Quantum Techniques in Machine Learning 2023

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

(Phys. Rev. Research 5, 033059 (2023))

Nov 24, 2023

Koichi Miyamoto

Center for Quantum Information and Quantum Biology, Osaka University

Markov chain Monte Carlo method & Metropolis-Hastings algorithm

- Markov chain Monte Carlo method (MCMC)
  - sample from the *target distribution* P by generating a chain of samples $x_0 \to x_1 \to x_2 \to \cdots$ s.t. the distribution of x_i converges to P
  - often used for Bayesian inference (e.g. parameter optimization in machine learning)
- Metropolis-Hastings method (MH): a widely-used kind of MCMC
  - given the ith sample x_i, the $(i+1)$th one x_{i+1} is chosen as follows
    1. randomly draw a candidate $\tilde{x}_{i+1}$ from the *proposal distribution* $T(x_i,\cdot)$
    2. calculate the *acceptance ratio* $A(x_i, \tilde{x}_{i+1}) = \min\left\{1, \frac{P(\tilde{x}_{i+1})T(\tilde{x}_{i+1},x_i)}{P(x_i)T(x_i,\tilde{x}_{i+1})}\right\}$
    3. set $x_{i+1} = \tilde{x}_{i+1}$ with prob. $A(x_i, \tilde{x}_{i+1})$, or $x_{i+1} = x_i$ with prob. $1 - A(x_i, \tilde{x}_{i+1})$
- Convergence rate of MCMC
  - # of iterations for the chain to converge to P: $\tilde{O}(1/\Delta)$ ¶
    - ✓spectral gap $\Delta = 1 - |\lambda_1|$
      λ_1= (the eigenvalue of the transition matrix W with the 2nd largest modulus)

¶ in terms of total variation distance (Levin & Peres, “Markov chains and mixing times” (2017))

Quantum algorithm for MCMC

- Quantum simulated annealing (QSA)†
  - generates P-encoding state $|P\rangle := \sum_x \sqrt{P(x)}|x\rangle$ querying the *quantum walk operator* U $\tilde{O}(1/\sqrt{\Delta})$ time
    → **quadratic speedup** compared to classical MCMC $\tilde{O}(1/\Delta)$
- For MH, a concrete implementation of U is given¶
  - acts on a system of 2 registers R_S, R_M and 1 qubit R_C
  - $U = RV^\dagger B^\dagger SFBV$
    - ✓ $V|x\rangle_{R_S}|0\rangle_{R_M} = |x\rangle_{R_S}\sum_{\Delta x}\sqrt{T(x, x+\Delta x)}|\Delta x\rangle_{R_M}$ (Δx: possible move)
    - ✓ $B|x\rangle_{R_S}|\Delta x\rangle_{R_M}|\phi\rangle_{R_C} = |x\rangle_{R_S}|\Delta x\rangle_{R_M} \otimes \begin{pmatrix} \sqrt{1-A(x,x+\Delta x)} & -\sqrt{A(x,x+\Delta x)} \\ \sqrt{A(x,x+\Delta x)} & \sqrt{1-A(x,x+\Delta x)} \end{pmatrix}|\phi\rangle_{R_C}$
    - ✓ $F|x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C} = |x\rangle_{R_S}|\Delta x\rangle_{R_M}|0\rangle_{R_C}, F|x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C} = |x+\Delta x\rangle_{R_S}|\Delta x\rangle_{R_M}|1\rangle_{R_C}$
    - ✓ $S|\Delta x\rangle_{R_M}|0\rangle_{R_C} = |\Delta x\rangle_{R_M}|0\rangle_{R_C}, S|\Delta x\rangle_{R_M}|1\rangle_{R_C} = |-\Delta x\rangle_{R_M}|1\rangle_{R_C}$
    - ✓ $R = I_{R_S} \otimes \left(2|0\rangle\langle 0|_{R_M} \otimes |0\rangle\langle 0|_{R_C} - I_{R_M} \otimes I_{R_C}\right)$

† Harrow & Wei, SODA 2020 ¶ Lemieux et al., Quantum 4, 287 (2020)

Issue: target distribution calculated via summation of many terms

- e.g., optimization of the parameter x in a statistical model with a large data set $\mathcal{D}$ in the Bayesian approach
  - ➢ we want to optimize the posterior distribution of x:
    $P(x|\mathcal{D}) \propto P_0(x)P(\mathcal{D}|x)$ ($P_0(x)$: prior distribution)
    - ✓likelihood $P(\mathcal{D}|x) = \exp(L_{\mathcal{D}}(x))$
      log-likelihood $L_{\mathcal{D}}(x) = \frac{1}{M}\sum_{i=0}^{M-1} \ell_i(x), M \gg 1$ ← **sum of many terms**
      $\ell_i(x)$: contribution from the ith data point in $\mathcal{D}$

- **Can we run MH with speeding up the summation by a quantum algorithm?**

Our idea: speed up the summation by QMCI

- We use **quantum Monte Carlo integration (QMCI)**† for $L_{\mathcal{D}}(x) = \frac{1}{M}\sum_{i=0}^{M-1} \ell_i(x)$, and incorporate it into QSA

- QMCI
  - a quantum algorithm to calculate an expectation of a random variable (and a sum as a special case)
  - calculate $L_{\mathcal{D}}(x) = \frac{1}{M}\sum_{i=0}^{M-1} \ell_i(x)$ querying the oracle O_ℓ to compute ℓ_i
    $O_\ell|x\rangle|i\rangle|0\rangle = O_\ell|x\rangle|i\rangle|\ell_i(x)\rangle$
  - for accuracy ϵ, the query number is $\tilde{O}(\sigma/\epsilon)$
    - ✓ σ^2: the variance of the terms ℓ_i
      $\sigma^2 \coloneqq \max_x \frac{1}{M}\sum_{i=0}^{M-1} \ell_i^2(x) - \left(\frac{1}{M}\sum_{i=0}^{M-1} \ell_i(x)\right)^2$
    - ✓ **quadratic speedup** compared to classical Monte Carlo integration: $\tilde{O}(\sigma^2/\epsilon^2)$

† Montanaro, Proc. R. Soc. A, 471(2181):20150301 (2015)

Drawback of using QMCI

- QMCI outputs $L_{\mathcal{D}}$ with an **error**
  → error in the acceptance ratio A
  → the chain converges to the distribution P' different from the target P

➢ If we use A' s.t. $\max_{x,y}|A(x,y) - A'(x,y)| \leq \epsilon$,
$\|P - P'\|_{TV} = \tilde{O}(\epsilon/\Delta)$ ¶
(TV: total variation distance)

⬇

➢ For $\|P' - P\|_{\mathrm{TV}} \leq \epsilon$, it is sufficient that the error in $L_{\mathcal{D}}$ is $\tilde{O}(\epsilon\Delta)$

⬇

➢ query complexity in QMCI: $\tilde{O}(\sigma/\epsilon\Delta)$

¶ Alquier et al., Statistics and Computing 26, 29 (2016)

Result 1: Generating $|P\rangle$ by QSA with $L_{\mathcal{D}}$ calculated by QMCI

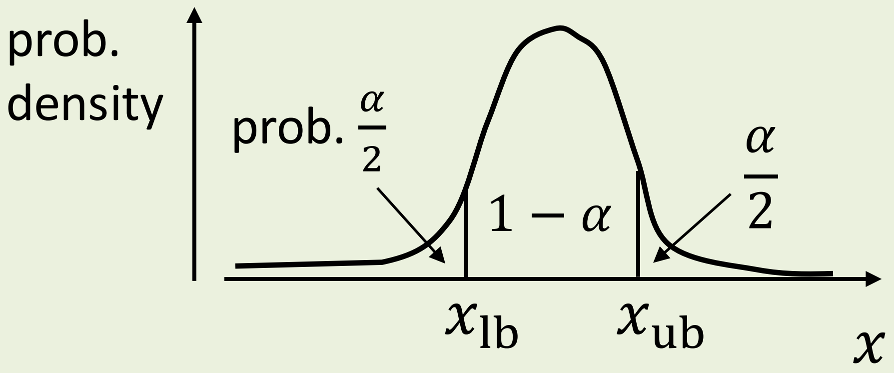
- Theorem (informal)
  - Suppose that we are given the oracle O_ℓ to compute ℓ_i.
    There is a quantum algorithm that outputs an ϵ-approximation of $|P\rangle := \sum_x \sqrt{P(x)}|x\rangle$, making $\tilde{O}\left(\sigma/\epsilon\Delta^{3/2}\right)$ queries to O_ℓ.

- In the *exact QSA*, in which $L_{\mathcal{D}} = \frac{1}{M}\sum_{i=0}^{M-1} \ell_i(x)$ is calculated as the definition (by M-time iterations of calculating $\ell_i(x)$ and adding it), the query number is $\tilde{O}\left(M/\Delta^{1/2}\right)$
  → QMCI improves the scaling on M, the number of terms, in compensation for ϵ, Δ

Estimation of the credible interval

- QSA outputs $|P\rangle$, but we want not a quantum state but some statistics on P as classical data

- Typical quantity of interest: credible interval (CI) of a parameter in a statistical model
  - $100(1-\alpha)\%$ CI for x:
    $[x_{\rm lb}, x_{\rm ub}]$ s.t. $P(x < x_{\rm lb}) = \frac{\alpha}{2}, P(x_{\rm ub} < x) = \frac{\alpha}{2}$



- Given the oracle to generate $|P\rangle$ by QSA, we can estimate $x_{\rm lb}, x_{\rm ub}$ as follows
  - We can calculate the cumulative distribution function (CDF) $\Phi(a) := P(x < a)$ as $\Phi(a) = E_P[\mathbf{1}_{x<a}] = \sum_x \mathbf{1}_{x<a} P(x)$, the expectation of the indicator function $\mathbf{1}_{x<a}$ in P, by QMCI
  - Calculating $\Phi(a)$ like this, we find $x_{\rm lb}, x_{\rm ub}$ by bisection (or other root-finding methods)

Result 2: CI estimation by QSA with QMCI

- Theorem (informal)
  - Suppose that we are given the oracle O_ℓ to compute ℓ_i.
    There is a quantum algorithm that outputs estimates on $x_{\mathrm{lb}}, x_{\mathrm{ub}}$ with accuracy ϵ (in terms of the CDF), making $\tilde{O}\left(\sigma/\epsilon^2\Delta^{3/2}\right)$ queries to O_ℓ.

- Based on the exact QSA, the query number is $\tilde{O}\left(M/\epsilon\Delta^{1/2}\right)$
  → QMCI improves the scaling on M, the number of terms, in compensation for ϵ, Δ

When QSA with QMCI is beneficial: σ sublinear w.r.t. M

- We defined $L_{\mathcal{D}}(x)=\frac{1}{M}\sum_{i=0}^{M-1}\ell_i(x)$

  → Note that the prefactor $1/M$, which does not always exist

  ➢ e.g., in the model parameter estimation with M independent data points, $L_{\mathcal{D}}(x)=\sum_{i=0}^{M-1}\ell_i(x)$ with ℓ_i not depending on M

- By redefining $\ell_i(x)\rightarrow M\ell_i(x)$, we can write $L_{\mathcal{D}}$ in the form $L_{\mathcal{D}}(x)=\frac{1}{M}\sum_{i=0}^{M-1}\ell_i(x)$, but this leads to $\sigma=O(M)$, which causes the advantage of QSA with QMCI to disappear

  ➢ Complexity of generating $|P\rangle$: $\tilde{O}\left(\sigma/\epsilon\Delta^{3/2}\right)=\tilde{O}\left(M/\epsilon\Delta^{3/2}\right)$

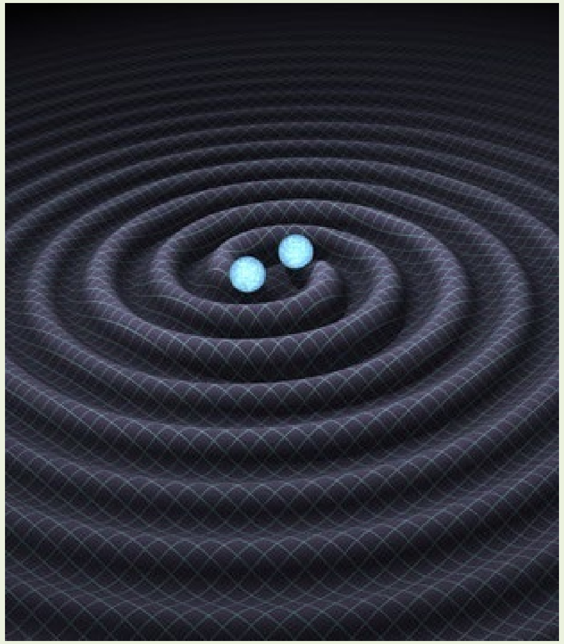
    → compared to the exact QSA ($\tilde{O}\left(M/\Delta^{1/2}\right)$), the scaling on M is same and those on ϵ,Δ are worse

- QSA with QMCI can be beneficial if **σ scales on M sublinearly**

  → e.g. **parameter estimation in a gravitational wave detection experiment**

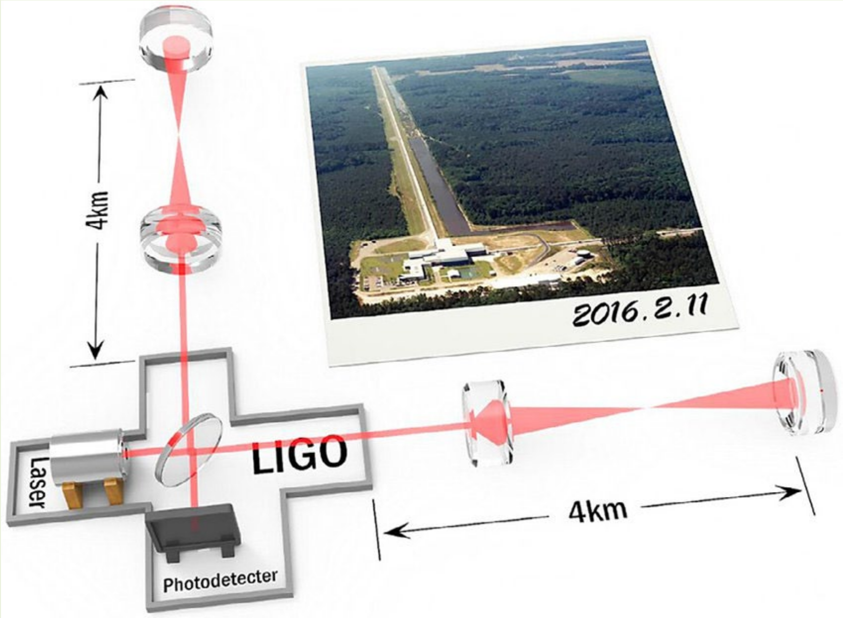
Gravitational wave

- Gravitational wave (GW)
  - wave of spacetime distortion caused by an extreme astrophysical event such as a merger of black holes (BHs)
  - detected by laser interferometers such as LIGO in US



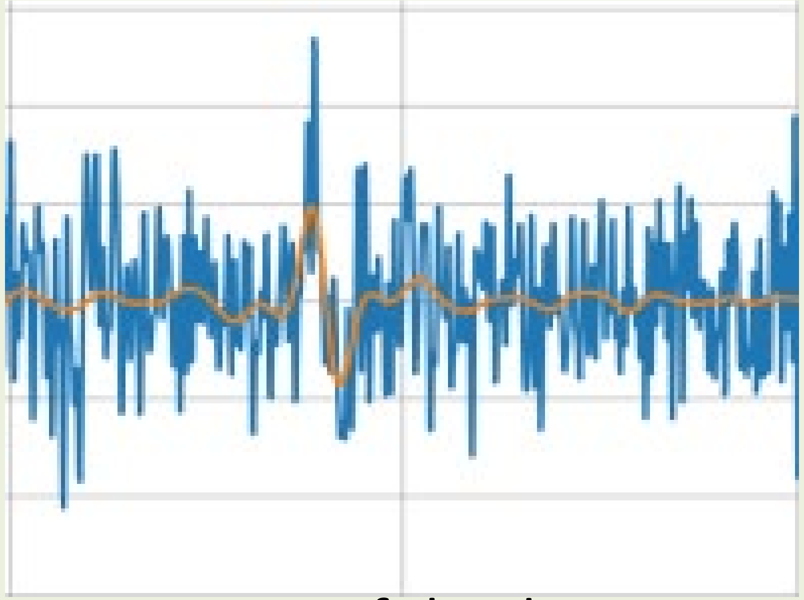
GW from BH merger

www.ligo.caltech.edu



Laser interferometer

Zuo+., Opt. Lasers Eng. 135, 106187 (2020)



output of the detector
(noise ≫ GW signal)

Morras+, Phys. Dark Universe 35, 100932 (2022)

Figures are for illustration purposes.

GW parameter estimation by QSA with QMCI

- We estimate GW parameters (e.g. BH's mass) from the signal in the detector output $s(t)$
  - If we have $s(t)$ as M-point time-series data with interval Δt,

$$L_{\mathcal{D}}(x) = \frac{1}{M/2}\sum_{i=1}^{M/2-1} \ell_i(x)\,, \ell_i(x) = \mathrm{Re}\left(\frac{4\tilde{h}^*(f_i,x)\tilde{s}(f_i)}{S_n(f_i)\Delta\mathrm{t}}\right) \text{¶}$$

    - ✓ ℓ_i: contribution from the Fourier mode with frequency $f_i = i/M\Delta t$
    - ✓ $\tilde{h}(f,x)$: Fourier-transformed theoretical waveform of GW depending on parameters x
    - ✓ $\tilde{s}$: Fourier-transf. of s, S_n: noise power spectrum
  - In this case, $\sigma = O(\sqrt{M})$, so QSA with QMCI can be beneficial
    - ✓ $\sigma = O(\sqrt{M})$ is due to the situation that random noise dominates over the GW signal in the detector output
- QSA with QMCI estimates the CI of a GW parameter with $\tilde{O}(M^{1/2}/\Delta^{3/2}\epsilon^2)$ queries to O_ℓ

¶ Some terms omitted.

Summary

- MCMC, especially MH is a widely used technique, e.g. Bayesian inference including parameter optimization in machine learning.
- QSA provides quadratic speedup with respect to spectral gap Δ compared with classical MH.
- We focused on another point, calculation of the log-likelihood $L_{\mathcal{D}}$ as a sum of many terms. We proposed speeding up the summation by QMCI and incorporated it into QSA.
- We consider not only generating the quantum state $|P\rangle$ but also extracting a quantity of interest, a credible interval.
- We present GW parameter estimation as an example where QSA with QMCI is beneficial.

- Summary of query complexity

Task	QSA with QMCI	Exact QSA	Classical MH
Generating $\|P\rangle$	$\tilde{O}(\sigma/\Delta^{3/2}\epsilon)$	$\tilde{O}(M/\Delta^{1/2})$	N/A
CI estimation (general)	$\tilde{O}(\sigma/\Delta^{3/2}\epsilon^2)$	$\tilde{O}(M/\Delta^{1/2}\epsilon)$	$\tilde{O}(M/\Delta\epsilon^2)$
CI estimation (GW)	$\tilde{O}(M^{1/2}/\Delta^{3/2}\epsilon^2)$	$\tilde{O}(M/\Delta^{1/2}\epsilon)$	$\tilde{O}(M/\Delta\epsilon^2)$