nature portfolio

Peer Review File

Efficient learning of ground and thermal states within phases of matter

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Problem statement:

The authors consider two variants of the problem of statistically estimating local (as well as some non-local) properties of quantum states at thermal equilibrium. These problems are motivated by the need for efficient schemes that can infer physically-relevant quantities from experimentally-prepared high-dimensional quantum many-body states.

(1) Gibbs state tomography: In the first variant, one can perform measurements on multiple copies of the thermal (Gibbs) state of a quantum system at a known temperature. The goal is to obtain a representation of this state from which extensive quantities such as energy or Renyi entropy can be inferred.
(2) Learning thermal phases: In the second variant, the measurement data obtained from various thermal states within a phase of matter (i.e. for instances with varying Hamiltonian parameters) are used to predict the local properties across the whole phase for unobserved instances.

In both problems the objective is to find the sample complexity: the fewest number of copies of the state required for precise estimates of (extensive or local) observables.

Strengths:

Improvement over previous works:

Assuming certain physically-motivated conditions on the thermal state, this work improves on previously known results for both of the statistical problems (1) and (2) mentioned above.

In problem (1), previously the number of required samples to estimate observables to a given precision scaled polynomially in the number of qudits n and exponentially in the size of the support of the observable. This result exponentially improves this sample complexity to polylog(n) and a polynomial dependency on the observables' support size.

In problem (2), this work establishes an improved dependency of the required number of samples on the precision parameter epsilon from exponential exp(1/epsilon) to quasi-polynomial exp(polylog(1/epsilon)).

These improvements drastically reduce the sample complexity of problems (1) and (2), hopefully making them more practically relevant.

Technical novelty:

This mentioned improvement in the sample complexity stems from a technical insight of the authors that uses the recently-introduced quantum Wasserstein distance. The large sample complexity in prior works is based on learning a very detailed description of the thermal state that is not needed when one only cares about estimating physically-relevant quantities and not certain worst-case global properties.

The authors obtain a refined notation of accuracy when learning thermal states that is just enough to infer physical properties, resulting in exponentially-improved sample complexity. This new insight may have further applications in the statical learning of quantum systems and beyond.

Weaknesses:

Even though the authors obtain improvement in the sample complexity, the performance with respect to other measures is still not favorable. This includes:

- Exponential running time: the number of samples needed for the Gibbs state tomography of non-commuting systems is polylog(n), but the running time of the algorithm that processes the measurement data scales exponentially exp(n) with the number of qudits.

- A concurrent work [LHT+23] on learning thermal phases (problem (2) above) obtains an unconditional result with respect to the prior distribution over the samples, but this manuscript still requires this distribution to satisfy the anti-concentration over the local marginals which seems hard to relax in the current approach as explained in Remark IV.7 of SI.

Summary:

In summary, this manuscript considers a well-motivated problem in statistical learning of quantum states, and by devising novel technical and conceptual tools, obtains an exponentially-improved sample complexity compared to past works. The insight from this paper can lead to further advancements on this topic and should be of interest to the broad readership of Nature Communications. As such, I recommend acceptance.

Comments for the authors:

- The result of Anshu [Anshu] on commuting Hamiltonians is repeatedly used in the manuscript. But given that [Anshu] is an unpublished note that hasn't been refereed, I'd recommend reproducing the relevant part of this result, at least minimally, in the supplementary notes.

- On page 5: You mentioned "Such observables can arise if we e.g. evolve a strictly local observable by a Hamiltonian with algebraically decaying interactions" as a reason for considering quasi-local observables, but it seems your techniques in the manuscript rely on

locality of the Hamiltonian terms (as opposed to algebraically decaying tails).

- If you were to sacrifice polylog(n) sample complexity to get poly(n) instead, does the result of [AAKS21] (without assuming the Markov property) allow you to improve the error eps to 1/eps^2 instead of your current 1/eps^4 scaling? For limited number of qudits n, a quadratic improvement in the accuracy epsilon may be more important in practice than an improved dependency on n.

- In Definition II.2, the Wasserstein distance between two n qubit quantum states ρ1, ρ2 should be between ρ0, ρ1.

- In Theorem II.3, you should mention that the assumption of decay of correlations is still needed for the (ii) case.

- The \mathcal{S} versus S in Eq. II.9 may confuse the readers, especially since S is also reserved for the von Neumann entropy and also supp(O_i) later on.

- More generally, the notion's introduced below II.9 are hard to unpack and refer to future Lemma III.5 that has not been discussed. I had to read this multiple times and go back to this part after reading the rest of the manuscript to grasp its content. I suggest modifying this part and also include explanations to aid the reader.

- In Corollary II.4. and also prior to that, the authors refer to the 'uniform' decay of correlations and 'uniform' Markov condition without specifying what they mean by uniform. I suggest you should introduce this notion when defining the exponential decay of correlations in Eq. II.6.

- It would also be helpful if the definition of the Markov property is given earlier on. The authors define CMI on Page 7, which seems to be the place to also define the decay of CMI (potentially even moved to earlier in the main).

- As mentioned before, the notation for entropy S(A) and region S(r_s) is conflicting in

Corollary II.4. For instance, what does S = A refer to below the equation? The writing of this part needs adjustments.

- Below Corollary II.4, it is mentioned that learning x up to ell_infty results in a multiplicative error in estimating the entropic quantities. It would be helpful to specify the scaling of this error with epsilon and n.

- In Figure 2, the notation seems to have changed from |_\mathcal{S}(r) to |_A(r).

- On page 12, there is a list of “Examples of classes of commuting states that satisfy exponential decay of correlations” but not all items are commuting.

- The same page, item 2, should be above a threshold temperature.

- In Proposition III.3 in the SI, where do you use the uniform assumption for the decay of correlations?

- It would be helpful to mention in Proposition III.12 that the bound on W_1 is obtained assuming decay of correlations and Markov property of sigma(beta, x).

- Can you explain in more detail how in the SI, Eq. (III.27) is obtained from Eq. (III.26)?

- It would be good to emphasize that regions ABC in Def III.9 may not partition the whole lattice as this seems to have been used in deriving Eq. (III.34) and below it.

- Two equations below Eq. (III.35), the RHS should be D(sigma(beta, x), sigma(beta, hat{x})) instead of D(rho, sigma(beta, x)).

- Can you add more details where/why the eps^4 dependency shows up in Corollary III.13?

Reviewer #2 (Remarks to the Author):

In this paper the authors considered two tasks: learning a Gibbs state from measurement data, and learning observable expectation values as a function of Hamiltonian parameters within a phase. For the first task, they proposed a method that learns the Gibbs state so that all appropriately normalized quasi-local observables can be computed to precision eps with sample complexity scaling logarithmically in the system size, assuming exponential decay of correlation. Their method works for non-commuting Hamiltonians and at constant temperature, unlike previous works that achieve a similar scaling for commuting Hamiltonians or high temperature. For the second task, they proposed a method to estimate local observable expectation values using the empirical average of samples whose local parameters are close to that of the target state. Their method achieves a much improved precision dependence compared to the previous state-of-the-art result in [HKT+22]. Their method works both for thermal states at constant temperature satisfying an exponential decay of correlation and for gapped ground states. The unifying theme of both methods is the continuity of the target observable as captured by Definition II.1 of a Lipschitz observable, and the Wassersterin distance defined through duality.

These methods significantly advance our understanding of these learning problems, and may inspire more applications of using Wasserstein distance in other learning tasks. I believe this manuscript should be accepted after certain modifications. Below are more detailed comments:

1. Can the authors comment on the possibility of learning observables in the presence of multiple phases? [HKT+22] numerically demonstrated that this is possible when there is data in all phases. This may be difficult to prove but may be something worth mentioning, at least as a future direction.

2. The authors emphasized that they obtain “point-wise guarantees” or their result “applies to the worst case”, as compared to [HKT+22] that only gives average guarantees. I think this comparison is a little unfair unless the authors provide the context that the training data come from a distribution that anti-concentrates (as pointed out at the end of Section I, but I think this should also be mentioned in the abstract and other places where this advantage is claimed).

3. The authors state that certain problems pertaining phases of matter are “computationally intractable in general”, and then say that having access to data can “conceivably speed up these computations”. This seems not very relevant to the present scenario, because the works cited here seem all to be about the difficulty of identifying phase boundary and generalizing to the thermodynamic limit, neither of which was addressed in this paper.
4. The notion of a Lipschitz observable features prominently in this paper. I think the authors should provide more intuition about this concept beyond the one sentence after Definition II.1.
5. In the sentence after Eq. (II.4), the authors state that the estimation is accurate “up to a multiplicative error”. It may be more accurate to say the error is relative to the Lipschitz norm of O. By multiplicative one would imagine it is relative to the expectation value of O.
6. In Eq. (II.8), I think it is worthwhile to explicitly state the temperature dependence in some way. It seems to me that this equation would not hold for zero temperature even if there is exponential decay of correlation (consider a classical Hamiltonian corresponding to SAT, a little change in the parameter may result in totally different ground states, even if all ground states are product states).
7. The authors should introduce what uniform clustering and uniform Markov means in the main text. They are rather important for understanding the results.
8. In the first paragraph in Section II.A.1, the authors write that “we make use of the continuity of the entropy functional with respect to the Wasserstein distance, mentioned in Equation (II.6).” Eq. (II.6) does not seem to be the correct equation.
9. The authors may wish to compare their result with the work “Reconstructing quantum states from local data” by Swingle and Kim. Their method allows sample efficient reconstruction of the ground state of a gapped local Hamiltonian, with error measured in trace distance. This means their method can also sample-efficiently estimate the entropy of any subregion without having an exponential dependence on the size of the region.
10. The first sentence in the proof ideas of Theorem II.5 is very confusing. Part of the confusion comes from the fact that xj is used both to represent a component of the parameter vector x (in the first paragraph of Section II.A) and a sample (in Theorem II.5). Different notations should be used for these two completely different things. The authors should also explain what an “enlargement” is. I also cannot understand xj and hj(xj) appearing at the same time. If xj here means a sample, then I have no idea what hj is. If xj

here means a component, then I cannot see how there can be an intersection between a set of components with [x-eps,x+eps]^m.
11. There seems to be a typo in the equation in Definition II.6. I imagine that we would want the right-hand side of the equation, which gives an upper bound of the approximation error, to go to zero when r goes to infinity. But in this equation it seems that this does not happen due to the eta(S) term, which greatly puzzles me. From the bottom of Page 18 in the arXiv version I think instead of |S|f(r) + eta(S), it should be |S|f(r+eta(S))?
12. In the main text there are many references to equations and theorems that are in the supplementary materials (basically whenever Section III is involved). Whenever there is a reference to the supplementary material it should be stated.

Reviewer #3 (Remarks to the Author):
In this manuscript, the authors explore provably efficient learning of local quantum-state properties within the same phase with an as small as possible number of samples [e.g., a polynomial function of log(system size) and precision parameters]. Two classes of physically relevant quantum states are considered: the Gibbs thermal states and gapped ground states of local Hamiltonians, where the number of parameters in the considered models are in a same order of the system size. Significant improvements on existing literature results are obtained, respectively. The key to these new findings is to fully exploit geometric localities of observables and of the quantum states within same phases of matter.

I particularly appreciate that, in the second task, by imposing geometric locality through exponentially decaying correlations or the generalized approximate local indistinguishability (GALI) condition, the effective dimension of the parameter space is reduced to ploylog(1/error), leading to a quasi-polynomial scaling in the uniform precision parameters for the number of samples, representing an exponential improvement with respect to the best existing result, and with a stronger recovery guarantee.

Based on the above arguments, I think the results in this paper advance significantly the research of provably efficient machine learning for quantum many-body physics. But I still have a few concerns and comments in the following. If they can be satisfactorily addressed

in the revised version, I would like to support the publication of the manuscript on Nature Communications.

1. Besides the theorems of existence type and the corresponding constructive proofs presented in this manuscript, can the authors give simple comments on the lower bound of the sample complexity? Also, the computational complexity better to be discussed.

2. In the definition of the GALI condition (Definition II.6), is it sufficient that f(r) only becomes zero in the limit of infinite r? If I understand correctly, an exponentially decaying f(r) is necessary to obtain a quasi-polynomial sample complexity in this manuscript.

3. Can the authors give demonstrations/ numerical experiments for the proposed efficient learning algorithms on a few physical examples? This will benefit the physics/chemistry community and attract more general audience, but in case that it could be quite time consuming, this could be optional and left for future work.

4. The hyperlinks for sections in the main text cannot be located correctly. For example, the "(see Proposition IV.3)" in page 9; "Section V" and "definition V.1", and "Section V B" in page 10; and possibly other places. The referenced sections can be in the Supplemental material; Better to check their consistence.

5. Some identical symbols are used for different physical quantities, for instance, the lattice length L is the same as the Lipschitz observable L; the support S and the entropy S, etc. Better to use different symbols.

6. Better to use "Quantum Wasserstein distance" in Definition II.2. It is actually the quantum Wasserstein distance of order one.

7. Refs. [DCGR19] and [DCGR20] have the same authors, the same titles, and the same journals and article numbers, but different years of publication? They might be replicative.

8. There are a few typos found in the supplemental material, for example, in Eq. (III.4), I think it should be “L – E(L)”, where the expectation E(L) was dropped. Also, the “3.\epsilon^2” in Propositions IV.3 and VII.1; The “- \nv \tau” in the last two lines of page 6 could be “- \nv r”; and possibly others. Better to double check the formulas and numbers.

Response to Reviewers: Efficient learning of ground &thermal states within phases of matter

We thank the reviewers for their details comments on the paper. We have attempted to address all the comments below or otherwise give justification for them.

REVIEWER COMMENTS

Reviewer 1

Weaknesses:

Even though the authors obtain improvement in the sample complexity, the performance with respect to other measures is still not favorable. This includes:

- *Exponential running time: the number of samples needed for the Gibbs state tomography of non-commuting systems is polylog(n), but the running time of the algorithm that processes the measurement data scales exponentially exp(n) with the number of qudits.*

Reply: the referee is correct that, at the time of submission, the scaling in computational complexity for general models was unfavorable. However, we stress that we reduced the problem of learning the Gibbs state in W_1 to that of learning the parameters of the underlying Hamiltonian (a problem usually called Hamiltonian learning). The recent months have seen significant progress on obtaining better solutions to this problem. Indeed, Tang et al recently obtained computationally and sample-efficient algorithms for exactly this problem after the submission, albeit in the regime with polynomial sample complexity, not logarithmic as in our case. Furthermore, older results by Tang et al. work in the logarithmic regime for noncommutative models at high temperatures. Thus, these results give us confidence that we will see log-time algorithms for the Hamiltonian learning problem in the near future, with direct implications to our result. As such, we believe this shortcoming of our method to be temporary.

- *A concurrent work [LHT+23] on learning thermal phases (problem (2) above) obtains an unconditional result with respect to the prior distribution over the samples, but this manuscript still requires this distribution to satisfy the anti-concentration over the local marginals which seems hard to relax in the current approach as explained in Remark IV.7 of SI.*

Reply: The referee is correct. But we stress that, although we do have stronger assumptions, we also get stronger recovery guarantees than [LHT+23]. Whereas they only obtain a good recovery on average over the parameter space, we obtain uniformly good recovery over all parameters. Thus, it is difficult to directly compare both results.

Comments for the authors:

1. *The result of Anshu [Anshu] on commuting Hamiltonians is repeatedly used in the manuscript. But given that [Anshu] is an unpublished note that hasn't been refereed, I'd recommend reproducing the relevant part of this result, at least minimally, in the supplementary notes.*

Reply: The work from [Anshu] has now been reproduced in the final section of the appendix. In this section, we have clearly stated that the work is not novel, and is a reproduction of [Anshu].

2. *On page 5: You mentioned "Such observables can arise if we e.g. evolve a strictly local observable by a Hamiltonian with algebraically decaying interactions" as a reason for considering quasi-local observables, but it seems your techniques in the manuscript rely on locality of the Hamiltonian terms (as opposed to algebraically decaying tails).*

   Reply: We believe that some confusion might have arisen from using the word Hamiltonian in two different contexts. First, there is the family of Hamiltonians associated with the Gibbs state. For these, we do need to enforce strict locality for our results to hold, at least in their current version. The second question is what observables we can recover by having a W_1 norm bound. The point of this example was to emphasize that we can indeed recover information that goes beyond few-body marginals and that this can include physically motivated examples. For instance, if we evolve the state by a Hamiltonian with long tails (i.e. algebraically decaying correlations), but note that this Hamiltonian is unrelated to the Hamiltonian we use to describe our family of states.

3. *If you were to sacrifice polylog(n) sample complexity to get poly(n) instead, does the result of [AAKS21] (without assuming the Markov property) allow you to improve the error eps to $1/eps^2$ instead of your current $1/eps^4$ scaling? For limited number of qudits n, a quadratic improvement in the accuracy epsilon may be more important in practice than an improved dependency on n.*

   Reply: Yes, if we are happy with a polynomial scaling in the sample complexity, then we can combine our results with those of [AAKS21]. Note that the results of [AAKS21] allow us to learn the parameters in ℓ_2 norm up to ϵ with $\text{poly}(n)\epsilon^{-2}$ samples. We can then convert this into a ℓ_1, use Corollary III.4 and estimate the W_1 distance up to ϵ. Note that this will still give a polynomial advantage when compared to an application of Pinker's inequality.

4. *In Definition II.2, the Wasserstein distance between two n qubit quantum states $\rho 1$, $\rho 2$ should be between $\rho 0$, $\rho 1$.*

   Reply: This has now been corrected.

5. *In Theorem II.3, you should mention that the assumption of decay of correlations is still needed for the (ii) case.*

   Reply: This has been added.

6. *The $\mathcal{S}$ versus S in Eq. II.9 may confuse the readers, especially since S is also reserved for the von Neumann entropy and also supp(O_i) later on.*

   Reply: We have changed the notation from S and R to $\mathcal{S}$ and $\mathcal{R}$ respectively.

7. *More generally, the notions introduced below II.9 are hard to unpack and refer to future Lemma III.5 that has not been discussed. I had to read this multiple times and go back to this part after reading the rest of the manuscript to grasp its content. I suggest modifying this part and also include explanations to aid the reader.*

   Reply: We have given a longer explanation of the notation, and have brought some additional equations from the Supplementary Material into this section to clarify the terms.

8. *In Corollary II.4. and also prior to that, the authors refer to the 'uniform' decay of correlations and 'uniform' Markov condition without specifying what they mean by uniform. I suggest you should introduce this notion when defining the exponential decay of correlations in Eq. II.6. add definitions and explain uniformity.*

   Reply: we have added a sentence in section II.A (where the first use of "uniform" occurs) explaining its meaning. Essentially uniform in this context implies the constants are independent of system size.

9. *It would also be helpful if the definition of the Markov property is given earlier on. The authors define CMI on Page 7, which seems to be the place to also define the decay of CMI (potentially even moved to earlier in the main).*
   Reply: Full definitions of CMU and clustering are given earlier in both the main text and the appendix.

10. *As mentioned before, the notation for entropy S(A) and region $S(r_s)$ is conflicting in Corollary II.4. For instance, what does $S = A$ refer to below the equation? The writing of this part needs adjustments.*

    Reply: We have adjusted the notation. Now we no longer use S to refer to a region, but only R. This should clear all confusion.

11. *Below Corollary II.4, it is mentioned that learning x up to ℓ_∞ results in a multiplicative error in estimating the entropic quantities. It would be helpful to specify the scaling of this error with epsilon and n.*

    Reply: We added the following sentence clarifying the scaling: "Furthermore, we remark that for many classes of Gibbs states, including high-temperature and commuting models [**? ? ?**], we can obtain such an estimate with $\mathcal{O}(\epsilon^{-2}\text{polylog}(n))$ samples, and it remains an important open problem to establish such a result in general for quantum models."

12. *In Figure 2, the notation seems to have changed from $|_{\mathcal{S}}(r)$ to $|_A(r)$.*

    Reply: This has now been remade with the new notation.

13. *On page 12, there is a list of "Examples of classes of commuting states that satisfy exponential decay of correlations" but not all items are commuting.*

    Reply: we have removed the work "commuting" — this was intended to be a more generic list which contains commuting Hamiltonians as a specific example. Furthermore, we added a comment on the implications of recent results of Tang et al on Hamiltonian learning for our work.

14. *The same page, item 2, should be above a threshold temperature.*

    Reply: We have changed this.

15. *In Proposition III.3 in the SI, where do you use the uniform assumption for the decay of correlations?*

    Reply: we use it twice: first, to invoke Lemma III.2 after Eq. III.8 and also to control the covariance in the next estimate. Note the phrase "where the second line follows from the condition of decay of correlations after it".

16. *It would be helpful to mention in Proposition III.12 that the bound on W_1 is obtained assuming decay of correlations and Markov property of sigma(beta, x).*

    Reply: This is now stated explicitly in the Proposition statement.

17. *Can you explain in more detail how in the SI, Eq. (III.27) is obtained from Eq. (III.26)?*

    Reply: We have added intermediary steps for the derivation by expanding the relative entropy.

18. *It would be good to emphasize that regions ABC in Def III.9 may not partition the whole lattice as this seems to have been used in deriving Eq. (III.34) and below it.*

    Reply: We added the sentence " We emphasize that ABC can be a proper subset of the lattice." after the definition to emphasize this.

19. *Two equations below Eq. (III.35), the RHS should be D(sigma(beta, x), sigma(beta, hatx)) instead of D(rho, sigma(beta, x)).*

    Reply: Indeed, we have corrected this now.

20. *Can you add more details where/why the ϵ^4 dependency shows up in Corollary III.13?*

    Reply: we added an explanation in the remark below.

Reviewer 2

1. *Can the authors comment on the possibility of learning observables in the presence of multiple phases? [HKT+22] numerically demonstrated that this is possible when there is data in all phases. This may be difficult to prove but may be something worth mentioning, at least as a future direction.*

   Reply: In the case of multiple phases, our methods would work in the bulk, but fail at the boundary. We have added a discussion of this in both the main article and in the appendix. We have also added a remark in the appendix.

2. *The authors emphasized that they obtain "point-wise guarantees" or their result "applies to the worst case", as compared to [HKT+22] that only gives average guarantees. I think this comparison is a little unfair unless the authors provide the context that the training data come from a distribution that anti-concentrates (as pointed out at the end of Section I, but I think this should also be mentioned in the abstract and other places where this advantage is claimed).*

   Reply: It is our understanding that [HKT+22] requires that their samples be drawn uniformly in parameter space, hence their assumptions on the distribution the samples are drawn from are stronger than ours (which just requires anticoncentration). See Lemma 3 section F.3 of the arXiv version, https://arxiv.org/abs/2106.12627 as a reference.

3. The authors state that certain problems pertaining phases of matter are "computationally intractable in general", and then say that having access to data can "conceivably speed up these computations". This seems not very relevant to the present scenario, because the works cited here seem all to be about the difficulty of identifying phase boundary and generalizing to the thermodynamic limit, neither of which was addressed in this paper.

   Reply: We assume the reviewer is talking about references [Amb14, WB21, BCGW22].

Amb14: is concerned with showing that measuring local observables on low energy states is computationally intractable, and so we should not expect an algorithm which takes a classical description of a Hamiltonian to be able to predict local observables of low energy states in general. However, the family of Hamiltonians Amb14 proves this for are extremely perverse.

WB21: is concerned with a similar task Amb14, except the family of Hamiltonians considered is now a translationally invariant system on a lattice. The problem studied is then to learn the expectation values of observables in parameter space. Since the order parameter of a phase transition is a local observable, the hardness of predicting local observables implies the hardness of predicting the phase boundary.

BCGW22: shows that predicting the expectation value of local observables of Gibbs states is computationally intractable for $\beta = O(1)$ temperatures.

4. *The notion of a Lipschitz observable features prominently in this paper. I think the authors should provide more intuition about this concept beyond the one sentence after Definition II.1.*

   Reply: We added the following paragraph below the definition: "Lipschitz observables encompass a variety of physically motivated observables; indeed, for most physically motivated extensive quantities, it is not the case that flipping one qubit should substantially change the value of the expectation value. For instance, if we consider the energy with respect to a local Hamiltonian on a regular lattice, then acting with a unitary on one of the qubits can only change the energy by a fixed amount. This is in contrast with arbitrary observables, where one local change can lead to drastic changes in the expectation value. Thus, in some sense, Lipschitz observables capture the notion that for well-defined physical quantities, a local operation should only have a limited effect."

5. *In the sentence after Eq. (II.4), the authors state that the estimation is accurate "up to a multiplicative error". It may be more accurate to say the error is relative to the Lipschitz norm of O. By multiplicative one would imagine it is relative to the expectation value of O.* **Reply:** We have rephrased that paragraph and explained why this scaling of the error often leads to a multiplicative error. **Reply:** I spotted one occurrence of us forgetting the β^{-1} term in Anurag's algorithm. It would be good to double-check if there are not others.

6. *In Eq. (II.8), I think it is worthwhile to explicitly state the temperature dependence in some way. It seems to me that this equation would not hold for zero temperature even if there is exponential decay of correlation (consider a classical Hamiltonian corresponding to SAT, a little change in the parameter may result in totally different ground states, even if all ground states are product states).*

   Reply: We have clarified that the bound has a linear scaling with β and added it. Thus, for the case of SAT, where we expect that $\beta \sim n$ to approximate the ground state with Gibbs, the statement is vacuous even for a ℓ_1 norm difference of constant order, as the W_1 is at most n.

7. *The authors should introduce what uniform clustering and uniform Markov means in the main text. They are rather important for understanding the results.*

   Reply: We have added these definitions to the main text, before the statement of the main results.

8. *In the first paragraph in Section II.A.3, the authors write that "we make use of the continuity of the entropy functional with respect to the Wasserstein distance, mentioned in Equation (II.6)." Eq. (II.6) does not seem to be the correct equation.*

   Reply: We have remove this reference and instead replaced it with an explicit equation.

9. *The authors may wish to compare their result with the work "Reconstructing quantum states from local data" by Swingle and Kim. Their method allows sample efficient reconstruction of the ground state of a gapped local Hamiltonian, with error measured in trace distance. This means their method can also sample-efficiently estimate the entropy of any subregion without having an exponential dependence on the size of the region.*

   Reply: We have added a brief discussion in Section III.B.

10. *The first sentence in the proof ideas of Theorem II.5 is very confusing. Part of the confusion comes from the fact that xj is used both to represent a component of the parameter vector x (in the first paragraph of Section II.A) and a sample (in Theorem II.5). Different notations should be used for these two completely different things. The authors should also explain what an "enlargement" is. I also cannot understand xj and hj(xj) appearing at the same time. If xj here means a sample, then I have no idea what hj is. If xj here means a component, then I cannot see how there can be an intersection between a set of components with* $[x - eps, x + eps]^m$.

    Reply: This has now been changed. Subscript now refers to samples and superscript now refers to the terms describing the Hamiltonian.

11. *There seems to be a typo in the equation in Definition II.6. I imagine that we would want the right-hand side of the equation, which gives an upper bound of the approximation error, to go to zero when r goes to infinity. But in this equation it seems that this does not happen due to the eta(S) term, which greatly puzzles me. From the bottom of Page 18 in the arXiv version I think instead of —S—f(r) + eta(S), it should be —S—f(r+eta(S))?*

    Reply: We have written out and explicit form of decay for the function f in the definition (exponential decay) and removed the η term. This additional parameter was only used in one of our derivations and likely overcomplicated the presentation, as it did not make much sense to state this condition in full generality.

12. *In the main text there are many references to equations and theorems that are in the supplementary materials (basically whenever Section III is involved). Whenever there is a reference to the supplementary material it should be stated.*

    We have updated the references and equations accordingly.

Reviewer 3

1. *Besides the theorems of existence type and the corresponding constructive proofs presented in this manuscript, can the authors give simple comments on the lower bound of the sample complexity? Also, the computational complexity better to be discussed.*
   Reply: Regarding the sample complexity for learning Gibbs states: we believe that obtaining lower bounds for the sample complexity of learning states in the Wasserstein distance is an important open question left by our work and we hope to address it in the future. However,

it should be noted that RSF22 showed that even for product states a polynomial number of samples is required to get a good recovery in trace distance. Thus, if we focus on comparing to the trace distance, it is already clear that our techniques yield exponential speedups and we already comment on this in the introduction of the text. Regarding learning phases: to the best of our knowledge, no lower bounds are available in this setting and it is unclear what they should be like. We added a comment to the conclusion stating this is an important open problem. Regarding the computational complexity:

Reply: We have discussed the complexity of the phase-learning algorithm after the proof of the Theorem "Learning algorithm for quantum Gibbs States". Furthermore, we have discussed after Theorem II.5 how there is no straightforward connection between the computational complexity of simulating a family of Gibbs states and the computational complexity of learning them given access to samples. For instance, for classical models, it is possible to learn them efficiently for any fixed temperature. However, it is not expected to be easy to sample from certain (classical) Gibbs states at constant temperature.

2. *In the definition of the GALI condition (Definition II.6), is it sufficient that f(r) only becomes zero in the limit of infinite r? If I understand correctly, an exponentially decaying f(r) is necessary to obtain a quasi-polynomial sample complexity in this manuscript.*

   Reply: The authors thank the careful reviewer for spotting this unprecise statement. Indeed in order to get quasi-polynomial sample complexity, GALI must be satisfied with exponentially decaying f. We amended the Theorem in the main text accordingly, and left the most general statement in the appendix.

3. ✓*Can the authors give demonstrations/ numerical experiments for the proposed efficient learning algorithms on a few physical examples? This will benefit the physics/chemistry community and attract more general audience, but in case that it could be quite time consuming, this could be optional and left for future work.*

   Reply: Regrettably, none of the authors are good programmers and as such we feel that performing and checking a numerical experiment would be a significant time investment. We leave this for future work.

4. *The hyperlinks for sections in the main text cannot be located correctly. For example, the "(see Proposition IV.3)" in page 9; "Section V" and "definition V.1", and "Section V B" in page 10; and possibly other places. The referenced sections can be in the Supplemental material; Better to check their consistence.*

   Ok, do it.

5. *Some identical symbols are used for different physical quantities, for instance, the lattice length L is the same as the Lipschitz observable L; the support S and the entropy S, etc. Better to use different symbols.*

   Reply: We have changed the S to R in the text to avoid this confusion.

6. *Better to use "Quantum Wasserstein distance" in Definition II.2. It is actually the quantum Wasserstein distance of order one.*

   Reply: We agree. This has been changed as requested.

7. *Refs. [DCGR19] and [DCGR20] have the same authors, the same titles, and the same journals and article numbers, but different years of publication? They might be replicative.*

   Reply: Thank you for catching this error. It has now been corrected.

8. *There are a few typos found in the supplemental material, for example, in Eq. (III.4), I think it should be "L – E(L)", where the expectation E(L) was dropped. Also, the "3.ϵ^2" in Propositions IV.3 and VII.1; The "- \nv τ" in the last two lines of page 6 could be "- \nv r"; and possibly others. Better to double check the formulas and numbers.*

   Reply: III.4 has been corrected. The "." in Propositions IV.3 and VII.1 have been removed. Unfortunately, the authors were not able to understand what the reviewer was referring to regarding proposition IV.3.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

In the revised version, the authors have improved the exposition of their manuscript following the referee reports. They have also addressed issues in the definition of GALI condition, uniform clustering, relative errors, and notational inconsistencies.

My assessment of the conceptual novelty and relevance of the results remain mostly the same. Considering the revisions made and the feedback from other referees, I would recommend accepting the paper.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed most of my comments. There is one remaining issue: I mentioned that the authors should discuss the paper ``Reconstructing quantum states from local data'' by Swingle and Kim, which they responded by adding a paragraph at the end of Section III.B. But instead of citing this paper they cited another paper by Swingle and McGreevy [SM16], which deals with a different (though related) topic. I think this must have been a mistake.

Reviewer #3 (Remarks to the Author):

In general, my previous concerns have been safely removed in the revised manuscript and with the reply letter. Now I recommend the paper's publication on Nature Communications without further delay.

Just two additional minor suggestions (which could be revised in the proof stage):

1. In the first page of the SI, S is still used to refer to a region instead of R;
2. In Definition II.2, the "Quantum Wasserstein distance" was not used as I suggested, but this could be optional. Just point out the minor inconsistency between the revised main text and the reply to my Comment 6.

Second Response to Reviewers : Efficient learning of ground & thermal states within phases of matter

We thank the reviewers for their careful proofreading. We believe we have now addressed all of the relevant errors.

REVIEWER 2

1. *I mentioned that the authors should discuss the paper "Reconstructing quantum states from local data" by Swingle and Kim, which they responded by adding a paragraph at the end of Section III.B. But instead of citing this paper they cited another paper by Swingle and McGreevy [SM16], which deals with a different (though related) topic. I think this must have been a mistake.*

   This has now been changed – thank you for catching this!

REVIEWER 3

1. *In the first page of the SI, S is still used to refer to a region instead of R;*

   This has now been addressed.

2. *In Definition II.2, the "Quantum Wasserstein distance" was not used as I suggested, but this could be optional. Just point out the minor inconsistency between the revised main text and the reply to my Comment 6.*

   Thank you for catching this and apologies for the confusion. his has now been addressed and changed to Quantum Wasserstein Distance.