SUPPLEMENTARY MATERIAL

PRELIMINARIES

We now provide some terminology and notations related to polynomials and polynomial optimization using sum-of-squares. An n-variate real polynomial P is a sum of finitely many terms of the form $c_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and $c_\alpha \in \mathbb{R}$. The monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is also denoted by x^α, and the polynomial P can be written as $P(x) = \sum_{\alpha \in \mathbb{N}^n} c_\alpha x^\alpha$ where $c_\alpha \neq 0$ only for finitely many α. The degree of a monomial x^α is $|\alpha| := \alpha_1 + \ldots + \alpha_n$, and the degree of a polynomial is the maximum degree of all its monomials with non-zero coefficients.

A n-variate polynomial $P = \sum_\alpha c_\alpha x^\alpha$ with degree $\leq d$ can be associated with its coefficient vector (c_α) as a point in $\mathbb{R}^{s_n(d)}$, where $s_n(d) := \binom{n+d}{d} = O(n^d)$ (which can be seen by counting the monomials x^α with $\alpha \in \mathbb{N}^n$ and $|\alpha| = \alpha_1 + \ldots + \alpha_n \leq d$).

A set $K \subseteq \mathbb{R}^n$ is said to be a *(basic closed) semi-algebraic set* if there exist n-variate polynomials $g_1, \ldots, g_m$ such that

$$K = \{x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for all } i \in [m]\}.$$

SUM-OF-SQUARES RELAXATIONS

A polynomial P is said to be a sum-of-squares (s.o.s) if there exist some $m \geq 1$ and n-variate polynomials $G_1, \ldots, G_m$ such that $P = G_1^2 + \ldots + G_m^2$. The set of polynomials $G := \{G_1, \ldots, G_m\}$ is said to be a *sum-of-squares decomposition* of P. The degree of the s.o.s decomposition is defined to be $\deg(G) := \max_{i \in [m]} \deg(G_i)$. A polynomial P is said to be a degree-d sum-of-squares if it has a s.o.s decomposition of degree $\leq d$. Clearly, a polynomial which is degree d s.o.s has degree $\leq 2d$, and the degree of a s.o.s representation for a degree $\leq 2d$ polynomial (if it exists) is at most d.

It is easy to see that every s.o.s polynomial is non-negative or positive semi-definite (p.s.d), but the converse (every p.s.d polynomial is s.o.s) is not true except in very specific cases (univariate polynomials, quadratics, bivariate quartics), as proved by Hilbert (Hilbert 1888).

However, we can construct a sound, but incomplete, verifier for the non-negativity (p.s.d-ness) of a given polynomial by checking whether the polynomial has degree-d sum-of-squares decomposition (for appropriately large d). Shor (Shor 1987) showed that the question of whether a given polynomial f has a degree-d sum-of-squares decomposition is equivalent to the feasibility of a semidefinite program (SDP) with $O(n^{2d})$ variables and $O(n^d)$ constraints. For constant d, such an SDP (which we may call the degree-d s.o.s relaxation) can be solved in poly(n) time.

Let $[x]_d$ denote the $s_n(d)$-length vector of all n-variate monomials with degree $\leq d$, according to some monomial ordering. Say,

$$[x]_d := (1 \;\; x_1 \cdots x_n \;\; x_1^2 \;\; x_1x_2 \cdots x_1x_n \quad x_n^2 \cdots x_1^d \cdots x_n^d).$$

Let f be a n-variate real polynomial with $\deg(f) \leq 2d$. That is,

$$f = \sum_{\substack{\alpha \\ |\alpha| \leq 2d}} c_\alpha x^\alpha = c^\top [x]_{2d} \quad \text{for some } c \in \mathbb{R}^{s_n(2d)}.$$

Theorem 1 ((Shor 1987)). *f is degree-d s.o.s if and only if there exists a symmetric positive semidefinite matrix $Q \in \mathbb{R}^{s_n(d) \times s_n(d)}$ such that $f = [x]_d^\top Q [x]_d$, coefficient-wise. That is, $c_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta,\gamma}$ for all α such that $x^\alpha \in [x]_{2d}$, and β, γ such that $x^\beta, x^\gamma \in [x]_d$.*

To perform (unconstrained) polynomial optimization — i.e. finding the global minimum $f^* := \inf_{x \in \mathbb{R}^n} f(x)$ of a given polynomial function f — using sum-of-squares, Shor (Shor 1987) formulated a sequence of sum-of-squares relaxation SDPs (which have increasing size/complexity as the degree d increases). The degree-d SDP finds $f_{\text{sos}}^{(d)} := \sup \gamma$, s.t. $f - \gamma$ is a degree-d s.o.s (which implies that γ is a lower bound for f).

$$\begin{gathered}
\max_Z \; -A^{(\mathbf{0})} \circ Z, \quad \text{subject to} \\
A^{(\alpha)} \circ Z = c_\alpha \;\; (\text{where } A^{(\alpha)}_{\beta,\gamma} = 1 \text{ if } \beta + \gamma = \alpha \text{ and } 0 \text{ othewise.}) \\
(\text{for all } \alpha \neq \mathbf{0} \in \mathbb{N}_d^n) \\
Z \succeq 0, \; Z \in \mathbb{S}^{s_n(d)}(\mathbb{R})
\end{gathered}$$

The dual of the above SDP is

$$\begin{gathered}
\min_y \; c^\top y, \quad \text{subject to} \\
\sum_\alpha y_\alpha \cdot A^{(\alpha)} \succeq 0, \; y_{\mathbf{0}} = 1, \; y \in \mathbb{R}^{s_n(2d)}
\end{gathered}$$

In the above SDP, Z may be interpreted as $Z \equiv Q - \gamma \cdot E_{11}$, where $Q \in \mathbb{S}^{s_n(d)}(\mathbb{R})$ is a symmetric p.s.d matrix such that $[x]_d^\top Q [x]_d = f$ (as in Theorem 1), and E_{11} denotes the elementary matrix with a 1 in the (first row, first column) and zeros elsewhere.

This implies, of course, that the objective is $-A^{(\mathbf{0})} \circ Z = -Z_{\mathbf{0},\mathbf{0}} = \gamma - c_{\mathbf{0}}$; maximizing it is equivalent to maximizing γ, and the s.o.s lower bound $f_{\text{sos}}^{(d)} := \gamma$ may be recovered as $\gamma = c_{\mathbf{0}} - Z_{\mathbf{0},\mathbf{0}}$.

This hierarchy of SDPs gives a sequence of increasing lower bounds $f_{\text{sos}}^{(1)} \leq f_{\text{sos}}^{(2)} \leq f_{\text{sos}}^{(3)} \leq \ldots$ for f, where we define $f_{\text{sos}}^{(d)} := -\infty$ if the degree-d SDP is infeasible. It is also possible in some cases (with s.o.s relaxations of sufficiently high degree) to extract a *certificate* x^* for the lower bound (i.e. an $x^* \in \mathbb{R}^n$ such that $f(x^*) = f_{\text{sos}}^{(d)}$). Clearly, the existence of such a certificate implies that the sum-of-squares hierarchy has reached the actual optimum, i.e. $f_{\text{sos}}^{(d)} = f^*$. However, this will not occur for all polynomials (and hence the s.o.s-based non-negativity verifier will always be *incomplete*). In the unconstrained minimization case, we only need to check s.o.s relaxations of degree $\leq 2d$. But such a degree upper-bound is not known for the constrained case, which is described below.

Finally, it was shown by Lasserre (Lasserre 2001) and Parrilo (Parrilo 2000) independently that it is possible to lower-bound a polynomial optimization problem over a basic closed semi-algebraic set $K \subseteq \mathbb{R}^n$ by using semidefinite relaxations — a (basic closed) semi-algebraic set $K \subseteq \mathbb{R}^n$ is the intersection of the solution sets of finitely many non-strict inequalities of real polynomials. This is done by applying results from real algebraic geometry known as *Positivstellensatz*.

If $\mathbb{K} = \{x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for all } i = 1, \ldots, m\}$ is a *compact* semi-algebraic set, then we can get sum-of-squares lower bounds (using Putinar's Positivstellensatz) for the constrained polynomial optimization problem of finding $f_{\mathbb{K}}^* := \inf_{x \in \mathbb{K}} f(x)$ as follows:

Let $v_j := \lceil \deg(g_j)/2 \rceil$, and let

$$d \geq d_0 := \max(\lceil \deg(f)/2 \rceil, v_1, \ldots, v_m).$$

Then

$$f_{\text{sos}}^{(d)} := \sup \gamma, \text{ s.t.}$$
$$f - \gamma = \sum_{j=1}^{m} \sigma_j g_j, \text{ where } \sigma_j \text{ is degree-}(d - v_j) \text{ s.o.s}$$

Similar to the unconstrained case, this gives a sequence of increasing lower bounds for f^*.

References

[1] David Hilbert. "Ueber die Darstellung definiter Formen als Summe von Formenquadraten". In: *Mathematische Annalen* 32.3 (1888), pp. 342–350.

[2] Jean B. Lasserre. "Global Optimization with Polynomials and the Problem of Moments". In: *SIAM J. on Optimization* 11.3 (2001), pp. 796–817.

[3] Pablo A Parrilo. "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization". PhD thesis. Caltech, 2000.

Algorithm 1 Individual bias verification for kernelized RBF models.

procedure FIND-BIAS-RBF
  Let $f = \sum_{i=1}^{M} w_i y_i \exp(-\gamma \|x - x_i\|^2)$, with $c < w_i < C$ for all $i \in [M]$.
  Let $\mathcal{S}^+$ be the subset of $\{(x_i, y_i)\}_{i=1}^{M}$ with $y_i = 1$ and $\mathcal{S}^-$ be the subset with $y_i = -1$.
  Let $L := \emptyset$.
  Construct the set V_p:

$$V_p := \{(v, v') \mid v, v' \text{ are feasible for } x_D,\, x'_D \text{ and } |v_i - v'_i| \leq \varepsilon_j \ \forall i \in D \cap S_j \ \forall j \in [t]\}$$

  for all $(v, v') \in V_p$ **do**
    for all $x_r \in \mathcal{S}^+$ **do**
      for all $x_s \in \mathcal{S}^-$ **do**
        Solve this optimization problem to get x^*, x'^*:

$$\min_{\text{valid } x, x'} \frac{1}{2}\left(\sum_{u \in \mathcal{S}^+} w_u \|x' - x_u\|^2 + \sum_{v \in \mathcal{S}^-} w_v \|x - x_v\|^2\right)$$
$$\text{subject to}$$
$$x_{rk} - D_r \leq x_k, x'_k \leq x_{rk} + D_r \text{ and}$$
$$x_{sk} - D_s \leq x_k, x'_k \leq x_{sk} + D_s, \text{ for all } k$$
$$|x_i - x'_i| \leq \varepsilon_j \ \ \forall i \in S_j \cap \overline{D} \ \forall j \in [t]$$
$$x_D = v \text{ and } x'_D = v'$$

        if $f(x'^*) \geq \varepsilon$ and $f(x^*) \leq -\varepsilon$ **then**
          Output (x^*, x'^*) and **return**
        else
          Add $f(x'^*) - f(x^*)$ to L.
  Output the lower bound $L^* := \min L$.

[4] Naum Z Shor. "Class of global minimum bounds of polynomial functions". In: *Cybernetics and Systems Analysis* 23.6 (1987), pp. 731–734.