SOS presentation: SOS is not obviously automatizable, even approximately

Paper written by Ryan O'Donnell

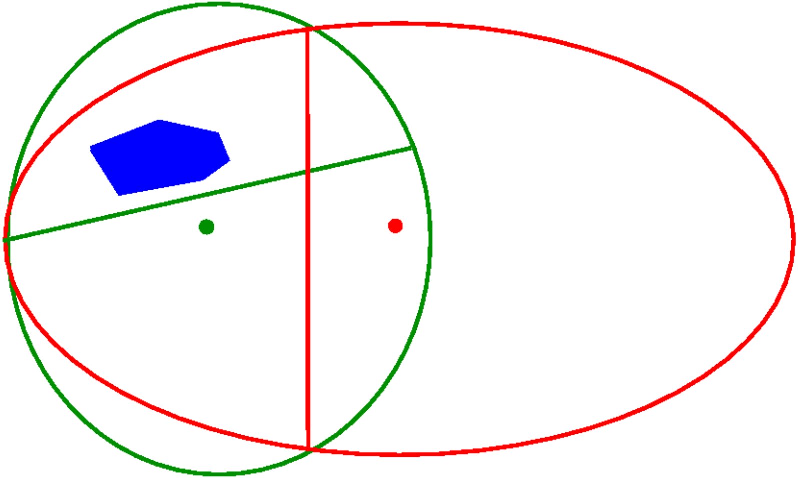
November 21, 2017

Introduction

We will only look at feasibility, no optimization.
We look at the Ellipsoid algorithm for solving SDP.

- Needs a polynomial time separation oracle for the constraints.
  PSD-ness constraint has a polynomial time separation oracle.
- Needs technical assumptions on solution space.

Ellipsoid algorithm



Ellipsoid algorithm

Technical assumptions for Ellipsoid algorithm

$V =$ feasible region for a given convex optimization problem.
Parameters:

1. R: radius of initial L_2-norm ball containing V
2. r: number such that

$$V \neq \emptyset \iff V \text{ contains some } L_2\text{-norm ball of radius } r$$

Ellipsoid algorithm

Ellipsoid algorithm:

- Start with the ball of size R (initial ellipsoid).
- Repeatedly find a violated constraint for the center, and construct the next ellipsoid based on that.

Ellipsoid algorithm

Ellipsoid algorithm:

- Start with the ball of size R (initial ellipsoid).
- Repeatedly find a violated constraint for the center, and construct the next ellipsoid based on that.

Termination:

1. Center of ellipsoid is feasible.
2. Volume gets too small, so there is no solution.

The Ellipsoid algorithm runs in time polynomial in $\log(R/r)$.

Ellipsoid algorithm running time

For linear programming, let

$$L = \text{the total number of bits in all coefficients together}$$

Can always take, without loss of generality, $R = O(2^L)$.
Usually however, $r = 0$.
Can modify the problem by changing

$$a_i x = b_i \implies -\epsilon \leq a_i x - b_i \leq \epsilon$$

for small enough ϵ.
Therefore, running time on LP is polynomial in L.

Degree-d SOS running time

Degree-d SOS can typically be formulated in $n^{O(d)}$ bits.
So $L = n^{O(d)}$, but what are r and R?
The paper gives an example where every SOS proof has very large coefficients:

$$\text{Very large} = 2^{\Omega(2^n)}$$

If we start with an ellipsoid centered at $\mathbf{0}$, then $R = 2^{\Omega(2^n)}$.
It seems that $r = 0$ (?)

Preliminary observation

SDP solutions can need doubly exponential coefficients:

$$x_1 = 2, x_{i+1} = x_i^2 \forall i$$

Solution:

$$x_n = 2^{2^{n-1}}$$

The example with large coefficients

Given the constraints

$$\begin{aligned} 2x_1y_1 &= y_1, & 2x_2y_2 &= y_2, & 2x_3y_3 &= y_3, & 2x_ny_n &= y_n \\ x_1^2 &= x_1, & x_2^2 &= x_2, & x_3^2 &= x_3, & x_n^2 &= x_n \\ y_1^2 &= y_2, & y_2^2 &= y_3, & y_3^2 &= y_4, & y_n^2 &= 0 \end{aligned}$$

Prove that $p_n(x, y) = x_1 + x_2 + x_3 + \ldots + x_n - 2y_1 \geq 0$.

The example with large coefficients

Given the constraints

$$\begin{aligned}
2x_1y_1 &= y_1, & 2x_2y_2 &= y_2, & 2x_3y_3 &= y_3, & 2x_ny_n &= y_n\\
x_1^2 &= x_1, & x_2^2 &= x_2, & x_3^2 &= x_3, & x_n^2 &= x_n\\
y_1^2 &= y_2, & y_2^2 &= y_3, & y_3^2 &= y_4, & y_n^2 &= 0
\end{aligned}$$

Prove that $p_n(x, y) = x_1 + x_2 + x_3 + \ldots + x_n - 2y_1 \geq 0$.

Solution by hand

- Solve the second row to get $x_i \in \{0, 1\} \, \forall i$.
- Solve the third row to get $y_i = 0 \, \forall i$.

The example with large coefficients

Given the constraints

$$\begin{aligned}
2x_1y_1 &= y_1, & 2x_2y_2 &= y_2, & 2x_3y_3 &= y_3, & 2x_ny_n &= y_n \\
x_1^2 &= x_1, & x_2^2 &= x_2, & x_3^2 &= x_3, & x_n^2 &= x_n \\
y_1^2 &= y_2, & y_2^2 &= y_3, & y_3^2 &= y_4, & y_n^2 &= 0
\end{aligned}$$

Prove that $p_n(x, y) = x_1 + x_2 + x_3 + \ldots + x_n - 2y_1 \geq 0$.
How do we show that SOS needs large coefficients?
We focus on degree-2 SOS here.

The example with large coefficients

Given the constraints

$$
\begin{array}{rrrr}
2x_1y_1 = y_1, & 2x_2y_2 = y_2, & 2x_3y_3 = y_3, & 2x_ny_n = y_n \\
x_1^2 = x_1, & x_2^2 = x_2, & x_3^2 = x_3, & x_n^2 = x_n \\
y_1^2 = y_2, & y_2^2 = y_3, & y_3^2 = y_4, & y_n^2 = 0
\end{array}
$$

Prove that $p_n(x, y) = x_1 + x_2 + x_3 + \ldots + x_n - 2y_1 \geq 0$.

Working "mod the ideal"

Solve

$$p_n(x, y) \equiv \sum_j \ell_j(x, y)^2 \mod (K)$$

Where K is the set of equations above and (K) the generated ideal.

The example with large coefficients

Given the constraints

$$\begin{aligned}
2x_1y_1 &= y_1, & 2x_2y_2 &= y_2, & 2x_3y_3 &= y_3, & 2x_ny_n &= y_n\\
x_1^2 &= x_1, & x_2^2 &= x_2, & x_3^2 &= x_3, & x_n^2 &= x_n\\
y_1^2 &= y_2, & y_2^2 &= y_3, & y_3^2 &= y_4, & y_n^2 &= 0
\end{aligned}$$

Prove that $p_n(x,y) = x_1 + x_2 + x_3 + \ldots + x_n - 2y_1 \geq 0$.

Solution

$$p_n(x,y) \equiv \sum_i (x_i - 2^{2^{i-1}} y_i)^2 \mod (K)$$

The example with large coefficients

$$p_n(x, y) \equiv \sum_j \ell_j(x, y)^2 \mod (K)$$

1. We ignore the "cross terms" $x_i x_j$, $x_i y_j$ and $y_i y_j$ $(i \neq j)$.
   They do not "mix" with the rest through the ideal.

The example with large coefficients

$$p_n(x,y) \equiv \sum_j \ell_j(x,y)^2 \mod (K)$$

1. We ignore the “cross terms” $x_i x_j$, $x_i y_j$ and $y_i y_j$ $(i \neq j)$. They do not “mix” with the rest through the ideal.
2. ℓ_j must have zero constant terms.

Proof

The constant term is of the form $\sum_j c_j^2$ and is not reduced by the ideal, and $p_n(x,y)$ has zero constant term.

The example with large coefficients

Therefore, if $\ell_j = \sum_i a_{ij} x_i + \sum_i b_{ij} y_i$,

$$\sum_j \ell_j(x,y)^2 \equiv \sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2) \quad \text{mod } (K, \text{crossterms})$$

where $A_i = \sqrt{\sum_j a_{ij}^2}$, $B_i = \sqrt{\sum_j b_{ij}^2}$ and $M_i = \sum_j a_{ij} b_{ij}$.
Note that by Cauchy-Schwarz, $|M_i| \leq A_i B_i$.

The example with large coefficients

The constraints were

$$\begin{aligned} 2x_1y_1 &= y_1, & 2x_2y_2 &= y_2, & 2x_3y_3 &= y_3, & 2x_ny_n &= y_n \\ x_1^2 &= x_1, & x_2^2 &= x_2, & x_3^2 &= x_3, & x_n^2 &= x_n \\ y_1^2 &= y_2, & y_2^2 &= y_3, & y_3^2 &= y_4, & y_n^2 &= 0 \end{aligned}$$

Therefore

$$\sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2) \equiv \sum_i (A_i^2 x_i + M_i y_i + B_i^2 y_{i+1}) \mod (K)$$

where $y_{n+1} = 0$.

The example with large coefficients

$$\sum_i x_i - 2y_1 \equiv \sum_i (A_i^2 x_i + M_i y_i + B_i^2 y_{i+1}) \mod (K, \text{crossterms})$$

Now we can drop the "mod".
This implies

1. $A_i = 1$ for all i.
2. $M_1 = -2$.
3. $M_{i+1} = -B_i^2$.

Combining this with $|M_i| \leq A_i B_i$, we get $B_1 \geq 2$ and $B_{i+1} \geq B_i^2$.
Therefore, $B_n \geq 2^{2^{n-1}}$.
So, the largest coefficient is doubly exponential.

Part 2: Even approximately

Degree-2 SOS proofs of the approximate version

$$p_n(x, y) \geq -o_n(1)$$

needs coefficients of size $2^{\Omega(2^n)}$.
It turns out, we can look at $p_n(x, y) \geq -0.01$.

Analysis of approximate case

We can still disregard cross-terms $x_k x_{k'}$, $x_k y_{k'}$ and $y_k y_{k'}$ $(k \neq k')$.
But linear functions may have non-zero constant terms.
Therefore, if $\ell_j = \sum_i a_{ij} x_i + \sum_j b_{ij} y_i + c_j$, $\sum \ell_j(x,y)^2$ becomes

$$\sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2 + 2U_i x_i + 2V_i y_i) + C^2$$

where A_i, B_i and M_i are as before, $U_i = \sum a_{ij} c_j$, $V_i = \sum b_{ij} c_j$,
and $C = \sqrt{\sum_j c_j^2}$.

Analysis of approximate case

$\sum \ell_j(x,y)^2$ becomes

$$\sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2 + 2U_i x_i + 2V_i y_i) + C^2$$

where A_i, B_i and M_i are as before, $U_i = \sum a_{ij} c_j$, $V_i = \sum b_{ij} c_j$, and $C = \sqrt{\sum_j c_j^2}$.

1. By Cauchy-Schwarz, $|U_i| \leq A_i C$ and $|V_i| \leq B_i C$.
2. Reducing modulo the ideal, we get

$$\sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2 + 2U_i x_i + 2V_i y_i) + C^2 =$$

$$\sum_i ((A_i^2 + 2U_i) x_i + (M_i + 2V_i) y_i + B_i^2 y_{i+1}) + C^2$$

Analysis of approximate case

$$\sum_i (A_i^2 x_i^2 + 2M_i x_i y_i + B_i^2 y_i^2 + 2U_i x_i + 2V_i y_i) + C^2 =$$
$$\sum_i ((A_i^2 + 2U_i) x_i + (M_i + 2V_i) y_i + B_i^2 y_{i+1}) + C^2 =$$
$$\sum x_i - 2y_1 + 0.01$$

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \,\forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \,\forall i$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \, \forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \, \forall i$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \, \forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \, \forall i$.
5. $C = 0.1$ and $|U_i| \leq 0.1A_i$, $|V_i| \leq 0.1B_i$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1\,\forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2\,\forall i$.
5. $C = 0.1$ and $|U_i| \leq 0.1A_i$, $|V_i| \leq 0.1B_i$.
6. $A_i^2 - 0.2A_i \leq 1 \implies A_i \leq 1.2\,\forall i$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \, \forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \, \forall i$.
5. $C = 0.1$ and $|U_i| \leq 0.1A_i$, $|V_i| \leq 0.1B_i$.
6. $A_i^2 - 0.2A_i \leq 1 \implies A_i \leq 1.2 \, \forall i$.
7. $|M_1| \geq 2 - 0.2B_1$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \, \forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \, \forall i$.
5. $C = 0.1$ and $|U_i| \leq 0.1A_i$, $|V_i| \leq 0.1B_i$.
6. $A_i^2 - 0.2A_i \leq 1 \implies A_i \leq 1.2 \, \forall i$.
7. $|M_1| \geq 2 - 0.2B_1$.
8. $|M_{i+1}| \geq B_i^2 - 0.2B_{i+1} \, \forall i$.

Analysis of approximate case

1. $C^2 = 0.01$,
2. $A_i^2 + 2U_i = 1 \, \forall i$,
3. $M_1 + 2V_1 = -2$,
4. $M_{i+1} + 2V_{i+1} = -B_i^2 \, \forall i$.
5. $C = 0.1$ and $|U_i| \leq 0.1A_i$, $|V_i| \leq 0.1B_i$.
6. $A_i^2 - 0.2A_i \leq 1 \implies A_i \leq 1.2 \, \forall i$.
7. $|M_1| \geq 2 - 0.2B_1$.
8. $|M_{i+1}| \geq B_i^2 - 0.2B_{i+1} \, \forall i$.

Now combine with $|M_i| \leq A_i B_i \leq 1.2B_i$ to get

1. $1.2B_1 \geq 2 - 0.2B_1$.
2. $1.2B_{i+1} \geq B_i^2 - 0.2B_{i+1} \, \forall i$.

So $B_i \geq 1.4(2/1.4^2)^{2^{i-1}}$, which is doubly exponential.

Analysis of approximate case: Archimedean constraints

The paper goes on to show that after adding the constraints $x_i^2 \leq 1$ and $y_i^2 \leq 1$, we still need doubly-exponential coefficients.

SOS with only booleanness constraints

If the *only* constraints are that *all* variables must be boolean ($x_i^2 = x_i$), then Ellipsoid runs in $n^{O(d)}$ time on the SOS problem. (up to additive error 2^{-n^c} for arbitrary constant c)