CHAPTER 9

Learning Gibbs States: Low Temperature

In the previous chapter, we saw an algorithm for learning Gibbs states at temperatures above some absolute constant depending on the geometry of the Hamiltonian. Although the algorithm we considered breaks down at lower temperatures, there is no *a priori* reason why there shouldn't exist any algorithm that succeeds in that regime. Indeed, one can show that at least *information-theoretically*, one can learn at arbitrary temperatures with sample complexity scaling exponentially in β and inversely in $\mathrm{poly}(\beta)$ [**AAKS21**]. For a while, it was an open question whether one could achieve this rate with a *computationally* efficient algorithm. This was resolved in a recent breakthrough of [**BLMT24**] which gave an algorithm with runtime and sample complexity $\mathrm{poly}(m,(1/\epsilon)^{2^\beta})$, where m is the number of terms in the Hamiltonian; this doubly exponential dependence on β was subsequently improved to singly exponential dependence by [**Nar24**]. These papers rely on a powerful algorithmic framework called *sum-of-squares programming*; unfortunately, a complete exposition of this approach would take us too far afield, and instead we will consider a different algorithm due to the very recent work of [**CAN25**]. This last paper gave an alternative algorithm with better dependence on the system size, and using an arguably more intuitive approach.

1. Technical Preliminaries

Throughout, fix a Hamiltonian

$$H = \sum_a \lambda_a P_a = \sum_\eta \eta \Pi_\eta \,,$$

where η ranges over the distinct eigenvalues ("energies") of H, and $\Pi_\eta = |\eta\rangle\langle\eta|\,\eta$ denotes projection to the eigenspace corresponding to eigenvalue η. Throughout, $\rho \propto e^{-\beta H}$ will denote its Gibbs state.

Any such Gibbs state naturally induces the following inner product which generalizes the classical L_2 inner product with respect to a probability measure:

Definition 147 (KMS inner product)**.** *Given operators A, B and a density matrix ρ, their **KMS inner product** is given by*

$$\langle A, B\rangle_\rho = \mathrm{tr}(A\rho^{1/2}B^\dagger\rho^{1/2})\,.$$

*This induces the **KMS norm** $\|A\|_\rho^2 \triangleq \langle A, B\rangle_\rho$.*

We will often be interested in *differences* between energies $\eta - \eta'$:

Definition 148 (Bohr frequencies and Bohr decomposition)**.** *The set of **Bohr frequencies**, denoted $B(H)$, of a Hamiltonian H consist of all differences $\eta - \eta'$ between eigenvalues of H. We will always use the letter ν, possibly with superscripts,* *to denote Bohr frequencies, and* $\sum_\nu$ *to denote* $\sum_{\nu \in B(H)}$ *when* H *is clear from context.*

Any operator A *can naturally be decomposed into blocks* $\Pi_{\eta'} A \Pi_\eta$ *corresponding to different pairs of eigenspaces, and the Bohr frequencies give a natural set of "bands" for grouping together these blocks. Given operator* A *and Bohr frequency* ν*, define*

$$A_\nu = \sum_{\eta' - \eta = \nu} \Pi_{\eta'} A \Pi_\eta \,.$$

Counterintuitively, the algorithm we will describe for learning Gibbs states, which are inherently *static* objects, will arise from reasoning about *dynamics* associated to the Hamiltonian. These were introduced briefly in Section 3. We will discuss their relevance to the learning algorithm and analysis in Section 2, but for now we simply recall their definition:

Definition 149 (Time evolution)**.** *Given a time* $t \in \mathbb{R}$ *and a Hermitian* H*, the associated* ***time evolution*** *operator is* e^{-iHt}*.*[1] *Under the* ***Schrödinger picture****, a state* ρ_0 *undergoing time evolution becomes* $\rho_t = e^{-iHt}\rho_0 e^{iHt}$ *at time* t*. Dually, one can consider the time evolution of* observables*. Under the* ***Heisenberg picture****, an observable* A_0 *undergoing time evolution becomes* $A_t = e^{iHt} A_0 e^{-iHt}$ *at time* t*. We will adopt some shorthand for the latter: given an operator* A*, define*

$$A_H(t) \triangleq e^{iHt} A e^{-iHt} \,.$$

While time evolution is defined with $t \in \mathbb{R}$ *(indeed, this is essential for* e^{iHt} *to be unitary), we can also consider conjugating operators by the Gibbs state instead of by* e^{iHt}*, which would correspond to* imaginary t*, to get*

$$e^{-\beta H} A e^{\beta H} \,.$$

We will refer to this as ***imaginary time evolution*** *and occasionally abuse notation by writing this as* $A_H(i\beta)$*.*

The Bohr decomposition behaves nicely under time evolution and imaginary time evolution:

Lemma 150. *For any* $t \in \mathbb{C}$*,*

$$A_H(t) = \sum_\nu e^{i\nu t} A_\nu \,.$$

Proof. By definition $e^{iHt} = \sum_\eta e^{i\eta t}\Pi_\eta$ has the same eigenvectors as H, so $\Pi_\eta e^{-iHt} = e^{-i\eta t}\Pi_\eta$ and $\Pi_{\eta'} e^{iHt} = e^{i\eta' t}$. Therefore, for any η, η' for which $\eta' - \eta = \nu$, we have

$$e^{iHt}\Pi_{\eta'} A \Pi_\eta e^{-iHt} = e^{i\nu t}\Pi_{\eta'} A \Pi_\eta \,,$$

from which the claim follows by linearity. □

[1] Here we have flipped the sign from what was defined in Section 3 as it is more convenient for some of the subsequent calculations.

2. Learning by Exploiting Detailed Balance

The starting point for the proof is the **Kubo-Martin-Schwinger (KMS) condition**, which can be thought of as a quantum analogue of detailed balance.

Theorem 151 (KMS condition)**.** *Let $\rho' = e^{-\beta H'}/tr(e^{-\beta H'})$ for some Hamiltonian H'. The equation*

$$\mathrm{tr}(\rho' A_H(t) O) = \mathrm{tr}(\rho' O A_H(t + i\beta)) \tag{57}$$

holds for all operators O and A and all times $t \in \mathbb{C}$[2] *if and only if $H' = H + c\mathrm{Id}$ for some absolute constant $c \in \mathbb{R}$.*

While this statement is incredibly powerful, as we will see, its proof is rather trivial.

Proof. One can readily verify that for $\rho = e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$,

$$\begin{aligned}\mathrm{tr}(\rho A_H(t) O) &= \frac{1}{Z}\mathrm{tr}(e^{iH(t+i\beta)} A e^{-iHt} O) \\ &= \frac{1}{Z}\mathrm{tr}(e^{iH(t+i\beta)} A e^{-iH(t+i\beta)} e^{-\beta H} O) \\ &= \mathrm{tr}(A_H(t+i\beta)\rho O) \\ &= \mathrm{tr}(\rho O A_H(t+i\beta))\,.\end{aligned}$$

To show the converse, note that Eq. (57) holding for all O, A, t is equivalent to the condition that

$$\rho' A_H(t) = A_H(t + i\beta)\rho' = \rho A_H(t) \rho^{-1} \rho'$$

for all A, t. This in particular implies that $A = (\rho^{-1}\rho')^{-1} A (\rho^{-1}\rho')$ for all A, which implies that the mixed states ρ and ρ' are equal, as desired. □

In other words, the only state that satisfies the KMS condition with respect to H is the Gibbs state. The rest of this section is about making this insight quantitative in order to extract out a learning algorithm. This requires answering two questions. First, if $\rho' \propto e^{-\beta H'}$ is *close*, in an appropriate sense, to satisfying the KMS condition, does that imply H' is close to H? Second, how does one computationally efficiently find an H' for which this is the case?

2.1. A KMS alternative that sees locality

The above argument that the only state that satisfies the KMS condition with respect to H is the Gibbs state is unfortunately rather *global* in nature as it involves multiplying by inverses of matrix exponentials. This is a horribly ill-conditioned operation as many of the eigenvalues of ρ are exponentially small. In this section, we will outline an approach to making this argument more "local" by carefully designing O and A.

First, let us slightly shift perspectives by switching the roles of H, ρ and H', ρ' in Theorem 151. For convenience, let us also replace t by $t - i\beta/2$ just so that instead of $(t, t + i\beta)$, we get $(t - i\beta/2, t + i\beta/2)$. This gives rise to the following equivalent formulation of the KMS condition:

---

[2]The KMS condition was originally devised as an alternative characterization of Gibbs states that can extend to *infinite dimensions*; in those contexts one has to be a bit careful about issues of analyticity and t is taken to be real, but in finite dimensions this is not an issue.

Theorem 152 (Dual KMS)**.** *Given a Hamiltonian H', operators O and A, and time $t \in \mathbb{C}$, define the observable*

$$\Delta[\![H'; O, A, t]\!] \coloneqq O A_{H'}(t + i\beta/2) - A_{H'}(t - i\beta/2) O \tag{58}$$

measuring the extent to which some equation in the KMS condition is violated. Then

$$\mathrm{tr}(\rho \Delta[\![H'; O, A, t]\!]) = 0$$

for all O, A, t if and only if $H = H' + c\mathrm{Id}$ for some absolute constant c.

If we had some way of enumerating over all H', O, A, t, then we might imagine using our copies of the Gibbs state ρ to estimate all of these observable values and hope that if we have found some H' for which the corresponding observable values $\mathrm{tr}(\rho\Delta^*_{H';O,A,t})$ were all very small, then $H' \approx H$.

Thus far, we haven't done anything new on top of Theorem 151, but we are now in a position to try making things more local. The first key insight towards doing so is to realize that there is a certain weighted combination of the observables $\Delta[\![H'; O, A, t]\!]$'s for *local* operators O and A which is small if and only if H and H' are close (up to additive shift).

Theorem 153 (Identifiability equation)**.** *For any operators O and A, we have*

$$\frac{\beta}{2}\langle O, [A, H - H']\rangle_\rho = \mathrm{tr}(\rho \overline{\Delta}[\![H'; O, A]\!])\,,$$

for

$$\overline{\Delta}[\![H'; O, A]\!] \coloneqq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Delta[\![H'; O^\dagger_H(t), A, t]\!]\, g_\beta(t)\, \mathrm{d}t$$

where $g_\beta(t) \coloneqq \frac{2}{\beta} g(2t/\beta)$ for

$$g(t) \triangleq -\frac{\pi^{3/2}}{2\sqrt{2}(1 + \cosh(\pi t))}$$

(see Figure 1 for a plot – the particular functional form is not important, but the fact that it is rapidly decaying is).

We will prove this in Section 2.2. Although this result is stated in terms of general operators O and A, the following result morally tells us that it suffices to consider 1-local Paulis A and $O = [A, H - H']$ and motivates why we consider $\langle O, [A, H - H']\rangle_\rho$ in the first place:

Lemma 154. *Suppose $H = \sum_a \lambda_a P_a$ and $H' = \sum_a \lambda_a E'_a$ are Hamiltonians with the same set of couplings. If $\frac{1}{2^n}\|[A, H - H']\|_F^2 \le \epsilon^2$ for all three Pauli operators $A \in \{X_i, Y_i, Z_i\}$ acting solely on the i-th qubit, then $|\lambda_a - \lambda'_a| \le \epsilon$ for every term P_a acting on qubit i.*

Proof. For any A, we have

$$\frac{1}{2^n}\|[A, H - H']\|_F^2 = \frac{1}{2^n}\mathrm{tr}([A, [A, H - H']](H - H'))\,.$$

If $A \in \{X_i, Y_i, Z_i\}$ and P is some Pauli operator, then $[A, P] = 0$ if P acts as A or Id on the i-th qubit, and otherwise $[A, P]$ is the operator which is identical to P off of the i-th qubit and equal to some signed Pauli on the i-th qubit. Moreover, in this case $[A, [A, P]] = 4P$.

We conclude that $\sum_{A\in\{X_i,Y_i,Z_i\}}[A,[A,H-H']] = 8\sum_{a\sim i}(\lambda_a - \lambda'_a)P_a$, where the sum is over terms a which act on i, and thus

$$3\epsilon^2 \geq \frac{1}{2^n}\sum_{A\in\{X_i,Y_i,Z_i\}} \mathrm{tr}([A,[A,H-H']](H-H')) = 8\sum_{a\sim i}(\lambda_a-\lambda'_a)^2\,,$$

from which the claim follows. □

Remark 155. *There is a slight but nontrivial catch that Lemma 154 pertains to the* Frobenius *norm of* $[A, H-H']$*, whereas Theorem 153 involves the* KMS *norm. For local operators like* $[A, H-H']$*, these can be related up to a* $e^{\mathrm{poly}(\beta)}$ *factor using ideas from Section 3 below. Proving this would take us too far afield, and we defer the interested reader to* [**CAN25**, Lemma III.6].

Modulo this remark, combining Theorem 153 and Lemma 154, we conclude that if H' was such that $\mathrm{tr}(\rho\overline{\Delta}[\![H';O,A]\!])$ was small for all 1-local Paulis A and $O = [A, H-H']$, then this would ensure that H and H' are equivalent. As we saw in the proof of Lemma 154, $[A, H-H']$ only consists of terms a which act on the i-th qubit, and for local Hamiltonians this is a constant number of terms. Furthermore, the operator $O^\dagger_H(t)$ that appears in the definition of $\overline{\Delta}[\![H';O,A]\!]$ is also approximately local, because intuitively the locality of H ensures that the time-evolved operator $O^\dagger_H(t)$ doesn't "spread out" too much in a short amount of time - this is the content of **Lieb-Robinson bounds**, which we discuss in Section 4.

There are however two important challenges remaining to "localizing" the KMS condition into something that can be algorithmically useful. First, the observables $\overline{\Delta}[\![H';O,A]\!]$ require knowledge of H, at the very least in order to write down $O^\dagger_H(t)$. Second, recall from the definition of Δ in Eq. (58) that they still involve the scary-looking imaginary-time-evolved operators $A_{H'}(t\pm i\beta/2)$. Imaginary time evolution involves conjugating by a fractional power of $e^{-\beta H'}$, which again might have exponentially small eigenvalues. So it would appear that we still haven't sidestepped the need to invert by ill-conditioned matrices.

The former issue is not so bad: even without knowing H, we can simply enumerate over guesses of the Hamiltonian in a way that we make precise in Section 4. The latter is the more fundamental issue, and we deal with this in Section 3 using a subtle **regularization trick** from the literature on quantum Gibbs sampling.

2.2. Proof of identifiability equation

In this section we prove Theorem 153. The key technical tools will be a *nested* Bohr decomposition with respect to the Bohr frequencies of H and H'.

Given Hamiltonians H_1 and H_2 with Bohr frequencies $B(H_1) = \{\nu_1\}$ and $B(H_2) = \{\nu_2\}$ and an operator A, we will use the following "double" decomposition:

$$(A_{\nu_1})_{\nu_2} = \sum_{\eta'_2-\eta_2=\nu_2}\ \sum_{\eta'_1-\eta_1=\nu_1} \Pi_{\eta'_2}\Pi_{\eta'_1}A\Pi_{\eta_1}\Pi_{\eta_2}\,.$$

Here η_1,η'_1 (resp. η_2,η'_2) denote eigenvalues of H_1 (resp. η_2), and the Π's are projectors to the corresponding eigenspaces.

This double decomposition gives us a way to analyze objects like the commutator on the left-hand side of the identifiability equation.

Lemma 156 (Calculations with double decomposition)**.** *The following identities hold:*

$$[A, H_2 - H_1] = -\sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2}(\nu_2 - \nu_1)\,.$$

$$e^{H_2}e^{-H_1}Ae^{H_1}e^{-H_2} - e^{-H_2}e^{H_1}Ae^{-H_1}e^{H_2} = \sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\,.$$

Proof. We have

$$\begin{aligned}
[A, H_1] &= AH_1 - H_1A \\
&= \sum_{\nu_1} \sum_{\eta'-\eta=\nu_1} \Pi_{\eta'}A\Pi_\eta H_1 - H_1\Pi_{\eta'}A\Pi_\eta \\
&= \sum_{\nu_1} \sum_{\eta'-\eta=\nu_1} (\eta - \eta')\Pi_{\eta'}A\Pi_\eta \\
&= -\sum_{\nu_1} \nu_1 A_{\nu_1}\,,
\end{aligned}$$

and similarly for $[A, H_2]$. So the first part follows.

For the second part,

$$e^{-H_1}Ae^{H_1} = \sum_{\nu_1} \sum_{\eta_1'-\eta_1=\nu_1} e^{-(\eta_1'-\eta_1)}\Pi_{\eta_1'}A\Pi_{\eta_1} = \sum_{\nu_1} e^{-\nu_1}A_{\nu_1}\,,$$

and similarly

$$e^{H_2}A_{\nu_1}e^{-H_2} = \sum_{\nu_2} e^{\nu_2}(A_{\nu_1})_{\nu_2}\,,$$

so

$$e^{H_2}e^{-H_1}Ae^{H_1}e^{-H_2} = \sum_{\nu_1,\nu_2} e^{\nu_2-\nu_1}(A_{\nu_1})_{\nu_2}$$

and similarly

$$e^{-H_2}e^{H_1}Ae^{-H_1}e^{H_2} = \sum_{\nu_1,\nu_2} e^{\nu_1-\nu_2}(A_{\nu_1})_{\nu_2}\,.$$

Note that $e^{\nu_2-\nu_1} - e^{\nu_1-\nu_2} = 2\sinh(\nu_2 - \nu_1)$, so the second part follows. $\square$

As sinh is a bijective function, the above lemma gives a crucial link between commutator differences and interleaved imaginary time evolution differences, formalized as follows:

Lemma 157 (Commutator difference in time domain)**.**

$$[A, H_2 - H_1] =$$
$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \Big[e^{H_2}e^{-H_1}A_{H_1}(t)e^{H_1}e^{-H_2} - e^{-H_2}e^{H_1}A_{H_1}(t)e^{-H_1}e^{H_2}\Big]_{H_2}(-t)\cdot g(t)\,\mathrm{d}t$$

for

$$\hat{g}[\omega] = -\frac{\omega}{2\sinh[\omega]}$$

and

$$g(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \hat{g}[\omega]e^{-i\omega t}\,\mathrm{d}t = -\frac{\pi^{3/2}}{2\sqrt{2}(1+\cosh(\pi t))}\,.$$

PROOF. We have

$$
\begin{aligned}
[A, H_2 - H_1] &= -\sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2}(\nu_2 - \nu_1) \\
&= \sum_{\nu_1,\nu_2} \hat{g}(\nu_2 - \nu_1) \cdot (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1) \\
&= \frac{1}{\sqrt{2\pi}} \sum_{\nu_1,\nu_2} \int_{-\infty}^{\infty} g(t) e^{-i(\nu_2-\nu_1)t} (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\pi}} \sum_{\nu_1,\nu_2} \int_{-\infty}^{\infty} e^{-i\nu_2 t} g(t) (A_{\nu_1} e^{i\nu_1 t})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\,\mathrm{d}t\,.
\end{aligned}
$$

where in the third step we used Fourier inversion. Using Lemma 150 and Lemma 156, the claimed identity follows. □

Note that the function $g(t)$ is rapidly decaying, see Figure 1 below.

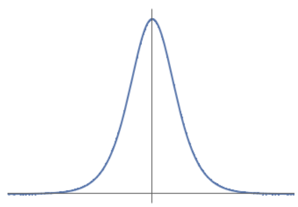

FIGURE 1. Plot of $g(t)$ from Theorem 153

We can now complete the proof of the identifiability equation.

PROOF OF THEOREM 153. Let ρ and ρ' denote the Gibbs states for H and H'. Taking $H_1 = \beta H'/2$ and $H_2 = \beta H/2$ in Lemma 157, consider the first term in the integral. Using that time evolution via H commutes with left- and right-multiplication by $\sqrt{\rho}$, we find that the first term equals

$$
\begin{aligned}
&\operatorname{tr}(\sqrt{\rho} O^\dagger \sqrt{\rho} [\sqrt{\rho^{-1}} \sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \sqrt{\rho}]_{\beta H/2}(-t)) \\
&\quad = \operatorname{tr}(O^\dagger [\sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \rho]_{\beta H/2}(-t)) \\
&\quad = \operatorname{tr}(O^\dagger_{\beta H/2}(t) \cdot \sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \rho) \\
&\quad = \operatorname{tr}(O^\dagger_H(t\beta/2) \cdot \sqrt{\rho'} A_{H'}(t\beta/2) \sqrt{\rho'^{-1}} \rho)\,,
\end{aligned}
$$

where in the penultimate step we pushed the reverse time evolution onto $O^\dagger$. Similarly

$$
\begin{aligned}
&\operatorname{tr}(\sqrt{\rho} O^\dagger \sqrt{\rho} [\sqrt{\rho} \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'} \sqrt{\rho^{-1}}]_{\beta H/2}(-t)) \\
&\quad = \operatorname{tr}(O^\dagger [\rho \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'}]_{\beta H/2}(-t)) \\
&\quad = \operatorname{tr}(O^\dagger_{\beta H/2}(t) \cdot \rho \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'}) \\
&\quad = \operatorname{tr}(O^\dagger_H(t\beta/2) \cdot \rho \sqrt{\rho'^{-1}} A_{H'}(t\beta/2) \sqrt{\rho'})\,.
\end{aligned}
$$

We conclude that

$$
\frac{\beta}{2} \langle O, [A, H - H'] \rangle_\rho = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \operatorname{tr}(\rho \Delta [\![H'; O^\dagger_H(t\beta/2), A, t\beta/2]\!])\,\mathrm{d}t\,.
$$

The identifiability equation follows by a change of variable $t \mapsto 2t/\beta$. □

3. Regularization

3.1. Preliminaries

Definition 158 (Operator Fourier transform). *Given a Hamiltonian H and an operator A, define the* ***operator Fourier transform (FT)*** $\hat{A}_H$ *by*

$$\hat{A}_H[\omega] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A_H(t) e^{-i\omega t} f(t)\, \mathrm{d}t\,,$$

where $f(t) = e^{-\sigma^2 t^2}\sqrt{\sigma\sqrt{2/\pi}}$ is a Gaussian filter. The "regularizing" role of $f(t)$ will be become clearer in the sequel. Its Fourier transform $\hat{f}[\omega] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-i\omega t} f(t)\,\mathrm{d}t$ satisfies $\hat{f}[\omega] = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}}\exp(-\omega^2/4\sigma^2)$.

Note that the operator FT commutes with imaginary time evolution:

$$e^{\beta H}\hat{A}_H[\omega]e^{-\beta H} = (\widehat{e^{\beta H} A e^{-\beta H}})_H[\omega]$$

Taking the operator FT of both sides of Lemma 150 results in the following useful identity:

$$\hat{A}_H[\omega] = \sum_\nu A_\nu \hat{f}[\omega - \nu]\,.$$

In other words, the operator FT gives "soft" access to the components in the Bohr decomposition of A. We have a corresponding "soft" Bohr decomposition, by Fourier duality.

Lemma 159. *For any operator A and Hermitian H,*

$$A = C_\sigma \int_{-\infty}^{\infty} \hat{A}_H[\omega]\,\mathrm{d}\omega$$

for $C_\sigma := \frac{1}{\sqrt{2\sigma\sqrt{2\pi}}}$.

Importantly, a straightforward calculation shows that the Gaussian filter ensures the operator FT decays exponentially in the frequency ω:

Lemma 160. *For any frequency ω and operator A satisfying $\|A\|_{\mathsf{op}} \le 1$,*

$$\hat{A}_H[\omega] = e^{-\beta\omega + \sigma^2\beta^2} e^{\beta H}\hat{A}_H[\omega - 2\sigma^2\beta]e^{-\beta H}$$

To see why this is useful, note that because $\|\hat{A}_H[\omega]\|_{\mathsf{op}} \le \hat{f}(0) = O(\sigma^{-1/2})$, this ensures that $\|e^{\beta H}\hat{A}_H[\omega']e^{-\beta H}\|_{\mathsf{op}} \lesssim e^{\sigma^2\beta^2 + \beta\omega'}\sigma^{-1/2}$. Crucially, the right-hand scales *exponentially in the frequency ω'*, rather than exponentially in the system size! In contrast, norm of the imaginary time-evolved observable $\|e^{\beta H}Ae^{-\beta H}\|_{\mathsf{op}}$ can scale exponentially in the system size.

3.2. Truncating the identifiability observable

Using Lemma 159, we can decompose A in $\langle O, [A, H - H']\rangle_\rho$ into low-frequency and high-frequency terms under operator FT with respect to H' and apply the