary search with precision $1/L_{\max}$. Here, $|\langle P_{\beta_i}|P_{\beta'}\rangle|^2$ is computed by ANAE with accuracy $1/10e^2$ and failure probability $\eta/l_{\max}L_{\max}$, with $|\widetilde{P_{\beta_i}}\rangle$ used instead of $|P_{\beta_i}\rangle$.
4: **if** at least one ANAE in line 3 returns a flag 0 **then**
5: Output flg $= 0$ and stop.
6: **end if**
7: Let the result in line 3 be β_{i+1}.
8: **if** $\beta_{i+1} = 1$ **then**
9: Output $\beta_0 = 0, \beta_1, \ldots, \beta_{i+1}$ and flg $= 1$, and stop.
10: **end if**
11: Generate $|\widetilde{P_{\beta_{i+1}}}\rangle$ from $|\widetilde{P_{\beta_i}}\rangle$ by A$\frac{\pi}{3}$AA with accuracy $\frac{\epsilon}{l_{\max}}$ and overlap $\frac{9}{10e^2}$.
12: **end for**
13: Output flg $= 0$.

Again, we can use only approximations of reflection operators. Reference [5] showed that, with probability at least $1 - \delta$, NAE using approximate reflection operators instead of exact ones output an ϵ approximation of $|\langle\Pi'|\Pi\rangle|^2$ for stationary distributions Π and Π' of some Markov chains. We hereafter call this approximate NAE (ANAE) with accuracy ϵ and failure probability δ. The following theorem states on its complexity.

Theorem 9 (Theorem 9 in Ref. [5]). Consider Markov chains $\mathcal{C}_1$ and $\mathcal{C}_2$ on a finite state space Ω with stationary distributions Π_1 and Π_2 and spectral gaps lower bounded by $\Delta \in (0, 1)$. Suppose that the state $|\Pi_1\rangle$ is given. Then, there is a quantum algorithm with the following properties:

(i) With probability at least $1 - \delta$, it outputs an ϵ approximation of $|\langle\Pi_1|\Pi_2\rangle|^2$ and a flag 1, and restores the state $|\Pi_1\rangle$.

(ii) Otherwise, output a flag 0.

(iii) It uses the controlled quantum walk operators for $\mathcal{C}_1$ and $\mathcal{C}_2$ $O(\frac{1}{\epsilon\sqrt{\Delta}}\log(\frac{1}{\epsilon})\log(\frac{1}{\delta}))$ times.

Then, combining the above building blocks yields the method proposed in Ref. [5]. It is summarized as Algorithm 2.

Here, $l_{\max} := \sqrt{\bar{L}\log\bar{L}}$, where $\bar{L} := \mathbb{E}_{P_0}[L(x)]$, and $L_{\max} := \max_{x\in\Omega} L(x)$.

The complexity of this algorithm is stated in Theorem 10.

Theorem 10 (*Theorem 10 in Ref. [5]*). Algorithm 2 yields output (A) with probability at least $1 - \eta$, calling operators in $\{U_\beta \mid \beta \in (0, 1]\}$
$$O\left(\frac{l_{\max}}{\sqrt{\Delta_{\min}}}\left(\log^2 l_{\max} + \log L_{\max}\log\left(\frac{l_{\max}L_{\max}}{\eta}\right)\right)\right) \tag{32}$$
times. For the obtained $\beta_0, \beta_1, \ldots, \beta_l$, using A$\frac{\pi}{3}$AA with accuracy ϵ and overlap $\frac{9}{10e^2}$ for Markov chains with stationary distributions $P_{\beta_0}, P_{\beta_1}, \ldots, P_{\beta_l}$, we can generate the state
$$|\widetilde{P_{\beta_l}}\rangle := |P_{\beta_l}\rangle|0\rangle^{\otimes n_{\text{anc}}} + |\xi\rangle, \tag{33}$$
where $n_{\text{anc}} = O(\log(\frac{1}{\Delta_{\min}})\log(\frac{l_{\max}}{\epsilon}))$ and $|\xi\rangle$ is an unnormalized state such that $\||\xi\rangle\| \leqslant \epsilon$. In this process, operators $U_{\beta_1}, \ldots, U_{\beta_l}$ are called
$$O\left(\frac{l_{\max}}{\sqrt{\Delta_{\min}}}\log^2\left(\frac{l_{\max}}{\epsilon}\right)\right) \tag{34}$$
times.

E. Quantum Monte Carlo integration

Reference [27] presented a quantum algorithm to calculate an expected value of a random variable, which we call QMCI in this paper.

Theorem 11 (*Theorem 2.3 in Ref. [27]*). Let P be a probability distribution on a finite sample space $\Omega \subset \mathbb{R}^d$. Suppose that we have a quantum circuit O_P on a two-register system that acts as $O_P|0\rangle|0\rangle = \sum_{x\in\Omega}\sqrt{P(x)}|\phi_x\rangle|x\rangle$, where $|\phi_x\rangle$ is some state on the first register. Also suppose that, for a function $F : \Omega \to [0, 1]$, we have a quantum circuit O_F on a two-register system that acts as $O_F|x\rangle|0\rangle = |x\rangle|F(x)\rangle$ for any $x \in \Omega$. Then, for any $\epsilon \in \mathbb{R}_+$ and $\delta \in (0, 1)$, there is a quantum algorithm that, with probability at least $1 - \delta$, outputs an ϵ approximation of $\mu_F := \sum_{x\in\Omega} P(x)F(x)$, making
$$O\left(\frac{1}{\epsilon}\log\delta^{-1}\right) \tag{35}$$
uses of O_P and O_F.

The above theorem is on a version of the algorithm for a bounded integrand F. Reference [27] presented another version for an integrand with a bounded variance. We now present a modification of this so that it can be used in QSA. Namely, we aim to obtain not an approximate value of an expectation but a quantum state in which approximations are encoded, making no measurement. Besides, we concentrate on the situation that we compute the mean of a finite number of real numbers, which is sufficient for our purpose.

Theorem 12. Let M be a positive integer and $\mathcal{X}$ be a set of M real numbers, $X_0, \ldots, X_{M-1}$, whose mean is $\mu := \frac{1}{M}\sum_{i=0}^{M-1} X_i$ and sample variance satisfies $\frac{1}{M}\sum_{i=0}^{M-1} X_i^2 - \mu^2 \leqslant \sigma^2$ with some $\sigma \in \mathbb{R}_+$. Suppose that we are given access to a unitary operator $O_{\mathcal{X}}$ that acts as
$$O_{\mathcal{X}}|i\rangle|0\rangle = |i\rangle|X_i\rangle, \tag{36}$$
for any $i \in [M]_0$. Then, for any $\epsilon \in \mathbb{R}_+$ and $\delta \in (0, 1)$, we have access to a unitary operator $O^{\text{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ that acts on a system of two registers R_1 and R_2 as
$$O^{\text{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}|0\rangle_{R_1}|0\rangle_{R_2} = |0\rangle_{R_1}|\tilde{\mu}\rangle_{R_2} + \gamma|\psi\rangle_{R_1,R_2}, \tag{37}$$

where $\tilde{\mu}$ is an ϵ approximation of μ, $|\psi\rangle_{R_1,R_2}$ is a state on the entire system, and $\gamma \in \mathbb{C}$ satisfies $|\gamma|^2 \leqslant \delta$. In $O^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$, $O_{\mathcal{X}}$ is used

$$O\left(\frac{\sigma}{\epsilon}\log^{3/2}\left(\frac{\sigma}{\epsilon}\right)\log\log\left(\frac{\sigma}{\epsilon}\right)\log\left(\frac{1}{\delta}\right)\right) \tag{38}$$

times. The total qubit number in the system of R_1 and R_2 is of order

$$O\left(\left(\log M+\log\left(\frac{\sigma}{\epsilon}\right)\right)\log\left(\frac{\sigma}{\epsilon}\right)\log\log\left(\frac{\sigma}{\epsilon}\right)\log\delta^{-1}\right). \tag{39}$$

Although this theorem resembles Theorem 5 in Ref. [23], there is the following difference. $O^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ in Ref. [23], which we rename $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$, generates a superposition of $|y_1\rangle, |y_2\rangle, \ldots$, where $\{y_i\}$ are real numbers close to μ. On the other hand, the state in Eq. (37) is almost equal to a product state of $|\tilde{\mu}\rangle$, a computational basis state corresponding to one approximation of μ, and $|0\rangle$, except for a small residual term $\gamma|\psi\rangle$. This is realized by combining $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ and rounding. Including this point, the proof of Theorem 12 is presented in Appendix B.

III. PROPOSED ALGORITHM

Now, let us present the proposed algorithm, the quantum MH algorithm with the target distribution estimated by QMCI.

A. Modified quantum walk operator

We start from generalizing the quantum walk operator in Eq. (16) for Ising models to that for the Markov chain generated by Algorithm 1 with a general finite state space $\Omega \subset \mathbb{R}^d$. We define

$$U = RV^\dagger B^\dagger SFBV. \tag{40}$$

This acts on a system of two quantum registers R_{S} and R_{M}, which now have a sufficient number of qubits to represent real vectors, and a qubit R_{C}. V acts on the system of R_{S} and R_{M} as

$$V|x\rangle_{R_{\mathrm{S}}}|0\rangle_{R_{\mathrm{M}}} = |x\rangle_{R_{\mathrm{S}}}\sum_{\Delta x\in\Omega_x}\sqrt{T(x,x+\Delta x)}|\Delta x\rangle_{R_{\mathrm{M}}} \tag{41}$$

for any $x \in \Omega$, where

$$\Delta\Omega_x := \{\Delta x \in \mathbb{R}^d \mid x+\Delta x \in \Omega\} \tag{42}$$

is the set of all the possible jumps in a transition from x. B acts as

$$\begin{aligned} &B|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} \\ &\quad = |x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}} \\ &\qquad \otimes \begin{pmatrix} \sqrt{1-A(x,x+\Delta x)} & -\sqrt{A(x,x+\Delta x)} \\ \sqrt{A(x,x+\Delta x)} & \sqrt{1-A(x,x+\Delta x)} \end{pmatrix}|\phi\rangle_{R_{\mathrm{C}}} \end{aligned} \tag{43}$$

for any $x \in \Omega$, $\Delta x \in \Omega_x$, and state $|\phi\rangle$ on R_{C}. In this operation, we compute $A(x, x+\Delta x)$ on an ancillary register using $|x\rangle_{R_{\mathrm{S}}}$ and $|\Delta x\rangle_{R_{\mathrm{M}}}$ as inputs, and use the rotation gate on R_{C} with the angle controlled by the ancillary register. F makes a state transition, which is implemented by an adder gate controlled by R_{C}, that is,

$$\begin{aligned} &F|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} \\ &\quad = \begin{cases} |x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle = |0\rangle_{R_{\mathrm{C}}} \\ |x+\Delta x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|1\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle_{R_{\mathrm{C}}} = |1\rangle_{R_{\mathrm{C}}}. \end{cases} \end{aligned} \tag{44}$$

The unitary S, for which Eq. (16) has no counterpart, acts on the system of R_{M} and R_{C} to flip the sign of the value on R_{M} under the control by R_{C}:

$$S|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} = \begin{cases} |\Delta x\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle_{R_{\mathrm{C}}} = |0\rangle_{R_{\mathrm{C}}} \\ |-\Delta x\rangle_{R_{\mathrm{M}}}|1\rangle_{R_{\mathrm{C}}} & \text{if } |\phi\rangle_{R_{\mathrm{C}}} = |1\rangle_{R_{\mathrm{C}}}. \end{cases} \tag{45}$$

In other words, S converts the jump from x to y to the inverse jump from y to x. We can consider that an identity operator is contained in Eq. (16) as a counterpart for S, since any spin flip is the inverse transform of itself. Finally, R is same as R_{IS} in Eq. (21). U in Eq. (40) also has the following property, the same as U_{IS} in Eq. (16).

Theorem 13. Consider the Markov chain generated by Algorithm 1 and denote by Δ its spectral gap. Define

$$\mathcal{A} := \mathrm{span}\{|x\rangle_{R_{\mathrm{S}}}|0\rangle_{R_{\mathrm{M}}}|0\rangle_{R_{\mathrm{C}}}|x\in\Omega\}, \quad \mathcal{B} := V^\dagger B^\dagger SFBV\mathcal{A}. \tag{46}$$

Then, on $\mathcal{A}+\mathcal{B}$, $|P\rangle$ is the unique eigenstate of U with eigenvalue 1, and any other eigenvalue is written as $e^{i\theta}$ with $\theta \in \mathbb{R}$ such that $|\theta| \geqslant \arccos(1-\Delta)$.

This is proven in Appendix A. Note that, as seen in the proofs in Appendix A, S is needed so that U holds the property stated in this theorem.

Let us consider how to implement the building-block operators in U. F and S are an addition and a sign flip controlled by the qubit R_{C}, respectively. Various quantum circuits for arithmetic have been proposed so far (see Ref. [40] as a review on circuits for four arithmetic operations and Refs. [41,42] as studies on circuits for elementary functions), and making them controlled is straightforward. R is an operator that multiplies -1 to the state vector when all the qubits in R_{M} and R_{C} take $|0\rangle$ and thus implemented with a multicontrolled Pauli Z gate. V is a circuit for loading a probability distribution into a quantum state, which has been also studied widely so far. If T can be calculated by some arithmetic, V can be implemented by the so-called Grover-Rudolph method [43], using a logarithmic number of arithmetic circuits with respect to the number of grid points for discrete approximation. Recently, some methods that avoid usage of arithmetic circuits have been proposed [44–48], including variational ones such as the quantum generative adversarial network [49–56].

Compared with these operators, B can be costly in some situations. Specifically, calculating the target distribution P, which is needed to evaluate the acceptance ratio, can be costly. For example, in parameter estimation in GW detection experiments, which was mentioned in the Introduction and will be explained in more detail in Sec. IV, P is obtained via calculating the log-likelihood function. It is determined by GW parameters and detector output data and evaluated as a sum of many terms that correspond to contributions from various frequency modes of the data. Naively calculating and summing up these terms leads to a large number of operations proportional to the number of terms. More generally, a similar

issue can arise in big-data analysis, specifically when we estimate parameters of a statistical model based on a lot of independent sample data and the log-likelihood is a sum of contributions from them.

B. Approximate quantum walk operator via calculating the target distribution by quantum Monte Carlo integration

Then, we are motivated to develop some faster way to calculate P in the aforementioned situation. We consider whether QMCI can be used to speed up summation of many terms in calculation of P.

We start from presenting the setup we consider. We make the following assumption.

Assumption 1. For every $x \in \Omega$, P is written as

$$P(x) = P_0(x)e^{-L(x)}. \tag{47}$$

Here, P_0 is a probability distribution on Ω. $L : \Omega \to \mathbb{R}_+$ is called the negative log-likelihood and written as

$$L(x) = L_{\mathrm{sum}}(x) + \ell_0(x) + C \tag{48}$$

with $\ell_0 : \Omega \to \mathbb{R}$, C a constant independent of x, and

$$L_{\mathrm{sum}}(x) := \frac{1}{M}\sum_{i=0}^{M-1} \ell(i, x), \tag{49}$$

where $M \in \mathbb{N}$ and $\ell : [M]_0 \times \Omega \to \mathbb{R}$. Besides, we are given the quantum circuit O_ℓ, which acts on a three-register system as

$$O_\ell|x\rangle|i\rangle|0\rangle = |x\rangle|i\rangle|\ell(i, x)\rangle \tag{50}$$

for any $i \in [M]_0$ and $x \in \Omega$. Moreover, we are given $\sigma \in \mathbb{R}_+$ such that

$$\frac{1}{M}\sum_{i=0}^{M-1}(\ell(i, x))^2 - \left(\frac{1}{M}\sum_{i=0}^{M-1}\ell(i, x)\right)^2 \leqslant \sigma^2 \tag{51}$$

for any $x \in \Omega$.

This assumption is threefold. The first part, the form of P, is in line with the aforementioned situation, where the log-likelihood contains a sum of many terms. The second one is availability of the quantum circuit O_ℓ to calculate the terms ℓ, which is used in QMCI. For large M, O_ℓ is the circuit queried most, and thus we hereafter focus on the number of queries to this as a metric of the complexity of our algorithm. The third one, the boundedness of the variance of ℓ, is needed to bound the error in QMCI.

Note that the form of L_{sum} is in fact an average rather than a sum. This is just for convenience in applying QMCI to computing it. Also note that the order of σ can depend on the term number M. For example, if L_{sum} is a sum of contributions from M independent samples, which applies to many cases in estimating parameters of statistical models, putting an overall factor $1/M$ and redefining $M\ell$ as ℓ leads to the form in Eq. (49), but this makes the order of ℓ $O(M)$ if it is originally independent of M.

Hereafter, we denote by $\mathcal{C}_L$ the Markov chain generated by Algorithm 1 with P written as Eq. (47) with L.

We also assume the availability of the quantum circuits to generate the states that encode the proposal distribution T and the prior distribution P_0 in amplitudes.

Assumption 2. We are given quantum circuits V that acts as Eq. (41).

Assumption 3. We are given quantum circuits O_{P_0} that acts as Eq. (31).

Furthermore, we assume that we can use a quantum circuit to compute the acceptance ratio $A(x, y)$, given estimates $\hat{L}_x$ and $\hat{L}_y$ of $L_{\mathrm{sum}}(x)$ and $L_{\mathrm{sum}}(y)$.

Assumption 4. We are given the quantum circuit O_{AR} that acts as

$$\begin{aligned} &O_{\mathrm{AR}}|x\rangle|y\rangle|\hat{L}_x\rangle|\hat{L}_y\rangle|0\rangle \\ &\quad = |x\rangle|y\rangle|\hat{L}_x\rangle|\hat{L}_y\rangle\left|\frac{P_0(y)T(y, x)\exp(-(\hat{L}_y + \ell_0(y)))}{P_0(x)T(x, y)\exp(-(\hat{L}_x + \ell_0(x)))}\right\rangle \end{aligned} \tag{52}$$

for any $x, y \in \Omega$ and $\hat{L}_x, \hat{L}_y \in \mathbb{R}$.

In many cases, the formulas for P_0 and T are explicitly given with elementary functions and thus O_{AR} is implemented with arithmetic circuits.

Under these assumptions, Theorem 12 leads to the following lemma.

Lemma 1. Let Ω be a finite subset of $\mathbb{R}^d$ and P be a distribution on it. Under Assumptions 1 and 4, for any $\delta, \epsilon \in (0, 1)$, we have access to a unitary operator $\tilde{\tilde{B}}_{\delta,\epsilon}$ on a system of three registers R_{S}, R_{M}, and R_{A} and a qubit R_{C} such that, for any $x \in \Omega$, $\Delta x \in \Delta\Omega_x$, and state $|\phi\rangle_{R_{\mathrm{C}}}$ on R_{C},

$$\begin{aligned} &\tilde{\tilde{B}}_{\delta,\epsilon}|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}}|0\rangle_{R_{\mathrm{A}}} \\ &\quad = \tilde{B}_\epsilon(|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}})|0\rangle_{R_{\mathrm{A}}} + \gamma_{x,\Delta x,\delta,\epsilon}|\Psi\rangle. \end{aligned} \tag{53}$$

Here,

$$\begin{aligned} &\tilde{B}_\epsilon|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}} \\ &\quad = |x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}} \\ &\qquad \otimes \begin{pmatrix} \sqrt{1 - \tilde{A}(x, x + \Delta x)} & -\sqrt{\tilde{A}(x, x + \Delta x)} \\ \sqrt{\tilde{A}(x, x + \Delta x)} & \sqrt{1 - \tilde{A}(x, x + \Delta x)} \end{pmatrix}|\phi\rangle_{R_{\mathrm{C}}}, \end{aligned} \tag{54}$$

where $\tilde{A} : \Omega \times \Omega \to \mathbb{R}$ is written as

$$\tilde{A}(x, y) = \min\left\{1, \frac{P_0(y)e^{-\tilde{L}(y)}T(y, x)}{P_0(x)e^{-\tilde{L}(x)}T(x, y)}\right\} \tag{55}$$

with $\tilde{L} : \Omega \to \mathbb{R}_+$ such that

$$\max_{x\in\Omega}|\tilde{L}(x) - L(x)| \leqslant \epsilon. \tag{56}$$

$|\Psi\rangle$ is some state on the entire system. $\gamma_{x,\Delta x,\delta,\epsilon} \in \mathbb{C}$ satisfies $|\gamma_{x,\Delta x,\delta,\epsilon}| \leqslant \delta$. $\tilde{\tilde{B}}_{\delta,\epsilon}$ makes queries to O_ℓ, whose number is of order (38). The qubit number in the entire system is of order (39).

Proof. Because of Theorem 12, given O_ℓ, we have access to a unitary operator $O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}$ on the system of three registers that, for any $x \in \Omega$, acts as

$$O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}|x\rangle|0\rangle|0\rangle = |x\rangle(|0\rangle|\tilde{L}_{\mathrm{sum}}(x)\rangle + \gamma_{x,\frac{\delta^2}{16}}|\psi_x\rangle), \tag{57}$$

where $\gamma_{x,\frac{\delta^2}{16}} \in \mathbb{C}$ satisfies $|\gamma_{x,\frac{\delta^2}{16}}|^2 \leqslant \frac{\delta^2}{16}$, $|\psi_x\rangle$ is some state on the system of the second and third registers, and $\tilde{L}_{\mathrm{sum}} : \Omega \to \mathbb{R}$

satisfies

$$\max_{x\in\Omega} |\tilde{L}_{\rm sum}(x) - L_{\rm sum}(x)| \leqslant \epsilon. \tag{58}$$

Equipped with this, we can construct the quantum circuit for the following operation on the system of $R_{\rm S}$, $R_{\rm M}$, $R_{\rm C}$, and ancillary registers $R_{{\rm A},1}, \ldots, R_{{\rm A},6}$:

$$\begin{aligned}
&|x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|0\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|0\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}}\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|0\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}}\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}\big(|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}} + \gamma_{x,\frac{\delta^2}{16}}|\psi_x\rangle_{R_{{\rm A},2},R_{{\rm A},3}}\big)|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}}\\
&\quad:= |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}} + \gamma_{x,\frac{\delta^2}{16}}\big|\Psi^{(1)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}\big(|0\rangle_{R_{{\rm A},4}}|\tilde{L}_{\rm sum}(x+\Delta x)\rangle_{R_{{\rm A},5}} + \gamma_{x+\Delta x,\frac{\delta^2}{16}}|\psi_{x+\Delta x}\rangle_{R_{{\rm A},4},R_{{\rm A},5}}\big)|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}}\\
&\qquad + \gamma_{x,\frac{\delta^2}{16}}\big|\Psi^{(2)}_{x,\Delta x}\big\rangle\\
&\quad:= |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|\tilde{L}_{\rm sum}(x+\Delta x)\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}} + \gamma'_{x,\Delta x}\big|\Psi^{(3)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|\tilde{L}_{\rm sum}(x+\Delta x)\rangle_{R_{{\rm A},5}}|\tilde{A}(x,x+\Delta x)\rangle_{R_{{\rm A},6}}|\phi\rangle_{R_{\rm C}} + \gamma'_{x,\Delta x}\big|\Psi^{(4)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|\tilde{L}_{\rm sum}(x+\Delta x)\rangle_{R_{{\rm A},5}}|\tilde{A}(x,x+\Delta x)\rangle_{R_{{\rm A},6}}\\
&\qquad\otimes\begin{pmatrix}\sqrt{1-\tilde{A}(x,x+\Delta x)} & -\sqrt{\tilde{A}(x,x+\Delta x)}\\ \sqrt{\tilde{A}(x,x+\Delta x)} & \sqrt{1-\tilde{A}(x,x+\Delta x)}\end{pmatrix}|\phi\rangle_{R_{\rm C}} + \gamma'_{x,\Delta x}\big|\Psi^{(5)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|\tilde{L}_{\rm sum}(x+\Delta x)\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}\\
&\qquad\otimes\begin{pmatrix}\sqrt{1-\tilde{A}(x,x+\Delta x)} & -\sqrt{\tilde{A}(x,x+\Delta x)}\\ \sqrt{\tilde{A}(x,x+\Delta x)} & \sqrt{1-\tilde{A}(x,x+\Delta x)}\end{pmatrix}|\phi\rangle_{R_{\rm C}} + \gamma'_{x,\Delta x}\big|\Psi^{(6)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|x+\Delta x\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|0\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}\\
&\qquad\otimes\begin{pmatrix}\sqrt{1-\tilde{A}(x,x+\Delta x)} & -\sqrt{\tilde{A}(x,x+\Delta x)}\\ \sqrt{\tilde{A}(x,x+\Delta x)} & \sqrt{1-\tilde{A}(x,x+\Delta x)}\end{pmatrix}|\phi\rangle_{R_{\rm C}} + \gamma''_{x,\Delta x}\big|\Psi^{(7)}_{x,\Delta x}\big\rangle\\
&\quad\to |x\rangle_{R_{\rm S}}|\Delta x\rangle_{R_{\rm M}}|0\rangle_{R_{{\rm A},1}}|0\rangle_{R_{{\rm A},2}}|0\rangle_{R_{{\rm A},3}}|0\rangle_{R_{{\rm A},4}}|0\rangle_{R_{{\rm A},5}}|0\rangle_{R_{{\rm A},6}}\\
&\qquad\otimes\begin{pmatrix}\sqrt{1-\tilde{A}(x,x+\Delta x)} & -\sqrt{\tilde{A}(x,x+\Delta x)}\\ \sqrt{\tilde{A}(x,x+\Delta x)} & \sqrt{1-\tilde{A}(x,x+\Delta x)}\end{pmatrix}|\phi\rangle_{R_{\rm C}} + \gamma''_{x,\Delta x}\big|\Psi^{(8)}_{x,\Delta x}\big\rangle\\
&\quad=: |\tilde{\Phi}_{x,\Delta x}\rangle,
\end{aligned} \tag{59}$$

where $|\Psi^{(1)}_{x,\Delta x}\rangle, \ldots, |\Psi^{(8)}_{x,\Delta x}\rangle$ are some states on the entire system and $\gamma'_{x,\Delta x}, \gamma''_{x,\Delta x} \in \mathbb{C}$. In Eq. (59), we use an adder circuit at the first arrow. At the second and third arrows, we use $O_{L_{\rm sum},\epsilon,\frac{\delta^2}{16},\sigma}$ on the system of $R_{\rm S}$, $R_{{\rm A},2}$, and $R_{{\rm A},3}$ and that of $R_{{\rm A}_1}$, $R_{{\rm A},4}$, and $R_{{\rm A},5}$, respectively. At the fourth arrow, we use $O_{\rm AR}$ to compute $\tilde{A}(x,x+\Delta x)$ as Eq. (55) with $\tilde{L} = \tilde{L}_{\rm sum} + \ell_0 + C$, which satisfies Eq. (56) because of Eq. (58). The fifth arrow is by the Y rotation $\left(\begin{smallmatrix}\cos\frac{\varphi}{2} & -\sin\frac{\varphi}{2}\\ \sin\frac{\varphi}{2} & \cos\frac{\varphi}{2}\end{smallmatrix}\right)$ on $R_{\rm C}$ with the rotation angle specified by $R_{{\rm A},6}$, which is implemented as follows [57]: we compute $\varphi = 2\arcsin(\sqrt{\tilde{A}(x,x+\Delta x)})$ onto another ancillary register using arithmetic circuits [40–42] and apply fixed-angle Y-rotation gates controlled by qubits in that ancillary register to $R_{\rm C}$. At the sixth arrow, we perform the inverse of the operation at the fourth arrow. The seventh arrow is by the inverses of the operations at the second and third arrows, which act as

$$\begin{aligned}
&\big(O_{L_{\rm sum},\epsilon,\frac{\delta^2}{16},\sigma}\big)^\dagger|x\rangle_{R_{\rm S}}|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}}\\
&\quad= \big(O_{L_{\rm sum},\epsilon,\frac{\delta^2}{16},\sigma}\big)^\dagger|x\rangle_{R_{\rm S}}\\
&\qquad\otimes\big(|0\rangle_{R_{{\rm A},2}}|\tilde{L}_{\rm sum}(x)\rangle_{R_{{\rm A},3}} + \gamma_{x,\frac{\delta^2}{16}}|\psi_x\rangle_{R_{{\rm A},2},R_{{\rm A},3}}\big)\\
&\qquad - \gamma_{x,\frac{\delta^2}{16}}\big(O_{L_{\rm sum},\epsilon,\frac{\delta^2}{16},\sigma}\big)^\dagger|\psi_x\rangle_{R_{{\rm A},2},R_{{\rm A},3}}\\
&\quad= |x\rangle_{R_{\rm S}}|0\rangle_{R_{{\rm A},2}}|0\rangle_{R_{{\rm A},3}} - \gamma_{x,\frac{\delta^2}{16}}\big(O_{L_{\rm sum},\epsilon,\frac{\delta^2}{16},\sigma}\big)^\dagger|\psi_x\rangle_{R_{{\rm A},2},R_{{\rm A},3}}
\end{aligned} \tag{60}$$

and, similarly,

$$\begin{aligned}&\left(O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}\right)^\dagger|x+\Delta x\rangle_{R_{\mathrm{A},1}}|0\rangle_{R_{\mathrm{A},4}}|\tilde{L}_{\mathrm{sum}}(x+\Delta x)\rangle_{R_{\mathrm{A},5}}\\&\quad=|x+\Delta x\rangle_{R_{\mathrm{A},1}}|0\rangle_{R_{\mathrm{A},4}}|0\rangle_{R_{\mathrm{A},5}}\\&\qquad-\gamma_{x+\Delta x,\frac{\delta^2}{16}}\left(O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}\right)^\dagger|\psi_{x+\Delta x}\rangle_{R_{\mathrm{A},4},R_{\mathrm{A},5}}.\end{aligned}\tag{61}$$

At the last arrow, we perform the inverse of the operation at the first arrow.

Then, let us show that $|\tilde{\Phi}_{x,\Delta x}\rangle$ is in the form of Eq. (53), with $R_{\mathrm{A},1},\dots,R_{\mathrm{A},6}$ collectively seen as R_{A}. Since we have seen that Eq. (56) holds, it is sufficient to check $|\gamma''_{x,\Delta x}|\leqslant\delta$. This is done as follows. We see that $\gamma'_{x,\Delta x}$, which is introduced by two $O'_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}$s in the second and third arrows in Eq. (59), is bounded as

$$|\gamma'_{x,\Delta x}|\leqslant\left|\gamma_{x,\frac{\delta^2}{16}}\right|+\left|\gamma_{x+\Delta x,\frac{\delta^2}{16}}\right|\leqslant\sqrt{\frac{\delta^2}{16}}+\sqrt{\frac{\delta^2}{16}}=\frac{\delta}{2}.\tag{62}$$

Similarly, applying $(O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma})^\dagger$ twice at the eighth arrow in Eq. (59) increases this by at most $\frac{\delta}{2}$:

$$|\gamma''_{x,\Delta x}|\leqslant|\gamma'_{x,\Delta x}|+\frac{\delta}{2}\leqslant\delta.\tag{63}$$

Thus, we have $|\gamma''_{x,\Delta x}|\leqslant\delta$.

The statements on the number of queries to O_ℓ and the qubit number immediately follow from Theorem 12, which gives the bounds on the query number and qubit number in $O_{L_{\mathrm{sum}},\epsilon,\frac{\delta^2}{16},\sigma}$ as Eqs. (38) and (39). ■

We now define the approximate quantum walk operator,

$$\tilde{\tilde{U}}_{\delta,\epsilon}:=\tilde{R}\tilde{V}^\dagger\tilde{\tilde{B}}^\dagger_{\frac{\delta}{2},\epsilon}\tilde{S}\tilde{F}\tilde{\tilde{B}}_{\frac{\delta}{2},\epsilon}\tilde{V}\tag{64}$$

on the system of R_{S}, R_{M}, R_{A}, and R_{C}, with $\tilde{R}$, $\tilde{V}$, $\tilde{S}$, and $\tilde{F}$ defined as $R\otimes I_{R_{\mathrm{A}}}$ and so on. We then have the following lemma immediately.

Lemma 2. Let Ω be a finite subset of $\mathbb{R}^d$ and P be a distribution on it. Under Assumptions 1, 2 and 4, for any $\delta,\epsilon\in(0,1)$, we have access to a unitary operator $\tilde{\tilde{U}}_{\delta,\epsilon}$ on a system of three registers R_{S}, R_{M}, and R_{A} and a qubit R_{C}, which is given as Eq. (64), and, for any $x\in\Omega$, $\Delta x\in\Delta\Omega_x$, and state $|\phi\rangle_{R_{\mathrm{C}}}$ on R_{C}, acts as

$$\begin{aligned}&\tilde{\tilde{U}}_{\delta,\epsilon}|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}}|0\rangle_{R_{\mathrm{A}}}\\&\quad=(\tilde{U}_\epsilon|x\rangle_{R_{\mathrm{S}}}|\Delta x\rangle_{R_{\mathrm{M}}}|\phi\rangle_{R_{\mathrm{C}}})|0\rangle_{R_{\mathrm{A}}}+\tilde{\gamma}_{x,\Delta x,\phi,\delta,\epsilon}|\tilde{\Psi}\rangle,\end{aligned}\tag{65}$$

where $|\tilde{\Psi}\rangle$ is some state on the entire system, $\tilde{\gamma}_{x,\Delta x,\phi,\delta,\epsilon}\in\mathbb{C}$ satisfies $|\tilde{\gamma}_{x,\Delta x,\phi,\delta,\epsilon}|\leqslant\delta$, and $\tilde{U}_\epsilon:=RV^\dagger\tilde{B}^\dagger_\epsilon SF\tilde{B}_\epsilon V$. $\tilde{\tilde{U}}_{\delta,\epsilon}$ makes a number of order (38) of calls to O_ℓ and uses a number of order (39) of qubits.

Note that $\tilde{\tilde{U}}_{\delta,\epsilon}$ has two types of differences from the exact quantum walk operator U. First, $\tilde{\tilde{U}}_{\delta,\epsilon}$ does not exactly act as a quantum walk operator because it generates the residual term $\tilde{\gamma}_{x,\Delta x,\phi,\delta,\epsilon}|\tilde{\Psi}\rangle$ in Eq. (65). Second, even if there were no residual term, $\tilde{\tilde{U}}_{\delta,\epsilon}$ would not be the quantum walk operator for the Markov chain $\mathcal{C}_L$ we consider but that for another one, $\mathcal{C}_{\tilde{L}}$, because of the error in the approximation $\tilde{L}$ of the exact negative log-likelihood L. This difference makes the stationary distribution differ from the target distribution P. Nevertheless, we can use $\tilde{\tilde{U}}_{\delta,\epsilon}$, controlling these differences by taking sufficiently small δ and ϵ.

C. Quantum simulated annealing with the approximate quantum walk operator

Now, we can construct an approximation of the phase gate $R^\omega_{|P\rangle}$ using this $\tilde{\tilde{U}}_{\delta,\epsilon}$ instead of the exact quantum walk operator U.

Lemma 3. Let $\delta,\epsilon\in(0,1)$. Under Assumptions 1, 2, and 4, consider a Markov chain $\mathcal{C}_L$. Denote its transition matrix by W and its spectral gap by Δ. Denote by κ the condition number of the matrix Q such that $Q^{-1}WQ$ is diagonal. Let ω be a complex number with unit modulus. Then, we have access to a unitary operator $\tilde{\tilde{R}}^\omega_{L,\delta,\epsilon}$ that has the following properties:

(i) $\tilde{\tilde{R}}^\omega_{L,\delta,\epsilon}$ acts on a system of R_{S} and n_{anc} ancillary qubits. Here,

$$\begin{aligned}n_{\mathrm{anc}}=O\Bigg(&\log\left(\frac{1}{\Delta}\right)\log\left(\frac{1}{\delta}\right)+\left(\log M+\log\left(\frac{\sigma}{\epsilon'}\right)\right)\\&\times\log\left(\frac{\sigma}{\epsilon'}\right)\log\log\left(\frac{\sigma}{\epsilon'}\right)\log\left(\frac{1}{\delta\sqrt{\Delta}}\right)\Bigg),\end{aligned}\tag{66}$$

where

$$\epsilon':=\min\left\{\epsilon,\frac{\Delta}{16\sqrt{\max_{y\in\Omega}\sum_{x\in\Omega\setminus\{y\}}T_{xy}\kappa}}\right\}.\tag{67}$$

(ii) $\tilde{\tilde{R}}^\omega_{L,\delta,\epsilon}$ uses O_ℓ

$$O\left(\frac{\sigma}{\epsilon'\sqrt{\Delta}}\log^{3/2}\left(\frac{\sigma}{\epsilon'}\right)\log\log\left(\frac{\sigma}{\epsilon'}\right)\log\left(\frac{1}{\delta\sqrt{\Delta}}\right)\log\left(\frac{1}{\delta}\right)\right)\tag{68}$$

times.

(iii) For any state $|\Xi\rangle$ on R_{S},

$$\tilde{\tilde{R}}^\omega_{L,\delta,\epsilon}|\Xi\rangle|0\rangle^{\otimes n_{\mathrm{anc}}}=\left(R^\omega_{|\tilde{P}\rangle}|\Xi\rangle\right)|0\rangle^{\otimes n_{\mathrm{anc}}}+|\xi\rangle.\tag{69}$$

Here, $R^\omega_{|\tilde{P}\rangle}$ is a unitary defined as Eq. (26), where $\tilde{P}$ is a distribution on Ω in the form of

$$\tilde{P}(x)\propto P_0(x)e^{-\tilde{L}(x)}\tag{70}$$

with some function $\tilde{L}:\Omega\to\mathbb{R}_+$ satisfying

$$\max_{x\in\Omega}|\tilde{L}(x)-L(x)|\leqslant\epsilon',\tag{71}$$

and $|\xi\rangle$ is an unnormalized state vector with $\||\xi\rangle\|\leqslant\delta$.

To prove this, we use the following lemma on the spectral gap of the Markov chain with replacement of L with $\tilde{L}$. The proof of this is presented in Appendix C.

Lemma 4. Let $\tilde{L}:\Omega\to\mathbb{R}_+$ be a function on Ω and denote by $\tilde{\Delta}$ the spectral gap of $\mathcal{C}_{\tilde{L}}$. Let Δ and κ be the same as Lemma 3. Then, if $\epsilon:=\max_{x\in\Omega}|\tilde{L}(x)-L(x)|\leqslant\frac{1}{4}$,

$$\tilde{\Delta}\geqslant\Delta-16\left(\max_{y\in\Omega}\sum_{x\in\Omega\setminus\{y\}}T_{xy}\right)^{1/2}\kappa\epsilon\tag{72}$$

holds.

This lemma means that, although the MH Markov chain with $\tilde{L}$ instead of L is different from the original one, the change of the spectral gap is small if $\tilde{L}$ is close to L.

Then, the proof of Lemma 3 is as follows.

Proof of Lemma 3. Note that, because of Lemma 4 and Eq. (71), the spectral gap $\tilde{\Delta}$ of $\mathcal{C}_{\tilde{L}}$ satisfies $\tilde{\Delta} \geqslant \frac{\Delta}{2}$. Then, because of Theorem 5, if we had access to $\tilde{U}_{\epsilon'}$, we could construct $\tilde{R}^{\omega}_{|\tilde{P}\rangle,\frac{\delta}{2}}$ making $O(\frac{\log(1/\delta)}{\sqrt{\Delta}})$ uses of $\tilde{U}_{\epsilon'}$, with $\tilde{P}$ having the stated property. In reality, we can use $\tilde{\tilde{U}}_{\delta',\epsilon'}$, an approximation of $\tilde{U}_{\epsilon'}$ with some $\delta' \in (0,1)$. Recalling Lemma 2, we see that the unitary $\mathcal{R}_{\delta'}$ we obtain by using $\tilde{\tilde{U}}_{\delta',\epsilon'}$ instead of $\tilde{U}_{\epsilon'}$ in construction of $\tilde{R}^{\omega}_{|\tilde{P}\rangle,\frac{\delta}{2}}$ acts as $\mathcal{R}_{\delta'}|\Xi\rangle|0\rangle^{\otimes n_{\rm anc}} = \tilde{R}^{\omega}_{|\tilde{P}\rangle,\frac{\delta}{2}}|\Xi\rangle|0\rangle^{\otimes n_{\rm anc}} + |\xi'\rangle$, where $\||\xi'\rangle\| = O(\delta'\frac{\log(1/\delta)}{\sqrt{\Delta}})$. Thus, there exists $\delta' = \Theta(\frac{\delta\sqrt{\Delta}}{\log(1/\delta)})$ that makes $\||\xi'\rangle\| \leqslant \frac{\delta}{2}$. Since $\|\tilde{R}^{\omega}_{|\tilde{P}\rangle,\frac{\delta}{2}}|\Xi\rangle|0\rangle^{\otimes n_{\rm anc}} - (R^{\omega}_{|\tilde{P}\rangle}|\Xi\rangle)|0\rangle^{\otimes n_{\rm anc}}\| \leqslant \frac{\delta}{2}$, $\mathcal{R}_{\delta'}$ with this δ' is in fact $\tilde{\tilde{R}}^{\omega}_{L,\delta,\epsilon}$ that satisfies Eq. (69).

The statement on the qubit number follows since constructing $\tilde{R}^{\omega}_{|\tilde{P}\rangle,\frac{\delta}{2}}$ with $\tilde{U}_{\epsilon'}$ uses $O(\log(\frac{1}{\Delta})\log(\frac{1}{\delta}))$ qubits and using $\tilde{\tilde{U}}_{\delta',\epsilon'}$ instead of $\tilde{U}_{\epsilon'}$ adds $O((\log M + \log(\frac{\sigma}{\epsilon'}))\log(\frac{\sigma}{\epsilon'})\log\log(\frac{\sigma}{\epsilon'})\log(\delta')^{-1})$ qubits, whose sum is of order (66).

The upper bound (68) on the number of queries to O_ℓ is obtained by substituting $\delta' = \Theta(\frac{\delta\sqrt{\Delta}}{\log(1/\delta)})$ for δ and ϵ' for ϵ in Eq. (38), which yields the query number in one $\tilde{\tilde{U}}_{\delta',\epsilon'}$, and multiplying $O(\frac{\log(1/\delta)}{\sqrt{\Delta}})$. ∎

We can use this approximate phase gate instead of the exact one in QSA. Before we make a statement on this approximate QSA, let us make some preparation. First, we make the following assumptions.

Assumption 5. There exists $\Delta_{\min} \in (0,1)$ such that, for any $\beta \in (0,1]$, the spectral gap of the Markov chain $\mathcal{C}_{\beta L}$ is equal to or larger than $\Delta_{\min}$.

Assumption 6. There exists $\kappa_{\min} \in \mathbb{R}_+$ such that, for any $\beta \in (0,1]$, the condition number of the matrix Q_β that diagonalizes the transition matrix W_β for the Markov chain $\mathcal{C}_{\beta L}$, which means $Q_\beta^{-1}W_\beta Q_\beta$ is diagonal, is equal to or smaller than $\kappa_{\min}$.

Besides, we also present the following lemma, whose proof is presented in Appendix D.

Lemma 5. Consider the Markov chains $\mathcal{C}_L$ and $\mathcal{C}_{\tilde{L}}$ with $L : \Omega \to \mathbb{R}_+$ and $\tilde{L} : \Omega \to \mathbb{R}_+$. Then, for their stationary distributions $P \propto P_0 e^{-L}$ and $\tilde{P} \propto P_0 e^{-\tilde{L}}$,

$$\|\tilde{P} - P\|_{\rm TV} \leqslant 8\left(\left\lceil \frac{\log(2\sqrt{P_{\min}})}{\log(1-\Delta)}\right\rceil + \frac{1}{\Delta}\right) \times \max_{x\in\Omega} |\tilde{L}(x) - L(x)| \tag{73}$$

holds, where Δ is the spectral gap of $\mathcal{C}_L$ and $P_{\min} := \min_{x\in\Omega} P(x)$.

Then, we have the following theorem.

Theorem 14. Suppose that Assumptions 1 through 6 are satisfied. Then, for any $\epsilon, \delta \in (0,1)$, there exists an algorithm that makes

$$O\left(\frac{\sigma l_{\max}}{\epsilon''\sqrt{\Delta_{\min}}} \log^{3/2}\left(\frac{\sigma}{\epsilon''}\right) \log\log\left(\frac{\sigma}{\epsilon''}\right) \log\left(\frac{1}{\sqrt{\Delta_{\min}}}\right) \times \left(\log l_{\max} + \log L_{\max} \log\left(\frac{l_{\max} L_{\max}}{\delta}\right)\right)\right) \tag{74}$$

queries to O_ℓ, where

$$\epsilon'' := \min\left\{ \frac{\Delta_{\min}\epsilon}{8\left(\Delta_{\min}\left\lceil \frac{\log(2\sqrt{P_{\min}})}{\log(1-\Delta_{\min})}\right\rceil + 1\right)}, \times \frac{\Delta_{\min}}{16\sqrt{\max_{y\in\Omega}\sum_{x\in\Omega\setminus\{y\}} T_{xy}\kappa_{\min}}}, \frac{\bar{L}}{2}\right\}, \tag{75}$$

and, with probability at least $1-\delta$, outputs a sequence $\tilde{\beta}_0 = 0 < \tilde{\beta}_1 < \cdots < \tilde{\beta}_{l-1} < \tilde{\beta}_l = 1$ with the following properties:

(i) $l \leqslant l_{\max}$.

(ii) Given this sequence, we have a unitary operator that generates a state $|\widetilde{\tilde{P}}\rangle$ ϵ-close to $|\tilde{P}\rangle|0\rangle^{\otimes n_{\rm anc}}$, where $\tilde{P}$ is a probability distribution on Ω such that $\|\tilde{P} - P\|_{\rm TV} \leqslant \epsilon$ and

$$n_{\rm anc} = O\left(\log\left(\frac{1}{\Delta_{\min}}\right)\log\left(\frac{l_{\max}}{\epsilon''}\right) + \left(\log M + \log\left(\frac{\sigma}{\epsilon''}\right)\right)\log\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right) \times \log\left(\frac{l_{\max}}{\epsilon''\sqrt{\Delta_{\min}}}\right)\right). \tag{76}$$

In that operator, O_ℓ is called

$$O\left(\frac{\sigma l_{\max}}{\epsilon''\sqrt{\Delta_{\min}}} \log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{l_{\max}}{\epsilon\sqrt{\Delta_{\min}}}\right) \times \log^2\left(\frac{l_{\max}}{\epsilon}\right)\right) \tag{77}$$

times.

For ϵ such that

$$\epsilon \leqslant \min\left\{\frac{1}{2\sqrt{\max_{y\in\Omega}\sum_{x\in\Omega\setminus\{y\}} T_{xy}\kappa_{\min}}}, \frac{4\bar{L}}{\Delta_{\min}}\right\} \times \left(\Delta_{\min}\left\lceil \frac{\log(2\sqrt{P_{\min}})}{\log(1-\Delta_{\min})}\right\rceil + 1\right), \tag{78}$$

Eq. (77) becomes

$$\tilde{O}\left(\frac{\sigma\bar{L}^{1/2}}{\epsilon\Delta_{\min}^{3/2}}\right). \tag{79}$$

Proof of Theorem 14. First, note that, for any $\beta \in (0,1]$, $\tilde{\tilde{R}}^{\omega}_{\beta L,\delta',\epsilon''}$ is equal to $\tilde{R}^{\omega}_{|\tilde{P}_\beta\rangle,\delta'}$, where $\delta' \in (0,1)$ and $\tilde{P}_\beta$ is a distribution on Ω in the form of $\tilde{P}_\beta \propto P_0 e^{-\beta\tilde{L}}$ with some function $\tilde{L} : \Omega \to \mathbb{R}_+$ satisfying

$$\max_{x\in\Omega} |\tilde{L}(x) - L(x)| \leqslant \epsilon''. \tag{80}$$

On the other hand, according to Ref. [5], given $\tilde{R}^{\omega_{\pi/3}}_{|\tilde{P}_\beta\rangle,\delta'}$ and $\tilde{R}^{-1}_{|\tilde{P}_\beta\rangle,\delta'}$ with some $\delta' = \Theta(1)$ for any $\beta \in (0,1]$, Algorithm 2 outputs the sequence $\tilde{\beta}_0 = 0 < \tilde{\beta}_1 < \cdots < \tilde{\beta}_{l-1} < \tilde{\beta}_l = 1$ such that $l \leqslant \tilde{l}_{\max}$ and $|\langle\tilde{P}_{\tilde{\beta}_i}|\tilde{P}_{\tilde{\beta}_{i+1}}\rangle|^2 \geqslant \frac{9}{10e^2}$ with probability at least $1-\delta$, making $O(\tilde{l}_{\max}\log\tilde{l}_{\max})$ uses of operators in $\{\tilde{R}^{\omega_{\pi/3}}_{|\tilde{P}_\beta\rangle,\delta'} \mid \beta \in (0,1]\}$ and $O(\tilde{l}_{\max}\log\tilde{L}_{\max}\log(\frac{\tilde{l}_{\max}\tilde{L}_{\max}}{\delta}))$ uses of operators in $\{\tilde{R}^{\omega_{-1}}_{|\tilde{P}_\beta\rangle,\delta'} \mid \beta \in (0,1]\}$. Here, $\tilde{l}_{\max} := \mathbb{E}_{P_0}[\tilde{L}(x)]\log(\mathbb{E}_{P_0}[\tilde{L}(x)])$ and $\tilde{L}_{\max} := \max_{x\in\Omega}\tilde{L}(x)$, which

are of order $O(l_{\max})$ and $O(L_{\max})$, respectively, because of Eq. (80). Since we can construct $\tilde{\tilde{R}}^{\omega}_{\beta L,\delta',\epsilon''}$ similarly to $\tilde{\tilde{R}}^{\omega}_{L,\delta',\epsilon''}$ using O_ℓ, we can obtain the above $\{\tilde{\beta}_i\}$ by Algorithm 2. Because of Lemma 3, the number of queries to O_ℓ in $\tilde{\tilde{R}}^{\omega}_{\beta L,\delta',\epsilon''}$ is

$$O\left(\frac{\sigma}{\epsilon''\sqrt{\tilde{\Delta}_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{1}{\sqrt{\tilde{\Delta}_{\min}}}\right)\right). \tag{81}$$

Here, $\tilde{\Delta}_{\min}$ is a lower bound of the spectral gaps of $\{\mathcal{C}_{\beta\tilde{L}} \mid \beta \in (0,1]\}$, which satisfies $\tilde{\Delta}_{\min} \geqslant \frac{\Delta_{\min}}{2}$ because of Lemma 4 and Eq. (80). Combining the above discussions, we see that in finding $\tilde{\beta}_1, \ldots, \tilde{\beta}_{l-1}$ the total number of calls to O_ℓ is of order (74).

After that, as shown in Ref. [5], a series of A$\frac{\pi}{3}$AA generates $|\widetilde{\tilde{P}}\rangle$ ϵ-close to $|\tilde{P}\rangle$, where $\tilde{P} := \tilde{P}_1 \propto P_0 e^{-\tilde{L}}$. According to Ref. [5], in this procedure, we makes $O(l\log(\frac{l}{\epsilon}))$ uses of $\tilde{\tilde{R}}^{\omega}_{\tilde{\beta}_i L,\delta'',\epsilon''}$ with some $\delta'' = \Theta(\frac{\epsilon}{l\log(l/\epsilon)})$. The number of calls to O_ℓ in $\tilde{\tilde{R}}^{\omega}_{\tilde{\beta}_i L,\delta'',\epsilon''}$ is

$$O\left(\frac{\sigma}{\epsilon''\sqrt{\Delta_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{l}{\epsilon\sqrt{\Delta_{\min}}}\right)\times\log\left(\frac{l}{\epsilon}\right)\right) \tag{82}$$

because of Lemma 3, and multiplying $l\log(\frac{l}{\epsilon})$ to this and replacing l with its upper bound $l_{\max}$ yields the bound on the total query number in Eq. (77).

The statement on n_{anc} is obtained by substituting δ'' for δ and ϵ'' for ϵ' in Eq. (66).

Last, $\|\tilde{P} - P\|_{\text{TV}} \leqslant \epsilon$ is seen from Lemma 5 and Eq. (80). ■

D. Finding the credible interval

By the above method, we can get an approximation of the quantum state in which the target distribution P is encoded in amplitudes. However, in practice, our aim is not to get a quantum state but some statistics on P. Although it seems that the previous studies on quantum algorithms for MCMC have not focused on this point, this paper considers it. Concretely, as a quantity that we typically want, we consider the equal-tailed credible interval of a random variable that obeys P. Formally, writing $x \in \Omega \subset \mathbb{R}^d$ as $x = (x^{(1)}, \ldots, x^{(d)})$ and defining $\Omega^{(i)} := \{x^{(i)} | x \in \Omega\}$, we want $x^{(i)}_{\text{ub}}, x^{(i)}_{\text{lb}} \in \Omega^{(i)}$ that satisfy[2]

$$P\left(\left\{x_i > x^{(i)}_{\text{ub}}\right\}\right) = \frac{\alpha}{2}, \quad P\left(\left\{x_i < x^{(i)}_{\text{lb}}\right\}\right) = \frac{\alpha}{2} \tag{83}$$

with a credibility level $\alpha \in (0,1)$ for each $i \in [d]$. In other words, $x^{(i)}$ is in the interval $[x^{(i)}_{\text{lb}}, x^{(i)}_{\text{ub}}]$ with probability $1-\alpha$.

[2]In the current setting that Ω is discrete and so is each $\Omega^{(i)}$, $x^{(i)}_{\text{ub}}$ and $x^{(i)}_{\text{lb}}$ satisfying Eq. (83) may not exist. However, for simplicity, we now assume that such $x^{(i)}_{\text{ub}}$ and $x^{(i)}_{\text{lb}}$ exist. As long as the discretization is sufficiently fine as assumed in Sec. II A, it is reasonable to expect that there are $x^{(i)}_{\text{ub}}, x^{(i)}_{\text{lb}} \in \Omega^{(i)}$ such that $P(\{x_i > x^{(i)}_{\text{ub}}\})$ and $P(\{x_i < x^{(i)}_{\text{lb}}\})$ are much closer to $\frac{\alpha}{2}$ than the accuracy we require.

A typical example of this type of problem is parameter estimation by Bayesian inference: with P the posterior distribution of the parameters in some statistical model, we find the bound for each parameter in the above form.

Given the quantum circuit to approximately generate $|P\rangle$, a natural approach is finding $x^{(i)}_{\text{ub}}$ by binary search with the CDF $\Phi_P(a) := P(\{x^{(i)} > a\})$ computed by QMCI (and $x^{(i)}_{\text{lb}}$ is found in the same fashion). We hereafter elaborate this approach. First, we describe how to compute $\Phi_P(a)$.

Lemma 6. Suppose that Assumptions 1 through 6 are satisfied. Then, for any $i \in [d]$, $\epsilon, \delta \in (0,1)$, and $a \in \Omega^{(i)}$, there exists an algorithm that, with probability at least $1-\delta$, outputs an ϵ approximation of $\Phi_P(a)$, making

$$\begin{aligned} &O\left(\frac{\sigma l_{\max}}{\epsilon''\sqrt{\Delta_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{1}{\sqrt{\Delta_{\min}}}\right)\right. \\ &\quad\times\left(\log l_{\max} + \log L_{\max}\log\left(\frac{l_{\max}L_{\max}}{\delta}\right)\right) \\ &\quad+\frac{\sigma l_{\max}}{\epsilon\epsilon''\sqrt{\Delta_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{l_{\max}}{\epsilon\sqrt{\Delta_{\min}}}\right) \\ &\quad\left.\times\log^2\left(\frac{l_{\max}}{\epsilon}\right)\log\left(\frac{1}{\delta}\right)\right) \end{aligned} \tag{84}$$

uses of O_ℓ.

Proof. Because of Theorem 14, by Algorithm 2, we get $\beta_0, \ldots, \beta_l$, with which A$\frac{\pi}{3}$AA generates $|\widetilde{\tilde{P}_\star}\rangle$ $\frac{\epsilon}{9}$-close to $|\tilde{P}_\star\rangle|0\rangle^{n_{\text{anc}}}$ with n_{anc} of order (76). Here, $\tilde{P}_\star$ is a distribution on Ω such that $\|\tilde{P}_\star - P\|_{\text{TV}} \leqslant \frac{\epsilon}{3}$. We denote this A$\frac{\pi}{3}$AA by $\mathcal{V}_P$.

Note that $|\widetilde{\tilde{P}_\star}\rangle$ is written as follows:

$$|\widetilde{\tilde{P}_\star}\rangle = \sum_{x\in\Omega}\sqrt{\tilde{P}_\star(x)}|x\rangle|0\rangle^{\otimes n_{\text{anc}}} + \hat{\epsilon}\sum_{\hat{x}\in\hat{\Omega}}\sqrt{\hat{P}(\hat{x})}|\hat{x}\rangle|\psi_{\hat{x}}\rangle, \tag{85}$$

where $\hat{\epsilon} \in [0, \frac{\epsilon}{9})$, $\hat{\Omega}$ is a finite subset of $\mathbb{R}^d$ that may differ from Ω, $\hat{P}$ is a distribution on $\hat{\Omega}$, and $|\psi_{\hat{x}}\rangle$ is a state on n_{anc} ancillary qubits. $|\widetilde{\tilde{P}_\star}\rangle$ can be rewritten as

$$\begin{aligned} |\widetilde{\tilde{P}_\star}\rangle &= \sum_{x\in\Omega}|x\rangle\left(\sqrt{\tilde{P}_\star(x)}|0\rangle^{\otimes n_{\text{anc}}} + \hat{\epsilon}\sqrt{\hat{P}(x)}|\psi_x\rangle\right) \\ &\quad+\hat{\epsilon}\sum_{\hat{x}\in\hat{\Omega}\cap\overline{\Omega}}\sqrt{\hat{P}(\hat{x})}|\hat{x}\rangle|\psi_{\hat{x}}\rangle \\ &= \sum_{x\in\Omega}\sqrt{\tilde{P}'_\star(x)}|x\rangle|\tilde{\psi}_x\rangle + \hat{\epsilon}\sum_{\hat{x}\in\hat{\Omega}\cap\overline{\Omega}}\sqrt{\hat{P}(\hat{x})}|\hat{x}\rangle|\psi_{\hat{x}}\rangle. \end{aligned} \tag{86}$$

Here, $|\tilde{\psi}_x\rangle$ is a state on n_{anc} ancillary qubits and $\sqrt{\tilde{P}'_\star(x)} := \|\sqrt{\tilde{P}_\star(x)}|0\rangle^{\otimes n_{\text{anc}}} + \hat{\epsilon}\sqrt{\hat{P}(x)}|\psi_x\rangle\|$. Then, $\sum_{x\in\Omega}\tilde{P}'_\star(x) \leqslant 1$ follows from $\||\widetilde{\tilde{P}_\star}\rangle\| = 1$ and

$$\left|\sqrt{\tilde{P}'_\star(x)} - \sqrt{\tilde{P}_\star(x)}\right| \leqslant \hat{\epsilon}\sqrt{\hat{P}(x)} \tag{87}$$

follows from the triangle inequality.

On the other hand, we can implement a quantum circuit O_a^{comp} that acts as

$$O_a^{\mathrm{comp}}|x\rangle|0\rangle = |x\rangle|\mathbf{1}_{x>a}\rangle \tag{88}$$

using arithmetic circuits. Therefore, because of Theorem 11, the capability to generate $|\widetilde{\tilde{P}_\star}\rangle$ means that we can get an $\frac{\epsilon}{3}$ approximation $\tilde{\tilde{\Phi}}'(a)$ of

$$\tilde{\Phi}'(a) := \sum_{x\in\Omega} \tilde{P}'_\star(x)\mathbf{1}_{x>a} + \hat{\epsilon}^2 \sum_{\hat{x}\in\hat{\Omega}\cap\overline{\Omega}} \hat{P}(\hat{x})\mathbf{1}_{\hat{x}>a} \tag{89}$$

by QMCI with probability $1-\delta$. Let us see that $\tilde{\tilde{\Phi}}'(a)$ is an ϵ approximation of $\Phi(a) = \sum_{x\in\Omega} P(x)\mathbf{1}_{x>a}$. The difference between $\tilde{\Phi}'(a)$ and $\Phi(a)$ is bounded as

$$\begin{aligned}
|\tilde{\Phi}'(a) - \Phi(a)| &\leqslant |\tilde{\Phi}'(a) - \tilde{\Phi}(a)| + |\tilde{\Phi}(a) - \Phi(a)| \\
&\leqslant \sum_{x\in\Omega} \left|\tilde{P}'_\star(x) - \tilde{P}_\star(x)\right| \\
&\quad + \hat{\epsilon}^2 \sum_{\hat{x}\in\hat{\Omega}\cap\overline{\Omega}} \hat{P}(\hat{x}) + \frac{\epsilon}{3}.
\end{aligned} \tag{90}$$

Here, $\tilde{\Phi}(a) := \sum_{x\in\Omega} \tilde{P}_\star(x)\mathbf{1}_{x>a}$ and we used $|\tilde{\Phi}(a) - \Phi(a)| \leqslant \frac{\epsilon}{3}$ that follows from $\|\tilde{P}_\star - P\|_{\mathrm{TV}} \leqslant \frac{\epsilon}{3}$. The first term in Eq. (90) is bounded as

$$\begin{aligned}
&\sum_{x\in\Omega} |\tilde{P}'_\star(x) - \tilde{P}_\star(x)| \\
&= \sum_{x\in\Omega} |\sqrt{\tilde{P}'_\star(x)} - \sqrt{\tilde{P}_\star(x)}|(\sqrt{\tilde{P}'_\star(x)} + \sqrt{\tilde{P}_\star(x)}) \\
&\leqslant \sum_{x\in\Omega} \hat{\epsilon}\sqrt{\hat{P}(x)}(\sqrt{\tilde{P}'_\star(x)} + \sqrt{\tilde{P}_\star(x)}) \\
&\leqslant \hat{\epsilon}\left(\left(\sum_{x\in\Omega}\hat{P}(x)\right)^{1/2}\left(\sum_{x\in\Omega}\tilde{P}'_\star(x)\right)^{1/2}\right. \\
&\quad \left. + \left(\sum_{x\in\Omega}\hat{P}(x)\right)^{1/2}\left(\sum_{x\in\Omega}\tilde{P}_\star(x)\right)^{1/2}\right) \\
&\leqslant 2\hat{\epsilon} \\
&\leqslant \frac{2}{9}\epsilon,
\end{aligned} \tag{91}$$

where we use Eq. (87) at the first inequality and the Cauchy-Schwarz inequality at the second inequality. The second term in Eq. (90) is bounded as

$$\hat{\epsilon}^2 \sum_{\hat{x}\in\hat{\Omega}\cap\overline{\Omega}} \hat{P}(\hat{x}) \leqslant \hat{\epsilon}^2 \leqslant \hat{\epsilon} \leqslant \frac{\epsilon}{9}. \tag{92}$$

Consequently, $\tilde{\Phi}(a)$ is a $\frac{2}{3}\epsilon$ approximation of $\Phi(a)$, which means that $\tilde{\tilde{\Phi}}'(a)$ is an ϵ approximation of $\Phi(a)$.

Finally, let us check the query complexity. To get $\tilde{\beta}_0, \ldots, \tilde{\beta}_l$ with probability at least $1-\frac{\delta}{2}$ by Algorithm 2, we make a number of order (74) of calls to O_ℓ. After this, to get $\tilde{\tilde{\Phi}}'(a)$ with probability at least $1-\frac{\delta}{2}$ by QMCI, we call the circuit to generate $|\widetilde{\tilde{P}_\star}\rangle$ $O(\frac{1}{\epsilon}\log\delta^{-1})$ times, and one call to this circuit contains a number of order (77) of calls to O_ℓ. Summing up these, we see that the total number of calls to O_ℓ is of order (84). ■

Then, we reach the algorithm to find $x_{\mathrm{ub}}^{(i)}$.

Theorem 15. Suppose that Assumptions 1 through 6 are satisfied. Let $i \in [d]$, $\alpha \in (0,1)$, $\delta \in (0,1)$, and $\epsilon \in (0, \frac{\alpha}{2})$. Suppose that there exists $x^{(i)} \in \Omega^{(i)}$ such that

$$\frac{\alpha}{2} - \frac{\epsilon}{3} \leqslant \Phi(x^{(i)}) \leqslant \frac{\alpha}{2} + \frac{\epsilon}{3}. \tag{93}$$

Then, there exists an algorithm that, with probability at least $1-\delta$, outputs $\tilde{x}_{\mathrm{ub}}^{(i)} \in \Omega^{(i)}$ such that

$$\frac{\alpha}{2} - \epsilon \leqslant \Phi\big(\tilde{x}_{\mathrm{ub}}^{(i)}\big) \leqslant \frac{\alpha}{2} + \epsilon, \tag{94}$$

making

$$\begin{aligned}
&O\left(\frac{\sigma l_{\max}}{\epsilon''\sqrt{\Delta_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{1}{\sqrt{\Delta_{\min}}}\right)\right. \\
&\quad \times \left.\left(\log l_{\max} + \log L_{\max}\log\left(\frac{l_{\max}L_{\max}}{\delta}\right)\right)\right. \\
&\quad + \frac{\sigma l_{\max}}{\epsilon\epsilon''\sqrt{\Delta_{\min}}}\log^{3/2}\left(\frac{\sigma}{\epsilon''}\right)\log\log\left(\frac{\sigma}{\epsilon''}\right)\log\left(\frac{l_{\max}}{\epsilon\sqrt{\Delta_{\min}}}\right) \\
&\quad \times \left.\log^2\left(\frac{l_{\max}}{\epsilon}\right)\log\left(\frac{1}{\delta}\right)\log\left(\frac{\log|\Omega^{(i)}|}{\delta}\right)\log|\Omega^{(i)}|\right)
\end{aligned} \tag{95}$$

queries to O_ℓ.

For ϵ satisfying Eq. (78), Eq. (95) becomes

$$\tilde{O}\left(\frac{\sigma\bar{L}^{1/2}}{\epsilon^2\Delta_{\min}^{3/2}}\right). \tag{96}$$

Proof of Theorem 15. The algorithm is presented as Algorithm 3.

Then, let us show that this algorithm has a property stated in the theorem.

First, note that the loop in lines 11–21 ends in at most $n_{\max}$ iterations. To see this, denoting j_{ub} and j_{lb} at the end of the kth iteration by $j_{\mathrm{ub},k}$ and $j_{\mathrm{lb},k}$, respectively, we notice that

$$j_{\mathrm{ub},k+1} - j_{\mathrm{lb},k+1} \leqslant \frac{j_{\mathrm{ub},k} - j_{\mathrm{lb},k}}{2} + \frac{1}{2}, \tag{97}$$

which implies

$$j_{\mathrm{ub},k} - j_{\mathrm{lb},k} \leqslant 2^{-k}(n_i - 2) + 1. \tag{98}$$

Thus, $j_{\mathrm{ub},k} - j_{\mathrm{lb},k}$ becomes 2 or less in at least $\lceil\log_2(n_i-2)\rceil$ iterations, and, even if it becomes 2, the next iteration makes it 1. Therefore, the loop ends in $n_{\max}$ iterations by the condition $j_{\mathrm{ub}} - j_{\mathrm{lb}} = 1$, or earlier by the condition $|\tilde{\tilde{\Phi}}'(x_{j_{\mathrm{mid}}}^{(i)}) - \frac{\alpha}{2}| \leqslant \frac{2}{3}\epsilon$.

Let us consider the case that all the QMCIs in the algorithm, that in line 4 and those in the loop in lines 11–21, successfully outputs $\frac{\epsilon}{3}$ approximations of $\Phi(x_1^{(i)})$ and $\Phi(x_{j_{\mathrm{mid}}}^{(i)})$.

Algorithm 3. Algorithm to find $x_{\rm ub}^{(i)}$.

Require:
- Accuracy $\epsilon \in (0, 1)$
- Failure probability $\delta \in (0, 1)$
- Credibility level $\alpha \in (0, 1)$

1: By Algorithm 2, get $\tilde{\beta}_0, \ldots, \tilde{\beta}_l$, with which $\mathrm{A}\frac{\pi}{3}\mathrm{AA}$ generates a state $\frac{\epsilon}{27}$-close to $|\tilde{P}_\star\rangle|0\rangle^{\otimes n_{\rm anc}}$, where $n_{\rm anc}$ is of order (76) and $\tilde{P}_\star$ is a distribution on Ω satisfying $\|\tilde{P}_\star - P\|_{\rm TV} \leqslant \frac{\epsilon}{9}$.
2: Using the obtained $\tilde{\beta}_0, \ldots, \tilde{\beta}_l$, construct the above $\mathrm{A}\frac{\pi}{3}\mathrm{AA}$ as a quantum circuit $\mathcal{V}_P$.
3: Sort the elements of $\Omega^{(i)}$ in ascending order and denote them by $x_1^{(i)}, \ldots, x_{n_i}^{(i)}$, where $n_i := |\Omega^{(i)}|$.
4: By QMCI with $\mathcal{V}_P$, get an $\frac{\epsilon}{3}$ approximation $\tilde{\bar{\Phi}}'(x_1^{(i)})$ of $\Phi(x_1^{(i)})$ with failure probability $\delta' := \frac{\delta}{n_{\max}+1}$, where $n_{\max} := \lceil \log_2(n_i - 2) \rceil + 1$ (we do not need to compute $\Phi(x_{n_i}^{(i)})$ since it is zero).
5: **if** $|\tilde{\bar{\Phi}}'(x_1^{(i)}) - \frac{\alpha}{2}| \leqslant \frac{2}{3}\epsilon$ **then**
6: Output $x_1^{(i)}$ as $\tilde{x}_{\rm ub}^{(i)}$ and stop.
7: **else if** $\tilde{\bar{\Phi}}'(x_1^{(i)}) < \frac{\alpha}{2} - \frac{2}{3}\epsilon$ **then**
8: Stop with no output.
9: **end if**
10: Set $j_{\rm ub} = n_i$ and $j_{\rm lb} = 1$.
11: **repeat**
12: Set $j_{\rm mid} = \lceil \frac{j_{\rm ub}+j_{\rm lb}}{2} \rceil$.
13: By QMCI with $\tilde{V}_P$, get an $\frac{\epsilon}{3}$ approximation $\tilde{\bar{\Phi}}'(x_{j_{\rm mid}}^{(i)})$ of $\Phi(x_{j_{\rm mid}}^{(i)})$ with failure probability δ'.
14: **if** $|\tilde{\bar{\Phi}}'(x_{j_{\rm mid}}^{(i)}) - \frac{\alpha}{2}| \leqslant \frac{2}{3}\epsilon$ **then**
15: Output $x_{j_{\rm mid}}^{(i)}$ as $\tilde{x}_{\rm ub}^{(i)}$ and stop.
16: **else if** $\tilde{\bar{\Phi}}'(x_{j_{\rm mid}}^{(i)}) > \frac{\alpha}{2} + \frac{2}{3}\epsilon$ **then**
17: Set $j_{\rm lb} = j_{\rm mid}$.
18: **else** /* in this case, $\tilde{\bar{\Phi}}'(x_{j_{\rm mid}}^{(i)}) < \frac{\alpha}{2} - \frac{2}{3}\epsilon$ */
19: Set $j_{\rm ub} = j_{\rm mid}$.
20: **end if**
21: **until** $j_{\rm ub} - j_{\rm lb} = 1$
22: Output $x_{j_{\rm mid}}^{(i)}$.

This occurs with probability at least $(1 - \delta')^{n_{\max}+1} \geqslant \delta$. In these QMCIs, if we obtain $|\tilde{\bar{\Phi}}'(x^{(i)}) - \frac{\alpha}{2}| \leqslant \frac{2}{3}\epsilon$ for some $x^{(i)}$, $|\Phi(x^{(i)}) - \frac{\alpha}{2}| \leqslant \epsilon$ also holds because of $|\tilde{\bar{\Phi}}'(x^{(i)}) - \Phi(x^{(i)})| \leqslant \frac{\epsilon}{3}$. In fact, we get such $x^{(i)}$ with certainty under the condition that all the QMCIs succeed. This is seen by contradiction. Suppose that, under this condition, the loop in lines 11–21 ends with $j_{\rm ub} - j_{\rm lb} = 1$. This means that $\tilde{\bar{\Phi}}'(x_{j_{\rm lb}}^{(i)}) > \frac{\alpha}{2} + \frac{2}{3}\epsilon$ and $\tilde{\bar{\Phi}}'(x_{j_{\rm ub}}^{(i)}) < \frac{\alpha}{2} - \frac{2}{3}\epsilon$, which leads to

$$\Phi\left(x_{j_{\rm lb}}^{(i)}\right) > \frac{\alpha}{2} + \frac{\epsilon}{3}, \quad \Phi\left(x_{j_{\rm ub}}^{(i)}\right) < \frac{\alpha}{2} - \frac{\epsilon}{3}. \tag{99}$$

Since Φ is monotonically decreasing and there is no $\Omega^{(i)}$ element between $x_{j_{\rm lb}}^{(i)}$ and $x_{j_{\rm ub}}^{(i)}$, Eq. (99) contradicts with the assumption that Eq. (93) holds for some $x^{(i)} \in \Omega^{(i)}$.

In summary, with probability at least $1 - \delta$, $\tilde{x}_{\rm ub}^{(i)}$ satisfying Eq. (94) is output after either of the QMCIs.

The statement on the query complexity immediately follows from Lemma 6. The first term in Eq. (95) correspond to finding $\tilde{\beta}_0, \ldots, \tilde{\beta}_l$ and is similar to the first term in Eq. (84). The second term in Eq. (95) corresponds to QMCIs and is obtained by multiplying the number of QMCIs, which is of order $O(\log |\Omega^{(i)}|)$, to the second term in Eq. (84), and substituting δ' for δ. ■

Seemingly, the statement in Theorem 15 is tricky: it assumes the existence of $x^{(i)}$ for which $\Phi(x^{(i)})$ is $\frac{\epsilon}{3}$-close to $\frac{\alpha}{2}$, but only guarantees that the algorithm's output is ϵ-close. This is because of the erroneous nature of QMCI. Suppose that we search $x^{(i)}$ such that $|\Phi(x^{(i)}) - \frac{\alpha}{2}| \leqslant \epsilon$ and there exists $x^{(i)}$ that marginally satisfies this. Then, even if we require high accuracy in QMCI, it may output an estimate of $\Phi(x^{(i)})$ out of the ϵ neighborhood of $\frac{\alpha}{2}$, which makes us fail to notice that $x^{(i)}$ is what we want. We thus conduct QMCIs with accuracy $\frac{\epsilon}{3}$ and pick up $x^{(i)}$ with $\tilde{\bar{\Phi}}'(x^{(i)})$ $\frac{2}{3}\epsilon$-close to $\frac{\alpha}{2}$ as an answer. Under this policy, we never miss $x^{(i)}$ satisfying Eq. (93), since the $\frac{\epsilon}{3}$ approximation of $\Phi(x^{(i)})$ is never out of the $\frac{2}{3}\epsilon$ neighborhood of $\frac{\alpha}{2}$. Of course, we might pick up $x^{(i)}$ for which $|\Phi(x^{(i)}) - \frac{\alpha}{2}| > \frac{2}{3}\epsilon$, given the erroneous QMCI estimate of $\Phi(x^{(i)})$ accidentally lying in the $\frac{2}{3}\epsilon$ neighborhood of $\frac{\alpha}{2}$. Even if so, the chosen $x^{(i)}$ at least satisfies $|\Phi(x^{(i)}) - \frac{\alpha}{2}| \leqslant \epsilon$, since the $\frac{\epsilon}{3}$ approximation of a number distant from $\frac{\alpha}{2}$ by more than ϵ never lies in the $\frac{2}{3}\epsilon$ neighborhood of $\frac{\alpha}{2}$.

A similar discussion is found in consideration on setting the threshold of the SNR in the quantum algorithm for GW matched filtering proposed in Ref. [23].

E. Comparison with other approaches

We now make a comparison of the above method for finding the credible interval with other approaches. We compare the order of the number of queries to O_ℓ in the various approaches except logarithmic factors. Since the binary search adds only logarithmic factors, it is sufficient to consider the complexity of calculating the CDF within accuracy ϵ.

First, let us consider QSA *without* QMCI. That is, the state $|P\rangle$ that encodes the target distribution P is prepared via Algorithm 2 and $\mathrm{A}\frac{\pi}{3}\mathrm{AA}$ with the obtained $\{\beta_i\}$, with L computed by not QMCI but M-time iterated calculations and additions of ℓ. Then, using this state-preparing circuit as $\mathcal{V}_P$, we estimate $x_{\rm ub}^{(i)}$ by Algorithm 3. We call this the exact QSA approach. Note that the quantum walk operator U, which is now the exact one in Eq. (16), makes $O(M)$ calls to O_ℓ. Combining this with Eq. (34), we see that the number of calls to O_ℓ in generating $|P\rangle$ by QSA is

$$\tilde{O}\left(\frac{M\bar{L}^{1/2}}{\Delta_{\min}^{1/2}}\right). \tag{100}$$

Besides, considering the complexity of QMCI in Eq. (35), we estimate the total number of calls to O_ℓ in finding a credible interval in the exact QSA approach as

$$\tilde{O}\left(\frac{M\bar{L}^{1/2}}{\Delta_{\min}^{1/2}\epsilon}\right). \tag{101}$$

Next, let us consider the fully classical approach: on a classical computer, generating the Markov chain by the MH method in Algorithm 1, with L obtained by M-time iterative calculations. We now regard O_ℓ as a classical subroutine to compute ℓ. Based on the bound (11) on the step number in MCMC-based expectation estimation, the total number of

calls to O_ℓ in finding the credible interval is

$$\tilde{O}\left(\frac{M}{\Delta\epsilon^2}\right). \tag{102}$$

Also note that we do not need binary search with respect to the CDF in the classical approach. We can store the sampled states on a classical memory, and thus, sorting them and taking the $100(1-\frac{\alpha}{2})$th percentile yields an estimate on $x_{\rm ub}^{(i)}$.

Note that the complexity of the proposed method in Eq. (96) is not better than those of the exact QSA approach and the fully classical approach in Eqs. (101) and (102) with respect to the spectral gap and accuracy. On the other hand, unlike Eqs. (101) and (102), the complexity of the proposed method is not explicitly dependent on M, the number of terms in L, but on σ, the standard deviation of ℓ. Thus, the proposed method can be advantageous with respect to M if σ scales with M sublinearly, and this actually holds in the case of GW parameter estimation considered in Sec. IV.

IV. APPLICATION: PARAMETER ESTIMATION IN GRAVITATIONAL WAVE DETECTION EXPERIMENTS

As an application of the credible interval calculation method proposed above, we consider parameter estimation in GW experiments. Since the first detection in 2015 [14], GW events have been detected by laser interferometers such as LIGO and Virgo [15–17]. Given a GW event, we want to estimate the parameters of the GW, such as masses of the sources for a GW from a compact binary coalescence (CBC). For this purpose, Bayesian inference with MCMC is widely used (for a review, see Ref. [13]). This is a time-consuming task and thus a target of quantum speedup. Although a previous paper [26] considered the application of the quantum MH method to this task, unlike that paper focusing on the simulation-based comparison of the TTS between the existing classical and quantum MH methods, our aim is now the application of the QMCI-based method proposed above.

The problem is formulated as follows. Given the detector output $s(t)$ as time-series data with time length T and interval Δt, the negative log-likelihood for a point x in the parameter space is given as follows:

$$\begin{aligned} L(x) &= -2\mathrm{Re}(h(\cdot,x)|s) + (h(\cdot,x)|h(\cdot,x)) + C, \\ \mathrm{Re}(h(\cdot,x)|s) &= \frac{4}{M}\sum_{k=1}^{\frac{M}{2}-1} \Re\left(\frac{\tilde{h}^*(f_k;x)\tilde{s}(f_k)}{S_{\rm n}(f_k)\Delta t}\right), \\ (h(\cdot,x)|h(\cdot,x)) &= \frac{4}{M}\sum_{k=1}^{\frac{M}{2}-1} \frac{|\tilde{h}(f_k;x)|^2}{S_{\rm n}(f_k)\Delta t}. \end{aligned} \tag{103}$$

Here, $M=\frac{T}{\Delta t}$, $f_k := \frac{k}{T}$, the tilde represents the Fourier transform of a function of time, $h(t,x)$ is the GW waveform for x, $S_{\rm n}$ is the single-sided power spectrum density of the noise, and C is a term independent of x. Since $\tilde{h}(\cdot,x)$ and $S_{\rm n}$ are smooth functions evaluated by explicit formulas, we assume that $(h(\cdot,x)|h(\cdot,x))$ is approximated by the integral $4\int_0^\infty \frac{|\tilde{h}(f;x)|^2}{S_{\rm n}(f)}df$ and this is further approximated by some formula efficiently computable by arithmetic circuits. Then, L in Eq. (103) is in the form of Eq. (48). In fact, M can be as large as 10^6–10^{10} in typical cases [23], and thus we are motivated to apply our QMCI-based method in Sec. III to find credible intervals for GW parameters, regarding $-2\mathrm{Re}(h(\cdot,x)|s(t))$ as $L_{\rm sum}$ and $-4\mathrm{Re}(\frac{\tilde{h}^*(f_k;x)\tilde{s}(f_k)}{S_{\rm n}(f_k)\Delta t})$ as $\ell(k,x)$.

Note that other conditions to apply the proposed method are met. Usually, we have found a high SNR point in the parameter space by matched filtering conducted prior to parameter estimation, and thus we can set a parameter region to be searched, for example, a hyperrectangle around such a point. We can set Ω to the sufficiently dense discrete points in that region. Commonly, the prior distribution P_0 is set to uniform on Ω and the proposal distribution $T(x,\cdot)$ is set to some easy-to-sample one such as the normal distribution around x, which means Assumptions 2, 3, and 4 are satisfied. On the other hand, since the detector output is affected by the random noise and unable to be expressed as an analytic formula, O_ℓ is not implemented as a combination of arithmetic circuits. Nevertheless, if we assume the availability of quantum random access memory (QRAM) [58], we can implement O_ℓ using a QRAM that stores the values of $\tilde{s}(f_k)$, and thus Assumption 1 is satisfied. The preparation of such a QRAM takes $O(M)$ time, but this is needed only once at the very beginning of calculation.

Let us estimate the query complexity of credible interval calculation for GW parameters by the proposed method. To do so, we need to bound the variance σ^2 of terms in $\mathrm{Re}(h(\cdot,x)|s(t))$. According to Ref. [23], $\sigma = O(\gamma M^{1/2})$ with

$$\gamma := \max_{\substack{x\in\Omega \\ k\in[\frac{M}{2}-1]}} \frac{\tilde{h}(f_k,x)}{\sqrt{S_{\rm n}(f_k)\Delta t}}, \tag{104}$$

which is $O(1)$ in some cases if h is normalized so that $(h(\cdot,x)|h(\cdot,x))=1$ as in matched filtering [23]. If we get the highest SNR $\rho=(h(\cdot,x_\star)|s)$ at the parameter point $x_\star$ with $(h(\cdot,x_\star)|h(\cdot,x_\star))=1$ in matched filtering, Ω should be set in the neighborhood of the parameter that corresponds to the waveform $\rho h(\cdot,x_\star)$, which leads to $\gamma=O(\rho)$ and $\sigma = O(\rho M^{1/2})$. Then, Eq. (96) becomes

$$\tilde{O}\left(\frac{M^{1/2}\rho\bar{L}}{\epsilon^2\Delta_{\rm min}^{3/2}}\right). \tag{105}$$

Compared to the complexity of the exact QSA approach in Eq. (101), the proposed method provides the quadratic speedup with respect to M, in compensation for the worse scaling on ϵ and $\Delta_{\rm min}$.

V. SUMMARY

In this paper, with the usage in Bayesian inference in mind, we have considered the quantum version of the MH algorithm in the case that the target probability P is in the form of Eq. (47) and L is given as Eqs. (48) and (49) with large M, based on QSA. In such a case, calculating L takes the $O(M)$ query complexity naively, and thus we have proposed application of QMCI, which may speed up a costly summation. We have presented not only the procedure to generate the state that encodes P but also that for finding a credible interval of a parameter in a statistical model. Setting the accuracy in QMCI based on the result in Ref. [30] on the MH algorithm with

the perturbed acceptance ratio, we have derived the bound on the complexity, the number of calls to the quantum circuit to compute ℓ, as summarized in Table I. Comparing QSA with L calculated exactly, the complexity of the proposed method scales worse on the required accuracy ϵ and the spectral gap $\Delta_{\min}$. On the other hand, if σ, the standard deviation of ℓ, scales on M sublinearly, the proposed method is advantageous with respect to M. As an example in which this holds, we have considered estimation of GW parameters in a GW detection experiment. In this example, σ scales on M as $O(\sqrt{M})$ and this results in the complexity shown in Table I, which is quadratically smaller with respect to M compared to the exact QSA method.

One future direction of this work is searching other applications of the proposed framework, the combination of QSA and QMCI. The sublinear scaling of σ on M in GW parameter estimation, which makes the proposed method beneficial, stems from the nature of the problem: the random noise in the detector output overwhelming the GW signal and the power spectrum of this noise [23]. Thus, if a similar situation holds for other problems in the fields such as scientific experimental data analysis or signal processing, the proposed method might be also beneficial to them.

This work is a theoretical one and we have not created any code.

ACKNOWLEDGMENTS

This work is supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant No. JPMXS0120319794, JSPS KAKENHI Grant No. JP22K11924, and JST COI-NEXT program Grant No. JPMJPF2014.

APPENDIX A: PROPERTIES OF THE WALK OPERATOR

Here, we present the proof of Theorem 13 on the spectrum of the quantum walk operator. That of Theorem 4 is almost the same with S seen as I.

First, we present the following theorem on which our proof is based.

Theorem 16 (Theorem 1 in Ref. [10]). Let $\mathcal{H}$ be an N-dimensional Hilbert space. Let $\mathcal{A}$ ($\mathcal{B}$) be an n-dimensional subspace of $\mathcal{H}$ spanned by orthonormal vectors $u_1, \ldots, u_m$ ($v_1, \ldots, v_n$). Denote by $V_{\mathcal{A}}$ ($V_{\mathcal{B}}$) the $N \times m$ ($N \times n$) matrix whose ith column is u_i (v_i). Define $R_{\mathcal{A}} = 2V_{\mathcal{A}}V_{\mathcal{A}}^{\dagger} - I$ and $R_{\mathcal{B}} = 2V_{\mathcal{B}}V_{\mathcal{B}}^{\dagger} - I$. Then, on $\mathcal{A} + \mathcal{B}$, the unitary operator $R_{\mathcal{A}}R_{\mathcal{B}}$ has an eigenvalue 1 with multiplicity 1, and any other eigenvalue is either of $e^{2i\theta_1}, e^{-2i\theta_1}, \ldots, e^{2i\theta_l}, e^{-2i\theta_l}$ or -1, where $\theta_1, \ldots, \theta_l \in (0, \frac{\pi}{2})$ are written as $\theta_i = \arccos \lambda_i$ with singular values $\{\lambda_i\}$ of $V_{\mathcal{A}}^{\dagger}V_{\mathcal{B}}$ that lie in (0,1).

In the current case, $\mathcal{A}$ and $\mathcal{B}$ are defined as Eq. (46).

We also use the following lemmas.

Lemma 7. On $\mathcal{A}$, $\Pi_0 V^{\dagger}B^{\dagger}SFBV\Pi_0$ has the same eigenvalues as W including multiplicity, where Π_0 is the projector onto $\mathcal{A}$.

Proof. By a straightforward calculation, we see that, for any $x \in \Omega$, applying $\Pi_0 V^{\dagger}B^{\dagger}SFBV$ to $|x\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}$ yields

$$\left[\sum_{\Delta x \in \Omega_x \setminus \{\vec{0}_d\}} \sqrt{T(x, x+\Delta x)T(x+\Delta x, x)A(x, x+\Delta x)A(x+\Delta x, x)}|x+\Delta x\rangle_{R_S} + \left(1 - \sum_{\Delta x \in \Omega_x \setminus \{\vec{0}_d\}} T(x, x+\Delta x)A(x, x+\Delta x)\right)|x\rangle_{R_S}\right]|0\rangle_{R_M}|0\rangle_{R_C} = \sum_{y\in\Omega}\sqrt{W_{x,y}W_{y,x}}|y\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}, \tag{A1}$$

where $\vec{0}_d$ is the d-dimensional zero vector. This means that

$$\Pi_0 V^{\dagger}B^{\dagger}SFBV\Pi_0 = \sum_{x,y\in\Omega}\sqrt{W_{x,y}W_{y,x}}|y\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\langle x|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C}. \tag{A2}$$

Using the detailed balance condition $P(x)W_{x,y} = P(y)W_{y,x}$, which is satisfied in the MH algorithm (Exercise 3.1 of Ref. [8]), we have

$$\begin{aligned}\Pi_0 V^{\dagger}B^{\dagger}SFBV\Pi_0 &= \sum_{x,y\in\Omega}\sqrt{\frac{P(x)}{P(y)}}W_{x,y}|y\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\langle x|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C} \\ &= \sum_{x,y\in\Omega}(D_P W D_P^{-1})_{x,y}|y\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\langle x|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C},\end{aligned} \tag{A3}$$

where D_P is a diagonal matrix indexed by $x, y \in \Omega$ and its (x, x) entry is $\sqrt{P(x)}$. Thus, since $\Pi_0 V^{\dagger}B^{\dagger}SFBV\Pi_0$ is expressed as the conjugation of W by D_P, it has the same eigenvalues as W on $\mathcal{A}$. ■

Lemma 8. $|P\rangle$ is the eigenstate of $U_W = RV^{\dagger}B^{\dagger}SFBV$ with eigenvalue 1.

Proof. This is shown by a straightforward calculation. Applying FBV to $|P\rangle = \sum_{x\in\Omega}\sqrt{P(x)}|x\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}$ yields

$$\sum_{x\in\Omega}\sum_{\Delta x\in\Omega_x}\left(\sqrt{P(x)T(x, x+\Delta x)A(x, x+\Delta x)}|x+\Delta x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C} + \sqrt{P(x)T(x, x+\Delta x)(1 - A(x, x+\Delta x))}|x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C}\right). \tag{A4}$$

By using the detailed balance condition

$$P(x)T(x,x+\Delta x)A(x,x+\Delta x)=P(x+\Delta x)T(x+\Delta x,x)A(x+\Delta x,x) \tag{A5}$$

and substituting Δx and $x+\Delta x$ with $-\Delta x$ and x, respectively, in the first term in Eq. (A4), we get

$$\sum_{x\in\Omega}\sum_{\Delta x\in\Omega_x}\Big(\sqrt{P(x)T(x,x+\Delta x)A(x,x+\Delta x)}|x\rangle_{R_S}|-\Delta x\rangle_{R_M}|1\rangle_{R_C} + \sqrt{P(x)T(x,x+\Delta x)(1-A(x,x+\Delta x))}|x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C}\Big), \tag{A6}$$

and, by applying S to this, we obtain

$$\sum_{x\in\Omega}\sum_{\Delta x\in\Omega_x}\sqrt{P(x)T(x,x+\Delta x)}|x\rangle_{R_S}|\Delta x\rangle_{R_M} \otimes\big(\sqrt{A(x,x+\Delta x)}|1\rangle_{R_C}+\sqrt{(1-A(x,x+\Delta x))}|0\rangle_{R_C}\big). \tag{A7}$$

Thus, applying $V^\dagger B^\dagger$ to this yields $|P\rangle$. Applying R at last does not change $|P\rangle$. ■

Lemma 9. The restriction of $\Pi_0 V^\dagger B^\dagger SFBV\Pi_0$ to $\mathcal{A}$ is equal to $V_\mathcal{A}^\dagger V_\mathcal{B}$.

Proof. Label the elements in Ω with integers $1,\dots,|\Omega|$ and denote the kth element by x_k. Then, for $k,l\in[|\Omega|]$, the (k,l) entry of the restriction of $\Pi_0 V^\dagger B^\dagger SFBV\Pi_0$ to $\mathcal{A}$ is

$$\begin{aligned}&\langle x_k|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C}\Pi_0 V^\dagger B^\dagger SFBV\Pi_0|x_l\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\\ &\quad=\langle x_k|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C}V^\dagger B^\dagger SFBV|x_l\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}.\end{aligned} \tag{A8}$$

From the definitions of $V_\mathcal{A}$ and $V_\mathcal{B}$, we see that this is also the (k,l) entry of $V_\mathcal{A}^\dagger V_\mathcal{B}$. ■

Lemma 10. On $\mathcal{A}+\mathcal{B}$

$$(SF)^\dagger=SF. \tag{A9}$$

Proof. For any $x\in\Omega$ and $\Delta x\in\Omega_x$,

$$SFSF|x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C}=|x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C}, \tag{A10}$$

since both S and F are not activated if the state on R_C is $|0\rangle_{R_C}$. Besides, applying F, S, F, and S to $|x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C}$ in this order transforms the state as

$$\begin{aligned}&|x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C}\\ &\xrightarrow{F}|x+\Delta x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C}\\ &\xrightarrow{S}|x+\Delta x\rangle_{R_S}|-\Delta x\rangle_{R_M}|1\rangle_{R_C}\\ &\xrightarrow{F}|x\rangle_{R_S}|-\Delta x\rangle_{R_M}|1\rangle_{R_C}\\ &\xrightarrow{S}|x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C}.\end{aligned} \tag{A11}$$

Thus, we see that $SFSF$ acts as I for any state in the form of $|x\rangle_{R_S}|\Delta x\rangle_{R_M}|\phi\rangle_{R_C}$, where $|\phi\rangle_{R_C}$ is any state on R_C. This means that $SFSF=I$ and thus $(SF)^\dagger=SF$ on $\mathcal{A}+\mathcal{B}$. ■

Then, combining these lemmas, we can prove Theorem 13.

Proof of Theorem 13. Combining Theorem 16 with Lemmas 7 and 9, we see that $R_\mathcal{A}R_\mathcal{B}$ has eigenvalue 1 with multiplicity 1 and that any other eigenvalue is -1 or in the form of $\exp(\pm 2i\theta_l)$, where $\theta_1,\theta_2,\dots\in(0,\frac{\pi}{2})$ are written as $\theta_l=\arccos|\lambda_l|$ with $\{\lambda_l\}$, the eigenvalues of W with modulus less than 1.

On the other hand, $R_\mathcal{A}$ and $R_\mathcal{B}$ are now

$$R_\mathcal{A}=2\sum_{x\in\Omega}|x\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\langle x|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C}-I \tag{A12}$$

and

$$\begin{aligned}R_\mathcal{B}&=2\sum_{x\in\Omega}V^\dagger B^\dagger SFBV|x\rangle_{R_S}|0\rangle_{R_M}|0\rangle_{R_C}\langle x|_{R_S}\langle 0|_{R_M}\langle 0|_{R_C}\\ &\quad\times(V^\dagger B^\dagger SFBV)^\dagger-I\\ &=V^\dagger B^\dagger SFBVR_\mathcal{A}V^\dagger B^\dagger SFBV,\end{aligned} \tag{A13}$$

respectively. In Eq. (A13), we used Lemma 10. Note that, on $\mathcal{A}+\mathcal{B}$, R in Eq. (21) acts as $R_\mathcal{A}$ and $V^\dagger B^\dagger SFBVRV^\dagger B^\dagger SFBV$ acts as $R_\mathcal{B}$. Thus, $RV^\dagger B^\dagger SFBVRV^\dagger B^\dagger SFBV=U^2$ acts as $R_\mathcal{A}R_\mathcal{B}$. Therefore, on $\mathcal{A}+\mathcal{B}$, the eigenvalues of U are equal to the square root of those of $R_\mathcal{A}R_\mathcal{B}$. They include 1 or -1 with multiplicity 1, and, because of Lemma 8, it is in fact 1 with the corresponding eigenstate $|P\rangle$. Any other eigenvalue of U_W is $e^{\pm i\theta_l}$, $-e^{\pm i\theta_l}=e^{i(\pm\theta_l+\pi)}$, or $\pm i=e^{\pm\frac{\pi}{2}i}$, whose phase has modulus no less than

$$\arccos(\max\{|\lambda_l|\})=\arccos(1-\Delta) \tag{A14}$$

in any case. ■

APPENDIX B: DETAILS OF QUANTUM MONTE CARLO INTEGRATION

First, let us recall Theorem 5 in Ref. [23].

Theorem 17 (Theorem 5 in Ref. [23], modified). Let $M\in\mathbb{N}$ and $\mathcal{X}$ be a set of M real numbers, $X_0,\dots,X_{M-1}$, whose mean is $\mu:=\frac{1}{M}\sum_{i=0}^{M-1}X_i$ and sample variance satisfies $\frac{1}{M}\sum_{i=0}^{M-1}X_i^2-\mu^2\leqslant\sigma^2$ with some $\sigma\in\mathbb{R}_+$. Suppose that we are given an access to a unitary operator O_X that acts as Eq. (36) for any $i\in[M]_0$. Let $\epsilon\in(0,4\sigma)$ and $\delta\in(0,1)$. Then, we have an access to a unitary operator $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ that acts on a system of two registers as

$$\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}|0\rangle|0\rangle=\sum_{y\in\mathcal{Y}}\alpha_y|\phi_y\rangle|y\rangle. \tag{B1}$$

Here, $\mathcal{Y}$ is a finite set of real numbers that includes a subset $\tilde{\mathcal{Y}}$ consisting of ϵ approximations of μ, $\{\alpha_y\}_{y\in\mathcal{Y}}$ are complex numbers satisfying $\sum_{\tilde{y}\in\tilde{\mathcal{Y}}}|\alpha_{\tilde{y}}|^2\geqslant 1-\delta$, and $\{|\phi_y\rangle\}_{y\in\mathcal{Y}}$ are states on the first register. In $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$, queries to $O_\mathcal{X}$ are made, whose number is of order (38). $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ uses qubits whose number is of order (39).

We construct $O^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ in Theorem 12 using $\tilde{O}^{\mathrm{mean}}_{\mathcal{X},\epsilon,\delta,\sigma}$ and adding some operations afterward.

Proof of Theorem 12. Any $x \in \mathbb{R}$ can be written as $x = \sum_{i=-\infty}^{i=\infty} x_i 2^i$, where $\{x_i\}_{i\in\mathbb{Z}}$ are binaries (0 or 1), that is, the binary representation of x. We call x_i the ith bit of x. For $x \in \mathbb{R}$ and $a \in \mathbb{Z}$, we define

$$\lfloor x \rfloor_a := \sum_{i=a}^{\infty} 2^i x_i. \tag{B2}$$

Namely, $\lfloor x \rfloor_a$ is the rounding of x at the ath bit. We denote by O_a^{round} the operator for rounding: $O_a^{\text{round}}|x\rangle|0\rangle = |x\rangle|\lfloor x \rfloor_a\rangle$. This is simply implemented by copying the higher-order qubits in the first register to the second register with controlled-NOT gates.

Then, we can perform the following operation:

$$\begin{aligned} |0\rangle|0\rangle|0\rangle &\to \sum_{y\in\mathcal{Y}} \alpha_y |\phi_y\rangle|y\rangle|0\rangle \\ &\to \sum_{y\in\mathcal{Y}} \alpha_y |\phi_y\rangle|y\rangle|\lfloor y \rfloor_b\rangle =: |\Phi\rangle. \end{aligned} \tag{B3}$$

Here, we use $\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma}$ at the first arrow and O_b^{round} at the second arrow, where $b = \lfloor \log_2 \epsilon \rfloor$, $\epsilon' := 2^{b-1}$, and $\delta' := \delta/4$. $\mathcal{Y}$ is a finite set of real numbers that has a subset $\tilde{\mathcal{Y}}$ consisting of ϵ' approximations of μ and the complex numbers $\{\alpha_y\}_{y\in\mathcal{Y}}$ satisfy $\sum_{\tilde{y}\in\tilde{\mathcal{Y}}} |\alpha_{\tilde{y}}|^2 \geqslant 1 - \delta'$. Note that, for any ϵ' approximation y of μ, $y_b, y_{b+1}, \ldots$ and $\mu_b, \mu_{b+1}, \ldots$ are equal, respectively, since any discrepancy in the bth or higher-order bits means that $|y - \mu| \geqslant 2^b > \epsilon'$. Thus, we have

$$|\Phi\rangle = \left(\sum_{y\in\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle\right) \otimes |\lfloor \mu \rfloor_b\rangle + \sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle \otimes |\lfloor y \rfloor_b\rangle. \tag{B4}$$

Therefore, letting $|\tilde{\Phi}\rangle := (\sum_{y\in\mathcal{Y}} \alpha_y |\phi_y\rangle|y\rangle) \otimes |\lfloor \mu \rfloor_b\rangle$, we have

$$\begin{aligned} &\||\tilde{\Phi}\rangle - |\Phi\rangle\| \\ &= \left\| \sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle|\lfloor \mu \rfloor_b\rangle - \sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle \otimes |\lfloor y \rfloor_b\rangle \right\| \\ &\leqslant \left\| \sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle|\lfloor \mu \rfloor_b\rangle \right\| + \left\| \sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} \alpha_y |\phi_y\rangle|y\rangle \otimes |\lfloor y \rfloor_b\rangle \right\| \\ &\leqslant 2\sqrt{\sum_{y\in\mathcal{Y}\setminus\tilde{\mathcal{Y}}} |\alpha_y|^2} \\ &\leqslant 2\sqrt{\delta'} \\ &= \sqrt{\delta}. \end{aligned} \tag{B5}$$

This means that we can write

$$\begin{aligned} |\Phi\rangle &= |\tilde{\Phi}\rangle + \gamma|\tilde{\psi}\rangle \\ &= \left(\sum_{y\in\mathcal{Y}} \alpha_y |\phi_y\rangle|y\rangle\right) \otimes |\lfloor \mu \rfloor_b\rangle + \gamma|\tilde{\psi}\rangle, \end{aligned} \tag{B6}$$

where $\gamma := \||\Phi\rangle - |\tilde{\Phi}\rangle\| \leqslant \sqrt{\delta}$ and $|\tilde{\psi}\rangle := \frac{1}{\gamma}(|\Phi\rangle - |\tilde{\Phi}\rangle)$. Then, performing $(\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma})^\dagger$ on the first and second registers transforms $|\Phi\rangle$ to

$$|0\rangle|0\rangle|\lfloor \mu \rfloor_b\rangle + \gamma|\psi\rangle, \tag{B7}$$

where $|\psi\rangle$ is a state on the entire system. Since $\lfloor \mu \rfloor_b$ is an ϵ approximation of μ, we see that the above operation yields a state in the form of Eq. (37), with the first and second registers together seen as R_1 and the third one seen as R_2.

The number of queries to $O_{\mathcal{X}}$ in the entire process is that in $\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma}$ and $(\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma})^\dagger$, that is, the double of that in $\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma}$, which is of order (38) since $\epsilon' = \Theta(\epsilon)$ and $\delta' = \Theta(\delta)$. The number of qubits used in the entire process is also the same as $\tilde{O}^{\text{mean}}_{\mathcal{X},\epsilon',\delta',\sigma}$, and is of order (39). ■

APPENDIX C: PROOF OF LEMMA 4

We use the following theorem.

Theorem 18 (Theorem IIIa in Ref. [59]). Let $n \in \mathbb{N}$ and $B, \tilde{B} \in \mathbb{C}^{n\times n}$. Assume that B is diagonalizable and denote by Q the matrix that diagonalizes B: $Q^{-1}BQ$ is diagonal. Denote by κ the condition number of Q. Then, for each eigenvalue λ of B, there exists an eigenvalue $\tilde{\lambda}$ of $\tilde{B}$ that satisfies

$$|\tilde{\lambda} - \lambda| \leqslant \kappa \|B - \tilde{B}\|. \tag{C1}$$

We also use the following lemma.

Lemma 11. Define A as Eq. (1) with P in the form of Eq. (47), and $\tilde{A}$ as Eq. (55). Then, if $\epsilon := \max_{x\in\Omega} |\tilde{L}(x) - L(x)| \leqslant \frac{1}{4}$,

$$|\tilde{A}(x, y) - A(x, y)| \leqslant 8\epsilon \tag{C2}$$

holds for any $x, y \in \Omega$.

Proof. We consider the following two cases.

Case (i). For $x, y \in \Omega$ such that $\frac{P(y)T(y,x)}{P(x)T(x,y)} \leqslant 2$. Note that

$$|e^a - 1| \leqslant 2|a| \tag{C3}$$

holds for any $a \in [-1, 1]$. Since

$$\begin{aligned} &|(\tilde{L}(x) - L(x)) - (\tilde{L}(y) - L(y))| \\ &\quad\leqslant |\tilde{L}(x) - L(x)| + |\tilde{L}(y) - L(y)| \\ &\quad\leqslant 2\epsilon \\ &\quad\leqslant 1, \end{aligned} \tag{C4}$$

we have

$$\begin{aligned} &\left| \frac{P_0(y)e^{-\tilde{L}(y)}T(y,x)}{P_0(x)e^{-\tilde{L}(x)}T(x,y)} - \frac{P(y)T(y,x)}{P(x)T(x,y)} \right| \\ &\quad= \left| \frac{P(y)T(y,x)}{P(x)T(x,y)} e^{(\tilde{L}(x)-L(x))-(\tilde{L}(y)-L(y))} - \frac{P(y)T(y,x)}{P(x)T(x,y)} \right| \\ &\quad\leqslant 2|e^{(\tilde{L}(x)-L(x))-(\tilde{L}(y)-L(y))} - 1| \\ &\quad\leqslant 8\epsilon. \end{aligned} \tag{C5}$$

Since $\min\{1, \cdot\}$ is a 1-Lipschitz function on $\mathbb{R}$, we obtain Eq. (C2).

Case (ii). For $x, y \in \Omega$ such that $\frac{P(y)T(y,x)}{P(x)T(x,y)} > 2$. Because of Eq. (C4),

$$e^{(\tilde{L}(y)-L(y))-(\tilde{L}(x)-L(x))} \geqslant e^{-2\epsilon} \geqslant e^{-\frac{1}{2}} \geqslant \frac{1}{2} \tag{C6}$$

holds, which means

$$\frac{P_0(y)e^{-\tilde{L}(y)}T(y,x)}{P_0(x)e^{-\tilde{L}(x)}T(x,y)} = \frac{P(y)T(y,x)}{P(x)T(x,y)}e^{(\tilde{L}(x)-L(x))-(\tilde{L}(y)-L(y))} \geqslant 1 \tag{C7}$$

and thus $\tilde{A}(x,y) = 1 = A(x,y)$.

Thus, in both cases, Eq. (C2) holds. ■

Then, Lemma 4 is proven as follows.

Proof of Lemma 4. Theorem 18 implies that

$$\tilde{\Delta} \geqslant \Delta - \kappa\|\delta W\|, \tag{C8}$$

where $\delta W := \tilde{W} - W$ and W ($\tilde{W}$) is the transition matrix of $\mathcal{C}_L$ ($\mathcal{C}_{\tilde{L}}$).

Then, let us bound $\|\delta W\|$. To do this, we use a well-known inequality (Corollary 2.3.2 or Ref. [60]):

$$\|\delta W\| \leqslant \sqrt{\|\delta W\|_1\|\delta W\|_\infty}. \tag{C9}$$

We also have

$$\begin{aligned}\|\delta W\|_\infty &= \max_{x\in\Omega}\sum_{y\in\Omega}|\delta W_{x,y}| \\ &= \max_{x\in\Omega}\left(|\delta W_{x,x}| + \sum_{y\in\Omega\setminus\{x\}}|\delta W_{x,y}|\right) \\ &= \max_{x\in\Omega}\left(\left|\sum_{y\in\Omega\setminus\{x\}}T(x,y)(A(x,y)-\tilde{A}(x,y))\right|\right. \\ &\quad \left. + \sum_{y\in\Omega\setminus\{x\}}|T(x,y)(\tilde{A}(x,y)-A(x,y))|\right) \\ &\leqslant 2\max_{x\in\Omega}\sum_{y\in\Omega\setminus\{x\}}T(x,y)|\tilde{A}(x,y)-A(x,y)| \\ &\leqslant 16\epsilon\max_{x\in\Omega}\sum_{y\in\Omega\setminus\{x\}}T(x,y) \\ &\leqslant 16\epsilon,\end{aligned} \tag{C10}$$

where we used Lemma 11 at the second inequality. Similarly, we have

$$\begin{aligned}\|\delta W\|_1 &= \max_{y\in\Omega}\sum_{x\in\Omega}|\delta W_{x,y}| \\ &\leqslant 16\epsilon\max_{y\in\Omega}\sum_{x\in\Omega\setminus\{y\}}T(x,y) \\ &\leqslant 16\epsilon\max_{y\in\Omega}\sum_{x\in\Omega}T(x,y).\end{aligned} \tag{C11}$$

Combining Eqs. (C9)–(C11) with Eq. (C8), we obtain Eq. (72). ■

APPENDIX D: PROOF OF LEMMA 5

Proof of Lemma 5. If $\epsilon := \max_{x\in\Omega}|\tilde{L}(x) - L(x)| \leqslant \frac{1}{4}$, Eq. (C2) holds for any $x, y \in \Omega$ because of Lemma 11. Combining this with Eq. (15), we obtain Eq. (73).

If $\epsilon > \frac{1}{4}$, Eq. (73) holds trivially since the right-hand side is larger than 1 and the left-hand side is not larger than 1 by definition. ■

[1] R. Somma, S. Boixo, and H. Barnum, Quantum simulated annealing, arXiv:0712.1008.

[2] R. D. Somma, S. Boixo, H. Barnum, and E. Knill, Quantum Simulations of Classical Annealing Processes, Phys. Rev. Lett. **101**, 130504 (2008).

[3] P. Wocjan and A. Abeyesinghe, Speedup via quantum sampling, Phys. Rev. A **78**, 042336 (2008).

[4] M.-H. Yung and A. Aspuru-Guzik, A quantum-quantum metropolis algorithm, Proc. Natl. Acad. Sci. USA **109**, 754 (2012).

[5] A. W. Harrow and A. Y. Wei, Adaptive quantum simulated annealing for Bayesian inference and estimating partition functions, in *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (Association for Computing Machinery, New York, 2020), pp. 193–212.

[6] J. Lemieux, B. Heim, D. Poulin, K. Svore, and M. Troyer, Efficient quantum walk circuits for Metropolis-Hastings algorithm, Quantum **4**, 287 (2020).

[7] R. Campos, P. A. M. Casares and M. A. Martin-Delgado, Quantum Metropolis solver: A quantum walks approach to optimization problems, Quantum Machine Intelligence **5**, 28 (2023).

[8] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, RI, 2017).

[9] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, London, 1995).

[10] M. Szegedy, Quantum speed-up of Markov chain based algorithms, in *45th Annual IEEE Symposium on Foundations of Computer Science, Rome, Italy* (IEEE, New York, 2004), pp. 32–41.

[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[12] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika **57**, 97 (1970).

[13] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, Pub. Astron. Soc. Aust. **36**, e010 (2019).

[14] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW150914: The Advanced LIGO Detectors in the Era of First Discoveries, Phys. Rev. Lett. **116**, 131103 (2016).

[15] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. X **9**, 031040 (2019).

[16] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021).

[17] R. Abbott *et al.* (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv:2111.03606.

[18] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Gravitational waves from coalescing binaries: Detection strategies and Monte Carlo estimation of parameters, Phys. Rev. D **53**, 3033 (1996).

[19] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, Phys. Rev. D **53**, 6749 (1996).

[20] B. J. Owen and B. S. Sathyaprakash, Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement, Phys. Rev. D **60**, 022002 (1999).

[21] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85**, 122006 (2012).

[22] S. Gao, F. Hayes, S. Croke, C. Messenger, and J. Veitch, Quantum algorithm for gravitational-wave matched filtering, Phys. Rev. Res. **4**, 023006 (2022).

[23] K. Miyamoto, G. Morrás, T. S. Yamamoto, S. Kuroyanagi, and S. Nesseris, Gravitational wave matched filtering by quantum Monte Carlo integration and quantum amplitude amplification, Phys. Rev. Res. **4**, 033150 (2022).

[24] D. Veske, C. Tüysüz, M. Amico, N. T. Bronn, O. T. Lanes, Z. Márka, and S. Márka, Quantum encoding is suitable for matched filtering, arXiv:2204.04159.

[25] J. J. Postema, P. Bonizzi, G. Koekoek, R. L. Westra, and S. J. Kokkelmans, Hybrid quantum singular spectrum decomposition for time series analysis, arXiv:2212.08831.

[26] G. Escrig, R. Campos, P. A. M. Casares, and M. A. Martin-Delgado, Parameter estimation of gravitational waves with a quantum Metropolis algorithm, Class. Quantum Gravity **40**, 045001 (2023).

[27] A. Montanaro, Quantum speedup of Monte Carlo methods, Proc. R. Soc. A **471**, 20150301 (2015).

[28] S. Herbert, Quantum Monte Carlo integration: The full advantage in minimal circuit depth, Quantum **6**, 823 (2022).

[29] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, Quantum amplitude amplification and estimation, Contemp. Math. **305**, 53 (2002).

[30] P. Alquier, N. Friel, R. Everitt, and A. Boland, Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels, Stat. Comput. **26**, 29 (2016).

[31] F. J. Medina-Aguayo, A. Lee, and G. O. Roberts, Stability of noisy Metropolis–Hastings, Stat. Comput. **26**, 1187 (2016).

[32] D. Rudolf and N. Schweizer, Perturbation theory for Markov chains via Wasserstein distance, Bernoulli **24**, 2610 (2018).

[33] B. Parhami, *Computer Arithmetic: Algorithms and Hardware Designs* (Oxford University Press, Oxford, UK, 2010).

[34] D. Rudolf, Error bounds for computing the expectation by Markov chain Monte Carlo, Monte Carlo Methods Appl. **16**, 323 (2010).

[35] G. O. Roberts and J. S. Rosenthal, General state space Markov chains and MCMC algorithms, Probab. Surv. **1**, 20 (2004).

[36] F. Magniez, A. Nayak, J. Roland, and M. Santha, Search via quantum walk, SIAM J. Comput. **40**, 142 (2011).

[37] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[38] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. London A **454**, 339 (1998).

[39] L. K. Grover, Fixed-Point Quantum Search, Phys. Rev. Lett. **95**, 150501 (2005).

[40] E. Muñoz-Coreas and H. Thapliyal, Everything you always wanted to know about quantum circuits, in *Wiley Encyclopedia of Electrical and Electronics Engineering* (John Wiley & Sons, New York, 2022), pp. 1–17.

[41] M. K. Bhaskar, S. Hadfield, A. Papageorgiou, and I. Petras, Quantum algorithms and circuits for scientific computing, Quantum Inf. Comput. **16**, 197 (2016).

[42] T. Häner, M. Roetteler, and K. M. Svore, Optimizing quantum circuits for arithmetic, arXiv:1805.12445.

[43] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, arXiv:quant-ph/0208112.

[44] Y. R. Sanders, G. H. Low, A. Scherer, and D. W. Berry, Black-Box Quantum State Preparation without Arithmetic, Phys. Rev. Lett. **122**, 020502 (2019).

[45] S. Wang, Z. Wang, G. Cui, S. Shi, R. Shang, L. Fan, W. Li, Z. Wei, and Y. Gu, Fast black-box quantum state preparation based on linear combination of unitaries, Quantum Inf. Process. **20**, 1 (2021).

[46] S. Wang, Z. Wang, R. He, S. Shi, G. Cui, R. Shang, J. Li, Y. Li, W. Li, Z. Wei, and Y. Gu, Inverse-coefficient black-box quantum state preparation, New J. Phys. **24**, 103004 (2022).

[47] A. G. Rattew and B. Koczor, Preparing arbitrary continuous functions in quantum registers with logarithmic complexity, arXiv:2205.00519.

[48] J. Bausch, Fast black-box quantum state preparation, Quantum **6**, 773 (2022).

[49] P.-L. Dallaire-Demers and N. Killoran, Quantum generative adversarial networks, Phys. Rev. A **98**, 012324 (2018).

[50] C. Zoufal, A. Lucchi, and S. Woerner, Quantum generative adversarial networks for learning and loading random distributions, npj Quantum Inf. **5**, 103 (2019).

[51] H. Situ, Z. He, Y. Wang, L. Li, and S. Zheng, Quantum generative adversarial network for generating discrete distribution, Inf. Sci. **538**, 193 (2020).

[52] S. A. Stein, B. Baheri, D. Chen, Y. Mao, Q. Guan, A. Li, B. Fang, and S. Xu, Qugan: A quantum state fidelity based generative adversarial network, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2021), pp. 71–81.

[53] A. Anand, J. Romero, M. Degroote, and A. Aspuru-Guzik, Noise robustness and experimental demonstration of a quantum generative adversarial network for continuous distributions, Adv. Quantum Technol. **4**, 2000069 (2021).

[54] A. Assouel, A. Jacquier, and A. Kondratyev, A quantum generative adversarial network for distributions, Quantum Mach. Intell. **4**, 28 (2022).

[55] G. Agliardi and E. Prati, Optimal tuning of quantum generative adversarial networks for multivariate distribution loading, Quantum Rep. **4**, 75 (2022).

[56] S. Kasture, O. Kyriienko, and V. E. Elfving, Protocols for classically training quantum generative models on probability distributions, arXiv:2210.13442.

[57] D. J. Egger, R. García Gutiérrez, J. C. Mestre, and S. Woerner, Credit risk analysis using quantum computers, IEEE Trans. Comput. **70**, 2136 (2021).

[58] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Random Access Memory, Phys. Rev. Lett. **100**, 160501 (2008).

[59] F. L. Bauer and C. T. Fike, Norms and exclusion theorems, Numer. Math. **2**, 137 (1960).

[60] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 2013).