What to do if a problem is hard? $LP \subseteq QCQP \subseteq SOCP \subseteq SDP$

For NP-hard problems:

- consider special instances of a problem (could be easier...)
- do a convex relaxation and look for bounds (what is the gap?)
- approximation algorithms (e.g. involving randomness)

Polynomial nonnegativity and Sum-of-Squares

Consider a polynomial optimization problem

$$\begin{cases} \min & p(x) \\ \text{s.t.} & x \in K = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0,\ h_i(x) = 0\} \end{cases}$$

p, g_i, h_i - multivariate polynomials

$$\Updownarrow$$

$$\begin{cases} \max & \gamma \\ \text{s.t.} & p(x) - \gamma \geq 0 \quad \forall x \in K \end{cases}$$

$$\begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{pmatrix}^T M \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix}$$

Recall: • $p(x)$ is deg 2, $g_i(x), h_i(x)$ are deg 1 $\mapsto$ NP-hard

• $\begin{cases} \min & x^T M x \\ \text{s.t.} & x \geq 0 \end{cases}$ is NP-hard

however

• $\begin{cases} \min & x^T M x \\ \text{s.t.} & x^T x \leq 1 \end{cases}$ is "easy" (eigenvalue problem is in P)

$x^T M x \leftarrow v(x)^T M v(x)$ if $v(x) = (x_1^2, x_2^2, \ldots, x_n^2)^T$

So, we know $v(x)^T M v(x)$ is nonnegative...

SOS

If $p(x) = \sum_{i=1}^{k} q_i^2(x)$

sum-of-squares of other polynomials is always ≥ 0!

Can we test this sufficient condition independently?

Yes!

Thm Let $p(x)$ have n variables and degree $2d$. It can be written as SOS $\Longleftrightarrow$ There exists a PSD matrix Q : $p(x) = z^T Q z$, where

$z = [1, x_1, x_2, \dots, x_n, x_1x_2, \dots, x_n^d]$

all monomials up to degree d

Proof: • if Q exists, $Q = V^T V$ (Cholesky)

$p(x) = z^T V^T V z = \|Vz\|^2$ ← sum of squares!

• if $p(x) = \sum q_i^2(x)$, then

$p(x) = \sum_i [a_i^T z(x)]^2 = z^T(x) \cdot \left(\sum_i a_i a_i^T\right) \cdot z(x)$ ∎

This is an SDP problem!

$Q \succeq 0$ and coefficients of Q are coming from the coefficients of p.

Example: $p(x) = x^2y^4 + x^4y^2 + 1 - 3x^2y^2$ $\quad n = 2$, $2d = 6$, $d = 3$

$xy^2 \qquad xy^2 \cdot x$

$z = (1, x, y, xy, x^2, y^2, x^2y, xy^2, x^3, y^3)$ $\qquad x^2y = x^2 \cdot xy = x \cdot x^2y$

Q is 10×10

$$\begin{array}{cc}
 & \begin{array}{cccccccccc} 1 & x & y & xy & x^2 & y^2 & x^2y & & & \end{array} \\
z^T & \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & & & & & & & & \\
0 & & 0 & & & & & & & \\
0 & & & -3 & & & 0 & & & \\
0 & & & & & & & & & \\
0 & & & & & & & & & \\
0 & & & & & & 1 & & & \\
0 & & & & & & & 1 & & \\
0 & & 0 & & & & & & 0 & \\
0 & & & & & & & & & 0
\end{bmatrix} z
\end{array}$$

this is not SDP :)

So, checking nonnegativity is harder than SOS!

But it corresponds in some cases, including $n = 1$ (univariate), $d = 1$ (quadratic), or $n = 2, d = 2$

Motzkin polynomial is nonnegative: $\frac{x^2y^4+x^4y^2+1}{3} \geq x^2y^2$ (AMGM)

Exercise: prove that if $x^2+y^2=1 \Rightarrow x+y \leq \sqrt{2}$ using SOS ideas

(with constraint)

$(x+y)^2 = x^2+y^2+2xy$

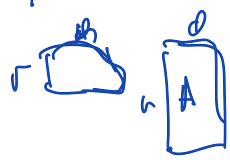
$$\sqrt{2}-x-y = \underbrace{\frac{x^2+y^2}{\sqrt{2}}}_{=1} - x - y + \frac{1}{\sqrt{2}}$$

$$= \frac{(x-y)^2}{2\sqrt{2}} + \frac{(x+y)^2}{2\sqrt{2}} - (x+y) + \frac{1}{\sqrt{2}}$$

$$= \frac{(x-y)^2}{2\sqrt{2}} + \frac{1}{2\sqrt{2}}(x+y-\sqrt{2})^2 \geq 0$$

A problem can be hard because it is large-scale (even in P)!

An idea: do dimension reduction and solve a smaller problem instead:



Random projections

- $Ax=b \longrightarrow SAx=Sb$ $\quad x=(SA)^+Sb$
- $\min\|b-Ax\|_2 \longrightarrow x_{opt}=A^+b$, $\tilde{x}_{opt}=(SA)^+Sb$

$\min\|Sb-SAx\|_2$

Are they close?

Do they give similar error?

Yes if matrix S · approximately preserves distances between the points

· while doing dimension reduction

("approximate isometry")

Luckily, many (random) matrices satisfy this property

and some of them can be "applied" fast.

Def A (random) matrix $S\in\mathbb{R}^{k\times n}$ forms a Johnson-Lindenstrauss transform JLT(ε,δ,d) if for any set of d points V with prob $\geq 1-\delta$

$$(1-\varepsilon)\|x\|_2^2 \leq \|Sx\|_2^2 \leq (1+\varepsilon)\|x\|_2^2 \quad \forall\ x\in V.$$

Then, for target dimension $r \sim \frac{d \log d}{\varepsilon^2}$ ←

$$\|b - A\tilde{x}_{opt}\|_2 \leq (1+\varepsilon)\ \|b - A x_{opt}\|$$

$$\|x - \tilde{x}_{opt}\|_2 \leq \sqrt{\varepsilon} \cdot \kappa(A)\ \|x_{opt}\|_2$$

$\kappa(A) = \frac{\sigma_{max}}{\sigma_{min}}$ condition number of A

(Sarlos, Mahoney, ... ~2006) See [M]

+ Pilanci & Wainright (constrained least squares, SVM, ...)

Can we sketch other optimization problems?

LP → simplex algorithm has instances with exponential running time

LP → interior point methods scale poorly

Linear Feasibility Problem (LFP). Given $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. Decide whether there exists $x \in \mathbb{R}^n$ such that $Ax = b \wedge x \geq 0$.

⇕

Cone Membership (CM). Given $b, a_1, \ldots, a_n \in \mathbb{R}^m$, decide whether $b \in \mathsf{cone}\{a_1, \ldots, a_n\}$.

Projected cone membership $Tb \overset{?}{\in} \text{cone}\{Ta_1, \ldots, Ta_n\}$
[Liberti et al 2015]

Sketching SDPs

$$\begin{cases} \max & \mathrm{Tr}(CX) \\ & \mathrm{Tr}(A_i X) \leq b_i \\ & X \succeq 0 \end{cases}$$

Sketching symmetric matrix X:

$\Phi(X) = SXS^T$, S is a JL transform

<u>Note</u>: We cannot sketch all SDP's!

Operator norm needs $O(n^2)$ lower bound for a fixed constant factor approximation via linear sketch [Woodruff '14]

In general, we can sketch SDP's with data having small Shatten-1 norms...

Claim: if $m \geq \sum rk(Q_i)$, S is a (ε, δ, m)-JL transform [Bluhm, Franca, '18]

$$\mathbb{P}\left(\forall i,j \quad \mathrm{Tr}(SQ_iS^TSQ_jS^T) - \mathrm{Tr}(Q_iQ_j) < 3\varepsilon \|Q_i\|_1 \cdot \|Q_j\|_1 \right) \leq 1-\delta$$

$\sum |\lambda_i|$

Sketched SDP:
$$\begin{cases} \max & \mathrm{Tr}(SCS^Ty) \\ \text{s.t.} & \mathrm{Tr}(SA_iS^Ty) \leq b_i + \mu \|A_i\|_1 \\ & y \succeq 0 \end{cases} \rightarrow \alpha_S$$

$\mu = 3\varepsilon\eta$, $\eta \geq \mathrm{Tr}(X^*)$

- $\alpha_S + 3\varepsilon \|X^*\|_1 \|C\|_1 \geq \alpha$ — one application of a claim
- Lower bound can be proved and depends on stability of a problem

$$\frac{\alpha_S}{1+3\varepsilon k\eta} \leq \alpha \qquad (\eta = tr X^*, \; k = \max_i \|A_i\|_1)$$

Proved via duality and relaxing original SDP.

Application of sketching in particular problem instances:

[Mixon, Xie '20] Clustering $\rightarrow \begin{cases} \max & x^TBx \\ & x^T \cdot 1 = 0 \\ & x = \{\pm 1\}^n \end{cases} \rightarrow \begin{cases} \max & tr(BX) \\ & \mathrm{diag}\, X = 1 \\ & X \succeq 0 \end{cases}$

Sketch-and-solve approach

taking random subsets of x_i's (graph vertices)

Stochastic Block model

[Yurtsever, Tropp et al, Udell] – low-rank solutions for SDP's

– fast low-rank approximation can be also found via sketching

Randomized SVD