IPCO 2019 Summer School

David Hershkowitz

May 29, 2019

These are the notes I took while at the IPCO 2019 summer school. These notes were written while trying to keep up with the talks and so are not free from errors. Cheers!

Contents

1 An Introduction to Semidefinite Programming for Combinatorial Optimization (Sam Burer)

1.1 Definitions

- Looking at symmetric matrices: ensures real eigenvalues
- Trace Inner product $M \cdot N$: multiply corresponding entries of matrices and sum up
- Induced Frobenius norm: $||M||_F := \sqrt{M \cdot M}$

Matrix is PSD iff

- $v^T X v \geq 0$ for every $v \in \mathbb{R}^p$
- $\lambda_{min} \geq 0$
- X = VV^T for some V: get to choose number of columns in V; number of columns, roughly speaking, controls the rank of the matrix
- Every principal submatrix has determinant ≥ 0

PSD matrices is a convex set of all symmetric matrices

$A \succeq 0$ means PSD

PSD matrices is a "proper" cone: i.e. closed, convex, pointed, full-dimenstional

Self-dual, meaning if you take the PSD cone and take all matrices with positive inner product (as defined above) you get back all PSD matrices

Question

What is the dimension of S^p_+?

Ambient dimesnion is p^2; but symmetry takes away $\binom{p}{2}$ degrees of freedom so dimension is $\binom{p+1}{2}$

Question

A diagonal matrix is PSD if entries are non-negative

Spectral Decomposition Theorem

$x \succeq$ iff there is an orthogonal V and non-negative diagonal D s.t. $X = VDV^T$

I.e. if you do the spectral decomposition can look at the diagonal matrix and see if it is PSD

Will assume LP as a background henceforth

1.2 Optimization

Given

Dimension n and m

Cost matrix C (must be symmetric), constraint matrices $A_1, \ldots, D_m$ and right side $b \in \mathbb{R}^m$

Let $X \in S^n$ denote our variable (we are optimizing over matrices)

Primal SDP problem

$P^* := infC \cdot X$ subject to $A_i \cdot X = b_i$ and X is PSD; like LP over a different type of cone

Gave an example of such a program and optimized it; point is that can reduce it to determinant conditions and then it looks very quadratic

LP is a special case of SDP

Write down a standard form LP and then can turn it into an SDP by applying $Diag$ operator that turns vector into a diagonal matrix

Second-Order Cone Programming is a Special case of SDP

SOCP is based on $||x|| \leq t$; this condition can be written as an SDP

So $LP \subseteq SOCP \subseteq SDP$

Dual SDP Problem

$d^* := \sup b^T y$ s.t. $C - \sum_i y_i A_i \succeq 0$

Did out dual of earlier problem

Weak Duality Theorem for SDP

The proof is the same as the LP duality theorem

As a corollary if a dual solution equal to value of primal solution then get that both are optimal

Also get complementary slackness-like condition

Also get $p^* \geq d^*$

Strong Duality

Don't always get strong duality; gives an example where it's not true (also true for SOCP)

So can ask under what coniditions does strong duality hold

- For that look at positive definite; definitions are as before in PSD but now inequalities are strict; in Cholesky decomposition the V must be invertible; denote it S^p_{++}

Suppose you have a feasible primal and dual; if there is a positive definite primal solution then you get strong duality

Remark

Algorithmic papers and results often assume both (P) and (D) have interior; but in any given application should double check

1.3 Algorithms

Cannot expect to solve an LP exactly or SDPs exactly

Gives example of an SDP where optimal solution is irrational despite rational input

Hence, for a user-specified $\epsilon > 0$, a reasonable goal is to find an ϵ-approximate solution

Going to focus on this goal for the dual

I.e. will grow feasible set where need to find a solution such that

ϵ-feasible: ϵ wiggle room for constraints

ϵ-optimal: optimal solution within ϵ

Ellipsoid Method

Given

$\epsilon > 0$

v vector encording

$\sigma =$ length of v and $\Sigma :=?$ measure length of problem

Optimize dual s.t. add two norm constraint on y: $||y||_2 \leq \Sigma^\sigma$

Then introduce ϵ-feasibility relaxation to get dual (D_ϵ) whose optimal value is ϵ-optimal

Define parameter $\theta \in (0, 1)$ such that $\theta \to 0$

If problem (D) is feasible then feasible set of our relaxation. . .

Theorem: ellipsoid method works with $O(m^2 \log(1/\theta))$ iterations where each iteration needs poly(n, m) FLOPs; roughly $\log(1/\epsilon)$

Interior-Point Methods

This is what you'll find in code that you download

Basic idea is to move from interior of set towards an optimal solution

Given

$\epsilon > 0$

Same measures of size of problem

Will also assume both primal and dual are interior feasible which implies strong duality and the "central path" exists

Initial interior primal-dual solution

Theorem: primal-dual shortstep method needs poly iterations and $\log(1/\epsilon)$

Compared to ellipsoid the number of iterations is less (though the cost per iteration is roughly the same)

Software

Only 1 commercial package for SDPs: Mosek; free academic licenses

Other Points

Interior point methods slow down for large instances; many methods to address this often by exploiting problem structure

E.g. low rank methods for SDPs solve SDPs by solving them as non-linear problems E.g. if there is an optimal solution then there is an optimal solution whose rank is bounded by about $\sqrt{m}$

Worked well in practice; surprising because took something non-convex then turned it into something convex

1.4 Classic Applications

Given a quadratic program can create an SDP relaxation of it (in a way similar to how on does this for LPs); the Shore relaxation

Max Stable Set

Undirected graph, α is max size of a stable set (NP-hard to calculate α), $\bar{\chi}$ is the min size of a clique cover (also NP-hard)

Easy to show $\alpha \leq \bar{\chi}$

E.g. for pentagon have $\alpha = 2$ and $\chi = 3$

So do IP formulation of max stable set problem and then turn it into a quadratic program (wouldn't normally think to do this); take SDP relaxation and you get Lovasz's θ number as the optimal solution

Have $\alpha \leq \theta \leq \bar{\chi}$

We don't really know how to compute α or $\bar{\chi}$ but we can compute θ

If think about perfect graphs (where $\alpha = \bar{\chi}$) we then can compute these both in poly time

Max Cut

Want to find a subset of vertices with maximum number of edges leaving

NP-hard

Can formulate it as a quadratic problem; name a variable for each vertex where -1 means in one subset and 1 means its in the other and then sum over edges $\frac{1}{2}(1 - x_i x_j)$; take SDP relaxation to get Goemans-Williams relaxation

GW gave a .87856-approximation algorithm for max cut

1. Solve SDP to obtain X^*
2. Compute a factorization $X^* = R^*(R^*)^T$
3. Randomly generate a vector uniform on the unit sphere
4. Round according to inner product with random vector

Lift and Project

Suppose a polyhedron contained in the 01 box

Intersect continuous relaxation with convex hull to get P^{01}, what can we say about P^{01}

Homogenization of P: introduces a variable x_0 which asserts $x_0 \geq 0$ and $AX \leq x_0 b$: pulls original polytope into higher dimension

Observation: $x \in P$ implies $(x_j, x_j x)$ is in the homogenization

Idea: linearize $x_j x$ by introducing a new variable y

The lifting procedure introduces this y and adds constraints in this higher dimension for some variable index j

Then project this lifted set back onto the original variable set

Egon + Gerard et al. show that this gets us closer to the convex hull: $P^{01} \subseteq BCC(P, j) \subseteq P$

Can rinse and repeat this: if do it up to nth level just get back the integral hull

Observation: the iterative BCC procedure depends on the order of the variables in the lifting: so an idea is to lift w.r.t. all the variables and then take the intersection

This operator is called $N_0(P)$; unfortunately this takes n steps in general to get back to integral hull

Can put this all into a matrix: get Lovasz-Schrijver hierarchy if you iterate this

Also an example where need to do all n operations to get to convex hull

—End of first day—

1.5 How to Convexify Nonconvex QP

In integer programming want to minimize linear objective subject to being in some feasible set. Key question is can we figure out what the convex hull of those integer points are.

Now imagine you're dealing with QPs. What is the analagous idea? Can rewrite the objective in terms of slightly larger border matrices.

But then what you can do is think about linearizing the matrix and the objective where lift border matrix into a higher feasible set.

Then can convexify in the feasible set.

Will let HG (holy grail) be this closed convex set.

HG is very non polyhedral; e.g. simple 2 dimensional example given.

Another example is if first variable is a scalar that is integral and you look at that variable square; taking convex hull of that set you get everything above a segmented parabola

This approach is not necessarily practical

- Characterizing HG is hard
- May make sense to just convexify over variables
- In fact, might be better to lift to a larger set of variables

Observation

Fully characterizing HG is equivalent to identifying "copositive" matrices over F

Three Types of Valid Inequalities

1. Explicit Quadratics

   Set F is defined by a quadrative inequality

2. Linear Conjunctions

   Can multiply two linear inequalities together which can be linearized

3. Linear Disjunctions

   Multiply linear inequalities but take opposite sign

1.6 Continuous Case

1. Unconstrained

   $F = \mathbb{R}^n$

   HG in this case is all PSD matrices

2. Linear Equations

   Need two types of inequalites: all **disjunctions** from PSD disjunctions and all **conjunctions** of the form $a_i^T x = b_i$ and $a_j^T x = b_j$

3. Nonnegative Orthant

   First case where it gets hard

   Don't know a full description of HG in this case

   DNN := doubly nonnegative; a first approximation to HG for this case

   Theorem: $HG = DNN$ iff $n \leq 3$

   HG is the dual cone of the copositive matrices

4. Nonnegative Orthant, Linear Equations, and Complementarities

   For HG in this case need copositivity; wouldn't necesarilly expect it in this case

5. Linear Inequalities

   Could also make equalities with slack variables; but wants to show how it looks without slack variables

   E.g. consider simple case of a box in two dimensions

6. Half-Ellipsoid

   To get HG need to use second-order cone programming (socp)

7. Swiss cheese case

   Intersection of ball with cuts with holes in it

   If none of the cuts and holes touch each other then HG is PSD $\cap$ RLT $\cap$ SOCRLT $\cap$$ another constraint

1.7 Mixed Binary Case

Copositivity is very important: non=negativity, linear equaitons, complmenetarity equations but also add a subset of variables is binary

HG hinges on copositivity in this case

1.8 Mixed Integer Case

1.8.1 Integer Lattice

Uses "split disjunctions"

HG in integer lattice is contained in these splits which are properly contained in PSD

1.8.2 Noninteger Lattice

Use "non-standard" splits

1.9 Final Thoughts

There is still a lot to do, especially for the integer case, for nonconvex QP

Convexification involves aspects of

- convex analysis
- polyhedral theory
- SDPs
- polynomial optimization

2 Discrepancy and Combinatorial Optimization (Nikhil Bansal)

Notes at: https://www.win.tue.nl/~nikhil/ipco1.pdf

Discrepancy Theory is a 100 year old field of mathematics

Originated with wanting to approximate something continous with something discrete

Will focus on a combinatorial aspect of things

2.1 Combinatorial Discrepancy

Universe of elements $U = [n]$ and subsets $S_1, \ldots, S_m$

Want to color elements red and blue so S_i is as "balanced as possible"

Let $x \in \{-1, 1\}^n$ be a coloring

$disc(S, x) = \max_i disc(S_i, x)$ where $disc(S_i, x) = |\sum_{i:x_i \in S_i} x_i|$

And so can define discrepancy of set system S as $disc(S) = \min_x disc(S, x)$

Can view incidence matrix and then you're trying to minimize the row sum

So for an arbitrary matix A can define $disc(A) = \min_{x \in \{-1,1\}^n} |Ax|_\infty$

2.2 Applications

Rounding: Thm[Lovasz, Spencer, V. '86]

Given Axb and $x \in \mathbb{R}^n$ can round x to $\tilde{x}$ such that $|x - \tilde{x}|_\infty \leq 1$

So that $|Ax - A\tilde{x}|_\infty \leq$ hereditary discrepancy(A)

What is the hereditary discrepancy? A more robust definition of discrepancy

$herdisc(A) = \max_{S \subseteq [n]} disc(A_{|s})$

proof: suppose $frac(x_i) \in \{0, 1/2\}$ then round according to the best coloring given by the hereditary discrepancy; namely letting s be all half integral variables

can then do this bit by bit

thm [rothvoss]: for bin packing: got OPT $+ \log$ OPT using discrepancy theory

steinitz problem: $v_1, \ldots, v - n \in \mathbb{r}^d$ such that $\sum v_i = \vec{0}$ and $|v_i| \leq 1$

question is does there exist a permutation π such that $\max_{k \in [n]} |v_{\pi(1)} + v_{\pi(2)} + \ldots| \leq ??$

steniitz showed you can get an upper bound of $2d$ for any norm

an application of papadimitriou to rounding

what does this have to do with discrepancy?

define discrepancy for prefix sums as follows

$v_1, \ldots, v_n$ ordered. find signs $x_1, \ldots, x_n \in \{-1, 1\}$ such that $\max_w |\sum_{i=1}^{w} x_i v_i|$ is minmized; let $f(d)$ be this result

theorem: steinitz constant $\leq f(d)$

proof idea: start with some ordering for steinitz problem. can create a new ordering π' such that $S(\pi') \leq (S(\pi) + f(d))/2$ so can repeatedely reorder until get down to $f(d)$.

Numerical integration example. Suppose have some weird shape. Number of points in a rectangle is roughly the same as the area.

How is this related to discrepancy? Start with a dense placement of points; trivially try that for any rectangle the number of points proportional to the area. Now want to color points so that have roughly equal number of pluses and minuses in every rectangle; this lets you refine your number of points down to fewer points: i.e. get rid of minuses and double weight of pluses. Using discrepancy to iteratively sparsify your system.

2.3 Techniques

1. Linear algebraic: developed in last 100 years
2. Partial coloring: developed in the '80s
3. Convex geometry: late 90's

A lot of work in about 2010 in making 2+3 algorithmic using SDPs and random walks

Today will look at first 2

2.3.1 Linear Algebraic Method

Start with coloring and will iteratively round

Beck and Fiala showed that if each element lies in at most k sets then the discrepancy is at most $2k - 1$ using linear algebraic methods

Idea is as follows. Consider a set of fractional variables. Call a set S_i dangerous if the number of 1s of S_i restricted to floating elements is larger than k

Will look at matrix B of dangerous sets and will solve for $By = 0$. Need to show this system has a non-zero solution: e.g. show that the number of rows is larger than the number of floating elements.

Claim: the number of rows in $B < f$ because total number of 1s is at most fk (because of above sparsity assumptions). Update now some a new 1.

When a set is dangerous it has discrepancy 0. Once no longer dangerous has at most k variables that go from in worst case $-.99$ to 1

We say Steinitz $\leq$ disc of prefix sums

Thm: disc. of prefix sums $\leq 2d$

Proof: Just need to start with 0 coloring and give a way to update it so as to not mess up prefix colorings. Look at first $d+1$ prefixes and let this be your B matrix.

Claim: disc. norm of any prefix $\leq d$

Stenitz l_∞-norm. LA method gives $O(d)$. Best we know is $O(\sqrt{D}\sqrt{\log(n)})$

2.3.2 Partial Coloring Technique

Developed in the '80s by Beck, Spencer and independently by Gluskin (whose proof we'll see)

Go back to set system view of things: could try and randomly color each element with probability $1/2$. For a paritucular set S in expectation it has discrepancy 0 How bad can this be? $S : \{1, \ldots, n\}$. Discrepancy will be $\sqrt{n}$. S is roughly Gaussian with standard deviation $\sqrt{n}$.

Example: Suppose given a set system with n sets and n elements. One can give simple examples which show that discrepancy is at least $1/2\sqrt{n}$.

If you do a random coloring can get $O(\sqrt{n}\sqrt{\log n})$ discrepancy. Look at a fixed set S_i. $\Pr(disc(S_i) \geq c\sqrt{n}) \leq e^{-c^2/2}$. So set c to be $\sqrt{\log n}$ and union bound over all sets.

Spencer shows that discrepancy of this set system is $6\sqrt{n}$; he invented the partial coloring method to show this.

Suppose want discrepancy of Δ_S for set S

Stated a more general union bound result.

Partial Coloring Lemma: Δ_i for set S_i If Δ_i satisfy some conidition that looks like what you need but slacker (given in the notes) then there exists a partial coloring that colors at least $n/10$ elements s.t. $disc(S_i) \leq \Delta_i$

E.g. Partial coloring gives you something when $\Delta_s = n^{1/4}$ whereas union bound doesn't.

A particular example

Beck Fiala Problem Thm: can show discrepancy is at most $\sqrt{k}\log n$ if every element lies in at most k sets using partial coloring. Proof given in notes in Theorem 4.4.

Main idea of proof of partical coloring lemma

Will take a geometric view of discrepancy. Define polytope P of all colorings that satisfy the discrepancy of every row is at most Δ. Thus, the discrepancy of A is just the smallest Δ such that some integral $\{-1,1\}^n$ point is in P.

Thm [Gluskin]: given a symmetric body $K \in \mathbb{R}^n$ with $\gamma_n(K) \geq 2^{-n/5}$ then it contains an $x \in \{-1,0,1\}^n$ with at least $n/10$ non-zero coordinates. γ is the gaussian measure, related to the volume of the body. Roughly the measure condition is saying that your convex body needs to be at least as large as the unit cube; this is where all the integral points lie so this makes sense intuitively.

Proof sketch: Consider taking our body and placing it at every point in the integral lattice. A simple calculation shows that if you add up the measures of all the shifted bodies is more than $2^n/2$. So there must be some y in K_x and $K_{x'}$ (two different shifts of polytope). x and x' must differ in at least $n/10$ coordinates. Get \$x-x' = k = k'\$for $y = x + k = x' + k'$ for $k, k' \in K$ and so $(x - x')/2 \in K$. And so if look at $(x - x')/2$ it will be in $\{-1, 1, 0\}^n$.

—End of first day of talks—

Gave a recap of first day

2.4 Recap of first day

2.4.1 Hereditary Discrepancy

Discrepancy useful measure of complexity of a set system but not robust so people look at hereditary discrepancy

Usually techniques for bounding discrepancy translate to hereditary discrepancy

2.4.2 Rounding

Application of Lovasz that bounds loss in feasability in terms of hereditary discrepancy

Rothvoss has an application for bin packing

Hereditary discrepancy is 1 iff matrix A is TU

Rounding proof of Lovasz is not constructive; can find it constructively with a $O(\sqrt{\log m \log n})$ loss in error

2.4.3 Ordering with Small Prefix

Quickly reviewed

2.4.4 Steinitz Problem

Can rearrange permutation to bring down max discrepancy

2.4.5 Sparsification

How well can you approximate a region by discrete points?

Randomly sample points—called Monte Carlo integration. Sparsify using discrepancy to do so with fewer points.

2.4.6 Tusnady's Problem

Not mentioned yesterday.

Given a grid with n points and want to color points so that any axis parallel rectangle has low discrepancy.

Random gives about $O(\sqrt{n}\sqrt{\log n})$

Long line of work getting it down to $O(\log^1 .5n)$

2.5 Questions Around Discrepancy Bounds

2.5.1 Combinatorial

Want to show good colorings exist

1. Linear algebra (yesterday) Start with all 0 colorings, update one step at a time. If a variable reaches -1 or 1, fix it forever. Find an update vector by solving some linear system . Cleverness of method is in choosing linear system.

   Beck Fiala has B in the linear system be all rows with size $> k$. Then force such rows to have discrepancy 0 but once thay have $\leq k$ size you don't control them at all.

2. Partial coloring (yesterday) Invented by Spencer to solve Erdos problem.

   Random coloring gives $O(\sqrt{n \log n})$

   Discrepancy is all about beating a random coloring because you can always do this

   Can always demand discrepancy 0 for about $\Omega(n)$ rows with partial coloring lemma

   Combines strength of probability and linear algebra

   (a) Proving Partial Coloring Lemma

      Started looking at this geometrically yesterday

      Consider polytope of all fractional colorings which achieve discerpancy t. Question is how large does t have to be so polytope contains an integral point.

      Gluskin showed a symmetric, convex body with large Gaussian colume contains a point with many integral points

      The guassian volume of a cube is about 2^{-n} in n dimensions

      Sketched a proof that looks at all 2^n shifts of our polytope. Volume computation shows that have exponential volume across all shifted polytopes. So some point lies in multiple copies. Uses this to come up with integral points with large hamming distance.

      Will now get partial coloring lemma from this lemma

      So question is basically for which choice of Δ_is is the Gaussian volume of our polytope large

      Sidak's theorem: symmetrix convex body K intersected with a slab has Gaussian measure larger than their product. (In fact, Royen'14 showed its true for any two convex bodies, not just one and a slab)

      Applying this inductively shows that Gaussian measure of our polytope is just the product of the Gaussian measure of a bunch of slabs

      So now just need to understand what the measure of slab looks like

      Lemma: $\gamma_n(Slab) = \exp(-g(\lambda))$ This can be shown by bucketing a Gaussian

      But for some problems like Beck Fiala you lose a $\log n$

      Iterating loses a $O(\log n)$ usually

      Partial coloring lemma is tight for case where $\Delta = \sqrt{k}$ where every set has size k

3. Banaszczyk'98

   Produces full coloring directly; uses deep convex geometric result

   Gave a brief history for Tusnady, Steinitz and Beck-Fiala problems: 1 then 2 then 3 each give better results

2.5.2 Algorithmic

Find coloring in poly time

Partial coloring is now constructive

- Bansal'10: SDP + random walk
- Lovett Neja'12: Random walk + linear algebra
- Rothvoss'14: sample and project (geometric)
- Many others now

Tailor made algorithm for Beck-Fiala

General Banaszczyk algorithm as of last year

1. Algorithmic Partial Coloring

   Useful view

   Think of independent random coloring as a random walk in the $\{-1, +1\}^n$ cube

   If hit a face, then stick to that fact and stay there

   Easy to convince yourself that the result is completely random

   With partial coloring we wanted to find a coloring that satisfied discrepancy inequalities. For each such inequality there is a slab. Could use these slabs as a barrier

   Lovett Meka's algorithm

   Random walk where in each coordinate change according to tiny Gaussian a) Fix j if $x_j = \pm 1$ b) If row a_i gets tight; move in subspace so as to not violate discrepancy

   So really just need to show that if you choose the right \$λi\$s you will eventually reach a point of all integral points

   Idea to show this progress: walk makes progress as long as dimension is $\Omega(n)$. It's good if you hit an integral constraint but bad if you hit a slab constraint; so you don't want to hit too many slab constraints.

   After $\frac{10}{\gamma^2}$ steps: $\Omega(n)$ variables must have hit a ± 1. On the other hand, a slab gets tight with $\exp(-\lambda_i^2)$ probability which is small enough for our purposes.

   Another Algorithm

   Rothvoss algorithm for general convex bodies

   Pick a random y, return closest point x in $K \cap [-1, 1]^n$

   Idea: measure concentration

   Another simple algorithm by Eldan, Singh'14

2. Banaszczyk

   Annoying loss of $O(\log n)$ to Use Partial Coloring to Get Full Coloring

   Ideal case is that at most $n/10$ big $(> 10k)$ sets

2.5.3 Approximating Discrepancy / Lower Bounds

Vector Discrepancy

Can give an integer program for vector discrepancy. But in LP relaxation can set all x to 0.

So move to SDP. But this discrepancy problem is hard to approximate; namely to distinguish up to a $\Omega(\sqrt{n})$ factor. So SDPs shouldn't help.

But SDPs can still be useful. Can define **hereditary vector discrepancy** of matrix A. Can show that a poly time algorithm with bound on hereditary discrepancy in terms of hereditary vector discrepancy.

Algorithm

Sticky random walk

At each step of walk, formulate SDP on unfixed variables.

SDP is feasible

Gaussian rounding -> step of walk

Properties of walk: high variance -> quick convergance

Low variance for discrepancy on sets -> low discrepancy

Approximating Herdisc

Discrepancy hard to approximate, but can we approximate hereditary discrepancy

Not even clear hereditary discrepancy is in NP because what would the witness be?

Matousek gave a lower bound in erms of the "determinant lower bound"

But wasn't clear what the upper bound is. Matousek showed this was tight up to logs. Idea: is to use SDP duality.

But still no approximation because not clear how to compute determinant lower bound

But recently can approximate

3 Sums of Squares in Polynomial Optimization (Joao Gouveia)

3.1 Unconstrainted Polynomial Optimization and Nonnegativity

Given a polynomial p, minimize it over the reals.

Even simpler what if p is univariate?

Almost equivalent to detecting real roots; somewhat easy but not trivial.

Sturm's Sequence

Given univariate poynomial p of degree d, the Sturm sequence is $P_0 = P, P_1 = P_0'$ and P_{i+1} is negative remainder of P_i divided by P_{i-1}

Sturm's Theorem: to locate roots using this sequence take Sturm sequence and evaluate it at any point χ, now count the number of times the sign changes in this sequence to get w. The number of zeros in interval $(a, b]$ is $w(a) - w(b)$.

Can use this to certify if a polynomial has no roots; i.e. $w(+\infty) = w(-\infty)$ and $p(0) > 0$. This is called a "certificate of positivity".

Observations

1. We could locate zeros of the derivative
2. We could search for the maximal λ such that $p - \lambda$ has no real roots
3. In practice, fastest way is using "Descartes rule of signs"

The General Case

Theorem: for polynomials of degree 4 deciding if $p^* = 0$ is NP-hard

Proof: for instance partition problem of breaking a set of integeres into two equal sum sets is the same as asking if a polynomial has a min of 0.

For odd degree polynomials $p^* = -\infty$ and for degree 2 solve a linear system and check the Hessian.

So this is a hard problem so why are we trying to solve this?

Classic problem: Distance Graph Realization Problem.

Have a graph. Have some distance information d_{ij} for $\{i, j\} \in E$. Want to place points n in $\mathbb{R}^k$ that realize these distances approximately.

Can express this as a polynomial.

Independence Number of a Graph

If want to check if independence number is $\geq t$ then can also check if some polynomial is non-negative

Re-Framing the POP

Let $P[x]$ be all nonnegative polynomials. Thus, could just ask for $p(x) - \lambda$ to be non-negative

Advantages: certificates of non-negativeity imply optimization schemes.

3.2 Sum of Squares and Nonnegativity

A simple ceritifcate for nonnegativity of a polynomial is being a sum of squares. i.e.

$p(x) = \sum_i (h_i(x))^2$

Let $\Sigma[x]$ be all such polynomials.

Univariate Polynomials Revisited

Proposition: a univariate polynomial is nonnegative iff it is sos

Proof: any real root of a nonegative polynomial must have even multiplicity as the zero must be a local minimum. Then can just apply fundamental theorem of algebra.

Motzkin's Example

But e.g. Motzkin's example is a nonegative polynomial in two variables but is not a sum of squares

Proof that it is nonnegative follows by AM-GM

To show that it's not SOS we need some more stuff.

The Newton Polytope of a polynomial $N(p)$ is the convex hull of the vectors of exponents of the monomials of p (1 vector for each term)

Lemma: if $p = \sum_i h_i^2$ then $N(p)$ is the convex hull of $N(h_i^2)$. In particular $2N(h_i) \subseteq N(p)$. Gave proof sketch based on contradicting the fact that a vertex is a vertex (i.e. a convexity proof).

Hilbert's Theorem

When do SOS work perfectly?

Hilbert's Theorem: sums of squares are the same as non-zero polynomials for univariate, quadratic and bivariate quartic ($n = 2$ and $d = 2$) polynomials.

In fact, an iff and only if: Blekherman shows that the volue of the SOS divided by the non-negative polynomials goes to 0: almost no nonnegative polynomials are SOSs

Summary: certify nonnegativity to optimize. SOS is a way of certifying nonegativity

The case for SOS

SOS is simple and very versatile

Gives a short proof of nonnegativity

They can be "strengthened"

We can find them "efficiently"

3.3 Sums of Squares and Semidefinite Programming

Quick summary of SDPs

Application: quadratic polynomial is nonegative iff it is SOS

Can write any quadratic q as nonegative as a matrix Q quadraticed with a vector having Q PSD. But then can use Gram decomp on Q to write the polynomial as a sum of squares. The reverse direction is the same.

The second part of this idea still works for general sums of squares

Proposition: let x_d be the vector of all monomials of degree at most degree d. p is SOS iff there is a PSD matrix Q such that $x_d^T Q x_d = p$. Gave proof.

Rank of Q will correspond to the number of squares, though Q is not unique.

So now can try and determine if p is SOS by writing an SDP

But complexity grows with n^d because n^d monomialsish

SOS Relaxation

Now instead of looking for largest λ to subtract and keep non-negative can look for largest λ to subtract and be SOS

Another example to show details. Found λ but didn't necesarilly find minimizer

Gave some libararies for SOS and SDPs

3.4 Strengthening Sums of Squares

So SOS can fail badly for certifying certain polynomials.

Can try a stronger certificate

Theorem: Every nonnegative polynomial can be written as a SOS of rational functions

In other words, if $p(x) \geq 0$ there are h_i and g_i such that $p(x) \sum_i (h_i/g_i)^2$

Finding such a certificate is not easy in general though

Idea: consider a uniform denominator instead of many g_i. A usual candidate would be $q(x) = (1+x_1^2+x_2^2+\ldots+x_n^2)$ for the inverse of the uniform denominator

If apply to Motzkin this will actually certify its nonnegativity

It kind of always works Reznick['95]: If $\inf p(x) > 0$ then there exists r s.t. $q(x)^r p(x)$ is a SOS; so have to divide by denominator many times

Can turn this into a hierarchy of relaxations

In particular, will depend on this r

Let $p_r^{SOS} = \sup \lambda$ such that $q^r(p(x) - \lambda)$ is in $\Sigma[x]$ (i.e. is an SOS)

As a corollary of Reznick if r is large enough this will eventually converge to the optimum, i.e. $p^* = \lim_{r\to\infty} p_r^{sos}$

Did out a simple example

—End of First Day—

3.5 The Moment Approach

Start all over with a new way of seeing problems

Will reforumlate unconstrained POP of integral polynomial over a probability measure where optimizing over every measure

Measure is just a weighted average so min will put all weight on min point

This can be seen as dual to our original problem

Can rewrite this integral as summing over all monomials

Integrating on each monomial is called the **moment sequence**

The integral is just $< \tilde{p}, \tilde{y} >$

Letting MOM be all relevant measures just want to minimize $y \in \text{Mom}(\mathbb{R}^n) < \tilde{p}, y >$

Characterizing $\text{Mom}(K)$ is a classic, hard problem. Only a few simple cases have solutions.

A necessary condition is that $< y, \tilde{p^2} > \geq 0$

Rewriting, just want moment matrix to be PSD: $M(y) \succeq 0$ which is to say that all truncated moment matrices must be PSD. Can truncate at degree of polynomial

This is

1. Still an SDP
2. Dual to the SOS version

Did out example using the moment approach

But now we **get the minimizer** unlike when we did SOS

3.6 Nonnegative certificates over the nonnegative orthant

Given **homogenous** polynomial p, want to find if $p \geq 0$ in the positive orthant

Why care? Testing copositivity of a matrix is exactly this

Reducing to SOS

Could just square every variable

Can now use SOS to search for these certificates

When used to check copositivity this is called the **Parrilo hierarchy**

Polya's Certificates

Let Δ^n denote the standard simplex $\{x \in \mathbb{R}^n | \sum_i x_i = 1\}$

If p has only nonnegative coefficients it is nonnegative over Δ^n

Polya showed that p is positive over Δ^n iff there is a k such that $(x_1+x_2+\ldots+x_n)^k p(x)$ has only positive coefficients

Can use this to generate LP relaxations of strict copositivity of matrices

General sense of proof

Proof boils down to fact that by perturbing polynomial by ϵ. As ϵ goes to 0 this converges uniformly

3.7 Constrained Polynomial Optimization

Want to optimizie polynomial given semialgebraic constraints: points where a bunch of poylnomials are nonnegative

Can get certificates of nonnegativity by expressing it as sum of products of constraint polynomials and sums of squares

Putinar: if S is arcamedian then such a certificate exists; roughly archimedian means system is compact

There is a catch: Checking membership is hard because there are no degree bounds on the sum of squares in the certificates

Gave a bad example

Thus, to search for certificates need to bound degrees.

Truncated Quadratic Module

Demands certificates as before but sums of squares have bounded degrees

This is what Lasserre Hierarchy is doing. In particular, ertificates in Lasserre are of this form.

Dealing with Equalities

Can just do inequalities with equalities but it's instructive

Commutative algebra free way: for polynomials set to 0 just let their multipliers be whatever

Commutative algebra way: work modulo the idea I generated by the polynomials set to 0

Stable Set Example

Modding out by the ideal in stable set is just to get the θ number (the degree 1 SOS relaxation)

Tensor Projection Approximation

Want to approximate a tensor by a rank 1 tensor

Can be seen as equivalent to a polynomial optiization problem but of high degree

3.8 A Few More Nonnegativity Certificates

- Linear programming is just an important case of constrained POP; nonnegativity certificate in this case just is **Farkas' Lemma**

So can think of SOS certificates as just generalizations of Farkas' Lemma

- Linear constraints optimizing general polynomial under affine constraints
- Schmudgen's certificates

Could do same thing as in semialgebraic sets but take all possible combinations of g_js and just require that at least 1 is nonnegative

Can get rid of archimedean condition and just use compactness here

-Sums of Binomial Squares

Sometimes SOS is too strong so can try for weaker certificate. Don't write as an SOS but as a sum of binomial squares (SOBS).

Clearly SOBS implies SOS.

But SOBS have a nice characterization

Can get a hierarchy out of this as well

3.9 Final Remarks

A series of citations on introductory references. Many exercises from first one.