Learning the Closest Product State

Ainesh Bakshi
Massachusetts Institute of Technology
Cambridge, USA
ainesh@mit.edu

John Bostanci
Columbia University
New York, USA
johnb@cs.columbia.edu

William Kretschmer
University of California, Berkeley
Berkeley, USA
kretsch@berkeley.edu

Zeph Landau
University of California, Berkeley
Berkeley, USA
zeph.landau@gmail.com

Jerry Li
University of Washington
Seattle, USA
jerryzli@cs.washington.edu

Allen Liu
Massachusetts Institute of Technology
Cambridge, USA
cliu568@mit.edu

Ryan O'Donnell
Carnegie Mellon University
Pittsburgh, USA
odonnell@cs.cmu.edu

Ewin Tang
University of California, Berkeley
Berkeley, USA
ewin@berkeley.edu

Abstract

We study the problem of finding a product state with optimal fidelity to an unknown n-qubit quantum state ρ, given copies of ρ. This is a basic instance of a fundamental question in quantum learning: is it possible to efficiently learn a simple approximation to an arbitrary state? We give an algorithm which finds a product state with fidelity ε-close to optimal, using $N = n^{\mathrm{poly}(1/\varepsilon)}$ copies of ρ and $\mathrm{poly}(N)$ classical overhead. We further show that estimating the optimal fidelity is NP-hard for error $\varepsilon = 1/\mathrm{poly}(n)$, showing that the error dependence cannot be significantly improved.

For our algorithm, we build a carefully-defined cover over candidate product states, qubit by qubit, and then demonstrate that extending the cover can be reduced to approximate constrained polynomial optimization. For our proof of hardness, we give a formal reduction from polynomial optimization to finding the closest product state. Together, these results demonstrate a fundamental connection between these two seemingly unrelated questions. Building on our general approach, we also develop more efficient algorithms in three simpler settings: when the optimal fidelity exceeds 5/6; when we restrict ourselves to a discrete class of product states; and when we are allowed to output a matrix product state.

CCS Concepts

• **Theory of computation** → **Machine learning theory**; **Quantum computation theory**.

Keywords

agnostic learning, product states, quantum state tomography, agnostic tomography

ACM Reference Format:
Ainesh Bakshi, John Bostanci, William Kretschmer, Zeph Landau, Jerry Li, Allen Liu, Ryan O'Donnell, and Ewin Tang. 2025. Learning the Closest Product State. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC '25), June 23–27, 2025, Prague, Czechia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3717823.3718207

STOC '25, Prague, Czechia

ACM ISBN 979-8-4007-1510-5/25/06
https://doi.org/10.1145/3717823.3718207

1 Introduction

When can we obtain a classical description of a complex quantum system? This problem, at the heart of quantum information theory, is one commonly faced by experimentalists: when we have a large, intricate quantum device, how can we tell what it is doing? Due to the exponentiality inherent to quantum mechanics, a generic system of n particles is described by a number of parameters scaling exponentially with n, so in general, an efficient description simply does not exist. However, real-world systems are not generic: the physics governing the device will suggest a corresponding model of the system, giving us a hint for how the system can be efficiently described.

Simultaneously, in real-world applications, the state which one is learning may not—and typically will not—exactly fall within a given model class, due to noise or other forms of imprecision in how our model represents the real world. In light of this, the natural question is to seek the best approximation to the underlying state within the prescribed model. Such an approximation can serve as a far more tractable proxy for the true state when it is complex to describe exactly. In this work, we consider the problem for the class of product states, arguably the most fundamental class of states to consider. Stated plainly, the question we ask is the following:

Can we efficiently learn the best product state approximation to any given state?

We formalize this problem as follows:

Problem 1 (Learning the closest product state)**.** Consider the set of n-qubit product states $\mathcal{P} = \{|\pi_1\rangle \otimes \cdots \otimes |\pi_n\rangle\}$, and let $\varepsilon, \delta > 0$ be error parameters. Given N copies of an arbitrary n-qubit state with density matrix ρ, output a classical description of a state $|\pi\rangle \in \mathcal{P}$ such that, with probability $\geqslant 1 - \delta$,

$$\langle\pi|\,\rho\,|\pi\rangle \geqslant \mathrm{OPT} - \varepsilon, \qquad \text{where } \mathrm{OPT} = \max_{|\pi\rangle \in \mathcal{P}} \langle\pi|\,\rho\,|\pi\rangle\ .$$

Product states are natural to study in this context for a number of reasons. Because of the locality inherent in physical systems, we commonly model physical systems with states exhibiting low entanglement. Chief among them are *mean-field theories*, which model systems as states which exhibit *zero entanglement*, e.g., product states [14]. The mean-field approximation plays a central role in domains relevant to quantum computing: in particular, in quantum chemistry, mean-field theories like Hartree-Fock theory and density functional theory are the standard algorithmic workhorses for understanding chemical processes [17]. In light of this, we can rephrase Problem 1 as asking for the best (pure) mean-field approximation to an arbitrary quantum state, and for the quality of that approximation. From this perspective, we believe that obtaining an efficient algorithm for Problem 1 will have important implications both for validating the effectiveness of these theories and for understanding their properties in real-world settings.

As an example application, physicists already run computations to solve Problem 1 in the setting where the input is not a quantum state, but a description of a condensed matter system. Collective entanglement of a multipartite state is often measured by OPT, the best fidelity of the state with a product state, also known as the *geometric measure of entanglement* [52]. Since its introduction in 2003, this entanglement measure has been used to understand a variety of condensed matter systems (see related work). An algorithm for Problem 1 can be used to compute the geometric measure of entanglement for states which are efficiently preparable on a quantum computer, by preparing copies of the state and running the algorithm to estimate OPT, giving an advantage when such states are classically intractable.

Despite the apparent simplicity of the problem, relatively little was known about the computational complexity of Problem 1. From a statistical point of view, one can obtain sample-efficient learners via classical shadow estimation [28] or shadow tomography [1], but these estimators require exponential runtime. On the other hand, efficient algorithms were known only for highly restricted versions of the problem [24]. This lack of efficient algorithms might be surprising, as when when the unknown state ρ is a product state, i.e. $\mathsf{OPT} = 1$, this task is easy: many algorithms work, including learning every register separately. However, these algorithms are brittle, and fail catastrophically when $\mathsf{OPT} < 0.99$. Even algorithms for the related problem of product state testing, initiated by the important work of Harrow and Montanaro [25], do not admit estimates of OPT when OPT is bounded away from 1. In contrast, one would hope to obtain efficient algorithms even when OPT is a small constant (say, 0.1): product states with constant fidelity are still great approximations, considering that almost all product states will have fidelity exponentially small in n.

Beyond specific applications, we hope that understanding this algorithmic task can shed light on a broader program in quantum learning theory. An emerging line of work has been studying "learning the closest state in a hypothesis class", also known as *agnostic tomography*: formally, this problem is Problem 1, except the class of product states $\mathcal{P}$ is replaced with a different hypothesis class C. Product states appear as a special case of several well-studied classes of quantum states, including states described by low-depth quantum circuits, matrix product states, and Gibbs states and ground states of local Hamiltonians. Understanding the computational complexity of agnostic tomography of product states is therefore an important stepping stone to building up to richer approximations. As we demonstrate below, it turns out that learning the closest product state is already a surprisingly deep problem.

1.1 Results

We answer the aforementioned question in the affirmative and provide the first efficient algorithm for agnostic tomography of product states:

Theorem 1.1 (Learning the closest product state). *There is an algorithm which, given as input $\varepsilon > 0$ and $N = n^{\mathrm{poly}(1/\varepsilon)}$ copies of an unknown n-qubit*[1] *state ρ, runs in time $\mathrm{poly}(N)$ and outputs the classical description of a pure product state $|\phi\rangle$ that, with probability at least 0.99, satisfies*

$$\langle\phi|\rho|\phi\rangle \geqslant \mathsf{OPT} - \varepsilon\,. \tag{1}$$

The algorithm also produces an estimate of OPT *to ε error.*

We pause to make several comments about this result. First, the regime we are primarily interested in is when ε is a constant (though possibly small). In this regime, our algorithm runs in polynomial time. This resolves an open question posed in [24].

Secondly, our result holds for all values of OPT, and not just OPT close to 1. The setting where OPT is a small constant (say, 0.1) is particularly challenging: in this regime, there may not be a unique closest product state. In this setting, our algorithm in fact actually outputs a net (albeit in a relatively weak sense) over *all* product states which are close to the unknown state ρ. Moreover, our algorithm does not need to know the value of OPT, nor does it need even a lower bound on OPT (though if OPT is large the algorithm's complexity improves). Note, however, that the guarantee on the fidelity of $|\phi\rangle$ with ρ is only nontrivial when $\mathsf{OPT} > \varepsilon$.

Finally, prior to this work, the only algorithms for this task were *sample*-efficient, but not time-efficient. For example, a polynomial number of random Clifford measurements suffices to estimate every fidelity with a product state $\langle\pi|\rho|\pi\rangle$ to ε error [28]. However, there are an exponential number of these product states, and computing even one fidelity from these randomized measurements requires exponential time [29].

Improved product state testing. Agnostic tomography of product states is closely related to the well-studied problem of product state testing [25], where the goal is to determine whether or not a state $|\psi\rangle$ is a product state, or has fidelity at most $1-\varepsilon$ with any product state. In the former case, the test should always accept, and in the latter, the test should reject with probability at least p, for some $p = \Theta(\varepsilon)$.

Our results shed new light on this problem: the celebrated tester of Harrow and Montanaro [25] exhibits a strange behavior, wherein their rejection probability satisfies $p \leqslant 1/2 + o(1)$, even when $\varepsilon \to 1$. That is, for some reason, the tester cannot distinguish the case where $|\psi\rangle$ has overlap roughly $1/2$ with some product state, versus the case where the state has overlap $\ll 1/2$ with any product state. Since our algorithm also produces an estimate of OPT to error ε, it improves upon the best-known guarantees for product state testing [48] in this "tolerant" [16] regime.

[1]For simplicity, we only consider when the local systems are qubits. We believe that the results should generalize to qudits without too much struggle.

Computational lower bounds. It is natural to ask whether or not one can hope for a running time for this problem which is polynomial in both n and $1/\varepsilon$. We complement our upper bound with a lower bound, demonstrating that our runtime is, in a qualitative sense, close to optimal:

Theorem 1.2 (Hardness of product state approximation). *Suppose there is an efficient quantum algorithm for solving the following problem: given* $\mathrm{poly}(n)$ *copies of an unknown, n-qubit mixed state* ρ*, with probability* $\geqslant 0.01$*, output* $|\psi\rangle$ *satisfying*

$$\langle\psi|\rho|\psi\rangle \geqslant \max_{|\pi\rangle\in\mathcal{P}} \langle\pi|\rho|\pi\rangle - \frac{1}{\mathrm{poly}(n)}\,.$$

Then $\mathsf{BQP} \supseteq \mathsf{NP}$.

In particular, this rules out algorithms with strongly polynomial dependencies on all parameters. We prove this hardness via a straightforward, polynomial-time reduction to an NP-complete problem. Consequently, this also rules out any algorithms that have sub-exponential dependence on $1/\varepsilon$, assuming the quantum analog of the exponential time hypothesis. We interpret this as saying that it is likely challenging to obtain substantial qualitative improvements to the runtime in Theorem 1.1.

We also remark that this hardness result demonstrates an interesting computational-statistical gap for the problem of finding the closest product state. Namely, classical shadow estimation [28] demonstrates that this regime can be solved sample efficiently, but on the other hand, our lower bound demonstrates that this rate cannot be matched by any efficient algorithm.

Approximate tensor optimization. The upper and lower bound are based on a new connection to the classical problem of *approximate tensor optimization.* Here, one is given a d-tensor $T \in (\mathbb{C}^n)^{\otimes d}$, and the goal is to find a unit vector $\vec{x} \in \mathbb{C}^n$ satisfying

$$T(\vec{x},\ldots,\vec{x}) \geqslant \max_{\|u\|_2=1} T(\vec{u},\ldots,\vec{u}) - \varepsilon\|T\|_F\,.$$

Our lower bound proceeds by direct reduction to this problem for $d = 4$, which is known to be NP-hard when $\varepsilon = 1/\mathrm{poly}(n)$ [23], and our upper bound works by reducing the problem to many different instances of constrained versions of this problem. This problem itself bears great resemblance to the problem of solving dense CSPs, and indeed, we believe the techniques we develop for constrained tensor optimization here may have applications to that setting as well.

Faster agnostic tomography of product states. In light of our lower bound, we ask whether there are simpler algorithms for agnostic tomography of product states, perhaps under additional assumptions. We show that this is true for three natural settings: (1) when the best product state approximation is quite good; (2) when the number of choices for each qubit is discrete; and (3) when the output is allowed to be a matrix product state.

First, we obtain a linear copy and nearly-quadratic time algorithm for agnostic tomography of product states as long as the fidelity of the optimal solution exceeds a fixed constant (namely, 5/6):

Theorem 1.3 (High-fidelity learning). *There is an algorithm that takes as input a parameter* $\varepsilon > 0$ *as well as* $N = O(n/\varepsilon)$ *copies of an n-qubit state* ρ*, and has the following guarantees: Provided* $\mathsf{OPT} > 5/6 + \varepsilon$*, it runs in* $O(Nn\log n)$ *time and outputs a pure product state* $|\psi\rangle$ *that satisfies*

$$\langle\psi|\rho|\psi\rangle \geqslant \mathsf{OPT} - \varepsilon\,,$$

(except with probability at most .01*).*

In other words, so long as the quality of the product approximation OPT exceeds 5/6, there is a strongly polynomial time algorithm for agnostic product state tomography. This stands in stark contrast to the state of affairs for general OPT, where the hardness result demonstrates such an algorithm is impossible. The threshold 5/6 naturally arises from our analysis, but it is an interesting open question to what extent it can be pushed.

We remark that the runtime dependence of the algorithm is linear in $1/\varepsilon$, even though it is easily seen that *estimating* OPT to $\pm\varepsilon$ requires $\Omega(1/\varepsilon^2)$ samples. For example, this lower bound holds even in the special case when $\rho = \mathsf{OPT}\,|0\rangle\langle 0| + (1-\mathsf{OPT})\,|1\rangle\langle 1|$ is a biased coin, and we want to distinguish whether (say) $\mathsf{OPT} = 0.9+\varepsilon$ or $\mathsf{OPT} = 0.9 - \varepsilon$. Our algorithm demonstrates that the task of *finding* a state whose fidelity is within ε of the optimum may be easier.

Second, we give an efficient algorithm for agnostic tomography, when the class of states is the set of product states where each qubit is drawn from a finite set of possible states:

Theorem 1.4 (Learning of a finite class of product states). *For* $k = 1,\ldots,n$*, let* $\mathcal{A}_k$ *denote a set of single qudit pure states satisfying* $|\mathcal{A}_k| \leqslant s$ *and* $|\langle\phi|\phi'\rangle| \leqslant 1-\delta$ *for all distinct* $|\phi\rangle, |\phi'\rangle \in \mathcal{A}_k$*. Let* $\mathcal{A} = \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$*, and for any n-qudit quantum state* ρ*, let* $\mathsf{OPT}_{\mathcal{A}} = \mathsf{OPT}_{\mathcal{A}}(\rho) = \max_{|\pi\rangle\in\mathcal{A}} \langle\pi|\rho|\pi\rangle$*. Then there is an algorithm which, given as input* $\varepsilon > 0$ *and* $N = \mathrm{poly}((ns)^{\log(1/\varepsilon)/\delta})$ *copies of an n-qudit state* ρ*, runs in* $\mathrm{poly}(N)$ *time and outputs the classical description of some* $|\psi\rangle \in \mathcal{A}$ *satisfying*

$$\langle\psi|\rho|\psi\rangle \geqslant \mathsf{OPT}_{\mathcal{A}} - \varepsilon\,,$$

(except with probability at most .01*).*

Stated plainly, so long as there are a finite set of possible states, and these states are all pairwise separated, then there is an efficient algorithm for agnostic tomography for this class of product states. We note that, similar to Theorem 1.1, our algorithm actually outputs all good solutions. This result also directly generalizes prior work of Grewal, Iyer, Kretschmer, and Liang [24], which studied the special case where each $\mathcal{A}_k$ is the set of 1-qubit stabilizer states. A very similar result was also obtained independently in [19], albeit with quite different techniques.

Third, we give an algorithm for learning a good matrix-product state approximation of a given state ρ. Matrix product states with small bond dimension can be used to efficiently describe systems of multiple particles where particles share a small (but non-zero) amount of entanglement, and are ubiquitous in quantum many-body physics [43, 47]. We give an algorithm for *agnostic* tomography of matrix product states.

Theorem 1.5 (Agnostic (improper) learning of matrix product states). *Let* n, d, r *be positive integers, and let* $\mathrm{MPS}_{n,d,r}$ *be the class of matrix product states on n qudits of local dimension d with bond dimension r. For any state* $\rho \in (\mathbb{C}^{d\times d})^{\otimes n}$*, let* $\mathsf{OPT}_r = \mathsf{OPT}_{n,d,r}(\rho) = \max_{|\phi\rangle\in\mathrm{MPS}_{n,d,r}} \langle\phi|\rho|\phi\rangle$ *be the maximum fidelity* *any such MPS has with ρ. There is an algorithm which, given as input $\varepsilon > 0$ and $N = \mathrm{poly}(n, d, r, 1/\varepsilon)$ copies of an unknown n-qudit state ρ, runs in time $\mathrm{poly}(N)$ and outputs the classical description of a matrix product state $|\widehat{\phi}\rangle$ of bond dimension $dn^2 \cdot \mathrm{poly}(r, 1/\varepsilon)$ such that*

$$\langle\widehat{\phi}|\rho|\widehat{\phi}\rangle \geqslant \mathsf{OPT}_r - \varepsilon,$$

(except with probability at most .01).

We can relate this task back to learning the closest product state by taking $r = 1$ and $d = 2$; then, $\mathrm{MPS}_{n,d,r}$ is the class of product states over qubits, and our algorithm is able to output a matrix product state with bond dimension $n^2 \mathrm{poly}(1/\varepsilon)$ whose fidelity with ρ is at least $\mathsf{OPT} - \varepsilon$. This gives an improper learner for product states, "improper" referring to our output not being a product state but instead a low-entanglement state. Our main result Theorem 1.1 is a *proper* learner for product states. In the error regimes of our lower bound, this gives an instance where improper agnostic learning is efficient, but proper agnostic learning is NP-hard. In general, the output of this algorithm is an MPS with a bond dimension of at least rn^2, which achieves a fidelity which is optimal with respect to MPSs with bond dimension r; this dependence on n in particular seems to be what makes this result more straightforward than proper learning of MPSs.

For this task, we recognize that the algorithm of Cramer, Plenio, Flammia, Somma, Gross, Bartlett, Landon-Cardinal, Poulin, and Liu [20] to learn an MPS also works when the input state is not an MPS, but merely has large constant fidelity with an MPS; our contribution is to generalize it to the agnostic case and perform the necessary analysis.

Techniques. All of these results, as well as Theorem 1.1, are all based on a common algorithmic framework, which may have applications more broadly. At a very high level, our algorithms sweep through the qubits one at a time, and generate a set of candidates for good solutions on the qubits seen so far. This cover is then used as the starting point for generating candidates over the subsequent qubits. The main algorithmic challenge is in making extending the cover efficient. In the case of Theorem 1.1, extending this cover is intimately connected to tensor optimization, as mentioned above. To achieve our faster algorithms Theorems 1.3 to 1.5, our key insight is that there are relatively simple and "greedy" techniques that allow us to extend this cover.

Our algorithms interleave classical computation with a particular quantum subroutine: the only way we access ρ is to perform tomography on various subspaces of subsystems, e.g. to estimate $\Pi\, \mathrm{tr}_S(\rho)\Pi$ for S some subset of qubits and Π a projector onto a subspace of $\mathrm{poly}(n)$ dimension (which can be represented efficiently with a quantum circuit). Our algorithms apply this subroutine to various choices of Π and S, which are adaptively chosen after classical computation on the output of the previous tomography routines. Since such a tomography subroutine can be performed with single-copy measurements, our algorithm can also be performed with only single-copy measurements. However, the adaptivity of this algorithm appears inherent: the classical shadows formalism [28] is the standard technique to allow algorithms like these to perform all of their measurements up-front, but doing this comes at the cost of exponential running time, which we cannot tolerate.

1.2 Related work

Concurrent work. In independent and concurrent work, Chen, Gong, Ye, and Zhang [19] give an algorithm for agnostic tomography of a finite set of product states, attaining a near-identical result to Theorem 1.4 via completely unrelated techniques. They also give an improved algorithm when the product states are stabilizer; we are also able to get a similar improvement in this setting.

Agnostic tomography. The notion of agnostic tomography was introduced by Grewal, Iyer, Kretschmer, and Liang [24], though similar notions have been considered under the notion "quantum hypothesis selection" [15] and in the PAC-learning setting; we refer to the survey [3] for a thorough discussion. Recent work has given agnostic tomography algorithms for stabilizer product states [24] and stabilizer states [19]. These algorithms use unrelated techniques.

Product state testing. A notable related algorithm is the product state test, which, using copies of a state ρ is able to distinguish the cases that $\mathsf{OPT} = 1$ from $\mathsf{OPT} \leqslant 1 - \varepsilon$. This algorithm, introduced by [38] and analyzed by Harrow and Montanaro [25], plays an important role in complexity theory, being used to prove that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ for $k > 2$. Though the algorithm suffices for testing, it cannot be used to estimate OPT when OPT is bounded away from 1 [25, Appendix B]. For similar reasons, it also does not seem to help with the task of finding good product states.

Product state approximations in other contexts. Our algorithm shows that it is possible to estimate the "geometric measure of entanglement" of a given pure state in polynomial time. This measure of entanglement, defined by Wei and Goldbart [50, 52], has seen significant investigation as a measure of multipartite entanglement. This interest comes from this measure's potential to capture aspects of entanglement in condensed matter systems which cannot be captured by the typical, bipartite measures of entanglement [41, 42, 51]. See the survey of De Chiara and Sanpera for further discussion [21]. However, research has been limited by computational intractability, so our work may give a possible avenue to expand its scope via quantum simulation.

Mean-field approximations also arise naturally in contexts where we want to understand things like ground states of many-body systems, and only have a handle on product states [14]. For example, to prepare ground states of many-body systems, current heuristic phase estimation methods have a running time which depends on the fidelity between the ground state and an input product state [33].

Agnostic learning of product distributions. In some ways, the problem we consider here is the quantum analog of the well-studied problem of agnostic learning of product distributions on the hypercube. In its most basic form, we are given samples from a distribution that is close to a product distribution over the hypercube, and the goal is to learn the optimal product distribution approximation. Efficient algorithms for this problem were given in [22, 32]. However, these algorithms only work when their version of OPT is sufficiently large; in classical learning theory, the regime when OPT is small is known as *list learning*, and efficient algorithms for list learning of product distributions are also known; see, e.g. [18, 31].

However, the guarantees they obtain are quite incomparable to ours, and their techniques do not have a meaningful parallel in the quantum setting.

Polynomial optimization. Polynomial optimization over the sphere is hard in general. Multiplicative approximations for optimizing low-degree polynomials in the worst case are well-understood (see [13] and references therein). However, polynomial optimization has still found prominent applications in classical learning problems in the last decade. The polynomials that naturally appear in these settings do not tend to be worst-case, and admit significantly better approximations. Optimizing low-degree polynomials (often subject to polynomial constraints) has become a key algorithmic primitive in dictionary learning [9], tensor decomposition [27], robustly learning Gaussian mixture models [5, 12, 34, 39] and private and list-decodable learning [6, 26, 30, 46]. These techniques have also found applications in quantum tasks, such as best separable state [10] and learning quantum Hamiltonians [7, 40]. An interesting feature of our algorithm, compared to this other work, is that we do not establish uniqueness of some strong structure arising from the underlying parameters. Instead, we output a (non-unique) cover over solutions, and use polynomial optimization as a subroutine to produce such a cover.

Optimizing low-degree polynomials over the hypercube also leads to approximation algorithms for constraint satisfaction problems on dense and low-threshold rank graphs [11, 35, 45] and high-dimensional expanders [2]. These results roughly proceed via solving a sum-of-squares relaxation of a polynomial maximization problem, and obtain additive error that scales proportional to ε times the ℓ_2-norm of the coefficients of the polynomial and runs in $n^{\mathrm{poly}(1/\varepsilon)}$ time. Similar techniques have also appeared in the context of refuting random CSP's [44].

A closely related problem is that of optimizing random polynomials over the sphere, which has deep connections to statistical physics and admits an additive-error guarantee under full replica-symmetry breaking [49]. While our optimization problem does not involve random polynomials, we show that we can optimize low-degree polynomials up to small additive-error efficiently.

2 Technical overview

We now cover the key technical ideas of our algorithms.

Why naive approaches fail. Given an n-qubit quantum state with density matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$, we want to find the product state that maximizes fidelity with ρ. The obvious algorithm that one might try to learn the closest product state is to take the best pure state approximation to each of its single-qubit subsystems. This algorithm works if ρ itself is a pure product state. However, the single-qubit subsystems do not contain enough information to deduce the best product state, even when the fidelity of ρ with the best product state is very close to 1. This phenomenon is why many naive approaches give exponentially poor approximations to the optimal value.

An illustrative example is to consider $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle$ is the state proportional to

$$|\psi\rangle \propto \sqrt{1-\varepsilon}\,|0^n\rangle + \sqrt{\varepsilon}\,|+^n\rangle$$

for some small constant ε. Because $\langle +^n|0^n\rangle = 2^{-n/2}$, $|\psi\rangle$ as written is exponentially close to normalized. The fidelity with the product state $|0^n\rangle$ can be computed explicitly:[2]

$$\langle 0^n|\rho|0^n\rangle = |\langle\psi|0^n\rangle|^2 = \left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon/2^n}}{\|\sqrt{1+\varepsilon}\,|0^n\rangle + \sqrt{\varepsilon}\,|+^n\rangle\|_2}\right)^2 \geqslant 1-\varepsilon\,.$$

In the limit of large n, the one-qubit density matrices of $|\psi\rangle$ all approach

$$\rho_i = \begin{bmatrix} 1-\varepsilon/2 & \varepsilon/2 \\ \varepsilon/2 & \varepsilon/2 \end{bmatrix}$$

We will see that there is a distinct state $|\psi'\rangle$ that is also ε-close to a product state, and has identical reduced density matrices, but for which $|0^n\rangle$ is a very bad product state approximation. Take an eigendecomposition of ρ_i as

$$\rho_i = p_1\,|v_1\rangle\langle v_1| + p_2\,|v_2\rangle\langle v_2|\ ,$$

with $p_1 > p_2$. The state

$$|\psi'\rangle = \sqrt{p_1}\,|v_1\rangle^{\otimes n} + \sqrt{p_2}\,|v_2\rangle^{\otimes n}$$

also has at least $1-\varepsilon$ fidelity with a product state (namely $|v_1\rangle^{\otimes n}$), and has all its local density matrices equal to ρ_i. However, calculation shows that $|\langle\psi'|0^n\rangle|^2$ decays exponentially to 0 in the limit of large n, because both $|v_1\rangle$ and $|v_2\rangle$ are constant-far from $|0\rangle$. So, there is not enough information in the one-qubit reduced density matrices to learn the best product state approximation.

Barriers to agnostic product tomography. The hard case above illuminates broader challenges inherent to this task. We are concerned with optimizing the fidelity $\langle\pi|\rho|\pi\rangle$ over the class of product states; however, fidelity is typically quite poorly behaved. For example, almost all product states have exponentially small fidelity with ρ, which is too small to detect, and the fidelity landscape can have many local optima which thwart local search algorithms, like those based on convex optimization. This ill-behavedness is a well-established phenomenon related to the "barren plateau" problem in quantum machine learning [37].

The regime where the optimal fidelity is a small constant like 0.1 is particularly challenging since, unlike the case where OPT is near 1, there are many well-separated globally optimal solutions. This lack of uniqueness presents basic issues for us: even if we manage to traverse the fidelity landscape and find many locally-optimal product states with fidelity 0.1, how can we conclude that we are done, and certify that there is no product state with fidelity 0.2?

Maintaining a cover over good product states. Our crucial insight is that we can efficiently maintain a cover over all product states that have large fidelity. This insight is enabled by the following observations:

(1) If a product state $|\pi\rangle$ has good fidelity with ρ, then its restriction to a subsystem S has good fidelity with the partial trace of ρ onto the subsystem: $\langle\pi|\rho|\pi\rangle \leqslant \langle\pi_S|\rho_S|\pi_S\rangle$.
(2) The number of product states with good fidelity with ρ and which have pairwise small fidelity with each other is small.

[2]It follows from one of our later results that the maximum product state fidelity with $|\psi\rangle$ is exponentially close to $1-\varepsilon$.

The first observation means that we do not have to optimize fidelity over the entire space of product states: just those which are extensions of good product states over a subsystem. In short, we can build a set of good product states qubit by qubit. The second observation means that, instead of maintaining *all* good product states, of which there could be exponentially many, it suffices to maintain a small number of well-separated good product states. In short, it suffices to maintain a cover.

A priori, it is even unclear whether a small cover over such product states exists. Our main technical contribution is to establish the existence of such a cover and demonstrate that it can be computed efficiently. Our algorithm starts with a cover over good product states for $\rho_{[1]}$, the state on qubit 1, and iteratively expands the cover a single qubit at a time. In particular, we show that given a cover for qubits $1, 2, \ldots, m-1$, extending it to qubit m can be reduced to polynomial optimization problems over the sphere with a dimension-independent number of ℓ_2 and ℓ_∞ constraints.

For the remainder of the section, we outline our approach to efficiently maintain a cover.

Fidelity and tangent distance. We begin by introducing a parametrization of product states which is used throughout the paper. For a n-dimensional vector of complex numbers, $\vec{z} \in \mathbb{C}^n$, we denote its corresponding product state by

$$|\pi_{\vec{z}}\rangle = \bigotimes_{i=1}^{n} \frac{|0\rangle + z_i\,|1\rangle}{\sqrt{1+|z_i|^2}}.$$

Looking ahead, we ultimately want to optimize over these z_i's, so we want a notion of cover which behaves nicely with respect to this parametrization. Fidelity is well-known to be an unwieldy notion of distance between quantum states and is typically hard to analyze. So, instead of constructing a cover directly using fidelity, we introduce an alternate measure between product states:

Definition 2.1 (Tangent distance)**.** Given two product states $|\pi_{\vec{z}}\rangle$ and $|\pi_{\vec{a}}\rangle$, the tangent distance between them is defined as

$$\mathrm{d}_{\tan}(|\pi_{\vec{z}}\rangle, |\pi_{\vec{a}}\rangle) = \left\| \frac{\vec{z}-\vec{a}}{1+\vec{z}^*\vec{a}} \right\|_2 = \Big(\sum_{i=1}^{n} \Big| \frac{z_i - a_i}{1+z_i^* a_i} \Big|^2 \Big)^{1/2}.$$

We call it "tangent distance" because, for a single qubit, this measure corresponds to $|\tan(\theta)|$, where θ is the angle between the two states on the Bloch sphere. This notion of distance satisfies several desiderata, including being invariant under single-qubit unitaries and being equal to $\|\vec{z}\|_2$ when $\vec{a} = \vec{0}$. Importantly, tangent distance can be related to fidelity as follows:

$$\log\left(\frac{1}{|\langle \pi_{\vec{z}}|\pi_{\vec{a}}\rangle|^2}\right) \leqslant \mathrm{d}_{\tan}(|\pi_{\vec{z}}\rangle, |\pi_{\vec{a}}\rangle)^2 \leqslant \frac{1}{|\langle \pi_{\vec{z}}|\pi_{\vec{a}}\rangle|^2} - 1. \quad (2)$$

Now, we can introduce our notion of cover under tangent distance[3]: a cover C over product states is (η, ε)-good for a state ρ if the following hold

- **Good fidelity:** For all product states $|\pi\rangle \in C$, $\langle \pi|\rho|\pi\rangle \geqslant \eta - \varepsilon$;
- **Separation:** For all distinct $|\pi\rangle, |\pi'\rangle \in C$, $\mathrm{d}_{\tan}(|\pi\rangle, |\pi'\rangle) \geqslant 2/\eta$;
- **Coverage:** For any product state $|\phi\rangle$ such that $\langle \phi|\rho|\phi\rangle \geqslant \eta$, there exists a $|\pi\rangle \in C$ such that $\mathrm{d}_{\tan}(|\phi\rangle, |\pi\rangle) \leqslant 3/\eta$.

[3]Though our algorithm naturally produces a cover with respect to tangent distance, one can use (2) to convert the guarantees to those involving fidelity.

We design an algorithm which, given η and $\varepsilon < \eta/3$, outputs a (η, ε)-good cover, where every product state $|\pi_{\vec{z}}\rangle$ in the cover is described by its parametrization $\vec{z}$. In particular, this gives a product state with fidelity $\geqslant \eta - \varepsilon$, assuming a product state with fidelity η exists. By performing binary search on η, one can use this to find a product state with fidelity $\geqslant \mathsf{OPT} - \varepsilon$, as stated in Theorem 1.1.

Existence of small covers. Our first step is to show that the size of an (η, ε)-good cover is at most $6/\eta$. Let $C = \{\langle \pi^{(i)} \rangle\}_i$ be an (η, ε)-good cover. For intuition, suppose the product states in the cover were not just well-separated but orthogonal. Then each captures a different part of the "mass" of ρ. That is,

$$1 = \mathrm{tr}(\rho) \geqslant \sum_i \langle \pi^{(i)}|\rho|\pi^{(i)}\rangle \geqslant |C|(\eta - \varepsilon) \geqslant |C|(2\eta/3),$$

where the last two inequalities use the good fidelity property of the cover and that $\varepsilon < \eta/3$, respectively. In general, $\sum_i \langle \pi^{(i)}|\rho|\pi^{(i)}\rangle$ is equal to $\mathrm{tr}(MM^\dagger \rho)$ for M the matrix whose columns are the states in the cover $|\pi^{(i)}\rangle$. Then,

$$|C|(2\eta/3) \leqslant \mathrm{tr}(MM^\dagger\rho) \leqslant \|MM^\dagger\|_{\mathrm{op}} = \|M^\dagger M\|_{\mathrm{op}} \leqslant 1 + |C|(\eta/2),$$

giving the bound $|C| \leqslant 6/\eta$. In the last step, we used the well-separated condition: the diagonal entries of $M^\dagger M$ are 1, the off-diagonal ones have magnitude at most $\eta/2$ by Eq. (2), and by the Gershgorin circle theorem the operator norm of $MM^\dagger$ is bounded by the maximum sum of magnitudes of any of its rows.

We can further show how to construct an (η, ε)-good cover algorithmically. We do this by iteratively forming an (η, ε)-good cover for $\rho_{[m]}$, the partial trace of ρ onto qubits 1 through m, for m from 1 to n. We can construct a good cover for $\rho_{[m]}$ greedily: start with C_m empty, and look for a violation of the coverage property. When we find one, add the corresponding $|\phi\rangle$ to C_m, and repeat. Because we know an (η, ε)-good cover on $m-1$ qubits C_{m-1}, we can restrict our search to just look over product states $|\phi\rangle$ whose first $m-1$ qubits are close in tangent distance to an element of C_{m-1}. This makes the problem of finding a violation tractable, since we only have to search in the neighborhood of some "root" product state. In particular, we show that it suffices to solve the following optimization problem.

$$\begin{aligned} \underset{\vec{z}\in\mathbb{C}^m}{\text{maximize}} \quad & \langle \pi_{\vec{z}}|\,\rho\,|\pi_{\vec{z}}\rangle \\ \text{subject to} \quad & \mathrm{d}_{\tan}(|\pi_{\vec{z}}\rangle, |\pi_{\vec{a}}\rangle) \geqslant 2/\eta \text{ for all } |\pi_{\vec{a}}\rangle \in C_m, \\ & \mathrm{d}_{\tan}(|\pi_{\vec{z}}\rangle, |\pi_{\vec{0}}\rangle) \leqslant 4/\eta. \end{aligned} \quad \text{(Tangent Cover)}$$

The precise soundness and completeness guarantees needed are shown in the full algorithm; the constraints allow for significant slack. Note that the second constraint is equivalent to $\|\vec{z}\|_2 \leqslant 4/\eta$.

Constructing covers and polynomial optimization. Now, we consider the task of solving (Tangent Cover). Solving this even in the simplest case is not straightforward. An example to keep in mind is the following: suppose we are adding our first state to C_n, which is currently empty. So, there are no "farness" constraints, the first kind of constraint in the program. Then, let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a superposition over computational basis strings with Hamming weight 0 and d:

$$|\psi\rangle = \sqrt{\gamma}\,|0^n\rangle + \sqrt{\frac{1-\gamma}{\binom{n}{d}}} \sum_{\substack{x\in\{0,1\}^n \\ |x|=d}} |x\rangle\,.$$

We are imagining, say, $\gamma = 0.9\eta$. Then $\langle 0^n|\rho|0^n\rangle = \gamma$, so our root state has good fidelity, but not quite enough to be a violation as we desire. (This can indeed happen; though $|0^n\rangle$ comes from the cover C_{n-1}, so $|0^{n-1}\rangle$ has fidelity at least $\eta - \varepsilon$, it is extended by one qubit, which can drop the fidelity to γ or lower.) However, for $\vec{z} = \frac{1}{\sqrt{n}}\vec{1}$,

$$\begin{aligned}\langle \pi_{\vec{z}}|\rho|\pi_{\vec{z}}\rangle &= (1+1/n)^{-n}\Big(\sqrt{\gamma} + \sqrt{\frac{\binom{n}{d}(1-\gamma)}{n^d}}\Big)^2 \\ &\underset{n\text{ large}}{\approx} \frac{1}{e}\Big(\sqrt{\gamma} + \sqrt{(1-\gamma)/d!}\Big)^2,\end{aligned}$$

which can be larger than η even for $d = \Theta(\log(1/\eta)/\log\log(1/\eta))$. Note that $\|\vec{z}\|_2 = 1 \leqslant 4/\eta$, so it is close enough to the root in (Tangent Cover), and our algorithm must be able to recognize this better solution of $|\pi_{\vec{z}}\rangle$. This demands knowledge of ρ in a (quite large) Hamming ball around the root product state. Further, by changing the signs of the $|x\rangle$'s in $|\psi\rangle$, (Tangent Cover) can encode dense d-CSP instances. This suggests that the right approach is a reduction to polynomial optimization.

First, we consider solving (Tangent Cover) when C_m is empty. We can reduce this to low-degree polynomial optimization over the sphere. We start by observing that it suffices to consider the projection of ρ on basis states of low Hamming weight. Concretely, let $\Pi_{\geqslant d}$ be the projection onto the subspace of Hamming weight at least $d = O(1/\eta + \log(1/\varepsilon))$. Then, we show that, provided $\|\vec{z}\| \leqslant 4/\eta$ as in (Tangent Cover),

$$\|\Pi_{\geqslant d}\,|\pi_{\vec{z}}\rangle\| \leqslant \varepsilon\,.$$

This is a Chernoff bound, since the squared norm is the probability that the n Bernoulli random variables sums to at least d, where the probabilities come from the $\vec{z}$'s. So, it suffices to perform state tomography for ρ on the space of low Hamming weight vectors $\rho_d = \Pi_{<d}\rho\Pi_{<d}$, which is computationally efficient because this subspace has dimension $O(n^d)$. We can use ρ_d in place of ρ in the objective because

$$\begin{aligned}|\langle\pi_{\vec{z}}|\,\rho\,|\pi_{\vec{z}}\rangle - \langle\pi_{\vec{z}}|\,\rho_d\,|\pi_{\vec{z}}\rangle| &= \Big|\mathrm{tr}\Big(\rho\big(|\pi_{\vec{z}}\rangle\!\langle\pi_{\vec{z}}| - \Pi_{<d}\,|\pi_{\vec{z}}\rangle\!\langle\pi_{\vec{z}}|\,\Pi_{<d}\big)\Big)\Big| \\ &\leqslant \||\pi_{\vec{z}}\rangle\!\langle\pi_{\vec{z}}| - \Pi_{<d}\,|\pi_{\vec{z}}\rangle\!\langle\pi_{\vec{z}}|\,\Pi_{<d}\|_{\mathrm{op}} \\ &\leqslant 2\||\pi_{\vec{z}}\rangle - \Pi_{<d}\,|\pi_{\vec{z}}\rangle\|_2 \\ &= 2\|\Pi_{\geqslant d}\,|\pi_{\vec{z}}\rangle\|_2 \leqslant \varepsilon\,.\end{aligned}$$

Further, once we have our estimate of ρ_d, the objective function is fully specified explicitly; the rest of the algorithm is classical. Because ρ_d is only supported on low Hamming weight, $\langle\pi_{\vec{z}}|\,\rho_d\,|\pi_{\vec{z}}\rangle$ is a low-degree polynomial up to a normalization factor:

$$\langle\pi_{\vec{z}}|\rho_d\,|\pi_{\vec{z}}\rangle = \underbrace{\frac{1}{\prod_{i\in[m]}(1+|z_i|^2)}}_{(3).(1)} \underbrace{\sum_{x,x'\in\{0,1\}^m} \langle x|\,\rho_d\,|x'\rangle\,(\vec{z}^*)^x(\vec{z})^{x'}}_{(3).(2)}\,. \tag{3}$$

(3).(2) is a degree-$2d$ polynomial in the z_i's and their complex conjugates. Further, when the $|z_i|$'s are small, we can approximate term (3).(1) by $\exp(-\|z\|_2^2)$, which is a bounded scalar variable that we can hardcode into our constraints. While the $|z_i|$'s won't always be small, we can guess the ones that are large, fix their value and use the above approximation on the rest. This is where we pick up an ℓ_∞ constraint on the z_i's, since we must enforce that entries which we do not guess are small. This reduces the algorithm to solving problems of the following form.

$$\max_{\substack{\|\vec{z}\|_2=1\\ \|\vec{z}\|_\infty\leqslant\mu}} p(\vec{z}) = \max_{\substack{\|\vec{z}\|_2=1\\ \|\vec{z}\|_\infty\leqslant\mu}} \sum_{x,x'\in\{0,1\}^m} \langle x|\,\rho_d\,|x'\rangle\,(\vec{z}^*)^x(\vec{z})^{x'}\,.$$

Optimizing low-degree polynomials over the sphere is known to be hard to approximate up to polynomial factors in the worst-case, even when the degree is 4 [8, 13]. However, in our case, $p(\vec{z})$ is not an arbitrary low-degree polynomial, but is quite 'small': the ℓ_2 norm of the coefficients $\langle x|\rho_d|x'\rangle$ is bounded, since it is $\|\rho_d\|_F \leqslant \|\rho\|_F \leqslant \mathrm{tr}(\rho) = 1$. We will show that, in this case, obtaining additive error that scales with ε admits a polynomial time algorithm. Additive-error approximations to max k-CSPs also admit a similar guarantee.

In the general case, we must also deal with the farness constraints in (Tangent Cover), $\mathrm{d}_{\tan}(|\pi_{\vec{z}}\rangle, |\pi_{\vec{a}}\rangle) \geqslant 2/\eta$ for a small number of $|\pi_{\vec{a}}\rangle$'s. Recall that $\mathrm{d}_{\tan}(|\pi_{\vec{a}}\rangle, |\pi_{\vec{z}}\rangle)^2 = \sum_{i\in[n]}|\frac{z_i-a_i}{1+z_i^*a_i}|^2$ by definition. We will not try to directly optimize over these constraints. We use a similar trick as before and show that when the $|z_i|$'s are small, $\mathrm{d}_{\tan}(|\pi_{\vec{a}}\rangle, |\pi_{\vec{z}}\rangle) \approx \|\vec{a} - \vec{z}\|_2$. These constraints essentially only enforce what $\vec{z}$ can be in the low-dimensional subspace spanned by the $\vec{a}$'s. So, we can guess the choice of $\vec{z}$ on this subspace (in addition to the coordinates for which $|z_i|$ is large), and for each guess, solve the problem with the guess hardcoded into the constraints. Putting all the steps together, we show that we can reduce the problem of extending the cover to a polynomial optimization problem, where the ℓ_2 norm of the coefficients is bounded by 1, subject to ℓ_2 and ℓ_∞ constraints.

Optimizing low-degree polynomials over the sphere. We now focus on the algorithmic problem of optimizing a low-degree polynomial over the sphere subject to ℓ_2 and ℓ_∞ constraints.

Our results for polynomial optimization can be thought of as analogs of maximizing dense k-CSPs, only the domain is the sphere instead of the hypercube. The underlying algorithms for max k-CSPs are either based on brute-force search over a dimension-independent number of variables followed by greedily completing the solution or global correlation rounding [2, 35, 36, 53]. One may expect the correlation rounding algorithms for max k-CSPs to generalize straightforwardly to optimize low-degree polynomials over the sphere up to additive error. However, the existing analysis [35] would merely translate to outputting a product state with fidelity $\Omega(\mathrm{OPT}) - \varepsilon$, as opposed to $\mathrm{OPT} - \varepsilon$. One can also try to extend the correlation rounding algorithm of Alev, Jeronimo and Tulsiani [2] to the sphere, but their algorithm obtains a doubly-exponential dependence on k. In contrast, our approach is closer in spirit to the brute-force style algorithm for max k-CSPs, and allows for additional ℓ_2 and ℓ_∞ constraints. Translating our algorithm back to optimizing dense polynomials over the hypercube, we can show that we obtain yet another algorithm that achieves additive error guarantees.

To get the key ideas across, we first consider the unconstrained polynomial optimization problem, reformulated as maximizing the injective norm of a tensor:

$$\max_{x\in\mathbb{C}^m, \|\vec{x}\|_2=1} \langle T, \vec{x}^{\otimes k}\rangle$$

for a tensor T with $\|T\|_F \leqslant 1$. While it is hard to obtain a multiplicative approximation to tensor optimization problems, we show that we can obtain an additive $\varepsilon \cdot \|T\|_F$ approximation in $n^{\mathrm{poly}(1/\varepsilon)}$ time. We begin by observing that there is a $\mathrm{poly}(1/\varepsilon)$-dimensional subspace such that projecting x onto this subspace suffices to obtain an $\varepsilon\|T\|_F$ approximation to the optimum objective value. To see this, we can unfold the tensor T along the first mode to obtain a $m \times m^{k-1}$ matrix M. We can then compute the singular value decomposition of M and let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m \geqslant 0$ be the resulting singular values. There are at most $r = \lceil 1/\varepsilon^2 \rceil$ singular values larger than $\varepsilon \cdot \|T\|_F$. Let $\Pi^{(1)}_{>\varepsilon}$ be the projection on the top-r subspace of M. Then,

$$\left|\left\langle \vec{x}, M(\mathrm{vec}(\vec{x}^{\otimes k-1}))\right\rangle - \left\langle \Pi^{(1)}_{>\varepsilon}\vec{x}, M(\mathrm{vec}(\vec{x}^{\otimes k-1}))\right\rangle\right| \leqslant \varepsilon\|T\|_F.$$

Now, we can repeat the same argument for the remaining modes to obtain projectors $\Pi^{(2)}_{>\varepsilon}, \ldots, \Pi^{(k)}_{>\varepsilon}$. Setting $\Pi_{>\varepsilon}$ to project onto the union of the spans of $\Pi^{(1)}, \ldots, \Pi^{(k)}$, we conclude that

$$\left|\left\langle T, \vec{x}^{\otimes k}\right\rangle - \left\langle T, \Pi_{>\varepsilon}\vec{x}^{\otimes k}\right\rangle\right| \leqslant k\varepsilon\,.$$

In short, *the polynomial can be approximated by projecting $\vec{x}$ onto a small-sized subspace.* To solve this tensor optimization problem, it suffices to brute-force over a fine-enough net on this constant-dimensional subspace and pick the vector that obtains the maximal value.

In general, we need to deal with an optimization problem that involves a dimension-independent number of ℓ_2 constraints and an ℓ_∞ constraint. We handle the ℓ_2 constraints by simply projecting onto the union of the subspace Π and the subspace corresponding to the span of the ℓ_2 constraints. It remains to handle the ℓ_∞ constraint, which takes the form $\|\vec{x}\|_\infty \leqslant \mu$ for some constant $\mu > 0$. Only $1/\mu^2$ of the coordinates can saturate the ℓ_∞ constraint. Since $\mu > 0$ is a constant, we can brute-force over which coordinates saturate the ℓ_∞ constraint. Ultimately, we can still reduce the constrained optimization problem to checking over a net in a constant-dimensional subspace. .

Hardness for agnostic product tomography. Our lower bound starts from the hardness of computing (asymmetric) tensor spectral norm for 4-tensors. In particular, for an $n \times n \times n \times n$ tensor T, computing the spectral norm to additive error $\|T\|_F/\mathrm{poly}(n)$ is NP-hard. We attain our result by reducing tensor optimization to product state learning, essentially inverting the reduction discussed earlier. The main idea is to set the unknown state $\rho = |\psi\rangle\langle\psi|$ where

$$|\psi\rangle = \frac{1}{\|T\|_F}\sum_{i,j,k,l\in[m]} T_{ijkl}\,|e_i\rangle\,|e_j\rangle\,|e_k\rangle\,|e_l\rangle\ .$$

Then, we can show that finding the $4m$-qubit product state

$$|\pi_{\vec{x}}\rangle\,|\pi_{\vec{y}}\rangle\,|\pi_{\vec{u}}\rangle\,|\pi_{\vec{v}}\rangle$$

with optimal fidelity is essentially equivalent to maximizing the tensor form $\langle T, \vec{x}\otimes\vec{y}\otimes\vec{u}\otimes\vec{v}\rangle$. The only additional difficulty is that this equivalence only holds if T is sufficiently flat; our reduction thus requires an additional step where we embed our input T in a larger space and randomly rotate it to make all its entries small without changing the optimal value.

Faster algorithms. In light of the lower bound, one can still ask what additional structure yields faster algorithms. We consider three additional settings: the high-fidelity regime (high overlap with a product state), a bounded number of choices for each qubit, and matrix-product states. In all of these settings, we follow the same overall strategy of sweeping over the qubits, but maintaining a cover becomes significantly easier:

(1) In the high-fidelity regime, the cover can be made to be only *one* state;
(2) In the finite-choices setting, we can simply maintain *all* good product states in the class, instead of a cover over them;
(3) In the MPS setting, we can make our cover one state, though one with a large bond dimension, in some sense capturing many good product states.

For this overview, we focus on the high-fidelity setting. Here, the optimal solution is unique and we do not require maintaining a net. Instead, we only need to maintain a single candidate product state as we sweep across the qubits, performing local updates until convergence.

To illustrate how and why local optimization works, suppose for simplicity that $\rho = |\psi\rangle\langle\psi|$ is a pure state; the mixed state case is similar but involves some additional parameters. Imagine that $|0^n\rangle$ is the current candidate product state (in some appropriately chosen basis), and consider what happens when we express $|\psi\rangle$ in the low-Hamming weight subspace:

$$|\psi\rangle = \alpha\,|0^n\rangle + \delta\,|v_1\rangle + \beta\,|v_{\geqslant 2}\rangle\ .$$

Above, we assume without loss of generality that $|v_1\rangle$ is a normalized state on the subspace of Hamming weight 1, $|v_{\geqslant 2}\rangle$ is a normalized state on the subspace of Hamming weight at least 2, and α, β, δ are all nonnegative reals. It is helpful to express $|\psi\rangle$ this way because δ quantifies local updates that we can make to improve fidelity. By rotating qubit i of our candidate product state away from $|0\rangle$, we can increase the product state fidelity from $|\langle 0^n|\psi\rangle|^2 = \alpha^2$ to $\alpha^2 + \delta^2|\langle e_i|v_1\rangle|^2$, where $|e_i\rangle$ is the string with 1 in position i and 0s elsewhere. Our goal, then, will be to establish that α^2 is close to OPT whenever δ is small, because it implies that local optimization converges efficiently: either δ is large, in which case we can increase the candidate fidelity by a substantial amount, or δ is small, in which case we are near the global optimum.

We prove this by bounding the contributions to product state fidelity from the Hamming weight 0, 1, and $\geqslant 2$ subspaces separately. Consider an arbitrary product state $|\pi\rangle$ that, when measured in the computational basis, gives $|0^n\rangle$ with probability $1-p$ and anything else with probability p. We observe that $|\pi\rangle$ places at most $O(p^2)$ probability on Hamming weight 2 or larger, and use this to upper bound the overlap between $|\pi\rangle$ and $|\psi\rangle$:

$$|\langle\pi|\psi\rangle| \leqslant \alpha(1-\Omega(p)) + \delta\sqrt{p} + O(\beta p)\,.$$

Working out the constants, we find that so long as $\alpha^2 \geqslant 2/3$, the $-\Omega(\alpha p)$ term dominates the $O(\beta p)$ term, leaving us with $|\langle\pi|\psi\rangle| \leqslant$

$\alpha+\delta\sqrt{p} \leqslant \alpha+\delta$. So, *every* product state satisfies $|\langle\pi|\psi\rangle|^2 \leqslant (\alpha+\delta)^2$, and therefore $\mathrm{OPT} \leqslant (\alpha+\delta)^2$.

The above analysis straightforwardly gives rise to a polynomial-time but suboptimal algorithm for finding the closest product state. We briefly summarize the additional tricks that are required to reduce the sample complexity to linear in n.

First, we observe that divide-and-conquer is more efficient than sweeping through one additional qubit at a time. So, the basic structure of the learning algorithm is:

(1) Recursively run the algorithm on the left and right halves of ρ, obtaining product states $|\pi_L\rangle$ and $|\pi_R\rangle$ that have fidelity at least $5/6$ with the respective halves.
(2) Take $|\pi\rangle = |\pi_L\rangle \otimes |\pi_R\rangle$, which satisfies $\langle\pi|\rho|\pi\rangle \geqslant 2/3$.
(3) Run local optimization on $|\pi\rangle$ until convergence.

Second, we improve the bound on $|\langle\pi|\psi\rangle|$ when α^2 is much larger than $2/3$. Namely, we show the alternative bound

$$|\langle\pi|\psi\rangle| \leqslant \alpha + O\left(\frac{\delta^2}{\alpha^2 - 2/3}\right),$$

which ultimately implies that local optimization needs fewer iterations to converge.

Third, we find that one can make larger improvements to the fidelity by updating all n qubits simultaneously, rather than one at a time. We take $\vec{z} \in \mathbb{C}^n$ to be the vector defined by $z_i = \langle e_i|\rho|0^n\rangle$, which captures the mass that ρ places on Hamming weight 1 that is coherent with $|0^n\rangle$. Then we show that a step from $|0^n\rangle$ to $|\pi_{\vec{z}/10}\rangle$ increases the fidelity with ρ by $\Omega(\|\vec{z}\|_2^2)$. Note that in the pure state case $\rho = |\psi\rangle\langle\psi|$, this $\vec{z}$ is precisely $\alpha\delta\,|v_1\rangle$, and therefore the fidelity improvement is $\Omega(\delta^2)$. For comparison, recall that the improvement from a single-qubit update was only $\delta^2 \max_{i\in[n]} |\langle e_i|v_1\rangle|^2$, which can be as small as δ^2/n.

Finally, we establish that it suffices to obtain a relative ℓ^2-error estimate of $\vec{z}$ in order to perform these local updates. In symbols, if we can produce an estimate $\vec{a}$ satisfying $\|\vec{a} - \vec{z}\|_2 \leqslant \|\vec{z}\|_2/3$ (say), then we show that $|\pi_{\vec{a}/10}\rangle$ also increases the fidelity by $\Omega(\|\vec{z}\|_2^2)$. This allows us to cut down some of the cost of the tomography subroutine by varying the error parameter throughout the algorithm, because we can afford to be sloppier when the step size is large.

3 Full Version

Due to space limitations, the technical sections of this work are omitted. The omitted sections can be found in the full version [4].

Acknowledgments

We thank Sitan Chen for helpful discussions at the early stages of this work.

AB is supported by the NSF TRIPODS program (award DMS-2022448). JB is supported by Henry Yuen's AFOSR (award FA9550-21-1-036) and NSF CAREER (award CCF-2144219). WK acknowledges support from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. ZL is supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator, and by NSF Grant CCF-2311733. AL is supported in part by an NSF GRFP and a Hertz Fellowship. RO is supported by ARO grant W911NF2110001 and by a gift from Google Quantum AI. ET is supported by the Miller Institute for Basic Research in Science, University of California, Berkeley.

References

[1] Scott Aaronson. 2020. Shadow tomography of quantum states. *SIAM J. Comput.* 49, 5 (Jan. 2020), STOC18–368–STOC18–394. https://doi.org/10.1137/18m120275x arXiv:1711.01053 [quant-ph]

[2] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. 2019. Approximating constraint satisfaction problems on high-dimensional expanders. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, Baltimore, MD, USA, 180–201. https://doi.org/10.1109/focs.2019.00021 arXiv:1907.07833 [cs.DS]

[3] Anurag Anshu and Srinivasan Arunachalam. 2023. A survey on the complexity of learning quantum states. *Nature Reviews Physics* 6, 1 (Dec. 2023), 59–69. https://doi.org/10.1038/s42254-023-00662-4 arXiv:2305.20069 [quant-ph]

[4] Ainesh Bakshi, John Bostanci, William Kretschmer, Zeph Landau, Jerry Li, Allen Liu, Ryan O'Donnell, and Ewin Tang. 2024. Learning the closest product state. (2024). arXiv:2411.04283 [quant-ph]

[5] Ainesh Bakshi, Ilias Diakonikolas, He Jia, Daniel M. Kane, Pravesh K. Kothari, and Santosh S. Vempala. 2022. Robustly learning mixtures of k arbitrary Gaussians. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC '22)*. ACM, New York, NY, USA, 1234–1247. https://doi.org/10.1145/3519935.3519953 arXiv:2012.02119 [cs.DS]

[6] Ainesh Bakshi and Pravesh K. Kothari. 2021. List-decodable subspace recovery: dimension independent error in polynomial time. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 1279–1297. https://doi.org/10.1137/1.9781611976465.78 arXiv:2002.05139 [cs.DS]

[7] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. 2024. Learning quantum Hamiltonians at any temperature in polynomial time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24)*. ACM, New York, NY, USA, 1470–1477. https://doi.org/10.1145/3618260.3649619 arXiv:2310.02243 [quant-ph]

[8] Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. 2012. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing (STOC'12)*. ACM, New York, NY, USA, 307–326. https://doi.org/10.1145/2213977.2214006 arXiv:1205.4484 [cs.CC]

[9] Boaz Barak, Jonathan A. Kelner, and David Steurer. 2015. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing (STOC '15)*. ACM, New York, NY, USA, 143–151. https://doi.org/10.1145/2746539.2746605 arXiv:1407.1543 [cs.DS]

[10] Boaz Barak, Pravesh K Kothari, and David Steurer. 2017. Quantum entanglement, sum of squares, and the log rank conjecture. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC '17)*. ACM, New York, NY, USA, 975–988. https://doi.org/10.1145/3055399.3055488 arXiv:1701.06321 [quant-ph]

[11] Boaz Barak, Prasad Raghavendra, and David Steurer. 2011. Rounding semidefinite programming hierarchies via global correlation. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE, Palm Springs, CA, USA, 472–481. https://doi.org/10.1109/focs.2011.95 arXiv:1104.4680 [cs.DS]

[12] Mikhail Belkin and Kaushik Sinha. 2015. Polynomial learning of distribution families. *SIAM J. Comput.* 44, 4 (Jan. 2015), 889–911. https://doi.org/10.1137/13090818x arXiv:1004.4864 [cs.LG]

[13] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. 2017. Weak decoupling, polynomial folds and approximate optimization over the sphere. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, Berkeley, CA, USA, 1008–1019. https://doi.org/10.1109/focs.2017.97 arXiv:1611.05998 [cs.DS]

[14] Fernando G. S. L. Brandão and Aram W. Harrow. 2016. Product-state approximations to quantum states. *Communications in Mathematical Physics* 342, 1 (Jan. 2016), 47–80. https://doi.org/10.1007/s00220-016-2575-1 arXiv:1310.0017 [quant-ph]

[15] Costin Bădescu and Ryan O'Donnell. 2024. Improved quantum data analysis. *TheoretiCS* Volume 3 (March 2024), 34 pages. https://doi.org/10.46298/theoretics.24.7 arXiv:2011.10908 [quant-ph]

[16] Clément L. Canonne. 2020. *A survey on distribution testing: Your data is big. But is it blue?* Number 9 in Graduate Surveys. Theory of Computing Library. 1–100 pages. https://doi.org/10.4086/toc.gs.2020.009

[17] Garnet Kin-Lic Chan. 2024. Spiers Memorial Lecture: Quantum chemistry, classical heuristics, and quantum advantage. *Faraday Discuss.* 254 (2024), 11–52. Issue 0. https://doi.org/10.1039/D4FD00141A

[18] Moses Charikar, Jacob Steinhardt, and Gregory Valiant. 2017. Learning from untrusted data. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC '17)*. ACM, New York, NY, USA, 47–60. https://doi.org/10.1145/3055399.3055491 arXiv:1611.02315 [cs.LG]
[19] Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. 2024. Stabilizer bootstrapping: a recipe for efficient agnostic tomography and magic estimation. (Aug. 2024). arXiv:2408.06967 [quant-ph]
[20] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. 2010. Efficient quantum state tomography. *Nature Communications* 1, 1 (Dec. 2010), 7 pages. https://doi.org/10.1038/ncomms1147 arXiv:1101.4366 [quant-ph]
[21] Gabriele De Chiara and Anna Sanpera. 2018. Genuine quantum correlations in quantum many-body systems: a review of recent progress. *Reports on Progress in Physics* 81, 7 (June 2018), 074002. https://doi.org/10.1088/1361-6633/aabf61 arXiv:1711.07824 [quant-ph]
[22] Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. 2019. Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput.* 48, 2 (Jan. 2019), 742–864. https://doi.org/10.1137/17m1126680 arXiv:1604.06443 [cs.DS]
[23] Shmuel Friedland and Lek-Heng Lim. 2017. Nuclear norm of higher-order tensors. *Math. Comp.* 87, 311 (Sept. 2017), 1255–1281. https://doi.org/10.1090/mcom/3239 arXiv:1410.6072 [cs.CC]
[24] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. 2024. Agnostic tomography of stabilizer product states. (April 2024). arXiv:2404.03813 [quant-ph]
[25] Aram W. Harrow and Ashley Montanaro. 2013. Testing product states, quantum Merlin-Arthur games and tensor optimization. *J. ACM* 60, 1 (Feb. 2013), 1–43. https://doi.org/10.1145/2432622.2432625 arXiv:1001.0017 [quant-ph]
[26] Samuel B. Hopkins, Gautam Kamath, Mahbod Majid, and Shyam Narayanan. 2023. Robustness implies privacy in statistical estimation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23)*. ACM, New York, NY, USA, 497–506. https://doi.org/10.1145/3564246.3585115 arXiv:2212.05015 [cs.DS]
[27] Samuel B. Hopkins, Jonathan Shi, and David Steurer. 2015. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory (Proceedings of Machine Learning Research, Vol. 40)*. PMLR, Paris, France, 956–1006. arXiv:1507.03269 [cs.LG]
[28] Hsin-Yuan Huang, Richard Kueng, and John Preskill. 2020. Predicting many properties of a quantum system from very few measurements. *Nature Physics* 16, 10 (June 2020), 1050–1057. https://doi.org/10.1038/s41567-020-0932-7 arXiv:2002.08953 [quant-ph]
[29] Richard Jozsa and Marrten Van Den Nest. 2014. Classical simulation complexity of extended Clifford circuits. *Quantum Info. Comput.* 14, 7 & 8 (May 2014), 633–648. https://doi.org/10.26421/QIC14.7-8-7 arXiv:1305.6190 [quant-ph]
[30] Sushrut Karmalkar, Adam Klivans, and Pravesh Kothari. 2019. List-decodable linear regression. In *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates, Inc., Vancouver, BC, Canada, 10 pages. arXiv:1905.05679 [cs.DS]
[31] Pravesh K. Kothari, Jacob Steinhardt, and David Steurer. 2018. Robust moment estimation and improved clustering via sum of squares. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC '18)*. ACM, New York, NY, USA, 1035–1046. https://doi.org/10.1145/3188745.3188970
[32] Kevin A. Lai, Anup B. Rao, and Santosh Vempala. 2016. Agnostic estimation of mean and covariance. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, New Brunswick, NJ, USA, 665–674. https://doi.org/10.1109/focs.2016.76 arXiv:1604.06968 [cs.DS]
[33] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M. Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, Michael Kastoryano, Ryan Babbush, John Preskill, David R. Reichman, Earl T. Campbell, Edward F. Valeev, Lin Lin, and Garnet Kin-Lic Chan. 2023. Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry. *Nature Communications* 14, 1 (April 2023), 7 pages. https://doi.org/10.1038/s41467-023-37587-6 arXiv:2208.02199 [physics.chem-ph]
[34] Allen Liu and Ankur Moitra. 2021. Settling the robust learnability of mixtures of Gaussians. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21)*. ACM, New York, NY, USA, 518–531. https://doi.org/10.1145/3406325.3451084 arXiv:2011.03622 [cs.DS]
[35] Pasin Manurangsi and Prasad Raghavendra. 2017. A birthday repetition theorem and complexity of approximating dense CSPs. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 78:1–78:15. https://doi.org/10.4230/LIPIcs.ICALP.2017.78 arXiv:1607.02986 [cs.CC]
[36] Claire Mathieu and Warren Schudy. 2008. Yet another algorithm for dense max cut: Go greedy. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (San Francisco, California) *(SODA '08)*. Society for Industrial and Applied Mathematics, USA, 176–182.
[37] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. 2018. Barren plateaus in quantum neural network training landscapes. *Nature Communications* 9, 1 (Nov. 2018), 6 pages. https://doi.org/10.1038/s41467-018-07090-4
[38] Florian Mintert, Marek Kuś, and Andreas Buchleitner. 2005. Concurrence of mixed multipartite quantum states. *Physical Review Letters* 95, 26 (Dec. 2005), 260502. https://doi.org/10.1103/physrevlett.95.260502 arXiv:quant-ph/0411127
[39] Ankur Moitra and Gregory Valiant. 2010. Settling the polynomial learnability of mixtures of Gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, Las Vegas, NV, USA, 93–102. https://doi.org/10.1109/focs.2010.15 arXiv:1004.4223 [cs.LG]
[40] Shyam Narayanan. 2024. Improved algorithms for learning quantum hamiltonians, via flat polynomials. (2024). arXiv:2407.04540 [quant-ph]
[41] Román Orús, Sébastien Dusuel, and Julien Vidal. 2008. Equivalence of critical scaling laws for many-body entanglement in the Lipkin-Meshkov-Glick model. *Physical Review Letters* 101, 2 (July 2008), 025701. https://doi.org/10.1103/physrevlett.101.025701 arXiv:0803.3151 [cond-mat.other]
[42] Román Orús and Tzu-Chieh Wei. 2010. Visualizing elusive phase transitions with geometric entanglement. *Physical Review B* 82, 15 (Oct. 2010), 155120. https://doi.org/10.1103/physrevb.82.155120 arXiv:0910.2488 [cond-mat.str-el]
[43] D. Perez-Garcia, F. Verstraete, M.M. Wolf, and J.I. Cirac. 2007. Matrix product state representations. *Quantum Information and Computation* 7, 5 & 6 (July 2007), 401–430. https://doi.org/10.26421/qic7.5-6-1 arXiv:quant-ph/0608197
[44] Prasad Raghavendra, Satish Rao, and Tselil Schramm. 2017. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC '17)*. ACM, New York, NY, USA, 121–131. https://doi.org/10.1145/3055399.3055417 arXiv:1605.00058 [cs.DS]
[45] Prasad Raghavendra and Ning Tan. 2012. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Kyoto, Japan, 373–387. https://doi.org/10.1137/1.9781611973099.33 arXiv:1110.1064 [cs.DS]
[46] Prasad Raghavendra and Morris Yau. 2020. List decodable learning via sum of squares. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Salt Lake City, UT, USA, 161–180. https://doi.org/10.1137/1.9781611975994.10 arXiv:1905.04660 [cs.DS]
[47] Ulrich Schollwöck. 2011. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics* 326, 1 (2011), 96–192. https://doi.org/10.1016/j.aop.2010.09.012
[48] Mehdi Soleimanifar and John Wright. 2022. Testing matrix product states. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 1679–1701. https://doi.org/10.1137/1.9781611977073.68 arXiv:2201.01824 [quant-ph]
[49] Eliran Subag. 2020. Following the ground states of full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics* 74, 5 (June 2020), 1021–1044. https://doi.org/10.1002/cpa.21922 arXiv:1812.04588 [math.PR]
[50] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. 1997. Quantifying entanglement. *Physical Review Letters* 78, 12 (March 1997), 2275–2279. https://doi.org/10.1103/physrevlett.78.2275 arXiv:quant-ph/9702027
[51] Tzu-Chieh Wei, Dyutiman Das, Swagatam Mukhopadyay, Smitha Vishveshwara, and Paul M. Goldbart. 2005. Global entanglement and quantum criticality in spin chains. *Physical Review A* 71, 6 (June 2005), 060305. https://doi.org/10.1103/physreva.71.060305 arXiv:quant-ph/0405162
[52] Tzu-Chieh Wei and Paul M. Goldbart. 2003. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Physical Review A* 68, 4 (Oct. 2003), 042307. https://doi.org/10.1103/physreva.68.042307 arXiv:quant-ph/0307219
[53] Grigory Yaroslavtsev. 2014. Going for speed: sublinear algorithms for dense r-CSPs. (2014). arXiv:1407.7887 [cs.DS]

Received 2024-11-04; accepted 2025-02-01