

Towards efficient and robust detection of entanglement structure

Jue Xu* and Qi Zhao†

(Dated: August 12, 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). We illustrate the advantages and limitations of machine learning and quantum algorithms.

CONTENTS

I. Introduction	1
II. Preliminaries	2
A. Entanglement detection	2
B. Tomography and trace estimation	7
III. Classical, data-powered, and quantum algorithms	8
A. Classical shadow and machine learning	9
B. Quantum trace (kernel) estimation	11
C. Variational (hybrid) quantum algorithms	12
D. Theoretic upper bounds and lower bounds	12
IV. Numerical simulation	13
A. Classification accuracy	13
B. Robustness to noise	13
C. Experiments	13
V. Conclusion and discussion	14
Acknowledgements	14
References	14
A. Machine learning background	15
1. Support vector machine	15
2. Neural network	15
B. Hardness assumptions	16

I. INTRODUCTION

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

* juexu@cs.umd.edu

† email

II. PRELIMINARIES

Notations: 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 (Hermitian matrices). Denote vector (matrix) \mathbf{x} , \mathbf{K} by boldface font. A simple (undirected, unweighted) graph $G = (V, E)$ is described by vertices V and edges E .

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 $\{|\mathbf{x}\rangle \equiv |x_1\rangle|x_2\rangle\cdots|x_n\rangle\}$ 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. shorthand $|\psi_A\rangle|\psi_B\rangle \equiv |\psi_A\rangle \otimes |\psi_B\rangle$. Hadamard basis $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$.

A. Entanglement detection

1. Bipartite entanglement

Large scale entanglement is the (main) resource of quantum advantages in quantum computation and communication. Firstly, we consider the simplest entanglement structure: bipartite separable case.

Definition 1 (density matrix). pure state $|\psi\rangle$; density matrix ρ (trace one, Hermitian, PSD) ...

Definition 2 (partial trace). partial trace; reduced density matrix $\rho_A = \text{Tr}_B(\rho_{AB})$

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

Definition 3 (bipartite separable). A pure state is (bi-)separable if it is in a tensor product form $|\psi_b\rangle = |\phi_A\rangle \otimes |\phi_B\rangle$, where $\mathcal{P}_2 = \{A, B\}$ 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??.

Consider a bipartite system AB with the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, where \mathcal{H}_A has dimension d_A and \mathcal{H}_B has dimension d_B , respectively. A state ρ_{AB} is *separable* if it can be written as a convex combination $\rho_{AB} = \sum_i \lambda_i \rho_{A,i} \otimes \rho_{B,i}$ with a probability distribution $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Otherwise, ρ_{AB} is entangled.

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 $S_b^{\mathcal{P}_2}$. entangled state?...

Rather than qualitatively determining (bi)separability, there are measures to quantify entanglement

Definition 4 (Schmidt measure). Consider the following bipartite pure state, written in Schmidt form:

$$|\psi\rangle = \sum_i^r \sqrt{p_i} |\phi_i^A\rangle \otimes |\phi_i^B\rangle \quad (1)$$

where $\{|\phi_i^A\rangle\}$ is a basis for \mathcal{H}_A and $\{|\phi_i^B\rangle\}$ for \mathcal{H}_B . 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 rank ?? $\text{SR}^A(\psi) = \text{rank}(\rho_\psi^A)$) Schmidt measure is minimum of $\log_2 r$ where r is number of terms in an expansion of the state in product basis.

Definition 5 (entropy). In quantum mechanics (information), the von Neumann *entropy* of a density matrix is $H_N(\rho) := -\text{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log(\lambda_i)$; 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](#))

Remark 1. (def partial trace ...) a pure (bipartite) state is entangled iff the reduced state $\rho^A = \text{Tr}_B(\rho)$ is mixed. The mixedness of this reduced state allows one to quantify the amount of entanglement in this state.

Definition 6 (entanglement entropy). The bipartite *von Neumann entanglement entropy* S is defined as the von Neumann entropy of either of its reduced density matrix ρ_A . For a pure state $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$, it is given by

$$E(\Psi_{AB}) = S(\rho_A) = -\text{Tr}(\rho_A \log \rho_A) = -\text{Tr}(\rho_B \log \rho_B) = S(\rho_B) \quad (2)$$

where $\rho_A = \text{Tr}_B(\rho_{AB})$ and $\rho_B = \text{Tr}_A(\rho_{AB})$ are the reduced density matrices for each partition. With Schmidt decomposition (Eq. (1)), the entropy of entanglement is simply $-\sum_i p_i^2 \log(p_i)$. the n th Renyi entropy, $S_n = \frac{1}{n-1} \log(R_n)$ where $R_n = \text{Tr}(\rho_A^n)$

Definition 7 (maximally entangled). a state vector is *maximally entangled* iff the reduced state at one qubit is maximally mixed, i.e., $\text{Tr}_A(|\psi\rangle\langle\psi|) = \frac{1}{2}$.

Classically, the hardness of determining the bipartite separability.

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

However, we do not know approximately correct complexity? quantum complexity? machine learning (data)? for entanglement () problem? multipartite?

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 $d_A d_B \leq 6$. Another widely used one is the k -symmetric extension hierarchy [15, 16], which is presently one of the most powerful criteria, but hard to compute in practice due to its exponentially growing complexity with k .[??]*

2. Multipartite 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.

Definition 8 (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, $\rho \notin S_b^{\mathcal{P}_2}, \forall \mathcal{P}_2 = \{A, \bar{A}\}$.

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

Definition 9 (genuine multipartite entanglement). 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 10 (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 11 (fully separable). An n -qubit pure state $|\psi_f\rangle$ is *fully separable* iff. An n -qubit pure state $|\psi_f\rangle$ is \mathcal{P} -fully separable iff it can be written as $|\psi_f\rangle = \otimes_i^m |\phi_{A_i}\rangle$. An n -qubit mixed state ρ_f is \mathcal{P} -fully separable iff it can be decomposed into a convex mixture of \mathcal{P} -fully separable pure states.

$$\rho_f = \sum_i p_i |\psi_f^i\rangle\langle\psi_f^i|, (\forall i)(p_i \geq 0, \sum_i p_i = 1). \quad (3)$$

P-bi-separable... $S_f^{\mathcal{P}} \subset S_b^{\mathcal{P}}$

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 2. 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 12 (Entanglement intactness, depth). the entanglement intactness of a state ρ to be m , iff $\rho \notin S_{m+1}$ and $\rho \in S_m$. When the entanglement intactness is 1, the state possesses **genuine multipartite entanglement**; and when the intactness is n , the state is **fully separable**. k -producible.

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

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

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

Schmidt rank, PPT criteria, entanglement witness

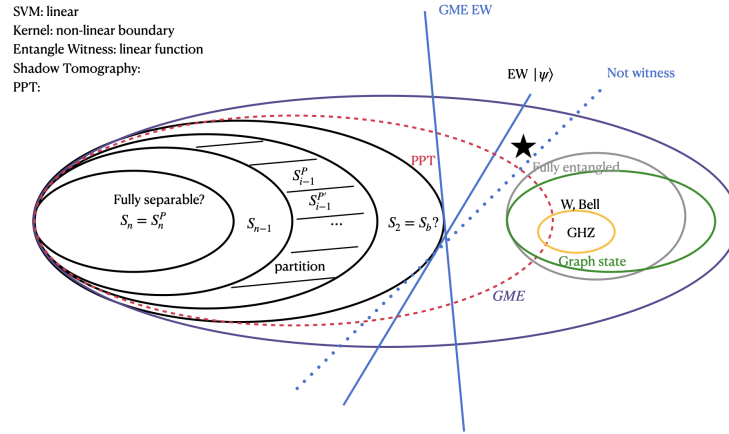


FIG. 1: (a) **entanglement witness**, **PPT criterion**, **SVM** (kernel)?. convex hull...

3. Entanglement witness

see Fig. 1 for relations. entanglement detection [7].

Definition 13 (entanglement witness). Given an (unknown? known target state) quantum state (density matrix) ρ , the *entanglement witness* \hat{W} is an observable such that

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

In a typical experiment one aims to prepare a pure state, $|\psi\rangle$, and would like to detect it as true multipartite entangled. While the preparation is never perfect, it is still expected that the prepared mixed state is in the proximity of $|\psi\rangle$. The usual way to construct entanglement witnesses using the knowledge of this state is

$$\hat{W}_\psi = c\mathbb{1} - |\psi\rangle\langle\psi| \quad (5)$$

where c is the smallest constant such that for every product state $\text{Tr}(\rho\hat{W}) \geq 0$

Remark 3. In order to measure the witness in an experiment, it must be decomposed into a sum of locally measurable operators. The number of local measurements in these decompositions seems to increase exponentially with the number of qubits.[??]

131 **Example 3** (entanglement witness for GHZ). three-qubit GHZ state [8]

$$\hat{W}_{\text{GHZ}_3} := \frac{3}{2}\mathbb{1} - \hat{\sigma}_x^{(1)}\hat{\sigma}_x^{(2)}\hat{\sigma}_x^{(3)} - \frac{1}{2}\left[\hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(2)} + \hat{\sigma}_z^{(2)}\hat{\sigma}_z^{(3)} + \hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(3)}\right] \quad (6)$$

132 This witness requires the measurement of the $\{\hat{\sigma}_x^{(1)}, \hat{\sigma}_x^{(2)}, \hat{\sigma}_x^{(3)}\}$ and $\{\hat{\sigma}_z^{(1)}, \hat{\sigma}_z^{(2)}, \hat{\sigma}_z^{(3)}\}$ settings. The projector based
 133 witness $\hat{W}_{\text{GHZ}_3} = \mathbb{1}/2 - |\text{GHZ}\rangle\langle\text{GHZ}|$ requires four measurement settings. detect genuine n -qubit entanglement close
 134 to GHZ_n

$$\hat{W}_{\text{GHZ}_n} = (n-1)\mathbb{1} - \sum_{k=1}^n S_k^{(\text{GHZ}_n)} \quad (7)$$

135 S_k is the stabilizer ... [9] Detecting Genuine Multipartite Entanglement with Two Local Measurements [8]

136 **Question 1.** how far/close to the target state (entanglement witness), noise limit?

137 It is natural to ask how nonlinear entanglement witness [10] and the [kernel](#) method (nonlinear boundary) in machine
 138 learning can be applied.

139 **Proposition 1.** Given a state $|\psi\rangle$, the *entanglement witness* operator \hat{W}_ψ can witness *genuine multipartite entangle-*
 140 *ment* near $|\psi\rangle$ with $c = 5/8$ in Eq. (5) that is, $\langle \hat{W}_\psi \rangle \geq 0$ for any separable state in S_b .

141 If the [fidelity](#) (quantum kernel?) of the prepared state ρ_{pre} with the target state $|\psi\rangle$, i.e., $\text{Tr}(\rho_{\text{pre}} |\psi\rangle\langle\psi|)$, exceeds
 142 $5/8$, ρ_{pre} possesses GME. It is generally difficult to evaluate the quantity $\text{Tr}(\rho_{\text{pre}} |\psi\rangle\langle\psi|)$ by the direct projection on
 143 $|\psi\rangle$, as it is an entangled state.

144 A usual approach for detecting entanglement is using Bell inequalities [??]

145 **Definition 14** (Bell inequality). CHSH inequality (game) ...; Bell inequalities for graph states $|\sum_{\sigma \in S} \langle \sigma \rangle| \leq C?...$

146 Another approach for detecting multipartite entanglement is using entanglement witnesses. Different Bell inequal-
 147 ities can be regarded as entanglement witness for different types of entanglement in a multi-party entangled state.
 148 These witnesses can be quite useful to detect entanglement in the vicinity of graph states.

149 **Problem 1** (separability). given (input) an **unknown** state, to determine (output) separability.

150 **Problem 2** (Entanglement witness with prior). Entanglement witness

- 151 • **Input:** a **known** state $|\psi\rangle$, with noise
- 152 • **Output:** separable or not ??? S_f^P ? S_b^P difficulty: multi(n)-partite, high-dimensional (qudit) [11], pure/mixed
 153 state, with/out prior knowledge, universal?

154 **Remark 4** (universal entanglement witness). [11] For example, a witness specifically designed for a four-qubit compact
 155 cluster state [16] confirms, when its expectation value is negative, the presence of that particular state having a very
 156 specific density function, while a positive measured expectation value of that operator only provides information that
 157 the tested state is not a compact cluster state. Indeed, the same witness, if applied to a four-qubit linear cluster or
 158 Greenberger-Horne-Zeilinger (GHZ) [17] states, would result in a positive measured expectation value, even though
 159 these two states are both highly entangled [17, 18]. Hence, a witness is a threshold test that can only detect the
 160 presence of a specific state. In contrast to an entanglement monotone (e.g. the entanglement entropy [6]), which
 161 determines the amount of entanglement, a witness cannot be used to quantify entanglement.

162 **Problem 3** (Certify entanglement). Multipartite entanglement-structure detection

- 163 • **Input:** an (actual) state ρ' from experiment that is close to a **known/target** (general multipartite) state $|\psi\rangle$,
 164 certain partition?
- 165 • **Output:** the certified lower-order entanglement among several subsystems could be still useful for some quantum
 166 information tasks. entanglement structure

4. Graph state

graph state is an important (large?) class of multipartite states in quantum information. Typical graph states include cluster (lattice) 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) [12].

Definition 15 (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

$$|G\rangle := \prod_{(i,j) \in E} cZ_{(i,j)} |+\rangle^{\otimes n} \quad (8)$$

Remark 5. An n -partite(qubit) graph state can also be uniquely determined by n independent stabilizers, $S_i := X_i \otimes_{j \in n} Z_j$, which commute with each other and $\forall i, S_i |G\rangle = |G\rangle$. The graph state is the unique eigenstate with eigenvalue of $+1$ for all the n stabilizers. As a result, a graph state can be written as a product of stabilizer projectors, $|G\rangle\langle G| = \prod_{i=1}^n \frac{S_i + 1}{2}$. stabilizer formalism?;

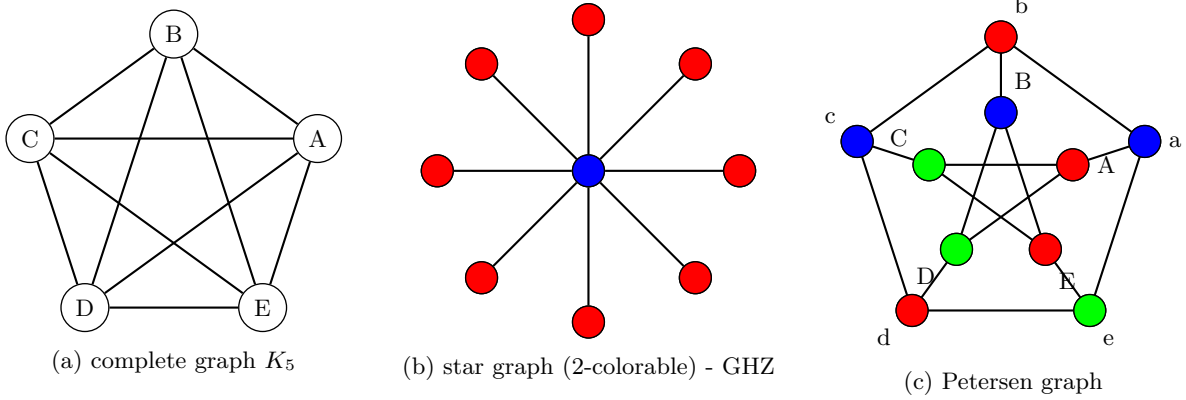


FIG. 2: graph states

Example 4 (graph states). Any connected graph state is fully entangled state. The GHZ state corresponds to the star graph and the complete graph (Fig. 2). This is easily seen by applying Hadamard unitaries $\hat{U}_H^{V \setminus a}$ to all but one qubit a in the GHZ-state, which yields the star graph state with a as the central qubit. (line, ring; hypercube, Petersen graph; cluster state in two dimensions, which corresponds to a rectangular lattice.) The Petersen graph is not LC-equivalent to its isomorphism (exchanging the labels at each end of the five “spokes”). However, the lists of Schmidt ranks (or, equivalently, the connectivity functions) of these graphs coincide. The class of CSS (error correction) states corresponds to the class of 2-colorable graphs. [13]

Remark 6 (?). The entanglement entropy $S(\rho_A)$ equals the rank of the adjacency matrix of the underlying bipartite graph, which can be efficiently calculated. For graph states, the reduced density matrices can be represented efficiently in terms of their stabilizer elements or their adjacency matrix.

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

$$\text{Tr}(|G\rangle\langle G| \rho_f) \leq \min_{\{A, \bar{A}\}} 2^{-S(\rho_A)} \quad (9)$$

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

Remark 7. LU, LC equivalence, local operations and classical communication (LOCC)

the entanglement in a graph state is related to the topology of its underlying graph.

Proposition 3 (Entanglement of graph state). [13]. *witness; bounds; graph property? vertex cover? Hamiltonian cycle of a graph state?*

generalize [14] stabilizer state, neural network state [15]?

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

$$\hat{W} = \frac{C}{2^N} \mathbb{1}_V - |G\rangle\langle G| \quad (10)$$

Let $|G\rangle$ be a graph state corresponding to a connected graph. Then

$$\hat{W}_1^{ab} = \mathbb{1}_V - K_a - K_b \quad (11)$$

is an entanglement witness for the $|G\rangle$ that detects entanglement in the reduced state $\rho_G^A (A = N_a \cup N_b \cup \{a, b\})$ with only two measurement settings and thus can rule out full separability of the total graph state???. The entanglement witness

$$\hat{W}_2 = (N-1)\mathbb{1}_V - \sum_{a \in V} K_a \quad (12)$$

detects *genuine multipartite entanglement*.

Question 2. for which case, C is hard to compute? non-stabilizer state? SWAP?

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

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.

$$\text{SR}_{\max}(G) \leq E_S(|G\rangle) \leq PP(G) \leq VC(G). \quad (13)$$

???

B. Tomography and trace estimation

The brute force approach is to fully characterize a system by performing quantum state tomography and calculating separability measures from the recovered density matrix. Intuitively, a general tomography [16] that extract (recover) all information of a state requires exponential copies (samples/measurements).

Problem 4 (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 $S_i \equiv \text{Tr}(\hat{\sigma}_i \rho)$

$$\rho = \frac{1}{2^n} \sum_{i_1, i_2, \dots, i_n=0}^3 S_{i_1, i_2, \dots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \dots \otimes \hat{\sigma}_{i_n} \quad (14)$$

However, tomography is experimentally and computationally demanding; for a state consisting of N particles, with each residing in a d -dimensional Hilbert space, we would have to perform $M = \mathcal{O}(d^{2N})$ measurements.

Theorem 4 (lower bound of *full tomography*?[17]). 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 16 (fidelity). Given a pair of states (target ρ and prepared ρ'), Uhlmann fidelity $F(\rho, \rho') := \text{Tr}(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}) \equiv \|\sqrt{\rho}\sqrt{\rho'}\|_1$, where $\sqrt{\rho}$ denotes the positive semidefinite square root of the operator ρ . For any mixed state ρ and pure state $|\psi\rangle$, $F(\rho, |\psi\rangle\langle\psi|) = \sqrt{\langle\psi|\rho|\psi\rangle} \equiv \sqrt{\text{Tr}(\rho|\psi\rangle\langle\psi|)}$ which can be obtained by the Swap-test[?].

Definition 17 (distance). For mixed states, trace distance $d_{\text{tr}}(\rho, \rho') := \frac{1}{2} \|\rho - \rho'\|_1$. For pure states, $d_{\text{tr}}(|\psi\rangle, |\psi'\rangle) := \frac{1}{2} \|\psi\psi^\dagger - \psi'\psi'^\dagger\|_1 = \sqrt{1 - |\langle\psi|\psi'\rangle|^2}$. fidelity and trace distance are related by the inequalities

$$1 - F \leq D_{\text{tr}}(\rho, \rho') \leq \sqrt{1 - F^2} \quad (15)$$

Definition 18 (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 $\|A\|_{\text{Tr}} \equiv \|A\|_1 := \text{Tr}(|A|) \equiv \text{Tr}(\sqrt{A^\dagger A})$, $|A| := \sqrt{A^\dagger A}$, $p = 1$; Frobenius norm $\|A\|_F := \sqrt{\text{Tr}(A^\dagger A)}$, $p = 2$; Hilbert-Schmidt norm $\|A\|_{HS} := \sqrt{\sum_{i,j} A_{ij}^2}$

In quantum mechanics, interesting properties are often linear functions of the underlying density matrix ρ . For example, the fidelity with a pure target state, entanglement witnesses fit this framework.

Problem 5 (trace estimation). related problems defined as follows

- **Input:** Given an observable (Hermitian) \hat{O} and (copies of) a mixed state ρ or several states (ρ', \dots, ρ_m) ,
- **Output:** with error ϵ measured by trace distance (fidelity...)
 - linear function (mostly): the expectation value $\langle \hat{O} \rangle = \text{Tr}(\hat{O}\rho)$; entanglement witness $\text{Tr}(\hat{W}\rho)$; shadow tomography $\text{Tr}(E_M \rho) = \mathbb{E}[E_M] = \mathbb{P}[E_i \text{ accept } \rho]??$; full tomography; two-point correlation $\langle O_i O_j \rangle$
 - nonlinear function: entropy (non-linear); quadratic $\text{Tr}(\hat{O}\rho_i \otimes \rho_j)$; fidelity $F(\rho, \rho')$, distance??;
 - multivariate: quantum kernel $\text{Tr}(\rho\rho')$; multivariate $\text{Tr}(\rho_1 \cdots \rho_m)$, nonlinear function?? linear;

Nevertheless, we usually only need specific properties of a target state rather than full classical descriptions about the state. This enables the possibility to shadow tomography.

Problem 6 (shadow tomography). *shadow tomography*

- **Input:** an unknown N -dimensional mixed state ρ , M known 2-outcome measurements E_1, \dots, 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 (bounds of shadow tomography [18]). *It is possible to do shadow tomography using $\tilde{O}(\frac{\log^4 M \cdot \log N}{\epsilon^4})$ copies. [no construction algorithm?] sample complexity lower bound $\Omega(\log(M) \cdot \epsilon^{-2})$,*

more details in Section III A 1

Remark 8 (compare shadow tomography and classical shadow [3]). While very efficient in terms of samples, Aaronson's procedure is very demanding in terms of quantum hardware — a concrete implementation of the proposed protocol requires **exponentially long quantum circuits** that act collectively on all the copies of the unknown state stored in a quantum memory??

III. CLASSICAL, DATA-POWERED, AND QUANTUM ALGORITHMS

In this paper, we focus on the entanglement structure detection for graph states.

Problem 7 (detect graph state entanglement structure?). problem without training data

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

with training data: **features:** classical shadow? raw data? quantum data, label: entangled?

Definition 19 (input model). several common input (encoding) models in quantum algorithms:

- **amplitude encoding:** given a normalized vector $\mathbf{x} \in \mathbb{R}^d$, the quantum state $|\mathbf{x}\rangle = \sum_z x_z |z\rangle$. need $\log(d)$ qubits for a data point; In general, it is hard to prepare such state. (subject to dequantization [19]). typical encoding method for quantum machine learning for classical problems (such as image classification).

- **unitary encoding:** quantum simulation (Hamiltonian); quantum random walk (adjacency matrix); oracle (controlled) unitary, e.g., quantum phase estimation; [graph state](#) encoding (discrete, efficient in space/time?, isomorphism?)

- **quantum data:** quantum state $|\psi\rangle$ or ρ from real-world experiments or designed quantum circuits \hat{U} . no input problem? more efficient? for quantum algorithms

Question 3. *how to relate graph state entanglement to [graph property test](#) ..??.*

Definition 20 (graph property). monotone ...

Example 5 (colorable). k -colorable is a graph property, i.e., allow for a coloring of the vertices with k colors such that no two adjacent vertices have the same color. A graph is bipartite iff 2-colorable. other graph properties: isomorphism; vertex cover; Hamiltonian cycle ...

Problem 8 (graph property test). **promise:** the input graph either has a property, or is ϵ -far from having the property, meaning that we must change at least an ϵ fraction of the edges to make the property hold.

Theorem 6 (bounds for graph property test).

Question 4. [20] *Is there any graph property which admits an exponential quantum speed-up? [21] depends on input model (query adjacency matrix/list)*

Graphs is another kind of data which is fundamentally different from a real value vector because of vertex-edge relation and graph isomorphism. So, graph kernel [22] need additional attention.

Definition 21 (graph kernel). given a pair of graphs (G, G') , *graph kernel* is $k(G, G') =$. quantum graph kernel $k(G, G') = |\langle G|G' \rangle|^2$?? [23]

A. Classical shadow and machine learning

related works

- separability classifier by classical neural network [24]: input: sythetic random density matrices; output: a classical classifier for [bipartite separable](#) (independent of state??). The idea is to feed the classifier by a large amount of sampled trial states (feature: synthetic density matrix with noise flatten as a real vector $\mathbf{x} \in \mathbb{R}^{d_A^2 d_B^2 - 1}$) as well as their corresponding class labels (separable or entangled by [PPT criterion](#), CHA), and then train the classifier to predict the class labels of new states that it has not encountered before. Previous methods **only detect a limited part of the state space**, e.g. different entangled states often require different [entanglement witness](#). In contrast, this classifier can handle a variety?? of input states once properly trained Fig. 1.
- [classical shadow](#) [3]: estimate entanglment witness (know target state, tripartite GHZ)
- classical SVM: An SVM allows for the construction of a hyperplane $\langle \hat{W} \rangle = \sum_k w_k \mathbf{x}_k$ that clearly delineates between separable states and the target entangled state (bipartite and **tripartite qubit and qudit**); this hyperplane is a **weighted sum of observables ('features')** whose coefficients are optimized during **the training of the SVM**. This method the ability to obtain witnesses that require only local measurements even when the target state is a **non-stabilizer state** W state (normally need nonlocal measurements). feature: $\mathbf{x}_{\vec{k}}$ expectation of Pauli strings. the training of an SVM is convex; if a solution exists for the given target state and ansatz, the optimal SVM 2 will be found. this SVM formalism allows for the programmatic removal of features, i.e., reducing the number of experimental measurements, in exchange for a lower tolerance to white noise, in a manner similar to [??].
- classical machine learning (SVM, NN) with [classical shadow](#) [25]??

1. Classical shadow

Inspired by Aaronson's shadow tomography [18], Huang et. al [3] introduce classical shadow. A classical shadow is a succinct classical description of a quantum state, which can be extracted by performing reasonably simple single-copy measurements on a reasonably small number of copies of the state. 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.

304 **Definition 22** (classical shadow). classical shadow ρ_{cs}

$$\rho_{cs} = \mathcal{M}^{-1} \left(U^\dagger \left| \hat{b} \right\rangle \left\langle \hat{b} \right| U \right) \quad (16)$$

305 such that we can predict the linear function with classical shadows

$$o_i = \text{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \text{Tr}(O_i \rho) \quad (17)$$

306 $\sigma_T(\rho(x_l))$ is the classical shadow representation of $\rho(x_l)$, a $2^n \times 2^n$ matrix that reproduces $\rho(x_l)$ in expectation over
307 random Pauli measurements.

$$\{x_l \rightarrow \sigma_T(\rho(x_l))\}_{l=1}^N \quad (18)$$

308 The classical shadow size required to accurately approximate all reduced r -body density matrices scales exponentially
309 in subsystem size r , but is independent of the total number of qubits n .

Algorithm III.1: Classical Shadow (tomography)

input : a density matrix ρ (many copies), observables $\hat{O}...$

output: classical shadow ρ_{cs}

```

310 1 for  $i = 1, 2, \dots, m$  do
      2   random Pauli measurements // a comment
      3   ... return?
311 4 return?

```

311 **Lemma 1.** *the variance*

$$\text{Var}[o] = \mathbb{E}[(o - \mathbb{E}[o])^2] \leq \left\| O - \frac{\text{Tr}(O)}{2^n} \mathbb{1} \right\|_{shadow}^2 \quad (19)$$

312 sample complexity

$$N_{tot} = \mathcal{O} \left(\frac{\log(M)}{\epsilon^2} \max_{1 \leq i \leq M} \left\| O_i - \frac{\text{Tr}(O_i)}{2^n} \mathbb{1} \right\|_{shadow}^2 \right) \quad (20)$$

313 **Theorem 7** (Pauli/Clifford measurements). *additive error ϵ , M arbitrary k -local linear function $\text{Tr}(\hat{O}_i \rho)$, lower*
314 *bound $\Omega(\log(M) 3^k / \epsilon^2)$ copies of the state ρ .*

315 *2. training data and classical kernel methods*

316 **Definition 23** (shadow kernel). given two density matrices (quantum states) ρ and ρ' , *shadow kernel* [3] is

$$k_{shadow}(S, S') := \exp \left(\sum \exp \left(\sum \right) \right) \quad (21)$$

317 power of data -

318 **Proposition 6** ([4]). *If a classical algorithm without training data can compute $f(x) = \langle x | \hat{U}_{QNN}^\dagger \hat{O} U_{QNN} | x \rangle$ (with*
319 *amplitude encoding) efficiently (poly time in ...) for any \hat{U}_{QNN} and \hat{O} , then **BPP** = **BQP** (which is believed unlikely).*

320 **Proposition 7** ([4]). *Training an arbitrarily deep quantum neural network \hat{U}_{QNN} with a trainable observable \hat{O} is*
321 *equivalent to training a **quantum kernel** method with kernel $k_Q(\mathbf{x}, \mathbf{x}') = \text{Tr}(\rho(\mathbf{x}) \rho'(\mathbf{x}'))$*

322 **Proposition 8** ([4]). *exist quantum advantage in machine learning (not significant, practical) ... discrete log,*
323 *factoring...*

Algorithm III.2: Classical learning (SVM) + classical shadow

input : classical shadow? (features) with label (training data)
output: entanglement structure? decision

```

1 for  $i = 1, 2, \dots, m$  do
2   kernel estimation                                     // classical kernel
3   SVM                                                    // SVM
4   return "?"
5 return w                                                // parameters of the separating hyperplane in the feature space

```

B. Quantum trace (kernel) estimation

1. Entanglement spectroscopy via quantum trace estimation

For $\text{Tr}(\rho_A^m)$, an important application of multivariate trace estimation [5] is to entanglement spectroscopy [26] - deducing the full set of eigenvalues of ρ_A . The smallest eigenvalue of diagnoses whether ψ_{AB} is separable or entangled [27]. **constant depth**

Definition 24 (entanglement spectroscopy). the concept of the entanglement spectrum which is the energy spectrum of the “entanglement Hamiltonian” \hat{H}_E defined through $\rho_A = \exp(\hat{H}_E)$. They pointed out that the largest eigenvalues of ρ_A [30] contain more universal signatures than the von Neumann entropy or S_2 alone. (SWAP trick, Quantum phase estimation) [26]:

Algorithm III.3: entanglement spectroscopy by ... quantum trace estimation

input : (copies of) density matrix (graph state?) ρ, \dots
output: spectrum of entanglement Hamiltonian

```

1 for  $i = 1, 2, \dots, m$  do
2   GHZ                                                    // prepare GHZ
3   parallel                                              // estimate real and imaginary part respectively
4   return  $\lambda$ 
5 return smallest eigenvalue of  $\rho_A$ 

```

Remark 9 ([5]). We remark that, an alternative way to estimate $\text{Tr}[\rho^k]$ for each $k \in [m]$ is by using the method of classical shadows to obtain ‘classical snapshots’ of ρ that can be linearly combined to obtain a classical random variable whose expectation is $\text{Tr}[\rho^k]$ (see Supplementary Material Section 6 of [3]). However, it is unclear to us if this method would offer savings in the quantum resources required, as the total number of times the quantum circuit needs to be run in the data acquisition phase should scale with the variance of the corresponding estimator. We do not know of a concise expression for this variance for arbitrary m . Indeed, calculating it for just a single value of m ($m = 2$) required four pages of calculations in [3].

2. Entanglement witness via quantum trace estimation

Algorithm III.4: entanglement witness by ... quantum trace estimation

input : (copies of) density matrix (graph state?) ρ , an entanglement witness (observable) \hat{W}
output: determine entangled structure??

```

1 for  $i = 1, 2, \dots, m$  do
2    $\hat{W}_i$                                                 // estimate entanglement witness by quantum circuit
3   return  $\text{Tr}(\hat{W}\rho)$ 
4 return entangled ? GME ? separable with certain partition?

```

3. Quantum kernel SVM

related work:

- quantum kernel method: estimate kernels by quantum algorithms (circuits) [28] [29]
- rigorous and robust quantum advantage of quantum kernel method in SVM [30]. group structured data [31]
- power of data in quantum machine learning [4]: input???

Definition 25 (quantum kernel). quantum kernel with quantum feature map $\phi(\mathbf{x}) : \mathcal{X} \rightarrow |\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x})|$

$$k_Q(\rho, \rho') := |\langle\phi(\mathbf{x})|\phi(\mathbf{x}')\rangle|^2 = \left| \langle 0 | \hat{U}_{\phi(\mathbf{x})}^\dagger \hat{U}_{\phi(\mathbf{x}')} | 0 \rangle \right|^2 = \text{Tr}(\rho\rho') \quad (22)$$

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);

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 entanglement witness (ansatz)

an ansatz for entanglement witness [32] (graph state entanglement)

$$\hat{W}_{\text{ansatz}} := \sum_{k_1, k_2, \dots, k_n} w_{k_1, k_2, \dots, k_n} \bigotimes_i^n \hat{\sigma}^{(k_i)}, \quad \hat{\sigma} \in \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z, \mathbb{1}_{2 \times 2}\} \quad (23)$$

c.f. full tomography (Stokes parameters) Eq. (14)

Algorithm III.5: ansatz + classical shadow + classical SVM

input : (copies of) ρ , an entanglement witness (observable) ansatz \hat{W}_a
output: classifier \mathbf{w}

```

1 for  $i = 1, 2, \dots, m$  do
2    $\rho_{cs}$                                      // classical shadow
3    $\mathbf{w}\mathbf{x}$                                      // optimize SVM
4 return parameters  $\mathbf{w}$  (SVM hyperplane)
```

2. Variational trace estimate (direct)

Variational Quantum Algorithms for Trace Distance and Fidelity Estimation [33]

Theorem 9. *On quantum computers, evaluating the trace distances is probably hard since even judging whether ρ and ρ' have large or small trace distance is known to be QSZK-complete [34], where QSZK (quantum statistical zero-knowledge) is a complexity class that includes BQP (bounded-error quantum polynomial time).*

D. Theoretic upper bounds and lower bounds

[3] [35] [4] [18]

1. Separations (complexity)

contrived problem (engineered dataset)? for exponential speedup

	gate/depth/computation	measurements/samples	query?	input/direct?
shadow tomography	exp circuit	Theorems 5 and 7	N/A	direct
entanglement witness	N/A	Proposition 1 (constant?)	convex?[36] [20]	indirect
classical shadow	N/A	Theorem 7	N/A	direct
classical ML + entanglement witness	??	efficient?	N/A	direct
quantum (variational) circuits	Theorem 8 (c-depth?)		property test	??

TABLE I: complexity (different measures) of different methods

2. Obstacles (practical)

IV. NUMERICAL SIMULATION

A. Classification accuracy

1. Data preparation

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

2. Hyperparameters and settings

We consider a set of different regularization parameters,...

3. Results

performance of different methods:

FIG. 3: comparison of

B. Robustness to noise

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

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}}) |G\rangle\langle G| + p_{\text{noise}} \frac{\mathbb{1}}{2^n} \quad (24)$$

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

Remark 10 ([2]). the largest noise tolerance p_{limit} just related to the **chromatic number** (graph property) of the graph.

Question 5. other noise (depolarization)? e.g., flip error, phase error

FIG. 4: robustness

find optimal (robustness) entanglement witness by classical machine learning (qunatum circuit?)

C. Experiments

future: experimental (photonic) implementation with a few qubits (generation, verification) [37]

V. CONCLUSION AND DISCUSSION

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 G. Toth, *Physics Reports* **474**, 1 (2009), [arXiv:0811.2803 \[cond-mat, physics:physics, physics:quant-ph\]](#).
- [8] G. Toth and O. Guehne, *Phys. Rev. Lett.* **94**, 060501 (2005), [arXiv:quant-ph/0405165](#).
- [9] G. Tóth and O. Gühne, *Phys. Rev. A* **72**, 022340 (2005).
- [10] O. Gühne and N. Lütkenhaus, *Phys. Rev. Lett.* **96**, 170502 (2006).
- [11] S. Sciara, C. Reimer, M. Kues, P. Roztock, A. Cino, D. J. Moss, L. Caspani, W. J. Munro, and R. Morandotti, *Phys. Rev. Lett.* **122**, 120501 (2019).
- [12] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, *Nature Phys* **5**, 19 (2009), [arXiv:0910.1116](#).
- [13] 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](#).
- [14] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, *Phys. Rev. A* **103**, 052426 (2021), [arXiv:2012.07606 \[quant-ph\]](#).
- [15] X. Gao and L.-M. Duan, *Nat Commun* **8**, 662 (2017), [arXiv:1701.05039 \[cond-mat, physics:quant-ph\]](#).
- [16] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.
- [17] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, *IEEE Trans. Inform. Theory*, 1 (2017).
- [18] 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](#).
- [19] E. Tang, *Phys. Rev. Lett.* **127**, 060503 (2021), [arXiv:1811.00414 \[quant-ph\]](#).
- [20] A. Montanaro and R. de Wolf, *A Survey of Quantum Property Testing* (2018), [arXiv:1310.2035 \[quant-ph\]](#).
- [21] S. Ben-David, A. M. Childs, A. Gilyén, W. Kretschmer, S. Podder, and D. Wang, *2020 IEEE 61st Annu. Symp. Found. Comput. Sci. FOCS*, 649 (2020), [arXiv:2006.12760](#).
- [22] N. M. Kriege, F. D. Johansson, and C. Morris, *Appl Netw Sci* **5**, 6 (2020), [arXiv:1903.11835 \[cs, stat\]](#).
- [23] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, *Pattern Recognition* **48**, 344 (2015).
- [24] 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 \[quant-ph\]](#).
- [25] 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 \[quant-ph\]](#).
- [26] S. Johri, D. S. Steiger, and M. Troyer, *Phys. Rev. B* **96**, 195136 (2017), [arXiv:1707.07658](#).
- [27] P. Horodecki and A. Ekert, *Phys. Rev. Lett.* **89**, 127902 (2002), [arXiv:quant-ph/0111064](#).
- [28] 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](#).
- [29] M. Schuld and N. Killoran, *Phys. Rev. Lett.* **122**, 040504 (2019), [arXiv:1803.07128 \[quant-ph\]](#).
- [30] Y. Liu, S. Arunachalam, and K. Temme, *Nat. Phys.* **17**, 1013 (2021), [arXiv:2010.02174 \[quant-ph\]](#).
- [31] 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 \[quant-ph\]](#).
- [32] E. Y. Zhu, L. T. H. Wu, O. Levi, and L. Qian, *Machine Learning-Derived Entanglement Witnesses* (2021), [arXiv:2107.02301 \[quant-ph\]](#).
- [33] R. Chen, Z. Song, X. Zhao, and X. Wang, *Quantum Sci. Technol.* **7**, 015019 (2022), [arXiv:2012.05768 \[math-ph, physics:quant-ph\]](#).
- [34] J. Watrous, *Quantum Computational Complexity* (2008), [arXiv:0804.3401 \[quant-ph\]](#).
- [35] H.-Y. Huang, R. Kueng, and J. Preskill, *Phys. Rev. Lett.* **126**, 190505 (2021), [arXiv:2101.02464 \[quant-ph\]](#).
- [36] S. Chakrabarti, A. M. Childs, T. Li, and X. Wu, *Quantum* **4**, 221 (2020), [arXiv:1809.01731 \[quant-ph\]](#).
- [37] 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).
- [38] A. Jacot, F. Gabriel, and C. Hongler, *Neural Tangent Kernel: Convergence and Generalization in Neural Networks* (2020), [arXiv:1806.07572 \[cs, math, stat\]](#).

Appendix A: Machine learning background

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

1. Support vector machine

Definition 26 (SVM). Given a set of (binary) labeled data, support vector machine (SVM) is designed to find a hyperplane (a linear function) such that maximize the margin between two partitions...

$$\max_{\mathbf{w}} \dots \quad (\text{A1})$$

a. kernel method

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

Definition 27 (kernel). In general, the kernel function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ measures the similarity between two input data points by an inner product

$$k(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \quad (\text{A2})$$

If the input $\mathbf{x} \in \mathbb{R}^d$ (conventional machine learning task, e.g., image classification), the feature map $\phi(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^n$ ($d < n$) 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 6 (kernels). Some common kernels: the polynomial kernel $k_{\text{poly}}(\mathbf{x}, \mathbf{x}') := (1 + \mathbf{x} \cdot \mathbf{x}')^q$ with feature map $\phi(\mathbf{x}) \dots$ The Gaussian kernel $k_{\text{gaus}}(\mathbf{x}, \mathbf{x}') := \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2)$ 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.

similarity measures? advantages? why? (isomorphism?)

Definition 28 (divergence). KL divergence (relative [entropy](#)): measure the distance (similarity) between two probability distributions:

$$D_{\text{KL}}(P||Q) := \sum P(x) \log(P(x)/Q(x)) \quad (\text{A3})$$

symmetric version: Jensen-Shannon divergence (machine learning)

$$D_{\text{JS}}(P||P') := \frac{1}{2}(D_{\text{KL}}(P||M) + D_{\text{KL}}(P'||M)) \equiv H_S(M) - \frac{1}{2}(H_S(P) + H_S(P')) \quad (\text{A4})$$

where $M = (P + P')/2$ and Shannon [entropy](#) H_S . Analogously, quantum Jensen-Shannon divergence D_{QJS} of two density matrices can be defined...

$$D_{\text{QJS}}(\rho||\rho') := H_V(\rho_M) - \frac{1}{2}(H_V(\rho) + H_V(\rho')) \quad (\text{A5})$$

as a quantum graph kernel (ρ induced by quantum random walk)

2. Neural network

a. neural network and kernel

Definition 29 (neural tangent kernel). neural tangent kernel [38]: proved to be equivalent to deep neural network [15] in the limit ...

$$k_{\text{NT}}(S_T(\rho_l), \tilde{S}_T(\rho_{l'})) = \left\langle \phi^{(\text{NT})}(S_T(\rho_l)), \phi^{(\text{NT})}(\tilde{S}_T(\rho_{l'})) \right\rangle \quad (\text{A6})$$

b. Stabilizer formalism

475

476 denote a group by \mathbb{G} and a subgroup \mathbb{H} .

477 **Definition 30** (Pauli group).

478 **Definition 31** (Clifford group).

479 **Definition 32** (Stabilizer).

c. quantum neural network

480

Appendix B: Hardness assumptions

481

482 **Definition 33** (NP). NP, NP-hard, NP-complete

483 **Definition 34** ($\#P$). $\#P$

484 **Definition 35** (QMA). QMA

485 **Definition 36** (BPP). BPP

486 **Definition 37** (BQP). BQP