Structured quantum learning via em algorithm for Boltzmann machines

arXiv:2507.21569

Masahito Hayashi

School of Data Science, The Chinese University of Hong Kong, Shenzhen

Shenzhen International Quantum Academy (SIQA),

Graduate School of Mathematics, Nagoya University

Joint work with Takeshi Kimura and Kohtaro Kato

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

深圳国际量子研究院
Shenzhen International Quantum Academy

NAGOYA UNIVERSITY

Contents

- Problems in current quantum machine learning
- Classical Machine Learning Lessons
- Idea for proposed method
- Semi-quantum Boltzmann machines and Quantum em algorithm

Problems in current quantum machine learning

(1) To find the good parameter to characterize the model, we use gradient descent.

Barren plateau problem: Gradient often vanishes before achieving the good solution.

(2) Fully visible Boltzmann machine does not have this problem. But, this model is too simple (Coopmans & Benedetti (2024)).

Classical Machine Learning Lessons: Need for Structural Optimization.

(1) To handle complex information, it is necessary to use models with hidden layers. In fact, many papers point out that models without hidden layers lack expressive power.

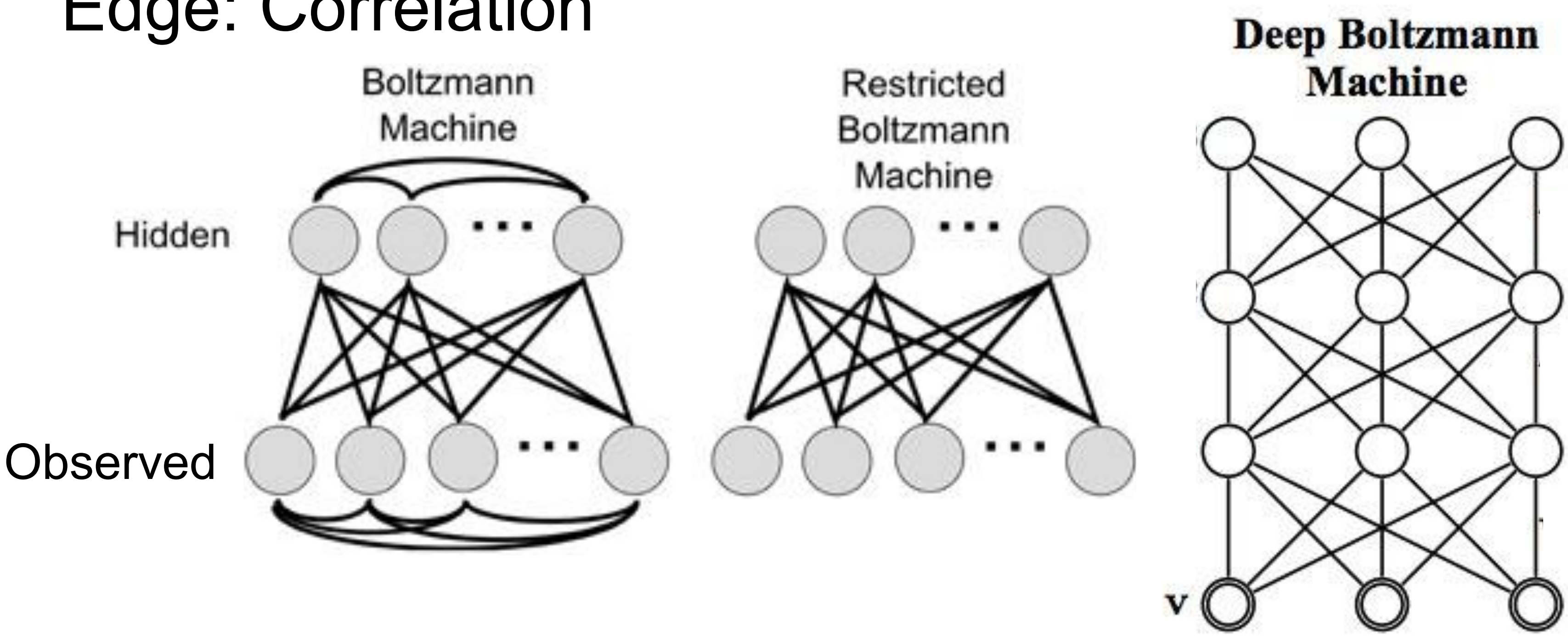
Boltzmann machine

Unsupervised learning

We need to estimate the hidden structure from observed data.

Node: Data point

Edge: Correlation



Classical Machine Learning Lessons: Need for Structural Optimization

(2) In such models, vanishing gradient problem occurs, preventing accurate parameter determination.

(3) To solve this problem, the current practice is to decompose the target optimization problem into easier optimization subproblems that reflect the hidden layer structure.

This is not a minor improvement like refining the gradient calculation.

Classical Machine Learning Lessons: Need for Structural Optimization

(4) Many of these approaches are considered variations of the standard EM algorithm

for cases with incomplete data. Although the EM algorithm and its variants can effectively avoid the vanishing gradient problem and they easily escape saddle points, the risk of converging to a local optimum generally exists in the classical case, but these algorithms still work well empirically.

Existing major studies on quantum ML

(1) Coopmans & Benedetti (Communications Physics, 2024): Fully visible quantum Boltzmann machine. Proposed quantum algorithm outperforms classical case. It lacks expressive power.

(2) Demidik *etal* (Communications Physics, 2025): It proposes semi-quantum Restricted Boltzmann machine. But, it employs gradient descent so that it cannot resolve Barren plateau problem

EM algorithm vs em algorithm

The concept used in EM algorithm cannot be extended to quantum case.

EM (Expectation-Maximization) algorithm	**em (e/m-projection algorithm) algorithm**
General method with hidden variable (Layered Boltzmann machine Hidden Markov, Missing Data)	General structure in information geometry by Amari Bregman divergence structure (Applicable beyond hidden variables, including rate-distortion theory)
E step (Expectation) M step (Maximization)	e-Projection along exponential family m-Projection along mixture family
Maximizes likelihood by iteratively handling hidden variables.	Minimize divergence between exponential family & mixture family
Quantum extension is difficult because of the difficulty of quantum extension of likelihood.	Quantum extension is possible based on quantum information geometry

Proposed method: semi-quantum Boltzmann machine

(1)Visible layer: classical system

(2)Hidden layer: quantum system, which has more expressive power.

(3)Applying the em algorithm by alternating e-step and m-step.

(4)e-step can be trivially done

(5)Most difficult part is m-step.

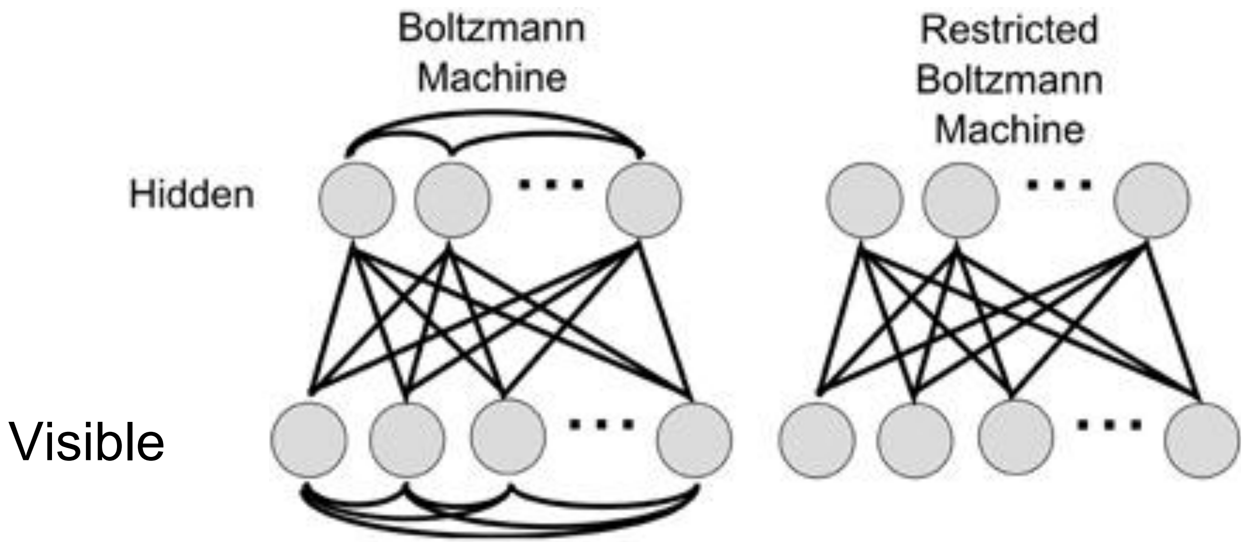
m-step in semi-quantum Boltzmann machine

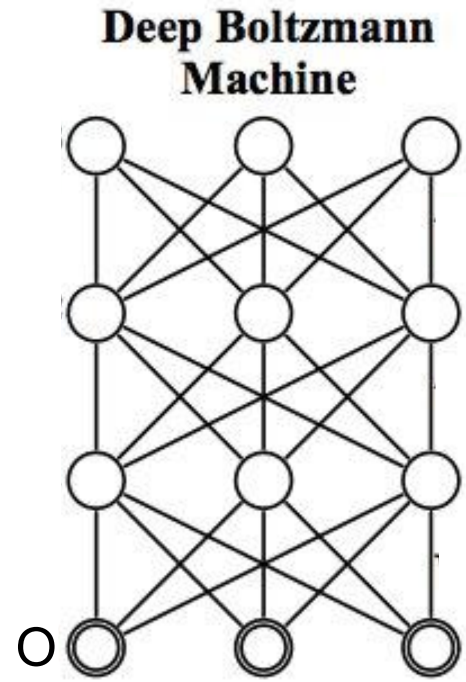
(1) m-step is equivalent to solving fully visible sq Boltzmann machine.

(2) In fully visible quantum Boltzmann machine the parameter can be determined via gradient descent. It can be done classically or quantumly. Coopmans & Benedetti's efficient method can be applied when QC is available.

Classical Boltzmann machine

Parametrization





$$\vec{\theta} = ((\theta^{k,j})_{(k,j)\in E}, (\theta^{k})_{k\in V}, (\theta^{j})_{j\in H})$$

$$\phi(\vec{\theta}) = \log \sum_{x_k, y_j \in \{-1,1\}} \exp\Big(\sum_{(k,j)\in E} \theta^{k,j} x_k y_j + \sum_{k\in V} \theta^{k} x_k + \sum_{j\in H} \theta^{j} y_j \Big)$$

Convex function

$$P_{XY,\vec{\theta}}(x,y) := \exp\Big(\sum_{(k,j)\in E} \theta^{k,j} x_k y_j + \sum_{k\in V} \theta^{k} x_k + \sum_{j\in H} \theta^{j} y_j - \phi(\vec{\theta}) \Big)$$

Which parameter explains observed data in the best way?

Parameters of hidden model explains visible structure

$$P_{XY,\vec{\theta}}(x,y) \to P_{X,\vec{\theta}}(x) := \sum_y P_{XY,\vec{\theta}}(x,y)$$

Given an observed empirical distribution $P_X(x)$

we find the closest distribution $P_{X,\vec{\theta}}(x)$

$$\arg\min_{\vec{\theta}} D(P_X \| P_{X,\vec{\theta}})$$

Maximum likelihood estimator

$D(P_X \| P_{X,\vec{\theta}})$ is not convex for $\vec{\theta}$ in general.

This minimization is very difficult!

em-algorithm

Amari

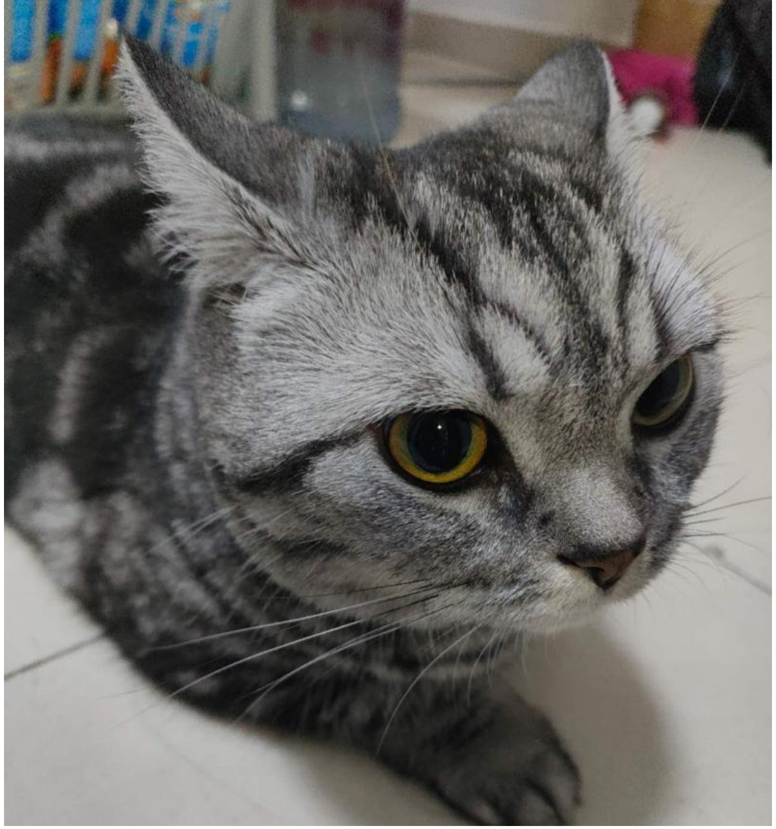
em-algorithm is a mathematical generalization of EM algorithm.

It is an alternative application of e-projection and m-projection, which are basic concepts of information geometry.

$$D(P_X \| P_{X,\vec{\theta}}) = \min_{P_{Y|X}} D(P_{Y|X} \times P_X \| P_{XY,\vec{\theta}})$$

Minimization is converted into alternating minimization.

$$\min_{\theta} D(P_X \| P_{X,\vec{\theta}}) = \min_{\theta, P_{Y|X}} D(P_{Y|X} \times P_X \| P_{XY,\vec{\theta}})$$



QUANTUM EXTENSION

Mixture family

$\mathcal{H}$: Hilbert space

$\mathcal{S}(\mathcal{H})$: Set of densities over $\mathcal{H}$

$X_1, \ldots., X_k$: linearly independent obervables on $\mathcal{H}$

Mixture family

$$\mathcal{M}(\vec{a}) := \{\rho \in \mathcal{S}(\mathcal{H}) \mid \mathrm{Tr}\rho X_i = a_i, i = 1, \ldots, k\}$$

$$\vec{a} := (a_1, \ldots, a_k)$$

Exponential family

$X_1, ..., X_k$: linearly independent observables on $\mathcal{H}$

$$\vec{\theta} := (\theta^1, ..., \theta^k)$$

Convex function

$$\phi(\vec{\theta}) := \log \mathrm{Tr} \exp(\log \rho + \sum_{i=1}^{k} \theta^i X_i)$$

$$\rho_{\vec{\theta}} := \exp(\log \rho + \sum_{i=1}^{k} \theta^i X_i - \phi(\vec{\theta}))$$

Exponential family

$$\mathcal{E}(P) := \{\rho_{\vec{\theta}} \in \mathcal{S}(\mathcal{H})\}$$

Pythagorean theorem

Kullback-Leibler divergence

$$D(\sigma\|\rho) := \mathrm{Tr}\,\sigma(\log\sigma - \log\rho)$$

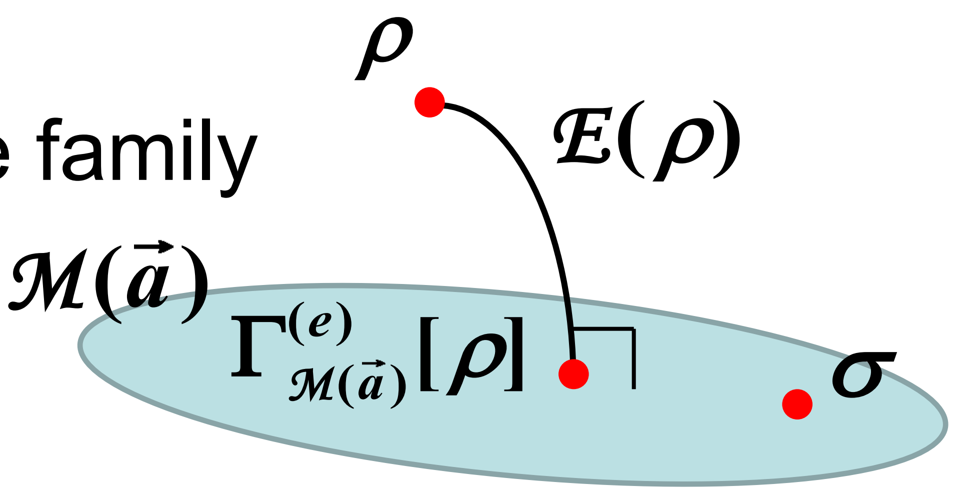
e-projection

$$\Gamma^{(e)}_{\mathcal{M}(\vec{a})}[\rho] := \mathop{\arg\min}_{\sigma\in\mathcal{M}(\vec{a})} D(\sigma\|\rho)$$

When $\sigma \in \mathcal{M}(\vec{a})$,

$$D(\sigma\|\rho) = D(\sigma\|\Gamma^{(e)}_{\mathcal{M}(\vec{a})}[\rho]) + D(\Gamma^{(e)}_{\mathcal{M}(\vec{a})}[\rho]\|\rho)$$

Mixture family



Pythagorean theorem

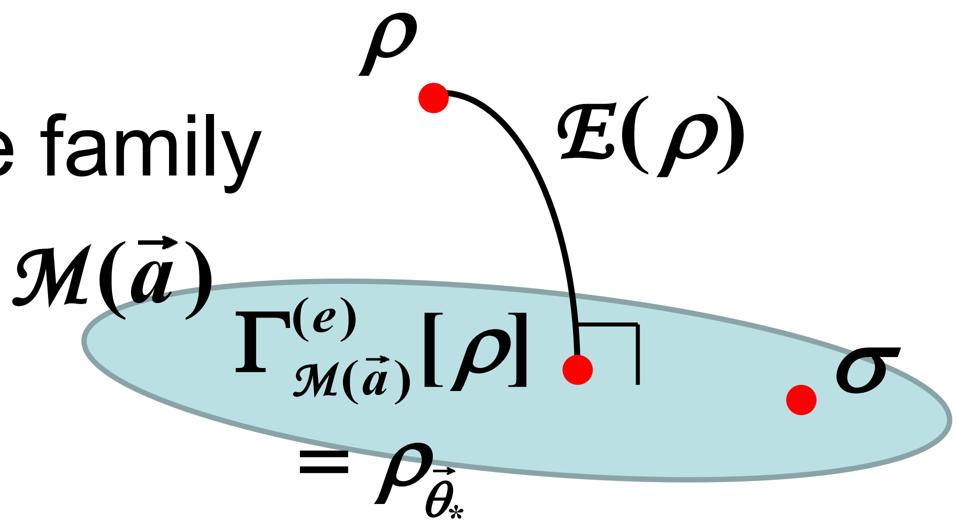
e-projection is calculated as

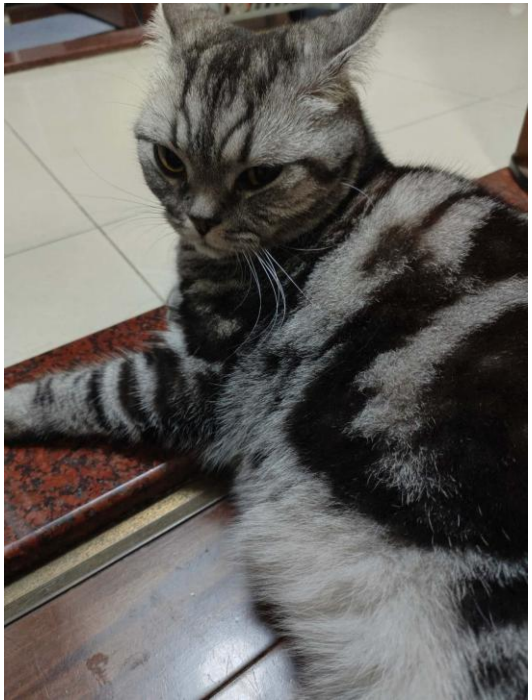
$$\Gamma^{(e)}_{\mathcal{M}(\vec{a})}[\rho] := \mathop{\arg\min}_{\sigma \in \mathcal{M}(\vec{a})} D(\sigma \| \rho) = \rho_{\vec{\theta}_*}$$

where $$\vec{\theta}_* := \mathop{\arg\min}_{\vec{\theta}} \phi(\vec{\theta}) - \sum_{i=1}^{k} \theta^i a_i$$

convex minimization!

Mixture family





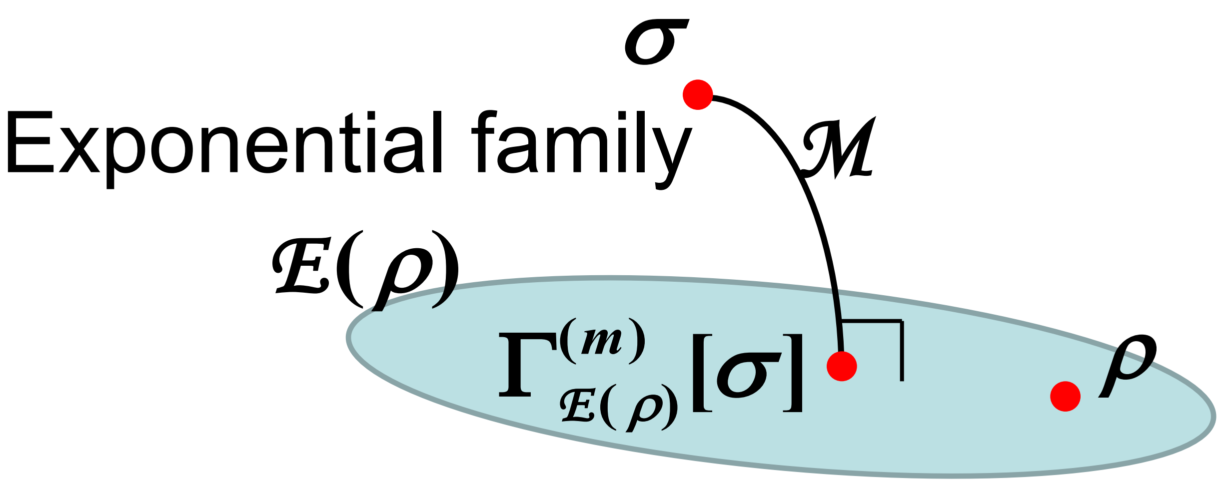
Pythagorean theorem

m-projection

$$\Gamma^{(m)}_{\mathcal{E}(\rho)}[\sigma] := \underset{\rho' \in \mathcal{E}(\rho)}{\arg\min}\, D(\sigma \| \rho')$$

When $\rho' \in \mathcal{E}(\rho)$,

$$D(\sigma \| \rho) = D(\sigma \| \Gamma^{(m)}_{\mathcal{E}(\rho)}[\sigma]) + D(\Gamma^{(m)}_{\mathcal{E}(\rho)}[\sigma] \| \rho)$$



Pythagorean theorem

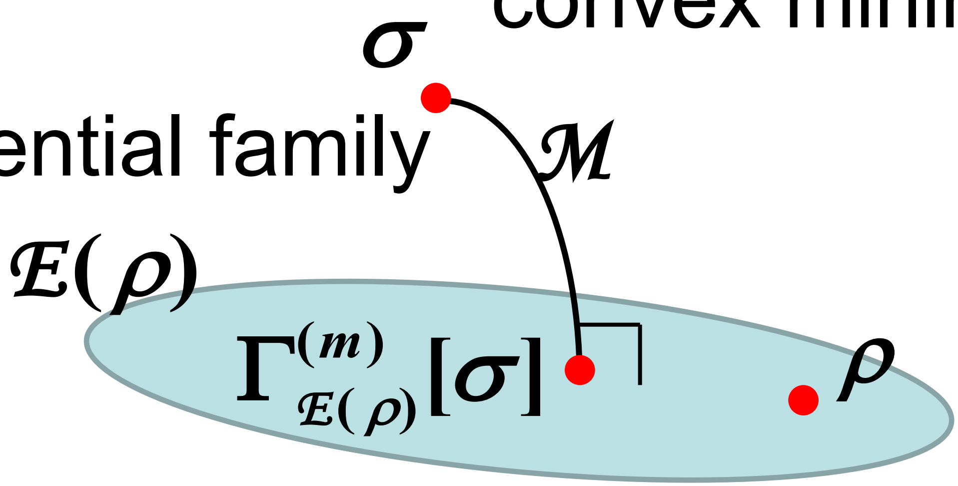
m-projection

$$\Gamma^{(m)}_{\mathcal{E}(\rho)}[\sigma] := \arg\min_{\rho' \in \mathcal{E}(\rho)} D(\sigma \| \rho') = \rho_{\vec{\theta}_{**}}$$

where
$$\vec{\theta}_{**} := \arg\min_{\vec{\theta}} \phi(\vec{\theta}) - \sum_{i=1}^{k} \theta^i a_i$$

convex minimization!

Exponential family



em-algorithm

$$\min_{\sigma \in \mathcal{E}} \min_{\rho \in \mathcal{M}} D(\rho \| \sigma)$$

em-algorithm is an iterative algorithm.

We set initial point $\sigma_{(1)} \in \mathcal{E}$

m-step $\rho_{(t+1)} := \arg\min_{\rho \in \mathcal{M}} D(\rho \| \sigma_{(t)})$

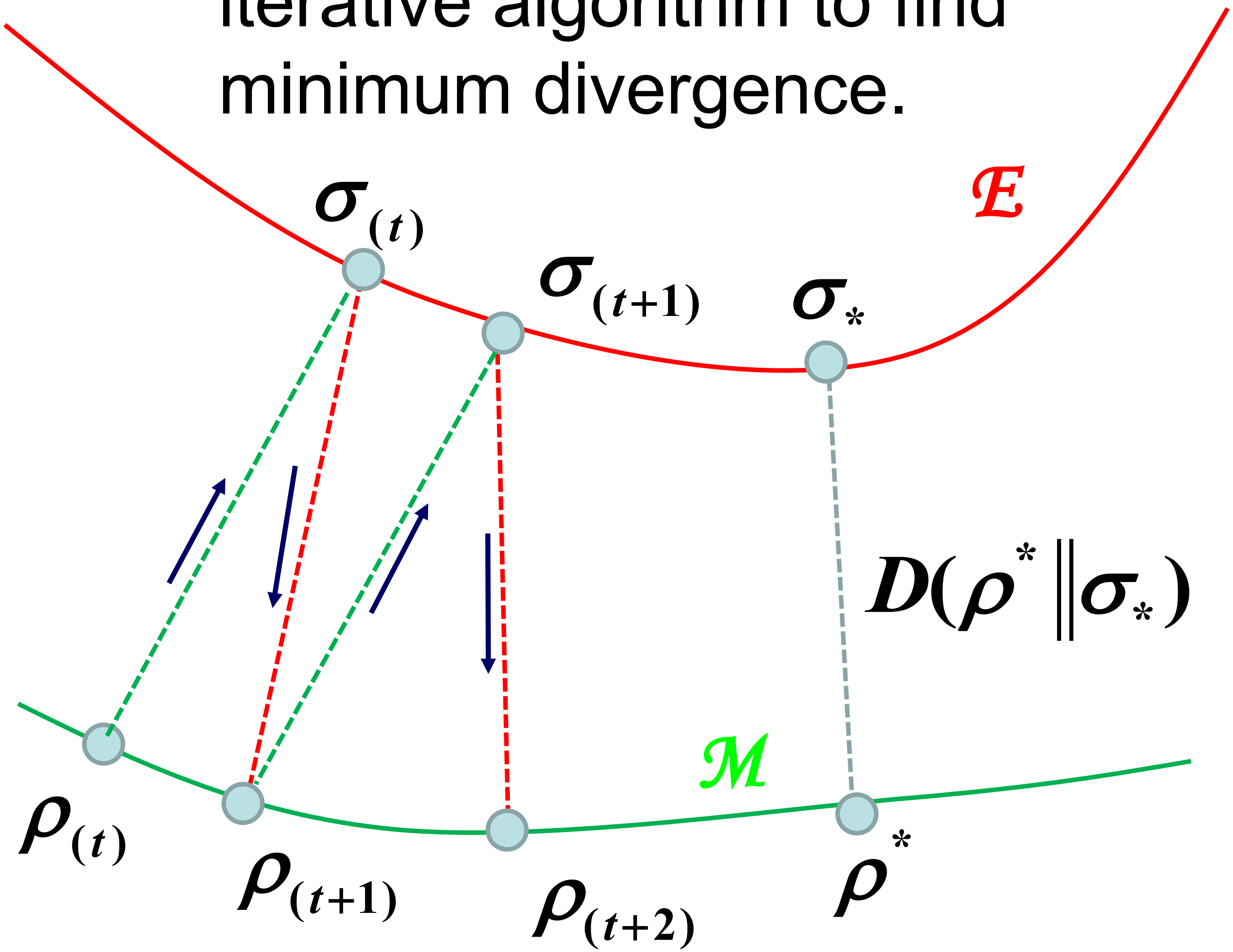
e-step $\sigma_{(t+1)} := \arg\min_{\sigma \in \mathcal{E}} D(\rho_{(t+1)} \| \sigma)$

$$D(\rho_{(t+1)} \| \sigma_{(t+1)}) \leq D(\rho_{(t+1)} \| \sigma_{(t)}) \leq D(\rho_{(t)} \| \sigma_{(t)})$$

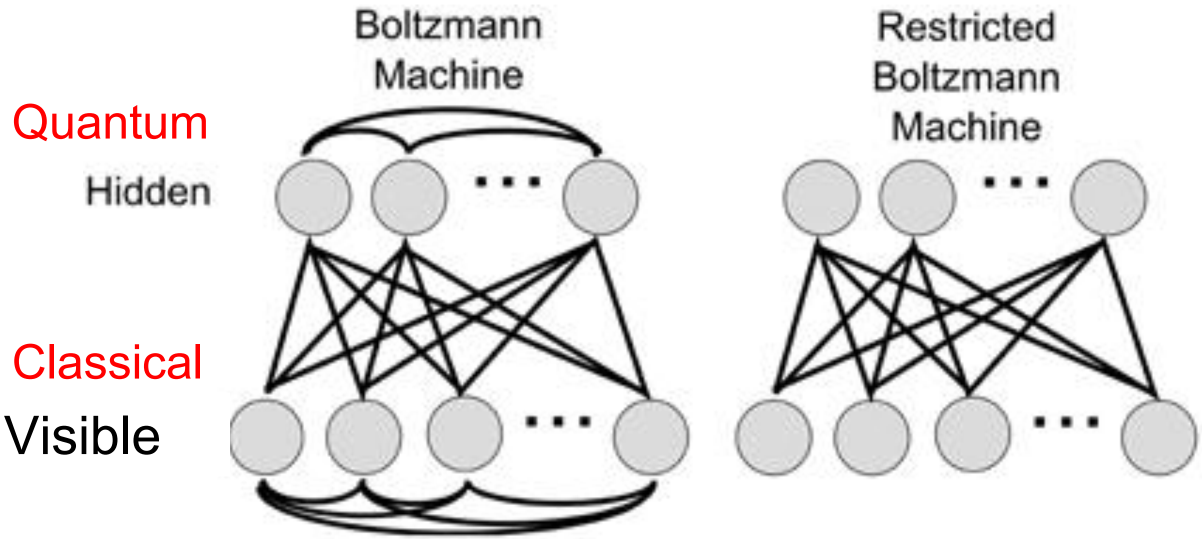
However, the convergence to the global minimum has not been discussed.

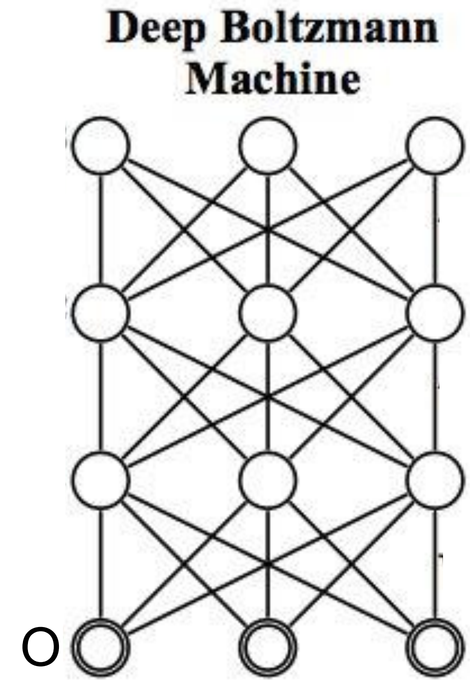
em-algorithm

iterative algorithm to find minimum divergence.



Semi-Quantum Boltzmann machine





Hamiltonian

$$H_{\vec{\theta}} = \sum_{k \in V} \theta_k^z \sigma_k^z + \sum_{(k,j) \in E} \theta_{k,j}^z \sigma_k^z \sigma_j^z + \theta_{k,j}^x \sigma_k^x \sigma_j^z + \sum_{k \in H} \theta_k^x \sigma_k^x + \theta_k^z \sigma_k^z$$

$$\vec{\theta} = ((\theta_k^z)_{k \in V}, (\theta_{k,j}^z, \theta_{k,j}^x)_{(k,j) \in E}, (\theta_k^x, \theta_k^z)_{k \in H})$$

$\rho_{\vec{\theta}} := \exp(H_{\vec{\theta}} - \phi(\vec{\theta})), \quad \phi(\vec{\theta}) := \log \mathrm{Tr} \exp(H_{\vec{\theta}})$ Convex function

Exponential family $\mathcal{E} := \{\rho_{\vec{\theta}}\}_{\vec{\theta}}$

sq Boltzmann machine is a key method for machine learning.

Semi-Quantum Boltzmann machine

$$H_{\vec{\theta}} = \sum_{k \in V} \theta_k^z \sigma_k^z + \sum_{(k,j)\in E} \theta_{k,j}^z \sigma_k^z \sigma_j^z + \theta_{k,j}^x \sigma_k^x \sigma_j^z + \sum_{j \in H} \theta_j^x \sigma_j^x + \theta_j^z \sigma_j^z$$

$$\vec{\theta} = ((\theta_k^z)_{k\in V}, (\theta_{k,j}^z, \theta_{k,j}^x)_{(k,j)\in E}, (\theta_k^x, \theta_k^z)_{k \in H})$$

$\rho_{\vec{\theta}} := \exp(H_{\vec{\theta}} - \phi(\vec{\theta})), \ \ \phi(\vec{\theta}) := \log \mathrm{Tr} \exp(H_{\vec{\theta}})$ Convex function

Exponential family $\mathcal{E} := \{\rho_{\vec{\theta}}\}_{\vec{\theta}}$

Mixture family

ρ_V : Distribution on observed system

$$\mathcal{M} := \{\rho_{HV} \mid \mathrm{Tr}_H \rho_{HV} = \rho_V\}$$

$$D(\rho_V \| \rho_{V,\vec{\theta}}) = \min_{\rho_{HV0} \in \mathcal{M}} D(\rho_{HV} \| \rho_{\vec{\theta}}), \ \rho_{V,\vec{\theta}} := \mathrm{Tr}_H \rho_{\vec{\theta}}$$

$$\min_{\vec{\theta}} D(\rho_V \| \rho_{V,\vec{\theta}}) = \min_{\vec{\theta}, \rho_{HV0} \in \mathcal{M}} D(\rho_{HV} \| \rho_{\vec{\theta}})$$

Find $\vec{\theta}_* := \operatorname*{argmin}_{\vec{\theta}_* \in \mathcal{E}} \min_{\rho \in \mathcal{M}} D(\rho \| \rho_{\vec{\theta}})$

Semi-Quantum Boltzmann machine

Mixture family

$$\mathcal{M} := \{\sum_x P_X(x)|x\rangle\langle x| \otimes \rho_{H|x}\}$$

Our aim

$$\vec{\theta}_* := \operatorname*{argmin}_{\vec{\theta}\in\mathcal{E}} \min_{\rho\in\mathcal{M}} D(\rho\|\rho_{\vec{\theta}})$$

$$\begin{aligned} & D(\rho\|\rho_{\vec{\theta}}) \\ &= \sum_x P_X(x)\mathrm{Tr}\rho_{H|x}\left(\log P_X(x)\rho_{H|x} - \log P_{X,\theta}(x)\rho_{H|x,\theta}\right) \\ &= \sum_x P_X(x)\left(\log P_X(x) - \log P_{X,\theta}(x)\right) \\ &\quad + \sum_x P_X(x)\mathrm{Tr}\rho_{H|x}\left(\log \rho_{H|x} - \log \rho_{H|x,\theta}\right) \\ &= D(P_X\|P_{X,\theta}) + \sum_x P_X(x) D(\rho_{H|x}\|\rho_{H|x,\theta}) \end{aligned}$$

e-step $\quad P_X = P_{X,\theta},\ \ \rho_{H|x} = \rho_{H|x,\theta} \quad \rho_{VH} = \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_{H|x}$

m-step

Convex minimization

$$\theta^{(t+1)} := \underset{\theta}{\operatorname{argmin}} D(\rho_{VH} \| \rho_{VH,\vec{\theta}}) = \arg\min_{\theta} \phi(\theta) - \operatorname{Tr} \rho_{VH} H_{\vec{\theta}}$$

$$H_{\vec{\theta}} = \sum_{k \in V} \theta_k^z \sigma_k^z + \sum_{(k,j) \in E} \theta_{k,j}^z \sigma_k^z \sigma_j^z + \theta_{k,j}^x \sigma_k^x \sigma_j^z + \sum_{j \in H} \theta_j^x \sigma_j^x + \theta_j^z \sigma_j^z$$

$$\rho_{\vec{\theta}} := \exp(H_{\vec{\theta}} - \phi(\vec{\theta})), \quad \phi(\vec{\theta}) := \log \operatorname{Tr} \exp(H_{\vec{\theta}})$$

We can use gradient descent

$$\frac{\partial}{\partial \theta_k^z} (\phi(\theta) - \operatorname{Tr} \rho_{VH} H_{\vec{\theta}}) = \operatorname{Tr} \rho_{VH,\vec{\theta}} \sigma_k^z - \operatorname{Tr} \rho_{VH} \sigma_k^z$$

$$\frac{\partial}{\partial \theta_j^x} (\phi(\theta) - \operatorname{Tr} \rho_{VH} H_{\vec{\theta}}) = \operatorname{Tr} \rho_{VH,\vec{\theta}} \sigma_j^x - \operatorname{Tr} \rho_{VH} \sigma_j^x$$

$$\frac{\partial}{\partial \theta_{k,j}^x} (\phi(\theta) - \operatorname{Tr} \rho_{VH} H_{\vec{\theta}}) = \operatorname{Tr} \rho_{VH,\vec{\theta}} \sigma_k^z \sigma_j^x - \operatorname{Tr} \rho_{VH} \sigma_k^z \sigma_j^x \qquad \text{etc}$$

$$\vdots$$

When size is not so large, it can be solved classically (analytically).

m-step with large size

$$\rho_{\vec{\theta}} := \exp(H_{\vec{\theta}} - \phi(\vec{\theta})), \quad \phi(\vec{\theta}) := \log \mathrm{Tr} \exp(H_{\vec{\theta}})$$

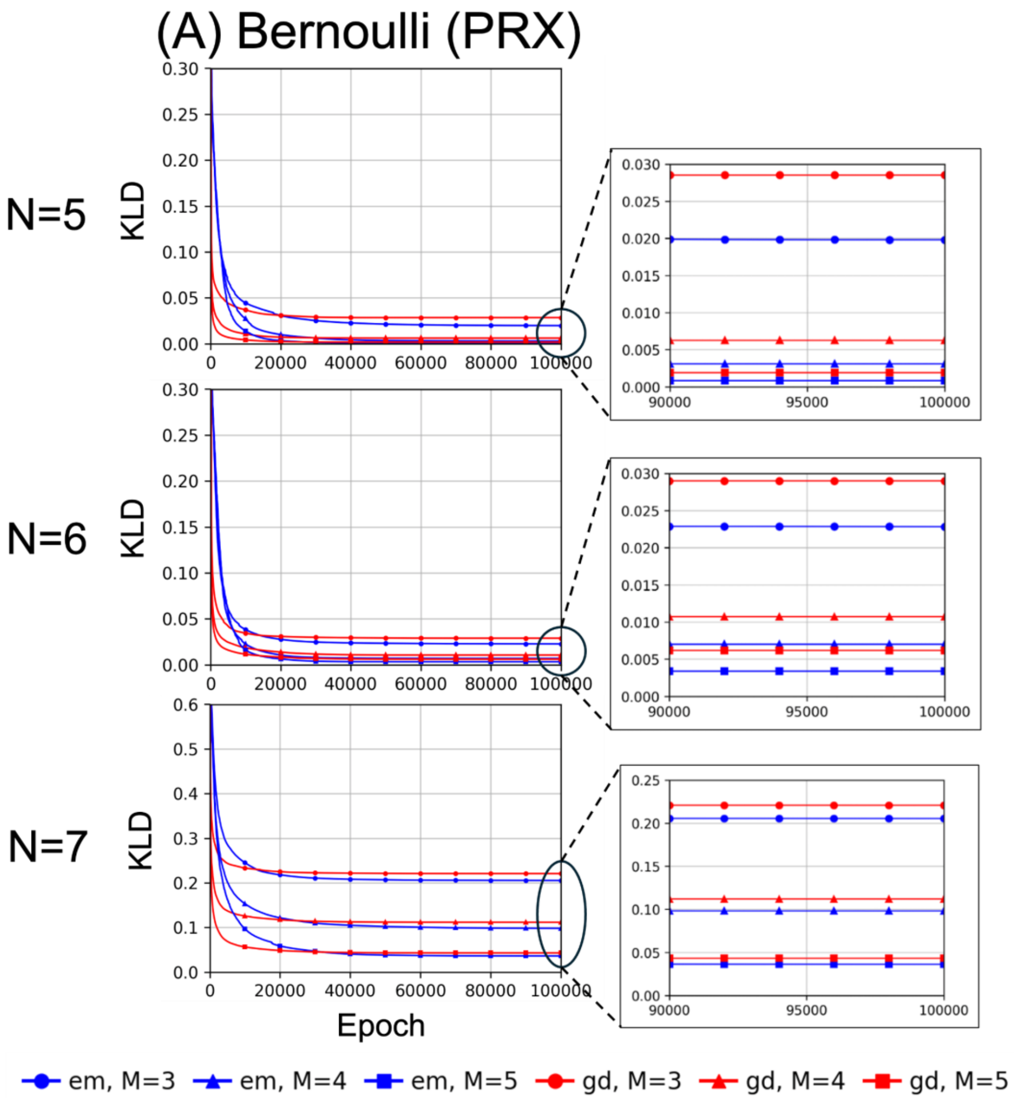
This model is fully visible quantum Boltzmann machine. We can use the method by Coopmans & Benedetti's.

Stochastic gradient descent + quantum Gibbs sampling.

In the classical case, the classical system requires certain structure like 2-partite structures etc. But, this method does not.

Recently, many experimental studies have been done for quantum Gibbs sampling.

Numerical calculation



bipartite graph structure between visible and hidden nodes.

N: Number of visible nodes

M: Number of hidden nodes

Conclusion

- Application of em algorithm to semiquantum Boltzmann Machine (sqBM).
- Generally, the objective function is non-convex. To resolve this, we divide the problem into e-step and m-step.
- e-step is easy, and m-step can be done with convex minimization.
- m-step is the same as fully visible quantum Boltzmann Machine.
- Fully visible quantum Boltzmann Machine has been already solved by Coopmans & Benedetti’s via stochastic gradient descent + quantum Gibbs sampling.

References

[1] Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B. & Melko, R. Quantum boltzmann machine. Physical Review X 8, 021050 (2018).

[2] Coopmans, L. & Benedetti, M. On the sample complex ity of quantum boltzmann machine learning. Communi cations Physics 7, 274 (2024).

[3] Demidik, M., Tuysuz, C., Piatkowski, N., Grossi, M. & Jansen, K. Expressive equivalence of classical and quan tum restricted boltzmann machines. Communications Physics 8, 413 (2025).

[4] S. Amari, "Information geometry of the EM and em algorithms for neural networks," Neural Networks **8**, 1379 – 1408 (1995).

[5] S. Amari, K. Kurata and H. Nagaoka, "Information geometry of Boltzmann machines," IEEE Transactions on Neural Networks, **3**, 260 – 271, (1992).

[6] Hayashi, M. Bregman divergence based em algorithm and its application to classical and quantum rate distortion theory. IEEE Transactions on Information Theory 69, 3460–3492 (2023).