On the sample complexity of quantum Boltzmann machine learning

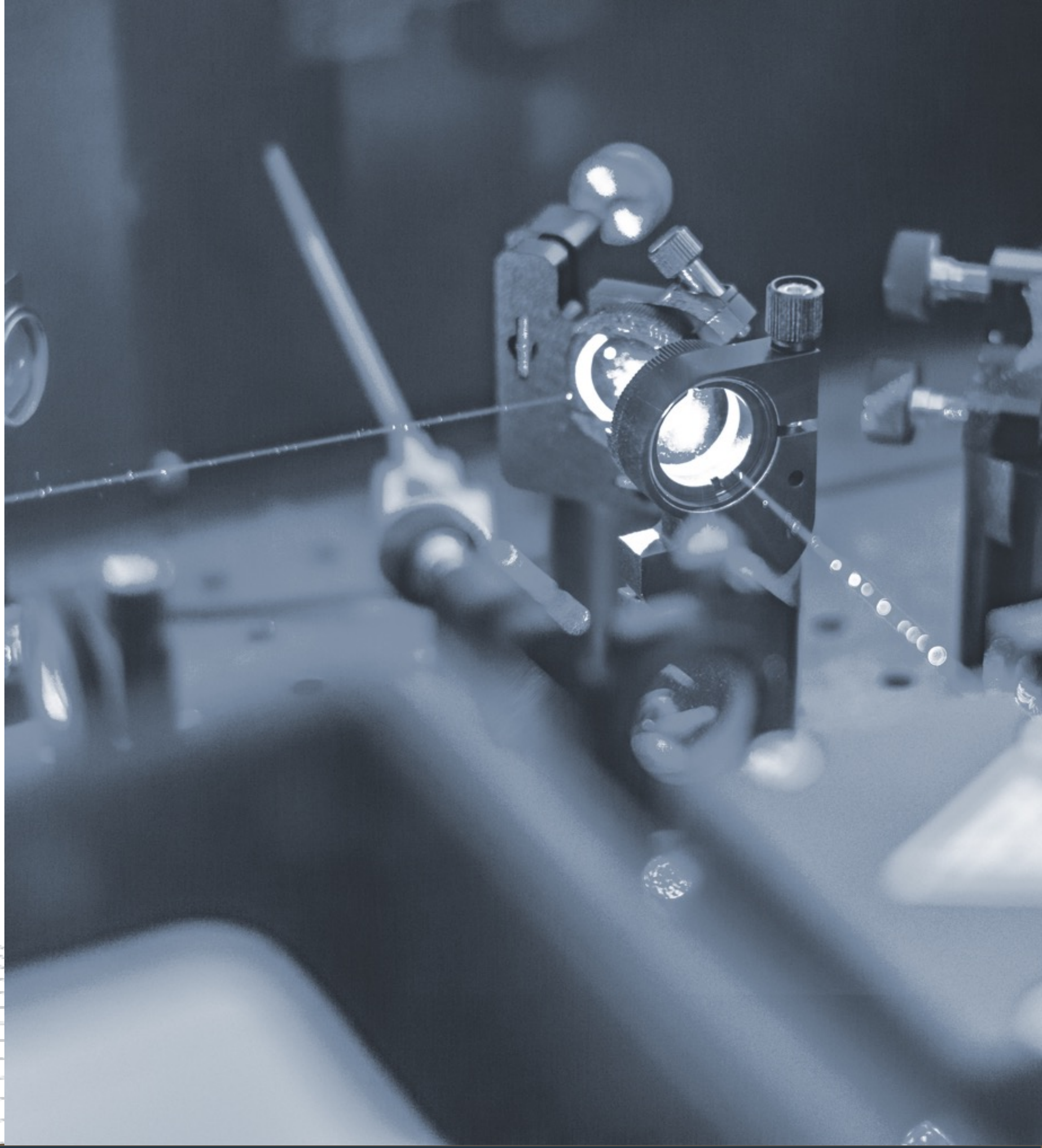
L. Coopmans and M. Benedetti [arXiv: 2306.14969]

Presented by:

Luuk Coopmans
21-11-2023

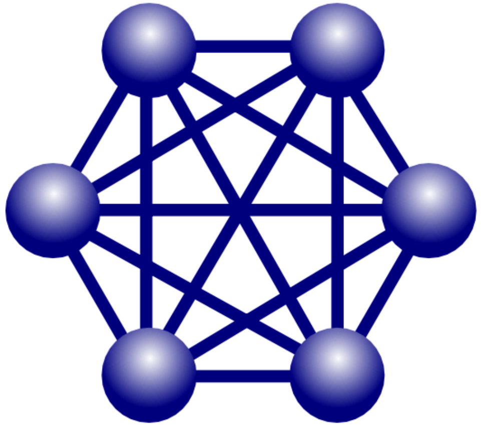
Outline of the Talk

- Introduction: what are QBMs?
- Main result: polynomial sample complexity
- Numerical verification
- Discussion and open problems



Introduction: what are quantum Boltzmann Machines

- A quantum Boltzmann machine (QBM) is a quantum machine learning model that can be used for generatively modelling classical and quantum data.

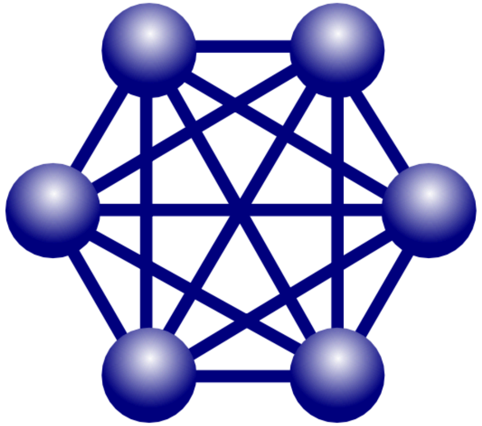


Introduction: what are quantum Boltzmann Machines

- A quantum Boltzmann machine (QBM) is a quantum machine learning model that can be used for generatively modelling classical and quantum data.

- $n-$qubit parameterised Gibbs state $\rho_\theta = \dfrac{e^{\mathscr{H}_\theta}}{\mathrm{Tr} e^{\mathscr{H}_\theta}}$,

- with ansatz/model Hamiltonian $\mathscr{H}_\theta = \sum_{i=1}^{m} \theta_i H_i$.



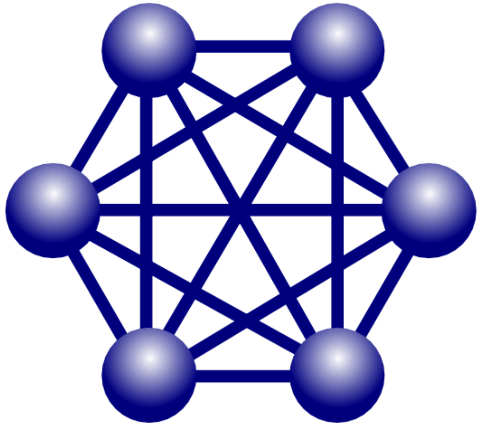
- Any set of (bounded) Hermitian operators $\{H_i\}_{i=1}^{m}$, e.g. Pauli operators.

Introduction: what are quantum Boltzmann Machines

- A quantum Boltzmann machine (QBM) is a quantum machine learning model that can be used for generatively modelling classical and quantum data.

- $n-$qubit parameterised Gibbs state $\rho_\theta = \frac{e^{\mathscr{H}_\theta}}{\mathrm{Tr} e^{\mathscr{H}_\theta}}$,

- with ansatz/model Hamiltonian $\mathscr{H}_\theta = \sum_{i=1}^{m} \theta_i H_i$.



- Any set of (bounded) Hermitian operators $\{H_i\}_{i=1}^{m}$, e.g. Pauli operators.

Note: focus here on fully visible QBMs

Introduction: goal and applications

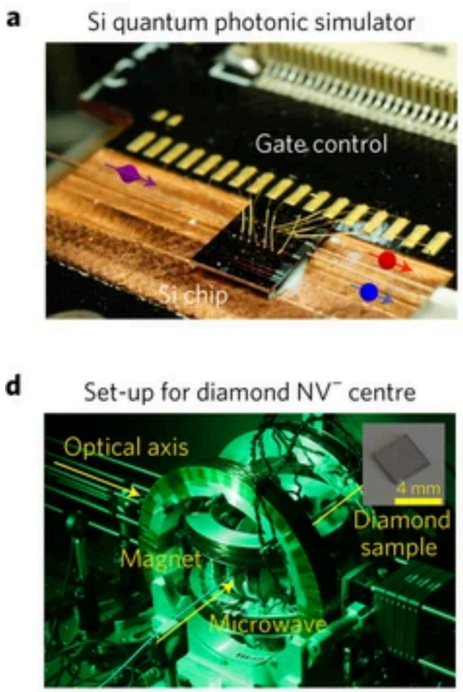
- Goal: optimise θ such that samples taken from the QBM $s' \sim \rho_\theta$ mimic samples from a given target dataset $\{s^\mu\}_{\mu=1}^M$.

-

Introduction: goal and applications

- Goal: optimise θ such that samples taken from the QBM $s' \sim \rho_\theta$ mimic samples from a given target dataset $\{s^\mu\}_{\mu=1}^M$.

- Applications:
  - Generative modelling of (binarised) classical data
    [Amin et al., 1601.02036], [Kappen, 1803.11278]
  - Learning physical models from experimental data
  - Learning the entanglement Hamiltonian
    [M.K. Joshi et al. arXiv 2306.00057 (2023), C. Kokail et al., PRL 127 (2021)]
  - Verification of quantum devices
    [J. Wang et al., Nature Physics 13 (2017)]
  - …



[J. Wang et al., Nature Physics 13 (2017)]

Introduction: goal and applications

- Goal: optimise θ such that samples taken from the QBM $s' \sim \rho_\theta$ mimic samples from a given target dataset $\{s^\mu\}_{\mu=1}^M$.

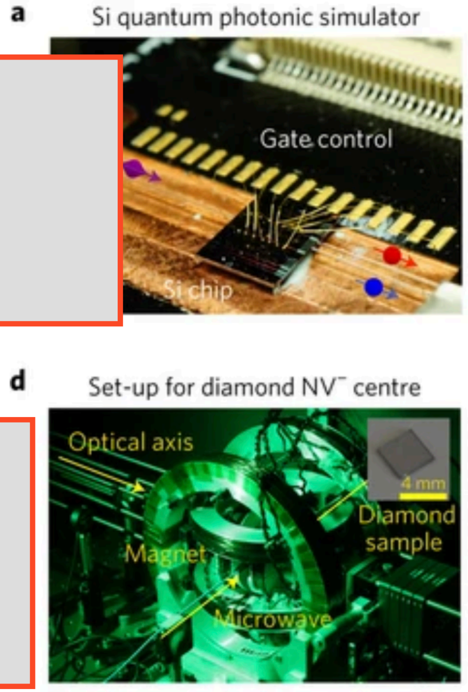
- Application

CAN QBMS BE TRAINED EFFICIENTLY?

  - Gene
    [Amin et al., 1601.02036], [Kappen, 1803.11278]
  - Learning physical models from experimental data
  - Learning tonian
    [M.K. Joshi et il et al., PR
  - Verificati
    [J. Wang et al
  - …

Sample complexity?

Time per sample?



[J. Wang et al., Nature Physics 13 (2017)]

Results: (Re)formulation of the learning problem

- Relative entropy loss function: $S(\eta||\rho_\theta) = \mathrm{Tr}(\eta \log \eta) - \mathrm{Tr}(\eta \log \rho_\theta)$, $\rho_\theta = \frac{e^{\mathcal{H}_\theta}}{\mathrm{Tr} e^{\mathcal{H}_\theta}}$

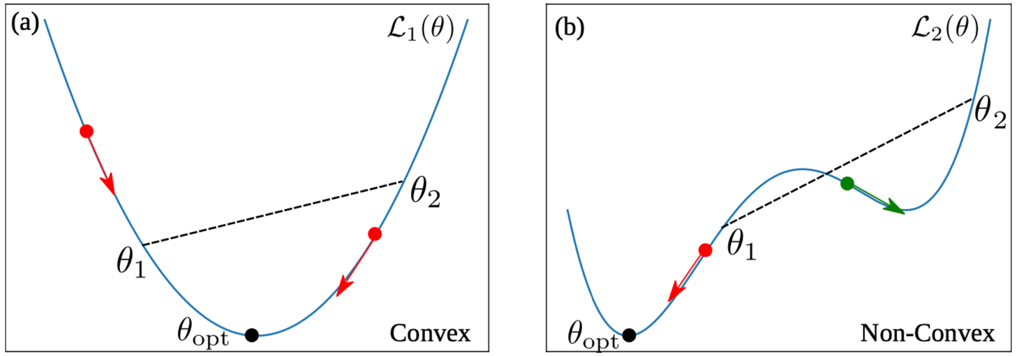
- Derivatives: $\frac{\partial S(\eta||\rho_\theta)}{\partial \theta_k} = \mathrm{Tr}[\rho_\theta H_k] - \mathrm{Tr}[\eta H_k]$.

Results: (Re)formulation of the learning problem

- Relative entropy <u>loss function</u>: $S(\eta||\rho_\theta) = \mathrm{Tr}(\eta\log\eta) - \mathrm{Tr}(\eta\log\rho_\theta)$, $\rho_\theta = \frac{e^{\mathscr{H}_\theta}}{\mathrm{Tr}e^{\mathscr{H}_\theta}}$

- Derivatives: $\frac{\partial S(\eta||\rho_\theta)}{\partial\theta_k} = \mathrm{Tr}[\rho_\theta H_k] - \mathrm{Tr}[\eta H_k]$.

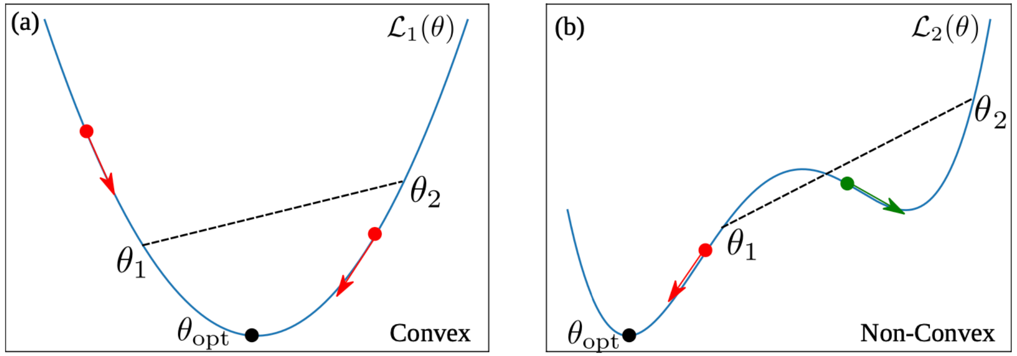
- Show that $S(\eta||\rho_\theta)$ is (strictly) <u>convex.</u>



Results: (Re)formulation of the learning problem

- Relative entropy loss function: $S(\eta||\rho_\theta) = \mathrm{Tr}(\eta \log \eta) - \mathrm{Tr}(\eta \log \rho_\theta)$, $\rho_\theta = \frac{e^{\mathscr{H}_\theta}}{\mathrm{Tr} e^{\mathscr{H}_\theta}}$
- Derivatives: $\frac{\partial S(\eta||\rho_\theta)}{\partial \theta_k} = \mathrm{Tr}[\rho_\theta H_k] - \mathrm{Tr}[\eta H_k]$.
- Show that $S(\eta||\rho_\theta)$ is (strictly) convex.
- => QBM trained for a θ^* such that:

$$|\mathrm{Tr}[\rho_{\theta*} H_k] - \mathrm{Tr}[\eta H_k]| \leq \epsilon, \forall k$$



Results: (Re)formulation of the learning problem

- Relative entropy <u>loss function</u>: $S(\eta||\rho_\theta) = \mathrm{Tr}(\eta \log \eta) - \mathrm{Tr}(\eta \log \rho_\theta)$, $\rho_\theta = \dfrac{e^{\mathcal{H}_\theta}}{\mathrm{Tr} e^{\mathcal{H}_\theta}}$

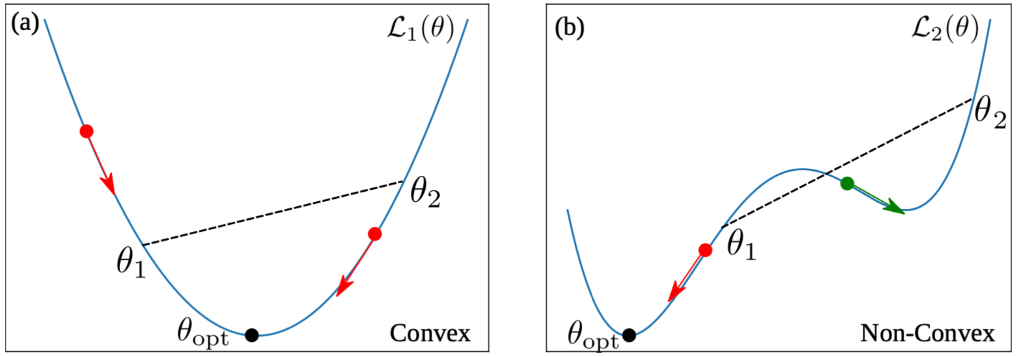
- Derivatives: $\dfrac{\partial S(\eta||\rho_\theta)}{\partial \theta_k} = \mathrm{Tr}[\rho_\theta H_k] - \mathrm{Tr}[\eta H_k]$.

- Show that $S(\eta||\rho_\theta)$ is (strictly) <u>convex.</u>

- => QBM trained for a θ^* such that:

$$|\mathrm{Tr}[\rho_{\theta*} H_k] - \mathrm{Tr}[\eta H_k]| \leq \epsilon, \forall k$$

Learning problem



Results: Polynomial sample complexity

- Optimize $S(\eta||\rho_\theta)$ with stochastic gradient descent.

Results: Polynomial sample complexity

- Optimize $S(\eta||\rho_\theta)$ with stochastic gradient descent.

Theorem 1: after $T = \dfrac{48\delta_0 m^2(\kappa^2+\xi^2)}{\epsilon^4}$ iterations, where $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\text{opt}}})$, we have solved the QBM learning problem to accuracy ϵ.

Each iteration requires $N \leq \mathcal{O}\left(\dfrac{1}{\kappa^4}\log\dfrac{m}{1-\lambda^{\frac{1}{T}}}\right)$ preparation of the Gibbs state ρ_θ.

Results: Polynomial sample complexity

- Optimize $S(\eta||\rho_\theta)$ with stochastic gradient descent.

Theorem 1: after $T = \frac{48\delta_0 m^2(\kappa^2 + \xi^2)}{\epsilon^4}$ iterations, where $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\mathrm{opt}}})$, we have solved the QBM learning problem to accuracy ϵ.

Each iteration requires $N \leq \mathcal{O}\left(\frac{1}{\kappa^4}\log\frac{m}{1-\lambda^{\frac{1}{T}}}\right)$ preparation of the Gibbs state ρ_θ.

- $m = \mathrm{Poly}(n)$ => Polynomial sample complexity!

Results: Polynomial sample complexity

- Optimize $S(\eta||\rho_\theta)$ with stochastic gradient descent.

Theorem 1: after $T = \dfrac{48\delta_0 m^2(\kappa^2 + \xi^2)}{\epsilon^4}$ iterations, where $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\text{opt}}})$, we have solved the QBM learning problem to accuracy ϵ.

Each iteration requires $N \leq \mathcal{O}\left(\dfrac{1}{\kappa^4}\log\dfrac{m}{1-\lambda^{\frac{1}{T}}}\right)$ preparation of the Gibbs state ρ_θ.

- $m = \text{Poly}(n)$ => Polynomial sample complexity!

No barren plateaus…

Results: reducing the sample complexity by pretraining

- Sample complexity depends on $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\text{opt}}})$.

Results: reducing the sample complexity by pretraining

- Sample complexity depends on $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\mathrm{opt}}})$.

Theorem 2: can reduce δ_0 by pretraining a sub-model defined by a subset of the operators $\{H_i\}_{i=1}^{\tilde{m}}$.

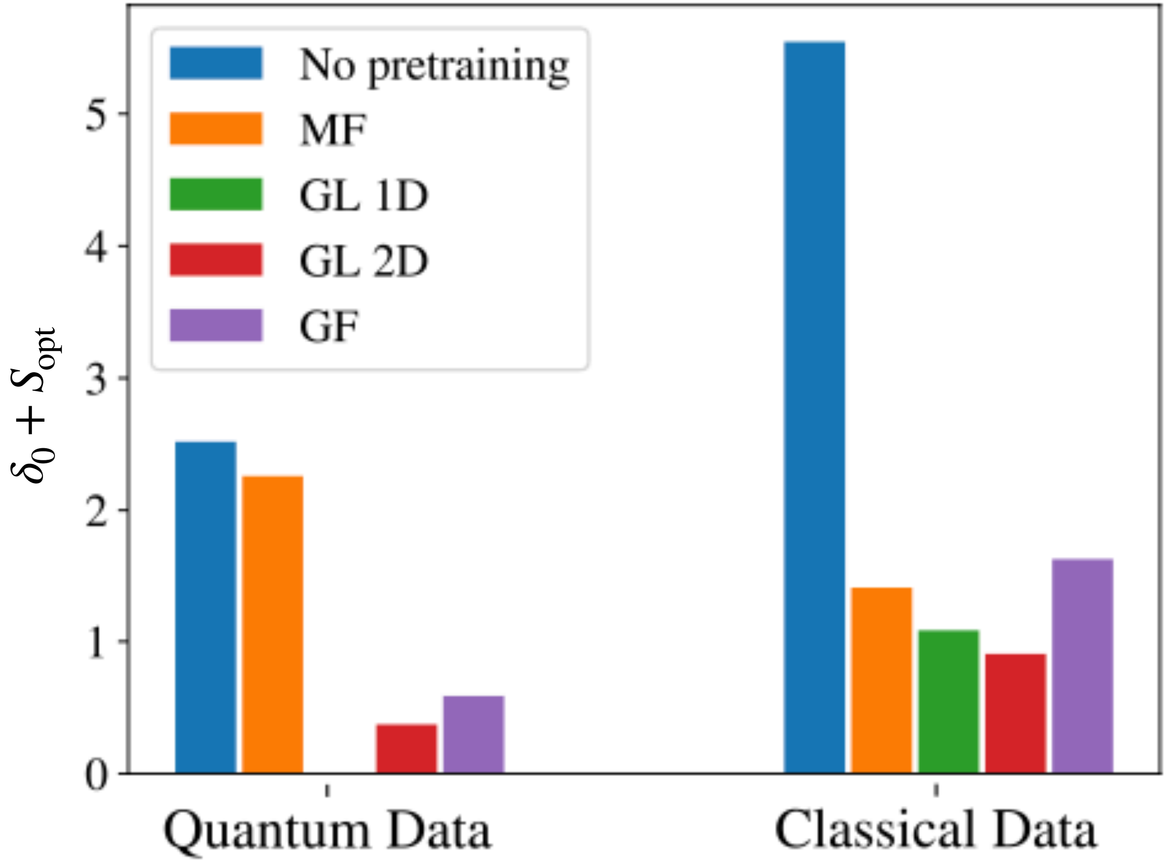
Results: reducing the sample complexity by pretraining

- Sample complexity depends on $\delta_0 = S(\eta||\rho_{\theta^0}) - S(\eta||\rho_{\theta^{\text{opt}}})$.

Theorem 2: can reduce δ_0 by pretraining a sub-model defined by a subset of the operators $\{H_i\}_{i=1}^{\tilde{m}}$.

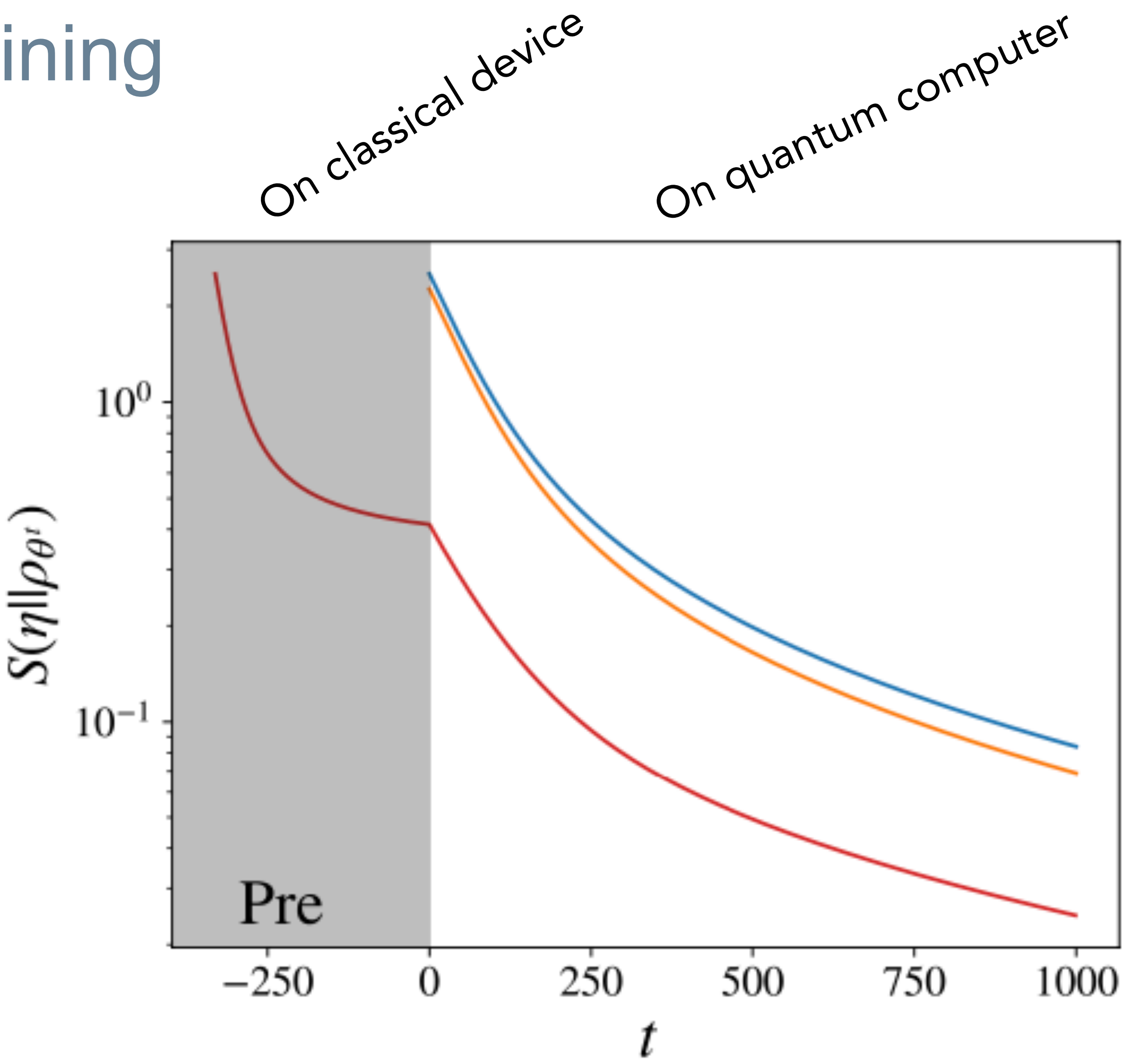
- Note δ_0 defined with respect to the larger model
- Examples: quantum mean-field (MF), Gaussian-Fermionic (GF), geometrically local (GL)

Numerical verification: pretraining

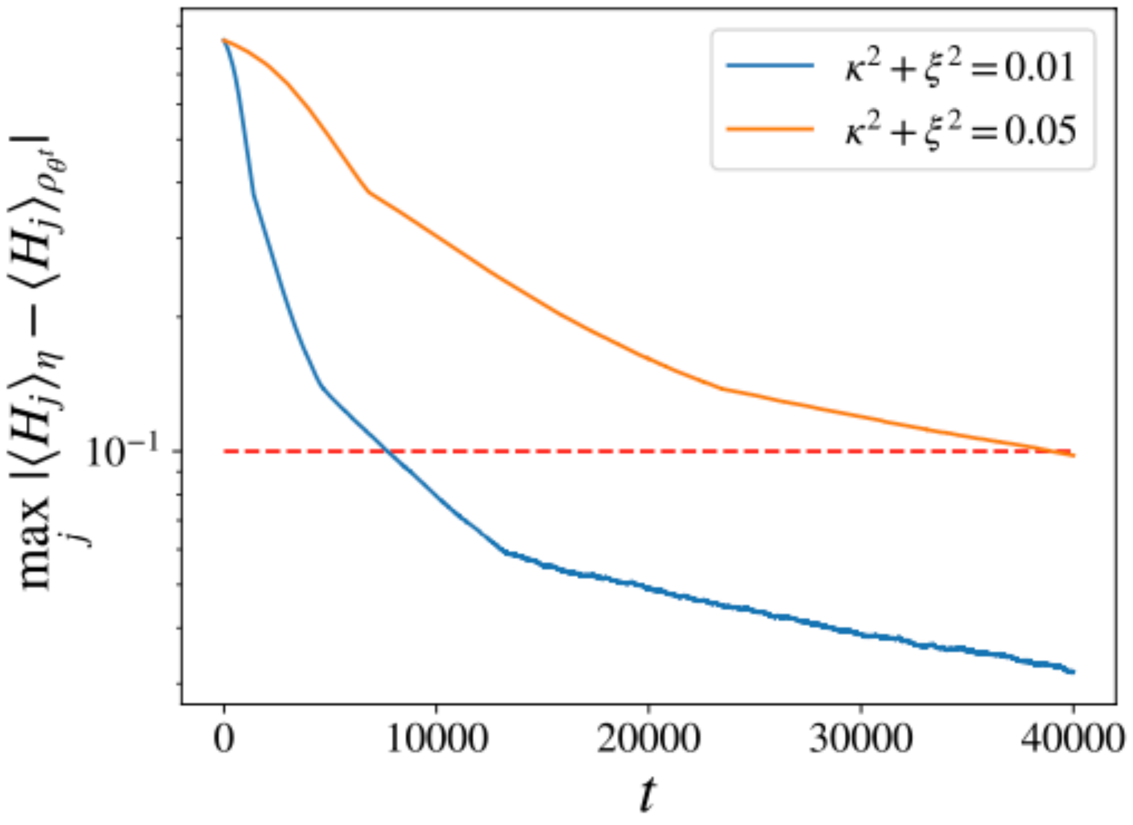


1D Heisenberg model

Binary time series data



Numerical verification: verifying sampling bound



Bound: $T = \mathcal{O}(10^9)$

Summary and discussion

- We have reformulated the QBM learning problem and proved that it can be solved with polynomial sample complexity.

- Pretraining on a subset of the parameters can only improve the sample complexity.

Summary and discussion

- We have reformulated the QBM learning problem and proved that it can be solved with polynomial sample complexity.
- Pretraining on a subset of the parameters can only improve the sample complexity.
- Time complexity: combine with Gibbs sampler of your choice (classical or quantum).

Summary and discussion

- We have reformulated the QBM learning problem and proved that it can be solved with polynomial sample complexity.

- Pretraining on a subset of the parameters can only improve the sample complexity.

- Time complexity: combine with Gibbs sampler of your choice (classical or quantum).

- Speedups in quantum Gibbs sampling translate to QBM learning.

  1) Generic case: quadratic speedup but possibly exponential time
  2) Efficient Gibbs samplers for specific cases??
     [O. Watts et al., 2310.07774], [Chen et al., 2311.09207]
  3) Resources for experimental demonstration??

Thank you for your attention!

Luuk Coopmans
QTML, 21-11-2023

See paper:
arXiv: 2306.14969