A SUM OF SQUARES APPROXIMATION OF NONNEGATIVE POLYNOMIALS

Jean B. Lasserre

LAAS-CNRS
7 Avenue du Colonel Roche, 31077 Toulouse, France

Abstract: We show that every real nonnegative polynomial f can be approximated as closely as desired by a sequence of polynomials $\{f_\epsilon\}$ that are sums of squares. Each f_ϵ has a simple and explicit form in terms of f and ϵ.

Keywords: Nonnegative polynomials, sums of squares.

1. INTRODUCTION

The study of relationships between *nonnegative* and *sums of squares* (s.o.s.) polynomials, initiated by Hilbert, is of real practical importance in view of numerous potential applications, notably in polynomial programming. Indeed, checking whether a given polynomial is nonnegative is a NP-hard problem whereas checking it is s.o.s. reduces to solving a (convex) Semidefinite Programming (SDP) problem for which efficient algorithms are now available. For instance, recent results in real algebraic geometry, most notably in [Schmüdgen, 1991], [Putinar, 1993], [Jacobi and Prestel, 2001] have provided s.o.s. representations of polynomials, positive on a compact semialgebraic set; the interested reader is referred to [Prestel and Delzell, 2001], and [Scheiderer, 2003] for a nice account of such results. This in turn has permitted to develop efficient SDP-relaxations in polynomial optimization; see e.g. [Lasserre, 2001, 2002], [Parrilo, 2003], [Schweighofer, 2004], and the many references therein. See also [Henrion and Lasserre, 2004] for control applications.

So, back to a comparison between nonnegative and s.o.s. polynomials, on the negative side, [Blekherman, 2004] has shown that if the degree is *fixed*, then the cone of nonnegative polynomials is much *larger* than that of s.o.s. However, on the positive side, a denseness result states that the cone of s.o.s. polynomials is *dense* in the space of polynomials that are nonnegative on $[-1,1]^n$ (for the norm $\|f\|_1 = \sum_\alpha |f_\alpha|$ whenever f is written $\sum_\alpha f_\alpha x^\alpha$ in the usual canonical basis); see e.g. Theorem 5, p. 122 in [Berg, 1980].

Contribution. We show that *every nonnegative* polynomial f is almost a s.o.s., namely we show that f can be approximated by a sequence of s.o.s. polynomials $\{f_\epsilon\}_\epsilon$, in the specific form

$$f_\epsilon \;=\; f + \epsilon \sum_{k=0}^{r_\epsilon} \sum_{j=1}^{n} \frac{x_j^{2k}}{k!}, \tag{1}$$

for some $r_\epsilon \in \mathbb{N}$, so that $\|f - f_\epsilon\|_1 \to 0$ as $\epsilon \downarrow 0$.

This result is in the spirit of the previous denseness result. However we here provide in (1) an *explicit* converging approximation with a very specific (and simple) form; namely it suffices to slightly perturbate f by adding a small coefficient $\epsilon > 0$ to each square monomial x_i^{2k} for all $i = 1, \ldots, n$ and all $k = 1, \ldots, r$, with r sufficiently large.

To prove this result we combine

- (generalized) **Carleman**'s sufficient condition for a moment sequence $\mathbf{y} = \{y_\alpha\}$ to have a *representing measure* μ (i.e., such that $y_\alpha = \int x^\alpha d\mu$ for all $\alpha \in \mathbb{N}^n$), and

- a **duality** result from convex optimization.

As a consequence, we may thus define a procedure to approximate the global minimum of a polynomial f. It consists in solving a sequence of SDP-relaxations which are simpler and easier to solve than those defined in [Lasserre, 2001].

2. NOTATION AND DEFINITIONS

For a real symmetric matrix A, the notation $A \succeq 0$ (resp. $A \succ 0$) stands for A positive semidefinite (resp. positive definite). The sup-norm $\sup_j |x_j|$ of a vector $x \in \mathbb{R}^n$, is denoted by $\|x\|_\infty$. Let $\mathbb{R}[x_1, \ldots, x_n]$ be the ring of real polynomials, and let

$$v_r(x) := (1, x_1, \ldots x_n, x_1^2, x_1 x_2, \ldots, x_n^r) \qquad (2)$$

be the canonical basis for the $\mathbb{R}$-vector space $\mathcal{A}_r$ of real polynomials of degree at most r, and let $s(r)$ be its dimension. Similarly, $v_\infty(x)$ denotes the canonical basis of $\mathbb{R}[x_1, \ldots, x_n]$ as a $\mathbb{R}$-vector space, denoted $\mathcal{A}$. So a vector in $\mathcal{A}$ has always *finitely* many zeros.

Therefore, a polynomial $p \in \mathcal{A}_r$ is written

$$x \mapsto p(x) \;=\; \sum_\alpha p_\alpha x^\alpha \;=\; \langle \mathbf{p}, v_r(x) \rangle, \qquad x \in \mathbb{R}^n,$$

(where $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$) for some vector $\mathbf{p} = \{p_\alpha\} \in \mathbb{R}^{s(r)}$, the vector of coefficients of p in the basis (2).

Extending $\mathbf{p}$ with zeros, we can also consider $\mathbf{p}$ as a vector indexed in the basis $v_\infty(x)$ (i.e. $\mathbf{p} \in \mathcal{A}$). If we equip $\mathcal{A}$ with the usual scalar product $\langle ., . \rangle$ of vectors, then for every $p \in \mathcal{A}$,

$$p(x) \;=\; \sum_{\alpha > \in \mathbb{N}^n} p_\alpha x^\alpha \;=\; \langle \mathbf{p}, v_\infty(x) \rangle, \qquad x \in \mathbb{R}^n.$$

Given a sequence $\mathbf{y} = \{y_\alpha\}$ indexed in the basis $v_\infty(x)$, let $L_\mathbf{y} : \mathcal{A} \to \mathbb{R}$ be the linear functional

$$p \mapsto L_\mathbf{y}(p) \;:=\; \sum_{\alpha \in \mathbb{N}^n} p_\alpha y_\alpha \;=\; \langle \mathbf{p}, \mathbf{y} \rangle.$$

Given a sequence $\mathbf{y} = \{y_\alpha\}$ indexed in the basis $v_\infty(x)$, the *moment* matrix $M_r(\mathbf{y}) \in \mathbb{R}^{s(r) \times s(r)}$ with rows and columns indexed in the basis $v_r(x)$ in (2), satisfies

$$[M_r(\mathbf{y})(1,j) \;=\; y_\alpha \;\text{ and }\; M_r(y)(i,1) \;=\; y_\beta] \\ \Rightarrow\; M_r(y)(i,j) \;=\; y_{\alpha+\beta}.$$

For instance, with $n = 2$,

$$M_2(\mathbf{y}) \;=\; \begin{bmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix}.$$

A sequence $\mathbf{y} = \{y_\alpha\}$ has a *representing* measure $\mu_\mathbf{y}$ if

$$y_\alpha \;=\; \int_{\mathbb{R}^n} x^\alpha \, d\mu_\mathbf{y}, \qquad \forall\, \alpha \in \mathbb{N}^n. \qquad (3)$$

In this case one also says that $\mathbf{y}$ is a *moment sequence*. In addition, if $\mu_\mathbf{y}$ is unique then $\mathbf{y}$ is said to be a *determinate* moment sequence.

The matrix $M_r(\mathbf{y})$ defines a bilinear form $\langle ., . \rangle_\mathbf{y}$ on $\mathcal{A}_r$, by

$$\langle q, p \rangle_\mathbf{y} \;:=\; \langle \mathbf{q}, M_r(\mathbf{y})\mathbf{p} \rangle \;=\; L_\mathbf{y}(qp), \quad q, p \in \mathcal{A}_r,$$

and if $\mathbf{y}$ has a *representing* measure $\mu_\mathbf{y}$ then

$$\langle \mathbf{q}, M_r(\mathbf{y})\mathbf{q} \rangle \;=\; \int_{\mathbb{R}^n} q(x)^2 \, \mu_\mathbf{y}(dx) \;\geq\; 0, \qquad (4)$$

so that $M_r(\mathbf{y}) \succeq 0$.

Next, given a sequence $\mathbf{y} = \{y_\alpha\}$ indexed in the basis $v_\infty(x)$, let $y_{2k}^{(i)} := L_\mathbf{y}(x_i^{2k})$ for every $i = 1, \ldots, n$ and every $k \in \mathbb{N}$. That is, $y_{2k}^{(i)}$ denotes the element in the sequence $\mathbf{y}$, corresponding to the monomial x_i^{2k}.

Of course not every sequence $\mathbf{y} = \{y_\alpha\}$ has a representing measure $\mu_\mathbf{y}$ as in (3). However, there exists a *sufficient* condition to ensure that it is the case. The following result stated in [Berg, 1980] is from [Nussbaum, 1966], and is re-stated here, with our notation.

Theorem 1. Let $\mathbf{y} = \{y_\alpha\}$ be an infinite sequence such that $M_r(\mathbf{y}) \succeq 0$ for all $r = 0, 1, \ldots$. If

$$\sum_{k=0}^{\infty} (\mathbf{y}_{2k}^{(i)})^{-1/2k} \;=\; \infty, \qquad i = 1, \ldots, n, \qquad (5)$$

then $\mathbf{y}$ is a determinate moment sequence.

The condition (5) in Theorem 1 is called *Carleman's condition* as it extends to the multivariate case the original Carleman's sufficient condition given for the univariate case.

3. PRELIMINARIES

Let B_M be the closed ball

$$B_M \;=\; \{x \in \mathbb{R}^n \mid \; \|x\|_\infty \;\leq\; M\}. \qquad (6)$$

Proposition 2. Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be such that $-\infty \;<\; f^* \;:=\; \inf_x f(x)$. Then, for every $\epsilon > 0$ there is some $M_\epsilon \in \mathbb{N}$ such that

$$f_M^* \;:=\; \inf_{x \in B_M} f(x) \;<\; f^* + \epsilon, \qquad \forall M \;\geq\; M_\epsilon.$$

Equivalently, $f^*_M \downarrow f^*$ as $M \to \infty$.

PROOF. Suppose it is false. That is, there is some $\epsilon_0 > 0$ and an infinite sequence sequence $\{M_k\} \subset \mathbb{N}$, with $M_k \to \infty$, such that $f^*_{M_k} \geq f^* + \epsilon_0$ for all k. But let $x_0 \in \mathbb{R}^n$ be such that $f(x_0) < f^* + \epsilon_0$. With any $M_k \geq \|x_0\|_\infty$, one obtains the contradiction $f^* + \epsilon_0 \leq f^*_{M_k} \leq f(x_0) < f^* - \epsilon_0$.

To prove our main result (Theorem 5 below), we first introduce the following related optimization problems.

$$\mathbb{P}: \qquad f^* := \inf_{x \in \mathbb{R}^n} f(x), \tag{7}$$

and for $0 < M \in \mathbb{N}$, the problem $\mathcal{P}_M$

$$\inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left\{ \int f\, d\mu \,|\, \quad \int \sum_{i=1}^n \mathrm{e}^{\mathrm{x_i^2}}\, \mathrm{d}\mu \leq \mathrm{n}\mathrm{e}^{\mathrm{M}^2} \right\}, \tag{8}$$

where $\mathcal{P}(\mathbb{R}^n)$ is the space of probability measures on $\mathbb{R}^n$. The respective optimal values of $\mathbb{P}$ and $\mathcal{P}_M$ are denoted $\inf \mathbb{P} = f^*$ and $\inf \mathcal{P}_M$, or $\min \mathbb{P}$ and $\min \mathcal{P}_M$ if the minimum is attained.

Proposition 3. Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be such that $-\infty < f^* := \inf_x f(x)$, and consider the two optimization problems $\mathbb{P}$ and $\mathcal{P}_M$ defined in (7) and (8) respectively. Then, $\inf \mathcal{P}_M \downarrow f^*$ as $M \to \infty$. If f has a global minimizer $x^* \in \mathbb{R}^n$, then $\min \mathcal{P}_M = f^*$ whenever $M \geq \|x^*\|_\infty$.

PROOF. Let $\mu \in \mathcal{P}(\mathbb{R}^n)$ be admissible for $\mathcal{P}_M$. As $f \geq f^*$ on $\mathbb{R}^n$ then it follows immediately that $\int f d\mu \geq f^*$, and so, $\inf \mathcal{P}_M \geq f^*$ for all M.

As B_M is closed and bounded, it is compact and so, with f^*_M as in Proposition 2, there is some $\hat{x} \in B_M$ such that $f(\hat{x}) = f^*_M$. In addition let $\mu \in \mathcal{P}(\mathbb{R}^n)$ be the Dirac probability measure at the point $\hat{x}$. As $\|\hat{x}\|_\infty \leq M$,

$$\int \sum_{i=1}^n \mathrm{e}^{\mathrm{x_i^2}}\, \mathrm{d}\mu \,=\, \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{e}^{(\hat{\mathrm{x}}_\mathrm{i})^2} \,\leq\, \mathrm{n}\mathrm{e}^{\mathrm{M}^2},$$

so that μ is an admissible solution of $\mathcal{P}_M$ with value $\int f\, d\mu = f(\hat{x}) = f^*_M$, which proves that $\inf \mathcal{P}_M \leq f^*_M$. This latter fact, combined with Proposition 2 and with $f^* \leq \inf \mathcal{P}_M$, implies $\inf \mathcal{P}_M \downarrow f^*$ as $M \to \infty$, the desired result. The final statement is immediate by taking as feasible solution for $\mathcal{P}_M$, the Dirac probability measure at the point $x^* \in B_M$ (with $M \geq \|x^*\|_\infty$). As its value is now f^*, it is also optimal, and so, $\mathcal{P}_M$ is solvable with optimal value $\min \mathcal{P}_M = f^*$.

Proposition 3 provides a rationale for introducing the following Semidefinite Programming (SDP) problems. Let $2r_f$ be the degree of f and for every $r_f \leq r \in \mathbb{N}$, consider the SDP problem

$$\mathbb{Q}_r \begin{cases} \quad \min_{\mathbf{y}} \; L_{\mathbf{y}}(f) \, (= \sum_\alpha f_\alpha y_\alpha) \\ \text{s.t.} \; M_r(\mathbf{y}) \,\succeq\, 0 \\ \quad \displaystyle\sum_{k=0}^r \sum_{i=1}^n \frac{y_{2k}^{(i)}}{k!} \,\leq\, n\mathrm{e}^{\mathrm{M}^2}, \\ \quad y_0 \,=\, 1, \end{cases} \tag{9}$$

and its associated *dual* SDP problem

$$\mathbb{Q}^*_r \begin{cases} \quad \max_{\lambda \geq 0, \gamma, q} \; \gamma - n\mathrm{e}^{\mathrm{M}^2} \lambda \\ \text{s.t.} \; f - \gamma \,=\, q - \lambda \displaystyle\sum_{k=0}^r \sum_{j=1}^n \frac{x_j^{2k}}{k!} \\ \quad q \quad \text{s.o.s. of degree } \,\leq 2r, \end{cases} \tag{10}$$

with respective optimal values $\inf \mathbb{Q}_r$ and $\sup \mathbb{Q}^*_r$ (or $\min \mathbb{Q}_r$ and $\max \mathbb{Q}^*_r$ if the optimum is attained, in which case the problems are said to be solvable). For more details on SDP theory, the interested reader is referred to the survey paper [VandenBerghe and Boyd, 1996].

The SDP problem $\mathbb{Q}_r$ is a relaxation of $\mathcal{P}_M$, and we next show that in fact

- $\mathbb{Q}_r$ is solvable for all $r \geq r_0$,

- its optimal value $\min \mathbb{Q}_r \to \inf \mathcal{P}_M$ as $r \to \infty$, and

- $\mathbb{Q}^*_r$ is also solvable with same optimal value as $\mathbb{Q}_r$, for every $r \geq r_f$.

This latter fact will be crucial to prove our main result in the next section. Let l_∞ (resp. l_1) be the Banach space of bounded (resp. summable) infinite sequences with the sup-norm (resp. the l_1-norm).

Theorem 4. Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be of degree $2r_f$, with global minimum $f^* > -\infty$, and let $M > 0$ be fixed. Then :

(i) For every $r \geq r_f$, $\mathbb{Q}_r$ is solvable, and $\min \mathbb{Q}_r \uparrow \inf \mathcal{P}_M$ as $r \to \infty$.

(ii) Let $\mathbf{y}^{(r)} = \{y_\alpha^{(r)}\}$ be an optimal solution of $\mathbb{Q}_r$ and complete $\mathbf{y}^{(r)}$ with zeros to make it an element of l_∞. Every (pointwise) accumulation point $\mathbf{y}^*$ of the sequence $\{\mathbf{y}^{(r)}\}_{r \in \mathbb{N}}$ is a determinate moment sequence, that is,

$$y^*_\alpha \,=\, \int_{\mathbb{R}^n} x^\alpha \, d\mu^*, \qquad \alpha \in \mathbb{N}^n, \tag{11}$$

for a unique probability measure μ^*, and μ^* is an optimal solution of $\mathcal{P}_M$.

(iii) For every $r \geq r_f$, $\max \mathbb{Q}^*_r = \min \mathbb{Q}_r$.

For a proof see [Lasserre, 2004].

So, one can approximate the optimal value f^* of $\mathbb{P}$ as closely as desired, by solving SDP-relaxations $\{\mathbb{Q}_r\}$ for sufficiently large values of r and M. Indeed, $f^* \leq \inf \mathcal{P}_M \leq f^*_M$, with f^*_M as in Proposition 2. Therefore, let $\epsilon > 0$ be fixed, arbitrary. By Proposition 3, we have $f^* \leq \inf \mathcal{P}_M \leq f^* + \epsilon$ provided that M is sufficiently large. Next, by Theorem 4(i), one has $\inf \mathbb{Q}_r \geq \inf \mathcal{P}_M - \epsilon$ provided that r is sufficiently large, in which case, we finally have $f^* - \epsilon \leq \inf \mathbb{Q}_r \leq f^* + \epsilon$.

Notice that the SDP-relaxation $\mathbb{Q}_r$ in (9) is simpler than the one defined in [Lasserre, 2001]. Both have the same variables $\mathbf{y} \in \mathbb{R}^{s(r)}$, but the former has *one* SDP constraint $M_r(\mathbf{y}) \succeq 0$ and one scalar inequality (as one substitutes y_0 with 1) whereas the latter has the same SDP constraint $M_r(\mathbf{y}) \succeq 0$ and one additional SDP constraint $M_{r-1}(\theta\mathbf{y}) \succeq 0$ for the localizing matrix associated with the polynomial $x \mapsto \theta(x) = M^2 - \|x\|^2$. This results in a significant simplification.

4. SUM OF SQUARES APPROXIMATION

Let $\mathcal{A}$ be equipped with the norm

$$f \;\mapsto\; \|f\|_1 \;:=\; \sum_{\alpha\in\mathbb{N}^n} |f_\alpha|, \qquad f \in \mathcal{A}.$$

Theorem 5. Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be nonnegative with global minimum f^*, that is,

$$0 \;\leq\; f^* \;\leq\; f(x), \qquad x \in \mathbb{R}^n.$$

(i) There is some $r_0 \in \mathbb{N}, \lambda_0 \geq 0$ such that, for all $r \geq r_0$ and $\lambda \geq \lambda_0$,

$$f + \lambda \sum_{k=0}^{r} \sum_{j=1}^{n} \frac{x_j^{2k}}{k!} \qquad \text{is a sum of squares.} \quad (12)$$

(ii) For every $\epsilon > 0$, there is $r_\epsilon \in \mathbb{N}$ such that,

$$f_\epsilon := f + \epsilon \sum_{k=0}^{r_\epsilon} \sum_{j=1}^{n} \frac{x_j^{2k}}{k!} \qquad \text{is a sum of squares.} (13)$$

Hence, $\|f - f_\epsilon\|_1 \to 0$ as $\epsilon \downarrow 0$.

For a proof see [Lasserre, 2004].

Remark 6. Theorem 5(ii) is a *denseness* result in the spirit of Theorem 5, p. 122 in [Berg, 1980] which states that the cone of s.o.s. polynomials is dense (also for the norm $\|f\|_1$) in the cone of polynomials that are nonnegative on $[-1, 1]^n$. However, notice that Theorem 5(ii) provides an *explicit* converging sequence $\{f_\epsilon\}$ with a simple and very specific form.

Ex: Let $f \in \mathbb{R}[x, y]$ be the Motzkin polynomial

$$(x, y) \mapsto f(x, y) := 1 + x^4y^2 + x^2y^4 - 3x^2y^2.$$

Then, as proved by B. Reznick,

$$f + (n-1)^{n-1}(xy)^{2n}/n^n, \quad \text{is s.o.s. } \forall n \geq 3.$$

REFERENCES

Berg C. (1980). The multidimensional moment problem and semi-groups, *Proc. Symp. Appl. Math.*, 37:110–124.

Blekherman G. (2004). There are significantly more nonnegative polynomials than sums of squares, Department of Mathematics, University of Michigan, Ann Arbor, USA.

Henrion D. and Lasserre J. B. (2004). Solving nonconvex optimization problems, *IEEE Contr. Syst. Mag.*, 24:72–83.

Jacobi T. and Prestel A. (2001). Distinguished representations of strictly positive polynomials, *J. Reine. Angew. Math.*, 532:223–235.

Lasserre J. B. (2001). Global optimization with polynomials and the problem of moments, *SIAM J. Optim.*, 11:796–817.

Lasserre J. B. (2002). Semidefinite programming vs. LP relaxations for polynomial programming, *Math. Oper. Res.*, 27:347–360.

Lasserre J. B. (2004). A sum of squares approximation of nonnegative polynomials, *Report # 04428*, July, LAAS-CNRS, Toulouse, France.

Nussbaum A. E. (1966). Quasi-analytic vectors, *Ark. Mat.*, 6:179–191.

Parrilo P. A. (2000). Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization, PhD thesis, Caltech, Pasadena, CA, USA.

Parrilo P. A. (2003). Semidefinite programming relaxations for semialgebraic problems, *Math. Prog. Ser. B*, 96:293–320.

Prestel A. and Delzell C. N. (2001). *Positive Polynomials*, Springer, Berlin.

Putinar M. (1993). Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.*, 42:969–984.

Scheiderer C. (2003). Positivity and sums of squares: A guide to some recent results, Department of Mathematics, University of Duisburg, Germany.

Schmüdgen K. (1991). The *K*-moment problem for compact semi-algebraic sets, *Math. Ann.*, 289:203–206.

Schweighofer M. (2004). Optimization of polynomials on compact semialgebraic sets, *SIAM J. Optim*, to appear.

Vandenberghe L. and Boyd S. (1996). Semidefinite programming, *SIAM Review*, 38:49-95.