Some applications of the Łojasiewicz inequality in optimization

Krzysztof Kurdyka, Stanisław Spodzieja

Gdańsk-Kraków-ódź-Warszawa Workshop in Singularity Theory
- a special session dedicated to the memory of StanisĹ,aw Łojasiewicz
12.12.2022 - 16.12.2022, Warszawa

Abstract

We show that if a polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ is nonnegative on a closed basic semialgebraic set

$$X = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \dots, g_r(x) \geq 0\},$$

where $g_1, \dots, g_r \in \mathbb{R}[x_1, \dots, x_n]$, then f can be approximated uniformly on compact sets by polynomials of the form

$$\sigma_0 + \varphi(g_1)g_1 + \cdots + \varphi(g_r)g_r,$$

where $\sigma_0 \in \mathbb{R}[x_1, \dots, x_n]$ and $\varphi \in \mathbb{R}[t]$ are sums of squares of polynomials. In particular, if X is compact, and $h(x) := R^2 - |x|^2$ is positive on X, then

$$f = \sigma_0 + \sigma_1 h + \varphi(g_1)g_1 + \cdots + \varphi(g_r)g_r$$

for some sums of squares $\sigma_0, \sigma_1 \in \mathbb{R}[x_1, \dots, x_n]$ and $\varphi \in \mathbb{R}[t]$.
We give stronger versions of known approximation and representation theorems with sums of squares of polynomials. Then, using the Łojasiewicz inequality, we give quantitative versions of these results and explain some applications to semidefinite optimization methods.

The presented results can be found in::

Kurdyka, Krzysztof; Spodzieja, Stanisław, *Convexifying positive polynomials and sums of squares approximation*. SIAM J. Optim. 25 (2015), no. 4, 2512–2536.

Krzysztof Kurdyka, Stanisďż˝aw Spodzieja, Anna Szlachcińska, *Metric properties of semialgebraic mappings*. Discrete Comput. Geom. 55 (2016), no. 4, 786–800.

Abdulljabar Naji Abdullah, Klaudia Rosiak, Stanisław Spodzieja, *Convexifying of polynomials by convex factor*, to appear as a chapter in Analytic in Algebraic Geometry 4.

Introduction

We denote by $\mathbb{R}[x]$ or $\mathbb{R}[x_1, \ldots, x_n]$ the ring of polynomials in $x = (x_1, \ldots, x_n)$ with coefficients in $\mathbb{R}$.
Important problems of real algebraic geometry are representations of nonnegative polynomials on closed semialgebraic sets.

Recall **Hilbert's 17th problem** (solved by E. Artin (1927)):
If $f \in \mathbb{R}[x]$ is nonnegative on $\mathbb{R}^n$, then

$$\text{(AH)} \qquad fh^2 = h_1^2 + \cdots + h_m^2 \quad \text{for some } h, h_1, \ldots, h_m \in \mathbb{R}[x],\ h \neq 0,$$

that is, f is a sum of squares of rational functions.

D. Hilbert (1888) proved that for $n \geq 2$ there are nonnegative polynomials on $\mathbb{R}^n$ which are not sums of squares of polynomials.
T. S. Motzkin (1967) gave an explicit example of such a polynomial, $f(x_1, x_2) = 1 + x_1^2 x_2^2 (x_1^2 + x_2^2 - 3)$, i.e., in the representation (AH) of f the degree of h must be positive.

Let $X \subset \mathbb{R}^n$ be a *closed basic semialgebraic set* defined by $g_1, \ldots, g_r \in \mathbb{R}[x]$, i.e.,

$$(0.1) \qquad X = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_r(x) \geq 0\}.$$

The *preordering* generated by $g_1, \ldots, g_r$ is defined to be

$$T(g_1, \ldots, g_r) = \Big\{ \sum_{e=(e_1,\ldots,e_r)\in\{0,1\}^r} \sigma_e g_1^{e_1} \cdots g_r^{e_r} : \sigma_e \in \sum \mathbb{R}[x]^2 \text{ for } e \in \{0,1\}^r \Big\},$$

where $\sum \mathbb{R}[x]^2$ denotes the set of sums of squares of polynomials.

When **the set X is compact**, a very important result was obtained by K. Schmüdgen (1991):

(Schmüdgen Positivstellensatz) Every strictly positive polynomial f on X belongs to the preordering $T(g_1, \ldots, g_r)$.

In the case of nonnegative polynomials C. Berg, J.P.R. Christensen and P. Ressel (1976) and J.B.Lasserre and T. Netzer (2007) proved that:

Aany polynomial f which is nonnegative on $[-1,1]^n$ can be approximated in the l_1-norm by sums of squares of polynomials.

The *l_1-norm of a polynomial* is defined to be the sum of the absolute values of its coefficients (in the usual monomial basis).

In general the Schmüdgen Positivstellensatz does not hold on noncompact sets. For a polynomial f positive on a noncompact set X the problem arises of approximation of f by elements of the preordering $T(g_1, \ldots, g_r)$ or of the *quadratic module*

$$P(g_1, \ldots, g_r) := \left\{\sigma_0 + \sigma_1 g_1 + \cdots + \sigma_r g_r : \sigma_i \in \sum \mathbb{R}[x]^2,\ i = 0, \ldots, r\right\}.$$

In this connection J. B. Lasserre (2008) proved that:

If $g_1, \ldots, g_r$ are concave polynomials such that $\operatorname{Int} X \neq \emptyset$, then any convex polynomial nonnegative on X can be approximated in the l_1-norm by polynomials from the set

$$L_c(g_1, \ldots, g_r) := \Big\{\sigma_0 + \lambda_1^2 g_1 + \cdots + \lambda_r^2 g_r : \sigma_0 \in \sum \mathbb{R}[x]^2 \textbf{ convex}, \\ \lambda_1, \ldots, \lambda_r \in \mathbb{R}\Big\}.$$

For $X = \mathbb{R}^n$ the approximation is uniform on compact sets.

Our contributions

We prove an analogue of the Schmüdgen and Putinar theorems for a smaller cone. Namely for $g \in \mathbb{R}[x]$ we put

$$\mathcal{K}(g, g_1, \ldots, g_r) := \Big\{\sigma_0 + \sigma_1 g + \sum_{i=1}^{r} \varphi(g_i) g_i : \sigma_0, \sigma_1 \in \sum \mathbb{R}[x]^2, \\ \varphi \in \sum \mathbb{R}[t]^2 \Big\},$$

where t is a single variable. Note that if we set

$$\Phi(g_1, \ldots, g_r) := \Big\{\varphi(g_1) g_1 + \cdots + \varphi(g_r) g_r : \varphi \in \sum \mathbb{R}[t]^2 \Big\},$$

then

$$\mathcal{K}(g, g_1, \ldots, g_r) = T(g) + \Phi(g_1, \ldots, g_r),$$

where $\mathcal{A} + \mathcal{B} = \{a + b : a \in \mathcal{A},\ b \in \mathcal{B}\}$.

We have

Theorem ((0.1) K. Kurdyka, S. Spodzieja (2015))

For a closed basic semialgebraic set X defined by $g_1, \dots, g_r \in \mathbb{R}[x]$ and a polynomial $f \in \mathbb{R}[x]$ the following conditions are equivalent:

(i) *f is nonnegative on X,*

(ii) *f can be uniformly approximated on compact sets by polynomials from the cone*

$$\mathcal{S}(g_1, \dots, g_r) := \sum \mathbb{R}[x]^2 + \Phi(g_1, \dots, g_r).$$

Moreover, f can be approximated by polynomials from $\mathcal{S}(g_1, \dots, g_r)$ in the l_1-norm.

From the above Theorem and the Schmüdgen Positivstellensatz we obtain

Corollary (0.2)

if X is a compact set and $g(x) := R^2 - |x|^2 \geq 0$ for $x \in X$, then

$$f \text{ is strictly positive on } X \quad \Longrightarrow \quad f \in \mathcal{K}(g, g_1, \ldots, g_r). \tag{0.2}$$

Lasserre relaxation method (2001)

Lasserre (2001) gave a method of minimizing a polynomial f on a compact basic semialgebraic set X of the form (0.1). More precisely, let

$$f^* := \inf\{f(x) : x \in X\}.$$

Then

$$f^* = \sup\{a \in \mathbb{R} : f(x) - a > 0 \text{ for } x \in X\},$$

and by Putinar's result (1993),

$$f^* = \sup\{a \in \mathbb{R} : f - a \in P(g_1, \dots, g_r)\},$$

or equivalently

$$f^* = \inf\{L(f) : L : \mathbb{R}[x] \to \mathbb{R} \text{ is linear, } L(1) = 1,\ L(P(g_1, \dots, g_r)) \subset [0, \infty)\}.$$

$$\text{Define} \qquad P_k(g_1,\dots,g_r) := \Big\{\sigma_0 g_0 + \cdots + \sigma_r g_r \in P(g_1,\dots,g_r) : \deg \sigma_i g_i \le k,\ i = 0,\dots,r\Big\},$$

where $g_0 = 1$. Lasserre considered the following optimization problems:

$$\text{maximize } a \in \mathbb{R} : f - a \in P_k(g_1,\dots,g_r),$$

$$\text{minimize } L(f) \text{ for } L : \mathbb{R}[x]_k \to \mathbb{R},\ \text{linear},\ L(1) = 1,\quad L(P_k(g_1,\dots,g_r)) \subset [0,\infty),$$

where $\mathbb{R}[x]_k := \{h \in \mathbb{R}[x] : \deg h \le k\}$. Set

$$a_k^* := \sup\{a \in \mathbb{R} : f - a \in P_k(g_1,\dots,g_r)\},$$

$$l_k^* := \inf\{L(f) : L : \mathbb{R}[x]_k \to \mathbb{R} \text{ linear},\ L(1) = 1,\ L(P_k(g_1,\dots,g_r)) \subset [0,\infty)\},$$

for sufficiently large $k \in \mathbb{N}$.

Lasserre proved that (a_k^*), (l_k^*) **are increasing sequences that converge to** f^* **and** $a_k^* \le l_k^* \le f^*$ **for** $k \in \mathbb{N}$.

Quantitative aspects of Theorem 0.1

In order to estimate the rate of convergence in Lasserre's relaxation method (2001) we show how to bound the degree of the polynomial φ in Theorem 0.1.

Assume now that X is a compact set of the form

$$X = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_r(x) \geq 0\}, \tag{0.3}$$

where $g_1, \ldots, g_r \in \mathbb{R}[x]$.

Choose $R > 0$ large enough so that $g_0(x) = R^2 - |x|^2$ is nonnegative polynomial on X.

We now take the cone $\mathcal{K}(g_0, \ldots, g_r)$.

By Theorem 0.1 we obtain

Remark

We may use the Lasserre algorithm for minimization of a polynomial f on a compact basic semialgebraic set X by using $\mathcal{K}(g, g_1, \ldots, g_r)$ instead of $P(g, g_1, \ldots, g_r)$. In fact, we can use the set $\mathcal{K}_k(g, g_1, \ldots, g_r)$ consisting of all $\sigma_0 + \sigma_1 g + \sum_{i=1}^r \varphi(g_i)g_i \in \mathcal{K}(g, g_1, \ldots, g_r)$ such that $\deg \sigma_0 \le k$, $\deg \sigma_1 g \le k$ and $\deg \varphi(g_i)g_i \le k$ for $i = 1, \ldots, r$. Consider the following optimization problems:

- maximize $a \in \mathbb{R}$ such that $f - a \in \mathcal{K}_k(g, g_1, \ldots, g_r)$,
- minimize $L(f)$ for $L : \mathbb{R}[x]_k \to \mathbb{R}$, linear, $L(1) = 1$, $L(\mathcal{K}_k(g, g_1, \ldots, g_r)) \subset [0, \infty)$.

Denote

$$u_k^* := \sup\{a \in \mathbb{R} : f - a \in \mathcal{K}_k(g, g_1, \ldots, g_r)\},$$
$$v_k^* := \inf\{L(f) : L : \mathbb{R}[x]_k \to \mathbb{R} \text{ linear}, L(1) = 1, L(\mathcal{K}_k(g, g_1, \ldots, g_r)) \subset [0, \infty$$

for sufficiently large $k \in \mathbb{N}$. **Then (u_k^*), (v_k^*) are increasing sequences that converge to f^* (by Corollary 0.2) and $u_k^* \le v_k^* \le f^*$ for $k \in \mathbb{N}$.** □

Quantitative Łojasiewicz inequality

Let $g_1, \ldots, g_r \in \mathbb{R}[x]$, and let $G : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$(0.4) \qquad G(x) = \max\{0, -g_1(x), \ldots, -g_r(x)\}, \quad x \in \mathbb{R}^n.$$

Then $X = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_r(x) \geq 0\} = G^{-1}(0)$. Moreover,

$$\operatorname{graph} G = Y_0 \cup Y_1 \cup \cdots \cup Y_r,$$

where $Y_0 = X \times \{0\}$,

$$Y_i = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y = -g_i(x),\ g_i(x) \leq 0,\ g_i(x) \leq g_j(x) \text{ for } j \neq i\},$$

for $i = 1, \ldots, r$. Note that each set Y_i, $i = 0, \ldots, r$, is defined by r inequalities and one equation. Let $d = \max\{\deg g_1, \ldots, \deg g_r\}$.

We now state the well-known *Łojasiewicz inequality* in a quantitative version proved by K. Kurdyka, S, Spodzieja and A. Szlachcińska (2014): there exist $C, \mathcal{L} > 0$ such that

$$G(x) \geq C \left(\frac{\operatorname{dist}(x, X)}{1 + |x|^d} \right)^{\mathcal{L}}, \quad x \in \mathbb{R}^n, \tag{0.5}$$

with

$$\mathcal{L} \leq d(6d - 3)^{n+r-1}. \tag{0.6}$$

It follows from (0.5) that for every $\rho > 0$ there exists $C_\rho > 0$ such that

$$G(x) \geq C_\rho \operatorname{dist}(x, X)^{\mathcal{L}} \quad \text{for any } x \in B(\rho), \tag{0.7}$$

where $B(\rho) = \{x \in \mathbb{R}^n : |x| \leq \rho\}$. Fix $R > 0$ such that $X \subset B(R)$. Assume that (0.7) holds with fixed $C' = C_R$ and $\mathcal{L}$.

Fact (1)

Let $\eta > 0$. *Set* $\delta_0 = C'\eta^{\mathcal{L}}$. *Then for any* $0 < \delta \leq \delta_0$,

$$\{x \in B(R) : g_i(x) \geq -\delta \text{ for } i = 1, \dots, r\} \subset \{x \in B(R) : \operatorname{dist}(x, X) \leq \eta\}.$$

Indeed, take $x \in B(R) \setminus X$ such that $g_i(x) \geq -\delta$ for $i = 0, \dots, r$. Let G be the function defined by (0.4). Hence by (0.7),

$$\delta \geq \max\{-g_1(x), \dots, -g_r(x)\} = G(x) \geq C' \operatorname{dist}(x, X)^{\mathcal{L}}.$$

Thus for $0 < \delta \leq \delta_0$ we deduce the assertion of Fact 1. □

For $\nu = (\nu_1, \ldots, \nu_n) \in \mathbb{N}^n$ we set $|\nu| = \nu_1 + \cdots + \nu_n$ and $a^\nu = a_1^{\nu_1} \cdots a_n^{\nu_n}$, where $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. For $h \in \mathbb{R}[x]$ of the form

$$h(x) = \sum_{j=0}^{d} \sum_{|\nu|=j} a_\nu x^\nu,$$

we define

$$\mathbb{A}(h, R) = \sum_{j=0}^{d} \sum_{|\nu|=j} |a_\nu| R^j, \qquad \mathbb{B}(h, R) = \sum_{j=1}^{d} \sum_{|\nu|=j} j |a_\nu| R^{j-1} \quad \text{for } R > 0.$$

Then for $x \in B(R)$ we have $|h(x)| \leq \mathbb{A}(h, R)$ and by the Euler formula for homogeneous functions, $|\nabla h(x)| \leq \mathbb{B}(h, R)$.

Proposition (04)

Let $f \in \mathbb{R}[x]$, *let* X *be a semialgebraic set of the form* (0.1) *such that* $X \subset B(R)$, $R > 0$, *and let* $g_1, \ldots, g_r \in \mathbb{R}[x]$ *be polynomials satisfying* (0.7) *with fixed* $C, \mathcal{L} > 0$. *Take* $M, A \in \mathbb{R}$ *such that*

$$M \geq \max\{1, \mathbb{A}(f, R), \mathbb{B}(f, R)\}, \quad A \geq \max\{1, \mathbb{A}(g_i, R)\} \quad \textit{for } i = 1, \ldots, r.$$

Take $\epsilon > 0$, *and set*

$$\varphi(t) = \left(\frac{1}{A}t - 1 + \frac{\delta}{2A}\right)^{2N},$$

where

$$0 < \delta \leq \min\left\{A, C\left(\frac{\epsilon}{2M}\right)^{\mathcal{L}}\right\},$$
$$N \geq \max\left\{\frac{(r-1)A-1}{2}, \frac{A(2M+1-\delta)}{\delta^2}, \frac{2rA-\epsilon}{2\epsilon}\right\}.$$

Proposition (Continuation of the proposition)

Then the function

$$h(x) = \sum_{i=1}^{r} \varphi(g_i(x))g_i(x) \in \Phi(g_1, \dots, g_r)$$

satisfies the following conditions:

$$0 \leq h(x) < \epsilon \quad \text{for } x \in X, \tag{0.8}$$

$$\forall_{|y| \leq R}\ \exists_{x \in X}\ f(y) - h(y) \geq f(x) - h(x) - \epsilon. \tag{0.9}$$

Remark

If we assume that $g_1, \dots, g_r$ are μ-strongly concave polynomials, i.e.,

$$g_i(y) \le g_i(x) + \langle y - x, \nabla g_i(x) \rangle - \frac{\mu}{2}|y - x|^2 \quad \text{for } x, y \in \mathbb{R}^n,$$

where $\mu > 0$ and $\langle \cdot\,, \cdot \rangle$ is the standard scalar product, then the assertion of Fact 1 holds with $\delta_0 = \eta^2\mu/2$. Hence, Proposition 0.4 holds with $0 < \delta \le \min\left\{A, \frac{\epsilon^2\mu}{8M^2}\right\}$.

Remark

We can use Proposition 0.4 to minimize a polynomial f on a compact basic semialgebraic set X. Let $X \subset \{x \in \mathbb{R}^n : |x| \leq R\}$. Then for any $\epsilon > 0$, we can effectively compute a polynomial $h(x) = \sum_{i=i}^{r} \varphi(g_i(x))g_i(x)$, where $\varphi \in \sum \mathbb{R}[t]^2$, such that

$$f^* - 2\epsilon \leq \inf\{f(y) - h(y) : |y| \leq R\} \leq f^* + 2\epsilon.$$

To approximate f^*, we can minimize $f - h$ on $B(R)$. To this end we may compute

$$a_k^{**} := \sup\{a \in \mathbb{R} : f - h - a \in P_k(R^2 - |y|^2)\} \quad \text{for } k \in \mathbb{N}.$$

By the Putinar Theorem (or the Schmüdgen Theorem) we see that

$$a_k^{**} \to f^{**} \quad \text{as } k \to \infty,$$

where $f^{**} := \inf\{f(y) - h(y) : |y| \leq R\}$.

Remark (Continuation of the remark)

Minimization of $f - h$ on $B(R)$ is much simpler than minimizing f on X, because the set $B(R)$ is described by one inequality $R^2 - |x|^2 \geq 0$. In this case M. Schweighofer (2004) gave the rate of convergence of the sequence a_k^{**}:

$$f^{**} - a_k^{**} \leq \frac{c}{\sqrt[d]{k}}$$

for some constant $c \in \mathbb{N}$ depending on f and $R^2 - |y|^2$ and some constant $d \in \mathbb{N}$ depending on $R^2 - |y|^2$.

Some other application of Łojdiewicz's inequality

Lwt $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial, let $r \in \mathbb{R}$. Assume that

$$X_{f \leq r} := \{x \in \mathbb{R}^n : f(x) \leq r\}$$

is compact. Then there exists a continuous mapping $\kappa : X_{f \leq r} \to X_{f \leq r}$ such that

$$f(\kappa(\xi)) \leq f(\xi) \quad \text{for } \xi \in X_{f \leq r},$$

and the equality holds if and only if $\xi \in \Sigma_f \cap X_{f \leq r}$

$$X_{f \leq r} \ni \xi \mapsto \kappa^\nu(\xi) \to \kappa_*(\xi) \in \Sigma_f \cap X_{f \leq r} \qquad \text{as } \nu \to \infty,$$

where

$$\Sigma_f := \{x \in \mathbb{R}^n : \nabla f(x) = 0\}.$$

So, by the Dini criterion, the convergence is uniform.

Theorem

If the function f has only one singular value in $(-\infty, r]$, then the mapping $\kappa_* : X_{f \leq r} \ni \xi \mapsto \kappa_*(\xi) \in \Sigma_f$ is continuous.

Proof.

One can assume that 1 is the unique singular value of f in $(-\infty, r]$. Then

$$f(\kappa^{\nu}(\xi)) \to 1 = f(\kappa_*(\xi))$$

So, by Dini's criterion, the convercence is uniform.

Remark

By the Łojasiewicz gradient inequality

$$(0.10) \qquad |f(x) - f(\kappa_*(\xi))|^{\varrho} \le C|\nabla f(x)|$$

in a neighbourhood in $\mathbb{R}^n$ *of the set* $f^{-1}(f(\kappa_*(\xi)))$ *for some constants* $0 < \varrho < 1$ *and* $C > 0$*, we obtain the following Kurdyka-Łojasiewicz inequality*

$$(0.11) \quad \operatorname{dist}(\kappa_\nu(\xi), f^{-1}(f(\kappa_*(\xi)))) \le \frac{1}{C(1-\varrho)}(f(\kappa_\nu(\xi)) - f(\kappa_*(\xi)))^{1-\varrho}.$$

So, we have

$$|\kappa_\nu(\xi) - \kappa_*(\xi)| \le \sum_{k=\nu}^{\infty} |\kappa_{\nu+1}(\xi) - \kappa_\nu(\xi)|$$

$$\le \frac{1}{C(1-\varrho)} \sum_{k=\nu}^{\infty} \left[(f(\kappa_k(\xi)) - f(\kappa_*(\xi)))^{1-\varrho} - (f(\kappa_{k+1}(\xi)) - f(\kappa_*(\xi)))^{1-\varrho} \right]$$

$$= \frac{1}{C(1-\varrho)} (f(\kappa\nu(\xi)) - f(\kappa_*(\xi)))^{1-\varrho}.$$

So, the sequence κ_ν tends to κ_* uniformly on $X_{f \le r}$.

The end