Improved Semidefinite Programming Hierarchy for h_{Sep}
with tools from Algebraic Geometry

Aram W. Harrow [1] **Anand Natarajan** [1] Xiaodi Wu [2]

[1]MIT

[2]University of Oregon

QMA(2) Workshop, QuICS, Aug 3rd 2016
arXiv:1506.08834

The problem h_{Sep}

Definition (Separable states)

A bipartite state $\rho \in \mathrm{D}\left(\mathcal{X} \otimes \mathcal{Y}\right)$ is *separable* if

$$\rho \in \mathrm{conv}(\{\sigma_X \otimes \sigma_Y : \sigma_X \in \mathrm{D}\left(\mathcal{X}\right), \sigma_Y \in \mathrm{D}\left(\mathcal{Y}\right)\}).$$

Let $\mathsf{Sep} \stackrel{\text{def}}{=} \{ \text{ separable states } \}$.

Definition (h_{Sep})

Given a Hermitian operator M over $\mathcal{X} \otimes \mathcal{Y}$,

$$h_{\mathsf{Sep}}(M) := \max_{\rho \in \mathsf{Sep}} \mathrm{Tr}[M\rho].$$

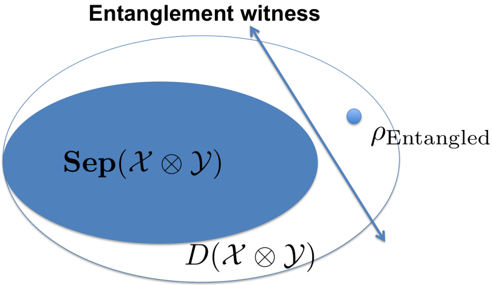
The problem WOpt(M, ϵ) is to compute an ϵ-additive approximation of $h_{\mathsf{Sep}}(M)$.

Entanglement Testing

Definition (Weak Membership)

WMem($\epsilon, \|\cdot\|$) : for any $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, decide whether $\rho \in$ Sep or $\|\rho -$ Sep$\| \geq \epsilon$.

- By "GLS theorems" in convex optimization, can solve WMem($\epsilon, \|\cdot\|_1$) efficiently with an oracle to WOpt(M, ϵ). (Uses ellipsoid method)
- Any M with $h_{\text{Sep}}(M) \leq \lambda_{\max}(M)$ is an *entanglement witness*.



Applications

- Finding the maximum acceptance probability of a QMA(2) machine $\Leftrightarrow$ solving WMem($\epsilon, \|\cdot\|_1$) for M arising from poly-time quantum circuit.
- Many more (mean-field aproximations, output entropies of channels, etc.)

Algorithms and Hardness

- Complexity is a function of *dimension* n and *accuracy* ϵ.
- Algorithms based on Sum-of-Squares [DPS] and ϵ-nets, [SW, BH, ...].
- **When $\epsilon = \Omega(1)$:**
  - Need time $n^{\Omega(\log n)}$ assuming Exponential-Time Hypothesis [HM]. (Unconditionally for SoS hierarchy [HNW]).
  - Exists $n^{O(\log(n)/\epsilon^2)}$ algorithm for **1-LOCC M** [BCY, BH].
- **When $\epsilon = 1/\operatorname{poly}(n)$:**
  - $\mathrm{WOpt}(\cdot, \epsilon)$ is NP-hard. [Gur, Ioa, Gha]
  - Exist $(n/\sqrt{\epsilon})^{O(n)}$ algorithms based on SDPs. [NOP]
- **What about scaling with ϵ?**

Why care about ϵ scaling?

- Surprising jumps in complexity in high-accuracy regime:
  - QIP jumps from PSPACE to EXP with inverse doubly exponential ϵ. [IKW]
  - QMA equals PSPACE for inverse-exponential ϵ. [FL]
  - QMA(2) equals NEXP for inverse-exponential ϵ. [Per]
- Better understand "brute-force algorithms" for quantum problems
  - Ex: **no known algorithm** to approximate entangled value of non-local game to constant factor.
- Could lead to algorithms that perform **better in practice**

Results

Theorem (Main)

There exists an algorithm that estimates $h_{\mathrm{Sep}(n)}(M)$ to error ϵ in time $\exp(\mathrm{poly}(n))\,\mathrm{poly}\log(1/\epsilon)$. similar for the multi-partite case.

- One algorithm based on **quantifier elimination**—less conceptually interesting, so skip for this talk.
- Other algorithm based on **SoS hierarchy with added constraints**, inspired by [Nie, NR].
- Algorithm is **exact**: only approximation comes from numerical error in SDP solver
- As a complexity result:

  $$\mathrm{QMA}_{\log}(2)[c, s = c - 1/\exp(\exp(n))] \subseteq \mathrm{DTIME}(\exp(\mathrm{poly}(n))).$$

Comparison with DPS

Advantages

- **Primal of SDP** leads to new monogamy relations relative to particular observables.
- **Dual of SDP** leads to a new class of entanglement witness.
- Analog of the exact convergence achievable for discrete optimization, e.g., SoS for Boolean CSPs. (Note: DPS does *not* converge at any finite level)

Disadvantages

- Unlike [DPS], set of entanglement witnesses could be *non-convex*.
- Hence, cannot solve WMem directly (need to use GLS reduction).

Proof Outline

- Treat WOpt as a constained polynomial optimization problem.
- Use **Karush-Kuhn-Tucker conditions** to restrict to *critical set* (technique from [Nie, NR])
- **Claim 1:** For generic M, critical set consists of $\exp(n)$-many discrete points. (From Bertini's thm)
- Algorithm: use SoS to approximately optimize over critical points by searching for **certificates of positivity.**
- **Claim 2:** For generic M, optimum has SoS certificate of degree $\exp(n)$. (From Claim 1 and Gröbner basis theory)
- (Continuity argument to handle non-generic M).

Simplification: the symmetric real case

By simple reductions [HM], we can restrict to states of the form $|\psi\rangle \otimes |\psi\rangle, |\psi\rangle \in \mathbb{R}^n$.

$$h_{\mathrm{ProdSym}(n)}(M) := \begin{array}{ll} \max\limits_{x \in \mathbb{R}^n} & f_0(x) := \sum\limits_{i_1, i_2, j_1, j_2} M_{(i_1, i_2),(j_1, j_2)} x_{i_1} x_{i_2} x_{j_1} x_{j_2} \\ \text{subject to} & f_1(x) := ||x||^2 - 1 = 0. \end{array} \tag{1}$$

REMARK: this is a *polynomial optimization* problem with a homogenous degree 4 objective and a degree 2 constraints.

Principle of Sum-of-Squares

One way to show that a polynomial $f(x)$ is *nonnegative* could be

$$f(x) = \sum a_i(x)^2 \geq 0.$$

Example

$$\begin{aligned} f(x) &= 2x^2 - 6x + 5 \\ &= (x^2 - 2x + 1) + (x^2 - 4x + 4) \\ &= (x-1)^2 + (x-2)^2 \geq 0. \end{aligned}$$

Such a decomposition is called a *sum of squares (SoS) certificate* for the non-negativity of f.

Principle of SoS : constrained domain

Definition (Variety)

A set $V \subseteq \mathbb{C}^n$ is called an *algebraic variety* if $V = \{x \in \mathbb{C}^n : g_1(x) = \cdots = g_k(x) = 0\}$.

Non-negativity of $f(x)$ on V can be shown by

$$f(x) = \sum a_i(x)^2 + \sum b_j(x) g_j(x) \geq 0.$$

Question: do all nonnegative polynomials on certain variety have a SoS certificate?

Putinar's Positivstellensatz

Definition (Ideal)

The *polynomial ideal* I generated by $g_1, \ldots, g_k \in \mathbb{C}[x_1, \ldots, x_n]$ is

$$I = \{\sum a_i g_i : a_i \in \mathbb{C}[x_1, \ldots, x_n]\} := \langle g_1, \cdots, g_k \rangle.$$

Theorem (Putinar's Positivstellensatz)

Under the Archimedean condition, if $f(x) > 0$ *on* $V(I) \cap \mathbb{R}^n$*, then*

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

for some choice of SoS polynomial $\sigma(x)$ *and* $g(x) \in I$*.*

SoS in Optimization

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0 \quad \forall i \end{aligned} \tag{2}$$

is equivalent to (under Archimedean condition)

$$\begin{aligned} \min \quad & \nu \\ \text{such that} \quad & \nu - f(x) = \sigma(x) + \sum_i b_i(x) g_i(x), \end{aligned} \tag{3}$$

where $\sigma(x)$ is SoS and $b_i(x)$ is any polynomial.

SoS hierarchy

- If $\sigma(x)$ and $b_i(x)$ can have *arbitrarily high* degrees, then the optimization problem (3) is equivalent to problem (2).
- By bounding the degrees, i.e., $\deg(\sigma(x))$, $\deg(b_i(x)g_i(x)) \leq 2D$ for some integer D, we get **SoS hierarchy at level** D.

$$\begin{aligned} \text{min} \quad & \nu \\ \text{such that} \quad & \nu - f(x) = \sigma(x) + \sum_i b_i(x) g_i(x), \end{aligned} \tag{4}$$

where $\sigma(x)$ is SoS and $b_i(x)$ is any polynomial and $\deg(\sigma(x))$, $\deg(b_i(x)g_i(x)) \leq 2D$.

DPS algorithm is SoS applied to h_{Sep}

Why is it an SDP?

Observation

- Any $p(x)$ (of degree $2D$) $= m^T Qm$, where m is the vector of monomials of degree up to $2D$ and Q is the coefficients.
- $p(x)$ is a SoS iff $Q \geq 0$.

$$\begin{aligned} \min_{\nu, b_{i\alpha} \in \mathbb{R}} \quad & \nu \\ \text{such that} \quad & \nu A_0 - F - \sum_{i\alpha} b_{i\alpha} G_{i\alpha} \geq 0. \end{aligned} \tag{5}$$

Complexity: $\text{poly}(m)\,\text{poly}\log(1/\epsilon)$, where $m = \binom{n+D}{D}$.

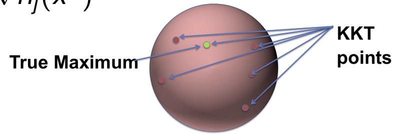
Karush-Kuhn-Tucker Conditions

For any optimization problem

$$\max f(x) \text{ s.t. } g_i(x) \leq 0, h_j(x) = 0, \forall i, j,$$

if x^* is a *local* optimizer, then $\exists \mu_i, \lambda_j$,

$$\begin{aligned} \nabla f(x^*) &= \sum \mu_i \nabla g_i(x^*) + \sum \lambda_j \nabla h_j(x^*) \\ g_i(x^*) &\leq 0, h_j(x^*) = 0, \\ \mu_i &\geq 0, \mu_i g_i(x^*) = 0. \end{aligned}$$



Remark: for convex optimization (our case), any global optimizer satisfies KKT.

Our case

Recall our optimization problem is

$$\max f_0(x) \text{ s.t. } f_1(x) = 0.$$

The KKT condition is $\nabla f_0(x) = \lambda \nabla f_1(x)$, which is equivalent to

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

$$g_{ij}(x) := \frac{\partial f_0(x)}{\partial x_i}\frac{\partial f_1(x)}{\partial x_j} - \frac{\partial f_0(x)}{\partial x_j}\frac{\partial f_1(x)}{\partial x_i} = 0, \quad \forall i, j$$

DPS+: Optimization with KKT constraints

$$
\begin{array}{lll}
\text{min} & \nu & \\
\text{such that} & \nu - f_0(x) \geq 0 & \\
 & f_1(x) = 0 & \\
\text{KKT} & g_{ij}(x) = 0 & \forall\, 1 \leq i \neq j \leq 2n
\end{array}
$$

- **DPS+ hierarchy:** SoS applied to above optimization problem
- Will show finite convergence when $D = \exp(\text{poly}(n))$. Then $m = \binom{n+D}{D} = \exp(\text{poly}(n))$. Thus the final time is $\exp(\text{poly}(n))\,\text{poly}\log(1/\epsilon)$.

KKT Ideal

Definition (KKT Ideal & Variety)

$$I_K = \left\{ v(x)f_1(x) + \sum h_{ij}(x)g_{ij}(x) \right\} =< f_1(x), g_{ij}(x) > .$$

$$V(I_K) = \left\{ x \in \mathbb{C}^{2n} : \forall p(x) \in I_K, p(x) = 0 \right\}$$

Definition (KKT Ideal to degree *m*)

$$\begin{aligned} I_K^m &= \{ v(x)f_1(x) + \sum h_{ij}(x)g_{ij}(x) : \deg(v(x)f_1(x)) \leq m, \\ &\quad \forall\, i, j, \deg(h_{ij}g_{ij}) \leq m \}. \end{aligned}$$

Main Technical Theorems

Theorem (Zero-dimensionality of generic I_K)

For a generic M, $|V(I_K)| < \infty$ and I_K is zero-dimensional.

Theorem (Degree bound)

There exists $m = O(\exp(\text{poly}(n)))$, s.t. for a generic M, $\epsilon > 0$,

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

where $\sigma(x)$ is SoS and $\deg(\sigma(x)) \leq m, g(x) \in I_K^m$.

Corollary (SDP solution)

For a generic M, $h_{\text{ProdSym}(n)}(M)$ can be estimated to error ϵ in time $\exp(\text{poly}(n))\text{poly}\log(1/\epsilon)$.

From generic to arbitrary M

Observations

- Generic Ms are *dense*, and value of SDP should be a continuous function of M.
- Issue: SDP might be *infeasible* for non-generic M.

Solution

- Switch to the dual SDP: satisfies Slater's condition, i.e, strictly feasible.
- For a generic M, by strong duality, $h_{\mathrm{ProdSym}(n)}(M) = OPT_{\mathrm{mom}}(M)$.
- For non-generic inputs M, use the continuity of the dual SDP.

Proof of Theorem 1

Let $\mathcal{U} = \{f_1(x) = 0\}, \mathcal{W} = \{\forall\, i, j, g_{ij} = 0\}$. then $V(I_K) \subseteq \mathcal{U} \cap \mathcal{W}$. It suffices to show $|\mathcal{U} \cap \mathcal{W}| < \infty$. Construct $\mathcal{A} = \mathcal{X} \cap \mathcal{U}$ s.t.

$\mathcal{A} \cap \mathcal{W} = \emptyset$ and $\dim(\mathcal{X}) = n - 1$. Note $\mathcal{W} \cap \mathcal{A} = (\mathcal{W} \cap \mathcal{U}) \cap \mathcal{X}$. By Bézout's theorem, two varieties with dimension sum$\geq n$ must intersect. Thus

$$\dim(\mathcal{W} \cap \mathcal{U}) + \dim(\mathcal{X}) = \dim(\mathcal{W} \cap \mathcal{U}) + n - 1 < n.$$

This implies $\dim(\mathcal{W} \cap \mathcal{U}) = 0$ and thus $|V(I_K)| < \infty$.

Proof of Theorem 1: construct $\mathcal{X}$

Let $\mathcal{X} = \{f_0(x) = \mu\}$ for generic (μ, M). $\dim(\mathcal{X}) = n - 1$.
By Bertini's theorem, $\dim(\mathcal{A}) = \dim(\mathcal{U} \cap \mathcal{X}) = n - 2$.

The Jacobian matrix $J_{\mathcal{A}} = \begin{pmatrix} \frac{\partial f_0}{\partial x_1} & \frac{\partial f_1}{\partial x_1} \\ \vdots & \vdots \\ \frac{\partial f_0}{\partial x_n} & \frac{\partial f_1}{\partial x_n} \end{pmatrix}$ has $\operatorname{rank}(J_{\mathcal{A}}) = 2$.

$\mathcal{W}$ by definition says $\operatorname{rank}(J_{\mathcal{A}}) = 1$. Thus no intersection!

Subtleties: varieties over **projective space**, so need to consider point at infinity

Proof of Theorem 2

- Goal: given SoS certificate

$$\nu - f_0(x) = \sigma(x) + g(x), \sigma(x) \in \mathrm{SoS}, g(x) \in I_K^m,$$

  upper bound $\deg(\sigma(x)), \deg(g(x))$.
- Idea: start from any SoS certificate and lower degree of $\sigma(x)$ using Gröbner basis.
- Then, show that remaining high-degree terms in $g(x)$ vanish
- **Gröbner basis: set of polynomials generating I_k such that *leading term* of any poly in I_k is divisible by leading term of a basis element.**
- [MR]:

$$|V(I_K)| < \infty \implies \max \deg\{\gamma_i\} \leq D = \exp(\mathrm{poly}(n)).$$

Proof of Theorem 2: SoS term

- Let $\{\gamma_i\}$ be a Gröbner basis for I_K.
- Take any SoS certificate:

$$v - f_0(x) = \sigma(x) + g(x). \text{ s.t. } \sigma(x) \text{ SoS }, g(x) \in I_K^m.$$

- Let $\sigma(x) = \sum s_a(x)^2$. By properties of Gröbner basis

$$s_a(x) = g_a(x) + u_a(x), \text{ s.t. } g_a(x) \in I_K, \deg(u_a(x)) \leq nD.$$

- Thus

$$v - f_0(x) = \sigma'(x) + g'(x), \deg(\sigma'(x)) \leq \exp(\text{poly}(n)), g' \in I_K.$$

Proof of Theorem 2: Ideal term

Suffices to show $g' \in I_K^m, m = \exp(\mathrm{poly}(n))$.

- Since $f_0(x)$ and $\sigma(x)$ have degree at most m, $\deg(g'(x)) \leq m$. Remains to show that $g'(x)$ can be obtained as sum of degree-m multiples of *original* generators g_{ij}.
- First step: decompose $g'(x)$ in Gröbner basis: $g'(x) = \sum t_k\gamma_k(x)$ with $\deg(t_k\gamma_k(x)) \leq m$.
- Second step: decompose Gröbner basis in terms of original generators: $\gamma_k(x) = \sum u_{ij}(x)g_{ij}(x)$ with $\deg(u_{ij}) \leq m$.
- Thus

$$g'(x) = \sum t_k u_{ij} g_{ij}(x), \deg(t_k u_{ij}) \leq m, \implies g'(x) \in I_K^m.$$

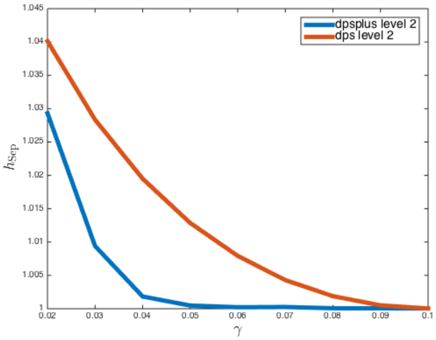
Numerical Results

DPS+ beats DPS at very high levels, but what about **low levels** of the hierarchy?

- DPS+ beats DPS at level 2, $n = 3$ for a family of measurements introduced by [DPS].

$$M_\gamma = \mathrm{Id} - (A_\gamma \otimes \mathrm{Id})Z(A_\gamma \otimes \mathrm{Id})$$
$$A_\gamma = \mathrm{diag}(1, 1/\gamma, \ldots, 1\gamma).$$

- For all $\gamma \in [0, 1]$, $h_{\mathrm{Sep}}(M_\gamma) = 1$.



Open Questions

DPS+

- Analyze the low levels of DPS+.
- Advantages of adding KKT conditions other than presented here.

Nonlocal games

- What is the correct version of KKT conditions for non-commutative polynomials?
- Can we have finite convergence for the commuting-operator game value? Would imply an upper bound on commuting-operator version of MIP* (*none known so far*)

SoS hierarchy

- Any other applications to quantum information?

Thank you!

Q & A