Title: Gibbs Sampling of Periodic Potentials on a Quantum Computer

Speakers: Arsalan Motamedi

Collection: New Frontiers in Machine Learning and Quantum

Date: November 23, 2022 - 4:00 PM

URL: https://pirsa.org/22110094

Abstract: "Motivated by applications in machine learning, we present a quantum algorithm for Gibbs sampling from continuous real-valued functions defined on high dimensional tori. We show that these families of functions satisfy a Poincare inequality. We then use the techniques for solving linear systems and partial differential equations to design an algorithm that performs zeroeth order queries to a quantum oracle computing the energy function to return samples
from its Gibbs distribution. We further analyze the query and gate complexity of our algorithm and prove that the algorithm has a polylogarithmic dependence on approximation error (in total variation distance) and a polynomial dependence on the number of variables, although it suffers from an exponentially poor dependence on temperature."

Gibbs Sampling
of Periodic Potentials on a Quantum Computer

Arsalan Motamedi, Pooya Ronagh

November 23, 2022

UNIVERSITY OF WATERLOO | IQC Institute for Quantum Computing

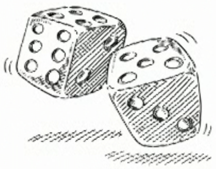
Outline

Problem definition

Given a (potential/energy) function $E : \mathbb{R}^d \to \mathbb{R}$, output samples according to the distribution:

$$e^{-E(x)}/Z$$

We denote this distribution by $\mathcal{G}$.



Source: istockphoto.com

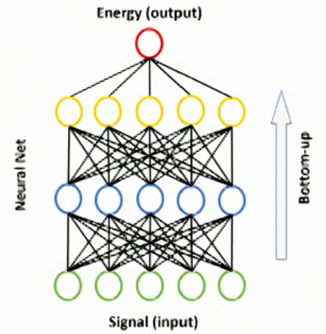
Outline

Energy-Based Models

Gradient estimation in the training of Energy-based models could be the reduced to Gibbs sampling[1].

$$\partial_\theta(LL) = -\nabla_\theta E + \mathbb{E}_{\mathcal{G}}\left[\nabla_\theta E\right]$$

Collect samples according to $\mathcal{G}$



Source: jasonlee1980hk.medium.com

[1] Song and Kingma 2021.

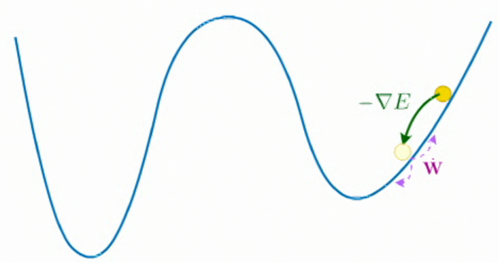
Outline

Langevin Dynamics

Consider the following Stochastic Differential Equation (aka Langevin Dynamics)

$$\mathrm{d}Y_t = -\nabla E(Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t$$

If E is confining, then, this process is ergodic, with the unique stationary-state being the Gibbs state[2].



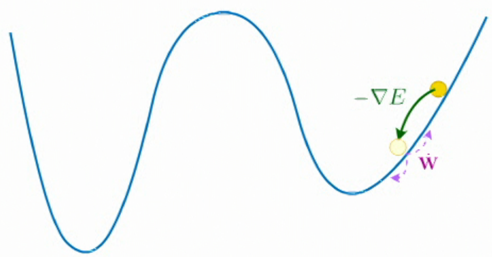
[2] Pavliotis 2014.

Langevin Dynamics

Consider the following Stochastic Differential Equation (aka Langevin Dynamics)

$$\mathrm{d}Y_t = -\nabla E(Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t$$

If E is confining, then, this process is ergodic, with the unique stationary-state being the Gibbs state[2].



How long should it run for?

[2] Pavliotis 2014.

Fast convergence

The convergence of this process is connected to a functional inequality, namely the *Poincaré inequality*, as follows:

$$\|\rho_t/\rho_\infty - 1\|_{L^2(\rho_\infty)} \leq e^{-\lambda t}\|\rho_0/\rho_\infty - 1\|_{L^2(\rho_\infty)}$$

$$\Updownarrow$$

$$\mathrm{Var}_{X\sim\mathcal{G}}[f(X)] \leq \lambda\,\mathbb{E}_{X\sim\mathcal{G}}\left[\|\nabla f(X)\|^2\right]$$

Source: Wikipedia

Poincaré inequality

$$\mathrm{Var}_{X\sim\mathcal{G}}[f(X)] \leq \lambda\,\mathbb{E}_{X\sim\mathcal{G}}\left[\|\nabla f(X)\|^2\right]$$

A few facts about functions satisfying PI:

- Any μ-strongly convex function satisfies PI with $\lambda = \mu$ (*Bakry-Emery criterion*).
- **KLS conjecture:** There is a universal PI constant for all log-concave, zero-mean Gibbs measures with an identity covariance matrix, which is independent of d[3].
- There are non-convex functions that satisfy a PI[4].
- Unimodal functions with strict minimum satisfy a PI[5].

[3] Alonso-Gutiérrez and Bastero 2015.
[4] Pavliotis 2014.
[5] Li and Erdogdu 2020.

Outline

Classical Algorithms

How to simulate Langevin Dynamics (LD) on a computer?

- Naïvely discretized LD may *not* provide the Gibbs distribution in the asymptotic limit[6]!
- However, adding a Metropolis-Hasting step will result in fast algorithms for Gibbs sampling, in cases where the energy function is strongly convex[7].

[6] Roberts and Tweedie 1996.

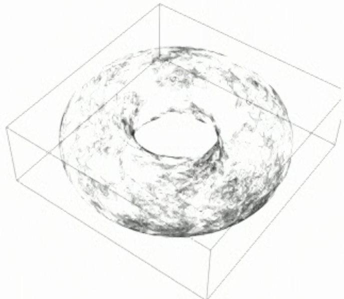
[7] Dwivedi et al. 2018.

Source: medium.com

Outline

Periodic Langevin Dynamics

Letting $X_t := (Y_t + \pi \mod 2\pi) - \pi$ will provide us with another stochastic process.



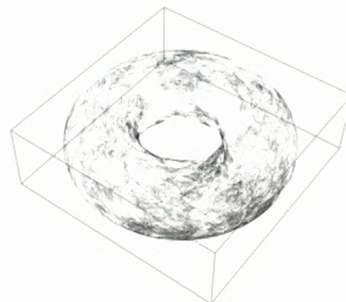
Source: stackexchange.com

If E is 2π-periodic and smooth, then X_t is a Markovian ergodic process with the unique stationary state being proportional to e^{-E} on the torus[8].

[8] Garcıa-Portugués et al. 2019.

Periodic Langevin Dynamics

Letting $X_t := (Y_t + \pi \mod 2\pi) - \pi$ will provide us with another stochastic process.



Source: stackexchange.com

If E is 2π-periodic and smooth, then X_t is a Markovian ergodic process with the unique stationary state being proportional to e^{-E} on the torus[8].

Proposition

For periodic Langevin dynamics, we have $\lambda = e^{\Delta}$ as the universal Poincaré constant, where Δ is the diameter of the range of E.

[8] Garcia-Portugués et al. 2019.

Outline

Main Result

Theorem (Very Informal)

There is a quantum algorithm sampling from the Gibbs distribution associated with the periodic potential E, making

$$\mathcal{O}\left(d^7\, e^{3\Delta/2}\, \mathrm{polylog}\left(\frac{de^{\Delta}}{\epsilon}\right)\right)$$

queries to an oracle for the energy function.

All other algorithms[9] that achieve high precision Gibbs sampling from classical potentials make assumptions about convexity or satisfaction of isoperimetric inequalities, whereas our algorithm requires mild analyticity conditions.

[9] Roberts and Tweedie 1996; Dwivedi et al. 2018; Chewi et al. 2021; Childs et al. 2022.

Outline

Connection to a PDE

The density of the process satisfies a forward Kolmogorov equation:

$$X_t = \text{s-mod}(Y_t) \Rightarrow \partial_t \rho(x,t) = \nabla \cdot \left(e^{-E(x)} \nabla \left(e^{E(x)} \rho(x,t)\right)\right)$$

Connection to a PDE

The density of the process satisfies a forward Kolmogorov equation:

$$X_t = \text{s-mod}(Y_t) \Rightarrow \partial_t \rho(x,t) = \nabla \cdot \left(e^{-E(x)} \nabla \left(e^{E(x)} \rho(x,t)\right)\right)$$

We simulate this evolution through the discretization of both space and time:

$$\frac{d}{dt} |\rho\rangle = L |\rho\rangle$$

via the following approximation:

Fourier Differentiation

Let $u : [-\pi, \pi] \to \mathbb{R}$ be periodic

$$\partial u \approx F^{-1} \left(ik \left(Fu\right)_k\right)$$

Then, use the machinery of Berry et al. 2017 to solve for $|\rho(T)\rangle$.

A quick detour

For a string $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{Z}_+^d$, let

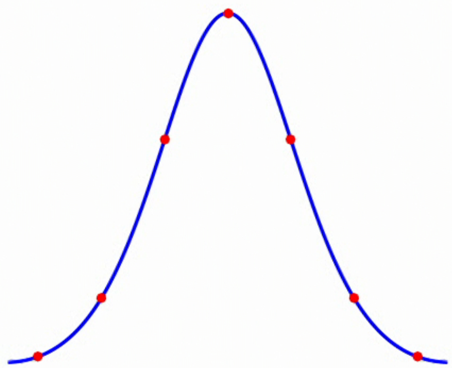
$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

Assume a periodic function $u : [-\pi, \pi]^d \to \mathbb{R}$. We say u is '*semi-analytic*', if there exist $C, a \geq 0$ such that:

$$\sqrt{\mathbb{E}\left(\sum_{\alpha:|\alpha|=m} |D^\alpha u(X)|^2\right)} \leq C\, a^m\, m!$$

Why Fourier?

Let $|u_N\rangle$ be a normalized state obtained from u, by taking N discretization points along each axis.



Why Fourier?

Let $|u_N\rangle$ be a normalized state obtained from u, by taking N discretization points along each axis.

Lemma

Assume u is (C, a)-semi-analytic and $N \geq 2ad$. Given $|\psi_N\rangle$, satisfying $\||\psi_N\rangle - |u_N\rangle\| \leq \delta$, there exists an efficient algorithm which samples from a distribution ϵ-close to u^2 where

$$\epsilon = \mathcal{O}\left(\delta + \frac{Ce^{-0.6N/a}}{\sqrt{\mathbb{E}\, u(X)^2}}\right)$$

Why Fourier?

Let $|u_N\rangle$ be a normalized state obtained from u, by taking N discretization points along each axis.

Lemma

Assume u is (C,a)-semi-analytic and $N \geq 2ad$. Given $|\psi_N\rangle$, satisfying $\||\psi_N\rangle - |u_N\rangle\| \leq \delta$, there exists an efficient algorithm which samples from a distribution ϵ-close to u^2 where

$$\epsilon = \mathcal{O}\left(\delta + \frac{Ce^{-0.6N/a}}{\sqrt{\mathbb{E}\, u(X)^2}}\right)$$

Furthermore, the Fourier method provides accurate estimates of the derivatives

Lemma

For a semi-analytic function, the Fourier approximation satisfies

$$\left\|\nabla^{(k)}u - \nabla_F^{(k)}u\right\| = \mathcal{O}\left(\|1\|\, C^k e^{-0.4N/a}\right)$$

for $k = 1, 2$.

Main Result

Combining all the aforementioned techniques and results, we prove the following theorem.

Theorem (Informal)

There is a quantum algorithm sampling from the Gibbs distribution associated with the periodic potential E, making

$$\mathcal{O}\left(a^4 d^7\, e^{3\Delta/2}\, \mathrm{polylog}\left(\frac{ade^{\Delta}C(1+lL)}{\epsilon}\right)\right)$$

queries to an oracle for the energy function. Also, the gate complexity of the algorithm is larger than the query complexity by a factor of $\mathrm{poly}(d)\,\mathrm{polylog}(Cade^{\Delta}(1+lL))/\epsilon)$.

Future Work

- Extending the work to other compact domains (Li and Erdogdu 2020).
- Estimation of partition functions (Childs et al. 2022).
- Uniform sampling from convex bodies (Lovász and Vempala 2004).
- Improving the upper bounds (Krovi 2022).
- Assuming access to first-order derivatives?

Thank You!