Restrictions on Quantum Advantage in Quantum Machine Learning

Apoorva D. Patel

IISc Quantum Technology Initiative, Bangalore
Visiting Professor, Quantum Research Park, IISc, Bangalore
Visiting Professor, ICTS-TIFR, Bangalore

International Conference on Quantum Information Science and Technology (ICQIST 2025)
CQuERE, TCG CREST, Kolkata
11 December 2025

Supported by: *DRDO Project "Quantum Algorithms and Simulations"*

Abstract

It is well-established that the extent of quantum advantage in algorithms is problem-dependent. For problems formulated in terms of probability distributions, characterising the input data as well as the targeted observables, quantum advantage can be obtained only when features of density matrices that are absent in classical probability distributions are exploited. This restriction is emphatically demonstrated in case of unsupervised machine learning, with illustrative examples and explicit solutions. Furthermore, the result brings out the features required in big data analysis, to obtain quantum or quantum-inspired advantage, with important constraints for both pattern recognition and sensing applications.

Quantum Potentialities and Limitations

Quantum Potentialities and Limitations

Fact 1: The extent of Quantum Advantage is problem dependent.

Some problems are easy (BQP), and some problems are hard (QMA).

Quantum theory was invented to explain physical phenomena that existing classical theories failed to explain. Quantum Advantage can be sizeable when such non-classical features are at the heart of the problems to be solved. Classical strategies are sufficient in many other situations, with not much to gain from quantum implementation. Look for Planck's constant.

Quantum Potentialities and Limitations

Fact 1: The extent of Quantum Advantage is problem dependent.

Some problems are easy (BQP), and some problems are hard (QMA).

Quantum theory was invented to explain physical phenomena that existing classical theories failed to explain. Quantum Advantage can be sizeable when such non-classical features are at the heart of the problems to be solved. Classical strategies are sufficient in many other situations, with not much to gain from quantum implementation. Look for Planck's constant.

Fact 2: Quantum dynamics is highly fragile against external disturbances.

It can work well in cooperative/protected settings, but not in hostile ones.

The need to shield quantum dynamics from unwanted disturbances (including those arising from control systems and measurements) makes quantum technology expensive. Consider economics.

Quantum Potentialities and Limitations

Fact 1: The extent of Quantum Advantage is problem dependent.

Some problems are easy (BQP), and some problems are hard (QMA).

Quantum theory was invented to explain physical phenomena that existing classical theories failed to explain. Quantum Advantage can be sizeable when such non-classical features are at the heart of the problems to be solved. Classical strategies are sufficient in many other situations, with not much to gain from quantum implementation. Look for Planck's constant.

Fact 2: Quantum dynamics is highly fragile against external disturbances.

It can work well in cooperative/protected settings, but not in hostile ones.

The need to shield quantum dynamics from unwanted disturbances (including those arising from control systems and measurements) makes quantum technology expensive. Consider economics.

Fact 3: Robust classical technology continues to improve, partly because of new ways of thinking offered by developments in quantum technology.

What is described as Quantum Advantage is a moving target. Quantum-inspired ideas emerge.

Quantum Potentialities and Limitations

Fact 1: The extent of Quantum Advantage is problem dependent.

Some problems are easy (BQP), and some problems are hard (QMA).

Quantum theory was invented to explain physical phenomena that existing classical theories failed to explain. Quantum Advantage can be sizeable when such non-classical features are at the heart of the problems to be solved. Classical strategies are sufficient in many other situations, with not much to gain from quantum implementation. Look for Planck's constant.

Fact 2: Quantum dynamics is highly fragile against external disturbances.

It can work well in cooperative/protected settings, but not in hostile ones.

The need to shield quantum dynamics from unwanted disturbances (including those arising from control systems and measurements) makes quantum technology expensive. Consider economics.

Fact 3: Robust classical technology continues to improve, partly because of new ways of thinking offered by developments in quantum technology.

What is described as Quantum Advantage is a moving target. Quantum-inspired ideas emerge.

Implication: Quantum technology will therefore be attractive only in cases where the advantage offered by it is sufficiently large to offset its cost.

Quantum technology will be practically useful as special-purpose devices, as custom subroutines in larger applications. Hybrid classical-quantum systems, with small quantum modules embedded in large classical peripherals, are also required for us to interpret quantum signals.

Machine Learning

Big data: Various types of sensors and detectors have become affordable and convenient. They collect huge amount of data (with space and time labels), which have to be analysed to make decisions.

Often there is no time and space to store the data;
interesting features must be extracted quickly and the rest discarded.

Examples: Astronomy, satellite imagery, surveillance, weather and climate, internet traffic, model system tests and validation (for design purposes), experimental observations, collider physics, genetic information.

Machine Learning

Big data: Various types of sensors and detectors have become affordable and convenient. They collect huge amount of data (with space and time labels), which have to be analysed to make decisions.

Often there is no time and space to store the data;
interesting features must be extracted quickly and the rest discarded.

Examples: Astronomy, satellite imagery, surveillance, weather and climate, internet traffic, model system tests and validation (for design purposes), experimental observations, collider physics, genetic information.

Analysis: Important steps to be (probabilistically) implemented are:
(1) pattern recognition, (2) classification, (3) recommendation systems.

Give the tedious job to a computer, together with some rules:

- Supervised learning (training data with known labels provided)
- Unsupervised learning (similar examples and analogies are provided)
- Reinforcement learning (reward functions are externally provided)

Machine Learning

Big data: Various types of sensors and detectors have become affordable and convenient. They collect huge amount of data (with space and time labels), which have to be analysed to make decisions.

Often there is no time and space to store the data;
interesting features must be extracted quickly and the rest discarded.

Examples: Astronomy, satellite imagery, surveillance, weather and climate, internet traffic, model system tests and validation (for design purposes), experimental observations, collider physics, genetic information.

Analysis: Important steps to be (probabilistically) implemented are:
(1) pattern recognition, (2) classification, (3) recommendation systems.

Give the tedious job to a computer, together with some rules:

- Supervised learning (training data with known labels provided)
- Unsupervised learning (similar examples and analogies are provided)
- Reinforcement learning (reward functions are externally provided)

Neural network methods are frequently used, heuristically without proof.
Available computational resources limit what problems can be tackled.

Unsupervised Learning

The task is to find the probability distribution of the data obtained in a specific scenario, which can then be used to make new predictions.
Initially one guesses the probability distribution based on analogies, and then evolves it (by tuning parameters) to improve the approximation.

Need: A criterion to measure difference of distributions, and an algorithm to reduce it.

Unsupervised Learning

The task is to find the probability distribution of the data obtained in a specific scenario, which can then be used to make new predictions.
Initially one guesses the probability distribution based on analogies, and then evolves it (by tuning parameters) to improve the approximation.

Need: A criterion to measure difference of distributions, and an algorithm to reduce it.

Difference between two probability distributions is measured using the Kullback-Leibler (KL) divergence, which is their relative entropy:

$$D_{KL}(p\|q) = \sum_x p(x)\log(p(x)/q(x)) \ .$$

It is the amount of information lost when $p(x)$ is used to approximate $q(x)$. It is asymmetric in $p(x)$ and $q(x)$, but it is always non-negative and vanishes only when $p(x)$ and $q(x)$ coincide.

Unsupervised Learning

The task is to find the probability distribution of the data obtained in a specific scenario, which can then be used to make new predictions.
Initially one guesses the probability distribution based on analogies, and then evolves it (by tuning parameters) to improve the approximation.

Need: A criterion to measure difference of distributions, and an algorithm to reduce it.

Difference between two probability distributions is measured using the Kullback-Leibler (KL) divergence, which is their relative entropy:

$$D_{KL}(p\|q) = \sum_x p(x) \log(p(x)/q(x)) \ .$$

It is the amount of information lost when $p(x)$ is used to approximate $q(x)$. It is asymmetric in $p(x)$ and $q(x)$, but it is always non-negative and vanishes only when $p(x)$ and $q(x)$ coincide.

So a parametrised $p(x)$ is fitted to the data $q(x)$, by minimising $D_{KL}(p||q)$.
When D_{KL} is expanded about its minimum, the leading quadratic form gives the Fisher information metric obeying the Cramér-Rao bound.

Boltzmann Machines

When the probability distribution arises from a physical process, one can expect $p(x) = e^{-\beta H(x)}/(\sum_{states} e^{-\beta H})$, in terms of a local Hamiltonian. When some degrees of freedom are not observed, they are summed over. That produces reduced probability distributions. The unobserved degrees of freedom are mapped to hidden variables of a neural network.

Boltzmann Machines

When the probability distribution arises from a physical process, one can expect $p(x) = e^{-\beta H(x)}/(\sum_{states} e^{-\beta H})$, in terms of a local Hamiltonian. When some degrees of freedom are not observed, they are summed over. That produces reduced probability distributions. The unobserved degrees of freedom are mapped to hidden variables of a neural network.

Boltzmann machines are stochastic spin-glass models with an external field, usually defined using Ising spins ($s_i = \pm 1$):

$$H = -\sum_{ij} w_{ij} s_i s_j - \sum_i b_i s_i,$$

where w_{ij} is the interaction strength between s_i and s_j, and b_i is the external bias field. Simulations evolve the spin network towards its thermal equilibrium state, starting from a high temperature (simulated annealing).

Boltzmann Machines

When the probability distribution arises from a physical process, one can expect $p(x) = e^{-\beta H(x)}/(\sum_{states} e^{-\beta H})$, in terms of a local Hamiltonian. When some degrees of freedom are not observed, they are summed over. That produces reduced probability distributions. The unobserved degrees of freedom are mapped to hidden variables of a neural network.

Boltzmann machines are stochastic spin-glass models with an external field, usually defined using Ising spins ($s_i = \pm 1$):

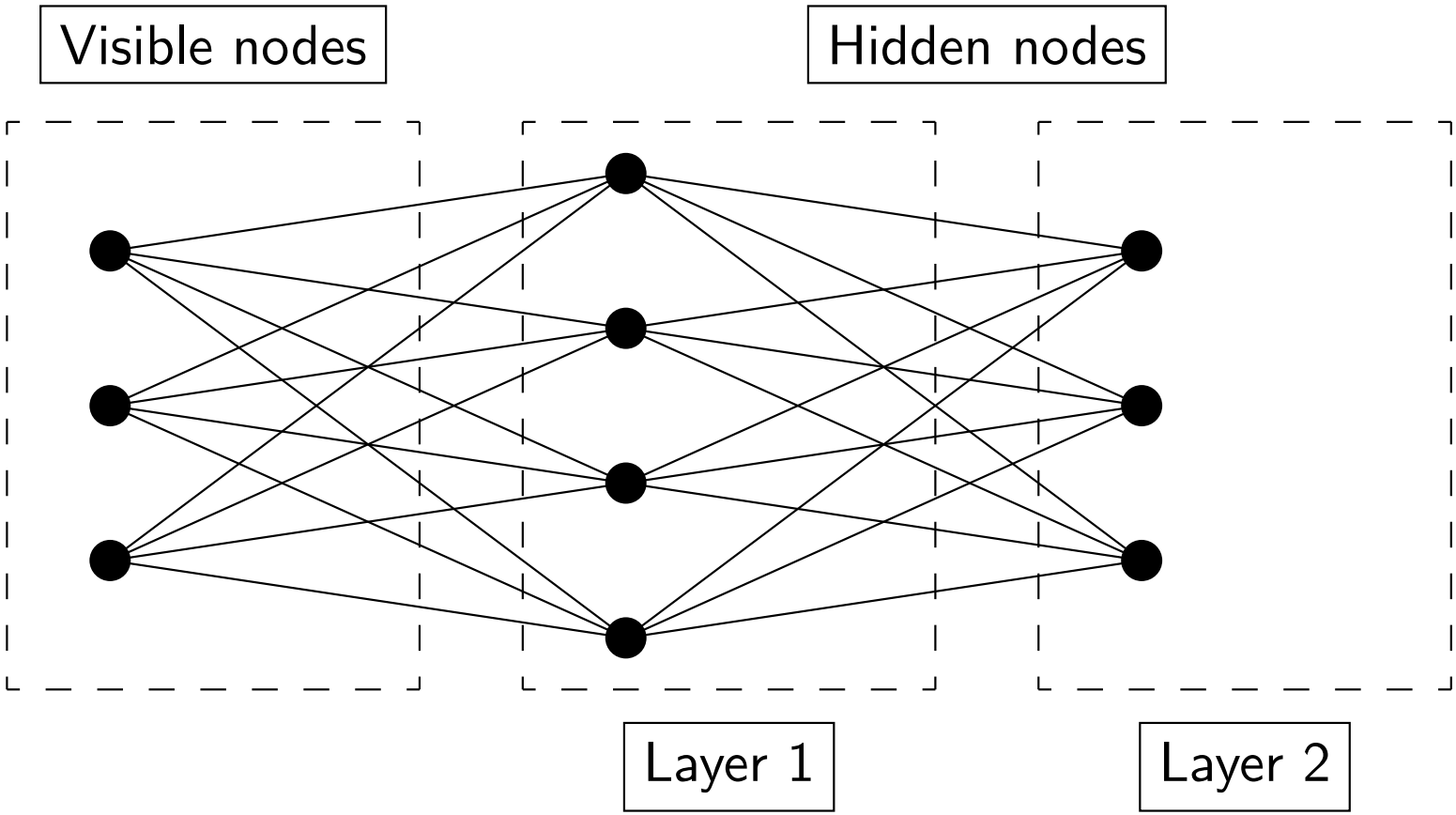
$$H = -\sum_{ij} w_{ij} s_i s_j - \sum_i b_i s_i,$$

where w_{ij} is the interaction strength between s_i and s_j, and b_i is the external bias field. Simulations evolve the spin network towards its thermal equilibrium state, starting from a high temperature (simulated annealing).

Restricted Boltzmann machines: Connections exist only between the observable and the hidden variables (layered structure with no connections among the observable ones or among the hidden ones).

Deep Boltzmann machines: There are multiple layers of hidden variables.

Restricted and Deep Boltzmann Machines



Simple Boltzmann machines are versions of the spin glass models of statistical mechanics, with restricted connectivity and variable weights. They can include external bias fields. Depending on the coupling values, they can exhibit regular/periodic/glassy/chaotic behaviour.

Quantum Generalisation

The quantum Boltzmann machine keeps the same network structure, but lets the inputs be quantum states and the nodes be qubits/qudits.

What additional capabilities would such a machine offer compared to the classical version?

Quantum Generalisation

The quantum Boltzmann machine keeps the same network structure, but lets the inputs be quantum states and the nodes be qubits/qudits.

What additional capabilities would such a machine offer compared to the classical version?

The classical probability distribution generalises to the quantum density matrix, with $\langle O \rangle = \sum_x p(x)O(x) \longrightarrow Tr(\rho O)$.

Since ρ is positive semi-definite, it can always be expressed as $e^{-\beta H}$.

The reduced probability distributions generalise to reduced density matrices obtained by partial trace over hidden variables.

The relative entropy generalises to $S(\rho\|\sigma) = Tr(\rho(\log\rho - \log\sigma))$, while remaining non-negative and vanishing only when ρ and σ coincide.

Quantum Generalisation

The quantum Boltzmann machine keeps the same network structure, but lets the inputs be quantum states and the nodes be qubits/qudits.

What additional capabilities would such a machine offer compared to the classical version?

The classical probability distribution generalises to the quantum density matrix, with $\langle O \rangle = \sum_x p(x)O(x) \longrightarrow Tr(\rho O)$.
Since ρ is positive semi-definite, it can always be expressed as $e^{-\beta H}$.
The reduced probability distributions generalise to reduced density matrices obtained by partial trace over hidden variables.
The relative entropy generalises to $S(\rho\|\sigma) = Tr(\rho(\log\rho - \log\sigma))$, while remaining non-negative and vanishing only when ρ and σ coincide.

The additional quantum feature is that ρ and O may not commute. When they do, they can be diagonalised simultaneously, and the analysis can be reduced to the classical case. So $[\rho, O] \neq 0$ is a necessary condition for the quantum Boltzmann machine to offer an advantage.

Quantum Generalisation

The quantum Boltzmann machine keeps the same network structure, but lets the inputs be quantum states and the nodes be qubits/qudits.

What additional capabilities would such a machine offer compared to the classical version?

The classical probability distribution generalises to the quantum density matrix, with $\langle O \rangle = \sum_x p(x)O(x) \longrightarrow Tr(\rho O)$.
Since ρ is positive semi-definite, it can always be expressed as $e^{-\beta H}$.
The reduced probability distributions generalise to reduced density matrices obtained by partial trace over hidden variables.
The relative entropy generalises to $S(\rho\|\sigma) = Tr(\rho(\log\rho - \log\sigma))$, while remaining non-negative and vanishing only when ρ and σ coincide.

The additional quantum feature is that ρ and O may not commute. When they do, they can be diagonalised simultaneously, and the analysis can be reduced to the classical case. So $[\rho, O] \neq 0$ is a necessary condition for the quantum Boltzmann machine to offer an advantage.

Note: The quantum advantage depends on the observable chosen, while the same $p(x)$ works for all observables in the classical case.

Quantum Advantage

The quantum advantage is expected to be the maximum when the norm of $[\rho, O]$ is the largest ($\|\rho\| \leq 1$ and $\|O\|$ can be fixed by convention).
The quantum Fisher information metric can be expressed in terms of $[\rho, O]$. The corresponding Cramér-Rao bound characterises the sensitivity of specific quantum states in determination of a given observable.

Quantum Advantage

The quantum advantage is expected to be the maximum when the norm of $[\rho, O]$ is the largest ($\|\rho\| \leq 1$ and $\|O\|$ can be fixed by convention). The quantum Fisher information metric can be expressed in terms of $[\rho, O]$. The corresponding Cramér-Rao bound characterises the sensitivity of specific quantum states in determination of a given observable.

Since both ρ and O can be diagonalised by unitary transformations, one can find the maximally sensitive observable O for a given state ρ, and the optimal probe ρ to measure a given observable O.

Quantum Advantage

The quantum advantage is expected to be the maximum when the norm of $[\rho, O]$ is the largest ($\|\rho\| \leq 1$ and $\|O\|$ can be fixed by convention).
The quantum Fisher information metric can be expressed in terms of $[\rho, O]$. The corresponding Cramér-Rao bound characterises the sensitivity of specific quantum states in determination of a given observable.

Since both ρ and O can be diagonalised by unitary transformations, one can find the maximally sensitive observable O for a given state ρ, and the optimal probe ρ to measure a given observable O.

Example:

For a single qubit, $\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma})$. For an O diagonal in the Z-basis, the norm of $[\rho, O]$ is maximum when $|\vec{r}| = 1$ in the equatorial plane. Starting from ρ_{diag} in the Z-basis, such a ρ is obtained by rotating ρ_{diag} about an axis in the equatorial plane by angle $\pi/2$.

Quantum Advantage

The quantum advantage is expected to be the maximum when the norm of $[\rho, O]$ is the largest ($\|\rho\| \leq 1$ and $\|O\|$ can be fixed by convention). The quantum Fisher information metric can be expressed in terms of $[\rho, O]$. The corresponding Cramér-Rao bound characterises the sensitivity of specific quantum states in determination of a given observable.

Since both ρ and O can be diagonalised by unitary transformations, one can find the maximally sensitive observable O for a given state ρ, and the optimal probe ρ to measure a given observable O.

Example:

For a single qubit, $\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma})$. For an O diagonal in the Z-basis, the norm of $[\rho, O]$ is maximum when $|\vec{r}| = 1$ in the equatorial plane. Starting from ρ_{diag} in the Z-basis, such a ρ is obtained by rotating ρ_{diag} about an axis in the equatorial plane by angle $\pi/2$.

Applications: Filter orientation for detecting electromagnetic wave polarisation, interferometer orientation to detect gravitational radiation.

Multi-qubit Version

The naive multi-qubit extension of this strategy maximises quantum advantage by evolving ρ using the transverse-field quantum Ising model:

$$H = -\sum_{ij} w_{ij}\sigma_i^{(z)}\sigma_j^{(z)} - \sum_i b_i\sigma_i^{(x)}.$$

Iterative algorithms have been developed to tune the hidden parameters of such an H so as to minimise the relative entropy $S(\rho\|\sigma)$. Such attempts have not found much quantum advantage.

Multi-qubit Version

The naive multi-qubit extension of this strategy maximises quantum advantage by evolving ρ using the transverse-field quantum Ising model:

$$H = -\sum_{ij} w_{ij}\sigma_i^{(z)}\sigma_j^{(z)} - \sum_i b_i\sigma_i^{(x)}.$$

Iterative algorithms have been developed to tune the hidden parameters of such an H so as to minimise the relative entropy $S(\rho\|\sigma)$. Such attempts have not found much quantum advantage.

On the other hand, the maximisation of quantum advantage can be explicitly carried out in any dimension, by maximising $\|[\rho, O]\|$ as follows.

Multi-qubit Version

The naive multi-qubit extension of this strategy maximises quantum advantage by evolving ρ using the transverse-field quantum Ising model:

$$H = -\sum_{ij} w_{ij}\sigma_i^{(z)}\sigma_j^{(z)} - \sum_i b_i\sigma_i^{(x)}.$$

Iterative algorithms have been developed to tune the hidden parameters of such an H so as to minimise the relative entropy $S(\rho\|\sigma)$. Such attempts have not found much quantum advantage.

On the other hand, the maximisation of quantum advantage can be explicitly carried out in any dimension, by maximising $\|[\rho, O]\|$ as follows.

Note: Only traceless parts of ρ and O contribute to $[\rho, O]$.
Work in the basis where O is diagonal. In n-dimensional Hilbert space, its traceless components can be expressed as a linear combination of $n-1$ independent Cartan generators. Explicitly, traceless part of O is $diag(a_1, \ldots, a_p, -b_1, \ldots, -b_q)$, $p+q=n$, $a_i, b_j \geq 0$, $\sum_i a_i = t = \sum_j b_j$.

Maximising Quantum Advantage

$\|[\rho, O]\|_\infty$ is then maximised by aligning ρ with the Cartan component $T = diag(\ldots, t, \ldots, -t, \ldots)$, with t in one of the p directions and $-t$ in one of the q directions, and all other components zero.

The remaining traceless contributions to O, orthogonal to T, have the form $diag(\ldots, 1, \ldots, 1, \ldots)$, and each of them has magnitude less than t.

The solution is $\rho = diag(\ldots, \frac{1}{2}, \ldots, \frac{1}{2}, \ldots) + c\vec{\sigma} \cdot \hat{n}_T$,
with $\hat{n}_T$ in the direction transverse to T and $c = \frac{1}{2}$.

Maximising Quantum Advantage

$\|[\rho, O]\|_\infty$ is then maximised by aligning ρ with the Cartan component $T = diag(\ldots, t, \ldots, -t, \ldots)$, with t in one of the p directions and $-t$ in one of the q directions, and all other components zero.

The remaining traceless contributions to O, orthogonal to T, have the form $diag(\ldots, 1, \ldots, 1, \ldots)$, and each of them has magnitude less than t.

The solution is $\rho = diag(\ldots, \frac{1}{2}, \ldots, \frac{1}{2}, \ldots) + c\vec{\sigma} \cdot \hat{n}_T$,
with $\hat{n}_T$ in the direction transverse to T and $c = \frac{1}{2}$.

The high degeneracy of the above choice can be lifted by choosing locations of t and $-t$ to match those of $max(a_i)$ and $max(b_j)$.
That maximises the spectral gap of the Cartan contributions,
and speeds up convergence of $\|[\rho, O]\|$ to its maximum.

Maximising Quantum Advantage

$\|[\rho, O]\|_\infty$ is then maximised by aligning ρ with the Cartan component $T = diag(\ldots, t, \ldots, -t, \ldots)$, with t in one of the p directions and $-t$ in one of the q directions, and all other components zero.

The remaining traceless contributions to O, orthogonal to T, have the form $diag(\ldots, 1, \ldots, 1, \ldots)$, and each of them has magnitude less than t.

The solution is $\rho = diag(\ldots, \frac{1}{2}, \ldots, \frac{1}{2}, \ldots) + c\vec{\sigma} \cdot \hat{n}_T$,
with $\hat{n}_T$ in the direction transverse to T and $c = \frac{1}{2}$.

The high degeneracy of the above choice can be lifted by choosing locations of t and $-t$ to match those of $max(a_i)$ and $max(b_j)$.
That maximises the spectral gap of the Cartan contributions,
and speeds up convergence of $\|[\rho, O]\|$ to its maximum.

ρ aligned with T can be efficiently prepared in an $m = \lceil \log_2 n \rceil$ qubit register, with $O(m)$ operations. (Permuting diagonal elements of O helps.)
Create the pure state in the direction of $max(a_i)$, and then rotate it by angle $\frac{\pi}{2}$ about the equatorial axis orthogonal to $\hat{n}_T$, in the two-dimensional subspace specified by $max(a_i)$ and $max(b_j)$.

Lessons

- Advantage of quantum Boltzmann machine is problem dependent.

Analytical maximisation of $\|[\rho, O]\|$ skips the need to minimise $S(\rho\|\sigma)$.

Lessons

- Advantage of quantum Boltzmann machine is problem dependent. Analytical maximisation of $\|[\rho, O]\|$ skips the need to minimise $S(\rho\|\sigma)$.
- The quantum advantage is maximum when ρ is a pure state, i.e. visible state is fully decoupled from the hidden degrees of freedom. (Generically, interaction with the environment reduces quantum effects.)

Lessons

- Advantage of quantum Boltzmann machine is problem dependent. Analytical maximisation of $\|[\rho, O]\|$ skips the need to minimise $S(\rho\|\sigma)$.
- The quantum advantage is maximum when ρ is a pure state, i.e. visible state is fully decoupled from the hidden degrees of freedom. (Generically, interaction with the environment reduces quantum effects.)
- Even when quantum correlations are allowed to spread to the hidden degrees of freedom, by letting ρ to be a mixed state, the spread is limited as per the Schmidt decomposition, i.e. n visible qubits can couple to at most n hidden (configurable) qubits,

$$|\psi\rangle_{vh} = \sum_i \sqrt{p_i}|i\rangle_v \otimes |i'\rangle_h$$ in the basis that diagonalises ρ_v.

(Beyond that, the advantage offered by Boltzmann machines is classical.)

Lessons

- Advantage of quantum Boltzmann machine is problem dependent. Analytical maximisation of $\|[\rho, O]\|$ skips the need to minimise $S(\rho\|\sigma)$.
- The quantum advantage is maximum when ρ is a pure state, i.e. visible state is fully decoupled from the hidden degrees of freedom. (Generically, interaction with the environment reduces quantum effects.)
- Even when quantum correlations are allowed to spread to the hidden degrees of freedom, by letting ρ to be a mixed state, the spread is limited as per the Schmidt decomposition, i.e. n visible qubits can couple to at most n hidden (configurable) qubits,
$|\psi\rangle_{vh} = \sum_i \sqrt{p_i}|i\rangle_v \otimes |i'\rangle_h$ in the basis that diagonalises ρ_v.
(Beyond that, the advantage offered by Boltzmann machines is classical.)
- The one-to-one map of the Schmidt decomposition also implies that restricted Boltzmann machines suffice for getting any quantum advantage.

Lessons

- Advantage of quantum Boltzmann machine is problem dependent. Analytical maximisation of $\|[\rho, O]\|$ skips the need to minimise $S(\rho\|\sigma)$.
- The quantum advantage is maximum when ρ is a pure state, i.e. visible state is fully decoupled from the hidden degrees of freedom. (Generically, interaction with the environment reduces quantum effects.)
- Even when quantum correlations are allowed to spread to the hidden degrees of freedom, by letting ρ to be a mixed state, the spread is limited as per the Schmidt decomposition, i.e. n visible qubits can couple to at most n hidden (configurable) qubits,
  $|\psi\rangle_{vh} = \sum_i \sqrt{p_i}|i\rangle_v \otimes |i'\rangle_h$ in the basis that diagonalises ρ_v.
  (Beyond that, the advantage offered by Boltzmann machines is classical.)
- The one-to-one map of the Schmidt decomposition also implies that restricted Boltzmann machines suffice for getting any quantum advantage.
- For measurement of a specific O, the optimal transverse ρ offers the best sensitivity as per the Cramér-Rao bound. It gives the eigenvector for the largest eigenvalue of the quantum Fisher information metric.

Further Comments

- Analytical (and hence classical) solutions have been found for many quantum problems (i.e. where nonzero Planck's constant appears).

Further Comments

- Analytical (and hence classical) solutions have been found for many quantum problems (i.e. where nonzero Planck's constant appears).
- Quantum advantage can be obtained only when features of density matrices that are absent in classical probability distributions are exploited.

Further Comments

- Analytical (and hence classical) solutions have been found for many quantum problems (i.e. where nonzero Planck's constant appears).
- Quantum advantage can be obtained only when features of density matrices that are absent in classical probability distributions are exploited.
- Examples of such features are: the off-diagonal elements of ρ, $\|[\rho, O]\| \neq 0$, quantum weights outside the real interval $[0, 1]$ (e.g. the "sign problem" in Wigner distributions and Feynman path integrals), dynamical correlations in quantum evolution arising from interference. They provide necessary conditions, but not sufficient ones.

Further Comments

- Analytical (and hence classical) solutions have been found for many quantum problems (i.e. where nonzero Planck's constant appears).
- Quantum advantage can be obtained only when features of density matrices that are absent in classical probability distributions are exploited.
- Examples of such features are: the off-diagonal elements of ρ, $\|[\rho, O]\| \neq 0$, quantum weights outside the real interval $[0, 1]$ (e.g. the "sign problem" in Wigner distributions and Feynman path integrals), dynamical correlations in quantum evolution arising from interference. They provide necessary conditions, but not sufficient ones.
- In case of machine learning, the restrictions can be bypassed, when the input data are quantum (have both magnitude and phase information). Their parallel correlated streams can be directly fed to a quantum processor, without intervening projective measurements.

Further Comments

- Analytical (and hence classical) solutions have been found for many quantum problems (i.e. where nonzero Planck's constant appears).
- Quantum advantage can be obtained only when features of density matrices that are absent in classical probability distributions are exploited.
- Examples of such features are: the off-diagonal elements of ρ, $\|[\rho, O]\| \neq 0$, quantum weights outside the real interval $[0, 1]$ (e.g. the "sign problem" in Wigner distributions and Feynman path integrals), dynamical correlations in quantum evolution arising from interference. They provide necessary conditions, but not sufficient ones.
- In case of machine learning, the restrictions can be bypassed, when the input data are quantum (have both magnitude and phase information). Their parallel correlated streams can be directly fed to a quantum processor, without intervening projective measurements.
- Coherent correlations between parallel data streams (equivalent to multiple copies of ρ) can provide an exponential quantum advantage.

[Huang, Kueng and Preskill, Phys. Rev. Lett. 126 (2021) 190505]