Measurement-Based Quantum Metropolis Algorithm

Jonathan E. Moussa
Molecular Sciences Software Institute, Virginia Tech, Blacksburg, Virginia 24060, USA

We construct a quantum version of a rejection-free classical Metropolis algorithm to prepare quantum thermal states. It samples thermal expectation values without interrupting its quantum Markov chain by driving transitions with observable measurements. Gaussian-filtered quantum phase estimation enables a thermalization timescale proportional to the reciprocal of temperature and the logarithm of biasing error.

We expect universal, efficient, unbiased simulations of quantum thermal states to require quantum algorithms on quantum computers [1]. Following the great successes of classical Markov chain Monte Carlo (MCMC) algorithms in sampling classical thermal states on classical computers [2], a natural progression is to consider quantum MCMC algorithms [3] for this purpose. The simplest approach is to extend an established classical MCMC algorithm such as the Metropolis algorithm [4] to the quantum case.

Two quantum Metropolis algorithms have already been proposed [5, 6], with two shared components and caveats. To prepare a quantum thermal state, $\rho_\beta = e^{-\beta H} / \text{tr}(e^{-\beta H})$, at a temperature $\beta^{-1}$, we need some form of access to the energies $E_a$ and stationary states $|\psi_a\rangle$ of its Hamiltonian,

$$H = \sum_a E_a |\psi_a\rangle \langle \psi_a|.$$  

Both algorithms use quantum phase estimation (QPE) [7] to access $E_a$ and $|\psi_a\rangle$ and then apply a unitary operation to drive transitions between $|\psi_a\rangle$. QPE uses Hamiltonian simulation time as a resource to reduce biasing errors to a magnitude that is inversely proportional to the maximum simulation time. Measurements of observables other than $H$ collapse $\rho_\beta$ and break the quantum Markov chain, and $\rho_\beta$ must be prepared again to repeat such measurements. Passive, unbiased sampling of thermal expectation values is a cornerstone of the classical Metropolis algorithm that its quantum extensions have failed to retain until now.

We construct a quantum Metropolis algorithm based on two revised components that retains the passive, unbiased sampling of thermal expectation values. Gaussian-filtered QPE (GQPE) defines a positive operator-valued measure (POVM) over $\omega$ that collapses an input state $|\Psi\rangle$ into

$$|\Psi(\omega)\rangle \propto \exp[-(\omega - H)^2/(2\gamma)] |\Psi\rangle$$  

and utilizes Hamiltonian simulation time more efficiently by reducing the magnitude of biasing errors exponentially in the maximum simulation time. The transitions between $|\psi_a\rangle$ are directly driven by preparation and measurement of the local stationary states $|\phi_a\rangle$ of a local observable,

$$L = \sum_a A_a |\phi_a\rangle \langle \phi_a| \otimes I.$$  

These measurements preserve the quantum Markov chain as $\rho_\beta$ mixes and subsequent measurements decorrelate.

To avoid the quantum complications of a rejection step [5], we use a rejection-free classical Metropolis algorithm [8] as the foundation for our quantum algorithm. Fig. 1 is the pseudocode of this stochastic algorithm that defines a conditional probability distribution, $M(b|a)$, from input to output state labels. We show that $M(b|a)$ obeys classical detailed balance in Appendix A, which guarantees that it preserves the thermal distribution $\langle \psi_a|\rho_\beta|\psi_a\rangle$ over input state labels. The conventional Metropolis algorithm [4] is recovered when $P(b|a)$ alternates between pairs of states, which causes the repeat-until-success loop to stop within two iterations and avoid protracted stopping times.

The quantum extension of Fig. 1 in Fig. 2 follows from using GQPE for indirect, imprecise energy measurements and a POVM that drives $P(b|a)$ transitions between $|\phi_a\rangle$ instead of $|\psi_a\rangle$. The quantum-extended pseudocode uses kets to denote quantum variables that are naturally passed by reference. Fig. 2 applies a POVM over $E$ and $a$ to $|\Psi\rangle$. 

FIG. 1: Rejection-free classical Metropolis algorithm [8] driven by sampling from the uniform distribution over $[0, 1]$, $U(0, 1)$, and a symmetric conditional probability distribution, $P(b|a)$.
For $\epsilon \ll 1$, the representative example of a single-energy Hamiltonian has a stopping-time distribution with a mean of $10 \log(1/\epsilon)$ iterations and a variance of $24/\epsilon$ [12].

Quantum thermalization using Figs. 2 and 3 occurs on a Hamiltonian simulation timescale set by $t_{\text{max}}$ in Eq. (4) multiplied by the repeat-until-success stopping timescale and the mixing timescale for the quantum Markov chain. Known Hamiltonian simulation algorithms can be used to relate resource estimates in simulation time to lower-level estimates such as a number of Hamiltonian oracle queries [13] or a circuit depth [14]. As in the classical Metropolis algorithm, we should anticipate long mixing timescales in difficult instances that might be reduced by careful design of $L$ and $P(b|a)$. The same considerations might apply to the stopping timescales for the quantum algorithm. In the simpler case of Fig. 1, the stopping time is related to the recurrence timescale of $P(b|a)$, which can be limited by restricting each $L$ to a small subsystem in Fig. 2. As a demonstration, we efficiently thermalize a transverse-field Ising model in Fig. 4 by sweeping over spin flips [12].

In conclusion, we have constructed and demonstrated a practical quantum Metropolis algorithm with efficient bias reduction and passive observable sampling. It serves as a foundation for the development of other quantum MCMC algorithms that retain the same benefits alongside further improvements. For example, coherent chain evolution [6] or quantum analogs of cluster updates [15] may accelerate quantum Markov chain mixing for difficult instances.

The Molecular Sciences Software Institute is supported by NSF Grant No. ACI-1547580. J. E. M. thanks Andrew Baczewski and Norm Tubman for useful discussions.

![Diagram](image-url)
APPENDIX A: DETAILED BALANCE

We unravel the statistical description $M(b|a)$ of Fig. 1 into all possible iteration numbers and transition paths,

$$M(b|a) = \sum_{n=1}^{\infty} \sum_{a_1, \ldots, a_{n-1}} M(b, a_{n-1}, \ldots, a_1|a),$$

$$\hat{M}(a_n, \ldots, a_1|a_0) = R^n_0 A^n_0 \prod_{i=1}^{n} P(a_i|a_{i-1}),$$

$$R^n_m = \max(m,n)-1 \prod_{i=\min(m,n)+1}^{\max(m,n)-1} (1 - A^n_m),$$

$$A^n_m = A(a_n, \ldots, a_m),$$

(A1)

where $R^n_m$ is the probability of reaching a path and $A^n_m$ is the probability of accepting it given a specific sequence of transitions. As in previous work [8], we require that pairs of statistical branches $\hat{M}$ corresponding to each transition path and its reverse independently obey detailed balance,

$$\hat{M}(a_n, \ldots, a_1|a_0) p_{a_0} = \hat{M}(a_0, \ldots, a_{n-1}|a_n) p_{a_n}$$

(A2)

for $p_a = e^{-\beta E_a}$, which is a stricter condition than detailed balance of $M(b|a)$. We can simplify Eq. (A2) further to

$$A^n_0 R^n_0 p_{a_0} = A^n_0 p_{a_0} p_{a_n}$$

(A3)

by factoring out path-reversal symmetric $P(b|a)$ terms in $\hat{M}$, which acts as a constraint on acceptance functions $A^n_m$. Subject to this constraint, we achieve the greedy maximum for each acceptance probability with the choice

$$A^n_m = \min\{ (R^n_m p_{a_m})/(R^n_0 p_{a_0}) , 1\},$$

(A4)

which is correspondingly implemented in Fig. 1.

The quantum version in Fig. 2 is a completely positive trace-preserving map $\mathcal{M}$ of $|\Psi\rangle$ when marginalized over $E$ and $a$ outputs. In a state transition basis, we unravel its description by iteration number and transition paths,

$$\mathcal{M}(\rho) = \sum_{a,b,c,d} \mathcal{M}^{ab}_{cd} |\psi_c\rangle \langle \psi_d| \rho_{|\psi_b\rangle} \langle \psi_b|,$$

$$\mathcal{M}^{ab}_{cd} = \sum_{n=1}^{\infty} \sum_{n_1, \ldots, n_{n-1}} \mathcal{M}^{ab}_{cd} n_1 \cdots n_{n-1} b_1 \cdots b_{n_1}$$

with

$$\mathcal{M}^{ab}_{cd} = n_{n-1} \cdots n_1 \prod_{i=1}^{n} \delta(n_i - a_{n-i}),$$

$$\Omega^n_m = (E_{a_n} \pm E_{b_m})/2,$$

$$\Omega^n_{max} = \Omega^n_{\min} + 1$$

$$\mathcal{R}^n_m(\Omega) = \prod_{i=\min(m,n)+1}^{\max(m,n)-1} [1 - A^n_m(\Omega)],$$

$$A^n_m(\Omega) := A(\omega_{a_n}, \ldots, \omega_{m+\text{sign}(n-m)}, \omega_m \pm \Omega),$$

$$\mathcal{T}_n = \sum_{c_n,d_n} P(c_n|d_n) \Theta_{c_n,d_n} A^n_{a_n a_{n-1}} A^n_{b_{n-1} b_n},$$

$$\Theta^{cd}_{ab} = \langle \psi_c | (|\psi_d\rangle \otimes I) \psi_d \rangle,$$

(A5)

where the reach and acceptance probabilities, $\mathcal{R}^n_m(0)$ and $A^n_m(0)$, now depend on sequences of GQPE measurement outcomes $\omega_n$ rather than direct state energies. Similar to the classical case, we guarantee quantum detailed balance [5] with a stricter condition on pairs of branches $\hat{M}$,

$$\mathcal{X}^{ab}_{da_1 \cdots a_0 \sqrt{P} a_0 P b_0} = \mathcal{X}^{ab}_{b_n \cdots b_0 \sqrt{P} a_n P b_n},$$

(A6)

We shift Boltzmann factors $e^{-\beta E}$ from unobserved $E_a$ to measured $\omega_n$ using the Gaussian integral identity

$$\int_{-\infty}^{\infty} d\omega \ f(\omega) e^{-\omega (\omega - E)^2}/\gamma \ E = e^{-\beta^2 \gamma/4}$$

$$\times \int_{-\infty}^{\infty} d\omega \ f(\omega + \beta \gamma/2) e^{-\beta \omega (\omega - E)^2}/\gamma,$$

(A7)

which allows us to factor out the path-reversal symmetric dependence on $a_n, b_n, c_n,$ and $d_n$ from Eq. (A6) to get

$$A^n_0(\beta \gamma/2) \mathcal{R}^n_0(\beta \gamma) e^{-\beta \omega} = A^n_0(\beta \gamma/2) \mathcal{R}^n_0(\beta \gamma) e^{-\beta \omega_n},$$

(A8)

The constraints on acceptance functions in Eqs. (A3) and (A8) are equivalent up to a $\beta \gamma/2$ bias correction.
We build a quantum circuit approximation of GQPE by expanding the POVM from Eq. (2) into a measurement of an ancillary continuous quantum variable \( \omega \),

\[
|\Psi(\omega)\rangle \otimes |\omega\rangle \propto \int_{-\infty}^{\infty} dt \frac{e^{i\omega t}}{\sqrt{2\pi}} g(t) e^{-iHt}|\Psi\rangle \otimes |\omega\rangle,
\]

\[
g(t) = \int_{-\infty}^{\infty} du \frac{e^{-i\omega u}}{\sqrt{2\pi}} e^{-\omega^2/(2\gamma)},
\]

that is prepared with a narrow Gaussian filter over \( \omega \), then Fourier transformed into a wide Gaussian filter over \( t \) and used as a control of Hamiltonian time evolution duration before its inverse Fourier transform back to \( \omega \). With only \( n \) ancillary qubits available, we approximate the integrals over continuous quantum variables by finite sums over

\[
\omega_j = \omega_{\text{max}} \frac{2j + 1 - 2^n}{2n}, \quad t_j = t_{\text{max}} \frac{2j + 1 - 2^n}{2n}, \tag{B1}
\]

for \( j \in \{0, \cdots, 2^n - 1\} \). If these sums are constrained by \( \omega_{\text{max}} t_{\text{max}} = 2^{n-1} - \pi / 2 \), then they can be implemented with the standard QFT as shown in Fig. 3. We group excess phase rotations with an initial ancillary state limited to \( r \) qubits,

\[
|F_r\rangle \propto \sum_{j=0}^{2^n-1} e^{-i(1-2^{-n})\bar{\omega}_j t_{\text{max}} - \bar{\omega}_j^2/(2\gamma)} |j\rangle \tag{B3}
\]

for \( \bar{\omega}_j = \omega_j + 2^{n-1} - 2^{-n-1} \). The preparation of \( |F_r\rangle \) is likely to require \( O(2^n) \) gates [17], thus we must carefully choose \( r \) alongside \( n \) and \( t_{\text{max}} \) to balance cost and accuracy.

Detailed balance of our quantum Metropolis algorithm is robust to this GQPE discretization because the identity in Eq. (A7) is also satisfied by infinite summations,

\[
\sum_{j \in \mathbb{Z}} f(\omega_j) e^{-(\omega_j - E)^2/\gamma - \beta E} = e^{-\beta^2 \gamma / 4} \times \sum_{j \in \mathbb{Z}} f(\omega_j + \beta \gamma / 2) e^{-\beta \omega_j - (\omega_j - E)^2/\gamma} \tag{B4}
\]

if \( \beta \gamma t_{\text{max}} / (2\pi) \) is an integer. This constrains the choice of \( \gamma \), and we choose \( \gamma = 2\pi / (\beta t_{\text{max}}) \) to minimize the effects of the bias correction in Fig. 2 in that \( E_{\text{new}} \) only needs to be one grid point lower than \( E \) on the finite energy grid to terminate the repeat-untill-success loop with certainty.

Since the discretization of POVM outcomes in Eq. (2) does not interfere with detailed balance, the biasing errors are distortions of its Gaussian form in Fig. 3,

\[
\tilde{G}(\omega) = \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{i(\omega - \bar{\omega}_j)(t - \bar{\omega}_k)/(2\gamma)} / 2^n \tag{B5}
\]

over the domain \( |\omega| \leq \Omega_{\text{max}} = (1 - 2^{-n}) \omega_{\text{max}} + E_{\text{max}} \) with \( E_{\text{max}} = \max_{a} |E_a| \). We quantify the errors in this Fourier approximation by its maximum deviation, which depends on four dimensionless parameters \( -\beta \max E_{\text{max}}/r \), \( \beta t_{\text{max}} \), \( r \), and \( n \) – in an empirical bound based on error asymptotics,

\[
\min_{\delta \in \mathbb{C}} \max_{\omega \in [-\Omega_{\text{max}}, \Omega_{\text{max}}]} |e^{-\omega^2/(2\gamma)} - (1 + \delta) \tilde{G}(\omega)| \leq \epsilon,
\]

\[
\epsilon = \exp[-\max\{\pi t_{\text{max}} / \beta, 2^{2(n-1)} \pi \beta / (4\max)\},
\]

\[
(2^n - 1) E_{\text{max}} t_{\text{max}} / \pi^2 \beta^2 (4\max) \}
\]

contingent on \( E_{\text{max}} \leq \omega_{\text{max}} [12] \). The resource estimate in Eq. (4) minimizes \( t_{\text{max}} \), \( r \), and \( n \) for a given \( \epsilon \).

The biasing error in each approximate GQPE operation accumulates over the iterations of the repeat-untill-success loop in Fig. 2 and between successive instances of Fig. 2 in a thermalizing quantum Markov chain. For every loop iteration, the Gaussian weights from \( G_n \) in Eq. (A5) incur an error bounded by \( 2\epsilon / (1 - \epsilon) \). When compounded over \( N \) iterations using the triangle inequality, this error bound grows to \( [(1 + \epsilon) / (1 - \epsilon)]^N - 1 \). It achieves saturation at \( N = [1 / \log_2(1 + \epsilon) / (1 - \epsilon)] \), and we then halt the loop without an increase in its boundable error. While a more detailed error analysis is beyond the scope of this Letter, Markov chains dissipate the effects of past biasing errors, and their stationary states only accumulate biasing errors over the mixing timescale rather than the chain length.

* Electronic address: godotalgorithm@gmail.com

[1] R. P. Feynman, Int. J. Theor. Phys. 21, 467 (1982).
[2] D. P. Landau and K. Binder, A Guide to Monte Carlo Simulations in Statistical Physics, (Cambridge University Press, Cambridge, 2015).
[3] B. M. Terhal and D. P. DiVincenzo, Phys. Rev. A 61, 022301 (2000).
[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).
[5] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Nature 471, 87 (2011).
[6] M.-H. Yung and A. Aspuru-Guzik, Proc. Nat. Acad. Sci. USA 109, 754 (2012).
[7] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. 83, 5162 (1999).
[8] A. Mira, Metron 59, 231 (2001).
[9] A. Paetznick and K. M. Svore, Quantum Inf. Comput. 14, 1277 (2014).
[10] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Phys. Rev. X 8, 041015 (2018).
[11] F. Fillion-Gourdeau, S. MacLean, and R. Laflamme, Phys. Rev. A 95, 042331 (2017).
[12] The software and data for all numerical results are available at https://github.com/godotalgorithm/quantum-metropolis.
[13] G. H. Low and I. L. Chuang, Quantum 3, 163 (2019).
[14] A. M. Childs and Y. Su, Phys. Rev. Lett. 123, 050503 (2019).
[15] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. 58, 86 (1987).
[16] D. Coppersmith, arXiv:quant-ph/0201067.
[17] V. V. Shende, S. S. Bullock, and I. L. Markov, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 25, 1000 (2006).