6.S997, Lecture 2

SoS Overview

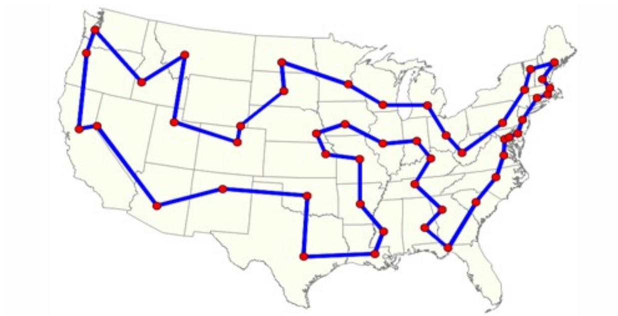
S. Hopkins

SoS is an *algorithmic toolkit* for solving *systems of polynomial inequalities.*

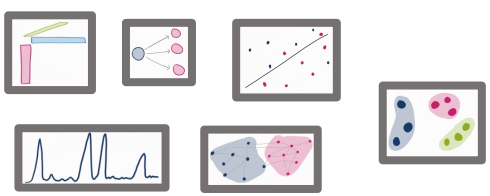
$$p_1(x_1, \dots, x_n) \geq 0, \dots p_m(x_1, \dots, x_n) \geq 0$$

Why polynomials?

Polynomials are ***extremely expressive***



Combinatorial optimization



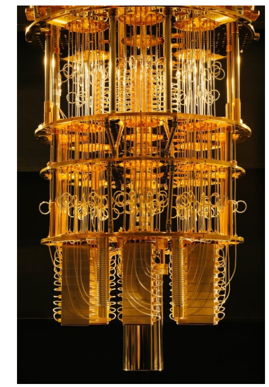
Statistics

Proofs to algorithms

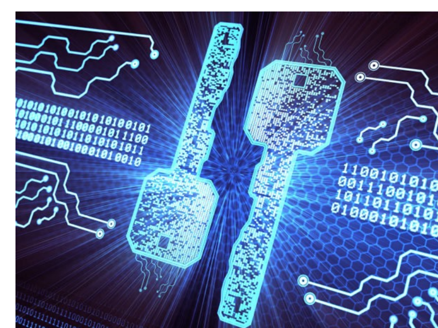
Polynomial Systems & Sum of Squares Method

convex programs subsuming LP, SDP, spectral methods

Robotics/optimal control



Quantum information



Cryptography


CAG
300

A (brief and opinionated) history

Early 1900s: Hilbert investigates relationship between nonnegative polynomials and squares.

1950s: Invention of linear programming

1960s: Krivine & Stengle prove that every nonnegative polynomial over a semialgebraic set can be certified nonnegative by an SoS proof

1970s: Ellipsoid method – convex programming in P

1987: Shor proposes precursor to SoS method, relating polynomial system solving to semidefinite/convex programming

1990s/2000s: LP, SDP, eigenvalue methods extensively investigated in theoretical computer science & optimization

2000s: Lasserre proposes “pseudoexpectation SDP” and Parrilo independently proposes “SoS proof SDP”.

2010s: SoS as a unifying view on LP, SDP, spectral algorithms, + extensive new applications

This course

Goal 1: familiarize you with SoS language and tools for theoretical analysis (no programming)

Goal 2: enable you to see possible uses of SoS in your own research (course project!)

Goal 3: see some beautiful algorithms

This course

TCS perspective:

qualitative: polynomial running times, large *n*

quantitative: accuracy guarantees

SoS as a high-level programming language for algorithm design

(We won't worry about "compiling down" to LP/SDP. And we won't worry about using the most lightweight algorithms possible – "just import all the libraries")

Prerequisites

Linear algebra, at the level of last lecture

matrices, eigenvalues, eigenvectors, quadratic forms, Cholesky decomposition

Probability & (today) Information Theory

every true fact about a constant-dimensional random variable is "trivial"

This course

We will cover some subset of:

Worst-case approximation algorithms (max-cut, last week)

Approx. algorithms for “structured” instances (today)

Algorithms for *random* instances (probably next week)

Statistical inference & robust statistics

Differentially private algorithms via SoS

”Fast” implementations of SoS

SoS view on computational complexity

Other topics?

Let’s review

Hypercube basics

Every $f : \{0,1\}^n \rightarrow \mathbb{R}$ can be uniquely represented as a multilinear polynomial

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot x^S$$

SoS Proof
(of nonnegativity on the hypercube)

$\vdash_d f \geq 0$: $f(x) = \sum_{i \leq n^d} p_i(x)^2$ for all $x \in \{0,1\}^n$
with $\deg p_i \leq d$

search for proofs in $n^{O(d)}$ time via SDP
(matrix representation of proofs)

every nonnegative f has $\vdash_{O(n)} f \geq 0$
for all f, $\vdash_{\deg f} f + \sum |\hat{f}(S)| \geq 0$

Pseudoexpectations

$\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$ which is:

(1) Linear

(2) Respects $x_i^2 = x_i$: for all S, $\widetilde{\mathbb{E}}\left[x^S x_i^2\right] = \widetilde{\mathbb{E}}[x^S]$

(3) Positive: $\widetilde{\mathbb{E}}[p^2] \geq 0$

(4) Normalized: $\widetilde{\mathbb{E}}[1] = 1$

represent as numbers $\left\{\widetilde{\mathbb{E}}[x^S]\right\}_{|S| \leq d}$

search for pseudoexpectation with $\widetilde{\mathbb{E}}[f] < 0$ in time $n^{O(d)}$

Very useful intuition: $\widetilde{\mathbb{E}}$ represents low-degree moments of distribution on the hypercube

(even though it doesn't…)

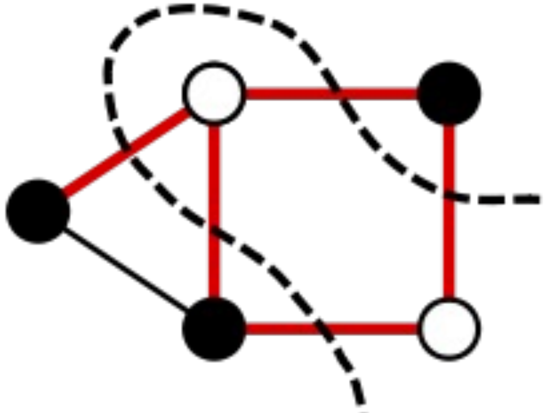
Duality

For every $f, d \geq \deg f$ (even), exactly one holds:

(1) $\vdash_d f \geq 0$

(2) Exists degree-d pseudoexpectation $\widetilde{\mathbb{E}}[f] < 0$

Max-Cut



$$G = (V, E)$$

Let $G(x) = \sum_{i \sim j} (x_i - x_j)^2$ = number of edges cut by x

Thm: for every G, $\vdash_2 G(x) \leq \frac{1}{0.878} \cdot \max_y G(y)$

Proof by rounding any $\tilde{\mathbb{E}}$ s.t. $\tilde{\mathbb{E}}[G] \geq \alpha$ to some y s.t. $G(y) \geq 0.878\, \alpha$.

(Also leads to algorithm for finding y)

Proof by rounding any $\tilde{\mathbb{E}}$ s.t. $\tilde{\mathbb{E}}[G] \geq \alpha$ to some y s.t. $G(y) \geq 0.878\,\alpha$.

Key idea: sample from Gaussian on $\mathbb{R}^n$ which has same mean and covariance as $\tilde{\mathbb{E}}$

Same key idea gives approximation algorithms for:

-- $\max_x x^\top A x$ for $A \succcurlyeq 0$

-- $\max_{x,y} x^\top A y$ ("cut norm"/Grothendieck)

and forms the basis for the best-known approximation algorithms for graph expansion

(Arora-Rao-Vazirani)