Polynomial Optimization

Simone Naldi

Université de Limoges
Master ACSYON
2017-2018

These are the notes for the cours “Semialgebraic Optimization” given jointly by Simone Naldi (Université de Limoges) and Didier Henrion (CNRS LAAS, Toulouse and Czech University, Prague) for the M2 of the Master ACSYON, Université de Limoges. In case of errors or typos, please send feedback to Simone Naldi at

`simone.naldi@unilim.fr`

We closely follow the book [5]. Other good references are [1] (treating different topics in conic and semialgebraic optimization), the book [4] focusing on the generalized moment problem, the original papers by J-B. Lasserre [3] and the PhD thesis of P. Parrilo [6]. A more general introduction to convex algebraic geometry is [2].

Contents

Lecture I

1.1. Introduction

The general goal of the course is to address a special class of optimization problems, defined with polynomial data. For instance, the objective function or the feasible set are defined by multivariate real polynomials such as

$$f = 3x^3y^2 - 5z^2 + yz - \sqrt{2}.$$

We use the standard notation for sets of natural ($\mathbb{N}$), integer ($\mathbb{Z}$), rational ($\mathbb{Q}$), real ($\mathbb{R}$) and complex ($\mathbb{C}$) numbers.

Definition 1 (Optimization Problem)**.** *Let $f, g_1, \ldots, g_m \colon \mathbb{R}^n \to \mathbb{R}$ be functions. The general Optimization Problem (OP) associated to $f, g_1, \ldots, g_m$ is*

$$\begin{array}{rcll} f^* & := & \inf & f(x) \\ & & s.t. & x \in S \end{array} \tag{1}$$

where $S = \{x \in \mathbb{R}^n : g_i(x) \geq 0, \forall i = 1, \ldots, m\}$. The set S is called the feasible set *and f is called the* objective function *of problem* (1).

We first remark that similar definitions in the optimization literature, for the feasible set S, could involve different types of constraints, but that our description with inequalities of type $\geq$ is general enough. Indeed:

- $g(x) \leq 0$ if and only if $-g(x) \geq 0$
- $g(x) = 0$ if and only if $g(x) \geq 0$ *and* $g(x) \leq 0$.

In particular, we will not distinguish between equality and inequality constraints. We also remark that in its whole generality S can possibly be

- *non-linear, non-convex* with *nodal points* on the boundary
- *finite or infinite*
- *bounded, unbounded*

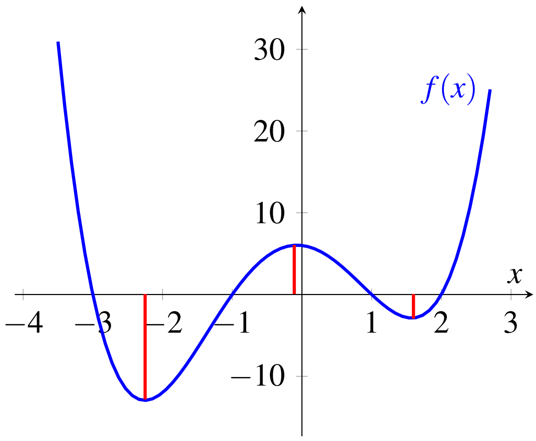
Example 2. Suppose that $g_1, \ldots, g_m$ are *concave functions*, that is for $i = 1, \ldots, m$

$$\forall x, y \in \mathbb{R}^n,\ \forall t \in [0,1] \Rightarrow g_i(tx + (1-t)y) \geq tg(x) + (1-t)g(y).$$

Then S is a convex set. By the way, this will *not* be the case in this course, generally speaking).

Typically, in optimization, we distinguish between *local* and *global* solutions.

Example 3. Let $f = (x+3)(x+1)(x-1)(x-2)$, $S = \mathbb{R}^n$. Then the goal is to minimise f over the whole domain $\mathbb{R}^n$. In this case f has one global minimizer between -3 and -1, a local maximizer between -1 and 1, and a local minimizer between 1 and 2. It has no global maximizer.



Local minima/maxima can be computed via *first-order conditions*, that is vanishing of the gradient vector:

$$\nabla f(x) = 0,$$

for functions $f \in C^1(S)$ (the space of continuously differentiable functions). By the way, one cannot distinguish between minima and maxima just looking at the first derivative. *Second-order conditions* on the Hessian of f

$$\mathrm{H}f(x) \succeq 0 \text{ or } \mathrm{H}f(x) \preceq 0,$$

(where $\succeq$, resp. $\preceq$, stands for positive semidefinite, resp. negative semidefinite, see Appendix) give local convexity/concavity for f; hence one could distinguish maxima and minima via second-order conditions (sufficient, not necessary).

In the special case of semialgebraic optimization, that is when data $f, g_1, \ldots, g_m$ are multivariate real polynomials (or possibly rational functions, but we will not treat this case in the course), we will see that one can approximate as close as possible (at fixed precision) the value f^*, and often compute it exactly. This course describes a method to address this special optimization problem.

1.2. Multivariate polynomials

Let $n \in \mathbb{N}$. In the whole notes, we use the x to denote the vector of n variables $x_1, \ldots, x_n$. For small n we will often use the letters $x, y, z, w \ldots$ to denote variables.

The *monomial* in x *with exponent* α is denoted by

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \text{ for } \alpha = (\alpha_1 \ldots \alpha_n) \in \mathbb{N}^n.$$

For instance, when $\alpha = (1,3,2) \in \mathbb{N}^3$ then one has $x^\alpha = x_1 x_2^3 x_3^2$. A *polynomial* is a finite linear combination of monomials, with coefficients in $\mathbb{R}$:

$$f = \sum_{\alpha \in I} f_\alpha x^\alpha, f_\alpha \in \mathbb{R}$$

where I is a finite subset of $\mathbb{N}^n$. The set of all real polynomials in x is denoted by $\mathbb{R}[x]$. The *degree* of x^α is $|\alpha| = \alpha_1 + \cdots + \alpha_n$. The *degree* of a polynomial $f = \sum_{\alpha \in I} f_\alpha x^\alpha$ is $\deg(f) = \max_{\alpha \in I} |\alpha|$. In particular, $\deg(x^\alpha) = |\alpha|$.

We denote by $\mathbb{R}[x]_d = \{f \in \mathbb{R}[x] : \deg(f) \leq d\}$, $d \in \mathbb{N}$. Both $\mathbb{R}[x]$ and $\mathbb{R}[x]_d$ are real vector spaces. The dimension of $\mathbb{R}[x]$ is infinite, while the dimension of $\mathbb{R}[x]_d$ equals the number of monomials x^α of degree at most d, that is the binomial coefficient

$$\dim \mathbb{R}[x]_d = \binom{n+d}{d}.$$

1.3. Semialgebraic optimization

We define the semialgebraic optimization problem. We start from the feasible set.

Definition 4 (Semialgebraic set)**.** *Let* $g_1, \ldots, g_m \in \mathbb{R}[x]$*. The* semialgebraic set *associated to* $g = (g_1, \ldots, g_m)$ *is the set* $S(g) = \{x \in \mathbb{R}^n : g_i \geq 0, \forall i = 1, \ldots, m\}$.

Examples of semialgebraic[1] sets are hyperplanes, polyhedra, circles, ellipsoids, spheres (*cf.* Figure 1.1)

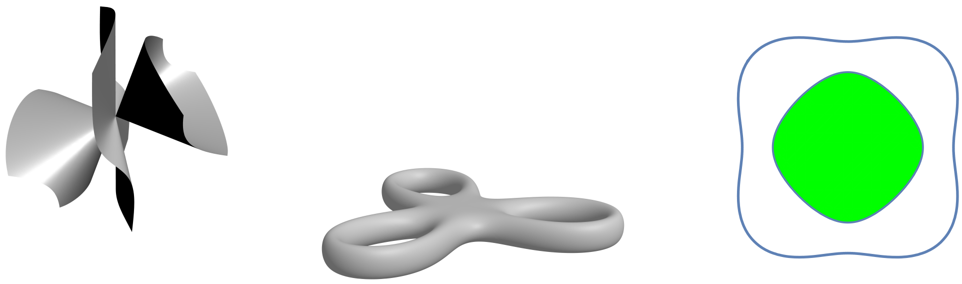


Figure 1.1. Some semialgebraic sets

As remarked above for general functions, equality constraints are allowed, and in this case the feasible set is called a *real algebraic set*: $\{x \in \mathbb{R}^n : g_i = 0, \forall i = 1, \ldots, m\}$.

Semialgebraic sets are the feasibles sets of the semialgebraic optimization problem, that we are going to define.

Definition 5 (Semialgebraic Optimization)**.** *Let* $f, g_1, \ldots, g_m \in \mathbb{R}[x]$*. The Semialgebraic Optimization Problem (SAO) associated to* $f, g_1, \ldots, g_m$ *is*

$$\begin{array}{rll} f^* := & \inf & f(x) \\ & s.t. & x \in S(g). \end{array} \tag{2}$$

1.4. Standard examples

Problem (2) models many important classes of optimization problems:

1. *Linear Programming (LP)* is the special case of Problem (2), when all polynomials are of degree 1:

$$\begin{array}{rll} f^* := & \inf & c^T x \\ & \text{s.t.} & Ax = b, x \geq 0. \end{array} \tag{3}$$

   Above, $x \geq 0$ means entry-wise positivity ($x_i \geq 0$ for all i). The feasible set is a polyhedron (a special semialgebraic set).

[1] In real algebraic geometry, the sets defined in Definition 4 are called *basic* semialgebraic sets, and *semialgebraic sets* are union of basic semialgebraic sets. In this course, we consider only basic semialgebraic sets, and we avoid the term basic.

2. *Quadratic Programming (QP)*, involving polynomials of degree 2:

$$\begin{array}{rlll} f^* & := & \inf & x^T C x + c^T x \\ & & \text{s.t.} & x^T P_j x + q_j^T x + r_j \geq 0, j = 1, \ldots, m \end{array} \tag{4}$$

   where C, P_j are real symmetric matrices and q_j are real vectors and $r_j \in \mathbb{R}$.

3. *Boolean Programming (BP)* where $S(g) = \{0,1\}^n$. $S(g)$ is semialgebraic since it is defined by the polynomial equalities $x_i(x_i - 1) = 0$.

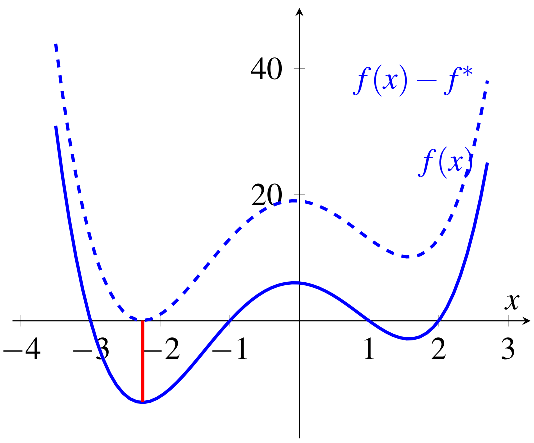
4. *Mixed-Integer Programming (MIP)*, that is LP with additional constraints of type $S(g) = [-N, N] \cap \mathbb{Z}^n$, solution to the polynomial equalities

$$(x_i + N)(x_i + N - 1) \cdots (x_i - N) = 0.$$

Hence (SAO) can be non-linear, non-convex; the feasible set can be unbounded. It is in general a hard problem, more precisely NP-hard, which means that there does not exist a polynomial-time algorithm solving all (SAO) problems. **1h30**

1.5. Equivalence with positivity

Suppose that the polynomial f in (2) is bounded from below on $S(g)$ ($f^* > -\infty$) and that $S(g)$ is non-empty ($f^* < \infty$). This is equivalent to the fact that $f^* \in \mathbb{R}$. Remark that in this case the polynomial $g(x) = f(x) - f^*$ takes only nonnegative values on $S(g)$. The next Figure shows the polynomial f in Example 3 and $f - f^*$.



Definition 6 (Positivity)**.** *Let $f, g_1, \ldots, g_m \in \mathbb{R}[x]$. We say that f is nonnegative on $S := S(g)$ if $f(x) \geq 0$ for all $x \in S$. We say that f is positive (or strictly nonnegative) if $f(x) > 0$ for all $x \in S$. The set of nonnegative polynomials on S is denoted by $\mathscr{P}(S)$, and for $d \in \mathbb{N}$, $\mathscr{P}_d(S) = \mathscr{P}(S) \cap \mathbb{R}[x]_d$.*

For instance, let $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ be the circle of radius 1 around the origin. Then the polynomial $f = x^2(1 - x^2 - y^2)$ is nonnegative on S. Indeed, it is the product of the "clearly globally nonnegative" polynomial x^2, with a polynomial which is also "clearly nonnegative over S", by definition of S (that is $1 - x^2 - y^2$).

Proposition 7 (Equivalence)**.** *Let $f, g_1, \ldots, g_m \in \mathbb{R}[x]_{\leq d}$. Then*

$$\begin{array}{ll} \inf & f(x) \\ s.t. & x \in S(g) \end{array} = \begin{array}{ll} \sup & \lambda \\ s.t. & f - \lambda \in \mathscr{P}_d(S(g)). \end{array} \tag{5}$$

Proof. Let f^* be the left term of the equality, and s^* the right one. If $f^* = -\infty$, then f is unbounded from below on S, hence $\{\lambda : f - \lambda \in \mathscr{P}_d(S)\} = \emptyset$, which implies that $s^* = -\infty = f^*$. Analogously if $f^* = +\infty$, then $S = \emptyset$ and every polynomial is nonnegative on S, hence $s^* = +\infty$.

Suppose now that $f^* \in \mathbb{R}$. Then $f - f^*$ is nonnegative on S, that is $f^* \leq s^*$. Now let λ be such that $f - \lambda$ is nonnegative on S. Hence $f^* \geq \lambda$, and passing to the supremum one gets $f^* = s^*$. ■

Hence we have the equivalence between polynomial positivity on S and polynomial optimization. We will typically distinguish two cases:

- $S(g) = \mathbb{R}^n$ (global case of SAO)
- $S(g) \subsetneq \mathbb{R}^n$ (local case of SAO)

In the local case, compactness or almost equivalent algebraic assumptions can be made in order to obtain characterizations of positivity (classically known as Positivstellensätze).

Proposition 8. *If $f \in \mathscr{P}(\mathbb{R}^n)$, then its degree is even.*

Proof. If f has odd degree, then there is an index $j \in \{1, \ldots, n\}$ such that

$$f = c_j x_j^{2k+1} \prod_{i \neq j} x_i^{k_i},$$

that is one variable must appear with an odd power. Then evaluating all other variables to 1, one gets that the resulting polynomial is non-zero and of odd degree. Hence it must change sign, and the same for f. ■

Definition 9. *A polynomial f is a sum of squares (SOS for short) if $f = f_1^2 + \cdots + f_m^2$ for some polynomials f_i. The set of sum-of-squares polynomials with n unknowns is denoted by Σ_n and its degree$-d$ part with $\Sigma_{n,d} = \Sigma_n \cap \mathbb{R}[x]_d$.*

Two of the exercises ask to prove that the cone of nonnegative polynomials, the cone of sums of squares and their degree$-d$ parts are closed convex cones with non-empty interiors.

1.6. Nonnegative polynomials and sums of squares

Every sum-of-squares polynomial is nonnegative by construction. What about the converse? Given a nonnegative polynomial, is it a sum of squares of polynomials? If yes, how such a decomposition can be computed?

Hilbert solved the first question in a celebrated result of 1888, inspired by a question posed by Minkowski in his doctoral dissertation. In this section we will partially prove Hilbert's theorem. We start with the easiest example of univariate polynomials.

Proposition 10. *A univariate polynomial is globally nonnegative if and only if it is a sum of squares. That is $\mathscr{P}(\mathbb{R}) = \Sigma_1$.*

Proof. Since f has real coefficients, then its roots are conjugate pairs of complex numbers (exercise). Moreover, by positivity the real roots have even multiplicity $(2,4,6\ldots)$. If R (resp. C) is the set of real (resp. complex non-real) roots of f (for R counted multiplicity), then

$$f = c \cdot \prod_{\alpha \in C} (x-\alpha)(x-\overline{\alpha}) \prod_{\beta \in R} (x-\beta)^2$$

where $c \geq 0$. Then $f = h\overline{h}g^2$ by properly defining polynomials $h \in \mathbb{C}[x]$ and $g \in \mathbb{R}[x]$. Hence we deduce $f = (\Re(h)^2 + \Im(h)^2)g^2 = (g\Re(h))^2 + (g\Im(h))^2$, where $\Re$ and $\Im$ return the real and imaginary part. ∎

Remark that the proof of Proposition 10 implies that two squares are always sufficient to describe all nonnegative polynomials.

Example 11. The polynomial $f = (x^2+1)(x-2)^2$ is nonnegative, and indeed its unique real root has multiplicity 2. Factorizing $f = (x-i)(x+i)(x-2)^2$ yields the polynomials h and g in the proof of Proposition 10, for instance $h = x+i$ and $g = x-2$, and the corresponding sum of squares decomposition:

$$f = (x(x-2))^2 + (x-2)^2.$$

We consider a second class of multivariate polynomials, those of degree 2. A quadratic form $f \in \mathbb{R}[x]_2$ is identified by its defining symmetric matrix M, for which

$$f = x^T M x, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Proposition 12. *A quadratic form $f \in \mathbb{R}[x]_2$ is nonnegative if and only if it is a sum of squares (of linear forms). That is, $\mathscr{P}(\mathbb{R}^n)_2 = \Sigma_{n,2}$.*

Proof. Suppose that $f \in \mathscr{P}(\mathbb{R}^n)_2$. Then remark that $f = x^T M x$ where M is a positive semidefinite symmetric matrix. By the Spectral Theorem there exists an orthogonal matrix P and a diagonal matrix D (with nonnegative entries) such that $M = P^T D P$. Hence one gets

$$f = x^T M x = x^T P^T D(Px) = (Px)D(Px) = \sum_i d_i(\ell_i(x))^2 = \sum_i (\sqrt{d_i}\ell_i(x))^2$$

where $\ell_i(x)$ are the elements of the vector Px. ∎

We finally state (without complete proof) Hilbert's 1888 theorem.

Theorem 13 (Hilbert 1888)**.** *The equality $\mathscr{P}(\mathbb{R}^n)_{2d} = \Sigma_{n,2d}$ holds if and only if*

- $n = 1$ *(univariate polynomials)*
- $d = 1$ *(quadratic forms)*
- $n = 2$ *and* $d = 2$ *(plane quartics)*

The original proof by Hilbert was not totally constructive. By the way he succeded in proving that in all cases ($d \geq 3$ and $n \geq 2$, or $n \geq 3$) there exist nonnegative polynomials that are not sums of squares. The first explicity examples were given by Motzkin and Robinson in the fifties and the sixties:

Motzkin polynomial $$M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$$

Robinson polynomial $$R(x,y,z) = y^2(1-x^2)^2 + x^2(1-y^2)^2 + x^2y^2(x^2+y^2-2)^2$$

We often call a sum of squares decomposition of a nonnegative polynomial an *algebraic certificate of positivity* or simply, a *certificate*. We will also give certificates for local positivity, that is, for polynomials that take nonnegative values when restricted on a given semialgebraic set.

While not every nonnegative polynomial is a sum of squares of polynomials, by generalizing the certificate to rational functions (hence embedding $\mathbb{R}[x]$ in $\mathbb{R}(x)$) one gets the following powerful result.

Theorem 14 (Artin 1927)**.** *Every nonnegative polynomial is a sum of squares of rational functions.*

For instance, the Motzkin polynomial is not SOS but has the following *Artin representation* (as sum of squares of rational functions):

$$M(x,y) = \left(\frac{xy(1-x^2)}{x^2+y^2}\right)^2 + \left(\frac{y^2(1-x^2)}{x^2+y^2}\right)^2 + \left(\frac{x^2(1-y^2)}{x^2+y^2}\right)^2 + \\ + \left(\frac{xy(1-y^2)}{x^2+y^2}\right)^2 + \left(\frac{x^2y(x^2+y^2-2)}{x^2+y^2}\right)^2 + \left(\frac{xy^2(x^2+y^2-2)}{x^2+y^2}\right)^2.$$

$\boxed{\text{3h00}}$

References

[1] Anjos, M. F., and Lasserre, J.-B. *Handbook on semidefinite, conic and polynomial optimization*. International Series in Operations Research & Management Science, vol. 166. Springer US, 2012.

[2] Blekherman, G., Parrilo, P., and Thomas, R. *Semidefinite optimization and convex algebraic geometry*, vol. 13. Siam, 2013.

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

[4] Lasserre, J.-B., Ed. *Moments, positive polynomials and their applications*, vol. 1 of *Optimization Series*. Imperial College Press, London, 2010.

[5] Lasserre, J.-B. *An introduction to polynomial and semi-algebraic optimization*. Vol. 52. Cambridge University Press, 2015.

[6] Parrilo, P. A. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. Dissertation (Ph.D.), California Institute of Technology, 2000.