Strong duality in Lasserre's hierarchy for polynomial optimization

Cédric Josz, Didier Henrion

Strong duality in Lasserre's hierarchy for polynomial optimization

Cédric Josz[1,2], Didier Henrion[3,4,5]

Draft of December 1, 2014

Abstract

A polynomial optimization problem (POP) consists of minimizing a multivariate real polynomial on a semi-algebraic set K described by polynomial inequalities and equations. In its full generality it is a non-convex, multi-extremal, difficult global optimization problem. More than an decade ago, J. B. Lasserre proposed to solve POPs by a hierarchy of convex semidefinite programming (SDP) relaxations of increasing size. Each problem in the hierarchy has a primal SDP formulation (a relaxation of a moment problem) and a dual SDP formulation (a sum-of-squares representation of a polynomial Lagrangian of the POP). In this note, when the POP feasibility set K is compact, we show that there is no duality gap between each primal and dual SDP problem in Lasserre's hierarchy, provided a redundant ball constraint is added to the description of set K. Our proof uses elementary results on SDP duality, and it does not assume that K has an interior point.

1 Introduction

Consider the following polynomial optimization problem (POP)

$$\begin{array}{ll} \inf_x & f(x) := \sum_\alpha f_\alpha x^\alpha \\ \text{s.t.} & g_i(x) := \sum_\alpha g_{i,\alpha} x^\alpha \geq 0, \quad i = 1, \ldots, m \end{array} \tag{1}$$

where we use the multi-index notation $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for $x \in \mathbb{R}^n$, $\alpha \in \mathbb{N}^n$ and where the data are polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[x]$ so that in the above sums only a finite number of coefficients f_α and $g_{i,\alpha}$ are nonzero. Let K denote its feasibility set:

$$K := \{x \in \mathbb{R}^n \; : \; g_i(x) \geq 0, \; i = 1, \ldots, m\}$$

[1]French transmission system operator Réseau de Transport d'Electricité (RTE), 9 rue de la Porte de Buc, BP 561, F-78000 Versailles, France.

[2]INRIA Paris-Rocquencourt, BP 105, F-78153 Le Chesnay, France. cedric.josz@inria.fr

[3]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France. henrion@laas.fr

[4]Université de Toulouse, LAAS, F-31400 Toulouse, France.

[5]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

To solve POP (1), Lasserre [3, 4] proposed a semidefinite programming (SDP) relaxation hierarchy with guaranteed asymptotic global convergence provided an algebraic assumption holds:

Assumption 1 *There exists a polynomial* $u \in \mathbb{R}[x]$ *such that* $\{x \in \mathbb{R}^n : u(x) \geq 0\}$ *is bounded and* $u = u_0 + \sum_{i=1}^{m} u_i g_i$ *where polynomials* $u_i \in \mathbb{R}[x]$, $i = 0, 1, \ldots, m$ *are sums of squares (SOS) of other polynomials.*

Nie *et al.* [5] have proven that Assumption 1 also implies generically finite convergence, that is to say that for almost every instance of POP, there exists a finite-dimensional SDP relaxation in the hierarchy whose optimal value is equal to the optimal value of the POP. Assumption 1 can be difficult to check computationally (as the degrees of the SOS multipliers can be arbitrarily large), and it is often replaced by the following slightly stronger assumption:

Assumption 2 *The description of* K *contains a ball constraint, say* $g_m(x) = R^2 - \sum_{i=1}^{n} x_i^2$ *for some real number* R.

Indeed, under Assumption 2, simply choose $u = g_m$, $u_1 = \cdots = u_{m-1} = 0$, and $u_m = 1$ to conclude that Assumption 1 holds as well. In practice, it is often easy to see to it that Assumption 2 holds. In the case of a POP with a bounded feasible set, a redundant ball constraint can be added.

More generally, if the intersection of the sublevel set $\{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ with the feasible set of the POP is bounded for some feasible point x_0, then a redundant ball constraint can also be added. As an illustration, a reviewer suggested the example of the minimization of $f(x) = x_1^2 + x_2^2 - 3x_1x_2$ on the unbounded set defined on $\mathbb{R}^2$ by the constraint $g_1(x) = 1 - 3x_1x_2 \geq 0$. The intersection of the feasible set with the set defined by the constraint $f(x) \leq f(0) = 0$ is included in the ball defined by $g_2(x) = 1 - x_1^2 - x_2^2 \geq 0$ so that the POP can be equivalently defined on the bounded set $K = \{x \in \mathbb{R}^2 : g_1(x) \geq 0, \ g_2(x) \geq 0\}$.

Each problem in Lasserre's hierarchy consists of a primal-dual SDP pair, called SDP relaxation, where the primal corresponds to a convex moment relaxation of the original (typically nonconvex) POP, and the dual corresponds to a SOS representation of a polynomial Lagrangian of the POP. The question arises of whether the duality gap vanishes in each SDP relaxation. This is of practical importance because numerical algorithms to solve SDP problems are guaranteed to converge only where there is a zero duality gap, and sometimes under the stronger assumption that there is a primal or/and dual SDP interior point.

In [6, Example 4.9], Schweighofer provides a two-dimensional POP with no interior point for which Assumption 1 holds, yet a duality gap exists at the first SDP relaxation: $\inf x_1x_2$ s.t. $x \in K = \{x \in \mathbb{R}^2 : -1 \leq x_1 \leq 1, \ x_2^2 \leq 0\}$, with primal SDP value equal to zero and dual SDP value equal to minus infinity. This shows that a stronger assumption is required to ensure a zero SDP duality gap. A sufficient condition for strong duality has been given in [4]: set K should contain an interior point. However, this may be too restrictive: in the proof of Lemma 1 in [2] the authors use notationally awkward arguments involving truncated moment matrices to prove the absence of SDP duality gap for a certain set K that contains no interior point. This shows that the existence of an interior point is not necessary for a zero SDP duality gap. More generally, it is not possible to assume the existence of an interior point for POPs with explicit equality constraints, and a weaker assumption for zero SDP duality gap is welcome.

Motivated by these observations, in this note we prove that under the basic Assumption 2 on the description of set K, there is no duality gap in the SDP hierarchy. Our interpretation of this result, and the main message of this contribution, is that in the context of Lasserre's hierarchy for POP, a practically relevant description of a bounded semialgebraic feasibility set must include a redundant ball constraint.

2 Proof of strong duality

For notational convenience, let $g_0(x) = 1 \in \mathbb{R}[x]$ denote the unit polynomial. Define the localizing matrix

$$M_{d-d_i}(g_i y) := \left(\sum_{\gamma} g_{i,\gamma} y_{\alpha+\beta+\gamma}\right)_{|\alpha|,|\beta|\leq d-d_i} = \sum_{|\alpha|\leq 2d} A_{i,\alpha} y_\alpha$$

where d_i is the smallest integer greater than or equal to half the degree of g_i, for $i = 0, 1, \ldots, m$, and $|\alpha| = \sum_{i=1}^n \alpha_i$. The Lasserre hierarchy for POP (1) consists of a primal moment SDP problem

$$(P_d) \ : \ \begin{array}{ll} \inf_y & \sum_\alpha f_\alpha y_\alpha \\ \text{s.t.} & y_0 = 1 \\ & M_{d-d_i}(g_i y) \succeq 0, \ i = 0, 1, \ldots, m \end{array}$$

and a dual SOS SDP problem

$$(D_d) \ : \ \begin{array}{ll} \sup_{z,Z} & z \\ \text{s.t.} & f_0 - z = \sum_{i=0}^m \langle A_{i,0}, Z_i \rangle \\ & f_\alpha = \sum_{i=0}^m \langle A_{i,\alpha}, Z_i \rangle, \quad 0 < |\alpha| \leq 2d \\ & Z_i \succeq 0, \ i = 0, 1, \ldots, m, \quad z \in \mathbb{R} \end{array}$$

where $A \succeq 0$ stands for matrix A positive semidefinite, $\langle A, B \rangle = \text{trace}\, AB$ is the inner product between two matrices. The Lasserre hierarchy is indexed by an integer $d \geq d_{\min} := \max_{i=0,1,\ldots,m} d_i$. The primal-dual pair (P_d, D_d) is called the SDP relaxation of order d for POP (1). The size of the primal variable $(y_\alpha)_{|\alpha|\leq 2d}$ is $\left(\begin{array}{c} n+2d \\ n \end{array}\right)$ and the size of the dual variable Z_i is $\left(\begin{array}{c} n+d-d_i \\ n \end{array}\right)$.

Let us define the following sets:

- $\mathcal{P}_d$: feasible points for P_d;

- $\mathcal{D}_d$: feasible points for D_d;
- int $\mathcal{P}_d$: strictly feasible points for P_d;
- int $\mathcal{D}_d$: strictly feasible points for D_d;
- $\mathcal{P}_d^*$: optimal solutions for P_d;
- $\mathcal{D}_d^*$: optimal solutions for D_d.

Finally, let us denote by val P_d the infimum in problem P_d and by val D_d the supremum in problem D_d.

Lemma 1 int $\mathcal{P}_d$ *nonempty or* int $\mathcal{D}_d$ *nonempty implies* val $P_d =$ val D_d.

Lemma 1 is classical in convex optimization, and it is generally called Slater's condition, see e.g. [7, Theorem 4.1.3].

Lemma 2 *The two following statements are equivalent :*

1. *$\mathcal{P}_d$ is nonempty and* int $\mathcal{D}_d$ *is nonempty;*
2. *$\mathcal{P}_d^*$ is nonempty and bounded.*

A proof of Lemma 2 can be found in [8]. According to Lemmas 1 and 2, $\mathcal{P}_d^*$ nonempty and bounded implies strong duality. This result is also mentioned without proof at the end of [7, Section 4.1.2].

Lemma 3 *Under Assumption 2, set $\mathcal{P}_d$ is included in the Euclidean ball of radius*

$$\sqrt{\binom{n+d}{n}\sum_{k=0}^{d} R^{2k}}$$

centered at the origin.

Proof: If $\mathcal{P}_d = \emptyset$, the result is obvious. If not, consider a feasible point $(y_\alpha)_{|\alpha|\leq 2d} \in \mathcal{P}_d$. Let $k \in \{1,\dots,d\}$. In the SDP problem P_k, the localizing matrix associated to the ball constraint $g_m(x) = R^2 - \sum_{i=1}^n x_i^2 \geq 0$ reads

$$M_{k-1}(g_m y) = \left(\sum_\gamma g_{m,\gamma}\ y_{\alpha+\beta+\gamma} \right)_{|\alpha|,|\beta|\leq k-1}$$

with trace equal to

$$\begin{aligned}
\text{trace } M_{k-1}(g_m y) &= \textstyle\sum_{|\alpha|\leq k-1} \sum_\gamma g_{m,\gamma}\ y_{2\alpha+\gamma} \\
&= \textstyle\sum_{|\alpha|\leq k-1} \left(g_{m,0}\ y_{2\alpha} + \sum_{|\gamma|=1} g_{m,2\gamma}\ y_{2\alpha+2\gamma} \right) \\
&= \textstyle\sum_{|\alpha|\leq k-1} \left(R^2 y_{2\alpha} - \sum_{|\gamma|=1} y_{2(\alpha+\gamma)} \right) \\
&= \textstyle\sum_{|\alpha|\leq k-1} R^2 y_{2\alpha} - \sum_{|\alpha|\leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)} \\
&= R^2 (\textstyle\sum_{|\alpha|\leq k-1} y_{2\alpha}) + y_0 - \sum_{|\alpha|\leq k} y_{2\alpha} \\
&= R^2 \text{ trace } M_{k-1}(y) + 1 - \text{trace } M_k(y).
\end{aligned}$$

From the structure of the localizing matrix, it holds $M_{k-1}(g_m y) \succeq 0$ hence trace $M_{k-1}(g_m y) \ \geq\ 0$ and

$$\text{trace } M_k(y) \leq 1 + R^2 \text{ trace } M_{k-1}(y)$$

from which we derive

$$\text{trace } M_d(y) \leq \sum_{k=1}^{d} R^{2(k-1)} + R^{2d} \text{ trace } M_0(y) = \sum_{k=0}^{d} R^{2k}$$

since trace $M_0(y) = y_0 = 1$. The 2-norm $\|M_d(y)\|_2$, equal to the maximum eigenvalue of $M_d(y)$, is upper bounded by trace $M_d(y)$, the sum of the eigenvalues of $M_d(y)$, which are all nonnegative. Moreover the Frobenius norm satisfies

$$\begin{aligned}
\|M_d(y)\|_F^2 &:= \langle\ M_d(y)\ ,\ M_d(y)\ \rangle \\
&= \langle\ \textstyle\sum_{|\delta|\leq 2d} A_{0,\delta}\ y_\delta\ ,\ \sum_{|\delta|\leq 2d} A_{0,\delta}\ y_\delta\ \rangle \\
&= \textstyle\sum_{|\delta|\leq 2d}\ \langle A_{0,\delta}, A_{0,\delta}\rangle\ y_\delta^2 \quad \text{by orthogonality of matrices } (A_{0,\delta})_{|\delta|\leq 2d} \\
&\geq \textstyle\sum_{|\delta|\leq 2d}\ y_\delta^2 \quad \text{because } \langle A_{0,\delta}, A_{0,\delta}\rangle \geq 1
\end{aligned}$$

where matrices $(A_{0,\delta})_{|\delta|\leq 2d}$ can be written using column vectors $(e_\alpha)_{|\alpha|\leq d}$, containing only zeros apart from the value 1 at index α, via the explicit formula

$$A_{0,\delta} = \sum_{\substack{\alpha+\beta=\delta \\ |\alpha|,|\beta|\leq d}} e_\alpha e_\beta^T.$$

The proof follows then from

$$\sqrt{\sum_{|\delta|\leq 2d} y_\delta^2} \leq \|M_d(y)\|_F \leq \sqrt{\left(\begin{array}{c} n+d \\ n \end{array}\right)} \|M_d(y)\|_2 \leq \sqrt{\left(\begin{array}{c} n+d \\ n \end{array}\right)} \sum_{k=0}^{d} R^{2k}.$$

$\square$

Theorem 1 *Assumption 2 implies that* $-\infty < \mathrm{val}\, P_d = \mathrm{val}\, D_d$ *for all* $d \geq d_{\min}$.

Proof: Let $d \geq d_{\min}$. Firstly, let us consider the case when $\mathcal{P}_d$ is nonempty. According to Lemma 3, $\mathcal{P}_d$ is bounded and closed, and the objective function in P_d is linear, so we conclude that $\mathcal{P}_d^*$ is nonempty and bounded. According to Lemma 2, $\mathrm{int}\, \mathcal{D}_d$ is nonempty, and from Lemma 1, $\mathrm{val}\, P_d = \mathrm{val}\, D_d$.

Secondly, let us consider the case when $\mathcal{P}_d$ is empty. An infeasible SDP problem can be either weakly infeasible or strongly infeasible, see [10, Section 5.2] for definitions. Let us prove by contradiction that P_d cannot be weakly infeasible. If P_d is weakly infeasible, there must exist a sequence $(y^p)_{p \in \mathbb{N}}$ such that

$$\forall p \in \mathbb{N}\ , \quad \left\{ \begin{array}{l} 1 - \frac{1}{p+1} \leq y_0^p \leq 1 + \frac{1}{p+1} \\ \lambda_{\min}(M_{d-d_i}(g_i y^p)) \geq -\frac{1}{p+1},\ i = 0, 1, \ldots, m \end{array} \right.$$

where $\lambda_{\min}$ denotes the minimum eigenvalue of a symmetric matrix. According to the proof of Lemma 3, for all $1 \leq k \leq d$ and all real numbers $(y_\alpha)_{|\alpha| \leq 2d}$, one has

$$\mathrm{trace}\, M_{k-1}(g_m y) = R^2\, \mathrm{trace}\, M_{k-1}(y) + y_0 - \mathrm{trace}\, M_k(y).$$

Clearly, $\mathrm{trace}\, M_{k-1}(g_m y) \geq -\frac{c}{1+p}$ where $c := \left(\begin{array}{c} n+d \\ n \end{array} \right)$ denotes the size of the moment matrix $M_d(y)$. The following holds

$$\mathrm{trace}\, M_k(y^p) \leq R^2\, \mathrm{trace}\, M_{k-1}(y^p) + 1 + \frac{1+c}{1+p}$$

from which we derive

$$\mathrm{trace}\, M_d(y^p) \leq (1 + \frac{1+c}{1+p}) \sum_{k=0}^{d} R^{2k}.$$

Together with $\lambda_{\min}(M_d(y^p)) \geq -\frac{1}{1+p}$, this yields

$$\lambda_{\max}(M_d(y^p)) \leq (1 + \frac{1+c}{1+p}) \sum_{k=0}^{d} R^{2k} + \frac{c-1}{1+p}$$

where $\lambda_{\max}$ denotes the minimum eigenvalue of a symmetric matrix. Hence for all $p \in \mathbb{N}$, the spectrum of the moment matrix $M_d(y^p)$ is lower bounded by $l := -1$ and upper bounded by $u := (2+c) \sum_{k=0}^{d} R^{2k} + c - 1$. Therefore:

$$\sqrt{\sum_{|\delta| \leq 2d} (y_\delta^p)^2} \leq \|M_d(y^p)\|_F \leq \sqrt{c}\, \max(|l|, |u|)$$

The sequence $(y^p)_{p \in \mathbb{N}}$ is hence included in a compact set. Thus there exists a subsequence which converges towards $y^{\lim}$ such that $y_0^{\lim} = 1$ and $\lambda_{\min}(M_{d-d_i}(g_i y^{\lim})) \geq 0$, $i = 0, 1, \ldots, m$. The limit $y^{\lim}$ is thus included is $\mathcal{P}_d$, which is a contradiction.

SDP problem P_d is strongly infeasible which means that its dual problem D_d has an improving ray [10, Definition 5.2.2]. To conclude that val $D_d = +\infty$, all that is left to prove is that $\mathcal{D}_d \neq \emptyset$. Consider the primal problem P_d discarding all constraints but $y_0 = 1$, $M_d(y) \succcurlyeq 0$, and $M_{d-1}(g_m y) \succcurlyeq 0$. It is a feasible and bounded SDP problem owing to Lemma 3. According to Lemma 2, its dual problem must contain a feasible point (z, Z_0, Z_m) and hence $(z, Z_0, 0, \ldots, 0, Z_m) \in \mathcal{D}_d$.

□

3 Conclusion

We prove that strong duality always holds in Lasserre's SDP hierarchy for POPs whose description of the feasible set contains a ball constraint. To preclude numerical troubles with SDP solvers, we advise to systematically add a ball constraint, combined with an appropriate scaling so that all scaled variables belong to the unit sphere. Without scaling, numerical troubles can occur as well, but they are not due to the presence of a duality gap, see [1] and also the example of [9, Section 6].

Acknowledgements

The authors would like to thank the anonymous reviewer for insightful comments.

References

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

[2] D. Henrion, J. B. Lasserre. Inner approximations for polynomial matrix inequalities and robust stability regions. *IEEE Transactions on Automatic Control*, 57(6):1456–1467, (2012)

[3] J. B. Lasserre. Optimisation globale et théorie des moments. *C. R. Acad. Sci. Paris, Série I*, 331:929–934, (2000)

[4] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, (2001)

[5] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Mathematical Programming*, Ser. A, to appear, (2014)

[6] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization*, 15:805–825, (2005)

[7] A. Shapiro, K. Scheinberg. *Duality, optimality conditions and perturbation analysis.* In: H. Wolkowicz, R. Saigal, L. Vandenberghe (eds.). *Handbook of Semidefinite Programming. Theory, Algorithms, and Applications.* Kluwer Academic Publishers, Boston, (2000)

[8] M. Trnovská. Strong duality conditions in semidefinite programming. *Journal of Electrical Engineering*, 56:1–5, (2005)

[9] H. Waki, M. Nakata, M. Muramatsu. Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization. *Computational Optimization and Applications*, 53(3):823–844, (2012)

[10] E. de Klerk, T. Terlaky, K. Roos. *Self-Dual Embeddings.* In: H. Wolkowicz, R. Saigal, L. Vandenberghe (eds.). *Handbook of Semidefinite Programming. Theory, Algorithms, and Applications.* Kluwer Academic Publishers, Boston, (2000)