Frontiers of Quantum Complexity *Instructors: Anand Natarajan, Anurag Anshu*

Lecture 20: Hamiltonian Learning

November 13, 2025 *Scribe: Deepu Pradeep*

1 Hamiltonian Learning

Hamiltonian learning can be viewed as a modern formulation of one of the most basic tasks in the physical sciences: learning the laws of nature from experiments. In the context of quantum many-body physics and condensed matter, those "laws" are encoded in an underlying Hamiltonian H, and our goal is to infer or approximate H from experimentally accessible data.

Traditionally, physics proceeds by observing a system, performing controlled experiments, and proposing models that fit the data. Hamiltonian learning formalizes this process in a precise, information-theoretic and algorithmic way, asking:

- what kind of data do we need to collect,
- how many samples are required,
- and how efficiently can we reconstruct or approximate the Hamiltonian?

There are many natural ways to state the Hamiltonian learning problem, depending on what kind of access we assume to the quantum system. In this lecture, we focus primarily on:

- **Hamiltonian learning from Gibbs states:** we are given access to thermal states

$$\rho_\beta(H) = \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}$$

  at some inverse temperature β.

- **Hamiltonian learning from time evolution:** we are given access to the unitary dynamics e^{-itH} for various times t.

Both settings are physically natural, but they come with different advantages:

- **Gibbs states** are often easier to realize experimentally, since many-body systems tend to thermalize quickly, so preparing an approximate thermal state can be relatively straightforward.
- **Time evolution** is, in a sense, more "local" information about the Hamiltonian than the full Gibbs state, and can make certain learning tasks conceptually cleaner or more efficient.

1.1 Lecture Outline

In this lecture, we will discuss Hamiltonian learning through the following structure:

1. **Problem statement:** precise formulations of the Hamiltonian learning task in the above settings.
2. **Motivation:** why Hamiltonian learning is interesting from both a physics and complexity-theoretic perspective.
3. **Known results (and room for improvement):** what is currently known about sample complexity, robustness, and limitations.
4. **Learning methods:** algorithmic approaches to Hamiltonian learning, focusing on two guiding questions:
   (a) *Is a given class of Hamiltonians information-theoretically learnable from the available data?*
   (b) *If it is learnable, can we perform the learning task efficiently?*

Our goal is to understand not only whether Hamiltonians can be learned from realistic experimental data, but also how the structure of the Hamiltonian (e.g., locality, sparsity) and the type of access (Gibbs states vs. time evolution) influence the computational complexity of the learning problem.

2 Problem Statement: Learning Local Hamiltonians from Gibbs States

We consider an n-qubit quantum many-body system arranged on a lattice, or more generally on a low-degree interaction graph. Think, for example, of a 2D square lattice with spins on the vertices and local interactions on nearby sites.

2.1 Model Class

We fix a family of Hermitian basis operators

$$\{E_a\}_{a=1}^{m}, \qquad m = O(n),$$

where n is the number of qubits. For this lecture, it is convenient to take the E_a's to be *local Pauli operators* (Pauli strings) supported on a constant number of qubits:

$$\mathrm{supp}(E_a) \subseteq [n], \qquad |\mathrm{supp}(E_a)| = O(1).$$

The unknown Hamiltonian H is assumed to be a linear combination of these basis terms:

$$H = \sum_{a=1}^{m} \lambda_a E_a,$$

with uniformly bounded coefficients

$$\lambda_a \in [-1, 1] \quad \text{for all } a.$$

The learner knows the geometry (lattice / low-degree graph) and the basis $\{E_a\}$, but not the coefficients $\{\lambda_a\}$.

The restriction $m = O(n)$ enforces a low-degree interaction structure and rules out highly nonlocal or dense models such as the SYK model or Hamiltonians defined on a star graph. Currently, we do not know how to learn such models efficiently from Gibbs-state data.

2.2 Access Model

We assume black-box access to thermal states of the unknown Hamiltonian at a *known* inverse temperature β:

$$\rho_\beta(H) = \frac{e^{-\beta H}}{\mathrm{Tr}\big(e^{-\beta H}\big)}.$$

The learner can obtain

$$\rho_\beta(H)^{\otimes s},$$

i.e., s independent and identically distributed copies of the Gibbs state, and may perform quantum measurements on these copies together with classical post-processing of the outcomes.

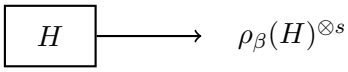

Figure 1: Black-box access: from the unknown Hamiltonian H, we can sample s independent copies of the Gibbs state $\rho_\beta(H)$ at known inverse temperature β.

2.3 Learning Task

Given an accuracy parameter $\varepsilon > 0$, the goal is to output an estimate

$$\lambda' = (\lambda'_1, \ldots, \lambda'_m)$$

such that, with high probability,

$$\forall a \in [m] \quad |\lambda'_a - \lambda_a| \le \varepsilon.$$

Equivalently, we wish to learn the Hamiltonian

$$H' = \sum_{a=1}^{m} \lambda'_a E_a$$

so that every local coupling is accurate to within additive error ε.

2.4 Quantities of Interest

This setup naturally leads to several complexity measures:

- **Sample complexity.** How many copies s of the Gibbs state $\rho_\beta(H)$ are information-theoretically necessary and sufficient to achieve accuracy ε?
- **Time (computational) complexity.** Given s samples, how much computation is required to output λ'? We are interested in algorithms whose running time is polynomial in n, m, and $1/\varepsilon$.
- **Locality of measurements.** From an experimental standpoint, it is highly desirable to use only *local* measurements, because joint measurements on many qubits are difficult in current platforms. Ideally, our measurement operators should act on at most a constant number of qubits and stay close to the locality structure of the Hamiltonian, i.e. to the supports $\mathrm{supp}(E_a)$ and their sizes $|\mathrm{supp}(E_a)|$.

Thus, the central question is: for local Hamiltonians on lattices or other low-degree graphs, can we learn the coefficients $\{\lambda_a\}$ efficiently—both in terms of the number of thermal-state samples and the computational cost—using only low-weight, experimentally feasible measurements?

3 Motivation

3.1 Effective Hamiltonians in the Lab

In real many-body systems, the "true" microscopic Hamiltonian is often not known exactly. Even if we believe that the dominant term is something simple—say a Heisenberg model on a lattice—there are typically additional, more complicated interactions:

- weak long-range couplings,
- cross-talk between qubits,
- or other device-specific imperfections.

These extra terms matter in practice: they change the dynamics, they affect thermal properties, and they show up as *noise* when we try to use the device as a quantum simulator.

Hamiltonian learning provides a systematic way to infer this *effective* Hamiltonian directly from experimental data. This is precisely the problem faced by experimental groups building quantum devices: given measurement data from their hardware, can they identify and compensate for the additional interactions that are not present in the ideal target model?

3.2 Learning H from Eigenstates and ETH

One concrete line of work is learning a Hamiltonian from (approximate) eigenstates. A question posed around 2017–2018 by Grover and Goodall (not the Grover of Grover's algorithm) asks:

> If we are given access to a *typical* eigenstate $|E\rangle$ of an unknown local Hamiltonian, can we reconstruct H?

The physical motivation comes from the **Eigenstate Thermalization Hypothesis (ETH)**. ETH states that for a typical highly excited eigenstate $|E\rangle$ of a chaotic Hamiltonian, the reduced density matrix on any small region R is approximately thermal:

$$\rho_R := \mathrm{Tr}_{\overline{R}}\left(|E\rangle\langle E|\right) \approx \frac{e^{-\beta H_R}}{\mathrm{Tr}\left(e^{-\beta H_R}\right)}.$$

Here H_R is an effective Hamiltonian acting only on region R, and $\overline{R}$ denotes the complement of R.

A heuristic picture is:

- The global universe might be in a single energy eigenstate of some huge Hamiltonian.
- Local observers only have access to reduced states on small regions R.
- Those reduced states look like Gibbs states of an *effective* local Hamiltonian on R.

Hamiltonian learning, in this context, asks whether we can reconstruct that effective H from local data coming from such eigenstates.

A related experimental motivation comes from work by Kokail *et al.*, who studied learning a Hamiltonian from (approximate) ground states and verified thermal-like properties in actual devices. This provides another concrete example where learning H is directly tied to understanding the physics of an engineered quantum system.

3.3 Computer-Science Motivation

There is also a strong computer-science analogue. If we drop the word "quantum," the same task becomes the problem of learning a *classical* Hamiltonian from samples of its Gibbs distribution. This is exactly the setting of:

- **Graphical models / Boltzmann machines**, where the Hamiltonian encodes pairwise or higher-order interactions on a graph.

By around 2015–2017, the classical theory of this problem was quite well-developed. Results by Bresler (2015), and later by Klivans and Meka (2017) and others, showed that one can:

- learn the parameters of a classical Hamiltonian from Gibbs samples,
- even when the interaction graph is *unknown*,
- as long as the underlying graph is sparse (bounded degree).

This is sometimes called **structured learning**. In this course we will not focus on recovering the unknown graph itself, but it serves as an important classical benchmark for what should be possible in the quantum setting.

3.4 Anti-Pseudorandomness: Noisy Circuits as Gibbs States

A final, more speculative motivation comes from the theory of pseudorandom quantum states. You may have seen constructions showing that low-depth quantum circuits (e.g., depth $O(\log n)$) can generate families of states that are *pseudorandom* (PRS): no efficient algorithm can distinguish them from Haar-random states, and in particular they are not learnable in any reasonable sense.

Now imagine taking such a circuit and adding realistic noise. The output of the noisy device can often be modeled as a *Gibbs state* of some effective Hamiltonian: the noise introduces enough entropy that the state looks thermal.

From this perspective, Hamiltonian learning can be seen as a kind of *anti-pseudorandomness* principle:

- pure low-depth circuits might generate pseudorandom states that are information-theoretically unlearnable,
- but once noise is present and the system looks like a Gibbs state, a successful Hamiltonian-learning algorithm would recover structure in the state and show that these noisy families are, in fact, learnable.

This interplay between physics (effective thermal descriptions of noisy devices) and complexity theory (pseudorandomness vs. learnability) is one of the key reasons Hamiltonian learning is an interesting problem at the interface of quantum information, condensed matter, and theoretical computer science.

4 Known Results

Information-theoretically, any algorithm for learning a local Hamiltonian from Gibbs-state samples must use at least

$$\Omega\left(\frac{\log n\, e^{\Omega(\beta)}}{\beta^2\, \varepsilon^2}\right)$$

copies of the Gibbs state to recover all coefficients to accuracy ε. The following table summarizes upper bounds that are known for different settings.

	Sample complexity	Time complexity	Measurement size (Pauli expectations)
Lattices	$O\left(\frac{\log n\, e^{\mathrm{poly}(\beta)}}{\beta^2\,\varepsilon^2}\right)$	$O\left(\frac{n\log n\, e^{\mathrm{poly}(\beta)}}{\beta^2\,\varepsilon^2}\right)$	$\mathrm{poly}\big(\beta\log(1/\varepsilon)\big)$
	$\mathrm{poly}(n)\,\frac{e^{\mathrm{poly}(\beta)}}{\beta^2\,\varepsilon^2}$	$2^{O(n)}$	$O(1)$
General graphs	$\mathrm{poly}(n)\,(1/\varepsilon)^{\beta^2}$	$\mathrm{poly}(n)\,(1/\varepsilon)^{\beta^2}$	$O\big(\beta^2\log(1/\varepsilon)\big)$
	$O\left(\frac{\log n}{\varepsilon^2}\right)$	$O\left(\frac{n\log n}{\varepsilon^2}\right)$	$O\big(\log(1/\varepsilon)\big)$

Table 1: Schematic upper bounds for learning local Hamiltonians from Gibbs states. Here n is the number of qubits, β the inverse temperature, and ε the target additive error per coefficient. The bottom row corresponds to the high-temperature regime $\beta < \Theta(1)$.

5 Learning Algorithms

A central question is whether the unknown Hamiltonian H is information-theoretically learnable with only $\mathrm{poly}(n)$ samples of its Gibbs state.

5.1 Identifiability via Jaynes' Principle

Recall **Jaynes' principle** from Problem Set 1: among all density matrices ρ consistent with a given collection of expectation values, the Gibbs state is the *unique* maximizer of the von Neumann entropy

$$S(\rho) = -\operatorname{Tr}(\rho\log\rho).$$

Fix the local observables $\{E_a\}_{a=1}^m$ as before, and suppose we are given numbers $\{\mu_a\}_{a=1}^m$ such that

$$\forall a \quad \operatorname{Tr}(E_a\rho) = \mu_a.$$

Jaynes' principle states that the solution to

$$\max_\rho\ S(\rho) \quad \text{s.t.} \quad \operatorname{Tr}(E_a\rho) = \mu_a \ \text{ for all } a$$

is a *Gibbs state* of the form

$$\rho_\beta(\lambda) = \frac{e^{-\beta\sum_a \lambda_a E_a}}{\operatorname{Tr}\big(e^{-\beta\sum_a \lambda_a E_a}\big)}$$

for some parameters $\lambda = (\lambda_1,\dots,\lambda_m)$, and this maximizer is unique.

In particular, if another state ρ' of the same form,

$$\rho' = \rho_\beta(\lambda') = \frac{e^{-\beta\sum_a \lambda'_a E_a}}{\operatorname{Tr}\big(e^{-\beta\sum_a \lambda'_a E_a}\big)},$$

satisfies the same constraints

$$\operatorname{Tr}(E_a\rho') = \operatorname{Tr}(E_a\rho_\beta(\lambda)) = \mu_a \quad \text{for all } a,$$

then Jaynes' principle forces $\rho' = \rho_\beta(\lambda)$, and hence

$$\lambda' = \lambda.$$

Thus, *knowing all the expectation values μ_a exactly is enough to pin down the underlying Hamiltonian parameters uniquely.*

5.2 Log-Partition Function and Convexity

There is another way to see this identifiability, which will be important for quantifying how many samples we need.

Define the **log-partition function**

$$Z(\lambda) = \mathrm{Tr}\Big(e^{-\beta \sum_a \lambda_a E_a}\Big), \qquad \psi(\lambda) = \log Z(\lambda).$$

The gradient of ψ with respect to λ_a gives the expectation values:

$$\mu_a(\lambda) := \mathrm{Tr}(E_a \rho_\beta(\lambda)) = -\frac{1}{\beta}\,\frac{\partial \psi(\lambda)}{\partial \lambda_a}.$$

Equivalently,

$$\nabla \psi(\lambda) = -\beta\, \mu(\lambda),$$

where $\mu(\lambda) = (\mu_1(\lambda), \ldots, \mu_m(\lambda))$.

The function $\psi(\lambda)$ is convex in λ, and in fact *strictly* convex under mild conditions. In one dimension, you can picture this as a curve $\lambda \mapsto \psi(\lambda)$ whose slope is $-\beta\mu(\lambda)$:

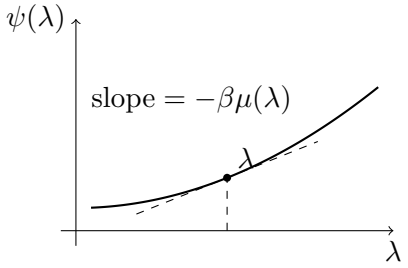

Figure 2: The log-partition function $\psi(\lambda)$ is convex, and its derivative gives the expectation value $\mu(\lambda)$.

Geometrically, the map

$$\lambda \longmapsto \mu(\lambda)$$

is the gradient of a convex function, so it is injective. Therefore, the vector of expectations $\mu = (\mu_a)_a$ determines a unique parameter vector λ.

In practice, however, we can only estimate the μ_a's from finitely many samples, so we never know them exactly. To handle this, we need not just convexity but *strong convexity*: the curvature of $\psi(\lambda)$ should be bounded away from zero so that small errors in μ translate to controlled errors in λ.

Theorem 5.1 (Informal stability of parameters)**.** *Suppose $\lambda, \lambda' \in \mathbb{R}^m$ and let $\mu(\lambda), \mu(\lambda')$ be the corresponding expectation vectors. In the regime we care about (local Hamiltonians on low-degree graphs), the log-partition function is strongly convex, and there exists a polynomial $p(\cdot)$ such that*

$$\sum_{a=1}^{m} (\lambda_a - \lambda'_a)^2 \;\le\; p(n) \sum_{a=1}^{m} \big(\mu_a(\lambda) - \mu_a(\lambda')\big)^2.$$

In particular, if each $\mu_a(\lambda)$ is estimated to within additive error δ, then the resulting estimate of λ is accurate to within $\mathrm{poly}(n)\,\delta$ *in ℓ_2-norm.*

This shows that, information-theoretically, *polynomially many* samples suffices to learn all the couplings $\{\lambda_a\}$: with $\mathrm{poly}(n, 1/\varepsilon)$ copies of $\rho_\beta(H)$ we can estimate each μ_a to accuracy $\varepsilon/\,\mathrm{poly}(n)$, and hence recover λ to accuracy ε.

5.3 An Algorithmic Viewpoint

The discussion above suggests an algorithmic formulation. Let $\widehat{\mu}_a$ denote empirical estimates of the expectations $\mu_a = \mathrm{Tr}(E_a \rho_\beta(\lambda))$ obtained from experiments. We would like to *solve for* parameters λ' such that

$$\forall a \quad \mathrm{Tr}\big(E_a \rho_\beta(\lambda')\big) \approx \widehat{\mu}_a.$$

Equivalently, we are trying to solve the system of nonlinear equations

$$\mathrm{Tr}\big(E_a \rho_\beta(\lambda')\big) = \mathrm{Tr}\big(E_a \rho_\beta(\lambda)\big) \quad \text{for all } a$$

for the unknown vector λ'.

From an information-theoretic perspective, this is fine: the strong convexity argument guarantees uniqueness and stability of the solution. However, computationally this system is problematic:

- The map $\lambda \mapsto \rho_\beta(\lambda)$ is highly nonlinear, involving a matrix exponential of a sum of operators.
- The resulting equations need not be low-degree polynomials in λ'.

Thus, directly solving this system is not known to be efficient, either in the quantum setting or even in the corresponding classical problems.

5.4 Classical Warm-Up: The Glauber / Ising Model

To see how one might make progress algorithmically, it is useful to first look at a classical special case where efficient learning is possible.

Consider the ferromagnetic Ising (Glauber) model on a graph $G = (V, E)$ with spins $z_i \in \{\pm 1\}$. The Hamiltonian has the form

$$H(z) = \sum_{(i,j)\in E} \lambda_{ij}\, z_i z_j,$$

and the associated Gibbs distribution at inverse temperature β is

$$\pi_\beta(z_1, \ldots, z_n) = \frac{1}{Z} \exp\left(-\beta \sum_{(i,j)\in E} \lambda_{ij}\, z_i z_j \right).$$

Let $N(i)$ denote the neighbors of vertex i. A key simplification in the classical setting is that the local conditional distribution of a single spin depends only on its neighbors:

$$\pi_\beta(z_i \mid z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n) = \pi_\beta(z_i \mid z_{N(i)}).$$

Writing

$$h_i(z_{N(i)}) := \sum_{j \in N(i)} \lambda_{ij} z_j,$$

a short calculation shows that

$$\pi_\beta(z_i \mid z_{N(i)}) = \frac{\exp\big(-\beta z_i h_i(z_{N(i)})\big)}{\exp\big(-\beta h_i(z_{N(i)})\big) + \exp\big(+\beta h_i(z_{N(i)})\big)}.$$

In particular,

$$\pi_\beta(z_i = +1 \mid z_{N(i)}) = \frac{\exp\big(-\beta h_i(z_{N(i)})\big)}{\exp\big(-\beta h_i(z_{N(i)})\big) + \exp\big(+\beta h_i(z_{N(i)})\big)} = \frac{1}{1 + \exp\big(2\beta h_i(z_{N(i)})\big)}.$$

That is, the conditional distribution at each site is just a simple logistic function of a linear combination of its neighbors.

From Gibbs samples of the Ising model, we can therefore:

- empirically estimate these local conditionals $\pi_\beta(z_i \mid z_{N(i)})$, and
- recover the couplings λ_{ij} by fitting a logistic regression at each vertex.

This yields efficient algorithms (both statistically and computationally) for learning classical Ising Hamiltonians on sparse graphs from Gibbs-state data, and serves as a guiding analogy for the more difficult quantum case we turn to next.

5.5 Quantum Perspective: Gibbs Samplers and Detailed Balance

In the quantum setting there is no literal notion of conditioning on all spins except one: we do not have a joint probability distribution over measurement outcomes that is independent of the measurement process. However, we *do* know how to *Gibbs sample*, i.e., how to construct dynamics that have the target Gibbs state as a fixed point. This suggests an operational way to phrase the learning problem.

Classical viewpoint revisited. Think of the classical Ising model again. Suppose we are given samples from the unknown Gibbs distribution $p^{(\lambda)}$. For any candidate parameter vector λ' we can construct the Glauber dynamics $M^{(\lambda')}$ whose stationary distribution is $p^{(\lambda')}$. Consider the following conceptual procedure:

(i) Start from a distribution P (in our case, $P = p^{(\lambda)}$) represented by samples.

(ii) Apply one (or several) steps of the Gibbs sampler $M^{(\lambda')}$ to obtain a new distribution Q.

If $\lambda' = \lambda$, then $p^{(\lambda)}$ is stationary for $M^{(\lambda')}$, so $Q = P$. On the other hand, if λ' differs substantially from λ, then the action of $M^{(\lambda')}$ moves P towards $p^{(\lambda')}$, and $\|Q - P\|_1$ becomes large. Thus learning λ can be viewed as finding the parameters for which the observed distribution is approximately a fixed point of the corresponding Gibbs sampler.

Quantum analogue. In the quantum case we can adopt the same philosophy. For a candidate Hamiltonian H' there exist *Gibbs-sampling Lindbladians* $\mathcal{L}_{H'}$ whose unique fixed point is the Gibbs state $\rho_\beta(H')$. In work of Anthony Chen et al., such $\mathcal{L}_{H'}$ are constructed to be *quasi-local*: they are still global operators, but the weight of long-range terms decays rapidly with their support size.

Conceptually, given samples from the unknown state $\rho_\beta(H)$, one could:

(i) guess a candidate Hamiltonian H' (equivalently, parameters λ');

(ii) apply the quantum channel $e^{t\mathcal{L}_{H'}}$ to copies of $\rho_\beta(H)$;

(iii) test whether the output state is still close to the input.

If $H' = H$, then $\rho_\beta(H)$ is a fixed point of $\mathcal{L}_{H'}$ and the state is invariant. If λ' is far from λ, the dynamics drive $\rho_\beta(H)$ towards $\rho_\beta(H')$, and the distance between input and output becomes large. This "learn by testing stationarity" picture motivates a quantum learning algorithm based on Gibbs samplers, although turning this into a *time-efficient* algorithm remains nontrivial.

The quasi-local Lindbladians are an important starting point, but they are still too global to yield efficient learning procedures. To push this picture further, we turn to the framework of detailed balance.

5.6 Classical and Quantum Detailed Balance

Classical detailed balance. For a Markov chain with transition matrix M and stationary distribution p, the detailed balance condition is

$$M_{x\leftarrow x'}\, p(x') = M_{x'\leftarrow x}\, p(x) \qquad \text{for all } x, x'.$$

Equivalently, writing $P = \mathrm{diag}(p)$, we have $PM = M^\top P$. A more symmetric way to express this is via the *discriminant matrix*

$$D := P^{1/2} M P^{-1/2},$$

which is symmetric if and only if the chain satisfies detailed balance.

A baby version of quantum detailed balance. In the quantum setting, a full definition of detailed balance for Lindbladians is more involved. For the learning problem, it is enough to consider the following "baby" condition that captures the same intuition.

Let $\rho(\lambda)$ and $\rho(\lambda')$ denote two Gibbs states associated with couplings λ and λ'. Consider the relation

$$\rho(\lambda')^{1/2} A \rho(\lambda')^{-1/2}\, \rho(\lambda) = \rho(\lambda)\, \rho(\lambda')^{-1/2} A \rho(\lambda')^{1/2} \qquad \text{for all operators } A. \tag{1}$$

We can rewrite this by multiplying on the left and right by $\rho(\lambda')^{-1/2}$:

$$AX = XA \quad \text{for all } A, \qquad X := \rho(\lambda')^{-1/2}\, \rho(\lambda)\, \rho(\lambda')^{-1/2}.$$

Thus X commutes with every operator, so X must be proportional to the identity: $X = c\,\mathbb{I}$. Taking the trace and using $\mathrm{Tr}(\rho(\lambda)) = \mathrm{Tr}(\rho(\lambda')) = 1$ shows that $c = 1$, hence $X = \mathbb{I}$ and

$$\rho(\lambda) = \rho(\lambda').$$

In other words, the "detailed balance" relation holds for all A if and only if the two Gibbs states coincide, and therefore $\lambda' = \lambda$.

From a learning perspective, this is encouraging: it says that the Gibbs state and the associated detailed-balance relations determine the couplings uniquely. However, the similarity transform

$$A \longmapsto \rho(\lambda')^{1/2} A \rho(\lambda')^{-1/2}$$

is extremely unpleasant to work with. Physically, it corresponds to *imaginary-time evolution* generated by H', which is highly nonlocal and difficult to implement efficiently. For condensed-matter physicists this operator is a nightmare, so while the formal characterization is clean, it does not directly yield a practical learning algorithm. However, there are ways to get around this by turning the equations into something useful, but we won't concern ourselves with how to do that in this class.

5.7 Connection to the KMS Condition

The "baby" quantum detailed balance relation

$$\rho(\lambda')^{1/2} A \rho(\lambda')^{-1/2}\, \rho(\lambda) = \rho(\lambda)\, \rho(\lambda')^{-1/2} A \rho(\lambda')^{1/2} \qquad \text{for all operators } A \tag{2}$$

is closely related to the *KMS condition* familiar from equilibrium statistical mechanics.

To see the connection, fix an arbitrary observable B and take the trace of (2) after multiplying by B on the left:

$$\mathrm{Tr}\big(B\, \rho(\lambda')^{1/2} A \rho(\lambda')^{-1/2} \rho(\lambda)\big) = \mathrm{Tr}\big(B\, \rho(\lambda)\, \rho(\lambda')^{-1/2} A \rho(\lambda')^{1/2}\big).$$

Define

$$C := \rho(\lambda')^{1/2} A \rho(\lambda')^{-1/2},$$

so that

$$\rho(\lambda')^{-1/2} A \rho(\lambda')^{1/2} = \rho(\lambda')^{-1} C\, \rho(\lambda').$$

Using cyclicity of the trace, we can rewrite the previous identity as

$$\mathrm{Tr}\big(BC\, \rho(\lambda)\big) = \mathrm{Tr}\big(B\, \rho(\lambda)\, \rho(\lambda')^{-1} C\, \rho(\lambda')\big) \qquad \text{for all } B. \tag{3}$$

Equation (3) has exactly the structure of a *KMS-type* condition: the operator C on the right-hand side is conjugated by the "imaginary-time evolution" generated by $\rho(\lambda')$.

To make this more explicit, specialize to $\lambda' = \lambda$ and write the Gibbs state as

$$\rho(\lambda) = \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}.$$

Then

$$\rho(\lambda)^{-1} C\, \rho(\lambda) = e^{\beta H} C\, e^{-\beta H},$$

which is precisely the Heisenberg evolution of C by imaginary time $i\beta$. In this notation, (3) becomes schematically

$$\mathrm{Tr}\big(\rho\, BC\big) = \mathrm{Tr}\big(\rho\, B\, C(i\beta)\big),$$

mirroring the standard KMS condition

$$\mathrm{Tr}\big(\rho\, A(t)B\big) = \mathrm{Tr}\big(\rho\, BA(t+i\beta)\big),$$

which characterizes thermal equilibrium states.

Thus the baby quantum detailed balance relation can be viewed as a finite-dimensional, Gibbs-state version of the KMS condition: the Gibbs state $\rho_\beta(H)$ is exactly the state with respect to which this imaginary-time invariance holds. Conceptually, this ties together three perspectives on thermal states:

- Jaynes' maximum-entropy principle;
- detailed balance for Gibbs samplers;
- and the KMS characterization of equilibrium via analytic continuation in imaginary time.