MIT-CTP/4587

An improved semidefinite programming hierarchy for testing entanglement

Aram W. Harrow, Anand Natarajan, and Xiaodi Wu

MIT Center for Theoretical Physics

We present a stronger version of the Doherty-Parrilo-Spedalieri (DPS) hierarchy of approximations for the set of separable states. Unlike DPS, our hierarchy converges exactly at a finite number of rounds for any fixed input dimension. This yields an algorithm for separability testing which is singly exponential in dimension and polylogarithmic in accuracy. Our analysis makes use of tools from algebraic geometry, but our algorithm is elementary and differs from DPS only by one simple additional collection of constraints.

I. INTRODUCTION

Entanglement is one of the key features that distinguishes quantum information from classical information. One particularly basic and important problem in the theory of entanglement is to determine whether a given mixed state ρ is entangled or separable. Via standard techniques of convex optimization, this problem is roughly equivalent to maximizing a linear function over the set of separable states [1, 2]. Indeed, it has close relations with a variety of problems, including estimating channel capacities, analyzing two-prover proof systems, finding the ground-state energy in the mean-field approximation, finding the least entangled pure state in a subspace, etc. as well as problems not obviously related to quantum mechanics such as planted clique, the unique games problem and small-set expansion [3].

However, there is no simple test for determining whether a state is entangled. Indeed not only are tests such as the PPT (positive partial transpose) condition known to have arbitrarily large error [4], but computational hardness results show that any test implementable in time polynomial in the dimension must be highly inaccurate, given the plausible assumption that 3-SAT requires exponential time [3, 5]. These limitations indicate that separability tests cannot be as efficient as, say, a test for correlation, or a calculation of the largest eigenvalue of a matrix.

The main open question is whether algorithms exist that match these hardness results, or whether further hardness results can be found. The two leading algorithmic frameworks are ϵ-nets and semidefinite programming (SDP) hierarchies. There are two regimes in which these come close to matching the known hardness results. Let n denote the dimension of the states we examine. Informally speaking, the well-studied regimes are the constant-error regime, where there are both algorithms and hardness results with time $n^{\Theta(\log n)}$ (although important caveats exist, discussed below), and the $1/\operatorname{poly}(n)$ regime, where the algorithms and hardness results together suggest that the complexity is exponential in n.

In this paper we consider the regime of much lower error. Specifically, if ϵ is the error allowed, we will focus on the scaling of error with ϵ rather than n. In other settings, such as infinite translationally invariant Hamiltonians, it is possible for the complexity to grow rapidly with $1/\epsilon$ even for fixed local dimension [6]. Another example closer to the current work is [7], which showed that approximating quantum interactive proofs to high accuracy (specifically with the bits of precision polynomial in the message dimension) corresponds to the complexity class EXP rather than PSPACE. However, for separability testing or for the corresponding complexity class $\mathsf{QMA}(2)$, we will give evidence that the complexity does not increase when ϵ becomes exponentially small in the dimension.[8]

Our main contribution is to describe a pair of classical algorithms for the separability problem. In the high-accuracy limit both run in time $\exp(\operatorname{poly}(n))\operatorname{poly}\log(1/\epsilon)$. One is based on quantifier elimination [9] and is simple, but does not appear to yield new insights into the problem. The second algorithm is based on an SDP hierarchy due to Doherty, Parrilo and Spedalieri (DPS) [10]. Like DPS, our algorithm runs in time $n^{O(k)}$ (or more precisely $\operatorname{poly}(\binom{n+k-1}{k})$) for what is called the k^{th} "level" of the hierarchy. As k is increased our algorithm, like that of DPS, becomes more accurate. Indeed, for any fixed value of k our algorithm performs at least as well as that of DPS. However, unlike DPS, our hierarchy always converges exactly in a finite number of steps, which we can upper bound by $\exp(\operatorname{poly}(n))$. Taking into account numerical error yields an algorithm again running in time $\exp(\operatorname{poly}(n)) \operatorname{poly}\log(1/\epsilon)$. Thus our algorithm is, for the first time, a single SDP hierarchy which matches or improves upon the best known performance of previous algorithms at each scale of ϵ.

The fact that our algorithm is a semidefinite program gives it further advantages. One very useful property of semidefinite programs is duality. In our algorithm, both the primal and dual problems have useful interpretations in terms of quantum information. On the primal side, our algorithm can be viewed as searching over symmetric mixed states over an extended system obtained by adding copies of the individual subsystems. In this light, our convergence bounds can be viewed as new monogamy relations: we show that if a state is symmetric under exchange of subsystems and satisfies certain other conditions, then if there are enough copies of each subsystem, then none of the subsystems can be entangled with each other. On the dual side, every feasible point of the dual is an entanglement witness operator. Indeed, our algorithm yields a new class of entanglement witnesses, as discussed in Section III D. Duality is also useful in practice, since a feasible solution to the dual can certify the correctness of the primal, and vice versa.

SDP hierarchies are also used for discrete optimization problems, such as integer programming [11]. In that case, it is known that the n^{th} level of most SDP hierarchies provides the exact answer to optimization problems on n bits (e.g. see Lemma 2.2 of [11]). By contrast, neither the DPS hierarchy nor the more general Sum-of-Squares SDP hierarchy will converge exactly at any finite level for general objective functions [10]. Our result can be seen as a continuous analogue of the exact convergence achievable for discrete optimization.

The main idea of our algorithm is that entanglement testing can be viewed as a convex optimization problem, and thus the solution should obey the KKT (Karush-Kuhn-Tucker) conditions. Thus we can WLOG add these as constraints. It was shown in [12] that for general polynomial optimization problems, adding the KKT conditions yields an SDP hierarchy with finite convergence. Moreover, the number of levels necessary for convergence is a function only of the number of variables and the degrees of the objective and constraint polynomials. However, the proof of convergence presented in [12] gives a very high bound on the number of levels (triply exponential in n or worse). In contrast, we obtain a bound in the number of levels that is singly exponential in n. We use tools from algebraic geometry (Bézout's and Bertini's Theorem) to show that generically, adding the KKT conditions reduces the feasible set of our optimization problem to isolated points. Then, using tools from computational algebra (Gröbner bases), we show that low levels of the SDP hierarchy can effectively search over this finite set. Although we use genericity in the analysis, our algorithm works for all inputs.

While some of these techniques have been used to analyze SDP hierarchies in the past, they have generally not been applied to the problems arising in quantum information. We hope that they find future application to understanding entanglement witnesses, monogamy of entanglement and related phenomena.

Our main contribution is an improved version of the DPS hierarchy which we describe in Section III. It is always at least as stringent as the DPS hierarchy, and in Theorem 3 we show that it outperforms DPS by converging exactly at a finite level, depending on the input dimension. We also present numerical evidence in Section III E that the improved hierarchy outperforms DPS even at the lowest nontrivial level for systems of small dimension.

II. BACKGROUND

A. Separability testing

This section introduces notation and reviews previous work on the complexity of the separability testing problem. Define $\mathrm{Sep}(n,k) := \mathrm{conv}\{|\psi_1\rangle\langle\psi_1|\otimes\cdots\otimes|\psi_k\rangle\langle\psi_k| : |\psi_1\rangle,\ldots,|\psi_k\rangle \in B(\mathbb{C}^n)\}$, where $\mathrm{conv}(S)$ denotes the convex hull of a set S (i.e. the set of all finite convex combinations of elements of S) and $B(V)$ denotes the set of unit vectors in a vector space V. States in $\mathrm{Sep}(n,k)$ are called separable, and those not in $\mathrm{Sep}(n,k)$ are entangled. Given a Hermitian matrix M, we define

$$h_{\mathrm{Sep}(n,k)}(M) := \max\{\mathrm{Tr}[M\rho] : \rho \in \mathrm{Sep}(n,k)\}. \tag{1}$$

We will often abbreviate $\mathrm{Sep} := \mathrm{Sep}(n,2)$ where there is no ambiguity. More generally if K is a convex set, we can define $h_K(x) := \max\{\langle x,y\rangle : y \in K\}$.

A classic result in convex optimization [1] holds that approximating h_K is roughly equivalent in difficulty to the weak membership problem for K: namely, determining whether $x \in K$ or whether $\mathrm{dist}(x,K) > \epsilon$ given the promise that one of these holds. This was strengthened in the context of the set Sep by Gharibian [13] to show that this equivalence holds when $\epsilon \leq 1/\mathrm{poly}(n)$. Thus, in what follows we will treat entanglement testing (i.e. the weak membership problem for Sep) as equivalent to the optimization problem in (1).

1. *Related problems*

A large number of other optimization problems are also equivalent to h_{Sep}, or closely related in difficulty. Many of these are surveyed in [3]. One that will particularly useful will be the optimization problem $h_{\mathrm{ProdSym}(n,k)}$, defined in terms of the set $\mathrm{ProdSym}(n,k) := \mathrm{conv}\{(|\psi\rangle\langle\psi|)^{\otimes k} : |\psi\rangle \in B(\mathbb{C}^n)\}$. In Corollary 14 of [3] (see specifically explanation (2) there) it was proven that for any n^2-dimensional M there exists M' with dimension $4n^2$ satisfying

$$h_{\mathrm{ProdSym}(2n,2)}(M') = \frac{1}{4} h_{\mathrm{Sep}(n,2)}(M). \tag{2}$$

Thus an algorithm for h_{ProdSym} implies an algorithm of similar complexity for h_{Sep}. In the body of our paper, we will describe an algorithm for the mathematically simpler h_{ProdSym}, with the understanding that it also covers the more widely used h_{Sep}.

We will not fully survey the applications of separability testing, but briefly mention two connections. First, $h_{\mathrm{Sep}(2^n,k)}$ is closely related to the complexity class $\mathsf{QMA_n(k)}$ in which k unentangled provers send n-qubit states to a verifier. If the verifier's measurement is M (which might be restricted, e.g. by being the result of a short quantum circuit) then the maximum acceptance probability is precisely $h_{\mathrm{Sep}(2^n,k)}(M)$. Thus the complexity of h_{Sep} is closely related to the complexity of multiple-Merlin proof systems. See [14] for a classical analogue of these proof systems, and a survey of recent open questions.

Second, h_{Sep} is closely related to the problems of estimating the $2 \to 4$ norm of a matrix, finding the least-expanding small set in a graph and estimating the optimum value of a unique game [15]. These problems in turn relate to the approximation complexity of constraint satisfaction problems, which are an extremely general class of discrete optimization problems. They are currently known only to be of intermediate complexity (i.e. only subexponential-time algorithms are known), and are the subject of intense research. One of leading approaches to these problems has been SDP hierarchies, but here too it is generally unknown how well these hierarchies perform or which features are important to their success.

2. Previous algorithms and hardness results

Algorithms and hardness results for estimating $h_{\mathrm{Sep}(n,2)}(M)$ can be classified by (a) the approximation error ϵ, and (b) assumptions (if any) for the matrix M. In what follows we will assume always that $0 \leq M \leq I$. Define 3-SAT$[m]$ to be the problem of solving a 3-SAT instance with m variables and $O(m)$ clauses. The exponential-time hypothesis (ETH) [16] posits that 3-SAT$[m]$ requires time $2^{\Omega(m)}$ to solve.

The first group of hardness results [4, 5, 17–19] for $h_{\mathrm{Sep}(n,2)}$ have $\epsilon \sim 1/\operatorname{poly}(n)$ and yield reductions from 3-SAT$[n]$. The strongest of these results [5] achieves this with $\epsilon \sim 1/n \operatorname{poly}\log(n)$. As discussed above, there are algorithms that come close to matching this. Taking $k = n/\sqrt{\epsilon}$ in the DPS hierarchy achieves error ϵ (see [20]) in time $(n/\sqrt{\epsilon})^{O(n)}$, which is $n^{O(n)}$ when $\epsilon = 1/\operatorname{poly}(n)$. An even simpler algorithm is to enumerate over an ϵ-net over the pure product states on $\mathbb{C}^n \otimes \mathbb{C}^n$. Such a net has size $(1/\epsilon)^{O(n)}$, which again would yield a run-time of $n^{O(n)}$ if $\epsilon = 1/\operatorname{poly}(n)$. Thus neither algorithm nor the hardness result could be significantly improved without violating the ETH. However, the value of ϵ in the hardness result could conceivably be reduced.

The second body of work has concerned the case when ϵ is a constant. Here the existing evidence points to a much lower complexity. Constant-error approximations for $h_{\mathrm{Sep}(n,\sqrt{n}\operatorname{poly}\log(n))}(M)$ were shown to be as hard as 3-SAT$[n]$ in [21] and in [22] this was shown to still hold when M is a Bell measurement (i.e. each system is independently measured and the answers are then classically processed). This was extended to bipartite separability in [3] which showed the 3-SAT$[n]$-hardness of approximating $h_{\mathrm{Sep}(\exp(\sqrt{n}\operatorname{poly}\log(n)),2)}(M)$ to constant accuracy. There it was shown that M could be taken to be separable (i.e. of the form $\sum_i A_i \otimes B_i$ with $A_i, B_i \geq 0$) without loss of generality. Scaling down this means that $h_{\mathrm{Sep}(n,2)}$ requires time $n^{\tilde{\Omega}(\log(n))}$ assuming the ETH. On the algorithms side, $O(\log(n)/\epsilon^2)$ levels of the DPS hierarchy are known [23–25] to suffice when M is a 1-LOCC measurement (i.e. separable with the extra assumption that $\sum_i A_i \leq I$). This also yields a runtime of $n^{O(\log(n)/\epsilon^2)}$, but does not match the hardness result of [3] because of the 1-LOCC assumption. Similar results are also achievable using ϵ-nets [26, 27]. One setting where the hardness result is known to be tight is when there are many provers. When M is implemented by $k-1$ parties measuring locally and sending a message to the final party, [25] showed that DPS could approximate the value of $h_{\mathrm{Sep}(n,k)}(M)$ in time $\exp(k^2 \log^2(n)/\epsilon^2)$. This nearly matches the hardness result of [22] described above. The same runtime was recently shown to work for a larger class of M in [28].

B. Sum-of-squares hierarchies

Here we review the general method of sum-of-squares relaxations for polynomial optimization problems. In this section, all variables are real and all polynomials have real coefficients, unless otherwise stated. To start with, let $g_1(x), \ldots g_k(x)$ be polynomials in n variables and define $V(I) = \{x \in \mathbb{R}^n : \forall_i g_i(x) = 0\}$. This notation reflects the fact that $V(I)$ is the variety corresponds to the ideal I generated by $g_1(x), \ldots, g_m(x)$; see Appendix A for definitions and more background on algebraic geometry.

Now given another polynomial $f(x)$, suppose we would like to prove that $f(x)$ is nonnegative for all $x \in V(I)$. One way to do this would be write f as

$$f(x) = \sum_j a_j(x)^2 + \sum_i b_i(x) g_i(x), \tag{3}$$

for polynomials $\{a_j(x)\}, \{b_i(x)\}$. The first term on the RHS is a sum of squares, and is thus non-negative everywhere, while the second term is zero everywhere on $V(I)$. Thus, if such a decomposition for $f(x)$ exists, it must be nonnegative on $V(I)$. Such a decomposition is thus called a *sum-of-squares (SOS) certificate* for the nonnegativity of f on $V(I)$.

A natural question to ask is whether all nonnegative polynomials on S have a SOS certificate. A positive answer to this question is provided under certain conditions by Putinar's Positivstellensatz [29]. One such condition is the *Archimedean condition*, which asserts that there exists a constant $R > 0$ and a sum-of-squares polynomial $s(x)$ such that

$$R - \sum_i x_i^2 - s(x) \in I. \tag{4}$$

Equivalently we could say that there is a SOS proof of $x \in V(I) \Rightarrow \sum_i x_i^2 \leq R$. This condition generally holds whenever $V(I)$ is a manifestly compact set. In this case, we have the following formulation of Putinar's Positivstellensatz from Theorem A.4 of [12].

Theorem 1 (Putinar)**.** *Let I be a polynomial ideal satisfying the Archimedean condition and $f(x)$ a polynomial with $f(x) > 0$ for all $x \in V(I) \cap \mathbb{R}^n$. Then there exists a sum-of-squares polynomial $\sigma(x)$ and a real polynomial $g(x) \in I$ such that*

$$f(x) = \sigma(x) + g(x).$$

Neither Putinar's Positivstellensatz, nor the Archimedean condition, put any bound on the degree of the SOS certificate. Now suppose we would like to solve a general polynomial optimization problem:

$$\begin{array}{ll} \max & f(x) \\ \text{subject to} & g_i(x) = 0 \ \forall i \end{array}. \tag{5}$$

We can rewrite this in terms of polynomial positivity as follows:

$$\begin{array}{ll} \min & \nu \\ \text{such that} & \nu - f(x) \geq 0. \\ \text{whenever} & g_i(x) = 0 \ \forall i \end{array} \tag{6}$$

Now, if the ideal $\langle \{g_i(x)\} \rangle$ generated by the constraints obeys the Archimedean condition, then Putinar's Positivstellensatz means that this problem is equivalent to

$$\begin{array}{ll} \min & \nu \\ \text{such that} & \nu - f(x) = \sigma(x) + \sum_i b_i(x) g_i(x), \end{array} \tag{7}$$

where $\sigma(x)$ is SOS and the polynomials $b_i(x)$ are arbitrary. If we allow $\sigma(x)$ and $b_i(x)$ to have arbitrarily high degrees, then the problem in this form is exactly equivalent to the original problem, but it involves optimizing over an infinite number of variables. However, if we limit the degrees, so that $\deg(\sigma(x)), \deg(b_i(x)g_i(x)) \leq 2D$ for some integer D, then we obtain a problem over a finite number of variables. As we increase D, we get a hierarchy of optimization problems over increasingly more variables, which must converge to the original problem.

It remains to show how to perform the optimization over a degree-$2D$ sum of squares certificate. It turns out that this optimization can be expressed as a *semidefinite program*. The idea is that any polynomial $g(x)$ of degree $2D$ can be represented as a quadratic form $m^T Q m$, where m is the vector of monomials of degree up to $2D$. Moreover, the polynomial $g(x)$ is SOS iff the matrix

Q of the corresponding quadratic form is positive semidefinite. One direction of this equivalence is as follows. If $g(x) = \sum_i h_i(x)^2$ then each $h_i(x) = \langle \vec{h}_i, m\rangle$ for some vector $\vec{h}_i$, and we have $Q = \sum_i \vec{h}_i \vec{h}_i^T$. The reverse direction follows from the fact that any psd Q can be decomposed in this way.

The SDP associated with the optimization in (7) is

$$\begin{aligned} &\min_{\nu, b_{i\alpha}\in\mathbb{R}} && \nu \\ &\text{such that} && \nu A_0 - F - \sum_{i\alpha} b_{i\alpha} G_{i\alpha} \succeq 0. \end{aligned} \tag{8}$$

Here A_0 is the matrix corresponding to the constant polynomial 1, F is the matrix corresponding to $f(x)$, α is a multi-index labeling monomials, and $G_{i\alpha}$ is the matrix representing the polynomial $x_1^{\alpha_1} \ldots x_n^{\alpha_n} g_i(x)$. These matrices have dimension $m \times m$, where m is the number of monomials of degree at most D. For n variables, $m = \binom{n+D}{D}$. There exist efficient algorithms to solve SDPs: if desired numerical precision is ϵ, and all feasible solutions have norm bounded by a constant R, then the running time for an SDP over $m \times m$ matrices is $O(\text{poly}(m)\,\text{poly}\log(R/\epsilon))$. For a more detailed discussion of SDP complexity, see e.g. [1].

These general techniques were applied to the separability testing problem by Doherty, Parrilo and Spedalieri in [10]. We refer to resulting SDP as the DPS relaxation. For a state ρ^{AB}, the level-k DPS relaxation asks whether there exists an extension $\tilde{\rho}^{A_1\ldots A_k B_1 \ldots B_k}$ invariant under left or right-multiplying by any permutation of the A or B systems and that remains PSD under transposing any subset of the systems. This latter condition is called Positivity under Partial Tranpose (PPT). It is straightforward to see that searching for such a $\tilde{\rho}$ can be achieved by an SDP of size $n^{O(k)}$. In [20] it was proven that the level-k DPS relaxation produces states within trace distance $O(n^2/k^2)$ of the set of separable states. Of course this bound is vacuous for $k < n$, but limited results are known in this case as well; cf. the discussion in II A 2.

Often weaker forms of DPS are analyzed. For example, we might demand only that an extension of the form $\tilde{\rho}^{AB_1\ldots B_k}$ exist, or might drop the PPT condition. Many proof techniques (e.g. those in [23] and followup papers) do not take advantage of the PPT condition, for example, although it is known that without it the power of the DPS relaxation will be limited (see e.g. [30]). Our approach will be to instead *add* constraints to DPS.

III. RESULTS

A. Separability as polynomial optimization

As discussed in Section II A 1, a number of problems in entanglement can be reduced to the problem $h_{\text{ProdSym}(n,d)}$:

$$\max_{\rho\in\text{ProdSym}(n,d)} \quad \text{Tr}[M\rho]. \tag{9}$$

Since $\text{ProdSym}(n,d)$ is a convex set, the maximum will be attained on the boundary, which is the set of pure product states $\rho = (|a\rangle\langle a|)^{\otimes d}$. We can rephrase the optimization in terms of the components of this pure product state.

$$\begin{aligned} &\max_{a\in\mathbb{C}^n} && \sum_{i_1\ldots i_k j_1 \ldots j_d} M_{(i_1\ldots i_d),(j_1\ldots j_d)} a_{i_1}^* \ldots a_{i_d}^* a_{j_1} \ldots a_{j_d} \\ &\text{subject to} && ||a||^2 = 1. \end{aligned} \tag{10}$$

This is an optimization problem over the complex vector space $\mathbb{C}^n$. We can convert it to a real optimization problem over $\mathbb{R}^{2n}$ by explicitly decomposing the complex vectors into real and imaginary parts. Since the matrix M is hermitian, the objective function in (10) is a real polynomial in the real and imaginary parts of a. Thus, we can write the problem as

$$\begin{aligned} \max_{x\in\mathbb{R}^{2n}} \quad & \sum_{i_1\ldots i_d j_1\ldots j_d} \tilde{M}_{(i_1\ldots i_d),(j_1\ldots j_d)} x_{i_1}\ldots x_{i_d} x_{j_1}\ldots x_{j_d} \\ \text{subject to} \quad & \|x\|^2 - 1 = 0 \end{aligned} \tag{11}$$

We will denote this problem by $h_{\mathrm{ProdSym}(\mathbb{R},2n,d)}(\tilde{M})$. Here the matrix $\tilde{M}$ has dimension $(2n)^d \times (2n)^d$. We can alternatively view $\tilde{M}$ as an object with $2d$ indices, each of which ranges from 1 to $2n$. We call this a *tensor of rank* $2d$. Without loss of generality, we can assume that $\tilde{M}$ is completely symmetric under all permutations of the indices. Henceforth, we will only work with real variables, so we will drop the tilde and just write M. For compactness' sake we will use the notation $\langle M, x^{\otimes 2d}\rangle$ to mean the contraction of M, viewed as a rank $2d$ tensor, with $2d$ copies of the vector x. In this notation, the problem $h_{\mathrm{ProdSym}(\mathbb{R},2n,d)}(M)$ becomes:

$$\begin{aligned} \max_{x\in\mathbb{R}^n} \quad & f_0(x) \equiv \langle M, x^{\otimes 2d}\rangle \\ \text{subject to} \quad & f_1(x) \equiv \|x\|^2 - 1 = 0 \end{aligned} \tag{12}$$

Our first algorithm for this problem uses quantifier elimination [9] to solve (12) in a black-box fashion. This yields an algorithm with runtime $d^{O(n)} \operatorname{poly}\log(1/\epsilon)$.

Theorem 2. *There exists an algorithm to estimate* (12) *to multiplicative accuracy* ϵ *in time* $d^{O(n)} \operatorname{poly}\log(1/\epsilon)$.

Estimating a number X to multiplicative accuracy ϵ means producing an estimate $\hat{X}$ satisfying $|X - \hat{X}| \le \epsilon |X|$, while additive accuracy ϵ means that $|X - \hat{X}| \le \epsilon$.

Proof. Assume WLOG that M is supported on the symmetric subspace and has been rescaled such that $\|M\| = 1$. Then

$$h_{\mathrm{ProdSym}(n,d)}(M) \ge \mathbb{E}_{|a\rangle} \operatorname{Tr}[M |a\rangle\langle a|^{\otimes d}] = \frac{\operatorname{Tr}[M]}{\binom{n+d-1}{d}} \ge \|M\| n^{-d}. \tag{13}$$

Thus it will suffice to achieve additive error $\epsilon' := \epsilon/n^d$.

Theorem 1.3.3 of [9] states that polynomial equations of the form

$$\exists x \in \mathbb{R}^n, g_1(x) \ge 0, \ldots, g_m(x) \ge 0 \tag{14}$$

can be solved using $(md)^{O(n)}$ arithmetic operations. Moreover if the $g_1, \ldots, g_m$ have integer coefficients with absolute value $\le L$ then the intermediate numbers during this calculation are integers with absolute value $\le L(md)^{O(n)}$. We can put (12) into the form (14) (with $m = O(1)$) by adding a constraint of the form $f_0(x) \ge \theta$ and then performing binary search on θ, starting with the *a priori* bounds $0 \le h_{\mathrm{ProdSym}}(M) \le \|M\| \le 1$. If we specify the entries of M to precision $\epsilon'/\operatorname{poly}(n)$ then this will induce operator-norm error $\le \epsilon'$, which implies error $\le \epsilon'$ in h_{ProdSym}. Thus we can take $L \le \operatorname{poly}(n)/\epsilon' \le n^{d+O(1)}/\epsilon$. Since arithmetic operations on numbers $\le L$ require $\operatorname{poly}\log(L)$ time, we attain the stated run-time. □

The advantage of this argument is that it is simple and yields an effective algorithm. However, SDP hierarchies have several advantages over Theorem 2. The dual of an SDP can be useful, and here corresponds to entanglement witnesses, as we discuss in III D. An SDP hierarchy can interpolate in runtime between polynomial and exponential, whereas the algorithm in Theorem 2 can only be run in exponential time. Finally the hierarchy we develop can be interpreted in terms of extensions of quantum states and therefore has an interpretation in terms of a monogamy relation, although developing this is something we leave for future work.

We now turn towards developing an improved SDP hierarchy for approximating h_{ProdSym} in a way that will be at least as good at DPS at the low end and will match the performance of Theorem 2 at the high end. The objective function and constraints in (12) are both smooth, so the maximizing point must satisfy the *Karush-Kuhn-Tucker (KKT) conditions*:

$$\operatorname{rank}\begin{pmatrix} \frac{\partial f_0(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_1} \\ \vdots & \vdots \\ \frac{\partial f_0(x)}{\partial x_{2n}} & \frac{\partial f_1(x)}{\partial x_{2n}} \end{pmatrix} < 2.$$

This rank condition is equivalent to the condition that all 2×2 minors of the matrix should be equal to zero. Each minor is a polynomial of the form

$$g_{ij}(x) = \frac{\partial f_0(x)}{\partial x_i}\frac{\partial f_1(x)}{\partial x_j} - \frac{\partial f_0(x)}{\partial x_j}\frac{\partial f_1(x)}{\partial x_i}. \tag{15}$$

Note that $\deg(g_{ij}(x)) = \deg(\langle M, x^{\otimes 2d}\rangle) = 2d$. If we add these conditions to (12), we get the following equivalent optimization problem:

$$\begin{aligned} \max_{x\in\mathbb{R}^{2n}} \quad & f_0(x) \\ \text{subject to} \quad & f_1(x) = 0 \\ & g_{ij}(x) = 0 \quad \forall\, 1 \le i, j \le 2n \end{aligned} \tag{16}$$

B. Constructing the Relaxations

We will now construct SDP relaxations for this problem. Our first step will be to express (16) in terms of polynomial positivity:

$$\begin{aligned} \min \quad & \nu \\ \text{such that} \quad & \nu\langle \mathbb{1}^{\otimes d}, x^{\otimes 2d}\rangle - f_0(x) \ge 0 \\ \text{whenever} \quad & f_1(x) = 0 \\ & g_{ij}(x) = 0 \qquad \forall\, 1 \le i, j \le 2n \end{aligned}$$

Here $\mathbb{1}$ is the identity matrix. Note that we have multiplied ν by $\langle \mathbb{1}^{\otimes d}, x^{\otimes 2d}\rangle = ||x||^{2d}$; we are free to do this because this factor is equal to 1 whenever the norm constraint is satisfied. Now, as we described in Section II B, we replace the positivity constraint with the existence of an SOS certificate.

$$\begin{aligned} \min \quad & \nu \\ \text{such that} \quad & \langle \nu\mathbb{1}^{\otimes d} - M, x^{\otimes 2d}\rangle = \sigma(x) + \phi(x) f_1(x) + \sum_{ij} \chi_{ij}(x) g_{ij}(x) \end{aligned} \tag{17}$$

Here σ is a sum of squares and ϕ, χ_{ij} are arbitrary polynomials. We can now produce a hierarchy of relaxations by varying the degree of the certificates $(\sigma, \phi, \chi_{ij})$ that we search over. Specifically, at the rth level of the hierarchy, the total degree of all terms in the SOS certificate is upper-bounded by $2(r + d)$.

1. Explicit SDPs

The formulation (17) of the hierarchy in terms of SOS polynomials will be the one we use for most of our analysis. However, there is an alternative formulation in terms of an explicit SDP over moment matrices, which is more convenient for some purposes. Before we derive it, we will first make some simplifications that will let us eliminate the polynomial $\phi(x)$. Suppose we are working at level r of the hierarchy, so all the terms in the certificate have degree at most $2(d+r)$. Without loss of generality, we can assume that all terms in $\phi(x)$ and $\chi_{ij}(x)$ have even degree [31]. Moreover, we claim that without loss of generality, all the polynomials χ_{ij} are homogeneous of degree $2r$. Indeed, suppose χ_{ij} contains a term a of degree $2(r-k)$. Then $a = ||x||^{2k}a + (1-||x||^{2k})a$. Since $f_1(x) = ||x||^2 - 1$ divides $||x||^{2k} - 1$ for all $k \geq 1$, this means we can replace a with $||x||^{2k}a$ and absorb the error term inside $\phi(x)$.

Now we can eliminate $\phi(x)$ using the following argument, which is based on Proposition 2 in [32]. Denote the LHS of (17) by $q(x)$ and observe that it is homogeneous of degree $2d$. Then since $f_1(x/\|x\|) = 0$ we have

$$\begin{aligned} q\left(\frac{x}{||x||}\right) &= \sigma\left(\frac{x}{||x||}\right) + \sum_{ij} \chi_{ij}\left(\frac{x}{||x||}\right) g_{ij}\left(\frac{x}{||x||}\right) \\ q(x)||x||^{2r} &= \sigma\left(\frac{x}{||x||}\right)||x||^{2(r+d)} + \sum_{ij} \chi_{ij}(x) g_{ij}(x) \end{aligned}$$

Since σ has degree at most $2(r+d)$, $\sigma'(x) \equiv \sigma\left(\frac{x}{||x||}\right)||x||^{2(r+d)}$ is a polynomial in x. Moreover, by expanding the $\sigma(x) = \sum_k a_k(x)^2$, one can check that $\sigma'(x) = \sum_a s_a^2(x)$ where each term s_a is homogeneous of degree $r+d$. We say that $\sigma'(x)$ is a sum of homogeneous squares. Thus, from a certificate of the form given in (17), we have constructed a new certificate of the form

$$q(x)||x||^{2r} = \sigma'(x) + \sum_{ij} \chi_{ij}(x) g_{ij}(x), \tag{18}$$

with σ' a sum of homogeneous squares. In this form we have eliminated the polynomial ϕ. Conversely, from any certificate of the form (18), we can produce a certificate in the form (17) as follows:

$$\begin{aligned} q(x)||x||^{2r} &= \sigma'(x) + \sum_{ij} \chi_{ij}(x) g_{ij}(x) \\ q(x) &= \sigma'(x) + q(x)(1-||x||^{2r}) + \sum_{ij} \chi_{ij}(x) g_{ij}(x) \end{aligned}$$

Since $1 - ||x||^2$ divides $1 - ||x||^{2r}$, this is indeed a certificate of the form given in (17). Thus, we have shown that the hierarchy (17) is equivalent to the following hierarchy.

$$\begin{aligned} &\min && \nu \\ &\text{such that} && \langle \nu \mathbb{1}^{\otimes(d+r)} - M \otimes \mathbb{1}^{\otimes r}, x^{\otimes 2(d+r)} \rangle - \sum_{ij} \chi_{ij}(x) g_{ij}(x) = \sigma(x). \end{aligned} \tag{19}$$

Here, $\chi_{ij}(x)$ is an arbitrary homogeneous polynomial of degree $2r$ and $\sigma(x)$ is a sum-of-homogeneous-squares polynomial of degree $2(d+r)$.

This SOS program can be written explicitly as an SDP, using the procedure described in Section II B. This would produce an SDP over $m \times m$ matrices where $m = \binom{2n+2(d+r)-1}{2(d+r)}$ is the number of monomials of degree $2(d+r)$. This SDP can be solved to accuracy ϵ in time $O(\mathrm{poly}(m)\,\mathrm{poly}\log(1/\epsilon))$.

However, in order to facilitate comparison with DPS, we will instead write an SDP over $(2n)^{d+r} \times (2n)^{d+r}$ matrices; this corresponds to treating different orderings of the variables in a monomial as distinct monomials. The redundant degrees of freedom will be removed by imposing symmetry constraints. Specifically, let the map $\mathcal{P}$ from tensors of rank $2k$ to matrices of dimension $(2n)^k$ be defined by

$$(\mathcal{P}A)_{(i_1 i_2 \ldots i_k),(i_{k+1} i_{k+2} \ldots i_{2k})} \equiv \frac{1}{(2k)!} \sum_{\pi \in \mathcal{S}_{2k}} A_{i_{\pi(1)} i_{\pi(2)} \cdots i_{\pi(2k)}},$$

where $\mathcal{S}_{2k}$ is the group of all permutations of $\{1, \ldots, 2k\}$. Then our SDP is

$$\begin{aligned} &\min && \nu \\ &\text{such that} && \mathcal{P}\left(\nu \mathbb{1}^{d+r} - M \otimes \mathbb{1}^{\otimes r} - \sum_{ij\alpha} \chi_{ij\alpha} A_\alpha \otimes \Gamma_{ij}\right) \succeq 0. \end{aligned} \tag{20}$$

Here, the indices ij label the KKT constraints, and the multi-index α labels all monomials of degree $2r$. The variable $\chi_{ij\alpha}$ is the coefficient of the monomial α in the polynomial χ_{ij}. The matrix A_α represents the monomial α, i.e. $\langle A_\alpha, x^{\otimes 2r} \rangle = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$. Finally, the matrix Γ_{ij} represents the KKT polynomial $g_{ij}(x)$, i.e. $\langle \Gamma_{ij}, x^{\otimes 2d} \rangle = g_{ij}(x)$.

Now we can at last write down the moment matrix version of the hierarchy by applying SDP duality to (20).

$$\begin{aligned} &\max_\rho && \langle \mathcal{P}(M \otimes \mathbb{1}^{\otimes r}), \rho \rangle \\ &\text{such that} && \rho \succeq 0 \\ & && \langle \mathcal{P}(A_\alpha \otimes \Gamma_{ij}), \rho \rangle = 0 \; \forall i, j, \alpha. \end{aligned} \tag{21}$$

In this program, the variable ρ is a matrix in $\mathbb{R}^{(2n)^{d+r} \times (2n)^{d+r}}$. Now we see the advantage of adding the redundant degrees of freedom in the SDP—just as in DPS, ρ can be interpreted as the density matrix over an extended quantum system. The main difference from DPS is the set of added constraints $\langle A_\alpha \otimes \Gamma_{ij}, \rho \rangle = 0$, which are the moment relaxations of the KKT conditions.

The SDP (21) is over $(2n)^{d+r} \times (2n)^{d+r}$ matrices, so if $r = O(\exp(n))$, we would naïvely expect it to have time complexity $O(\exp(\exp(n)\log(n)))$. This apparently large complexity is caused by the redundant degrees of freedom we added above. In practice, we can use the symmetry constraints enforced by $\mathcal{P}$ to eliminate the redundancy and bring the complexity back down to $\binom{2n+2(d+r)-1}{2(d+r)}^{O(1)}$, which is $O(\exp(n))$ when $r = O(\exp(n))$. This is discussed in more detail in Section IV of the original DPS paper [10].

C. Degree bounds for SOS certificates

In this section, we will show that for generic inputs, the SOS form of the hierarchy (17) converges exactly within $d^{O(n^2)}$ levels. In other words, we will show that generically, there exists a sum-of-squares certificate of degree $O(d^{\mathrm{poly}(n)})$. This is an algebraic statement, so it is useful to recast it in the language of polynomial ideals. We define the *KKT ideal* I_K to be the ideal generated by the polynomials g_{ij} and f_1. Likewise, define the *truncated* KKT ideal I_K^m to be

$$I_K^m = \left\{ v(x) f_1(x) + \sum_{ij} h_{ij}(x) g_{ij}(x) : \deg(v(x) f_1(x)) \leq m, \max_{i,j} \deg(h_{ij}(x) g_{ij}(x)) \leq m \right\}.$$

Then we claim

Theorem 3. *Let f_0, f_1, g_{ij} be as defined in* (12). *Then there exists $m = d^{O(n^2)}$ such that for generic M, if $\nu - f_0(x) > 0$ for all $x \in \mathbb{R}^{2n}$ such that $f_1(x) = 0$, then*

$$\nu - f_0(x) = \sigma(x) + g(x),$$

where $\sigma(x)$ is sum of squares, $\deg(\sigma(x)) \leq m$ *and $g(x) \in I_K^m$.*

The proof is in Section IV.

Corollary 4. *We can estimate $h_{\mathrm{ProdSym}(n,2)}$ to multiplicative error ϵ in time* $\exp(\mathrm{poly}(n)) \operatorname{poly}\log(1/\epsilon)$.

This follows from Theorem 3 and the fact that the value of semidefinite programs can be computed in time polynomial in the dimension, number of constraints and bits of precision (i.e. $\log 1/\epsilon$).

D. Entanglement detection

So far, we have restricted ourselves to optimization problems over the convex sets Sep and ProdSym. In practice, another very important problem is entanglement detection, i.e. testing whether a given density matrix is a member of Sep or ProdSym. In general, membership testing and optimization for convex sets are intimately related. There exist polynomial time reductions in both directions using the ellipsoid method, as described in Chapter 4 of [1]. Thus, our results immediately imply an algorithm of complexity $O(d^{\mathrm{poly}(n)} \operatorname{poly}\log(1/\epsilon))$ for membership testing in Sep.

There is, however, a more direct way to go from optimization to membership, using the notion of an *entanglement witness*. The idea is that to show that a given state ρ is not in Sep (resp. ProdSym), it suffices to find a Hermitian operator Z such that $\mathrm{Tr}[Z\rho] < 0$, but for all $\rho' \in$ Sep (resp. ProdSym), $\mathrm{Tr}[Z\rho'] \geq 0$. Such an operator Z is called an entanglement witness for ρ. The search for an entanglement witness can be phrased as an optimization problem:

$$\begin{aligned} \min_{Z} \quad & \mathrm{Tr}[Z\rho] \\ \text{such that} \quad & \mathrm{Tr}[Z\rho'] \geq 0 \ \forall \rho' \in \mathrm{ProdSym} \end{aligned} \tag{22}$$

If the optimum value is less than 0, then we know that ρ is entangled. Geometrically, an entanglement witness is a separating hyperplane between ρ and the convex set of separable states. Thus, because of the hyperplane separation theorem for convex sets, every entangled ρ must have some witness that detects it. However, finding the witness may be very difficult.

The witness optimization problem (22) is closely related to the problem h_{ProdSym}. In particular, suppose that for a measurement operator M, we know that $h_{\mathrm{ProdSym}}(M) < \nu$. Then $Z = \nu\mathbb{1} - M$ is a feasible point for (22). As a consequence of this, any feasible solution to the SOS form of either DPS or our hierarchy will yield an entanglement witness operator.

In the case of DPS, it turns out that this connection also yields an efficient way to search for a witness detecting a given entangled state. To see this, we consider the set of all possible witnesses generated by DPS at level r, for any measurement operator M. Through straightforward computations (see Section VI of [10]), one finds that this set is

$$\mathrm{EW}_{\mathrm{DPS}}(r) = \{\Lambda^\dagger(Z_0 + Z_1 + \cdots + Z_r) : Z_0 \succeq 0, Z_1^{T_1} \succeq 0 \ldots Z_r^{T_m} \succeq 0\}. \tag{23}$$

Here Λ is a certain fixed linear map and the superscripts $T_1, \ldots T_m$ indicate various partial transposes (i.e. permutations interchanging a subset of the row and column indices). The important thing to note is that this is a convex set; in fact, it has the form of the feasible set of a semidefinite program. Thus, given a state, it is possible to efficiently search for an entanglement witness detecting it using a semidefinite program.

Once we add the KKT conditions, the situation is not as convenient. The set of all entanglement witnesses at level r, denoted $\mathrm{EW}_{KKT}(r)$, is the set of Z for which $\exists \sigma(x), \chi_{ij}(x)$ such that

$$\begin{aligned}&\langle Z, x^{\otimes 2d}\rangle = \sigma(x) + \sum_{ij} \chi_{ij}(x) g_{ij}(x)\\&\deg(\sigma(x)) \leq r\\&\deg(\chi_{ij}(x) g_{ij}(x)) \leq r\end{aligned}$$

The important difference from DPS is that the polynomials $g_{ij}(x)$ come from the KKT conditions and thus *depend on* Z. This in particular means that $\mathrm{EW}_{KKT}(r)$ no longer has the form of an SDP feasible set, nor indeed is it necessarily convex. However, we also note that by Theorem 3, an open dense subset of all entanglement witnesses is contained in $\mathrm{EW}_{KKT}(r)$ for $r = n^{O(d^2)}$.

E. Numerical results

While our theoretical results show that adding the KKT conditions results an improvement at very high levels of the DPS hierarchy, we have also found numerical evidence of improvements even for low-dimensional systems at very low levels of the hierarchy. We compared the performance of the hierarchy with and without the KKT conditions at the second level (i.e. searching over SOS certificates of degree 6) on a family of measurements with local dimension 3. The measurements were obtained by the applying the construction in section VIII.A of [10] to the entanglement witness given in equation (69) of the same reference. Explicitly, they are given by

$$M_\gamma := \mathbb{1} \otimes \mathbb{1} - (A_\gamma^{-1} \otimes \mathbb{1}) Z (A_\gamma^{-1} \otimes \mathbb{1}), \qquad \gamma \in [0, 1], \tag{24}$$

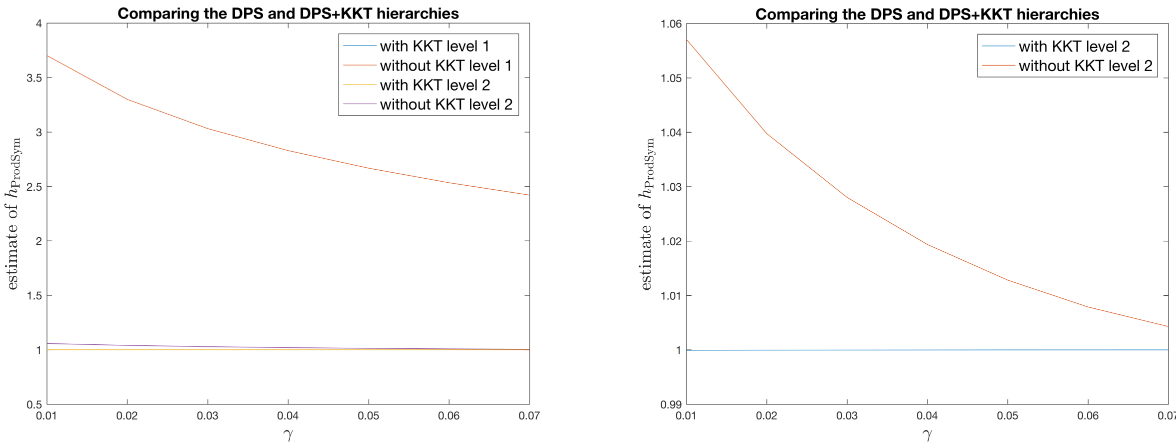
where

$$\begin{aligned}A_\gamma &= \mathrm{diag}(1, \gamma, \ldots, \gamma)\\Z &= 2(|00\rangle\langle 00| + |11\rangle\langle 11| + |22\rangle\langle 22|)\\&\qquad + |02\rangle\langle 02| + |10\rangle\langle 10| + |21\rangle\langle 21| - 3|\psi_+\rangle\langle\psi_+|\\|\psi_+\rangle &= \frac{1}{\sqrt{3}} \sum_{i=1}^{3} |ii\rangle.\end{aligned}$$

By construction, $h_{\mathrm{Sep}}(M_\gamma) \leq 1$ for all $\gamma \in [0, 1]$. However, it was shown in [10] that for sufficiently small γ, the optimum value of DPS applied to M_γ will be strictly greater than 1. Numerically, we find that this behavior occurs for $\gamma < 0.1$. In Figure 1, we plot the optimum value returned by the second level of the hierarchy for a range of values of γ between 0.01 and 0.07. We find that adding the KKT conditions substantially improves the convergence. The calculations were performed using YALMIP optimization package [33, 34], the SDP solver Mosek [35], and the SDP preprocessing package frlib [36].

IV. PROOFS

In this section, we will make use of a number of tools from algebraic geometry, which are described in Appendix A. At a high level, the proof will proceed as follows: first we show that for



(a) (b)

FIG. 1. Performance of hierarchy with and without KKT conditions, for the family of measurements in Eq. (24). The true value of h_{Sep} is ≤ 1 for all γ. Figure (a) shows the performance with and without KKT conditions at both level 1 (searching over degree-4 SoS certificates) and level 2 (degree-6 SoS certificates) for both variants of the hierarchy, while figure (b) shows only level 2. At level 1, the KKT constraints have no effect and both hierarchies yield the same relaxation. At level 2, the hierarchy with KKT immediately converges to the true value 1, within numerical error, while the hierarchy without KKT obtains upper bounds that are strictly greater than 1.

generic M, the KKT ideal I_K is zero-dimensional. This implies that a Gröbner basis of exponential degree can be found for I_K. We then complete the proof using a strategy due to Laurent (Theorem 6.15 of [37]): we start with a SOS certificate of high degree, and then use division by the Gröbner basis to reduce the degree. This will result in a SOS certificate whose degree is the same order as the degree of the Gröbner basis, thus proving the theorem.

A. Generic inputs

We will now show that, for generic M, the KKT ideal is zero-dimensional, using a dimension-counting argument based on the theorems in Section A 2. A similar result was proved in Proposition 2.1 (iii) of [38]. However, that result required both the objective function and the constraints to be generic. Since the norm constraint is fixed independent of the input M, this means we cannot apply the result of [38] directly. Nevertheless, we find that we can use a very similar argument.

Lemma 5. *For generic M, the KKT ideal I_K is zero dimensional.*

The intuition behind the proof is the same reason that the KKT conditions characterize optimal solutions. Roughly speaking the KKT conditions encode the fact that at an optimal solution one should not be able to increase the objective function without changing one or more of the constraint equations. This corresponds to a particular Jacobian matrix having less than full rank. Here we will see that this rank condition on a Jacobian directly implies that the set of solutions is zero dimensional.

Proof. For the proof we will find it is useful to move to complex projective space $\mathbb{P}^n$, parametrized by homogeneous coordinates $\tilde{x} = (x_0, x_1, \ldots, x_n)$. For a polynomial $p(x)$, we denote its homogenization by $\tilde{p}(\tilde{x})$. We also define the following projective varieties

$$\mathcal{U} = \{\tilde{x} : \tilde{f}_1(\tilde{x}) = 0\}$$
$$\mathcal{W} = \{\tilde{x} : \forall\, i, j, \tilde{g}_{ij}(\tilde{x}) = 0\}$$

The variety associated with the KKT ideal $V(I_K)$ is just the affine part of $\mathcal{U} \cap \mathcal{W}$. So it suffices to show that $\mathcal{U} \cap \mathcal{W}$ is finite. We will do this using a dimension counting argument. Specifically, we will construct a variety of high dimension that does not intersect $\mathcal{W}$. By Bézout's Theorem, this will give us an upper bound on the dimension of $\mathcal{W}$.

To find such a variety, consider the family $\mathcal{H}$ of all hypersurfaces $\mathcal{X}$ in $\mathbb{P}^n$ of the form $\{\tilde{f}_0(\tilde{x}) - \mu x_0^{2d} = 0\}$, parametrized by $\mu \in \mathbb{C}$ and the matrix $M \in \mathbb{C}^{n^2 \times n^2}$. Multiplying μ and M by a nonzero scalar leaves the associated hypersurface unchanged, so we can think of (M, μ) as a point in a projective space $\mathbb{P}^k$. We will be interested in the intersection $\mathcal{A} = \mathcal{X} \cap \mathcal{U}$ of a hypersurface $\mathcal{X}$ in this family with the feasible set $\mathcal{U}$. The Jacobian matrix $\tilde{J}_{\mathcal{A}}$ of such an intersection is given by

$$\tilde{J}_{\mathcal{A}} = \begin{pmatrix} \frac{\partial}{\partial x_0}(\tilde{f}_0(\tilde{x}) - \mu x_0^{2d}) & \frac{\partial \tilde{f}_1(\tilde{x})}{\partial x_0} \\ \frac{\partial \tilde{f}_0(\tilde{x})}{\partial x_1} & \frac{\partial \tilde{f}_1(\tilde{x})}{\partial x_1} \\ \vdots & \vdots \\ \frac{\partial \tilde{f}_0(\tilde{x})}{\partial x_n} & \frac{\partial \tilde{f}_1(\tilde{x})}{\partial x_n} \end{pmatrix}.$$

Let $J_{\mathcal{A}}$ denote the submatrix of $\tilde{J}_{\mathcal{A}}$ obtained by removing the first row. We claim that for a generic choice of M and μ, the matrix $J_{\mathcal{A}}$ is of rank 2 everywhere on $\mathcal{A}$. Since $\mathcal{W}$ is the set of points with rank $\tilde{J}_{\mathcal{A}}$, this implies that $\mathcal{A} \cap \mathcal{W} = \emptyset$.

Now, to prove the claim, we use Bertini's Theorem (Theorem 20). The variety $\mathcal{U}$ is smooth and has dimension $n-1$, and as long as $M \neq 0$, there are no points in common to all the hypersurfaces in $\mathcal{H}$. Thus, by Theorem 20, for a generic choice of $(M, \mu) \in \mathbb{P}^k$, the variety $\mathcal{A} = \mathcal{U} \cap \{\tilde{f}_0(\tilde{x}) - \mu x_0^{2d} = 0\}$ is smooth (has no singular points) and has dimension $n-2$. This means that $\tilde{J}_{\mathcal{A}}$ must have rank 2 everywhere on $\mathcal{A}$. By homogeneity, we know that if $\tilde{f}_0(\tilde{x}) - \mu x_0^{2d} = 0$, then $\tilde{f}_0(\lambda\tilde{x}) - \mu(\lambda x_0)^{2d} = 0$ for all $\lambda \neq 0$. If we take the derivative of this expression with respect to λ and set $\lambda = 1$, we get that $x_0 \frac{\partial}{\partial x_0}(\tilde{f}_0(\tilde{x}) - \mu x_0^{2d}) = -\sum_i x_i \frac{\partial}{\partial x_i} f_0(\tilde{x})$. Likewise we also find that $x_0 \frac{\partial}{\partial x_0}\tilde{f}_1(\tilde{x}) = -\sum_i x_i \frac{\partial}{\partial x_i} f_1(\tilde{x})$. So whenever $x_0 \neq 0$, the first row of $\tilde{J}_{\mathcal{A}}$ is in the span of the other rows. Hence, for $x_0 \neq 0$, $\text{rank}(\tilde{J}_{\mathcal{A}}) = 2$ implies that $\text{rank}(J_{\mathcal{A}}) = 2$ as well. This means that that the affine part $(x_0 \neq 0)$ of $\mathcal{A}$ does not intersect the affine part of $\mathcal{W}$. It only remains to check the part at infinity $(x_0 = 0)$. We know that since $\mathcal{A}$ is smooth, $\tilde{J}_{\mathcal{A}}$ has rank 2 here also. By direct evaluation, we see that the first row of $\tilde{J}_{\mathcal{A}}$ is zero when $x_0 = 0$, so $J_{\mathcal{A}}$ has rank 2 here as well. Therefore, $\mathcal{A}$ does not intersect $\mathcal{W}$ anywhere.

Now we complete the proof using a dimension-counting argument. Bézout's Theorem (Theorem 19) states that any two projective varieties in $\mathbb{P}^n$, the sum of whose dimensions is at least n, must have a non-empty intersection. Thus, since $\mathcal{W} \cap \mathcal{A} = (\mathcal{W} \cap \mathcal{U}) \cap \{\tilde{f}_0(\tilde{x}) = \mu x_0^{2d}\} = \emptyset$, we deduce that

$$\dim(\mathcal{W} \cap \mathcal{U}) + \dim(\{\tilde{f}_0(\tilde{x}) = \mu x_0^d\}) = \dim(\mathcal{W} \cap \mathcal{U}) + n - 1 < n.$$

This implies that $\mathcal{W} \cap \mathcal{U}$ has dimension 0, i.e. it is a finite set of points in $\mathbb{P}^n$. So $\mathcal{W} \cap \mathcal{U} \cap \{x_0 = 1\}$ is a finite set of points in $\mathbb{C}^n$. But this is precisely the variety associated with the KKT ideal, or rather its complex analogue. However, the fact that the KKT equations have a finite set of solutions in $\mathbb{C}^n$ implies that their set of solutions in $\mathbb{R}^n$ is also finite. Thus, the KKT ideal is zero-dimensional as claimed. $\square$

For the next result, we will want to consider the ideal generated by a homogenized version of the KKT conditions. For convenience sake, we would like all the generators to be homogeneous of the *same* degree. The polynomials $g_{ij}(x)$ are already homogeneous and have degree $2d$. The polynomial $f_1(x)$ is not homogeneous and has degree 2. So we will homogenize it and multiply it by $x_0^{2(d-1)}$ to make it also degree $2d$. This yields the following ideal

$$\tilde{I}_K = \left\langle g_{ij}(\tilde{x}), x_0^{2(d-1)} \tilde{f}_1(x) \right\rangle .$$

Lemma 6. *The ideal $\tilde{I}_K$ has a Gröbner basis in the degree ordering whose elements have degree $O(d^{\mathrm{poly}(n)})$. Moreover, each Gröbner basis element $\gamma_k(\tilde{x})$ can be expressed in terms of the original generators as $\gamma_k(\tilde{x}) = \sum_{ij} u_{ijk}(\tilde{x}) g_{ij}(\tilde{x}) + v_k(\tilde{x})(x_0^{2(d-1)} \tilde{f}_1(\tilde{x}))$ where $\deg(u_{ijk}(\tilde{x})), \deg(v_k(\tilde{x})) = O(d^{\mathrm{poly}(n)})$.*

Proof. Let D be the degree of the Gröbner basis. Since the KKT ideal is zero dimensional, the homogenized KKT ideal is one-dimensional (that is, $V(\tilde{I}_K)$ is one-dimensional when viewed as an affine variety in $\mathbb{C}^{n+1}$). So the result of Proposition 16 evaluated at $r = 1$ gives a bound $D = O(d^{n^2})$. Moreover, since the ideal is homogeneous, by Proposition 15 the Gröbner basis elements can be chosen to be homogeneous as well. We will denote this Gröbner basis of homogeneous polynomials as $\{\tilde{\gamma}_k(\tilde{x})\}$.

Now, we know that any given Gröbner basis element can be expressed in terms of the original generators from (15):

$$\tilde{\gamma}_k(\tilde{x}) = \sum_{ij} u_{ijk}(\tilde{x}) \tilde{g}_{ij}(\tilde{x}) + v_k(\tilde{x})(x_0^{2d} \tilde{f}_1(\tilde{x})),$$

where the polynomials $u_{ij}(\tilde{x})$ and $v_k(\tilde{x})$ could have arbitrarily high degree. Let the degree of $\tilde{\gamma}_k(\tilde{x})$ be $D_k \leq D$. Since it is homogeneous, all the terms on the RHS must be of degree D_k. Moreover, we know that $\tilde{g}_{ij}(\tilde{x})$ and $x_0^{2(d-1)} \tilde{f}_1(\tilde{x})$ are homogeneous of degree $2d$. Therefore, any terms in $u_{ijk}(\tilde{x})$ or $v_k(\tilde{x})$ with degree higher than D_k will result only in terms of degree higher than $D_k + 2d$ on the RHS. We know that these terms must cancel out to zero. Therefore, we can just drop all terms with degree higher than D_k from $u_{ijk}(\tilde{x})$ and $v_k(\tilde{x})$ and equality will still hold in the equation above. Thus, we have shown that every Gröbner basis element can be expressed in terms of the original generators with coefficients of degree at most D as desired. □

Now we prove Theorem 3. The argument is the same as case (i) of Theorem 6.15 in [37].

Proof. Let $\{\tilde{\gamma}_i(\tilde{x})\}$ be a degree-ordered Gröbner basis for $\tilde{I}_{KKT}$, as in the previous proposition. By dehomogenizing, we get a Gröbner basis $\{\gamma_i(x)\}$ for I_k. Since $1 - \sum_i x_i^2 \equiv 0 \pmod{I_{KKT}}$, the KKT ideal satisfies the Archimedean condition and Theorem 1 holds. Thus, there exists some $\sigma(x)$ SOS and $g(x) \in I_K$ such that $\nu - f_0(x) = \sigma(x) + g(x)$. Let us write $\sigma(x)$ explicitly as

$$\sigma(x) = \sum_a s_a(x)^2.$$

Since I_{KKT} is zero-dimensional, by Proposition 18, each term $s_a(x)$ can be written as $s_a(x) = \sum a_{ak}(x)\gamma_k(x) + u_a(x) \equiv g_a(x) + u_a(x)$, where $\deg(u_a(x)) \leq nD$ and $g_a(x) \in I_{KKT}$. If we substitute this decomposition into the expression for $\sigma(x)$, we get

$$\sigma(x) = \sum_a u_a(x)^2 + g'(x),$$

where $g'(x) \in I_{KKT}$. We can combine the terms in I_{KKT} to get the following expression for the SOS certificate:

$$\nu - f_0(x) = \sigma'(x) + g''(x),$$

where $g''(x) \in I_{KKT}$ and $\deg(\sigma'(x)) \leq 2nD = d^{O(n^2)}$. Now, the LHS of this expression has degree $2d < \deg(\sigma'(x))$, so $g''(x)$ must also have degree $d^{O(n^2)}$. By Proposition 14, it can be expressed as

$$g''(x) = \sum h_k(x)\gamma_k(x),$$

where $\deg(h_k(x)\gamma_k(x)) = d^{O(n^2)}$. Using Lemma 6, we can express this in terms of the original generators as

$$g''(x) = \sum_{ijk} h_k(x)u_{ijk}(x)g_{ij}(x).$$

We know that $\deg(u_{ijk}(x)) = d^{O(n^2)}$. Therefore, $g'(x) \in I_K^m$ for $m = d^{O(n^2)}$. This proves the theorem. □

B. An algorithm for all inputs

We have shown that for generic M, there exists a SOS certificate of low degree for the optimization problem (17). However, for nongeneric M, it is possible that no certificate of low degree exists, so the SOS formulation of the hierarchy may not converge within $d^{O(n^2)}$ levels. In this section, we will show that this problem goes away if we switch to the moment matrix formulation (21) of the hierarchy. We will show that this formulation converges in $d^{O(n^2)}$ levels for *any* input M. First, we show that the SDPs of the moment hierarchy are well behaved in the sense that they satisfy *Slater's condition* for any input M. This is the condition that either the primal or dual feasible set of the SDP should have a nonempty relative interior. To show this, we use the following result from [39, 40].

Proposition 7. *For a given SDP, let $\mathcal{P}, \mathcal{D}$, and $\mathcal{P}^*$ be the primal feasible set, dual feasible set, and set of primal optimal points, respectively. Then $\mathcal{P}$ and* $\mathrm{interior}(\mathcal{D})$ *are nonempty iff $\mathcal{P}^*$ is nonempty and bounded.*

In our case, let (21) be the primal and (17) be the dual. The primal feasible set is nonempty, since the true optimizing point for the unrelaxed problem h_{ProdSym} is always feasible. Moreover, primal feasible set is compact. Thus, the primal *optimal* set $\mathcal{P}^*$ is nonempty and bounded, and thus by Proposition 7, Slater's condition holds.

Slater's condition implies strong duality, so for generic M, (21) and (17) give the same optimum value. It also implies that the SDP value is a differentiable function of the input parameters. We use this to extend our results to non-generic inputs M.

Theorem 8. *For all input M, the hierarchy* (21) *converges to the optimum value of* (12) *at level $r = d^{O(n^2)}$.*

Proof. For a given M, let $f^*_{\mathrm{mom},r}(M)$ be the optimum value of the r-th level of the hierarchy (21). It is easy to see that $h_{\mathrm{ProdSym}}(M)$ is a continuous function of M [41]. We claim that $f^*_{\mathrm{mom},r}(M)$ is also continuous. Indeed, Theorem 10 of [42] states that if an SDP satisfies Slater's condition and has a nonempty bounded feasible set for all input parameters, then the optimum value is a differentiable function of the inputs. By the preceding discussion, these conditions hold for the moment hierarchy for all M, so $f^*_{\mathrm{mom},r}(M)$ is indeed continuous.

Now, by the remarks above, $h_{\mathrm{ProdSym}}(M) = f^*_{\mathrm{mom},r}(M)$ for all generic M. Recall from Section A that the set of generic M is an open, dense set, according to the standard topology. Thus, since both functions h_{ProdSym} and $f^*_{\mathrm{mom},r}$ are continuous and agree on an open dense subset, $h_{\mathrm{ProdSym}}(M) = f^*_{\mathrm{mom},r}(M)$ for *all* M. □

Corollary 9. *For all input M, $h_{\mathrm{ProdSym}}(M)$ can be approximated up to additive error ϵ in time $O(d^{\mathrm{poly}(n)} \operatorname{poly}\log(1/\epsilon))$.*

V. DISCUSSION AND OPEN QUESTIONS

Adding the KKT conditions provides a new way of sharpening the familiar DPS hierarchy for testing separability. We have given some evidence that its asymptotic performance is superior to that of the original DPS hierarchy. Indeed, [43] shows that even for constant n, a variant of the r^{th} DPS hierarchy has error lower-bounded by $\Omega(1/r)$. But our hierarchy converges in a constant number of steps for any fixed local dimension.

Does this mean that our hierarchy has other asymptotic improvements over the DPS hierarchy at lower values of r? We have seen already cases in which DPS dramatically outperforms the weaker r-extendability hierarchy. For example, if M is the projector onto an n-dimensional maximally entangled state, then its maximum overlap with PPT states is $1/n$ while its maximum overlap with r-extendable states is $\geq 1/r$. A more sophisticated example of this scaling based on an M arising from a Bell test related to the unique games problem is in [30]. One of the major open questions in this area is whether low levels of SDP hierarchies such as DPS can resolve hard optimizations problems of intermediate complexity such as the unique games problem [15].

ACKNOWLEDGMENTS

AWH was funded by NSF grant CCF-1111382 and CCF-1452616. AN was funded by a Clay Fellowship. AN also thanks Cyril Stark for helpful conversations. All three authors (AWH, AN and XW) were funded by ARO contract W911NF-12-1-0486.

---

[1] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, second corrected edition ed., Algorithms and Combinatorics, Vol. 2 (Springer, 1993).
[2] Y.-K. Liu, *The Complexity of the Consistency and N-representability Problems for Quantum States*, Ph.D. thesis, Univ. of California, San Diego (2007), 0712.3041.
[3] A. W. Harrow and A. Montanaro, J. ACM **60**, 3:1 (2013), 1001.0017.
[4] S. Beigi and P. W. Shor, J. Math. Phys. **51**, 042202 (2010), 0902.1806.
[5] F. L. Gall, S. Nakagawa, and H. Nishimura, Q. Inf. Comp. **12**, 589 (2012), 1108.4306.
[6] T. S. Cubitt, D. Perez-Garcia, and M. Wolf, "Undecidability of the spectral gap problem," (2014), in preparation.
[7] T. Ito, H. Kobayashi, and J. Watrous, in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12 (2012) pp. 266–275, 1012.4427.
[8] On the other hand, there is evidence that there is a separation between QMA(2) with constant error and with error scaling as an inverse polynomial in the dimension. [44].
[9] S. Basu, R. Pollack, and M.-F. Roy, J. ACM **43**, 1002 (1996).
[10] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, "A complete family of separability criteria," (2003), arXiv:quant-ph/0308032.

[11] B. Barak and D. Steurer, "Sum-of-squares proofs and the quest toward optimal algorithms," (2014), 1404.5236.

[12] J. Nie, "An exact Jacobian SDP relaxation for polynomial optimization," (2010), arXiv:1006.2418.

[13] S. Gharibian, QIC **10**, 343 (2010), 0810.4507.

[14] S. Aaronson, R. Impagliazzo, and D. Moshkovitz, in *Computational Complexity (CCC), 2014 IEEE 29th Conference on* (2014) pp. 44–55, 1401.6848.

[15] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou, in *Proceedings of the 44th symposium on Theory of Computing*, STOC '12 (2012) pp. 307–326, 1205.4484.

[16] R. Impagliazzo, R. Paturi, and F. Zane, in *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on* (IEEE, 1998) pp. 653–662.

[17] L. Gurvits, "Classical deterministic complexity of Edmonds' problem and quantum entanglement," (2003), arXiv:quant-ph/0303055.

[18] H. Blier and A. Tapp, in *First International Conference on Quantum, Nano, and Micro Technologies* (IEEE Computer Society, Los Alamitos, CA, USA, 2009) pp. 34–37, 0709.0738.

[19] A. Chiesa and M. A. Forbes, Chicago Journal of Theoretical Computer Science **2013** (2013), 1108.2098.

[20] M. Navascués, M. Owari, and M. B. Plenio, Phys. Rev. A **80**, 052306 (2009), 0906.2731.

[21] S. Aaronson, S. Beigi, A. Drucker, B. Fefferman, and P. Shor, Annual IEEE Conference on Computational Complexity **0**, 223 (2008), 0804.0802.

[22] J. Chen and A. Drucker, "Short multi-prover quantum proofs for SAT without entangled measurements," (2010), 1011.0716.

[23] F. G. S. L. Brandão, M. Christandl, and J. Yard, Comm. Math. Phys. **306**, 805 (2011), 1010.1750.

[24] K. Li and A. Winter, Comm. Math. Phys. **326**, 63 (2014), 1210.3181.

[25] F. G. S. L. Brandão and A. W. Harrow, in *Proceedings of the 45th annual ACM Symposium on theory of computing*, STOC '13 (2013) pp. 861–870, 1210.6367.

[26] Y. Shi and X. Wu, in *ICALP12* (Springer, 2012) pp. 798–809, 1112.0808.

[27] F. G. Brandão and A. W. Harrow, "Estimating injective tensor norms using nets," (2014), in preparation.

[28] K. Li and G. Smith, "Quantum de Finetti theorem measured with fully one-way LOCC norm," (2014), 1408.6829.

[29] M. Putinar, Indiana University Mathematics Journal **42**, 969 (1993).

[30] H. Buhrman, O. Regev, G. Scarpa, and R. de Wolf, in *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity*, CCC '11 (2011) pp. 157–166, 1012.5043.

[31] The LHS of (17) only contains even degree terms, as do $f_1(x)$ and $g_{ij}(x)$. Thus, any odd degree terms in $\phi(x)$ and $\chi_{ij}(x)$ must cancel each other, so they can all be removed.

[32] E. Klerk, M. Laurent, and P. Parrilo, in *Positive Polynomials in Control*, Lecture Notes in Control and Information Science, Vol. 312, edited by D. Henrion and A. Garulli (Springer Berlin Heidelberg, 2005) pp. 121–132.

[33] J. Löfberg, in *In Proceedings of the CACSD Conference* (Taipei, Taiwan, 2004).

[34] J. Löfberg, IEEE Transactions on Automatic Control **54**, 1007 (2009).

[35] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).* (2015).

[36] F. Permenter and P. Parrilo, "Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone," (2014), arXiv:1408.4685.

[37] M. Laurent, in *Emerging Applications of Algebraic Geometry*, The IMA Volumes in Mathematics and its Applications, Vol. 149, edited by M. Putinar and S. Sullivant (Springer New York, 2009) pp. 157–270.

[38] J. Nie and K. Ranestad, SIAM Journal on Optimization **20**, 485 (2009), http://dx.doi.org/10.1137/080716670.

[39] M. Trnovská, Journal of Electrical Engineering **56**, 1 (2005).

[40] C. Cédric Josz (INRIA), Didier Henrion (LAAS, "Strong duality in Lasserre's hierarchy for polynomial optimization," (2014), 1405.7334.

[41] One way to show this is to note that $h_{\text{ProdSym}}(M)$ is a norm of M.

[42] A. Shapiro, Mathematical Programming , 301 (1997).

[43] S. Strelchuk and J. Oppenheim, Phys. Rev. A **86**, 022328 (2012), 1207.1084.

[44] A. Pereszlényi, "Multi-prover quantum Merlin-Arthur proof systems with small gap," (2012), 1205.2761.

[45] D. Cox, J. little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, second edition ed., Undergarduate texts in mathematics (Springer, 1996).
[46] J. Harris, *Algebraic geometry: a first course*, Graduate texts in mathematics (Springer, 1992).
[47] B. Buchberger, Aequationes Mathematicae **4**, 374 (1970).
[48] E. W. Mayr and S. Ritscher, in *Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, ISSAC '10 (ACM, New York, NY, USA, 2010) pp. 21–27.

Appendix A: Algebraic Geometry

In this paper we will use some basic tools from algebraic geometry, which we define in this section. The material presented here can all be found in basic textbooks like [45, 46].

At the most basic level, algebraic geometry is about sets of zeros of polynomial functions. Throughout this paper, we will be working with polynomials in n complex variables $x_1, \dots, x_n$. We denote the ring of such polynomials by $\mathbb{C}[x_1, \dots, x_n]$. A fundamental concept in algebraic geometry is the polynomial ideal:

Definition 10. *The* polynomial ideal I *generated by polynomials* $g_1(x), \dots, g_k(x) \in \mathbb{C}[x_1, \dots, x_n]$ *is the set*

$$I = \left\{ \sum_{i=1}^{k} a_i(x) g_i(x) : a_i(x) \in \mathbb{C}[x_1, \dots, x_n] \right\}.$$

The polynomials $g_i(x)$ are called a *generating set* for the ideal, and we write $I = \langle g_1(x), \dots, g_k(x) \rangle$. Note that the same ideal can be generated by many different generating sets.

Another fundamental concept is the algebraic variety:

Definition 11. *A set* $V \in \mathbb{C}^n$ *is called an* (affine) algebraic variety *if* $V = \{x : u_1(x) = \cdots = u_k(x) = 0\}$ *for some polynomials* $u_1(x), \dots, u_k(x)$.

Every ideal I has an associated variety $V(I)$, which is the set of common zeros of all polynomials in I (or equivalently, the set of common zeros of all the generators of I for any generating set).

In this paper, we will be using some theorems concerning intersections of varieties. These properties are most conveniently stated not in $\mathbb{C}^n$, but in the complex *projective* space $\mathbb{P}^n$. There are several ways to define $\mathbb{P}^n$, but for our purposes it will be most convenient to use *homogeneous coordinates*: we define $\mathbb{P}^n$ as the set of all points $(x_0, x_1, \dots, x_n) \in \mathbb{C}^{n+1} - \{0\}$ up to multiplication by a nonzero constant. Thus, $(x_0, x_1, \dots, x_n)$ denotes the same point as $(\lambda x_0, \lambda x_1, \dots, \lambda x_n)$. Henceforth, we will denote the homogeneous coordinates using $\tilde{x}$. The hyperplane $x_0 = 0$ can be thought of as the set of "points at infinity."

We define a *homogeneous polynomial* to be the sum of monomial terms that are all of the same degree. Given any polynomial function $f(x)$ on $\mathbb{C}^n$ of degree d, we define its homogenization by $\tilde{f}(\tilde{x}) = x_0^d f(x_1/x_0, \dots, x_n/x_0)$. Using these concepts, we can define a *projective algebraic variety* as a set of the form $\mathcal{V} = \{\tilde{x} \in \mathbb{P}^n : \tilde{u}_1(\tilde{x}) = \cdots = \tilde{u}_k(\tilde{x}) = 0\}$, where $\tilde{u}_i(\tilde{x})$ are homogeneous polynomials. Given any affine variety in $\mathbb{C}^n$, we can produce a corresponding projective variety on $\mathbb{P}^n$ by homogenizing the defining polynomials. Likewise, we can go from a projective variety to an affine variety by dehomogenizing, i.e. intersecting with $\{x_0 = 1\}$.

In general, an algebraic variety may not be a smooth manifold in $\mathbb{C}^n$ or $\mathbb{P}^n$—it may have one or more singular points. A criterion for smoothness can be obtained from the Jacobian matrix associated with the variety. The Jacobian matrix of the variety $V = \{x \in \mathbb{C}^n : u_1(x) = \cdots =$

$u_k(x) = 0\}$ is given by

$$J = \begin{pmatrix} \frac{\partial u_1(x)}{\partial x_1} & \cdots & \frac{\partial u_k(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_1(x)}{\partial x_n} & \cdots & \frac{\partial u_k(x)}{\partial x_n} \end{pmatrix}.$$

A point $x \in V$ is a singular point if the matrix J has less than full rank at x. V is smooth if it has no singular points. The *codimension* of V (i.e. $n - \dim V$) is equal to the rank of J at nonsingular points. This also coincides with the intuitive meaning of dimension (from differential geometry) as applied to manifolds. If a variety on $\mathbb{C}^n$ or $\mathbb{P}^n$ has dimension $n-1$, we call it a *hypersurface*. Using the correspondence between ideals and varieties, we can also define the dimension of an ideal I as the dimension of the associated affine variety $V(I)$.

The last basic notion we will need is the idea of "genericity." To define this precisely in the context of algebraic geometry, we need to introduce the Zariski topology. This is the topology over $\mathbb{C}^n$ or $\mathbb{P}^n$ in which the closed sets are precisely the algebraic varieties. We say that a property over points in $\mathbb{C}^n$ or $\mathbb{P}^n$ is *generic* if it is true for a Zariski open dense subset of $\mathbb{C}^n$. Note that all Zariski closed sets are also closed in the standard topology, and therefore all Zariski open sets are open in the standard topology. So if a set is generic in the sense defined here, it is also open and dense in $\mathbb{C}^n$ under the standard topology.

1. Gröbner bases

We noted above that a polynomial ideal can have many different generating sets. However, there is a notion of a canonical generating set, called a *Gröbner basis*, that is computationally useful. To define it, we must first define the notion of a monomial ordering.

Definition 12. *A monomial ordering is any total ordering $\prec$ on the set of monomials satisfying the following:*

(i) If $a \prec b$, then for any monomial c, $ac \prec bc$.

(ii) Any nonempty subset of monomials has a smallest element (the well-ordering property).

An important class of monomial orderings is the degree orderings: these are the orderings in which if $deg(a) > deg(b)$, then $a \succ b$.

Once we have chosen a monomial ordering, for any polynomial $f(x)$ we can define the *leading term* $\mathrm{LT}(f(x))$ as the monomial term in $f(x)$ that is highest according to our chosen ordering. With these notions in place, we can define the Gröbner basis as follows.

Definition 13. *A collection of polynomials $\{g_1(x), \ldots, g_k(x)\}$ is a* Gröbner basis *of an ideal I if $I = \langle g_1(x), \ldots, g_k(x) \rangle$ and*

$$\langle \mathrm{LT}(g_1(x)), \ldots, \mathrm{LT}(g_k(x)) \rangle = \langle \{\mathrm{LT}(f(x)) : f(x) \in I\} \rangle.$$

Gröbner bases were introduced by Buchberger [47], who showed that every ideal has a finite Gröbner basis, and gave an algorithm to compute this basis for any given monomial ordering.

A key application of the Gröbner basis is in the *Gröbner basis division algorithm.* The output of this algorithm is described in the following proposition.

Proposition 14. *Let $f(x)$ be any polynomial, and I be an ideal with a degree-ordered Gröbner basis $\{g_1(x), \ldots, g_k(x)\}$. If D is the maximum degree of the Gröbner basis elements, then there exists a unique decomposition $f(x) = \sum a_i(x)g_i(x) + u(x)$, where $\deg(a_i(x)) \leq \deg(f(x))$, and no term of $u(x)$ is divisible by the leading term of a Gröbner basis element. Moreover, if $f(x) \in I$, then $u(x) = 0$.*

Proof. This is an immediate consequence of Proposition 1 in Section 2.6 and Theorem 3 in Section 2.3 of [45]. □

If an ideal is generated by homogeneous polynomials, then the degree-ordered Groebner basis can also be taken to be homogeneous.

Proposition 15. *Let $\tilde{I} = \langle \tilde{h}_1(\tilde{x}), \ldots \tilde{h}_k(\tilde{x})\rangle$ be an ideal generated by homogeneous polynomials, and $\{g_1(\tilde{x}), \ldots, g_k(\tilde{x})\}$ be a degree-ordered Gröbner basis for $\tilde{I}$. If we let $g_i'(\tilde{x})$ be the highest-degree terms of $g_i(\tilde{x})$, then $\{g_1'(\tilde{x}), \ldots, g_k'(\tilde{x})\}$ is also a degree-ordered Gröbner basis for I.*

Proof. We need to show that $\{g_1'(\tilde{x}), \ldots, g_k'(\tilde{x})\}$ is a generating set for $\tilde{I}$, and that the condition in Definition 13 still holds. The latter follows immediately from the fact that $\mathrm{LT}(g_i'(\tilde{x})) = \mathrm{LT}(g_i(\tilde{x}))$ for degree orderings. As for the former, suppose that $f(\tilde{x}) \in \tilde{I}$, meaning that $f(\tilde{x}) = \sum_i u_i(\tilde{x})\tilde{h}_i(\tilde{x})$. Let $P_d f$ denote the degree-d terms of $f(\tilde{x})$. Then $P_d f(\tilde{x}) = \sum_i (P_{d-\deg(\tilde{h}_i)} u_i(\tilde{x}))\tilde{h}_i(\tilde{x})$, so $P_d f(\tilde{x}) \in \tilde{I}$. Now, for any Gröbner basis element $g_i(\tilde{x})$, let $d < \deg(g_i(\tilde{x}))$. Since $P_d g_i(\tilde{x}) \in \tilde{I}$, by Proposition 14, $P_d g_i(\tilde{x}) = \sum a_{ij}(\tilde{x})g_j(\tilde{x})$, where the sum only contains Gröbner basis elements with degree at most d. Since $d < \deg(g_i(\tilde{x}))$, this means in particular that this sum does *not* include $g_i(\tilde{x})$. This implies that we can replace $g_i(\tilde{x})$ by $g_i(\tilde{x}) - P_d g_i(\tilde{x})$, and still have a generating set for $\tilde{I}$. By repeatedly applying this process, we can replace each $g_i(\tilde{x})$ by $g_i'(\tilde{x})$ and still have a generating set. Thus, $\{g_1'(\tilde{x}), \ldots, g_k'(\tilde{x})\}$ is indeed a Gröbner basis for $\tilde{I}$. □

The dimension of an ideal is related to properties of its Gröbner basis. For ideals of any dimension, the following bound on the degree of the Gröbner basis was shown in [48].

Proposition 16. *For an r-dimensional ideal generated by polynomials of degree at most d in n variables, with coefficients over any field, the Gröbner basis in any ordering has degree upper-bounded by*

$$2\left(\frac{1}{2}d^{n-r} + d\right)^{2^r}.$$

In the special case of zero-dimensional ideals, we further have the following property:

Proposition 17. *Let I be an ideal and $\{g_1(x), \ldots, g_k(x)\}$ a Gröbner basis for I. Then I is zero-dimensional iff for every variable x_i, there exists $m_i \geq 0$ such that $x_i^{m_i} = \mathrm{LT}(g(x))$ for some element $g(x)$ in the Gröbner basis.*

Proof. This is the equivalence (i) $\iff$ (iii) in Theorem 6 of Chapter 5 of [45]. □

This result enables us to bound the degree of the remainder term u in Proposition 14 above, when the ideal is zero dimensional.

Proposition 18. *If I is a zero-dimensional ideal over n variables, and it has a degree-order Gröbner basis whose maximum total degree is D, then the remainder $u(x)$ in Proposition 14 has degree at most $n(D-1)$.*

Proof. Suppose $u(x)$ contains a term with degree greater than $n(D-1)$. Then this term would be divisble by x_i^D for some variable x_i. However, since I is zero dimensional, by the above proposition there exists a Gröbner basis element g_j whose leading term is x_j^k for some $k < D$. Thus, we have found a term in $u(x)$ that is divisble by the leading term of a Gröbner basis element, which contradicts Proposition 14. $\square$

2. Intersections of varieties

Finally, we include two important theorems concerning the intersections of projective algebraic varieties. In full generality these theorems are much more powerful than we need; the statements we give here are tailored for our use, and are based on those in [38]. The first theorem is Bézout's Theorem, which says that two projective varities of sufficiently high dimension must intersect (the full version also bounds the number of components in the intersection):

Theorem 19 (Bézout)**.** *Suppose $\mathcal{U}$ and $\mathcal{V}$ are projective varieties in $\mathbb{P}^n$, and* $\dim(\mathcal{U})+\dim(\mathcal{V}) \geq n$. *Then $\mathcal{U}$ and $\mathcal{V}$ have a nonempty intersection.*

The second theorem is Bertini's Theorem. Roughly, this states that the intersection of a smooth variety with a "generic" hypersurface is also a smooth variety with dimension 1 lower. The precise statement is:

Theorem 20 (Bertini)**.** *Let $\mathcal{U}$ be a k-dimensional smooth projective variety in $\mathbb{P}^n$, and $\mathcal{H}$ a family of hypersurfaces in $\mathbb{P}^n$ parametrized by coordinates in a projective space $\mathbb{P}^m$. If there are no points common to all the hypersurfaces in $\mathcal{H}$, then for generic $\mathcal{A} \in \mathcal{H}$, the intersection $\mathcal{U} \cap \mathcal{A}$ is smooth and has dimension $k-1$.*