QMA-completeness of the Local Hamiltonian Problem - Part I

(Based on lecture notes by Henry Yuen)

Sanchayan Dutta

15 November, 2022

Table of Contents

Classical-Quantum Dictionary

As seen previously, there is a correspondence between concepts of computer science and the language of physicist.
These can be summarized in the following table:

Classical	**Quantum**
Constraint Satisfaction Problem (CSP)	Hamiltonian
Variables	Qubits
Constraints	Hamiltonian terms
Solution quality	Energy
Optimal solution	Ground state
P	BQP
NP	QMA
Cook-Levin SAT formula	Feynman-Kitaev Hamiltonian

Because of their analogy with k-local CSPs, we introduced k-local Hamiltonians and described some examples. The first one was the classical Ising model which was analogue to Max-Cut. This model is said to be *classical* since the Hamiltonian is diagonal in the computational basis. We now describe the quantum version of the Ising model.

A Hamiltonian H which is diagonal in a *product basis* $\mathcal{B}$ is sometimes described as classical, because the evolution of any initial state in $\mathcal{B}$ driven by such H does not give rise to quantum effects such as superposition and entanglement. Moreover, a measurement in $\mathcal{B}$ at any point during the evolution yields a single, deterministic result and does not entail wavefunction collapse.

The quantum Ising model

Consider N qubits (or spins) arranged on a ring, where we associate qubit $N+1$ with qubit 1. The transverse field Ising model describes a nearest-neighbor magnetic dipole interaction along the Z axis when all spins are subject to a transverse magnetic field along the X axis. The family of such Hamiltonians is given by:

$$H(g) = -\sum_{i=1}^{N} Z_i \otimes Z_{i+1} - g \sum_{i=1}^{N} X_i$$

where g is a real parameter corresponding to the strength of the transverse field, Z_i and Z_{i+1} indicate the Pauli Z matrix acting on qubits i and $i+1$, and X_i is the Pauli X matrix acting on the i-th qubit.

For every value g there is a corresponding Hamiltonian with properties that can vary vastly. For example, if $g = 0$ we find the classical Ising model, which is the same as the Max Cut Hamiltonian, but on a ring graph. The ground state of $H(g = 0)$ is therefore a simple unentangled basis state, or bit string.

Similarly, when $g \gg 1$, the transverse field part of H dominates and the ground state is again classical and given by $|+\rangle^{\otimes N}$, or all spins aligned with the magnetic field.

However, for $g \neq 0$, the system truly becomes quantum, because the Hamiltonian terms don't necessarily commute anymore, *e.g.* $[Z_i \otimes Z_{i+1}, X_i \otimes I_{i+1}] \neq 0$. The Hamiltonian matrix is therefore not diagonal in the computational basis.

The ground states of these Hamiltonians will in general exhibit quantum entanglement and all sorts of interesting phases of matter.

The quantum Heisenberg model

A generalization of the quantum Ising model is the Heisenberg model which includes magnetic dipole interactions along the X, Y and Z axes. Notably, it models the behavior of quantum magnetism in atomic systems. It's general Hamiltonian may be written as

$$H(J_x, J_y, J_z, g) = J_x \sum_i X_i X_{i+1} + J_y \sum_i Y_i Y_{i+1} + J_z \sum_i Z_i Z_{i+1} + g \sum_i X_i,$$

where J_x, J_y, J_z and g are real parameters and X_i, Y_i, and Z_i are Pauli matrices acting on qubit i.

Again, the ground state properties of $H(J_x, J_y, J_z, g)$ depend largely on the values of the parameters. As in the case of the quantum Ising model, setting $J_x = J_y = 0$ and $g = 0$ gives you the classical Max-Cut problem since the resulting Hamiltonian is diagonal. However, taking $J_x = J_y = J_z$ and $g = 0$, we find a non-diagonal Hamiltonian which we call "quantum Max-Cut".

In this case, the ground state becomes a non-trivial entangled state. To gain some intuition on how this arises, let's first focus on any pair of adjacent qubits. Along each edge $(i, i+1)$, the state of qubits i and $i+1$ that would get the lowest energy is the so-called singlet state

$$|\psi_-\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle \right)$$

which is one of the four Bell states. We can relate $|\psi_-\rangle$ to the EPR pair $|\phi_+\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle \right)$ by applying the Pauli X then Pauli Z operator on the first qubit. As the EPR pair, $|\psi_-\rangle$ is a maximally entangled state.

However, while each local term wants two consecutive qubits to be maximally entangled, this is impossible to satisfy globally: there is no quantum state where the state of qubit i and $i+1$ is maximally entangled, and the state of qubit $i+1$ and $i+2$ is also maximally entangled. This is a phenomenon called *monogamy of entanglement* that says that entanglement is not something that can be freely shared, unlike classical correlations (this is morally related to the no-cloning principle). Thus the ground state of the Heisenberg model gets really interesting because it's trying to satisfy all these local demands as best as possible, but it won't be able to do so perfectly.

It is worth noting that the 1D Ising model and Heisenberg model are rare examples of Hamiltonians that can be solved exactly using the Bethe ansatz. How exactly this is achieved is beyond the scope of this course and requires a lot beautiful mathematical physics – that's really a task for a condensed matter theory course.

Quantum Complexity Classes

We'll now take a step back and think about the problem more abstractly, and consider the task of solving *general* local Hamiltonians. Let's define the following decision problem, called k-$\mathrm{LOCAL\text{-}HAM}_{a,b}$ where $a < b$ are real numbers. The instances of k-$\mathrm{LOCAL\text{-}HAM}_{a,b}$ are going to be all k-local Hamiltonians $H = H_1 + \cdots + H_m$ such that the operator norm of each H_i is at most 1 (this is to ensure consistent normalization), and in the YES case the ground energy $\lambda_{\min}$ of H is at most a (i.e. the Hamiltonian's ground energy is "low"), and in the NO case the ground energy of H is at least b (the ground energy is "high"). We ignore all Hamiltonians that don't fall into either category. Formally, we write

$$\begin{aligned} &\text{YES instance} && \lambda_{\min}(H) \leq a \\ &\text{NO instance} && \lambda_{\min}(H) \geq b. \end{aligned}$$

How is a local Hamiltonian presented? Well, there's a lot of flexibility, but here's a convenient way to describe local Hamiltonians:

1. the number of qubits n,
2. the number of terms m,
3. for each term $i = 1, \ldots, m$, we write what subset $S_i \subseteq [n]$ of k qubits the i-th term H_i is going to non-trivially act on, and then a description of the $2^k \times 2^k$ matrix h_i such that $H_i = h_i \otimes I^{\otimes n-k}$.

If k is small and $m = \mathrm{poly}(n)$, then this presentation takes only $\mathrm{poly}(n)$ bits to describe.

So we've just defined a decision problem that models an important task in physics: try to figure out the ground state energies of different Hamiltonians. Of course, this is not the only thing one does – you'd also like to figure out what the ground state is, what properties it has, and so on, but as a start you first want to figure out the minimum eigenvalue.

We now turn to the complexity of quantum decision problems. We proceed in analogy with the classical case. For instance, just like classical CSPs have NP-completeness, the quantum analogue for quantum CSPs is QMA-completeness, as we will see later. Let's first investigate the quantum analogue of P, which is BQP.

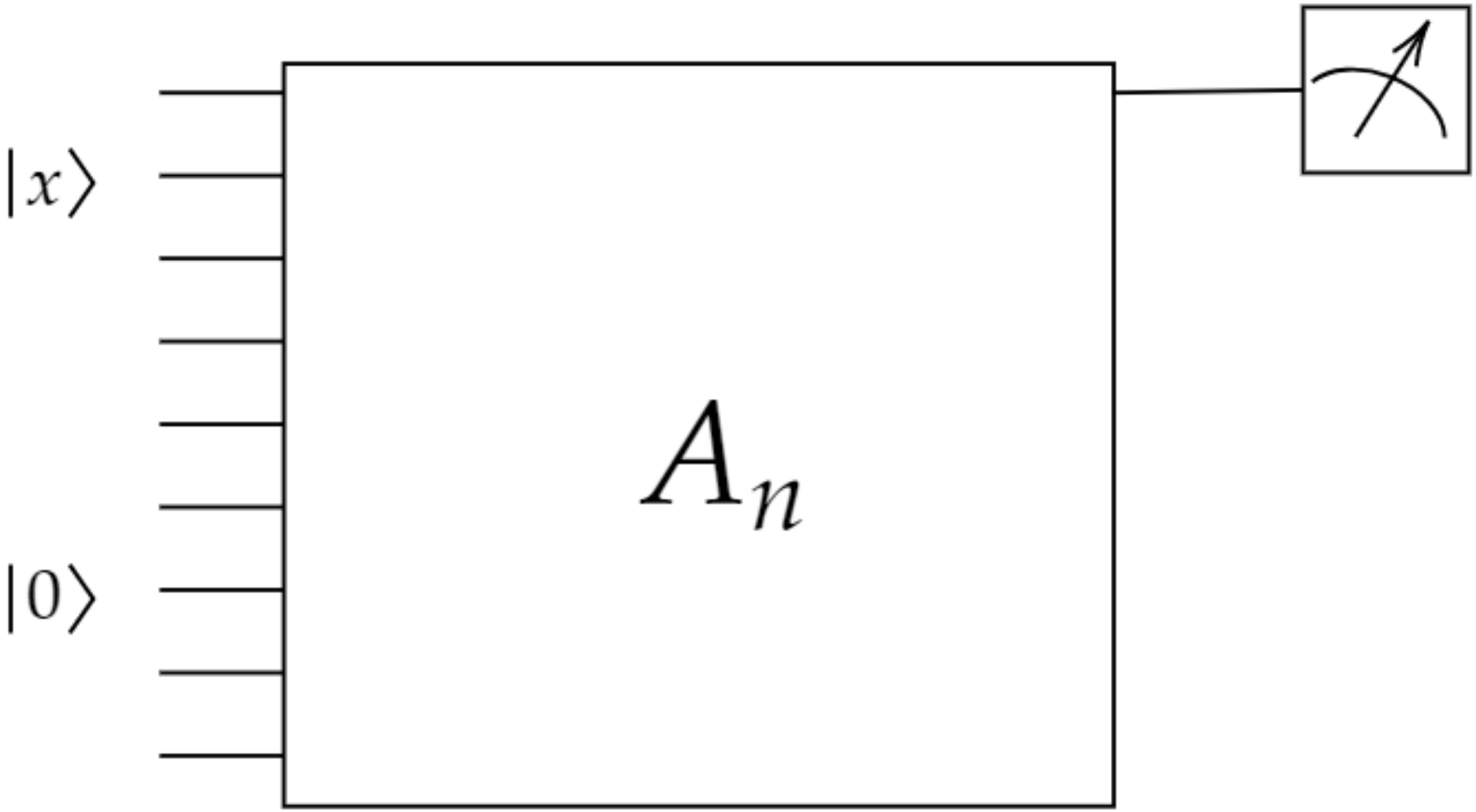
BQP

The quantum analogue of P is BQP (which stands for "Bounded-Error Quantum Polynomial Time"), which is the class of decision problems that can be decided by polynomial-time quantum algorithms with bounded error. That is, L is in BQP if there exists a polynomial time quantum algorithm A such that, if $x \in L_{yes}$, then $A(x) = 1$ with probability at least $2/3$, and if $x \in L_{no}$, then $A(x) = 1$ with probability at most $1/3$. The constants $2/3$ and $1/3$ are arbitrary, they can be set to any two distinct constants, say .99 and .01 – the gap between these numbers reflect the confidence you have in the output of the algorithm.

Quantum Algorithms

What do we mean by quantum algorithm? Well, there is something called a quantum Turing machine, but its definition is incredibly unwieldly and basically nobody uses it. Instead, people like to think about quantum circuits. So when we say polynomial-time quantum algorithm A, we really are referring to an infinite family of quantum circuits $\{A_n\}$ that are indexed by integers $n \in \mathbb{N}$. The circuits A_n have size at most $\mathrm{poly}(n)$, and when the input is some n-bit string x, we run the circuit A_n on input $|x\rangle$ and some ancilla qubits $|0\rangle$. The output bit of the circuit is obtained by measuring the first qubit at the end of the circuit in the standard basis.

An example of such circuit is shown in the figure on the next page. So basically there's a different circuit for each input length. We also insist that the infinite array of circuits $\{A_n\}$ are *uniformly generated*, meaning that there's a *classical* Turing machine M that, when given input n, outputs the description of the n-th circuit A_n. This is to ensure that all of the circuits A_n are all "related to each other," and it's unlike having a completely different algorithm for each input length. If you're not so familiar with this distinction of uniform-vs-non-uniform family of circuits, don't worry about it, it won't be so essential to what we're talking about.



BQP Verifier Circuit A_n acting on the input $|x\rangle$ and some ancilla qubits initialized in $|0\rangle$.

QMA

The quantum analogue of NP, called QMA (which stands for "Quantum Merlin-Arthur"), is naturally defined as the following: a decision problem L is in QMA if there exists a family of polynomial-sized *verifier circuits* $\{A_n\}$ such that if $x \in L_{yes}$, then there exists a polynomial-sized *proof state* $|\psi\rangle$ where $\Pr[A_n \text{ accepts } |x\rangle \otimes |\psi\rangle] \geq 2/3$, where $n = |x|$, the number of qubits needed to encode $|x\rangle$, and if $x \in L_{no}$, then *for all* proof states $|\psi\rangle$ we have $\Pr[A_n \text{ accepts } |x\rangle \otimes |\psi\rangle] \leq 1/3$.
In other words, to determine whether x is a YES or NO instance, the verifier circuit gets a proof "from the sky" to help it determine which is the case – with the twist that the *proof* is now a quantum state!

QMA

The name "Quantum Merlin-Arthur" is based on medieval folklore about King Arthur and the Knights of the Round Table. King Arthur gets advice from the all-powerful wizard Merlin, but doesn't necessarily trust everything Merlin says. So Arthur has to *verify* what Merlin tells him. We think of Arthur as being a polynomial-time verifier, and Merlin as a *prover* who can solve any (computational) problem he likes. Merlin tries to convince Arthur of some statement X by sending Arthur a proof, which Arthur then checks.

In Quantum Merlin-Arthur, Arthur is a polynomial-time *quantum verifier*, and Merlin can send *quantum* proofs to Arthur to verify.

A quantum notion of proofs

Quantum proofs define a really intriguing model of "mathematical proof". The traditional notion of a mathematical proof is that there's a formal statement X, which may or may not be true. Maybe X is something like "There are infinitely many primes" or "P $\neq$ NP". And someone tries to convince you that X is true by giving you a table of text π, that you can check line-by-line whether π is a valid deduction, using the standard mathematical axioms that we've learned in math class, of the statement X.

A quantum notion of proofs

However, with QMA, we've changed this notion of a proof: someone can try to convince you of the truth of some statement X by handing you a *quantum proof* $|\pi\rangle$. And to verify this proof, you can perform some quantum measurement on the state to determine whether you're convinced or not. However the quantum measurement do not need to resemble the traditional notion of line-by-line proof checking. It can be something more exotic, like first performing a Quantum Fourier Transform on $|\pi\rangle$.

Furthermore, the fact that the proof is a quantum state makes it very distinct from a traditional proof, because for one you can't *copy* the proof and share it with your friends, due to the No-Cloning Theorem. Also, it may be difficult to extract any information from the quantum proof other than the fact that X is true. Let's say you make a measurement on part of the quantum proof to try to "read" it. This will generally collapse the quantum proof, and change the state. So it's a really unusual notion of proof.

Despite their strangeness, we believe that quantum proofs can be much more efficient for verifying the truth of certain types of statements X than if you were forced to check an equivalent classical proof. Examples of such statements X include:

1. (**Local Hamiltonians**) "Local Hamiltonian H on n qubits has a ground state with energy less than $a = 0.1$."
2. (**Consistency of local density matrices**) "Here is a collection of two-qubit density matrices on every pair (i, j) of qubits: $\{\rho_{ij}\}$. There exists a global n-qubit state $|\psi\rangle$ where the reduced density matrix of $|\psi\rangle$ on qubits (i, j) is exactly ρ_{ij}."
3. (**Group Non-Membership**) "An element h of a finite group G is not in the subgroup generated by elements $g_1, \ldots, g_k \in G$". Here, think of G as having exponentially many elements.

Group Non-Membership Problem

As an example, we now detail this last problem.

Recall that a group G is a set with an identity element e, and it's closed under an invertible binary operation (that we'll call *multiplication*). If you have a subset of elements $\{g_1, \ldots, g_k\} \subseteq G$, the subgroup H generated by this subset is simply all possible products of the g_i's and their inverses.

Here is the problem, suppose you have some finite group G with $\exp(n)$ many elements, so you can represent each element $g \in G$ using $\mathrm{poly}(n)$-many bits. Furthermore, given two elements $g, h \in G$, you can efficiently multiply them together to obtain a representation of $gh \in G$ in $\mathrm{poly}(n)$ time, and also you can compute inverses g^{-1} in $\mathrm{poly}(n)$ time.

Now suppose you're given elements $\{g_1, \ldots, g_k\} \subseteq G$, and an element $h \in G$. One question you might be interested in is the **Group Membership Problem**: determine whether h is in the subgroup H generated by $\{g_1, \ldots, g_k\}$. How might someone convince you that $h \in H$?

Well, they could give you a sequence of products of the g_i's and their inverses that multiply to h. That would be fine, except this sequence of products might be exponentially long – so this wouldn't constitute a polynomial-sized proof.

Fortunately, there's a nontrivial result in group theory that ensures there is always a polynomial-length sequence of products of generators and their inverses that multiply to any given subgroup element h. So this implies that the Group Membership Problem is in NP.

What about the opposite problem, the **Group Nonmembership Problem**? Here, you're given generators for a subgroup H and an element h from the parent group G. How can someone convince you that h is *not* in H?

This seems to be a much harder problem, because it's basically asking for the *non-existence* of a way to multiply elements of H together to get h. Proving non-existence seems to be much harder than proving existence.

We don't know what the classical complexity of the Group Nonmembership (GNM) Problem is (it might even be in NP), but it turns out using *quantum proofs* it is possible to efficiently solve the GNM problem.

To Be Continued!