Combining SOS and Moment Relaxations with Branch and Bound to Extract Solutions to Global Polynomial Optimization Problems (or Why Error Bounds Matter)

Hesameddin Mohammadi
Matthew M. Peet
Cybernetic Systems and Controls Laboratory
School for Engineering of Matter, Transport and Energy
Arizona State University, Tempe, USA

SIAM Conference on Optimization, May 2017

Global Polynomial Optimization (GPO)

- **Global Polynomial Optimization** is defined as

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0 \qquad \text{for } i = 1, \cdots, m, \end{aligned}$$

where f , g_i $\in \mathbb{R}[\mathbf{x}]$ are all real valued polynomials.

The Decision Variable is $x \in \mathbb{R}^n$.

- Not Convex

Compare with **Optimization of Polynomials**

$$\begin{aligned} \underset{f\in\mathbb{R}^m[x]}{\text{minimize}} \quad & \mathcal{L}_1(f)(x)) \\ \text{subject to} \quad & \mathcal{L}_2(f)(x)) \geq 0 \qquad \text{for all } x \in \mathbb{R}^m, \end{aligned}$$

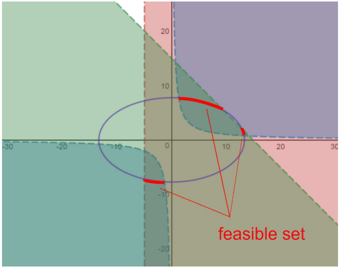
The Decision Variable is $f \in \mathbb{R}^m[x]$.

- Convex

Example: A 2-variate GPO problem

- Consider the optimization problem

$$\begin{aligned} \min_{x,y\in\mathbb{R}}. \quad & y \\ \text{subject to} \quad & x+5 \geq 0\,, \\ & x\,y - 10 \geq 0\,, \\ & 15 - x - y \geq 0\,, \\ & x^2 + 3y^2 - 180 = 0\,. \end{aligned}$$



Many problems can be formulated as a GPO

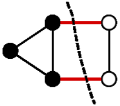
Combinatorial Examples :

Max-cut problem:
asks for a minimal edge-cut of maximum weight.

$$\min_{\mathbf{x}\in\mathbb{R}^n}. \quad \sum_{i,j\,:\,(v_i,v_j)\in E} (x_i\, x_j\ -1)\left(\frac{w_{ij}}{2}\right)$$

$$\text{s.t.} \quad x_i^2 = 1 \quad \text{for } i = 1,\cdots,n.$$

We want to know the CUT!
($x_i \in \{-1, 1\}$)



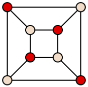
Stable set problem:
asks for a maximum independent set of vertices $S \subset V$.

$$\min_{\mathbf{x}\in\mathbb{R}^n}. \quad \sum_{i=1}^{n} -x_i$$

$$\text{s.t} \quad x_i\, x_j = 0 \quad \text{for } v_i v_j \in E,$$

$$\text{s.t} \quad x_i^2 - x_i = 0 \quad \text{for } i = 1,\ldots,n.$$

We want to know the Vertices!
($x_i \in \{0, 1\}$)



Greatest Lower Bound (GLB) problem

Let $S := \{x \in \mathbb{R}^n : g_i(x) \geq 0, \text{for } i = 1, \dots, m\}$ be a semialgebraic set.

GPO problem:

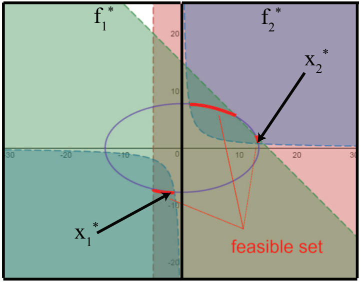
$$f^* := f(x^*) = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad \mathbf{x} \in S.$$

GLB problem:

$$\lambda^* := \underset{\lambda \in \mathbb{R}}{\text{maximize}} \quad \lambda$$
$$\text{subject to} \quad f(x) - \lambda \geq 0 \quad, \forall x \in S.$$

- Then $\lambda^* = f^*$ (iff $S \neq \emptyset$).
- GLB is convex and relatively tractable.
- Can we use GLB to solve GPO?

Suppose we can solve GLB exactly ($\lambda^* = f^*$)

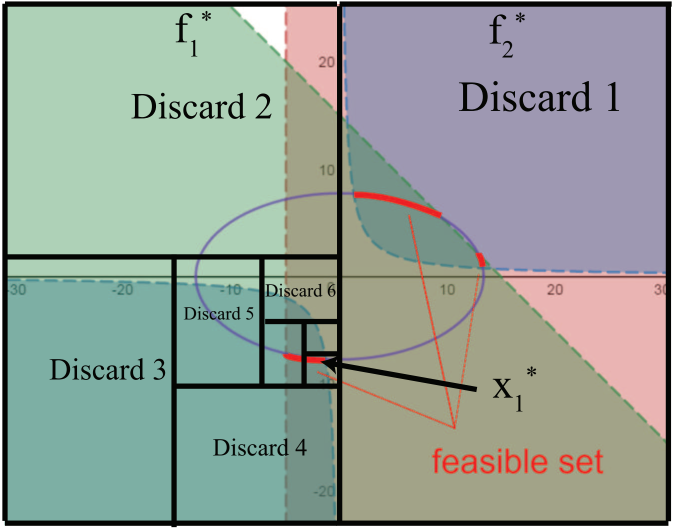


$$\lambda_1^* = f_1^* := \min_{x,y\in\mathbb{R}} \; y$$
subject to
$$\begin{aligned} & x+5 \ge 0\,, \\ & x\,y - 10 \ge 0\,, \\ & 15 - x - y \ge 0\,, \\ & x^2 + 3y^2 - 180 = 0\,. \end{aligned}$$
Extra $\rightarrow$ $-x(x+30) \ge 0\,.$

$$\lambda_2^* = f_2^* := \min_{x,y\in\mathbb{R}} \; y$$
subject to
$$\begin{aligned} & x+5 \ge 0\,, \\ & x\,y - 10 \ge 0\,, \\ & 15 - x - y \ge 0\,, \\ & x^2 + 3y^2 - 180 = 0\,. \end{aligned}$$
Extra $\rightarrow$ $x(30-x) \ge 0\,.$

Then we know something!: If $f_1^* < f_2^*$, then $x^* \in [-30, 0]$. Else, $x^* \in [0, 30]$

Repeat Until we get Desired Accuracy



Complexity For Bisection along the longest edge, after q iterations, the longest edge of the hypercube has decreased by a factor of $2^{-q/n}$.

CIAIM: Given an algorithm which solves GLB (*exactly!*) in $O(k)$ steps, then for any accuracy ϵ, we can design an algorithm which returns an $x \in \mathbb{R}^n$ such that

- $|x - x^*| \leq \epsilon$,
- $|f(x) - f(x^*)| < \epsilon$.

in $O(\log(1/\epsilon)k)$ steps.

ALGORITHM:

- Suppose the feasible set of the GPO problem, S, satisfies

$$\emptyset \neq S \subset A = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

- At every iteration, we have a hyper-rectangle $A = [a, b]$;
  1. Bisect $A = [a, b] = [a', b'] \cup [a'', b''] = A_1 \cup A_2$;
  2. Compute the Greatest Lower Bound of

$$\begin{aligned} \lambda_i^* := \max_{\lambda \in \mathbb{R}}. \quad & \lambda \\ \text{subject to} \quad & f(x) - \lambda > 0 \quad , \forall x \in S \cap A_i; \end{aligned}$$

  3. If $\lambda_1^* > \lambda_2^*$, set $A = A_1$, otherwise $A = A_2$;
  4. Goto 1 ;
  5. At termination, we choose any $x \in A$.

GLB as a SOSP

SOS Polynomials: $\Sigma_{\mathbf{S}} := \{s \in \mathbb{R}[\mathbf{x}] \,:\, s(x) = \sum_{i=1}^{l} (p_i(x))^2,\, p_i \in \mathbb{R}[\mathbf{x}],\, l \in \mathbb{N}\}$
Feasible Set (Nonempty, Compact): $S := \{\mathbf{x} \in \mathbb{R}^n \,:\, g_i(\mathbf{x}) \geq 0\}$.
The **Quadratic Module** generated by $g_1, \ldots, g_m$ is defined as

$$M := \{\sigma_0(\mathbf{x}) + \sum_{i=1}^{m} \sigma_i(\mathbf{x})\, g_i(\mathbf{x}) \mid \quad \sigma_i \in \Sigma_S\}.$$

- if $\quad h \in M, \quad$ then $\quad h(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in S$. **Obvious**
- if $\quad h(\mathbf{x}) > 0, \quad \forall \mathbf{x} \in S, \quad$ then $\quad h \in M$. **Putinar's Positivstellensatz**

↳ In this case, M must be Archimedean. Can add a redundant norm bound inequality if not ($g := R^2 - \sum_{j=1}^{n} x_i^2 \geq 0$).

The GLB problem is **Equivalent** to the following SOSP:

$$\lambda^* := \max_{\lambda \in \mathbb{R}}. \quad \lambda$$

$$\text{subject to} \quad f(\mathbf{x}) - \lambda \in M,$$

- Which is equivalent to an infinite dimensional semidefinite program.

$$f(\mathbf{x}) - \lambda = Z(\mathbf{x})^T (\Omega_0 + \sum_{i=1}^{m} \Omega_i g_i(\mathbf{x})) Z(\mathbf{x}), \quad \text{for some PSD matrices } \Omega_i \geq 0.$$

Truncating the SOSP

- Bounding the degree of SOS polynomials by $k \in \mathbb{N}$, we can define the degree-k bounded quadratic module as

$$(M)_k := \{\sigma_0(\mathbf{x}) + \sum_{i=1}^{m} \sigma_i(\mathbf{x})\, g_i(\mathbf{x}) | \quad \sigma_i \in \Sigma_S,\ \deg(\sigma_i) \leq k\}.$$

- Define the optimization problem (D_k) :

$$\begin{aligned} \lambda_k^* := \max_{\lambda \in \mathbb{R}}. \quad & \lambda \\ \text{subject to} \quad & f(\mathbf{x}) - \lambda \in (M)_k, \end{aligned}$$

.

  - (D_k) is a relaxation to the GLB problem.
  - (D_k) is a semidefinite program with $o((m+1)\binom{n+k}{k}^2)$ variables.

How Accurate is the Truncated SOSP?

Error bound on λ_k^*

For any $k \in \mathbb{N}$, (D_k) is a relaxation of the GLB problem.

1. The case $S \neq \emptyset$ [Nie, Scweighofer]:

$$0 \leq \lambda^* - \lambda_k^* \leq \frac{c_1}{\sqrt[c_2]{\log k}}$$

- $c_1,\ c_2$ are constants - depend on f (objective) and g_i (constraints).
- Recall λ^* is the optimal objective value of the GLB problem.
- Recall λ_k^* is the optimal objective value of the relaxation (D_k).

2. The case $S = \emptyset$:

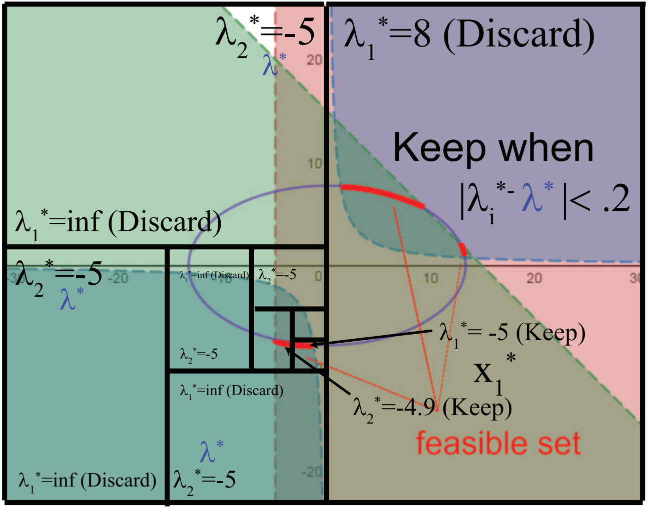
$$-1 \text{ test:} \qquad -1 \in M \implies$$

$$\exists k_0 \in \mathbb{N} \quad \text{s.t.} \quad \lambda_k^* = \infty\ , \quad \forall k > k_0\ ,$$

- There is no available bound on k_0 so far.

Modified B+B

For desired Accuracy, ϵ, let $L > n\log_2\left(\frac{r\sqrt{n}}{\epsilon}\right)$ (Iterations)(r = radius)



Trim a region B_i when $\lambda^*(B_i) > \lambda^*_{\min} + \frac{m\epsilon}{1+l}$

Modified Branch and Bound (using SOS)

Index the algorithms as $E_k,\ k \in \mathbb{N}$ acording to the degree of the GLB subroutine.

- Let $\epsilon > 0$: error tolerance. $L > n \log_2 \left(\frac{r\sqrt{n}}{\epsilon} \right)$: # of iterations.
- At iteration l, we have an active hyper-rectangle $A = [a, b]$ and a set of feasible rectangles $Z_i = \{[a_i, b_i]\}_i$ each with associated GLB λ_i.
  1. Bisect $A = [a, b] = [a', b'] \cup [a'', b''] = A_1 \cup A_2$
  2. Solve the k'th order SOS relaxation associated to the GLB problem

$$\lambda_i^* := \max_{\lambda \in \mathbb{R}} . \quad \lambda$$
$$\text{subject to} \quad f(x) - \lambda > 0 \quad , \forall x \in S \cap A_i.$$

  3. If $\lambda_i^* \leq \lambda^* + \frac{l\epsilon}{L}$, add A_i to Z.
  4. Set $A = Z_i$ where Z_i is of the smallest volume in Z.
  5. Set $\lambda^* = \lambda_j$ where Z_j has the lowest lower bound in Z.
  6. GOTO 1

FEASIBLE POINT: At termination, choose any $x^* \in A$.
CLAIM: $\exists y \in S$ such that $\|y - x^*\| \leq \epsilon$ and $f(y) - f^* \leq \epsilon$.

To Understand Convergence, Consider Implementation:

The GPO problem:

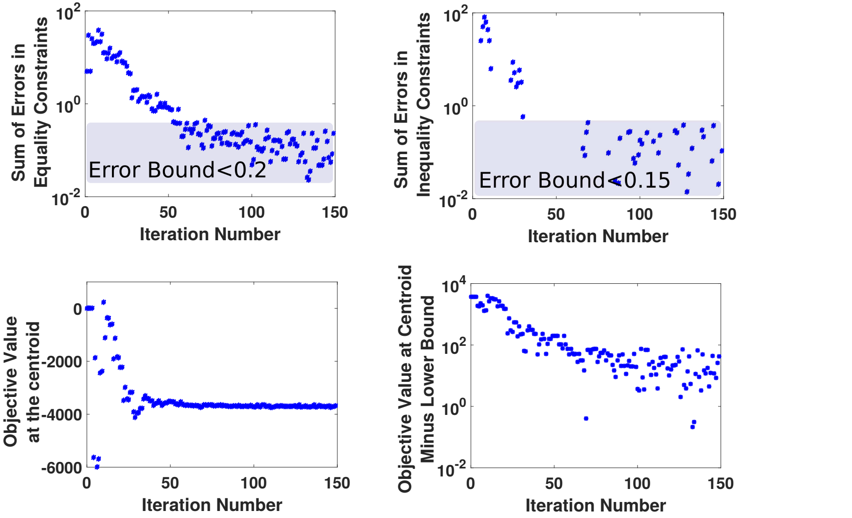
$$\begin{aligned}
\underset{\mathbf{x}\in\mathbb{R}^6}{min.} \quad & f(\mathbf{x}) = 7x_1x_5^3 + 6x_1x_5^2x_6 + 9x_2x_4^3 + 4x_2x_4x_5 + \\
& \qquad 3x_2x_5x_6 + x_3x_4x_5 \\
\text{subject to} \quad & g_1(\mathbf{x}) = 100 - (x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2) \geq 0 \\
& g_2(\mathbf{x}) = x_1^3 + x_2^2x_4 + x_3x_5^2 \geq 0 \\
& g_3(\mathbf{x}) = x_2^2x_1 + x_5^3 + x_4x_1x_2 \geq 0 \\
& h_1(\mathbf{x}) = x_1 + x_2^2 - x_3^2 + x_4x_5 = 0 \\
& h_2(\mathbf{x}) = x_5x_1 - x_4^2 = 0
\end{aligned}$$

In this example:

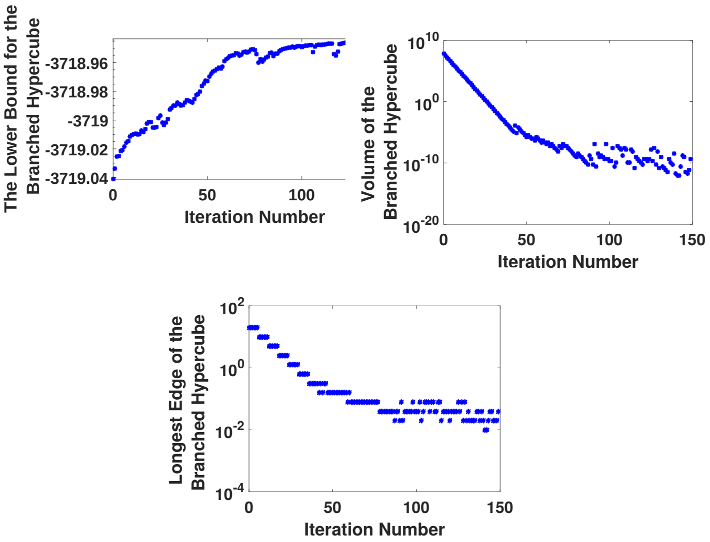
- There are $n = 6$ variables.

Choose parameter $\epsilon = 0.05$ with SOS relaxations of degree 5.

Numerical Implementation



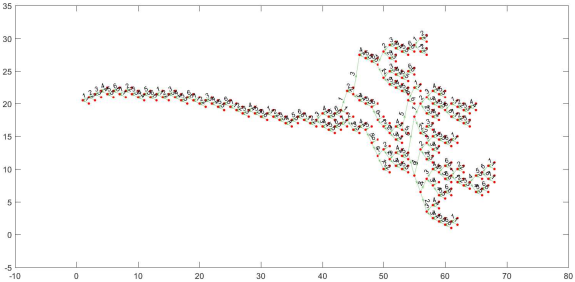
Numerical Implementation



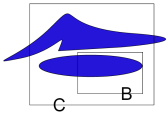
Numerical Implementation

The algorithm returns the following point:
$\hat{x}^* = [5.1416,\ 3.9307,\ 0.7568,\ -4.6777,\ 4.2676,\ -4.1504]$.

- All inequalities are feasible. The equality constraints h_1 and h_2 have errors of 0.0563 and 0.0610, respectively.
- The best valid lower bound is -3718.9. The objective value is $f(\hat{x}^*) = -3693.3$, within the error of 0.007%.
- The first 40 iterations result in a nested sequence of branchings.



CLAIM: The lower bounds are not decreasing



If we intersect the semialgebraic set

$$S := \{\mathbf{x} | g_i(\mathbf{x}) \geq 0\,,\ i = 1, \ldots, m\}$$

with the hypercubes

$$C := \{\mathbf{x} : (x_i - c_i)(d_i - x_i) \geq 0,\ i = 1, \ldots, n\},$$
$$B := \{\mathbf{x} : (x_i - a_i)(b_i - x_i) \geq 0,\ i = 1, \ldots, n\}$$

where $B \subset C$ and $C \cap S \neq \emptyset$ and define the GLB problems

$$\lambda_b^* := \min_{\lambda \in \mathbb{R}}. \quad -\lambda \qquad\qquad \lambda_c^* := \min_{\lambda \in \mathbb{R}}. \quad -\lambda$$

$$\text{s.t. } f(\mathbf{x}) - \lambda > 0,\ \forall\, \mathbf{x} \in S \cap B, \qquad\qquad \text{s.t. } f(\mathbf{x}) - \lambda > 0,\ \forall\, \mathbf{x} \in S \cap C,$$

then obviously

$$\lambda_b^* \geq \lambda_c^*.$$

How about the solution of the k'th order SOS relaxations?

PROOF: The lower bounds are not decreasing

For a fixed k, the sequential monoids are nested.

Lemma: If $a \leq c < d \leq b \in \mathbb{R}$, $(g_{a,b}) := (x-a)(b-x)$ and $(g_{c,d}) := (x-c)(d-x)$ then there exist α, β and $\gamma \in \mathbb{R}$, such that

$$(g_{a,b})(x) = \alpha(g_{c,d}(x)) + \beta(x+\gamma)^2\,, \quad \alpha, \beta \geq 0.$$

Therefore, if

$$M^{(k)}_{c\,d} := \Big\{ p \,:\, p = \sigma_0 + \sigma_1 g + \sigma_2(g_{c,d}),\ \sigma_i \in \Sigma_S,\ \mathsf{deg}(\sigma_i g) \leq k \Big\},$$

and

$$M^{(k)}_{a\,b} := \Big\{ p \,:\, p = \sigma_0 + \sigma_1 g + \sigma_2(g_{a,b}),\ \sigma_i \in \Sigma_S,\ \mathsf{deg}(\sigma_i g) \leq k \Big\},$$

we have $M^{(k)}_{c\,d} \subset M^{(k)}_{a\,b}$.

Main Result

Theorem: For any GPO problem

$$f^* := \min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x})$$
$$\text{subject to} \quad \mathbf{x} \in S = \{\mathbf{x} : g_i(\mathbf{x}) \geq 0\} \qquad \text{for } i = 1, \cdots, m,$$

where $S \neq \emptyset$ and S is compact and for any desired accuracy, $\epsilon > 0$,

- there exists a $k \in \mathbb{N}$,
- s.t. for $L > n \log_2 \left(\frac{r\sqrt{n}}{\epsilon}\right)$,

if the algorithm $E_k(L)$ returns the point $x^* \in A$, then

- there exists $y \in S$,
- $f(y) - f^* \leq \epsilon$,
- $\|y - x\| < \epsilon$,

where k depends on ϵ f, g_i.
OPEN Question: What is k???

Proof Outline

Given iteration bound, L, find the set of all possible hyper-rectangles, $\{S_i\}$

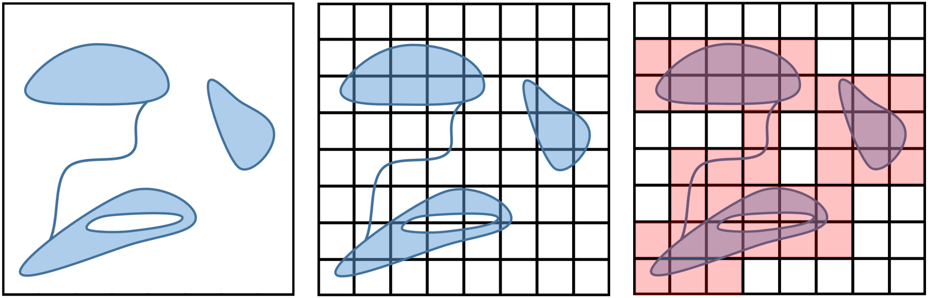
- Over each feasible set, we have a bound

$$\lambda^* - \lambda_k^* \leq \frac{c_1}{\sqrt[c_2]{logk}}$$

- This yields a degree bound for which we take the max over the set of all feasible sets.

The nested hypercube lemma and *increasing* error tolerance then ensures a sequence of nested hyper-rectangles.

Recall: the trim condition at iteration l: $\lambda_i^* \leq \lambda^* + \frac{l\epsilon}{L}$



Conclusion

- We proposed a hierarchy of Algorithms $E_k,\ k \in \mathbb{N}$ to extract solutions to the GPO problem

$$f^* := \min_{\mathbf{x}\in\mathbb{R}^n}. \quad f(\mathbf{x})$$
$$\text{subject to} \quad g_i(\mathbf{x}) \geq 0 \qquad \text{for } i = 1, \cdots, m,$$

  based on a combination of the BB and SOS relaxations.
- The computational-complexity of Algorithm E_k is
  - polynomial in k,
  - polynomial in the number of constraints,
  - linear in the number of branches l.
- For any scaler $\epsilon > 0$, there exists $k \in \mathbb{N}$ such that Algorithms E_k, in $O(\log(1/\epsilon))$ number of iterations, returns a point that is within the ϵ-distance of a feasible and ϵ-suboptimal point.
- For a fixed $k \in \mathbb{N}$, our numerical case study demonstrates convergence of E_k to a level of residual error which can then be decreased by increasing the degree.

future work: bound this residual error as a function of degree using available bounds on the error of SOS relaxations.

Thank you!