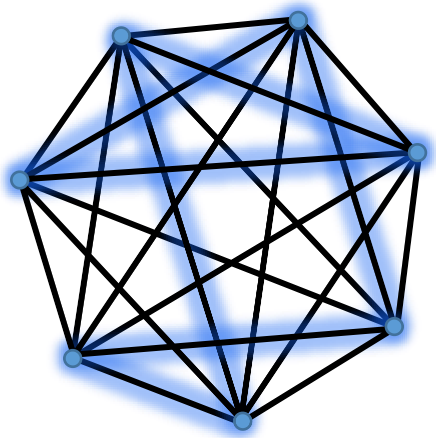
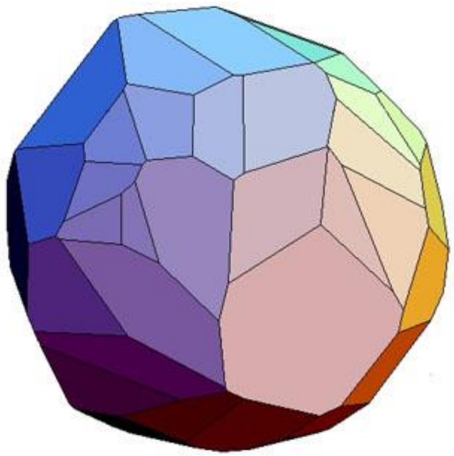
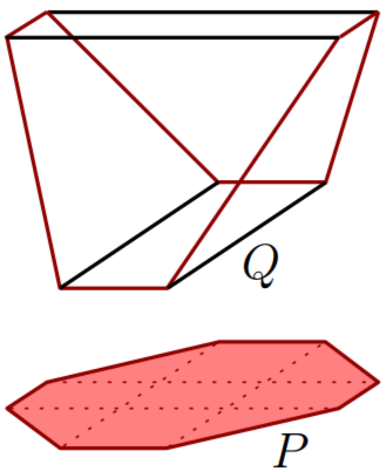
spectrahedral lifts of combinatorial polytopes

James R. Lee
University of Washington





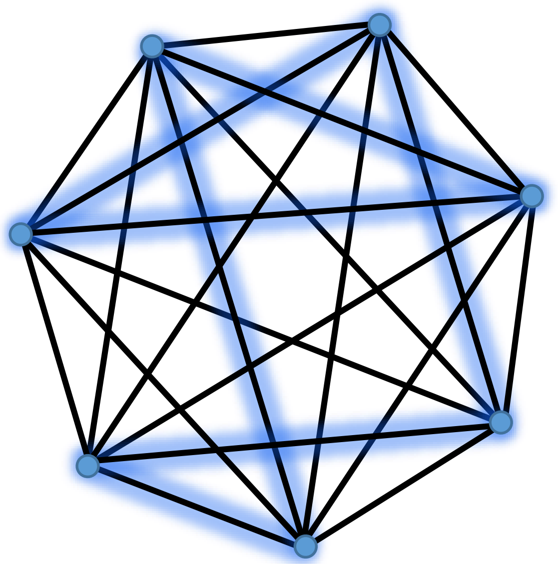


Prasad Raghavendra (UC Berkeley)   David Steurer (Cornell)

Traveling Salesman Problem:

Given n cities $\{1,2,\ldots,n\}$ and costs $c_{ij} \geq 0$ for traveling between cities i and j, find the permutation π of $\{1,2,\ldots,n\}$ that minimizies

$$c_{\pi(1)\pi(2)} + c_{\pi(2)\pi(3)} + \cdots + c_{\pi(n)\pi(1)}$$



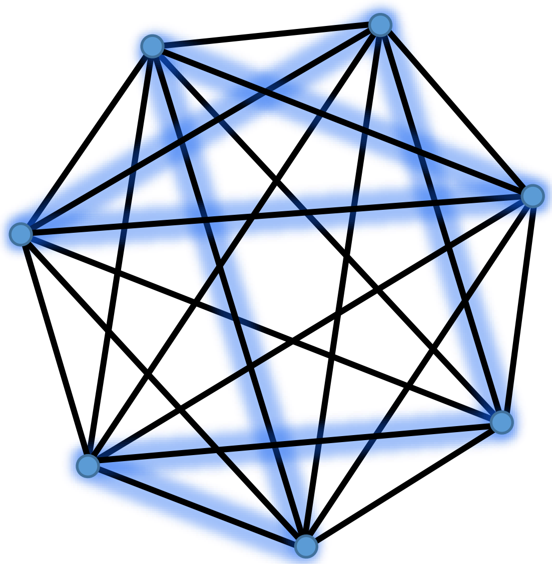
Attempts to solve the traveling salesman problem and related problems of discrete minimization have led to a revival and a great development of the theory of polyhedra in spaces of n dimensions, which lay practically untouched—except for isolated results—since Archimedes. Recent work has created a field of unsuspected beauty and power, which is far from being exhausted.

Gian Carlo Rota, 1969

Traveling Salesman Problem:

Given n cities $\{1,2,\dots,n\}$ and costs $c_{ij} \geq 0$ for traveling between cities i and j, find the permutation π of $\{1,2,\dots,n\}$ that minimizies

$$c_{\pi(1)\pi(2)} + c_{\pi(2)\pi(3)} + \cdots + c_{\pi(n)\pi(1)}$$



$$\mathrm{TSP}_n = \mathrm{conv}\left(\left\{1_\tau \in \{0,1\}^{\binom{n}{2}} : \tau \text{ is a tour}\right\}\right) \subseteq \mathbb{R}^{\binom{n}{2}}$$

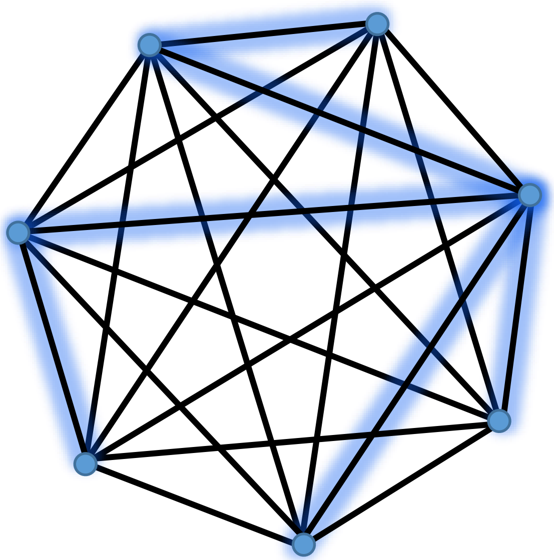
Can find an optimal tour by minimizing a linear function over TSP_n: $\min \{ \langle c, x \rangle : x \in \mathrm{TSP}_n \}$

Problem: TSP_n has exponentially (in n) many facets.

One can tell the same (short) story for many polytopes associated to NP-complete problems.

Minimum Spanning Tree:

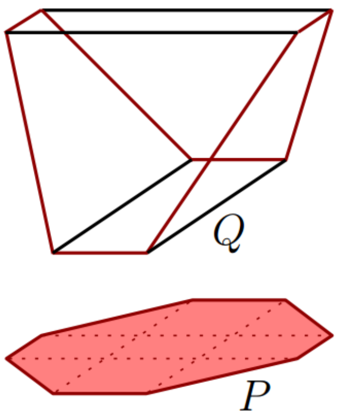
Given n cities $\{1,2,\ldots,n\}$ and costs $c_{ij} \geq 0$ between cities i and j, find a spanning tree of minimum cost.



$$\mathrm{ST}_n = \mathrm{conv}\left(\left\{1_\tau \in \{0,1\}^{\binom{n}{2}} : \tau \text{ is a spanning tree}\right\}\right)$$

Again, has exponentially (in n) many facets.

There is a **lift** of ST_n in n^3 dimensions with only $O(n^3)$ facets.



A **lift** Q of $P \subseteq \mathbb{R}^d$ is a polytope $Q \subseteq \mathbb{R}^N$ for $N \geq d$ such that Q linearly projects to P.

If we can optimize over Q, then we can optimize over P:

$$\max\{\langle c, x\rangle : x \in P\} \quad = \quad \max\{\langle \bar{c}, x\rangle : x \in Q\}$$

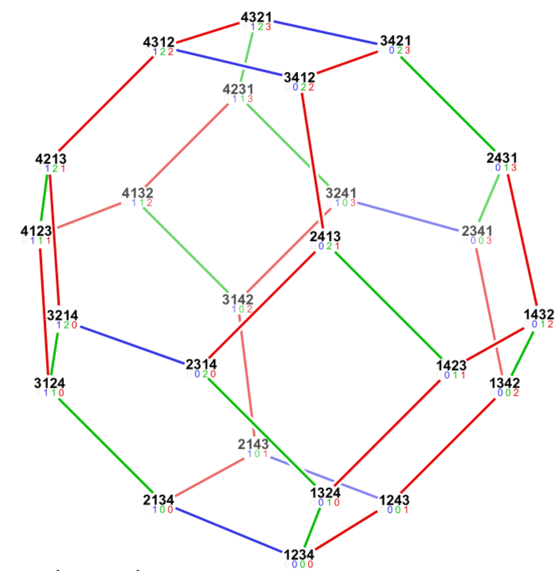
example: the permutahedron

Permutahedron:

$\Pi_n \subseteq \mathbb{R}^n$

Convex hull of vectors $(i_1, i_2, \ldots, i_n)$ such that $\{i_1, i_2, \ldots, i_n\} = \{1,2, \ldots, n\}$.

Number of facets is $2^n - 2$.



Let $\overline{\Pi}_n \subseteq \mathbb{R}^{n^2}$ be the convex hull of permutation matrices.

The map $A \mapsto (1, 2, \ldots, n)A$ is a linear projection from $\overline{\Pi}_n$ onto Π_n (i.e., $\overline{\Pi}_n$ is a lift of Π_n)

Birkhoff-von Neumann: $\overline{\Pi}_n$ is the set of matrices $A = (a_{ij})$ satisfying $a_{ij} \geq 0$ for all i, j with all row and column sums equal one. Thus $\overline{\Pi}_n$ has n^2 facets.

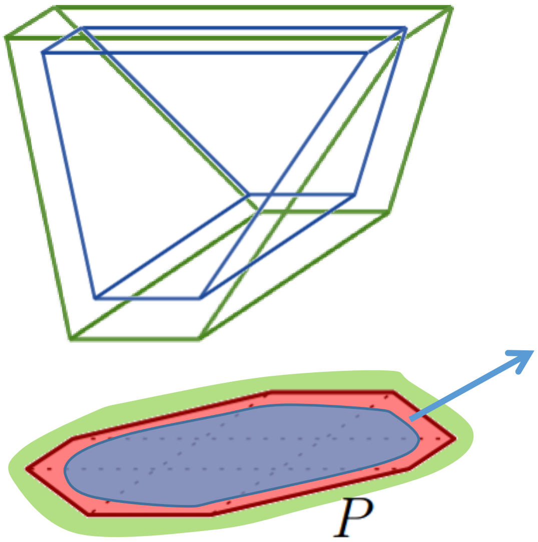
[Goemans 2013]: Smallest lift has $\Theta(n \log n)$ facets.

a general model of (small) linear programs

Powerful model of computation:

For a polytope $P \subseteq \mathbb{R}^d$, what is the minimal # of facets in a lift of P?

Even more powerful when we allow **approximation**:



Indication of power:

Integrality gaps for LPs often lead to NP-hardness of approximation.

Polynomial-size LPs for NP-hard problems would show that $\mathrm{NP} \subseteq \mathrm{P}/\mathrm{poly}$.
[Rothvoss 2013]

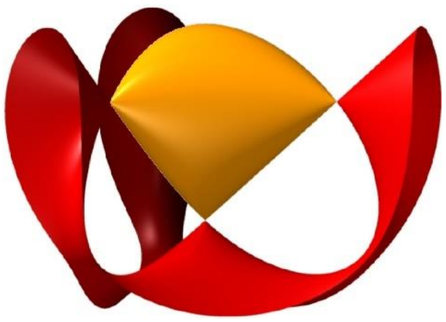
an analog for semi-definite programs

Semidefinite programs (aka spectrahedral lifts):

Let $\mathcal{S}_k^+$ denote the cone of $k \times k$ symmetric, positive semidefinite matrices.

A **spectrahedron** is the intersection $\mathcal{S}_k^+ \cap \mathcal{L}$ for some affine subspace $\mathcal{L}$.

This is precisely the feasible region of an SDP.



Definition: A polytope P admits a **PSD lift** of size k if P is a linear projection of a spectrahedron $\mathcal{S}_k^+ \cap \mathcal{L}$.

- Easy to see that minimal size of PSD lift is ≤ minimal size of polyhedral lift
- Assuming the Unique Games Conjecture, integrality gaps for SDPs translate directly into NP-hardness of approximation results.

[Khot-Kindler-Mossel-O'Donnell 2004, Austrin 2007, Raghavendra 2008]

a brief history of extended formulations

1980s: Swart tries to prove that $P = NP$ by giving a linear program for TSP.

1989: Yannakakis (the referee for Swart's submissions) shows that every *symmetric* LP for TSP must have exponential size.

. . .

2012: Fiorini, Massar, Pokutta, Tiwary, de Wolf show that *every* LP for TSP must have exponential size.

2013: Chan, L, Ragahvendra, Steurer show that no polynomial-size LP can approximate the MAX-CUT problem within a factor better than 2. [Goemans-Williamson 1998: SDPs can do factor ≈ 1.139.]

2014: Rothvoss shows that every LP for the matching polytope must have exponential size.

????: Spectahedral lifts?

[Briët-Dadush-Pokutta 2014]: Random $\{0,1\}$-polytopes do not have PSD lifts of size $e^{o(n)}$.

our results [L-Raghavendra-Steurer]

Lower bounds on PSD lift size

The TSP_n, CUT_n, and $\mathrm{STAB}(G_n)$ polytopes do not admit PSD lifts of size $c^{n^{2/11}}$
(for some constant $c > 1$ and some family $\{G_n\}$ of n-vertex graphs)

Lower bounds on PSD rank via sums-of-squares / Lasserre hierarchy

Approximation hardness for constraint satisfaction problems

For instance, no family of polynomial-size SDP relaxations can achieve better than a $7/8$-approximation for MAX 3-SAT.

slack matrices & psd factorizations

Consider a non-negative matrix $M \in \mathbb{R}_+^{m \times n}$

$$\mathrm{rank}(M) = \min \{ r : M_{ij} = \langle u_i, v_j \rangle, \qquad \{u_i, v_j\} \subseteq \mathbb{R}^r \}$$

$$\mathrm{rank}_+(M) = \min \{ r : M_{ij} = \langle u_i, v_j \rangle, \qquad \{u_i, v_j\} \subseteq \mathbb{R}_+^r \}$$

$$\mathrm{rank}_{\mathrm{psd}}(M) = \min \{ r : M_{ij} = \langle U_i, V_j \rangle, \qquad \{U_i, V_j\} \subseteq \mathcal{S}_+^r \}$$

where $\langle U, V \rangle = \mathrm{Tr}(U^T V)$ for $U, V \in \mathcal{S}_+^r$

Suppose $P \subseteq \mathbb{R}^d$ is a polytope defined by

$P = \{ x : \langle a_i, x \rangle \leq b_i : i = 1, \ldots, m\}$, and

$P = \mathrm{conv}(\{x_1, \ldots, x_n\})$

The corresponding **slack matrix** $M \in \mathbb{R}^{m \times n}$ is given by $M_{ij} = b_i - \langle a_i, x_j \rangle$

Yannakakis Factorization Theorem

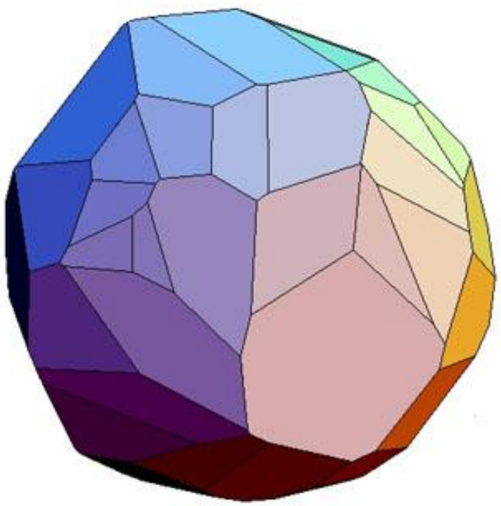
[Yannakakis 1989, Fiorini-Massar-Pokutta-Tiwary-de Wolf, Gouveia-Parrilo-Thomas 2011]:

The minimum size of a polyhedral (resp., PSD) lift for P is precisely $\mathrm{rank}_+(M)$ (resp., $\mathrm{rank}_{\mathrm{psd}}(M)$) for any slack matrix M of P.

polytopes as a set of "theorems"

Yannakakis Factorization Theorem
[Yannakakis 1989, Fiorini-Massar-Pokutta-Tiwary-de Wolf, Gouveia-Parrilo-Thomas 2011]:

The minimum size of a polyhedral (resp., PSD lift) for P is precisely $\mathrm{rank}_+(M)$ (resp., $\mathrm{rank}_{\mathrm{psd}}(M)$) for any slack matrix M of P.



Polytope P is a collection of valid linear inequalities.

Farkas' Lemma: Every valid linear inequality is a **conic** combination of defining inequalities.

Polyhedral lift size: Smallest set of "axioms" that generate all valid inequalities for P. **But we are allowed to use auxiliary variables.**

For the rest of the talk, restriction attention to the polytope $\mathrm{CUT}_n \subseteq \mathbb{R}^{\binom{n}{2}}$:
Convex hull of cuts in the complete graph.

valid linear inequalities for CUT_n

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic polynomial such that the restriction $f|_{\{0,1\}^n}$ is non-negative.

For instance: $f(x) = (x_1 + x_2 + \cdots + x_n - 1)^2$ or

$$f(x) = \left(x_1 + x_2 + \cdots + x_n - \frac{n}{2}\right)^2 - \frac{1}{4} \quad \text{for } n \text{ odd}$$

Then the inequality $\{ f(x) \geq 0 : x \in \mathbb{R}^n \}$ is a valid "linear" inequality, and every valid inequality is of this form.

[Can consider this as a linear inequality in $x \otimes x$.]

$$\mathrm{QML}_+^n = \{ f : \{0,1\}^n \to \mathbb{R}_+ : \deg f \leq 2\}$$

Let Q be a subspace of $L^2(\{0,1\}^n)$, and define

$$\mathrm{sos}(Q) = \mathrm{cone}(q^2 : q \in Q) \subseteq L^2(\{0,1\}^n)$$

minimal size of PSD lift for $\mathrm{CUT}_n \approx \min\{\dim Q : \mathrm{sos}(Q) \supseteq \mathrm{QML}_+^n \}$

minimal size of polyhedral lift for $\mathrm{CUT}_n = \min\{|Q| : \mathrm{sos}(Q) \supseteq \mathrm{QML}_+^n \}$

sums of (low-degree) squares

For $f : \{0,1\}^n \to \mathbb{R}_+$, define: $\deg_{\text{sos}}(f) = \min\{d : f \in \text{sos}(\mathcal{P}_d)\}$
where $\mathcal{P}_d$ is the space of degree $\leq d$ multi-linear polynomials on $\{0,1\}^n$

Easy: $\deg_{\text{sos}}(f) \leq n$. (Trivial case of the Krivine-Stengle Positivstellensatz.)

Fix a function $g: \{0,1\}^m \to \mathbb{R}_+$ and a number $n \geq m$.

For every subset $S \subseteq \{1, \ldots, n\}$ with $|S| = m$, let $g_S(x) = g(x|_S)$.

Let $\mathcal{F}(g,n) = \{ g_S : \{0,1\}^n \to \mathbb{R}_+ : |S| = m \}$.

Theorem: For any $g: \{0,1\}^m \to \mathbb{R}_+$ and $n \geq 2m$,

$$1 + n^d \geq \min\{\dim Q : \mathcal{F}(g,n) \subseteq \text{sos}(Q)\} \geq C\left(\frac{n}{\log n}\right)^{\frac{d-1}{2}}$$

where $d = \deg_{\text{sos}}(g)$ and $C = C(g) > 0$.

a "hard" quadratic function

Main Theorem [L-Raghavendra-Steurer 2015]:
For any $g: \{0,1\}^m \to \mathbb{R}_+$ and $n \geq 2m$,

$$1 + n^d \geq \min\{\dim Q : \mathcal{F}(g,n) \subseteq \mathrm{sos}(Q)\} \geq C\left(\frac{n}{\log n}\right)^{\frac{d-1}{2}}$$

where $d = \deg_{\mathrm{sos}}(g)$ and $C = C(g) > 0$.

Theorem [Grigoriev 2001]:
For every odd integer $m \geq 1$, the function $g : \{0,1\}^m \to \mathbb{R}_+$ given by

$$g(x) = \left(x_1 + \cdots + x_m - \frac{m}{2}\right)^2 - \frac{1}{4}$$

has $\deg_{\mathrm{sos}}(g) \geq \frac{m+1}{2}$.

Corollary: The minimal dimension of a PSD lift of CUT_n grows faster than any polynomial.

proof outline

Let $\mathcal{Q} \subseteq L^2(\{0,1\}^n)$ be a subspace of dimension k.

Pre-processing:

$\mathrm{sos}(\mathcal{Q}) \approx \widetilde{\mathrm{sos}}(\mathcal{Q}')$ for some "well-conditioned" $\mathcal{Q}'$ with $\dim \mathcal{Q} = \dim \mathcal{Q}'$

[John's theorem—factorization through a cone—via Briët-Dadush-Pokutta 2014]

Approximation:

Then $\widetilde{\mathrm{sos}}(\mathcal{Q}') \mathrel{\tilde{\subseteq}} \mathrm{sos}(\mathcal{P}_{O(\log k)})$

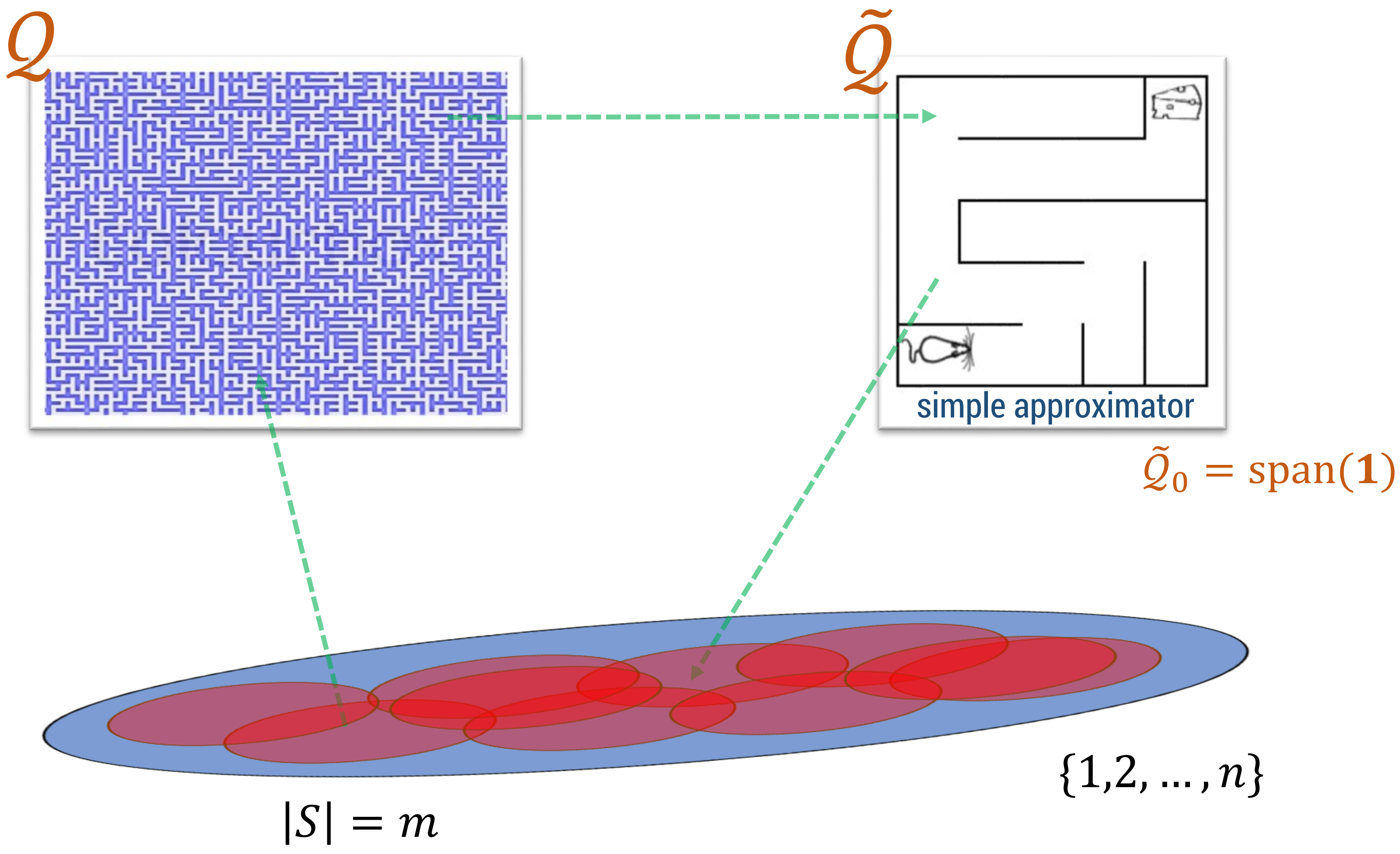
(with respect to a certain class of test functionals)

Degree reduction:

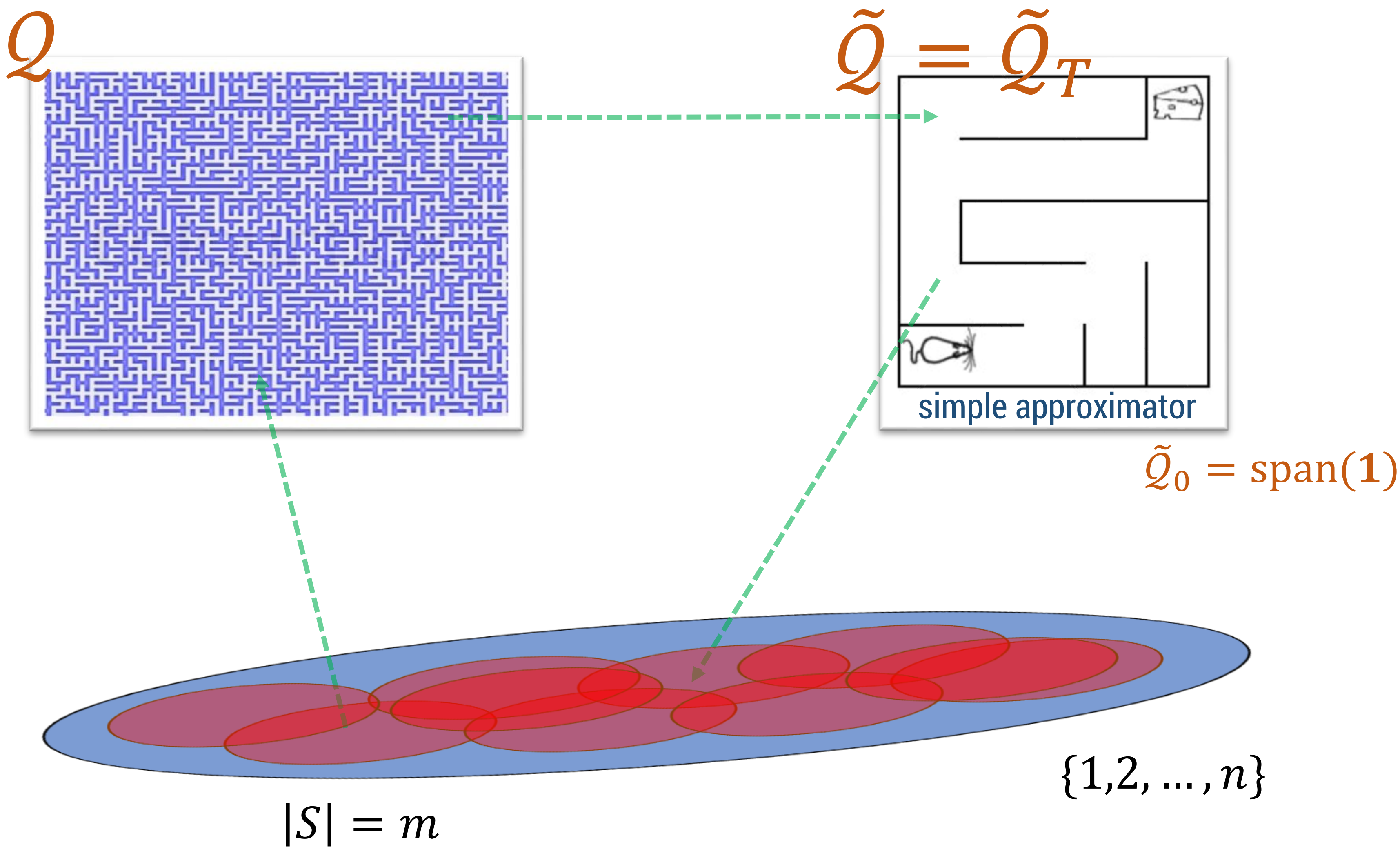
For a uniformly random subset $S \subseteq \{1,2,\ldots,n\}$ with $|S| = m$,

with high probability $\mathrm{sos}(\mathcal{Q}|_S) \mathrel{\tilde{\subseteq}} \mathrm{sos}\left(\mathcal{P}_{\frac{\log k}{\log n}}\right)$

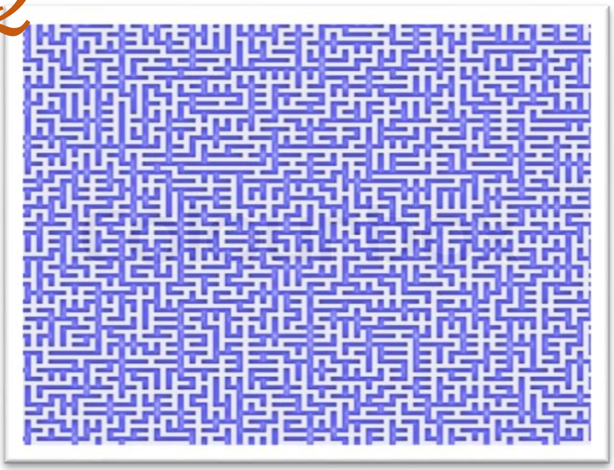
quantum learning



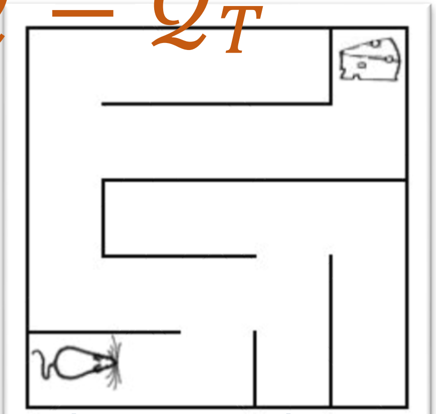
quantum learning



$\mathcal{Q}$



$\tilde{\mathcal{Q}} = \tilde{\mathcal{Q}}_T$



simple approximator

$\tilde{\mathcal{Q}}_0 = \mathrm{span}(\mathbf{1})$

$\dim(\mathcal{Q})$ small

$\Rightarrow$ $\mathcal{Q}$ has small von-Neumann entropy

$\Rightarrow$ learning cannot go on for too long

Theorem:

Consider a map $Q : \{0,1\}^n \to \mathcal{S}_+^k$ with $\mathbb{E}_x \operatorname{Tr}\big(Q(x)\big) = 1$

Want to approximate Q by a simpler map $\tilde{Q}$ (s.t. $\mathbb{E}_x \operatorname{Tr}\left(\tilde{Q}(x)\right) = 1$) with respect to a class $\mathcal{T}$ of test functionals $\Lambda : \{0,1\}^n \to \mathcal{S}^k$:

$$\left|\mathbb{E}_x \operatorname{Tr}\left(\Lambda(x)\left(Q(x) - \tilde{Q}(x)\right)\right)\right| \leq \varepsilon$$

If $\deg(\Lambda) \leq \kappa$ and $\max_x \|\Lambda(x)\| \leq \omega$ for all $\Lambda \in \mathcal{T}$ then there is a function

$R : \{0,1\}^n \to \mathcal{S}_+^k$ with $\mathbb{E}_x \operatorname{Tr}(R(x)^2) = 1$ such that

$$\deg(R) \leq O\left(\frac{\kappa\omega}{\varepsilon}\right) S\big(U_Q \,\|\, \mathcal{U}\big)$$

$$\left|\mathbb{E}_x \operatorname{Tr}\big(\Lambda(x)(Q(x) - R(x)^2)\big)\right| \leq \varepsilon \quad \forall \Lambda \in \mathcal{T}$$

$$U_Q = \mathbb{E}_x \left(e_x e_x^T \otimes Q(x)\right)$$

$$\mathcal{U} = \mathrm{Id}/\operatorname{Tr}(\mathrm{Id})$$

$$S(A \,\|\, B) = \operatorname{Tr}\big(A(\log A - \log B)\big)$$

construction of $\tilde{Q}$

Theorem:

Consider a map $Q : \{0,1\}^n \to \mathcal{S}_+^k$ with $\mathbb{E}_x \operatorname{Tr}\big(Q(x)\big) = 1$

Want to approximate Q by a simpler map $\tilde{Q}$ (s.t. $\mathbb{E}_x \operatorname{Tr}\left(\tilde{Q}(x)\right) = 1$) with respect to a class $\mathcal{T}$ of test functionals $\Lambda : \{0,1\}^n \to \mathcal{S}^k$:

$$\left|\mathbb{E}_x \operatorname{Tr}\left(\Lambda(x)\left(Q(x) - \tilde{Q}(x)\right)\right)\right| \leq \varepsilon$$

minimize $S(U_Q \,\|\, \mathcal{U})$ subject to these constraints

$$U_Q = \mathbb{E}_x\left(e_x e_x^T \otimes Q(x)\right)$$

Optimality dictates: $$Q^*(x) \propto \exp\left(\sum_{\Lambda \in \mathcal{T}} c_\Lambda \, \Lambda(x)\right)$$

If one can achieve $\sum_\Lambda |c_\Lambda|$ small, then the Taylor series of the exponential can be truncated at a low degree.

- Work modulo ideals other than $\{X_i^2 - X_i : i = 1, \dots, n\}$

  Perfect matchings

  Cliques in $G(n, p)$

- Sums of squares of sparse polynomials:

  Let $\mathcal{V}_s$ be the **subset** of multi-linear polynomials on $\{0,1\}^n$ that have at most s non-zero monomials.

  Can one find an explicit $f : \{0,1\}^n \to \mathbb{R}_+$ such that $f \notin \mathrm{sos}(\mathcal{V}_s)$?