Polynomial optimization (revisited): references, exercises, questions and answers

Didier Henrion[1,2]

Prague 1 December 2022

1 References

The moment-SOS hierarchy applied to polynomial optimization is described in [23, 24] and surveyed in [22], see also [26] for an introduction. It fits the framework of the generalized problem of moments [25, 21] that we also follow in the second and third lecture to deal with dynamical systems. See [12] for sketchy lecture notes for the first and third lectures.

See [2, 1, 5] for textbooks on convex optimization. Conic duality and infinite-dimensional optimization are covered in [28] and also [1]. For a concise account of background material on the approximation of cones of moments and positive polynomials, see [20, Chapter 2].

Asymptotic convergence of the moment-SOS hierarchy for POP, as proved originally in [24], relies on Putinar's solution [31] to the problem of moments based on a version of the Positivstellensatz, a representation of positive polynomials [33]. See also [29, 3] for positive polynomials and SOS.

Finite convergence, and global optimality certificate for POP is based on flat extensions of moment matrices [7, 8]. Generic finite convergence of the moment-SOS hierarchy was proved in [30].

Extraction of the global minimizers for POP was described in [15].

The GloptiPoly package for Matlab is described in [13, 16].

[1]CNRS; LAAS; Université de Toulouse, 7 avenue du colonel Roche, F-31400 Toulouse, France.
[2]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czechia.

2 Exercises

2.1 Exercise 1.1

2.1.1 Statement

Given a polynomial $p \in \mathbb{R}[x]$ and a compact set X, the polynomial optimization problem (POP)

$$\begin{array}{rcll} v^* & = & \min_x & p(x) \\ & & \text{s.t.} & x \in X \end{array}$$

is reformulated as the linear problem (LP)

$$\begin{array}{rcll} p^* & = & \inf_\mu & \langle p, \mu \rangle \\ & & \text{s.t.} & \langle 1, \mu \rangle = 1 \\ & & & \mu \in C(X)'_+ \end{array}$$

where $C(X)'_+$ is the cone of (Borel regular positive) measures on X, topologically dual to the cone of positive functions on X, and the duality

$$\langle f, \mu \rangle := \int_X f(x) d\mu(x)$$

is integration of a function f by a measure μ.

Prove that $v^* = p^*$ and that the LP has an optimal solution equal to the Dirac measure at any optimal solution of the POP.

2.1.2 Solution

For any feasible $\xi \in X$, it holds $p(\xi) = \langle p, \mu \rangle$ for the Dirac measure $\mu = \delta_\xi$, showing $v^* \geq p^*$. Conversely, as $p(x) \geq v^*$ for all $x \in X$, it holds $\langle p, \mu \rangle \geq \langle v^*, \mu \rangle = v^* \langle 1, \mu \rangle = v^*$ since μ is a probability measure, which shows that $p^* \geq v^*$. It follows that $v^* = p^*$ and that the infimum in the LP is attained by a Dirac measure $\mu = \delta_{x^*}$ where x^* is any global optimum of the POP.

2.2 Exercise 1.2

2.2.1 Statement

Dual to the primal LP

$$\begin{array}{rcll} p^* & = & \inf_\mu & \langle p, \mu \rangle \\ & & \text{s.t.} & \langle 1, \mu \rangle = 1 \\ & & & \mu \in C(X)'_+ \end{array}$$

is the LP

$$\begin{array}{rcll} d^* & = & \sup_{v \in \mathbb{R}} & v \\ & & \text{s.t.} & p - v \in C(X)_+. \end{array}$$

Derive the dual LP from the primal LP using convex duality. Prove that strong duality holds i.e. $p^* = d^*$. Give a graphical interpretation to the dual LP.

2.2.2 Solution

Let us build the Lagrangian $\ell(\mu, v, w) := \langle p, \mu \rangle - (\langle 1, \mu \rangle - 1)v - \langle w, \mu \rangle$ where $v \in \mathbb{R}$ is the Lagrange multiplier or dual variable associated to the linear constraint $\langle 1, \mu \rangle = 1$ and $w \in C(X)_+$ is the dual variable associated to the conic constraint $\mu \in C(X)'_+$. Rearrange the dual Lagrange function $d(v, w) := \inf_{\mu \in C(X)'} \ell(\mu, v, w) = v + \inf_{\mu \in C(X)'} \langle p - v - w, \mu \rangle$ and observe that it is bounded below only if $w = p - v$. The dual LP follows readily.

The dual LP seeks the largest lower bound on the graph of p on X, its supremum is attained at $d^* = \min_{x \in X} p(x)$, which is also the value attained by the primal LP as shown in Exercise 1.2. This shows strong duality.

An alternative proof of strong duality consists of applying [1, Theorem IV.7.2] after observing that the cone $\{(\langle 1, \mu \rangle, \langle p, \mu \rangle) : \mu \in C(X)'_+\}$ is closed in $\mathbb{R}^2$.

2.3 Exercise 1.3

2.3.1 Statement

Prove that deciding whether a polynomial is a sum of squares (SOS) can be reduced to semidefinite programming.

2.3.2 Solution

Let $\mathbb{R}[x]_d$ denote the vector space of polynomials of degree up to d in the indeterminates $x \in \mathbb{R}^n$. Let $\mathbb{N}^n_d := \{a \in \mathbb{N}^n : \sum_{k=1}^n a_k \leq d\}$ and $\mathbf{b} := (b_a)_{a \in N^n_d} \in \mathbb{R}[x]^{\binom{n+d}{n}}$ denote a basis for this space, so that every element $p \in \mathbb{R}[x]_d$ can be expressed as a linear combination

$$p = \sum_{a \in \mathbb{N}^n_d} p_a b_a = \mathbf{p}^T \mathbf{b}$$

with coefficient vector $\mathbf{p} := (p_a)_{a \in \mathbb{N}^n_d} \in \mathbb{R}^{\binom{n+d}{n}}$.

A polynomial $s \in \mathbb{R}[x]_{2d}$ is a sum of squares (SOS) if it can be expressed as a finite sum $s = \sum_i p_i^2$ for some $p_i = \mathbf{p}_i^T \mathbf{b} \in \mathbb{R}[x]_d$. Hence

$$s = \sum_i (\mathbf{p}_i^T \mathbf{b})^2 = \sum_i \mathbf{b}^T \mathbf{p}_i \mathbf{p}_i^T \mathbf{b} = \mathbf{b}^T S \mathbf{b}$$

where $S := \sum_i \mathbf{p}_i \mathbf{p}_i^T \in \mathbb{S}^{\binom{n+d}{n}}$ is a positive semidefinite matrix called Gram matrix. Given s, finding whether it is SOS amounts to finding whether it has a positive semidefinite Gram matrix S whose entries are linearly related to the coefficient vector of s. In other words, the SOS cone is a projection of the semidefinite cone.

2.4 Exercise 1.4

2.4.1 Statement

Given polynomials $p_k \in \mathbb{R}[x]$, $k = 0, 1, \ldots, m$ defining our compact set $X := \{x \in \mathbb{R}^n : p_k(x) \geq 0, k = 1, \ldots, m\}$ with $p_0(x) = 1$, $p_1(x) = R^2 - x^T x$ and integers $r \geq d$, define the truncated quadratic module

$$Q(X)_r := \{q \in \mathbb{R}[x]_d : q = \sum_{k=0}^{m} s_k p_k, \;\; s_k \in \Sigma[x], \;\; s_k p_k \in \mathbb{R}[x]_r\}$$

as a projection of the SOS cone $\Sigma[x]$, where $\mathbb{R}[x]_r$ denotes the vector space of polynomials of degree up to r.

Describe explicitly the moment cone relaxation $Q(X)'_r$ as the projection of a spectrahedron.

2.4.2 Solution

Each polynomial $q \in \mathbb{R}[x]_d$ can be identified with its vector of coefficients $\mathbf{q} = (q_a)_a \in \mathbb{R}^{\binom{n+r}{n}}$ in the basis $\mathbf{b} \in \mathbb{R}[x]^{\binom{n+r}{n}}$, as in Exercise 1.3. By definition, the dual of the truncated quadratic module is

$$Q(X)'_r := \{\mathbf{y} : \ell_{\mathbf{y}}(q) \geq 0, \; \forall q \in Q(X)_r\} \subset \mathbb{R}^{\binom{n+r}{n}}$$

where the duality is the Riesz functional

$$\ell_{\mathbf{y}}(q) := \langle q, \mathbf{y} \rangle = \sum_a q_a y_a.$$

The dual can be explicitly constructed

$$\begin{aligned} Q(X)'_r &= \{\mathbf{y} : \ell_{\mathbf{y}}(\textstyle\sum_{k=0}^{m} s_k p_k) \geq 0, \; \forall s_k \in \Sigma[x], \; \deg(s_k p_k) \leq r, \; k = 0, 1, \ldots, m\} \\ &= \{\mathbf{y} : \ell_{\mathbf{y}}(p_k p^2) \geq 0, \; \forall p \in \mathbb{R}[x], \;\; \deg(s_k p^2) \leq r, \; k = 0, 1, \ldots, m\} \\ &= \{\mathbf{y} : M_r(p_k \mathbf{y}) \text{ positive semidefinite}, \;\; k = 0, 1, \ldots, m\} \end{aligned}$$

since positivity of the quadratic form $p \mapsto \ell_{\mathbf{y}}(p_k p^2)$ is equivalent to positive semidefiniteness of the matrix representing this quadratic form in the basis $\mathbf{b}$, namely

$$M_r(p_k \mathbf{y}) := \ell_{\mathbf{y}}(p_k \mathbf{b}\mathbf{b}^T)$$

if we let the Riesz functional act entrywise on matrices. The set of vectors $\mathbf{y}$ such that the symmetric linear matrix $M_r(p_k \mathbf{y})$ is positive semidefinite describes a spectrahedron. This is also the case for $Q(X)'_r$ since the intersection of finitely many spectrahedra is a spectrahedron.

Matrix $M_r(p_k \mathbf{y})$ is called a localizing matrix. When $k = 0$ and hence $p_0 = 1$, it is called a moment matrix. In that form they were introduced in [23, 24]. Their construction is comprehensively described in [22, 26].

2.5 Exercise 1.5

2.5.1 Statement

Consider the primal conic problem

$$\begin{array}{rcll} p_r^* & = & \inf_y & \langle p, y\rangle \\ & & \text{s.t.} & \langle 1, y\rangle = 1 \\ & & & y \in Q(X)'_r \end{array}$$

and its dual conic problem

$$\begin{array}{rcll} d_r^* & = & \sup_{v\in\mathbb{R}} & v \\ & & \text{s.t.} & p - v \in Q(X)_r. \end{array}$$

Prove that strong duality holds: $p_r^* = d_r^*$.

2.5.2 Solution

In the case that X has non empty interior, it is easy to prove that the primal feasible set has an non-empty interior: just consider moment sequences of an atomic probability measure with enough atoms sampled in X. Strong duality follows then readily from Slater's qualification constraint. The case that X has empty interior is more tricky and requires some additional arguments. The proof relies on the assumption that $p_1(x) := R^2 - x^T x$, see [18] for details.

3 Questions and answers

Q: Any condition to get strong duality [in the moment-SOS relaxations of degree r] ?

A: Indeed there is a condition, the quadratic module $Q(X)_\infty$ must be Archimedean, i.e. containing the polynomial $R^2 - x^T x$ for R large enough. An easy way to ensure this is to enforce $g_1(x) := R^2 - x^T x$ for R large enough. This is without loss of generality since we assume that X is bounded. This is what is done in the solution to Exercise 1.5.

Q: So p_r^* should be a lower bound for p^* ?

A: Indeed, since we minimize over a larger set (a relaxation), in general p_r^* will be smaller than p^*.

Q: Could you show an example of Q_r for some small dimension?

A: This is tricky since the interesting cases are high dimensional. What can be done however is projecting Q'_r onto two-dimensional subspaces, each coordinate corresponding to a moment of a given degree. In the case of the set of moments of an invariant measure for the logistic map (a classical dynamical system), this was achieved in [11, Figure 2]. On this figure we see clearly that a two-dimensional projection of the moment relaxation of a linear slice of Q'_r becomes tighter for increasing values of r.

Q: How is the state of the theory of Lasserre hierarchy relaxations with respect to a possible application to convergent power series instead of polynomials?

A: Positivity certificates (Positivstellensäetze) for functions in a finitely generated algebra of functions, in particular semi-algebraic functions, were investigated in [27]. Psäetze are also available for classes of rational functions [9]. Examples of algebras of non-semi algebraic functions, approximated by power series, are given in [27, Section 2.2]. There may be broader generalizations, but then the question is whether they can be more useful that semi-algebraic functions.

Q: Out of the generic case (and maybe also in the generic case), is it possible to decide a priori an optimal bound for r so the computations are tractable and the approximations are good enough?

A: There are upper bounds on r in the case of combinatorial optimization (POP on a finite set X, i.e. a union of points), but they are not useful because the number of variables in the description of $Q(X)_r$ is an exponential function of n, the number of variables of the original POP. So in pratice, we solve the relaxations for small, increasing values of the relaxation degree r, and we check sufficient conditions for exactness of the relaxations, based on flat extensions of moment matrices, see e.g. [15] or [22, Chapters 5 and 6].

Q. If the constraints of the optimisation problem are rational functions, is there a way to relax the problem to a SDP problem?

A. Usually the set X is defined by polynomial constraints. If you have rational constraints, you can translate them into polynomial constraints. If the objective function is rational instead of polynomial, you can adapt the moment-SOS hierarchy, see e.g. [6] in the case of sparse rational optimization.

Q. What kind of convergence results (if any) are available if the semialgebraic set X is not compact?

A. Usually X is closed. If X is not bounded, then some additional regularity assumptions are required on the sequence of moments. Convergence of the hierarchy is ensured if the moments satisfy the Carleman condition [25, Section 3.4.2]. Roughly speaking, it means that the sequence of moments should not grow like $\exp(2d)$ or faster.

Q. When does this approach make sense as opposed to techniques using cylindrical algebraic decomposition?

A. Cylindrical algebraic decomposition or other computer algebra (symbolic computation) algorithms solve POP exactly, using integer arithmetic. The solution of the POP is encoded as algebraic numbers, i.e. roots of (high degree) univariate polynomials with (large) integer coefficients. This is typically very costly. The moment-SOS approach to solving POP relies on semidefinite programming, which is efficiently implemented in floating point arithmetic. It means that the computed solution is an approximation of the exact solution. One can solve semidefinite programs exactly using computer algebra algorithms (critical points methods), see e.g. [17] but this is also typically very costly. For a comparison of moment-SOS and computer algebra (Groëbner basis) algorithms for solving a POP arising in electrical engineering, see e.g. [10].

Q. How do these techniques scale, especially to large numbers of variables (and possibly constraints), possibly assuming the constraints are not very "difficult" (maybe have some common structure)?

A. The number of variables in the moment-SOS relaxations grows polynomially in the relaxation order r, but the exponent is n, the number of variables in the POP. For e.g. $n = 10$ only the few smallest relaxations can be solved. It is then essential to exploit the problem structure (sparsity, symmetry).

Q. Will we see some examples in the next lectures?

A. There will be indeed some simple examples in the next two lectures. Benchmark examples of continuous and discrete optimization problems are reported in [13]. Additional benchmark examples of polynomial systems of equations can be found in [14]. For a challenging benchmark from electrical engineering, with a comparison of the moment-SOS and computer algebra algorithms, see e.g. [10].

Q. Otherwise which are some references for the case of rational function constraints? Is there a software that does the SDP relaxation of this case?

A. Rational constraints can be reformulated as polynomial constraints. For references on the moment-SOS hierarchy for rational optimization, see e.g. the introduction of [6]. GloptiPoly can easily deal with rational functions and sparsity, as explained in [6, Section 4.2].

Q. If there is a group G acting on K and the cost function is invariant under G, can we use this information to make things easier?

A. Yes indeed, in the context of moment relaxations for POP, this program was initiated in [32]. There is currently a lot of activity in symmetry reduction for POP, especially within the POEMA network.

Q. Is it in some cases possible to combine problem sparsity with the symmetry approach [..], to reduce the problem even further

A. Yes in principle. For recent references on exploiting sparsity in POP see e.g. [34].

Q. In GloptiPoly, how is the absence of a user defined ball constraint handled ?

A. It is not handled, it is the user's responsibility to add a ball constraint. For a better scaling of the SDP relaxations, it is recommended that the radius of the ball is close to unity. Moreover, as shown in [18], adding a ball constraint ensures strong duality in the SDP relaxations. For numerical evidence that the ball constraint is required, see e.g. [15, Section 4] and [6, Section 4.1.2].

Q. Is the work on symmetric polynomials (i.e. replacing x with y and y with x gives the same polynomial), which may strongly reduce the space of polynomials to consider ?

A. Exploiting symmetry of the SDP relaxations indeed allows to solve problems that would be otherwise out of reach.

Q. Other than Lasserre hierarchy which are other common approaches to POP ?

A. Global optimization techniques (e.g. branch and bound schemes) can be applied in principle. Local optimization (e.g. Newton's method) can be applied as well, and combined with the moment-SOS hierarchy as follows. For minimizing a polynomial, local optimization returns a valid upper bound, whereas the moment-SOS hierarchy returns a valid lower bound at each relaxation order. If the upper bound matches closely with the upper bound, there is no need to refine Newton's method (e.g. by choosing different initial conditions) and there is no need to go deeper in the hierarchy. In the context of optimal control (covered during the third lectures of the series) applied to a problem in data science, this was achieved in [4] - a local solution is obtained with Pontryagin's Maximum Principle, and its global optimality is certified by the moment-SOS hierarchy and Hamilton-Jacobi-Bellman inequalities.

Q. Is there a general estimate for the necessary radius bound R, or just for specific problems?

A. Usually the radius bound R is known in applications. Physically relevant POPs have bounded variables. If nothing is known about the geometry of the feasibility set X, finding a bound on its radius, or even deciding whether X is empty or not, can be difficult as it reduces to certifying polynomial non-negativity. For example, if X is a spectrahedron, deciding emptiness is a difficult problem that can be solved with the moment-SOS hierarchy [19].

References

[1] A. Barvinok. A course in convexity. AMS, 2002.

[2] A. Ben-Tal, A. Nemirovski. Lectures on modern convex optimization. SIAM, 2001.

[3] G. Blekherman, P. A. Parrilo, R. R. Thomas. Semidefinite optimization and convex algebraic geometry. SIAM, 2013.

[4] B. Bonnard, M. Claeys, O. Cots, P. Martinon. Geometric and numerical methods in the contrast imaging problem in nuclear magnetic resonance. Acta Appl. Math. 135(1):5-45, 2015.

[5] S. Boyd, L. Vandenberghe. Convex optimization. Cambridge University Press, 2004.

[6] F. Bugarin, D. Henrion, J. B. Lasserre. Minimizing the sum of many rational functions. Mathematical Programming Computation, 8(1):83-111, 2016.

[7] R. E. Curto, L. A. Fialkow. Recursiveness, positivity, and truncated moment problems. Houston J. Math. 17:603-635, 1991.

[8] R. E. Curto, L. A. Fialkow. The truncated complex K-moment problem. Trans. Amer. Math. Soc. 352:2825-2855, 2000.

[9] G. Fichou, R. Quarez, J.-P. Monnier. Continuous functions on the plane regular after one blowing-up. Mathematische Zeitschrift, 285(1):287-323, 2017.

[10] S. Galeani, D. Henrion, A. Jacquemard, L. Zaccarian. Design of Marx generators as a structured eigenvalue assignment. Automatica 50(10):2709-2717, 2014.

[11] D. Henrion. Semidefinite characterisation of invariant measures for one-dimensional discrete dynamical systems. Kybernetika, 48(6):1089-1099, 2012.

[12] D. Henrion. Optimization on linear matrix inequalities for polynomial systems control. Les cours du C.I.R.M. 3(1):1-44, 2013.

[13] D. Henrion, J. B. Lasserre. GloptiPoly: global optimization over polynomials with Matlab and SeDuMi. ACM Transactions Math. Soft. 29:165-194, 2003.

[14] D. Henrion, J. B. Lasserre. Solving global Optimization Problems over Polynomials with GloptiPoly 2.1. In C. Bliek, C. Jermann, A. Neumaier (Editors). Global Optimization and Constraint Satisfaction. Lecture Notes on Computer Science 2861, Springer, 2003.

[15] D. Henrion, J. B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In D. Henrion, A. Garulli (Editors). Positive Polynomials in Control. Lecture Notes on Control and Information Sciences 312, Springer, 2005.

[16] D. Henrion, J. B. Lasserre, J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software 24(5):761-779, 2009.

[17] D. Henrion, S. Naldi, M. Safey El Din. Exact algorithms for linear matrix inequalities. SIAM J. Optim. 26(4):2512-2539, 2016.

[18] C. Josz, D. Henrion. Strong duality in Lasserre's hierarchy for polynomial optimization. Optimization Letters 1(10):3-10, 2016.

[19] I. Klep, M. Schweighofer. An exact duality theory for semidefinite programming based on sums of squares. Math. Oper. Res. 38(3):569–590, 2013.

[20] M. Korda. Moment-sum-of-squares hierarchies for set approximation and optimal control. PhD thesis, EPFL, 2016.

[21] M. Korda, D. Henrion, J. B. Lasserre. The moment-SOS hierarchy. World Scientific, 2020.

[22] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar, S. Sullivant (Editors). Emerging Applications of Algebraic Geometry. IMA Volumes in Mathematics and its Applications 149, Springer, 2009.

[23] J. B. Lasserre. Optimisation globale et théorie des moments. Comptes-rendus de l'académie des sciences, Série 1 - mathématiques, 331(11):929-934, 2000.

[24] J. B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. Optimization 11(3):796-817, 2001.

[25] J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, 2010.

[26] J. B. Lasserre. Introduction to polynomial and semi-algebraic optimization. Cambridge University Press, 2015.

[27] J. B. Lasserre, M. Putinar. Positivity and Optimization for Semi-Algebraic Functions. SIAM J. Optim. 20(6):3364–3383, 2010.

[28] D. Luenberger. Optimization by vector space methods. John Wiley and Sons, 1969.

[29] M. Marshall. Positive polynomials and sums of squares. AMS, 2008.

[30] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. Mathematical Programming, Ser. A, 146(1-2):97-121, 2014.

[31] M. Putinar. Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal 42:969-984, 1993.

[32] C. Riener, T. Theobald, A. L. Jansson, J. B. Lasserre. Exploiting symmetries in SDP relaxations for polynomial optimization. Math. Oper. Res. 38(1):122–141, 2013.

[33] K. Schmüdgen. The moment problem. Springer, 2017.

[34] T. Weisser, J. B. Lasserre, K. C. Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. Mathematical Programming Computation 10:1-32, 2018.