Semidefinite programming and computational aspects of entanglement **IHP Fall 2017**

Lecture 6: November 24

Lecturer: Aram Harrow *Scribe: Anand*

Rest of the course: algorithms (mostly SoS, a little ϵ-nets) for h_{Sep}.

We'll start with a high-level overview of the techniques we have. Recall that

$$\text{Sep}(d,k) = \text{conv}(\{\alpha_1 \otimes \ldots \otimes \alpha_k : \alpha_i \in D(\mathbb{C}^d)\}).$$

This is abbreviated as Sep when the dimension and number of parties is clear.

$$h_{\text{Sep}}(M) = \max_{\rho \in \text{Sep}} \text{tr}[M\rho] = \max_{\alpha_1,\ldots,\alpha_n \in D(\mathbb{C}^d)} \text{tr}\, M(\alpha_1 \otimes \ldots \otimes \alpha_k).$$

As motivation for this problem, here are some quantum problems that you could solve if you could compute h_{Sep} (see [2] for more):

- Membership in Sep.
- For a quantum channel $\mathcal{N}$ you can compute $\|\mathcal{N}\|_{1\to\infty}$ or $S_\infty^{\min}(\mathcal{N})$.
- Finding the least entangled state in a subspace.
- Ground energies of mean field Hamiltonians: $\lambda_{min}(\frac{1}{\binom{n}{2}} \sum_{i<j} H_{ij})$

And some non-quantum problems (see [1] for more):

- Maximizing a degree-4 polynomial over S^n.
- Computing the $\|A\|_{\ell_2\to\ell_4}$ for a tensor A.
- Small set expansion (closely related to unique games, see the earlier lectures).

And finally, this problem is interesting "because it is there": the best lower bound is quasipolynomial, so it might be one of the rare problems with complexity between polynomial and exponential.

6.1 Algorithms

6.1.1 De Finetti/monogamy of entanglement

The original de Finetti theorem was about infinite sequences of exchangeable random variables (e.g. infinite sequences of coin flips, where we don't care about the order in which the flips occur), and stated that any such sequence is a convex combination of iid sequences. This was generalized to the finite case by Diaconis and Friedman (essentially, they showed that sampling "with replacement" and "without replacement" give similar outcomes). To understand the quantum de Finetti theorem, we need to introduce the symmetric subspace.

Let S_n be the symmetric group on n letters. It acts on $(\mathbb{C}^d)^{\otimes n}$ by

$$P_d(\pi) = \sum_{i_1,\ldots,i_n \in [d]} \left|i_{\pi(1)},\ldots,i_{\pi(n)}\right\rangle \left\langle i_1,\ldots,i_n\right|.$$

The symmetric subspace is

$$\text{Sym}^n \mathbb{C}^d = \{|\psi\rangle \in (\mathbb{C}^d)^{\otimes n} : P_d(\psi)\,|\psi\rangle = |\psi\rangle \;\; \forall \pi \in S_n\}.$$

Theorem 1 (Quantum finite de Finetti theorem)**.** *Let $|\psi\rangle \in \mathrm{Sym}^n \mathbb{C}^d$, and let $\psi = |\psi\rangle\langle\psi|$. Then*

$$\mathrm{tr}_{n-k} \psi \approx \int_{\rho \in D(\mathbb{C}^d)} \rho^{\otimes k} \, d\mu(\rho),$$

where the approximation is up to error dk/n in fidelity.

A consequence of this theorem is that if $|\psi\rangle_{AB_1,\ldots,B_k} \in \mathbb{C}^{d_A} \otimes \mathrm{Sym}^n \mathbb{C}^{d_B}$, then the reduced state ψ_{AB} is close to separable:

$$\mathrm{dist}(\psi_{AB}, \mathrm{Sep}(d_A, d_B)) \leq \frac{d_B}{n}.$$

You can view this is a guarantee on a weak form of the SoS hierarchy, without the PPT constraints.

A further corollary for states with less symmetry: suppose $\rho_{A_1,\ldots,A_n} \in D((\mathbb{C}^d)^{\otimes n})$ is a mixed state satisfying

$$[\rho, P_d(\pi)] = 0 \; \forall \psi.$$

Then there exists a purification $|\psi\rangle_{A_1 B_1 \ldots A_n B_n} \in \mathrm{Sym}^n(\mathbb{C}^d \otimes \mathbb{C}^d)$, so applying the de Finetti theorem, one deduces that

$$\mathrm{tr}_{n-k} \rho \approx \int d\mu(\sigma) \, \sigma^{\otimes k},$$

where the error is now $d^2 k/n$ in fidelity.

Intuitively, this theorem is exploiting the phenomenon of "monogamy of entanglement." The idea is that if Alice is highly entangled with one of the Bobs, she cannot be highly entangled with any other. So the symmetry conditions enforce that Alice cannot be very entangled with *any* of the Bobs.

This is an algorithm for approximation h_{Sep} but not a great one, as the number of copies scales poorly with the dimension. Moreover we cannot hope for a better theorem of this form: there is a counterexample state. This is the "universal counterexample" in quantum information: the antisymmetric state.

$$|\phi\rangle = \frac{1}{\sqrt{n!}} \sum_{\pi \in S_n} \mathrm{sgn}(\pi) \, |\pi(1), \ldots, \pi(n)\rangle \in (\mathbb{C}^n)^{\otimes n}.$$

This state does not live in the symmetric subspace but its density matrix is symmetric. Moreover there is a "purification" of it in the symmetric subspace:

$$|\phi\rangle_{A_1,\ldots,A_n} |\phi\rangle_{B_1,\ldots,B_n} \in \mathrm{Sym}^n \mathbb{C}^{d^2}.$$

The reduced state onto two parties of the antisymmetric state is

$$\rho = \mathrm{tr}_{n-2} \phi = \frac{I - \mathrm{SWAP}}{n(n-1)} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{|ij\rangle - |ji\rangle}{\sqrt{2}} \frac{\langle ij| - \langle ji|}{\sqrt{2}}.$$

I claim that this state is very far from Sep. Indeed, let

$$M = \frac{I - \mathrm{SWAP}}{2}.$$

Then this measurement separates ρ from Sep:

$$\mathrm{tr}[M\rho] = 1$$

$$h_{\mathrm{Sep}}(M) = \max_{\alpha,\beta} \mathrm{tr}[M(\alpha \otimes \beta)] = \max_{\alpha,\beta} \frac{1 - \mathrm{tr}(\alpha\beta)}{2} = \frac{1}{2}.$$

We will come back to prove this theorem later in the course. For now, a final word on this theorem: for applications to mean-field Hamiltonians over e.g. spins, or any other situation where the local dimension of the systems is constant or small, this theorem can be useful, even though it not so helpful for h_{Sep} when the local dimension is large.

6.1.2 Variants of de Finetti (still not helpful for h_{Sep})

The first variant was developed for cryptography by Renner.

Theorem 2 (Exponential de Finetti theorem)**.** *For $|\psi\rangle$ in the symmetric subspace,*

$$\text{tr}_{n-k} \approx \int d\mu(\phi)\, \tau_\phi,$$

where the error scales as $n^{O(d)} \exp(-r(n-k)/k)$, and the states τ_ϕ are supported on the span of r-local operators acting on $|\phi\rangle^{\otimes k}$.

This is a relaxation of de Finetti because we only require the states to be "mostly" product. However the error scales much better with the number of subsystems traced out: you could take $r, n-k$ to both scale as $n^{2/3}$ and still get an exponential decrease in error.

This has been simplified to get something called the "postselection method" or "de Finetti reductions," by Christandl, König, and Renner, and subsequently other authors as well.

Theorem 3 (De Finetti reductions)**.** *If $|\psi\rangle$ in the symmetric subspace, then the following operator inequality holds:*

$$\psi \leq n^{O(d)}\, \mathbb{E}[\rho^{\otimes n}].$$

This is relevant to approximating $h_{\text{Sep}}(M)$ up to (large!) multiplicative error: useful for cryptography (where $n \gg d$ and $h_{\text{Sep}}(M)$ can be made extremely small), but not for optimization.

6.1.3 Different norms

Bipartite 1-LOCC Here is where we can make real progress from the optimization point of view. Let's say that a measurement M is 1-LOCC if

$$M = \sum_i A_i \otimes B_i, \quad A_i \geq 0, 0 \leq B_i \leq I, \sum_i A_i \leq I.$$

These measurements can be implemented by local operations and one-way classical communication from Alice to Bob. Based on this class of measurements, we can define a norm

$$\|\Delta\|_{1\text{-LOCC}} = \max_{M \in 1\text{-LOCC}} |\operatorname{tr} M\Delta|.$$

This norm obeys the inequalities

$$\|\Delta\|_2 \leq \|\Delta\|_{1\text{-LOCC}} \leq \|\Delta\|_1.$$

Further, let's define the set of k-extendible states as

$$k\text{-ext} = \{\rho_{AB} : \exists \sigma_{AB_1 \ldots B_k} \text{ s.t. } [I \otimes P_{d_B}(\pi), \sigma] = 0 \; \forall \pi \in S_k$$

. Note that

$$D(\mathbb{C}^{d_A d_B}) = 1\text{-ext} \supseteq 2\text{-ext} \supseteq 3\text{-ext} \cdots \supseteq \infty\text{-ext} = \text{Sep}\,.$$

In the 1-LOCC norm, there is a very strong de Finetti theorem:

Theorem 4 (Brandão Christandl Yard 1010.1750)**.**

$$\text{dist}_{1\text{-LOCC}}(k\text{-ext}, \text{Sep}) \leq \sqrt{\frac{\log d_A}{k}}.$$

Recall that the runtime of the k-ext hierarchy (a weak version of SoS) at level k scales as $d_A d_B^k$. The theorem of BCY implies that to obtain an ϵ-approximation to h_{Sep}, it suffices to take $k \approx \frac{\log d_A}{\epsilon^2}$, yielding a runtime of $\exp\left(\frac{(\log d_A)(\log d_B)}{\epsilon^2}\right)$. The original paper has a complicated proof, using ideas from many areas in quantum information, hypothesis testing, etc. A simpler proof can be found in Brandão-Harrow. Intuitively, 1-LOCC is easier than the general case because there is a classical message between the two quantum parties, and we can use classical techniques to analyze this message.

Multipartite 1-LOCC There is also a multipartite version of this norm. Consider measurements of the form

$$M = \sum_{i_1} A_{i_1} \otimes \sum_{i_2} B_{i_2}^{i_1} \otimes \sum_{i_3} C_{i_3}^{i_1,i_2} \otimes \dots.$$

These are a multipartite version of 1-LOCC, where Alice sends a message to Bob, Bob sends a message to Charlie, and so on. This class of measurements includes Bell-type measurements. The improved de Finetti theorem of Brandão and Harrow shows that

$$\mathrm{dist}_{1\text{-LOCC}}(k\text{−ext}, \mathrm{Sep}(d, \ell \text{ parties})) \leq \sqrt{\frac{\ell^2 \log d}{k}}.$$

This is encouraging because it matches exactly the hardness result of Chen and Drucker, who reduce 3-SAT on n variables to computing h_{Sep} up to constant ϵ for Bell measurements on $\sqrt{n}$ parties of $\tilde{O}(n)$ local dimension each.

Dimension-agnostic bounds If you instead consider measurements $M : A \to X$ where X represents a classical outcome register, then the BCY theorem becomes

$$\mathrm{dist}_{1\text{-LOCC}}(k\text{−ext}, \mathrm{Sep}) \leq \sqrt{\frac{\log d_A \log |X|}{k}}.$$

Can we get a de Finetti theorem that depends only on the outcome spaces of the measurements? Indeed, Brandão and Harrow give such a theorem. Recall in the setting nonlocal games, we were interested in conditional probability distributions

$$p(x, y|a, b)$$

realizable by the players in the game. The non-signalling condition implies that the marginal $p(x|a)$ is well-defined. Let's say that a distribution is k-extendible if there exists a non-signalling distribution

$$q(x, y_1, \dots, y_k|a, b_1, \dots, b_k) = q(x, y_{\pi(1)}, \dots, y_{\pi(k)}|a, b_{\pi(1)}, \dots, b_{\pi(k)}) \ \forall \pi,$$

such that

$$q(x, y|a, b) = p(x, y|a, b).$$

Then it can be shown that

$$\mathrm{dist}_{\text{product questions}}(p, LHV) \leq \sqrt{\frac{\log |X|}{k}},$$

where LHV denotes the set $\mathrm{conv}\{r(x|a)s(y|b)\}$ of distributions realizable using classical local hidden variables, and the distance is

$$\min_{\ell \in LHV} \max_{\mu_A, \mu_B} \sum_{a,b} \mu_A(a)\mu_B(b) \| p(\cdot, \cdot|a, b) - \ell(\cdot, \cdot|a, b) \|_1.$$

We can search over such q in time $\mathrm{poly}(|A| \cdot |X| \cdot (|B| \cdot |Y|)^k)$.

This result resembles log Sobolev inequalities: instead of bounding the norm of some error term (distance from the fixed point of Markov chain), one bounds the entropy.

The 2-norm? In a recent work, still "undigested" by us, Barak, Kothari, and Steurer '16 make progress on approximating $h_{\mathrm{Sep}}(M)$ by using the 2-norm. They show that one can distinguish between $h_{\mathrm{Sep}}(M) = 1$ and $h_{\mathrm{Sep}}(M) \leq \frac{1}{2}$ using $O(\sqrt{d} \log^2 d)$ levels of SoS (the full hierarchy, including the PPT constraints). In comparison, vanilla de Finetti needs d levels. To understand their improvement, we need to understand the proof of the de Finetti theorem. Roughly speaking, the proof imagines doing tomography on the $n-k$ systems being traced out, and showing that for most outcomes of the tomography process, the state remaining on the last k systems will be very close to product. BKS observe that, in the case that $h_{\mathrm{Sep}}(M) = 1$, this is overkill: it's sufficient for some outcomes of the tomography to succeed (in other words, one can use postselection/SLOCC operations in the tomography procedure). Besides this, BKS's analysis gains from switching between the S_1, S_2, and S_∞ norms, in a way that we still don't understand intuitively.

Local Hamiltonians Suppose instead of h_{Sep} I want to find

$$\lambda_{\min}\left(\frac{1}{|E|}\sum_{(i,j)\in E} H_{ij}\right),$$

where E is the edge set of some graph and each local term has norm $\|H_{ij}\| \leq 1$. Suppose further that E is a k-regular graph for large k. Then Brandão and Harrow showed that $\lambda_{\min}$ is close to the min energy over product states (this fits with the intuition from monogamy of entanglement). The error scales as $(d^2/k)^{1/3}$, where d is the local dimension of the system. Moreover, BH showed that if the Hamiltonian graph has not just high degree but also high expansion (or rather, high *threshold rank*), then there is an efficient algorithm to compute a product state approximation to the ground state of H.

This last result uses yet another variant of the sum of squares algorithm adapted for the local Hamiltonian problem. In this variant, the variables are the density matrices ρ_S of all sets S of ℓ qubits; or equivalently, the pseudoexpectations $\tilde{\mathbb{E}}[X]$ of all ℓ-local operators. The most interesting constraint on these pseudoexpectations is that

$$\tilde{\mathbb{E}}[X^\dagger X] \geq 0$$

for all $\ell/2$-local X. Note that X can be a linear combination of local terms touching sites from all over the system, so this constraint is a "global" one.

BH showed that if $\ell \geq \text{poly}(d, \frac{1}{\epsilon}, \text{rank}_\delta A)$, where $\delta = \text{poly}(\epsilon, 1/d)$ and $\text{rank}_\delta(A)$ is the threshold rank of A, i.e. the number of eigenvalues of A greater than δ, then the SoS hierarchy gives a good approximation to the ground energy. This follows a paper from Barak, Raghavendra, and Steurer.

In summary, we saw five examples of variants of h_{Sep} and related problems where algorithms similar SoS achieve nontrivial performance. Moreover, for most of these, there are no matching hardness results, leaving open the possibility that these algorithms perform much better than we can prove at present.

6.2 Next time, and a prequel

In the remaining lectures, we'll discuss two further topics: ϵ-net algorithms, and applications to the 2-4 norm and the small set expansion problem. But even before we get into this, it's useful to consider a classical analog of these problems as a "prequel".

Let Δ_d be the probability simplex over $[d]$, and let $f(x)$ be a degree k polynomial. We consider the problem

$$\max_{x\in\Delta_d} f(x).$$

This is easy for $k = 1$ and already NP-hard for $k = 2$. The proof of this is a beautiful result of Motzkin and Strauss: if A is the adjacency matrix of a graph on d vertices, and w is the clique number of this graph, then

$$\max_{x\in\Delta_d} x^T A x = 1 - \frac{1}{w}.$$

This tells us not only that the exact problem is NP-hard, but also how hard approximations are.

If you want a fun math problem for the weekend, try to prove the above. Next time, we'll explain the proof of this, and classical algorithms for this problem. We'll then conclude by proving the quantum de Finetti algorithms and discussing applications.

References

[1] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 307–326, 2012, arXiv:1205.4484.

[2] A. W. Harrow and A. Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *J. ACM*, 60(1):3:1–3:43, Feb. 2013, arXiv:1001.0017.