A Quantum Algorithm for Gibbs Sampling

Sriram Gopalakrishnan

August 2021

1 Introduction

Statistical Mechanics (SM) has been extremely successful in our understanding of nature, and has been adapted to several different areas of science. The beauty of equilibrium SM is the fact that a small number of fundamental quantities can be used to derive a plethora of interesting physical properties. There are three basic mathematical objects that can for one: be formalized in different ways to different physical systems, and additionally: be used to derive an abundance of other important system properties such as entropy, free energy and so on. These basic objects are: the density of states, the occupancy distribution, and the partition function.

Given a classical or quantum system described by a many-body Hamiltonian H, the eigen-spectrum of H characterizes the *density of states* of the system, which is a function $g(\lambda)$ that counts how many eigenstates have eigenvalue λ, also called *degeneracy*. We then have an *occupancy distribution* $p(\lambda)$ that describes how likely it is for a single quasiparticle to occupy a state with eigenvalue λ, or in the context of quantum information: how likely it is for a mixed state to be in a pure state with eigenvalue λ when measured in the energy basis. Lastly, the *partition function* Z is the sum

$$Z = \sum_{\lambda : p(\lambda) g(\lambda) > 0} e^{-\beta\lambda} = \text{tr}(e^{-\beta H}) \tag{1}$$

which is a sum of exponentials over states that are allowed and can be occupied, and where *inverse temperature* $\beta = \frac{1}{k_B T}$ is a measure of the equilibrium temperature T of the many-body system.

Empirically, at very low temperatures, the occupancy distribution $p(\lambda)$ is strongly skewed towards the *ground state* of H. Now, in a highly ideal situation, if we "slowly heat" the system to equilibriate at inverse temperature β, the occupancy distribution is known to be well described by a *gibbs distribution*

$$p(\lambda) = \frac{e^{-\beta\lambda}}{Z} \tag{2}$$

This process is known as *thermalization.* In the context of quantum computing, we perform unitary transformations on a system of artificial spins. At an elementary layer of abstraction, the computation can be thought of as time-evolving an Ising-like Hamiltonian defined on a graph $G = (V, E)$ describing allowed two-qubit couplings. Not all two-qubit couplings are allowed due to spatial restrictions on the electromagnetic fields used to implement these gates. Now, one may argue in a similar fashion as the previous para that if we "slowly heat" a quantum computer from its ground state (say $|0^n\rangle$) to a finite inverse temperature β, then the density matrix of the system should evolve to a *gibbs state* ρ_G defined as

$$\rho_G = \frac{e^{-\beta H}}{\text{tr}(e^{-\beta H})} = \sum_{j=0}^{N-1} \frac{e^{-\beta E_j}}{Z} |\psi_j\rangle \langle\psi_j| \tag{3}$$

where the Hamiltonian is expanded in its eigenspace as $H = \sum_j E_j |\psi_j\rangle \langle\psi_j|$ and $N = 2^n$ (Note that the gibbs state is a mixed state, and can be thought of as an ensemble of pure states weighted by the gibbs distribution).

Sadly, this is a bad argument for one overarching reason: interacting many-qubit systems are never in equilibrium. Even at a fixed temperature, the moment we turn on couplings on the graph G we mentioned earlier, we begin time evolving the system's state with the unitary e^{-iHt} if we created a Hamiltonian H. Besides, this is a simplistic view of time evolution given the presence of many unwanted degrees of freedom in a realistic system.

However, that said, it is very natural to consider the following artificial thermalization setting at a fixed near-zero temperature introduced by Poulin and Wocjan in 2009 [1]. Suppose we as problem solvers are given a mathematical description of a Hamiltonian H as a *frustration-free* projector decomposition

$$H = \sum_{k=1}^{K} \alpha_k \Pi_k \tag{4}$$

So each Π_k is hermitian and satisfies $\Pi_k^2 = \Pi_k$, and the frustration-free property means that the ground state of H is also a ground state of Π_k for all k. We will require this property for technical reasons as we'll see later on. The sum here runs over $K = \text{poly}(n)$ terms. Also suppose that we are given a quantum computer in the state $|0^n\rangle$ with the ability to incorporate polynomially many ancillas into our computation.

Then, formally, the computational task of *gibbs state preparation* is to give a description of a unitary V such that

$$\hat{\rho} = \text{tr}_a \left(V(|0^n\rangle \langle 0^n| \otimes |0\rangle \langle 0|_a) V^\dagger \right) \tag{5}$$

satisfies

$$\frac{1}{2} \|\hat{\rho} - \rho_G\|_1 \leq \epsilon \tag{6}$$

Basically, performing the computation on n qubits plus a certain number of ancillas "a" and then tracing out (measuring) all the ancillas should leave us with an n-qubit mixed state that is ϵ-close to the gibbs state ρ_G (equation 3) in trace norm for an arbitrary given precision ϵ.

In previous work, Poulin and Wocjan used Phase Estimation (PE) to develop an algorithm for this problem with gate complexity scaling as $O\left(\frac{\beta^6}{\epsilon^3}\sqrt{\frac{N}{Z}}\text{polylog}(\epsilon^{-1})\right)$. In the present work [2], the authors use several ideas to develop an algorithm that scales as $O\left(\sqrt{\frac{N\beta}{Z}}\text{polylog}\left(\frac{1}{\epsilon}\sqrt{\frac{N\beta}{Z}}\right)\right)$.

In the next section, we'll see how this algorithm works.

2 The Algorithm

The line of reasoning that motivates the approaches used in this algorithm can be explained as follows.

First, it is to be understood that in order to construct an approximation of the gibbs state $e^{-\beta H}$ (upto normalization), we wish to develop a gate based approximation to the non-unitary matrix $e^{-\frac{\beta H}{2}}$. The rationale here is that applying $e^{-\frac{\beta H}{2}}$ to a "suitably chosen" pure state can give us a *purification* (HJW theorem) of the gibbs state $e^{-\beta H}$. And hence, a partial trace over the ancillas of that purification will give us the required mixed state $e^{-\beta H}$. Of course we'll have to see what that "suitably chosen" pure state is and account for a unitary constructing it from $|0^n\rangle$ plus ancillas.

So the basic goal of the algorithm now is to construct a gate based approximation of $e^{-\frac{\beta H}{2}}$. As already mentioned, this is a non-unitary matrix in general. However, a powerful result from fourier analysis known as the Hubbard-Stratonovich transformation can be used to accurately approximate $e^{-\frac{\beta H}{2}}$ as a *linear combination of unitaries*. The transformation can be invoked as follows:

$$e^{-\frac{\beta H}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dy\ e^{-\frac{y^2}{2}}\ e^{-iy\sqrt{\beta H}} \tag{7}$$

Notice that inside the integral, we have a term like $e^{-iy\sqrt{\beta H}}$ which is unitary. Although this is an integral over a continuous variable y, we can easily discretize it, which we consider in Lemma 1.

The thing to worry about with the term $e^{-iy\sqrt{\beta H}}$ is the $\sqrt{H}$ inside it. In the projector decomposition $H = \sum_{k=1}^{K} \alpha_k \Pi_k$, we do not in general enforce that the projectors Π_k be mutually orthogonal. So it is difficult in general to compute $\sqrt{H}$ in the same Hilbert space as H.

Here is where the frustration-free property introduced in the context of Spectral Gap Amplification [3] comes handy. We can construct a $\sqrt{H}$ in a larger Hilbert space, as the matrix $\widetilde{H}$, defined as

$$\widetilde{H} = \sum_{k=1}^{K} \sqrt{\alpha_k} \Pi_k \otimes (|k\rangle\langle 0| + |0\rangle\langle k|)_{a_1} \tag{8}$$

where we introduce an ancillary register a_1 of size $O(\log K)$ qubits. It turns out that if H is frustration-free, then $\widetilde{H}$ has the same set of eigenstates as $H \otimes I$, although with a larger spectral gap.

In particular, if H has spectral gap Δ, and $\widetilde{H}$ has spectral gap $\widetilde{\Delta}$, then $\widetilde{\Delta} \geq \Omega(\sqrt{\Delta})$ according to Spectral Gap Amplification. This square root equivalence of eigenstates allows one to note that

$$(\widetilde{H})^2 |\phi\rangle \otimes |0\rangle_{a_1} = (H|\phi\rangle) \otimes |0\rangle_{a_1} \quad \text{for all} \quad |\phi\rangle \in \mathbb{C}^{2^n} \tag{9}$$

Hence, in a sense, the action of $\widetilde{H}$ is equivalent to $\sqrt{H}$ in this larger Hilbert space. So we can replace $\sqrt{H}$ in equation (7) by $\widetilde{H}$ and subsequently discretize the integral as a linear combination of unitaries.

Lemma 1: Consider the following discretization of equation (7) as the matrix X_β

$$X_\beta = \frac{1}{\sqrt{2\pi}} \sum_{j=-J}^{+J} c_j \exp\left(-iy_j \sqrt{\beta}\ \widetilde{H}\right)$$

where we use a step size denoted δy and total interval length $2J\delta y$. Also, $c_j = e^{-\frac{y_j^2}{2}}\delta y$ and $y_j = j\delta y$. Then, if $J = \Theta(\sqrt{\|H\|\,\beta}\log(\epsilon^{-1}))$, $\log(\epsilon^{-1}) \geq 4$, and $\|H\|\,\beta \geq 4$, we will have an L_2 norm ϵ-closeness as follows

$$\left\| \left(X_\beta - (e^{-\beta H/2} \otimes I)\right) |\phi\rangle|0\rangle_{a_1} \right\| \leq \epsilon/2 \quad \text{for all} \quad |\phi\rangle\,|0\rangle_{a_1} \in \mathbb{C}^N \otimes \mathbb{C}^K \tag{10}$$

We skip the proof, but it may be studied independently in the reference.

Essentially, the lemma states that the discretization is a good approximation to $e^{-\beta H/2}$ if certain mild constraints on the required precision ϵ, the spectral norm of H, and the number of terms J, are satisfied.

Hence, we now have an approximation X_β to $e^{-\beta H/2}$ that is a linear combination of unitaries $\{e^{-iy_j\sqrt{\beta}\widetilde{H}}\}$ weighted by coefficients $\{c_j\}$ that is computed classically.

So our next task is to get a gate based approximation W_j of each unitary $\exp\left(-iy_j\sqrt{\beta}\ \widetilde{H}\right)$ in the linear combination. Notice that each unitary entails time-evolving $\widetilde{H}$ for time $y_j\sqrt{\beta}$ for each j. Hence we invoke a suitable Hamiltonian Simulation algorithm as a subroutine for this purpose.

To proceed, we first express $\widetilde{H}$ as a linear combination of unitaries. If we denote $M_k := (|k\rangle\langle 0| + |0\rangle\langle k|)_{a_1}$, then note by Euler's formula that

$$M_k = \frac{i}{2}\left(e^{-i\pi M_k/2} - e^{i\pi M_k/2}\right) \tag{11}$$

Also note that for any projector Π_k, we have a unitary U_k such that

$$\Pi_k = \frac{1}{2}(I + U_k) \tag{12}$$

If we substitute equations (11) and (12) into our definition of $\widetilde{H}$ in equation (8), we can express $\widetilde{H}$ as a linear combination of unitaries expressible in terms of known quantities in the problem.

$$\widetilde{H} = \sum_{k=1}^{K} \sqrt{\alpha_k}\Pi_k \otimes M_k \equiv \sum_{k=1}^{\widetilde{K}} \widetilde{\alpha}_k \widetilde{U}_k \tag{13}$$

It is easy to note that $\widetilde{K} = 4K$ and $\widetilde{\alpha}_k = \sqrt{\alpha_{\lfloor k/4 \rfloor}}$, while $\widetilde{U}_k$ is a unitary directly expressible in terms of U_k and M_k.

Now we will use reference [4] to perform efficient Hamiltonian simulation. The algorithm assumes the ability to implement the unitary $\widetilde{Q}$ defined as

$$\widetilde{Q} = \sum_{k=1}^{\widetilde{K}} \widetilde{U}_k \otimes |k\rangle\langle k|_{a_2} \tag{14}$$

where we introduce a second set of ancillas a_2 of size $O(\log K)$ qubits. The method uses queries to $\widetilde{Q}$ to approximate $e^{-i\widetilde{H}t}$. We will not go into further detail here, but [BCCKS 14] can be used to develop gate constructions of unitaries W_j that approximate $\exp\left(-iy_j\sqrt{\beta}\ \widetilde{H}\right)$ for each j, with gate complexity that scales linearly with K and logarithmically with the evolution time. Specifically, the gate complexity of implementing each W_j will be

$$O\left((K\ +\ C_U \log K)\tau \frac{\log(\tau/\epsilon)}{\log\log(\tau/\epsilon)}\right) \tag{15}$$

where $\tau = |t|\sum_k \widetilde{\alpha}_k$, and C_U is the gate complexity of implementing each $\widetilde{U}_k$.

So at this point, we know both the coefficients $\{c_j\}$ and gate constructions of unitaries $\{W_j\}$ such that $X_\beta = \sum_j c_j W_j$ is ϵ-close to $e^{-\beta H/2}$ over its action on the class of states we are interested in. It is worth noting that W_j acts on a vector space of dimension $O(n + \log K)$

We now have to develop an approximate gate construction of $X_\beta = \sum_j c_j W_j$, which is a linear combination of unitaries. For this, we invoke the *Linear Combination of Unitaries (LCU)* algorithm [5] (which we studied in assignment 1) as a subroutine.

[The LCU protocol generally works by assuming the ability to implement the unitary

$$SW = \sum_{j=-J}^{+J} W_j \otimes |j\rangle \langle j|_{a_3} \tag{16}$$

where we introduce our third set of ancillas a_3 of size $|a_3| = O(\log(2J+1))$ qubits.

It will also assume the ability to implement the unitary B that acts like

$$B\,|0^{|a_3|}\rangle = \frac{1}{\sqrt{\sum_j c_j}} \sum_j \sqrt{c_j}\,|j\rangle \tag{17}$$

so that it can implement $R := (I \otimes B^\dagger) SW (I \otimes B)$ on an $O(n + \log(K(2J+1)))$ dimensional register. Uses of R along with Amplitude Amplification can be used to prepare $\sum_j c_j W_j |0\rangle$ in the first $O(n + \log K)$ dimensional register]

However, in this work, the LCU protocol is implemented in a more sophisticated setting. This is where the "suitably chosen" pure state we mentioned in the start comes into the picture.

The first step in the construction of the unitary V of equation (5) will involve preparing a maximally entangled state

$$|\phi\rangle = \frac{1}{\sqrt{N}} \sum_{\sigma=0}^{N-1} |\sigma\rangle|\sigma\rangle_{a_4} \tag{18}$$

where we introduce a set of ancillas a_4 of size n qubits. The above state is equivalent to the maximally entangled state $(\sum_j |\psi_j\rangle\,|\psi_j\rangle)$ for the eigenbasis of a positive Hermitian H.

We then implement $X_\beta\,|\phi\rangle$ via the LCU protocol introducing three more sets of ancillas a_1, a_2 and a_3 as previously discussed.

Then it can be seen that $X_\beta\,|\phi\rangle\,|0\rangle$ approximates a purification of the gibbs state given by $e^{-\beta H/2}\,|\phi\rangle\,|0\rangle$, and hence tracing out all the ancillas gives us an approximation of the Gibbs state.

3 Conclusion

In summary, the authors have developed a quantum algorithm for state preparation with an effective gate complexity of $O\left(\sqrt{\frac{N\beta}{Z}}\text{polylog}\left(\frac{1}{\epsilon}\sqrt{\frac{N\beta}{Z}}\right)\right)$. Although they do boast a polynomial improvement in β, the key scaling with $\sqrt{N}$ is still exponential. Nevertheless, they find that the dependence on precision ϵ is only logarithmic, which corresponds to an exponential improvement over previous work by Poulin and Wocjan [1].

References

[1] David Poulin and Pawel Wocjan. "Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer". In: *Physical review letters* 103.22 (2009), p. 220502.

[2] Anirban Narayan Chowdhury and Rolando D Somma. "Quantum algorithms for Gibbs sampling and hitting-time estimation". In: *arXiv preprint arXiv:1603.02940* (2016).

[3] Rolando D Somma and Sergio Boixo. "Spectral gap amplification". In: *SIAM Journal on Computing* 42.2 (2013), pp. 593–610.

[4] Dominic W Berry et al. "Exponential improvement in precision for simulating sparse Hamiltonians". In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing.* 2014, pp. 283–292.

[5] Dominic W Berry et al. "Simulating Hamiltonian dynamics with a truncated Taylor series". In: *Physical review letters* 114.9 (2015), p. 090502.