Towards efficient entanglement structure detection

Jue Xu* and Qi Zhao†

(Dated: August 7, 2022)

Verification (detection) of entanglement structure is an indispensable step for practical quantum computation (communication). In this work, we compare complexity and performance of several recently-developed methods, including entanglement witness methods, shadow tomography, classical machine learning, and quantum algorithms (circuits). illustrate the advantages and limitations of machine learning and quantum algorithms.

CONTENTS

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [], quantum communication [], and quantum cryptography []. It is essential to benckmark (characterize) multipartite entanglement structures of target states. We review the recently developed methods: entanglement witness [2], shadow tomography [3], classical machine learning [4], and quantum (variational/circuit) algorithms [5].

II. PRELIMINARIES

Notations: The (classical) training data (for supervised learning) is a set of m data points $\left\{(\mathbf{x}^{(i)}, y^{(i)})\right\}_{i=1}^{m}$ where each data point is a pair $(\mathbf{x}, y)$. Normally, the input (e.g., an image) $\mathbf{x} := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ is a vector where d is

* juexu@cs.umd.edu
† email

the number of *features* and its *label* $y \in \Sigma$ is a scalar with some discrete set Σ of alphabet/categories. For simplicity and the purpose of this paper, we assume $\Sigma = \{-1, 1\}$ (binary classification).

a graph $G = (V, E)$ is described by vertices V and edges E. denote a group by $\mathbb{G}$ and a subgroup $\mathbb{H}$. The hats on the matrices such as $\hat{A}$, $\hat{H}$, ρ (omitted), $\hat{O}$, $\hat{W}$, emphasize that they play the roles of operators. Denote vector (matrix) $\mathbf{x}$, $\mathbf{K}$ by boldface font.

For specific purpose, we use different basis (representations) for quantum states. One is the computational basis $\{|z\rangle\}$ with $z \in [2^n]$ where n is the number of qubits, while another useful one is the binary representation of computational basis $\{|\mathbf{x}\rangle \equiv |x_1, x_2, \ldots, x_n\rangle\}$ with $x_j \in \{0, 1\}$. For simplicity, we let $N \equiv 2^n$ and $|\mathbf{0}\rangle \equiv |0^n\rangle \equiv |0\rangle^{\otimes n}$ if no ambiguity. $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$

A. Entanglement detection

Large scale entanglement is the (main) resource of quantum advantages in quantum computation and communication.

1. *Bipartite entanglement*

Definition 1 (Entangled state)**.** Consider a n-partite (subsystem) system $\mathcal{H} = \bigotimes_i^n \mathcal{H}_i$, separable states or product states are i.e.,

$$|\Psi\rangle = \bigotimes_i |\psi_i\rangle \tag{1}$$

entangled pure state is a quantum state that cannot be written as a (tensor) product state (inseparable). For (generalize) mixed states, a mixed entangled state is a convex combination of entangled pure state, that is

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|, \forall i, p_i \geq 0, \sum_i p_i = 1 \tag{2}$$

...

Many methods [...] have been developed to determine whether a state is separable.

Definition 2 (bipartite separable)**.** A pure state is (bi-)separable if it is in a tensor product form $|\psi_b\rangle = |\phi_A\rangle \otimes |\phi_{\bar{A}}\rangle$, where $\mathcal{P}_2 = \{A, \bar{A}\}$ is a bipartition of the qubits in the system. A mixed state is separable if it can be written as a mixing of pure separable states. Note that each separable state $|\psi_b\rangle$ in the summation can have different bipartitions. The separable state set is denoted as S_b. There is another restricted way for the extension to mixed states. A state is $\mathcal{P}_2$-separable, if it is a mixing of pure separable states with a same partition $\mathcal{P}_2$, and we denote the state set as $S_b^{\mathcal{P}_2}$.

Instead of qualitatively determining entanglement, quantify entanglement

Definition 3 (Schmidt coefficient/rank/measure)**.** consider the following pure state on system AB, written in Schmidt form:

$$|\psi\rangle = \sum_i \sqrt{p_i} |\phi_i^A\rangle \otimes |\phi_i^B\rangle \tag{3}$$

where $\{|\psi_1^A\rangle\}$ is a basis for $\mathcal{H}_A$ and ... The strictly positive values $\sqrt{p_i}$ in the Schmidt decomposition are its *Schmidt coefficients*. The number of Schmidt coefficients, counted with multiplicity, is called its *Schmidt rank*, or Schmidt number. Schmidt measure is minimum of log r where r is number of terms in an expansion of the state in product basis.

Example 1. The Schmidt measure for any multi-partite GHZ states is 1, because there are just two terms. ... 1D, 2D, 3D-cluster state is $\lfloor \frac{N}{2} \rfloor$. .. of tree is the size of its minimal vertex cover.

Definition 4 (entropy)**.** In quantum mechanics (information), the von Neumann *entropy* of a density matrix is $H_N(\rho) := -\operatorname{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log(\lambda_i)$; In classical information (statistical) theory, the Shannon entropy of a probability distribution P is $H_S(P) := -\sum_i P(x_i) \log P(x_i)$. relative entropy (divergence)

Definition 5 (maximally entangled)**.** a state vector is maximually entangled iff the reduced state at one qubit is maximally mixed, i.e., $\operatorname{Tr}_a(|\psi\rangle\langle\psi|) = \frac{1}{2}$.

2. entanglement structures

For multipartite quantum systems, it is crucial to identify not only the presence of entanglement but also its detailed structure. An identification of the entanglement structure may thus provide us with a hint about where imperfections in the setup may occur, as well as where we can identify groups of subsystems that can still exhibit strong quantum-informationprocessing capabilities.

Given a n-qubit quantum system and its partition into m subsystems, the *entanglement structure* indicates how the subsystems are entangled with each other. In some specific systems, such as distributed quantum computing[] quantum networks[] or atoms in a lattice, the geometric configuration can naturally determine the system partition. Therefore, it is practically interesting to study entanglement structure under partitions.

GME is the strongest form of entanglement, that is, all qubits in the system are indeed entangled with each other

Definition 6 (fully entangled)**.** An n-qubit quantum state ρ is a *fully entangled*, if it is outside of the separable state set $S_b^{\mathcal{P}_2}$ for any partition, $\rho \notin S_b^{\mathcal{P}_2}, \forall \mathcal{P}_2 = \left\{A, \bar{A}\right\}$.

Definition 7 (genuine entangled)**.** A state possesses *genuine multipartite entanglement* (GME) if it is outside of S_2, and is (fully) n-separable if it is in S_n. A state possesses $\mathcal{P}$-genuine entanglement if it is outside of $S_b^{\mathcal{P}}$. A state ρ possesses $\mathcal{P}$-genuine entanglement iff $\rho \notin S_b^{\mathcal{P}}$.

Compared with genuine entanglement, multipartite entanglement structure still lacks a systematic exploration, due to the rich and complex structures of n-partite system. Unfortunately, it remains an open problem of efficient entanglement-structure detection of general multipartite quantum states.

Definition 8 (Multipartite state)**.** denote the partition $\mathcal{P}_m = \{A_i\}$ and omit the index m when it is clear from the context.

define fully- and biseparable states with respect to a *specific partition* $\mathcal{P}_m$

Definition 9 (fully separable)**.** An n-qubit pure state $|\psi_f\rangle$ is *fully separable* iff. An n-qubit pure state $|\psi_f\rangle$ is *P-fully separable* iffit can be written as $|\psi_f\rangle = \otimes_i^m |\phi_{A_i}\rangle$. An n-qubit mixed state ρ_f is P-fully separable iff it can be decomposed into a convex mixture of P-fully separable pure states P-bi-separable... $S_f^P \subset S_b^P$

By going through all possible partitions, one can investigate higher level entanglement structures, such as entanglement intactness (non-separability), which quantifies how many pieces in the n-partite state are separated.

Remark 1. P-... can be viewed as generalized versions of regular fully separable, biseparable, and genuinely entangled states, respectively. In fact, when $m = n$, these pairs of definitions are the same. By definitions, one can see that if a state is P_m-fully separable, it must be m-separable. Of course, an m-separable state might not be P_m-fully separable, for example, if the partition is not properly chosen.

entanglement structure measures. To benchmark our technological progress towards the generation of largescale genuine multipartite entanglement, it is thus essential to determine the corresponding entanglement depth.

Definition 10 (Entanglement intactness, depth)**.** the entanglement intactness of a state ρ to be m, iff$\rho \notin S_{m+1}$ and $\rho \in S_m$. k-producible

When the entanglement intactness is 1, the state is genuine entangled; and when the intactness is n, the state is fully separable.

Example 2 (GHZ)**.** bipartite: Bell states; nontrivial multipartite: tripartite. GHZ state: $|\mathrm{GHZ}\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$ (eight-photon) produce the five different entangled states (one from each entanglement structure/partition?):

$$|\mathrm{GHZ}_8\rangle, |\mathrm{GHZ}_{62}\rangle, |\mathrm{GHZ}_{44}\rangle, |\mathrm{GHZ}_{422}\rangle, |\mathrm{GHZ}_{2222}\rangle.$$

Schmidt rank, PPT criteria, entanglement witness

3. Entanglement witness

Theorem 1 ([6])**.** *The weak membership problem for the convex set of separable normalized bipartite density matrices is NP-Hard.* ***Input****: ??*

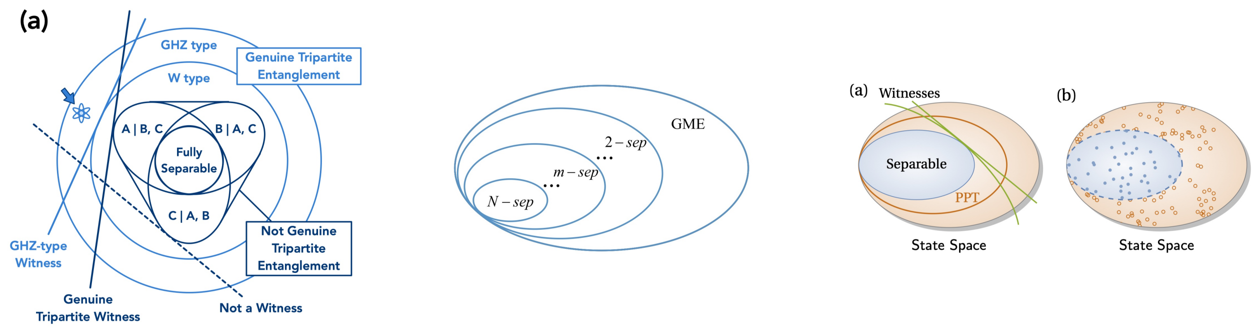


FIG. 1: (a) entanglement witness, PPT criteria, SVM (kernel)?. (c) convex hull...

Question 1. *specific cases? approximately correct? quantum complexity? machine learning (data)?*

Theorem 2 (PPT criterion)**.** *the positive partial transpose (PPT) criterion, saying that a separable state (bipartite separable) must have PPT?. Note, it is only necessary and sufficient when $d_A d_B \leq 6$.*

see Fig. 1

Definition 11 (entanglement witness)**.** Given an (unknown) quantum state (density matrix) ρ, the *entanglement witness* $\hat{W}$ is an obseverable such that

$$\mathrm{Tr}\left(\hat{W}\rho\right) \geq 0, \forall \text{ separable} \; ; \quad \mathrm{Tr}\left(\hat{W}\rho\right) < 0, \; \text{for some entangled} \tag{4}$$

It is natural to ask nonlinear entanglement witness [7] kernel ML

Proposition 1. *Given a state $|\psi\rangle$, the entanglement witness operator $\hat{W}_\psi$ can witness (genuine entangled) genuine multipartite entanglement near $|\psi\rangle$*

$$\hat{W}_\psi = \frac{5}{8}\mathbb{1} - |\psi\rangle\langle\psi| \tag{5}$$

with $\left\langle \hat{W}_\psi \right\rangle \geq 0$ for any separable state in S_b.

If the fidelity (quantum kernel?) of the prepared state ρ_{pre} with the target state $|\psi\rangle$, i.e., $\mathrm{Tr}(\rho_{pre} |\psi\rangle\langle\psi|)$, exceeds $5/8$, ρ_{pre} possesses GME. It is generally difficult to evaluate the quantity $\mathrm{Tr}(\rho_{pre} |\psi\rangle\langle\psi|)$ by the direct projection on $|\psi\rangle$, as it is an entangled state.

Problem 1 (Entanglement witness with prior)**.** with/out prior knowledge

- **Input**: a **known** state $|\psi\rangle$, with noise
- **Output**: separable or not ??? (decision problem?? find problem)

4. Graph state

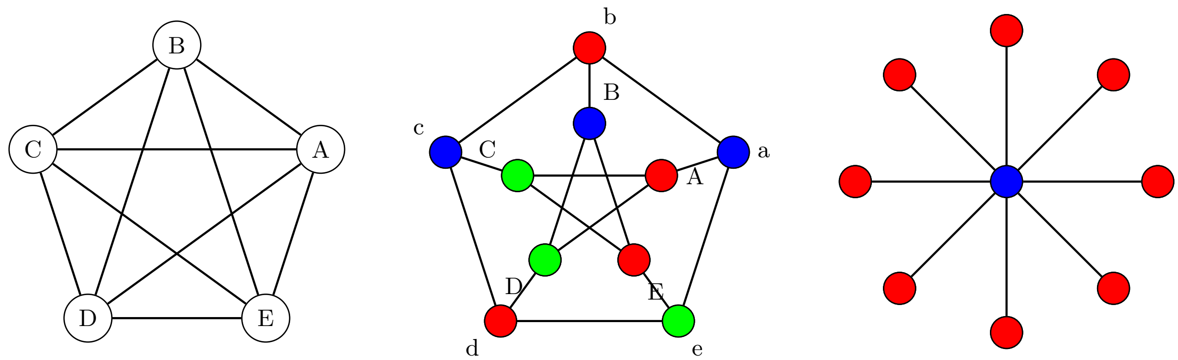
graph state is an important (large?) class of multipartite states in quantum information. Typical graph states include cluster states, GHZ states, and the states involved in error correction (toric code?). It worth noting that 2D cluster state is the universal resource for the measurement based quantum computation (MBQC) [8].

Definition 12 (graph state)**.** Given a simple graph (undirected, unweighted, no loop and multiple edge) $G = (V, E)$, a graph state is constructed as from the initial state $|+\rangle^{\otimes n}$ corresponding to n vertices. Then, apply controlled-Z gate to every edge, that is

$$|G\rangle = \prod_{(i,j)\in E} \mathsf{cZ}_{(i,j)} |+\rangle^{\otimes n} \tag{6}$$

Remark 2. An n-partite graph state can also be uniquely determined by n independent stabilizers, $S_i := X_i \bigotimes_{j\in n} Z_j$, which commute with each other and $\forall i, S_i |G\rangle = |G\rangle$.?? The graph state is the unique eigenstate wtih eignevalue of $+1$ for all the n stabilizers. As a result, a graph state can be writteb as a product of stailizer projectors, $|G\rangle\langle G| = \prod_{i=1}^{n} \frac{S_i+\mathbb{1}}{2}$. stabilizer formalism?;

Example 3 (graph states)**.** line; GHZ (star); complete graph, hypercube, Petersen graph; cluster state



(a) complete graph K_5 (b) Petersen graph (c) star - GHZ

Observation 1. Any connected graph state is fully entangled m-particle state

Problem 2 (Certify entanglement)**.** Multipartite entanglement-structure detection

- **Input**: Given a state close to a **known** (general multipartite) state $|\psi\rangle$, certain partition?
- **Output**: the certified lower-order entanglement among several subsystems could be still useful for some quantum information tasks. entanglement structure

Remark 3. The entanglement entropy $S(\rho_A)$ equals the rank of the adjacency matrix of the underlying bipartite graph, which can be efficiently calculated.

Proposition 2 ([2])**.** *Given a graph state $|G\rangle$ and a partition $\mathcal{P} = \{A_i\}$, the fidelity between $|G\rangle$ and any fully separable is upper bounded by*

$$\mathrm{Tr}(|G\rangle\langle G|\,\rho_f) \leq \min_{\{A,\bar{A}\}} 2^{-S(\rho_A)} \tag{7}$$

where $S(\rho_A)$ is the von Neumann entropy of the reduced density matrix $\rho_A = \mathrm{Tr}_{\bar{A}}(|G\rangle\langle G|)$.

Theorem 3. *k local measurements. Here, k is the chromatic number of the corresponding graph, typically, a small constant independent of the number of qubits.*

Proposition 3 (Entanglement of graph state)**.** *[9]. witness; bounds; graph property? vertex cover? Hamiltona cycle of a graph state?*

generalize [10] stabilizer state, neural network state [11]?

Proposition 4 (Entanglement witness for graph state)**.**

Proposition 5 (Bounds to the Schmidt measure of graph states)**.** *For any graph state $|G\rangle$, the Schmidt measure E_A is bounded from below by the maximal Schmidt rank $SR_{\max}$ and from above by the Pauli persistency PP or the minimal vertex cover, i.e.*

$$SR_{\max}(G) \leq E_S(|G\rangle) \leq PP(G) \leq VC(G). \tag{8}$$

???

B. Shadow tomography

Intuitively, a general tomography [12] that extract (recover) all information about a state requires exponential copies (samples/measurements).

Problem 3 (full tomography)**.** In contrast to shadow tomography, we refer to *full tomography* here

- **Input**: Given a **unknown** N-dimensional mixed state ρ
- **Output**: a complete description? of ρ (decomposition coefficients) with error? Stokes parameter $S_i \equiv \mathrm{Tr}(\hat{\sigma}_i \rho)$

$$\rho = \frac{1}{2^n} \sum_{i_1, i_2, \ldots, i_n = 0}^{3} S_{i_1, i_2, \ldots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \cdots \otimes \hat{\sigma}_{i_n} \tag{9}$$

Theorem 4 (lower bound of full tomography?[13])**.** *Known fundamental lower bounds [66, 73] state that classical shadows of exponential size (at least) $T = \Omega(2^n/\epsilon^2)$ are required to ϵ-approximate ρ in trace distance.*

Definition 13 (fidelity)**.** Given a pair of states (target and prepared), Uhlmann fidelity $F(\rho, \rho') := \mathrm{Tr}\big(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}\big)$

Definition 14 (distance)**.** trace distance $d_{\mathrm{tr}}(\rho, \rho') := \frac{1}{2}\|\rho - \rho'\|_1$

Definition 15 (norm)**.** Schatten p-norm $\|x\|_p := (\sum_i |x_i|^p)^{1/p}$. Euclidean norm l_2 norm; Spectral (operator) norm $\|\mathbf{x}\|_\infty$; Trace norm $\|A\|_{\mathrm{Tr}} \equiv \|A\|_1 := \mathrm{Tr}\Big(\sqrt{A^\dagger A}\Big)$, $p = 1$; Frobenius norm $\|A\|_F := \sqrt{\mathrm{Tr}(A^\dagger A)}$, $p = 2$; Hilbert-Schmidt norm $\|A\|_{HS} := \sqrt{\sum_{i,j} A_{ij}^2}$

Problem 4 (trace estimation)**.** related problems defined as follows

- **Input**: Given an observable (Hermitian) $\hat{O}$ and (copies of) a mixed state ρ ($\rho', \ldots, \rho_m$) in (trace one, Hermitian, PSD),
- **Output**: the expectation value $\Big\langle \hat{O} \Big\rangle = \mathrm{Tr}\Big(\hat{O}\rho\Big)$ with error ϵ in trace distance. entanglement witness, linear function
- **Output**: fidelity $F(\rho, \rho')$, distance
- **Output**: quantum kernel $\mathrm{Tr}(\rho\rho')$
- **Output**: multivariate $\mathrm{Tr}(\rho_1 \cdots \rho_m)$, nonlinear function
- **Output**: shadow tomography $\mathrm{Tr}(E_M \rho) = \mathbb{E}[E_M] = \mathbb{P}[E_i \text{ accept } \rho]$??
- **Output**: full tomography

Nevertheless, we usually only need specific properties of a target state rather than all information about the state. This enables the possibility to . Inspired by Aaronson's shadow tomography [14], Huang et. al [3]

Problem 5 (shadow tomography)**.** *shadow tomography*

- **Input**: an **unknown** N-dimensional mixed state ρ, M known 2-outcome measurements $E_1, \ldots, E_M$
- **Output**: estimate $\mathbb{P}[E_i \text{ accept } \rho]$ to within additive error ϵ, $\forall i \in [M]$, with $\geq 2/3$ success probability.

Theorem 5 ([14])**.** *It is possible to do shadow tomography using $\tilde{\mathcal{O}}(\frac{\log^4 M \cdot \log N}{\epsilon^4})$ copies. [no construction algorithm?] sample complexity lower bound $\Omega(\log(M) \cdot \epsilon^{-2})$,*

$$\{x_l \to \sigma_T(\rho(x_l))\}_{l=1}^{N} \tag{10}$$

$\sigma_T(\rho(x_l))$ is the classical shadow representation of $\rho(x_l)$, a $2^n \times 2^n$ matrix that reproduces $\rho(x_l)$ in expectation over random Pauli measurements.

Definition 16 (classical shadow)**.** classical shadow

$$\rho_{cs} = \mathcal{M}^{-1}\Big(U^\dagger \left|\hat{b}\middle\rangle\middle\langle\hat{b}\right| U\Big) \tag{11}$$

such that we can predict the linear function with classical shadows

$$o_i = \mathrm{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \mathrm{Tr}(O_i \rho) \tag{12}$$

The classical shadow attempts to approximate this expectation value by an empirical average over T independent samples, much like Monte Carlo sampling approximates an integral. The classical shadow size required to accurately approximate all reduced r-body density matrices scales exponentially in subsystem size r, but is independent of the total number of qubits n.

Algorithm II.0: Classical Shadow (tomography)

```
input  : a density matrix ρ, ..
output: classical shadow
for i = 1, 2, ..., m do
    random Pauli measurements                    // a comment
    return "?"
return?
```

Lemma 1. *the variance*

$$\mathrm{Var}[o] = \mathbb{E}[(o - \mathbb{E}[o])^2] \leq \left\| O - \frac{\mathrm{Tr}(O)}{2^n}\mathbb{1} \right\|^2_{shadow} \tag{13}$$

sample complexity

$$N_{tot} = \mathcal{O}\left(\frac{\log(M)}{\epsilon^2} \max_{1\leq i\leq M} \left\| O_i - \frac{\mathrm{Tr}(O_i)}{2^n}\mathbb{1} \right\|^2_{\text{shadow}}\right) \tag{14}$$

Theorem 6 (Pauli/Clifford measurements)**.** *additive error ϵ, M arbitrary k-local linear function $\mathrm{Tr}\Big(\hat{O}_i\rho\Big)$, $\Omega(\log(M)3^k/\epsilon^2)$ copies of the state ρ.*

III. CLASSICAL, DATA-POWERED, AND QUANTUM ALGORITHMS

We consider the problem, entanglement structure dectection for graph states

Problem 6 (?data)**.** problem without training data

- **Input**: a graph G encoding in a graph state $|G\rangle$; adjacency matrix?
- **Output**: entanglement structure

with training data: **features**: classical shadow? raw data? quantum data, label: entangled?
input encoding (model):

- amplitude encoding: given a normalized vector $\mathbf{x} \in \mathbb{R}^d$, the quantum state $|\mathbf{x}\rangle = \sum_z^d x_z |z\rangle$. need $\log(d)$ qubits for a data point; dequantization [15]
- quantum data $|\psi\rangle$ or ρ: quantum state from real-world experiments or quantum circuits $\hat{U}$. no input problem? more efficient?
- graph state encoding: graph state, discrete, efficient? space (time), isomorphism?

A. Classical algorithms

1. Classical shadow and kernel methods

separability classifier by neural network [16]. rigorous quantum advantage of quantum kernel method in SVM [17]. classical machine learning with classical shadow [18].

Graphs is another kind of data which is fundamentally different from a real value vector because of vertex-edge relation, graph isomorphism. So, graph kernel [19] need additional attention.

Definition 17 (graph kernel)**.** given a pair of graphs (G, G'), *graph kernel* is $k(G, G') =$. quantum graph kernel $k(G, G') = |\langle G|G'\rangle|^2$?? [20]

Definition 18 (quantum kernel)**.** quantum kernel with quantum feature map $\phi(\mathbf{x}) : \mathcal{X} \to |\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x})|$

$$k_Q(\rho, \rho') := |\langle \phi(\mathbf{x})|\phi(\mathbf{x}')\rangle|^2 = \left| \langle 0|\hat{U}^\dagger_{\phi(\mathbf{x})} \hat{U}_{\phi(\mathbf{x}')}|0\rangle \right|^2 = \mathrm{Tr}(\rho\rho') \tag{15}$$

where $\hat{U}_{\phi(\mathbf{x})}$ is a quantum circuit or physics process that encoding an input $\mathbf{x}$. In quantum physics, quantum kernel is also known as transition amplitude (quantum propagator);

Definition 19 (shadow kernel)**.** given two density matrices (quantum states) ρ and ρ', *shadow kernel* [3] is

$$k_{\text{shadow}}(S, S') := \exp\left(\sum \exp\left(\sum\right)\right) \tag{16}$$

Definition 20 (neural tagent kernel)**.** neural tagent kernel [21]: proved to be equivalent to deep neural network [11] in the limit ...

$$k_{\text{NTK}}\left(S_T(\rho_l), \tilde{S}_T(\rho_{l'})\right) = \left\langle \phi^{(NTK)}(S_T(\rho_l)), \phi^{(NTK)}(\tilde{S}_T(\rho_l)) \right\rangle \tag{17}$$

similarity measures? advantages? why? (isomorphism?)

Definition 21 (divergence)**.** KL divergence (relative entropy): measure the distance (similarity) between two probability distributions:

$$D_{\text{KL}}(P||Q) := \sum P(x) \log(P(x)/Q(x)) \tag{18}$$

symmetric version: Jensen-Shannon divergence (machine learning)

$$D_{\text{JS}}(P||P') := \frac{1}{2}(D_{\text{KL}}(P||M) + D_{\text{KL}}(P'||M)) \equiv H_S(M) - \frac{1}{2}(H_S(P) - H_S(P')) \tag{19}$$

where $M = (P + P')/2$ and Shannon entropy H_S. Analogously, quantum Jensen-Shannon divergence D_{QJS}

power of data -

Theorem 7 (informal [4])**.** *data learning*

- *machine learning (strictly) more powerful than BPP*
- *exist quantum advantage in machine learning (not significant, practical)*

Algorithm III.1: Classical learning (SVM) + classical shadow

input : labeled features (data), classical shadow?
output: entanglement structure? decision

```
for i = 1, 2, ..., m do
    kernel estimation                                  // a comment
    return "?"
return ?
```

B. Quantum trace (kernel) estimation

Theorem 8 ([5]). *multivariate trace estimation can be implemented in constant quantum depth, with only linearly-many controlled two-qubit gates and a linear amount of classical pre-processing*

C. Variational (hybrid) quantum algorithms

1. Variational quantum kernel estimation (hybrid)

an ansatz for entanglement witness

$$\hat{W}_a := \sum_{\{i\dots\}} a_{..} \bigotimes_{i}^{n} \hat{\sigma}^{(i)}, \quad \hat{\sigma} \in \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z, \mathbb{1}_{2\times 2}\} \tag{20}$$

Algorithm III.2: Entanglement witness by ...

input : (copies of) density matrix ρ
output: determine entangled structure??

```
for i = 1, 2, ..., m do
    Ŵ_i                                  // this is a comment
    return "separable?"
return entangled ?
```

2. Variational trace estimate (direct)

find optimal entanglement witness (qunatum circuit?) [22] [23] [24]

D. Theoretic upper bounds and lower bounds

[4] [3] [14] [25] [17]

Definition 22 (graph property)**.** monotone

Problem 7 (graph property test)**.**

	gate/depth/computation	measurements/samples	query?	necessary?sufficient
shadow tomography		Theorems 5 and 6	N/A	
indirect? direct (no prior), promise				
entanglement witness (Section II A 3)		constant	convex?	
classical ML + SC (Section III A 1)				
quantum (variational) circuits	c-depth?			

TABLE I: complexity measures of different methods

1. Separations

contrived problem (engineered dataset)? for exponential speedup

2. Obstacles

IV. NUMERICAL SIMULATION

A. Classification accuracy

1. Data preparation

multi-partite entangled state: generate synthetic (engineered) data from (random graph?). separable state from randomly

We consider a set of different regularization parameters,

2. Results

performance of different methods:

B. Robustness to noise

tradeoff between (white noise) tolerance (robustness) and efficiency (number of measurements).

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}}) |G\rangle\langle G| + p_{\text{noise}} \frac{\mathbb{1}}{2^n} \tag{21}$$

p_{noise} indicates the robustness of the algorithm (witness).

Remark 4. the largest noise tolerance p_{limit} just related to the **chromatic number** of the graph.[??] graph property

V. CONCLUSION AND DISCUSSION

todo: experiment (generation, verification) [26]

Acknowledgements

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), arXiv:quant-ph/0702225.
[2] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, npj Quantum Inf **5**, 83 (2019).
[3] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].
[4] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Nat Commun **12**, 2631 (2021), arXiv:2011.01938 [quant-ph].
[5] Y. Quek, M. M. Wilde, and E. Kaur, Multivariate trace estimation in constant quantum depth (2022), arXiv:2206.15405 [hep-th, physics:quant-ph].
[6] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement (2003), arXiv:quant-ph/0303055.
[7] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).
[8] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, Nature Phys **5**, 19 (2009), arXiv:0910.1116.
[9] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. V. den Nest, and H.-J. Briegel, Entanglement in Graph States and its Applications (2006), arXiv:quant-ph/0602096.
[10] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, Phys. Rev. A **103**, 052426 (2021), arXiv:2012.07606 [quant-ph].
[11] X. Gao and L.-M. Duan, Nat Commun **8**, 662 (2017), arXiv:1701.05039 [cond-mat, physics:quant-ph].
[12] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.
[13] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inform. Theory , 1 (2017).

[14] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018) pp. 325–338, arXiv:1711.01053.
[15] E. Tang, Phys. Rev. Lett. **127**, 060503 (2021), arXiv:1811.00414 [quant-ph].
[16] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Phys. Rev. A **98**, 012315 (2018), arXiv:1705.01523 [quant-ph].
[17] Y. Liu, S. Arunachalam, and K. Temme, Nat. Phys. **17**, 1013 (2021), arXiv:2010.02174 [quant-ph].
[18] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems (2021), arXiv:2106.12627 [quant-ph].
[19] N. M. Kriege, F. D. Johansson, and C. Morris, Appl Netw Sci **5**, 6 (2020), arXiv:1903.11835 [cs, stat].
[20] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, Pattern Recognition **48**, 344 (2015).
[21] A. Jacot, F. Gabriel, and C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks (2020), arXiv:1806.07572 [cs, math, stat].
[22] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, Covariant quantum kernels for data with group structure (2021), arXiv:2105.03406 [quant-ph].
[23] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019), arXiv:1804.11326.
[24] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019), arXiv:1803.07128 [quant-ph].
[25] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **126**, 190505 (2021), arXiv:2101.02464 [quant-ph].
[26] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, Y.-C. Liang, X. Ma, Y.-A. Chen, and J.-W. Pan, Phys. Rev. X **8**, 021072 (2018).

Appendix A: Machine learning background

1. SVM

Definition 23 (SVM)**.** support vector machine (SVM) is designed to find a hyperplane (a linear function) such that maximize the margin ...

kernel nonlinear boundary. map to a higher dimensional (feature) space, in which data is linearly separable.

Definition 24 (kernel)**.** In general, the kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ measures the similarity between two input data points by an inner product

$$k(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \tag{A1}$$

If the input $\mathbf{x} \in \mathbb{R}^d$ (conventional machine learning task, e.g., image classification), the feature map $\phi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^n$ ($d < n$) from a low dimensional space to a higher dimensional space. The corresponding kernel (Gram) matrix $\mathbf{K}$ should be a positive, semidefinite (PSD) matrix.

Example 4 (kernels)**.** Some common kernels: the polynomial kernel $k_{\mathrm{poly}}(\mathbf{x}, \mathbf{x}') := (1 + \mathbf{x} \cdot \mathbf{x}')^q$ with feature map ... the Gaussian kernel $k_{\mathrm{gaus}}(\mathbf{x}, \mathbf{x}') := \exp\left(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2\right)$ with an infinite dimensional feature map $\phi(\mathbf{x})$. An important feature of kernel method is that kernels can be computed efficiently without evaluating feature map (might be infinite dimension) explicitly.

Appendix B: Hardness assumptions

BQP,BPP