Modern quantum complexity theory

Chinmay Nirkhe

University of Washington | CSE 599c Spring 2026

Contents

Chapter 1

Open Questions

1.1 Approach

Overview. The course is organized around open problems and research directions. Quantum complexity theory is not a finished story: many of its most basic structural questions remain unanswered, and the aim is to develop intuition for those questions rather than to present a closed body of results.

The course includes assigned problems, but participation is the main component of the grade. No separate project or formal scribing requirement is assumed, though volunteer scribes are welcome.

North-star principle. It is useful to adopt one or two guiding questions as a personal "North star" for the quarter. Each lecture can then be read in relation to those questions.

1.2 Hamiltonian Complexity

A central object is a local Hamiltonian

$$H = \sum_{i=1}^{m} h_i,$$

where each term h_i acts on k qudits and satisfies $\|h_i\| = 1$. Write

$$\lambda_{\min}(H)$$

for the minimum eigenvalue of H.

The local Hamiltonian problem asks us to distinguish between the cases

$$\lambda_{\min}(H) \le \alpha \qquad \text{and} \qquad \lambda_{\min}(H) \ge \beta,$$

under the promise that

$$\beta - \alpha = \frac{1}{\mathrm{poly}(n)}.$$

This problem is QMA-complete.

The local Hamiltonian problem is the quantum analogue of constraint satisfaction. Since the PCP theorem is the central structural result for classical CSPs, it is natural to ask for a quantum analogue.

Quantum PCP conjecture. There exists a constant $\varepsilon > 0$ such that the local Hamiltonian problem remains QMA-complete even when the promise gap is linear in the number of terms, namely

$$\beta - \alpha = \varepsilon m.$$

A central theme is to understand why the classical PCP proof has resisted quantization. This leads to a broader guiding question: are there natural subclasses of local Hamiltonians that are easier than the full LH problem?

1.3 Friends of QMA

Several nearby complexity classes and Hamiltonian subclasses help illuminate the structure of QMA.

1.3.1 Commuting local Hamiltonians

A first subclass is the commuting local Hamiltonian problem. Here one considers

$$H = \sum_i h_i \qquad \text{such that} \qquad [h_i, h_j] = 0 \text{ for all } i, j.$$

We use CLH for the commuting problem and LH for the general local Hamiltonian problem.

Because

$$\mathsf{CSP} \subseteq \mathsf{CLH} \subseteq \mathsf{LH},$$

we already know that CLH is NP-hard, and the classical PCP theorem implies NP-hardness for the corresponding approximation problem as well.

Open questions about CLH. What is the true complexity of commuting local Hamiltonians? Is CLH even MA-hard? More basically, is it even BPP-hard?

These questions are subtle. Since $\mathsf{BPP} \subseteq \mathsf{QMA}$, reductions from BPP problems to local Hamiltonians do exist. However, the known constructions rely on Kitaev's clock construction or on perturbation theory, and these constructions are inherently noncommutative.

There is some evidence that commuting Hamiltonians may be easier. In a suitable basis, each term h_i is

diagonal, and one may moreover assume without loss of generality that each h_i is a projection. Simultaneous diagonalization is easy for any small collection of terms, but finding a global decomposition is difficult in general.

For some families, the commuting local Hamiltonian problem is known to lie in NP. Two examples are:

- 2-local Hamiltonians on qudits of arbitrary local dimension;
- 2-dimensional local Hamiltonians on qubits.

Both results rely on the structural lemma of Bravyi and Vyalyi, which decomposes a CLH instance into a problem with an NP witness.

At the same time, many Hamiltonians arising from quantum error correction are commuting local Hamiltonians, so the commuting setting can still support highly nontrivial entanglement.

CLH **PCP question.** How hard is it to distinguish between

$$\lambda_{\min}(H) = 0 \qquad \text{and} \qquad \lambda_{\min}(H) > \varepsilon m?$$

A related question is whether this problem is strictly easier than, or just as hard as, distinguishing between

$$\lambda_{\min}(H) = 0 \qquad \text{and} \qquad \lambda_{\min}(H) \geq 1.$$

A possible advantage of the commuting setting is that some cloning-related obstacles to the general quantum PCP program may disappear for CLH.

1.3.2 Stoquastic Hamiltonians

Stoquastic Hamiltonians can be viewed as Hamiltonians with no "sign problem." Concretely, each local term h_i is real and has nonpositive off-diagonal entries in the standard basis.

This structure implies that a minimum-eigenvalue eigenvector can be chosen to have only nonnegative coordinates in the standard basis. With additional machinery, this yields an AM algorithm. Every MA promise problem reduces to a stoquastic Hamiltonian problem, and approximating stoquastic Hamiltonians is known to lie in NP.

Classification for 2-local Hamiltonians. Every 2-local Hamiltonian problem falls into one of four types,

$$\mathsf{P}, \qquad \mathsf{NP}, \qquad \mathsf{StoqMA}, \qquad \mathsf{QMA}.$$

Moreover,

$$\mathsf{MA} \subseteq \mathsf{StoqMA} \subseteq \mathsf{AM}.$$

Whether a similarly clean classification extends to other families of local Hamiltonians remains open.

1.3.3 QCMA

The class QCMA replaces the quantum witness of QMA by a classical witness, while Arthur remains a BQP verifier. Thus

$$\mathsf{MA} \subseteq \mathsf{QCMA} \subseteq \mathsf{QMA}.$$

There are known oracle separations between these classes, including classical oracle separations.

A central quantitative question is to determine for which functions q and t one has

$$\mathsf{QMA} \subseteq \mathsf{QCMA}[q, t].$$

Two coarse upper bounds are

$$\mathsf{QMA} \subseteq \mathsf{QCMA}[2^{O(n)}, O(\log n)] \cap \mathsf{QCMA}[0, 2^{O(n)}],$$

with the first coming from the PCP theorem together with $\mathsf{QMA} \subseteq \mathsf{PP}$, and the second coming from brute force.

Open question. Let QMA_ε denote the constant-gap energy-approximation version of the local Hamiltonian problem, that is, the "QPCP" version in which one approximates to energy εm. What is the best QCMA upper bound on QMA_ε?

This question interacts naturally with the quantum PCP conjecture.

1.3.4 QMA(2)

In $\mathsf{QMA}(2)$, Arthur receives two unentangled quantum proofs from two Merlins. The basic containments are

$$\mathsf{QMA} \subseteq \mathsf{QMA}(2) \subseteq \mathsf{NEXP}.$$

The power of unentangled proofs remains poorly understood. In particular, no oracle separation is known in either direction.

The Blier–Tapp protocol shows that NP has a two-proof system for 3-coloring with logarithmic-size proofs and inverse-polynomial soundness. By contrast, if

$$\mathsf{NP} \subseteq \mathsf{QMA}[\log n, 1/\mathrm{poly}(n)],$$

then

$$\mathsf{NP} \subseteq \mathsf{BQP}.$$

This highlights a sharp difference between ordinary logarithmic-size QMA proofs and the two-prover

setting.

1.3.5 Quantum games and PCP

Nonlocal games provide another route toward PCP-type questions. In a typical two-prover quantum game, Alice and Bob may share entanglement, they receive questions x and y from the verifier, they return answers a and b, and the verifier evaluates a predicate of the form

$$V(a, b \mid x, y).$$

Given such a game, one may ask for its optimal quantum value.

Theorem. Deciding the optimal quantum value of such a game is RE-complete. Equivalently,

$$\mathsf{MIP}^* = \mathsf{RE}.$$

In particular,

$$\mathsf{QMA} \subseteq \mathsf{MIP}^*.$$

One can then ask whether this inclusion admits a prover-efficient realization. For a QMA problem with witness $|\psi\rangle$, such a prover-efficient protocol does exist.

1.4 Interactive Proofs

An interactive proof consists of a sequence of quantum messages exchanged between Merlin and Arthur.

Theorem.

$$\mathsf{QIP} = \mathsf{QIP}(3) = \mathsf{IP} = \mathsf{PSPACE}.$$

In particular, every quantum interactive proof has a three-round quantum equivalent, and every PSPACE problem has a three-round protocol.

This raises two further questions: what is the power of two-round interaction, and what happens when the communication is classical but Arthur remains quantum?

Conjectural inclusion.

$$\mathsf{P}^{\#\mathsf{P}} \subseteq \mathsf{QIP}(2)$$

with classical proofs.

Since

$$\mathsf{BQP} \subseteq \mathsf{PSPACE} = \mathsf{IP},$$

there are certainly interactive proofs for quantum computation in principle. The harder question is whether there are *efficient* interactive proofs for BQP, or whether one can instead construct an oracle counterexample.

For the desired notion of efficiency, a yes-instance should admit a communication strategy computable in quantum polynomial time, whereas for a no-instance every communication strategy should be rejected with high probability.

1.5 State Complexity and Description Complexity

Some states are especially important in quantum computation, especially the ground states of local Hamiltonians. The question is not only whether one can decide properties of these states, but also how difficult it is to *describe* them.

When the ground energy is 0, the limiting operator

$$\lim_{\beta \to \infty} e^{-\beta H}$$

formally isolates the ground space. In that tautological sense, all information about the ground state is already encoded in H itself.

However, that description is not useful to a BQP machine unless $\mathsf{BQP} = \mathsf{QMA}$. A more operational notion of description is a circuit that prepares the state $|\psi\rangle$. This leads to the question: what is the minimum circuit size or depth required to generate $|\psi\rangle$?

Consequences. If $\mathsf{QMA} \neq \mathsf{QCMA}$, then there are superpolynomial circuit-depth lower bounds. Combining $\mathsf{QMA} \neq \mathsf{QCMA}$ with the QPCP perspective further points to superpolynomial circuit-depth lower bounds for all low-energy states.

At present, no robust lower bounds are known beyond roughly logarithmic depth.

1.6 Search-to-Decision and Cloning

1.6.1 Search-to-decision for QMA

Search-to-decision question. Given a QMA instance H, can one construct a ground state $|\psi\rangle$ using oracle access to QMA decision questions?

The classical analogy is search-to-decision for NP, where an oracle can be used to binary-search for an optimal proof. The quantum version is much less clear. One would like either a search-to-decision algorithm or a classical-oracle counterexample.

1.6.2 Complexity of cloning

Another question asks about the complexity of cloning ground states. Suppose one is given a Hamiltonian H together with one copy of a ground state $|\psi\rangle$. How hard is it to create a state $|\phi\rangle$ such that $|\phi\rangle$ is a ground state of

$$\mathbf{1} \otimes H + H \otimes \mathbf{1}?$$

The state $|\phi\rangle$ may itself be entangled if H is degenerate.

No complexity-theoretic lower bound for this problem is currently known. Nevertheless, public-key quantum money is built on the assumption that cloning is hard, so there are strong cryptographic motivations for studying this task.

1.7 Low-Depth Circuits, Queries, and Quantum Assumptions

1.7.1 Low-depth quantum circuits

Several low-depth or otherwise restricted models are relevant, including QNC_0, QAC_0, models with only a few non-Clifford gates, and the magic hierarchy. The corresponding open problem is whether one can prove state-preparation lower bounds or unitary lower bounds for these models. This topic is not expected to be a major focus of the course.

1.7.2 Query complexity

Query complexity is a setting where rigorous results can be proved. The highlighted problem is a unitary synthesis question: can an arbitrary unitary U on n qubits be implemented given access to ordinary gates together with superposition access to a Boolean function

$$f : \{0,1\}^{\mathrm{poly}(n)} \to \{0,1\}?$$

Restricted models are again relevant here, and only a one-query lower bound is currently known.

1.7.3 Cryptography from truly quantum assumptions

A final direction asks whether cryptography based on genuinely quantum assumptions can survive even under unusual complexity collapses. One example is hardness of cloning even in a world where

$$\mathsf{BQP} = \mathsf{QMA}.$$

A natural research program is to build oracle models exhibiting such phenomena.

Many relevant results are already known relative to *quantum* oracles. An open question is how far those results can be converted to *classical* oracles.

Chapter 2

QMA Upper Bounds, Clock Hamiltonians, Perturbation Theory, and One-Dimensional Local Hamiltonians

Overview. This chapter develops several standard tools in Hamiltonian complexity. It begins with two basic upper bounds for QMA, culminating in the inclusion

$$\mathsf{QMA} \subseteq \mathsf{PP}.$$

It then reviews Kitaev's clock-Hamiltonian construction and the reduction from quantum circuit satisfiability to the local Hamiltonian problem. Next it introduces perturbation theory, the projection lemma, and the effective-Hamiltonian expansion that underlies locality-reduction gadgets. The final part of the chapter turns to one-dimensional local Hamiltonians, tensor-network descriptions of quantum states, matrix product states, Schmidt rank, area laws, and the viable-space framework behind algorithms for gapped chains.

2.1 Upper Bounds for QMA

2.1.1 A trivial exponential-time upper bound

Theorem 2.1 (Trivial upper bound)**.**

$$\mathsf{QMA} \subseteq \mathsf{EXP}.$$

Proof. The local Hamiltonian problem is QMA-complete. An instance on n qubits is specified by a Hamiltonian

$$H \in \mathbb{C}^{2^n \times 2^n}.$$

Writing out this full matrix and diagonalizing it takes exponential time, so the minimum eigenvalue can be computed in EXP. Therefore every problem in QMA can be solved in exponential time. □

2.1.2 The class PP

Definition 2.2. *The class* PP *consists of decision problems solvable by a classical randomized computation such that*

$$\textit{yes instances satisfy } \mathbb{P}[\textit{accept}] \geq \frac{1}{2},$$
$$\textit{no instances satisfy } \mathbb{P}[\textit{accept}] < \frac{1}{2}.$$

Thus PP may be viewed as BPP without a promise gap. It is also closely related to counting complexity, and it is natural to think of it as a decision analogue of #P.

Exercise 2.3 (Counting formulation of PP)**.** Let

$$f : \{0,1\}^n \to \{-2^n - 1, \dots, 2^n - 1\}$$

be polynomial-time computable, and let

$$T \in [-2^{2n}, 2^{2n}].$$

Show that deciding whether

$$\sum_{x \in \{0,1\}^n} f(x) \leq T$$

can be done in PP.

2.1.3 The inclusion QMA ⊆ PP

Theorem 2.4.

$$\mathsf{QMA} \subseteq \mathsf{PP}.$$

Proof. Because the local Hamiltonian problem is QMA-complete, it suffices to prove the inclusion for a complete local Hamiltonian instance. After standard amplification and rescaling, we may assume the input is a Hamiltonian

$$H = \sum_{i=1}^{m} h_i$$

on n qubits with

$$0 \le h_i \le \mathbf{1},$$

and promise

$$\lambda_{\min}(H) \le 2^{-n} \qquad \text{or} \qquad \lambda_{\min}(H) \ge n^{-c}.$$

Define

$$H' := \mathbf{1} - \frac{H}{m},$$

so that

$$0 \le H' \le \mathbf{1}.$$

Then

$$\lambda_{\max}(H') = 1 - \frac{\lambda_{\min}(H)}{m}.$$

Hence, in a yes instance,

$$\lambda_{\max}(H') \ge 1 - \frac{2^{-n}}{m},$$

whereas in a no instance,

$$\lambda_{\max}(H') \le 1 - \frac{1}{mn^c}.$$

Now consider $(H')^t$. In a yes instance, as long as $t \ll m2^n$,

$$\lambda_{\max}\big((H')^t\big) \ge \left(1 - \frac{2^{-n}}{m}\right)^t \ge 1 - \frac{t2^{-n}}{m}.$$

In a no instance,

$$\lambda_{\max}\big((H')^t\big) \le \left(1 - \frac{1}{mn^c}\right)^t \le e^{-t/(mn^c)}.$$

Since $(H')^t \ge 0$, the trace is the sum of nonnegative eigenvalues, so in a yes instance

$$\operatorname{tr}\big((H')^t\big) \ge 1 - \frac{t2^{-n}}{m}.$$

In a no instance there are at most 2^n eigenvalues, each at most $e^{-t/(mn^c)}$, and therefore

$$\operatorname{tr}\big((H')^t\big) \le 2^n e^{-t/(mn^c)}.$$

Choosing, for example,

$$t = mn^{c+1},$$

we obtain the separation

$$\operatorname{tr}\big((H')^t\big) \ge 1 - \frac{n^{c+1}}{2^n} \qquad \text{for yes instances,}$$

and

$$\operatorname{tr}\big((H')^t\big) \le \left(\frac{2}{e}\right)^n \qquad \text{for no instances.}$$

Thus it remains to show that $\operatorname{tr}((H')^t)$ is computable in PP.

Write

$$H' = \sum_{i=1}^{m} \frac{1}{m}(\mathbf{1} - h_i) = \sum_{i=1}^{m} \frac{1}{m} h_i',$$

where $h_i' := \mathbf{1} - h_i$. Then

$$\begin{aligned}\operatorname{tr}\big((H')^t\big) &= \frac{1}{m^t} \operatorname{tr}\left(\left(\sum_{i=1}^{m} h_i'\right)^t\right) \\ &= \frac{1}{m^t} \sum_{x_1,\dots,x_t \in \{0,1\}^n} \sum_{i_1,\dots,i_t \in [m]} \left\langle x_1 | h_{i_1}' | x_2 | x_1 | h_{i_1}' | x_2 \right\rangle \left\langle x_2 | h_{i_2}' | x_3 | x_2 | h_{i_2}' | x_3 \right\rangle \cdots \left\langle x_t | h_{i_t}' | x_1 | x_t | h_{i_t}' | x_1 \right\rangle .\end{aligned}$$

Each summand is efficiently computable because every local term acts on only constantly many qubits. Therefore the trace is a sum of an efficiently computable function over exponentially many inputs, and the counting formulation above shows that it can be evaluated in PP. This proves

$$\mathsf{QMA} \subseteq \mathsf{PP}.$$

□

Heat intuition. Repeated powers of $\mathbf{1} - H/m$ suppress every excited eigenspace while preserving the ground space. In the limit,

$$\lim_{t\to\infty} \left(\mathbf{1} - \frac{H}{m}\right)^t$$

is proportional to the projector onto the ground space of H, provided the ground energy is separated from the rest of the spectrum.

2.2 Kitaev's Clock Hamiltonian

The local Hamiltonian problem is not only complete for QMA; it is complete via a remarkably structured construction. The starting point is a quantum verification circuit

$$C = g_T g_{T-1} \cdots g_1$$

acting on a witness register together with ancillas initialized to $|0^m\rangle$. The goal is to construct a Hamiltonian $H(C)$ such that

$$\begin{aligned} &\lambda_{\min}(H(C)) \text{ is small} \quad \text{if there exists } |\psi\rangle \text{ accepted with probability } \geq 1 - 2^{-O(n)}, \\ &\lambda_{\min}(H(C)) \text{ is large} \quad \text{if every } |\psi\rangle \text{ is accepted with probability } \leq 2^{-O(n)}. \end{aligned}$$

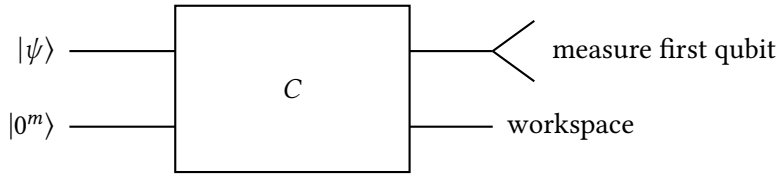

Figure 2.1: A quantum verifier with a witness register and ancillas initialized to $|0^m\rangle$.
A circuit box labeled C takes two inputs: a witness state ket psi and an ancilla register initialized to ket 0 superscript m. The first output qubit is measured, while the remaining output qubits form a workspace register.

2.2.1 History states and the propagation Hamiltonian

Introduce a time register storing an integer $t \in \{0, 1, \ldots, T\}$. With a binary clock, this requires $\lceil \log(T+1) \rceil$ qubits. The ground states we want are the *history states*

$$|\Phi\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle \otimes |\psi_t\rangle ,$$

where

$$|\psi_t\rangle = g_t g_{t-1} \cdots g_1 |\psi_0\rangle \qquad \text{and} \qquad |\psi_0\rangle = |\psi\rangle |0^m\rangle .$$

For each $t \in \{1, \ldots, T\}$, define the propagation term

$$h_t := \frac{1}{2}\Big(|t\rangle\langle t| \otimes \mathbf{1} - |t\rangle\langle t-1| \otimes g_t - |t-1\rangle\langle t| \otimes g_t^\dagger + |t-1\rangle\langle t-1| \otimes \mathbf{1} \Big).$$

Let

$$H_{\mathrm{prop}} := \sum_{t=1}^{T} h_t.$$

It is convenient to conjugate by the unitary

$$W := \sum_{t=0}^{T} |t\rangle\langle t| \otimes g_t g_{t-1} \cdots g_1,$$

with the convention that the $t = 0$ summand is $|0\rangle\langle 0| \otimes \mathbf{1}$. A direct calculation gives

$$W^\dagger h_t W = \frac{1}{2}\big(|t\rangle - |t-1\rangle\big)\big(\langle t| - \langle t-1|\big) \otimes \mathbf{1}.$$

Hence

$$W^\dagger H_{\mathrm{prop}} W = \frac{1}{2}\begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} \otimes \mathbf{1}.$$

This is the Laplacian of a path graph on vertices $0, 1, \dots, T$.

The zero-eigenspace is easy to identify. For every state $|\varphi\rangle$ in the computational register,

$$|\eta_0(\varphi)\rangle := \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle \otimes |\varphi\rangle$$

is a zero-eigenvector of $W^\dagger H_{\mathrm{prop}} W$. Moreover the first nonzero eigenvalue satisfies

$$\lambda_1 \geq \frac{c}{T^2}$$

for a universal constant $c > 0$. Undoing the conjugation, we obtain the following description.

Structure of H_{prop}. The Hamiltonian H_{prop} has ground energy 0, and its ground space is spanned by the history states

$$\frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle \otimes g_t g_{t-1} \cdots g_1 \, |\psi_0\rangle$$

as $|\psi_0\rangle$ varies. If Π_{prop} denotes the projector onto this history-state subspace, then

$$H_{\mathrm{prop}} \geq \frac{c}{T^2}\big(\mathbf{1} - \Pi_{\mathrm{prop}}\big).$$

2.2.2 Input and output checks

The propagation Hamiltonian only checks that the computation evolves consistently. It does not check that the computation begins with the correct ancillas or ends in an accepting state. These conditions are enforced by two additional terms.

To check the output, penalize states whose time register equals T and whose output qubit is $|0\rangle$:

$$H_{\text{out}} := |T\rangle\langle T|_{\text{time}} \otimes |0\rangle\langle 0|_1 .$$

To check the input, penalize states whose time register equals 0 and whose ancilla register is not initialized to $|0^m\rangle$:

$$H_{\text{in}} := \sum_{j=1}^{m} |0\rangle\langle 0|_{\text{time}} \otimes |1\rangle\langle 1|_{\text{ancilla}(j)} .$$

Every local term in $H_{\text{out}} + H_{\text{in}}$ is a projector, and all of these projectors commute. Consequently $H_{\text{out}} + H_{\text{in}}$ has integer spectrum. If Π_{inout} denotes the projector onto the subspace satisfying all input and output checks, then

$$H_{\text{out}} + H_{\text{in}} \geq \mathbf{1} - \Pi_{\text{inout}}.$$

We therefore consider the full Hamiltonian

$$H := H_{\text{prop}} + H_{\text{in}} + H_{\text{out}}.$$

2.2.3 Completeness

Suppose there exists a witness $|\psi\rangle$ such that the verifier accepts $|\psi\rangle\,|0^m\rangle$ with probability at least $1 - \epsilon$. Let

$$|\Phi\rangle := \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle \otimes g_t g_{t-1} \cdots g_1 \,|\psi, 0^m\rangle$$

be the associated history state. By construction,

$$\langle\Phi|\, H_{\text{prop}} \,|\Phi\rangle = 0 \qquad \text{and} \qquad \langle\Phi|\, H_{\text{in}} \,|\Phi\rangle = 0.$$

The only possible contribution comes from the output term, and the rejection probability at the final time is at most ϵ, so

$$\langle\Phi|\, H_{\text{out}} \,|\Phi\rangle \leq \epsilon.$$

Therefore

$$\lambda_{\min}(H) \leq \epsilon.$$

2.2.4 Soundness

Now suppose every witness is accepted with probability at most ϵ. Let θ be the angle between the two subspaces

$$\mathrm{im}(\mathbf{1}_{2^n} \otimes |0^m\rangle\langle 0^m|) \qquad \text{and} \qquad \mathrm{im}\big(C^\dagger(|1\rangle\langle 1| \otimes \mathbf{1})C\big)\,.$$

The maximum acceptance probability is then

$$\cos^2\theta \le \epsilon.$$

In particular, small acceptance probability means that the angle θ is close to $\pi/2$.

Since

$$H_{\mathrm{prop}} \ge \frac{c}{T^2}(\mathbf{1} - \Pi_{\mathrm{prop}}),$$

it is enough to lower bound

$$\lambda_{\min}\Big(\frac{c}{T^2}(\mathbf{1} - \Pi_{\mathrm{prop}}) + (\mathbf{1} - \Pi_{\mathrm{inout}})\Big)\,.$$

If γ denotes the angle between the subspaces Π_{prop} and Π_{inout}, then a standard geometric estimate gives

$$\lambda_{\min}\Big(\frac{c}{T^2}(\mathbf{1} - \Pi_{\mathrm{prop}}) + (\mathbf{1} - \Pi_{\mathrm{inout}})\Big) \ge \frac{2c}{T^2}\sin^2\Big(\frac{\gamma}{2}\Big)\,.$$

Thus the problem reduces to computing γ.

Because the angle is preserved by conjugation, we may work with

$$\gamma = \angle(\Pi_{\mathrm{prop}}, \Pi_{\mathrm{inout}}) = \angle\big(W^\dagger \Pi_{\mathrm{prop}} W,\ W^\dagger \Pi_{\mathrm{inout}} W\big).$$

In the rotated picture,

$$W^\dagger \Pi_{\mathrm{prop}} W$$

projects onto uniform superpositions over the clock, while $W^\dagger \Pi_{\mathrm{inout}} W$ enforces the input condition at time 0 and the output condition at time T. Consequently,

$$\begin{aligned}\cos^2\gamma &= \max_{|\psi\rangle} \big\langle h_\psi \big|\, \Pi_{\mathrm{inout}} \,\big| h_\psi \big\rangle \\ &= \frac{T-1}{T+1} + \frac{1}{T+1}\max_{|\psi\rangle}\Big(\langle\psi|\, C^\dagger(|1\rangle\langle 1| \otimes \mathbf{1})C\,|\psi\rangle + \langle\psi|\,(\mathbf{1}_{2^n} \otimes |0^m\rangle\langle 0^m|)\,|\psi\rangle\Big)\,.\end{aligned}$$

The optimal $|\psi\rangle$ lies halfway between the two relevant projectors, which are separated by angle θ. Hence

the final maximization equals

$$2\cos^2\left(\frac{\theta}{2}\right).$$

Using

$$\cos\theta = \sqrt{\epsilon} = 2\cos^2\left(\frac{\theta}{2}\right) - 1,$$

we obtain

$$\cos^2\gamma = \frac{T-1}{T+1} + \frac{1}{T+1}(1+\sqrt{\epsilon}) = 1 - \frac{1-\sqrt{\epsilon}}{T+1}.$$

Therefore

$$\sin^2\left(\frac{\gamma}{2}\right) = \frac{1-\cos^2\gamma}{2} = \frac{1-\sqrt{\epsilon}}{2(T+1)}.$$

Substituting into the geometric bound yields

$$\lambda_{\min}(H) = \Omega\left(\frac{1-\sqrt{\epsilon}}{T^3}\right).$$

In particular, for sufficiently small ϵ,

$$\text{yes instances satisfy } \lambda_{\min}(H) \le \epsilon, \qquad \text{no instances satisfy } \lambda_{\min}(H) \ge \Omega\left(\frac{1}{T^3}\right).$$

Clock-Hamiltonian conclusion. The Hamiltonian

$$H = H_{\text{prop}} + H_{\text{in}} + H_{\text{out}}$$

implements the reduction from quantum circuit satisfiability to local Hamiltonian using a binary clock, with locality $O(\log n)$. Replacing the binary clock by a unary clock and adding clock-check terms turns the construction into a 5-local Hamiltonian.

2.3 Why Perturbation Theory?

The clock-Hamiltonian construction already proves QMA-hardness of the local Hamiltonian problem, but the resulting Hamiltonians are still rather synthetic. Natural structural restrictions one might want include

- 2-locality,
- geometric locality,

- translational invariance.

A basic question is whether such more physical families can still be QMA-hard. The main tool for reducing general local Hamiltonians to such structured families is perturbation theory.

As a toy problem, suppose one is given a tensor-product interaction

$$h = h_1 \otimes h_2 \otimes \cdots \otimes h_k.$$

The goal is to construct a 2-local Hamiltonian whose low-energy subspace reproduces the effect of h. The rough picture is to introduce a new ancilla particle and couple each factor h_i to that ancilla.

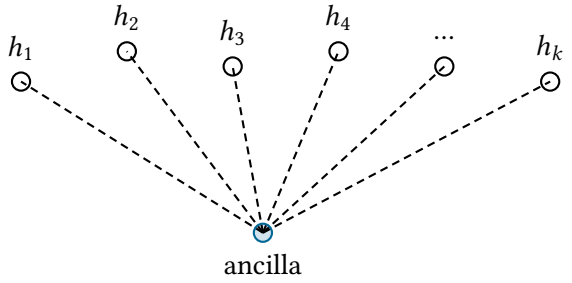

Figure 2.2: A mediator-particle picture for a perturbative gadget.
Several outer systems labeled h1, h2, h3, h4, ellipsis, and hk are all connected to one central ancilla particle. The diagram indicates that each local factor couples separately to the mediator, and the many-body interaction is recovered only in the low-energy effective theory.

2.4 The Projection Lemma

Let the Hilbert space decompose as

$$\mathcal{H} = S \oplus S^{\perp}.$$

Write $\Pi_{S^{\perp}}$ for the projector onto $S^{\perp}$, and let

$$H := \Delta \Pi_{S^{\perp}} + V,$$

where

$$-\mathbf{1} \leq V \leq \mathbf{1}.$$

For this section, write $\lambda(X)$ for the minimum eigenvalue of X.

Lemma 2.5 (Projection lemma)**.** *With the notation above,*

$$\lambda(V|_S) - \frac{1}{\Delta - 2} \leq \lambda(H) \leq \lambda(V|_S).$$

Proof. The upper bound is immediate: if $|v\rangle \in S$, then $\Pi_{S^\perp}|v\rangle = 0$, so

$$\langle v| H\!\!\downarrow |v\rangle = \langle v| V |v\rangle .$$

Minimizing over S gives

$$\lambda(H\!\!\downarrow) \leq \lambda(V|_S).$$

For the lower bound, let $|v\rangle$ be any unit vector, and decompose it as

$$|v\rangle = \alpha_1 |v_1\rangle + \alpha_2 |v_2\rangle ,$$

where $|v_1\rangle \in S$, $|v_2\rangle \in S^\perp$ are unit vectors, and $\alpha_1, \alpha_2 \geq 0$ with $\alpha_1^2 + \alpha_2^2 = 1$. Then

$$\begin{aligned}
\langle v| H\!\!\downarrow |v\rangle &= \langle v| V |v\rangle + \Delta\alpha_2^2 \\
&= (1-\alpha_2^2)\langle v_1| V |v_1\rangle + 2\alpha_1\alpha_2 \operatorname{Re}\langle v_1| V |v_2\rangle + \alpha_2^2 \langle v_2| V |v_2\rangle + \Delta\alpha_2^2 \\
&\geq \langle v_1| V |v_1\rangle - 2\alpha_2 - \alpha_2^2 + \Delta\alpha_2^2 \\
&= \langle v_1| V |v_1\rangle + (\Delta-1)\alpha_2^2 - 2\alpha_2 \\
&\geq \lambda(V|_S) + (\Delta-2)\alpha_2^2 - 2\alpha_2.
\end{aligned}$$

The final quadratic is minimized at $\alpha_2 = 1/(\Delta-2)$, giving

$$\langle v| H\!\!\downarrow |v\rangle \geq \lambda(V|_S) - \frac{1}{\Delta-2}.$$

Since $|v\rangle$ was arbitrary, the same bound holds for $\lambda(H\!\!\downarrow)$. □

Corollary 2.6. *If $\|V\| \leq K$, then after rescaling the same argument yields the error estimate*

$$\lambda(V|_S) - \frac{K^2}{\Delta - 2K} \leq \lambda(H\!\!\downarrow) \leq \lambda(V|_S).$$

The projection lemma controls the ground energy of a perturbed Hamiltonian. To understand perturbative gadgets, however, one needs more than the ground energy alone: the entire low-energy spectrum should be well approximated by an effective Hamiltonian acting on the low-energy subspace S.

2.5 The Effective Hamiltonian

Assume from now on that

$$\|V\| \ll \frac{\Delta}{2},$$

so the perturbation is small relative to the penalty scale. Write the block decomposition of V with respect to $S \oplus S^{\perp}$ as

$$V = \begin{pmatrix} A & B^{\dagger} \\ B & C \end{pmatrix}, \qquad A = \Pi_S V \Pi_S.$$

Then

$$H = \begin{pmatrix} A & B^{\dagger} \\ B & \Delta \mathbf{1} + C \end{pmatrix}.$$

Consider an eigenvector

$$|\Psi\rangle = |x\rangle + |y\rangle \in S \oplus S^{\perp}$$

with eigenvalue $\lambda < \Delta/2$. The eigenvalue equation becomes

$$\begin{aligned} A\,|x\rangle + B^{\dagger}\,|y\rangle &= \lambda\,|x\rangle\,, \\ B\,|x\rangle + (\Delta \mathbf{1} + C)\,|y\rangle &= \lambda\,|y\rangle\,. \end{aligned}$$

Rearranging the second equation gives

$$\big((\Delta - \lambda)\mathbf{1} + C\big)\,|y\rangle = -B\,|x\rangle\,.$$

Since λ is small compared with Δ, this operator is invertible, and therefore

$$|y\rangle = -\big((\Delta - \lambda)\mathbf{1} + C\big)^{-1} B\,|x\rangle\,.$$

Substituting back into the first equation yields the *effective Hamiltonian*

$$H_{\mathrm{eff}}(\lambda) := A - B^{\dagger}\big((\Delta - \lambda)\mathbf{1} + C\big)^{-1} B.$$

Thus $|\Psi\rangle = |x\rangle + |y\rangle$ is a low-energy eigenvector of H with eigenvalue λ precisely when

$$H_{\mathrm{eff}}(\lambda)\,|x\rangle = \lambda\,|x\rangle$$

and $|y\rangle$ is recovered from the resolvent formula above.

By construction,

$$\||y\rangle\| = O\!\left(\frac{\|V\|}{\Delta}\right)\||x\rangle\|,$$

so low-energy eigenvectors of H lie mostly in the subspace S. They do not lie *entirely* in S, however; the perturbation leaves a small tail in $S^{\perp}$, and this tail is exactly what produces the effective interaction.

To extract that interaction, expand the resolvent as a power series:

$$\big((\Delta-\lambda)\mathbf{1}+C\big)^{-1}=\frac{1}{\Delta}\mathbf{1}-\frac{C-\lambda\mathbf{1}}{\Delta^2}+\frac{(C-\lambda\mathbf{1})^2}{\Delta^3}-\cdots=\sum_{k\geq 0}(-1)^k\frac{(C-\lambda\mathbf{1})^k}{\Delta^{k+1}}.$$

Substituting into $H_{\text{eff}}(\lambda)$ gives

$$H_{\text{eff}}(\lambda)=A-\frac{1}{\Delta}B^\dagger B+\frac{1}{\Delta^2}B^\dagger(C-\lambda\mathbf{1})B+O\left(\frac{K^3}{\Delta^3}\right),$$

where K is a bound on $\|V\|$. The first three terms are the second-order effective Hamiltonian,

$$H_{\text{eff}}^{(2)}(\lambda):=A-\frac{1}{\Delta}B^\dagger B+\frac{1}{\Delta^2}B^\dagger(C-\lambda\mathbf{1})B.$$

Low-energy characterization. If $|\Psi\rangle$ is an eigenvector of $\not{H}$ with eigenvalue $\lambda\ll\Delta/2$, then $\Pi_S|\Psi\rangle$ is an eigenvector of $H_{\text{eff}}(\lambda)$. In the perturbative regime, this effective operator is well approximated by $H_{\text{eff}}^{(2)}(\lambda)$.

2.6 A Basic Perturbative Gadget

As an application, consider a target k-local term of the form

$$h_2h_1+h_1^\dagger h_2^\dagger,$$

where both h_1 and h_2 are $(k/2)$-local operators. Introduce one ancilla qubit and set

$$S=\text{span}\{|0\rangle\},\qquad S^\perp=\text{span}\{|1\rangle\}.$$

Define

$$V:=-|1\rangle\langle 0|\otimes h_1-|0\rangle\langle 1|\otimes h_1^\dagger-|1\rangle\langle 0|\otimes h_2^\dagger-|0\rangle\langle 1|\otimes h_2.$$

In block form,

$$V=\begin{pmatrix}0 & -(h_1+h_2^\dagger)\\ -(h_1^\dagger+h_2) & 0\end{pmatrix}.$$

Now consider

$$\not{H}:=\Delta|1\rangle\langle 1|+V.$$

The effective Hamiltonian on the $|0\rangle$-subspace is

$$\begin{aligned} H_{\text{eff}}(\lambda) &= -\left(\frac{1}{\Delta} + \frac{\lambda}{\Delta^2}\right) |0\rangle\langle 0| \otimes (h_1^\dagger + h_2)(h_1 + h_2^\dagger) + O\left(\frac{1}{\Delta^3}\right) \\ &= -\left(\frac{1}{\Delta} + \frac{\lambda}{\Delta^2}\right) |0\rangle\langle 0| \otimes \left(h_1^\dagger h_1 + h_2 h_2^\dagger + h_2 h_1 + h_1^\dagger h_2^\dagger\right) + O\left(\frac{1}{\Delta^3}\right). \end{aligned}$$

The cross terms reproduce the desired k-local interaction

$$h_2 h_1 + h_1^\dagger h_2^\dagger,$$

while the remaining terms are only $(k/2)$-local corrections. Thus the gadget captures the target interaction in the low-energy sector, at the price of introducing an ancilla and a large penalty scale Δ.

More generally, one can introduce an ancilla qudit of dimension d and build a similar clock-like gadget. Higher-order gadgets require a stronger choice of Δ and an analysis of dth-order terms in the expansion, but the same perturbative mechanism can be used to reduce the locality of d-local interactions.

2.7 Restricted Families and the One-Dimensional Setting

Perturbative gadgets can be combined with geometric and locality-preserving constructions to prove hardness for families of Hamiltonians that look much closer to physical models. Examples include the following:

(1) one-dimensional Hamiltonians on qudits of local dimension 12,

(2) Hamiltonians whose local terms are weighted $X \otimes X$ or $Z \otimes Z$ interactions on arbitrary interaction graphs,

(3) two-dimensional Hamiltonians on qudits of local dimension 6.

The broader goal is to understand when local Hamiltonian systems are hard and when they become algorithmically tractable. A natural next test case is the one-dimensional setting.

Consider a chain of n qudits of local dimension d, and let

$$H = \sum_{i=1}^{n-1} h_i,$$

where each h_i acts only on sites i and $i+1$. By grouping together constantly many neighboring spins into larger sites, more general finite-range one-dimensional interactions can often be reduced to this nearest-neighbor form.

Any efficient algorithm depending only on n and d would solve a QMA-complete problem. Thus one-dimensional algorithms must exploit an additional structural parameter. In the setting discussed here, that parameter is the spectral gap.

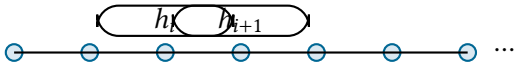

Figure 2.3: A one-dimensional nearest-neighbor Hamiltonian.
A line of sites is drawn from left to right. Two overlapping nearest-neighbor interaction terms, labeled h sub i and h sub i plus 1, each act on adjacent pairs of sites.

Roadmap for the one-dimensional algorithmic theory. The discussion proceeds in three steps:

(1) describe tensor networks and matrix product states as compact classical encodings of quantum states,

(2) explain why one-dimensional area laws bound the entanglement of gapped ground states,

(3) use these bounded-entanglement descriptions inside a viable-space construction.

Example 2.7 (Classical sanity check)**.** Every classical one-dimensional constraint satisfaction problem admits a polynomial-time algorithm for computing its minimum energy. Any quantum algorithm for one-dimensional Hamiltonians should be viewed as a noncommutative analogue of this dynamic-programming phenomenon.

2.8 Tensor Networks and Matrix Product States

2.8.1 Tensors and their interpretations

A tensor of rank k may be viewed as a function

$$M : [d_1] \times [d_2] \times \cdots \times [d_k] \to \mathbb{C}.$$

When all edge dimensions are equal to d, this is simply an array of d^k complex numbers. Diagrammatically, one draws a tensor as a vertex with k incident edges, each edge carrying one index.

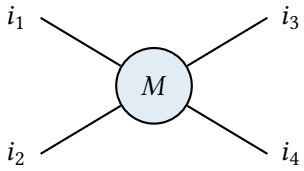

Figure 2.4: A rank-four tensor with four labeled indices.
A single circular node labeled M has four incident edges labeled i1, i2, i3, and i4. Each edge represents one tensor index.

Fix a Hilbert space $\mathcal{H} = \mathbb{C}^d$ with basis $|0\rangle, |1\rangle, \ldots, |d-1\rangle$. The same tensor can be interpreted in several equivalent ways.

First, it can be viewed as a vector in $\mathcal{H}^{\otimes k}$. For a rank-four tensor,

$$|M\rangle = \sum_{i_1,i_2,i_3,i_4} M_{i_1,i_2,i_3,i_4} |i_1\rangle |i_2\rangle |i_3\rangle |i_4\rangle .$$

Second, after partitioning the indices into input and output groups, it can be viewed as a linear map. For instance, if (i_1, i_2) are inputs and (i_3, i_4) are outputs, then

$$M(|i_1\rangle |i_2\rangle) = \sum_{i_3,i_4} M_{i_1,i_2,i_3,i_4} |i_3\rangle |i_4\rangle .$$

A different partition gives a different linear map. For example,

$$M(|i_3\rangle) = \sum_{i_1,i_2,i_4} M_{i_1,i_2,i_3,i_4} |i_1\rangle |i_2\rangle |i_4\rangle ,$$

so the same tensor can also be regarded as an operator $\mathcal{H} \to \mathcal{H}^{\otimes 3}$.

A pure n-qudit state is itself a tensor:

$$|\psi\rangle = \sum_{x_1,\ldots,x_n} \psi_{x_1,\ldots,x_n} |x_1\rangle |x_2\rangle \cdots |x_n\rangle .$$

Tensor-network notation is simply a way to factor this large tensor into smaller ones.

2.8.2 Join, tensor product, and contraction

If two tensors M and N have disjoint sets of indices, their join is just the tensor product $M \otimes N$.

If two tensors share an index, one may contract that index. In coordinates, if j is a shared label, then contraction replaces the pair of tensors by a new tensor whose coefficients are obtained by summing over j:

$$K_{\alpha,\beta} = \sum_j M_{\alpha,j} N_{j,\beta}.$$

Matrix multiplication is the most familiar example of contraction. A closed loop is another basic example: contracting the two indices of a square matrix M yields $\mathrm{tr}(M)$.

Contraction order is often flexible. Different sequences of pairwise contractions can represent the same final tensor, and choosing the contraction order well is crucial for efficient computation.

It is also useful to allow different edges to carry different dimensions. This lets one describe Hilbert spaces of the form

$$\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_n}$$

and to distinguish physical indices from internal bond indices. For example, if $i_1, i_2 \in [2]$ are physical indices

and $j \in [5]$ is a bond index, then contracting two tensors over j produces a two-site state with amplitudes

$$\psi_{i_1,i_2} = \sum_{j=1}^{5} M^{[1]}_{i_1,j} M^{[2]}_{i_2,j}.$$

2.8.3 Matrix product states

A state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ is a *matrix product state* of bond dimension D if there exist tensors

$$A^{[1]}, A^{[2]}, \ldots, A^{[n]}$$

such that

$$\psi_{i_1,\ldots,i_n} = A^{[1]}_{i_1} A^{[2]}_{i_2} \cdots A^{[n]}_{i_n},$$

where $A^{[1]}_{i_1}$ is a $1 \times D$ row vector, $A^{[n]}_{i_n}$ is a $D \times 1$ column vector, and the intermediate $A^{[t]}_{i_t}$ are $D \times D$ matrices.

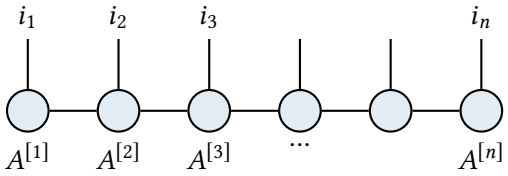

Figure 2.5: A matrix product state. The vertical edges are physical indices, and the horizontal edges are bond indices.

A chain of tensors A superscript 1 through A superscript n is connected left to right by horizontal bond edges. Each tensor also has one upward physical edge labeled by a site index.

Several basic examples are worth recording.

Exercise 2.8. Show that every product state

$$|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$$

is a bond-dimension-1 matrix product state.

Exercise 2.9. Show that the GHZ state

$$\frac{|0^n\rangle + |1^n\rangle}{\sqrt{2}}$$

is a bond-dimension-2 matrix product state.

Exercise 2.10. Show that the W-type state

$$\frac{|10\cdots0\rangle + |010\cdots0\rangle + \cdots + |0\cdots01\rangle}{\sqrt{n}}$$

is a bond-dimension-2 matrix product state.

Exercise 2.11. Show that every n-qudit state has an exact matrix product state representation of bond dimension at most $d^{n/2}$.

2.8.4 Cost of basic tensor-network computations

The cost of joining tensors is the cost of writing out the resulting tensor, which is proportional to the product of the dimensions of all legs of the output tensor.

For contractions, it is often best to reinterpret the tensors as linear maps and use matrix multiplication. If

$$A : \mathbb{C}^{d_1} \to \mathbb{C}^{d_2}, \qquad B : \mathbb{C}^{d_2} \to \mathbb{C}^{d_3},$$

then the contracted tensor BA can be computed in time

$$O(d_1 d_2 d_3).$$

With fast matrix multiplication, this may be improved to

$$O\big((d_1 d_2 d_3)^{\omega/3}\big).$$

These estimates already imply efficient algorithms for basic MPS tasks.

Lemma 2.12. *Given an n-site matrix product state of bond dimension D, the norm*

$$\langle\psi|\psi|\psi|\psi\rangle$$

can be computed in time $O(ndD^3)$. More generally, if O is an $O(1)$-local observable, then

$$\langle\psi|\, O\,|\psi\rangle$$

can be computed in time polynomial in n, d, and D.

Idea of the computation. Contract the network from left to right. At every stage one keeps only a bounded number of open legs, each of dimension at most d or D. This is the standard "bubble over the MPS" contraction picture. Since there are n local contractions, the total runtime is polynomial in n, d, and D, with the norm evaluation taking $O(ndD^3)$. □

2.9 Schmidt Rank and Entanglement Across a Cut

Let

$$\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B, \qquad \dim(\mathcal{H}_A) = d_A, \qquad \dim(\mathcal{H}_B) = d_B.$$

Every state $|\psi\rangle \in \mathcal{H}$ admits a Schmidt decomposition

$$|\psi\rangle = \sum_{i=1}^{\min(d_A,d_B)} c_i \, |v_i\rangle_A \, |u_i\rangle_B \,,$$

where $c_i \geq 0$, and $\{|v_i\rangle\}$, $\{|u_i\rangle\}$ are orthonormal families. This is the vectorized form of the singular value decomposition.

Definition 2.13. *The* Schmidt rank *of $|\psi\rangle$ across the bipartition $A|B$ is the number of nonzero Schmidt coefficients c_i.*

Two standard examples are:

$$|\mathrm{EPR}_d\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle \, |i\rangle \,,$$

which has maximal Schmidt rank d, and any product state

$$|\psi\rangle_A \otimes |\phi\rangle_B \,,$$

which has Schmidt rank 1.

The reduced density operators are immediately determined by the Schmidt decomposition:

$$\mathrm{tr}_A(|\psi\rangle\langle\psi|) = \sum_i c_i^2 \, |u_i\rangle\langle u_i| \,,$$
$$\mathrm{tr}_B(|\psi\rangle\langle\psi|) = \sum_i c_i^2 \, |v_i\rangle\langle v_i| \,.$$

Hence the entanglement entropy across the cut is

$$S(\psi_A) = S(\psi_B) = -\sum_i c_i^2 \log\left(c_i^2\right).$$

For one-dimensional systems, controlling the Schmidt rank and entropy across every cut is precisely what makes matrix product states effective.

2.10 Gapped One-Dimensional Hamiltonians and Area Laws

Theorem 2.14 (Landau, Vazirani, and Vidick, 2013)**.** *Let H be a nearest-neighbor Hamiltonian on a line of n qudits of local dimension d. Assume for simplicity that*

$$H = \sum_i h_i,$$

each h_i is a projector, the ground energy satisfies $\lambda_{\min}(H) = 0$, the ground state $|\Gamma\rangle$ is unique, and the first excited energy satisfies

$$\lambda_1(H) = \epsilon > 0.$$

Then there is an algorithm running in time

$$n^{c(d,\epsilon)}\mathrm{poly}\left(\frac{n}{\eta}\right)$$

that outputs a matrix product state $|\phi\rangle$ of bond dimension $n^{c(d,\epsilon)}$ such that

$$|\langle\phi|\Gamma|\phi|\Gamma\rangle|^2 \geq 1 - \eta.$$

Here

$$c(d,\epsilon) = 2^{O(\log^3 d/\epsilon)}.$$

This theorem has two immediate interpretations. First, it produces a classical description of the ground state itself, not just its energy. Second, it shows that gapped one-dimensional ground states have controlled entanglement: a state that admits a polynomial-bond-dimension MPS description cannot be highly entangled across every cut.

That entanglement control is supplied by an area law.

Theorem 2.15 (One-dimensional area law)**.** *Let H be a local Hamiltonian with interaction graph $G = (V, E)$, and let $|\Gamma\rangle$ be its ground state. For any subset $S \subseteq V$,*

$$S(\Gamma_S) = O_{d,\epsilon}\big(|E(S,\bar{S})|\big),$$

where ϵ denotes the spectral gap. In one dimension it suffices to consider intervals

$$S = \{1, \ldots, i\},$$

for which

$$S(\Gamma_S) \leq O\left(\frac{\log^3 d}{\epsilon}\right).$$

The original one-dimensional bound due to Hastings was exponentially worse in $\log d/\epsilon$; later work improved it to the form written above.

Interpretation of the area law. In one dimension, the boundary of an interval consists of only constantly many edges. The area law therefore bounds the entanglement entropy across every cut by

a function of d and ϵ, independent of n. This is the structural reason that matrix product states can efficiently approximate gapped ground states.

2.10.1 Linear algebra over a tractable subspace

Brute-force diagonalization solves the local Hamiltonian problem over the full Hilbert space $(\mathbb{C}^d)^{\otimes n}$. The key algorithmic move is to replace this enormous space by a carefully chosen span of matrix product states.

Lemma 2.16. *Let $|\phi_1\rangle, \ldots, |\phi_s\rangle$ be matrix product states of bond dimension D, and let*

$$S = \operatorname{span}\{|\phi_1\rangle, \ldots, |\phi_s\rangle\}.$$

Then one can compute the state $|\phi\rangle \in S$ minimizing $\langle\phi| H |\phi\rangle$ in time

$$O\left(s^3 + s^2 n d D^3\right).$$

Moreover, the minimizing vector can itself be represented as a matrix product state of bond dimension Ds.

Proof. Compute the Gram matrix G and energy matrix E defined by

$$G_{ij} = \left\langle \phi_i | \phi_j \middle| \phi_i | \phi_j \right\rangle,$$
$$E_{ij} = \langle\phi_i| H \left|\phi_j\right\rangle.$$

Each entry can be evaluated by contracting matrix product states, so the total preprocessing time is $O(s^2 n d D^3)$.

Every vector $|\phi\rangle \in S$ may be written as

$$|\phi\rangle = \sum_{i=1}^{s} \alpha_i |\phi_i\rangle.$$

Since the basis need not be orthonormal,

$$\||\phi\rangle\|^2 = \sum_{i,j} \overline{\alpha_i} \alpha_j G_{ij} = \langle\alpha| G |\alpha\rangle,$$

and similarly

$$\langle\phi| H |\phi\rangle = \sum_{i,j} \overline{\alpha_i} \alpha_j E_{ij} = \langle\alpha| E |\alpha\rangle.$$

Thus

$$\min_{\substack{|\phi\rangle \in S \\ \|\phi\|=1}} \langle\phi| H |\phi\rangle = \min_{\alpha \neq 0} \frac{\langle\alpha| E |\alpha\rangle}{\langle\alpha| G |\alpha\rangle}.$$

After discarding linear dependencies, G is positive definite. Write

$$G = B^\dagger B$$

with B invertible. The minimization reduces to the smallest eigenvalue of

$$(B^{-1})^\dagger E B^{-1},$$

which can be found in time $O(s^3)$.

Finally, once the minimizing coefficients α_i are known, the state

$$\sum_{i=1}^{s} \alpha_i \left|\phi_i\right\rangle$$

is converted into a single MPS by incorporating the index i into each bond. This multiplies the bond dimension by a factor of s, giving bond dimension Ds. □

2.11 Viable Spaces and the High-Level Algorithm

Definition 2.17. *A subspace $V \subseteq (\mathbb{C}^d)^{\otimes n}$ is* δ-viable *if there exists a state $|\phi\rangle \in V$ such that*

$$|\langle\phi|\Gamma|\phi|\Gamma\rangle| \geq 1 - \delta,$$

where $|\Gamma\rangle$ is the ground state.

For a contiguous interval

$$R = \{a, a+1, \ldots, b\} \subseteq [n],$$

a subspace $V \subseteq (\mathbb{C}^d)^{\otimes |R|}$ is δ-viable for R if there exists a global state

$$|\phi\rangle \in (\mathbb{C}^d)^{\otimes(a-1)} \otimes V \otimes (\mathbb{C}^d)^{\otimes(n-b)}$$

with

$$|\langle\phi|\Gamma|\phi|\Gamma\rangle| \geq 1 - \delta.$$

The algorithm seeks viable spaces that satisfy two competing requirements: they should contain a good approximation to the ground state, and they should admit a basis of efficiently contractible matrix product states.

Two elementary facts initiate the induction.

Fact 2.18. *For a single site $R = \{i\}$, the full space $\mathbb{C}^d$ is a 0-viable space, with basis $|0\rangle, |1\rangle, \ldots, |d-1\rangle$.*

Fact 2.19. *Let R and R' be adjacent intervals, and let V and V' be δ-viable spaces for R and R', respectively. If V is spanned by m matrix product states of bond dimension D, and V' is spanned by m' such states, then $V \otimes V'$ is a 2δ-viable space for $R \cup R'$, spanned by mm' matrix product states of bond dimension D.*

Naively iterating this tensor-product growth leads to exponential blowup in the dimension of the viable spaces. Moreover, the viability error deteriorates as intervals are merged. The rest of the algorithm is devoted to trimming the size and bond dimension while periodically restoring accuracy.

Two organizational strategies are relevant here. The 2013 Landau-Vazirani-Vidick approach grows the interval one site at a time, using the area law across the current cut. A later construction builds viable spaces using a binary tree. The one-site growth picture is the one emphasized here.

2.11.1 The four-step induction

Write

$$s = \dim(S_{i-1}), \qquad B = \text{bond dimension of each basis MPS}, \qquad \delta = \frac{c_\delta}{n},$$

for the current viable space S_{i-1} on the interval $\{1, \dots, i-1\}$.

The induction from $\{1, \dots, i-1\}$ to $\{1, \dots, i\}$ consists of four steps.

Step 1. Extension. Tensor each basis matrix product state with a basis state $|j\rangle$ on the new site. This multiplies the dimension of the viable space by d, while leaving the bond dimension essentially unchanged.

Step 2. Size trimming. Build an $\frac{c_\delta}{n}$-net over the possible boundary tensors at the new cut. Because the bond dimension is polynomially bounded, this net has polynomial size. The area law implies that one may keep only polynomially many representative states while preserving a constant level of viability.

Step 3. Bond trimming. Truncate small Schmidt coefficients across the cut in order to reduce the bond dimension back to a polynomial bound. This is the standard MPS compression step.

Step 4. Error reduction. Apply an approximate ground-space projector (AGSP). This raises the bond dimension and the size of the spanning family in a controlled way, but it restores the viability parameter to c_δ/n, so the next induction step can begin.

Tracking the parameters s, B, and δ through these steps shows that extension increases the size, trimming reduces it, and the AGSP repairs the approximation error.

Algorithmic philosophy. The viable-space framework alternates between three operations:

(1) expand the search space enough to capture the ground state on a larger interval,

(2) compress the description using the area law and Schmidt truncation,

(3) restore accuracy using an AGSP.

This is the one-dimensional analogue of balancing expressivity against computational manageability.

2.12 Beginning of the Area-Law Proof

To prove the one-dimensional area law across a cut

$$A = \{1, \dots, i\}, \qquad B = \{i+1, \dots, n\},$$

one begins by writing the ground state in Schmidt form:

$$|\Gamma\rangle = \sum_j c_j \left|\phi_A^j\right\rangle \left|\phi_B^j\right\rangle .$$

The continuation of the proof analyzes these Schmidt coefficients in order to bound

$$S(\Gamma_A) = -\sum_j c_j^2 \log\left(c_j^2\right)$$

by $O(\log^3 d/\epsilon)$.

Bibliography