Cambridge University Press
978-1-107-06057-9 - An Introduction to Polynomial and Semi-Algebraic Optimization
Jean Bernard Lasserre
Excerpt
More information

1

Introduction and message of the book

1.1 Why polynomial optimization?

Consider the global optimization problem:

$$\mathbf{P}: \qquad f^* := \inf_{\mathbf{x}} \{ f(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \} \tag{1.1}$$

for some feasible set

$$\mathbf{K} := \{ \mathbf{x} \in \mathbb{R}^n : \; g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}, \tag{1.2}$$

where $f, g_j : \mathbb{R}^n \to \mathbb{R}$ are some continuous functions.

If one is only interested in finding a *local* (as opposed to *global*) minimum then **P** is a Nonlinear Programming (NLP) problem for which several methods and associated algorithms are already available.

But in this book we insist on the fact that **P** is a *global* optimization problem, that is, f^* is the *global* minimum of f on **K**. In full generality problem (1.1) is very difficult and there is no general purpose method, even to approximate f^*.

However, and this is one of the messages of this book, if one now restricts oneself to *Polynomial Optimization*, that is, optimization problems **P** in (1.1) with the restriction that:

f and $g_j : \mathbb{R}^n \to \mathbb{R}$ are all polynomials, $j = 1, \ldots, m$,

then one may approximate f^* as closely as desired, and sometimes solve **P** exactly. (In fact one may even consider *Semi-Algebraic Optimization*, that is, when f and g_j are semi-algebraic functions.) That this is possible is due to the conjunction of two factors.

- On the one hand, *Linear Programming* (LP) and *Semidefinite Programming* (SDP) have become major tools of convex optimization and today's powerful LP and SDP software packages can solve highly nontrivial problems of relatively large size (and even linear programs of extremely large size).
- On the other hand, remarkable and powerful representation theorems (or positivity certificates) for polynomials that are positive on sets like $\mathbf{K}$ in (1.2) were produced in the 1990s by real algebraic geometers and, importantly, the resulting conditions can be checked by solving appropriate semidefinite programs (and linear programs for some representations)!

And indeed, in addition to the usual tools from *Analysis*, *Convex Analysis* and *Linear Algebra* already used in optimization, in Polynomial Optimization *Algebra* may also enter the game. In fact one may find it rather surprising that algebraic aspects of optimization problems defined by polynomials have not been taken into account in a systematic manner earlier. After all, the class of linear/quadratic optimization problems is an important subclass of Polynomial Optimization! But it looks as if we were so familiar with linear and quadratic functions that we forgot that they are polynomials! (It is worth noticing that in the 1960s, Gomory had already introduced some algebraic techniques for attacking (pure) linear integer programs. However, the algebraic techniques described in the present book are different as they come from *Real Algebraic Geometry* rather than pure algebra.)

Even though Polynomial Optimization is a restricted class of optimization problems, it still encompasses a lot of important optimization problems. In particular, it includes the following.

- Continuous convex and nonconvex optimization problems with linear and/or quadratic costs and constraints, for example

  $$\inf_{\mathbf{x}} \{ \mathbf{x}^T\mathbf{A}_0\mathbf{x} + \mathbf{b}_0^T\mathbf{x} : \ \mathbf{x}^T\mathbf{A}_j\mathbf{x} + \mathbf{b}_j^T\mathbf{x} - c_j \geq 0, \quad j = 1, \ldots, m \},$$

  for some scalars c_j, $j = 1, \ldots, m$, and some real symmetric matrices $\mathbf{A}_j \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{b}_j \in \mathbb{R}^n$, $j = 0, \ldots, m$.
- 0/1 optimization problems, modeling a Boolean variable $x_i \in \{0, 1\}$ via the quadratic polynomial constraint $x_i^2 - x_i = 0$. For instance, the celebrated MAXCUT problem is the polynomial optimization problem

$$\sup_{\mathbf{x}} \{ \mathbf{x}^T \mathbf{A} \mathbf{x} : \; x_i^2 - x_i = 0, \quad i = 1, \dots, n \},$$

where the real symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is associated with some given graph with n vertices.

- Mixed-Integer Linear and NonLinear Programming (MILP and MINLP), for instance:

$$\inf_{\mathbf{x}} \left\{ \mathbf{x}^T A_0 \mathbf{x} + \mathbf{b}_0^T \mathbf{x} \;:\; \begin{array}{l} \mathbf{x}^T \mathbf{A}_j \mathbf{x} + \mathbf{b}_j^T \mathbf{x} - c_j \geq 0, \; j = 1, \dots, m; \\ \mathbf{x} \in [-M, M]^n; \\ x_k \in \mathbb{Z}, \; k \in J \end{array} \right\},$$

for some real symmetric matrices $\mathbf{A}_j \in \mathbb{R}^{n \times n}$, vectors $\mathbf{b}_j \in \mathbb{R}^n$, $j = 0, \dots, m$, and some subset $J \subseteq \{1, \dots, n\}$. Indeed, it suffices to model the constraint $x_i \in [-M, M]$, $i \notin J$, with the quadratic inequality constraints $M - x_i^2 \geq 0$, $j \notin J$, and the integrality constraints $x_k \in \mathbb{Z} \cap [-M, M]$, $k \in J$, with the polynomial equality constraints:

$$(x_k + M) \cdot (x_k + M - 1) \cdots x_k \cdot (x_k - 1) \cdots (x_k - M) = 0, \qquad k \in J.$$

1.2 Message of the book

We have already mentioned one message of the book.

- Polynomial Optimization indeed deserves a special treatment because its algebraic aspects can be taken into account in a systematic manner by invoking powerful results from real algebraic geometry.

But there are other important messages.

1.2.1 Easyness

A second message of the book which will become clear in the next chapters, is that the methodology for handling polynomial optimization problems **P** as defined in (1.1) is rather simple and easy to follow.

- Firstly, solving a polynomial optimization problem (1.1) is trivially equivalent to solving

$$f^* = \sup_{\lambda} \{ \lambda : \; f(\mathbf{x}) - \lambda \geq 0, \quad \forall \mathbf{x} \in \mathbf{K} \}, \tag{1.3}$$

which, if f is a polynomial of degree at most d, is in turn equivalent to solving

$$f^* = \sup_{\lambda} \{ \lambda : f - \lambda \in C_d(\mathbf{K}) \}, \tag{1.4}$$

where $C_d(\mathbf{K})$ is the convex cone of polynomials of degree at most d which are nonnegative on $\mathbf{K}$. But (1.4) is a finite-dimensional *convex* optimization problem. Indeed, a polynomial $f : \mathbb{R}^n \to \mathbb{R}$ of degree d is encoded by its vector $\mathbf{f} \in \mathbb{R}^{s(d)}$ of coefficients (e.g. in the usual canonical basis of monomials), where $s(d) := \binom{n+d}{n}$ is the dimension of the vector space of polynomials of degree at most d (that is the number of monomials $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $\boldsymbol{\alpha} \in \mathbb{N}^n$, such that $\sum_{i=1}^n \alpha_i \leq d$). And so $C_d(\mathbf{K})$ is a finite-dimensional cone which can be viewed as (or identified with) a convex cone of $\mathbb{R}^{s(d)}$. Therefore (with some abuse of notation) (1.4) also reads

$$f^* = \sup_{\lambda} \left\{ \lambda : \mathbf{f} - \lambda \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \in C_d(\mathbf{K}) \right\}, \tag{1.5}$$

a convex conic optimization problem in $\mathbb{R}^{s(d)}$. Note in passing that the convex formulations (1.3) and (1.5) are proper to the global optimum f^* and are *not* valid for a local minimum $\hat{f} > f^*$! However, (1.5) remains hard to solve because in general there is no simple and *tractable* characterization of the convex cone $C_d(\mathbf{K})$ (even though it is finite dimensional).

Then the general methodology that we use follows a simple idea. We first define a (nested) increasing family of convex cones $(C_d^\ell(\mathbf{K})) \subset C_d(\mathbf{K})$ such that $C_d^\ell(\mathbf{K}) \subset C_d^{\ell+1}(\mathbf{K})$ for every ℓ, and each $C_d^\ell(\mathbf{K})$ is the projection of either a polyhedral cone or the intersection of a subspace with the convex cone of positive semidefinite matrices (whose size depends on ℓ). Then we solve the hierarchy of conic optimization problems

$$\rho_\ell = \sup_{\lambda} \left\{ \lambda : \mathbf{f} - \lambda \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \in C_d^\ell(\mathbf{K}) \right\}, \quad \ell = 0, 1, \ldots \tag{1.6}$$

For each fixed ℓ, the associated conic optimization problem is convex and can be solved efficiently by appropriate methods of convex optimization. For instance, by using some appropriate *interior points* methods, (1.6) can be solved to arbitrary precision fixed in advance, in time polynomial in its input size. As the $C_d^\ell(\mathbf{K})$ provide a nested sequence of inner approximations of $C_d(\mathbf{K})$, $\rho_\ell \leq \rho_{\ell+1} \leq f^*$ for every ℓ. And the $C_d^\ell(\mathbf{K})$ are chosen so as to ensure the convergence $\rho_\ell \to f^*$ as $\ell \to \infty$. So depending on which type of convex approximation is used, (1.6) provides a *hierarchy* of linear or semidefinite programs (of increasing size) whose respective associated sequences of optimal values both converge to the desired global optimum f^*.

- Secondly, the powerful results from Real Algebraic Geometry that we use to justify the convergence $\rho_\ell \to f^*$ in the above methodology, are extremely simple to understand and could be presented (without proof) in undergraduate courses of Applied Mathematics, Optimization and/or Operations Research. Of course their proof requires some knowledge of sophisticated material in several branches of mathematics but we will *not* prove such results, we will only *use* them! After all, the statement in Fermat's theorem is easy to understand and this theorem may be used with no need to understand its proof.

  For illustration and to give the flavor of one such important and powerful result, we will repeatedly use the following result which states that a polynomial f which is (strictly) positive on $\mathbf{K}$ (as defined in (1.2) and compact) can be written in the form

$$f(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \qquad \forall\, \mathbf{x} \in \mathbb{R}^n, \tag{1.7}$$

  for some polynomials σ_j that are *Sums of Squares* (SOS). By SOS we mean that each σ_j, $j = 0, \ldots, m$, can be written in the form

$$\sigma_j(\mathbf{x}) = \sum_{k=1}^{s_j} h_{jk}(\mathbf{x})^2, \qquad \forall\, \mathbf{x} \in \mathbb{R}^n,$$

  for finitely many polynomials h_{jk}, $k = 1, \ldots, s_j$.

  As one may see, (1.7) provides f with a *certificate* of its positivity on $\mathbf{K}$. This is because if $\mathbf{x} \in \mathbf{K}$ then $f(\mathbf{x}) \geq 0$ follows immediately from (1.7) as $\sigma_j(\mathbf{x}) \geq 0$ (because σ_j is SOS) and $g_j(\mathbf{x}) \geq 0$ (because $\mathbf{x} \in \mathbf{K}$), for all j. In other words, there is no need to check the positivity of f on $\mathbf{K}$ as one may read it directly from (1.7)!
- Finally, the convex conic optimization problem (1.4) has a *dual* which is another finite-dimensional convex conic optimization problem. And in fact this classical duality of convex (conic) optimization captures and illustrates the beautiful duality between *positive polynomials* and *moment problems*. We will see that the dual of (1.4) is particularly useful for extracting global minimizers of **P** when the convergence is finite (which, in addition, happens generically!). Depending on which type of positivity certificate is used we call this methodology the *moment-LP* or *moment-SOS* approach.

1.2.2 A general methodology

The class of polynomial optimization problems contains "easy" convex problems (e.g. Linear Programming and convex Quadratic Programming) as well as NP-hard optimization problems (e.g. the MAXCUT problem already mentioned).

> Still, the general methodology presented in this book *does not distinguish* between easy convex problems and nonconvex, discrete and mixed-integer optimization problems!

This immediately raises the following issues.

How effective can a general purpose approach be for addressing problems which can be so different in nature (e.g. convex, or continuous but possibly with a nonconvex and nonconnected feasible set, or discrete, or mixed-integer, etc.)?

- *Should we not specialize the approach according to the problem on hand, with ad hoc methods for certain categories of problems?*
- *Is the general approach reasonably efficient when applied to problems considered easy? Indeed would one trust a general purpose method designed for hard problems, and which would not behave efficiently on easy problems?*

Indeed, a large class of convex optimization problems are considered "easy" and can be solved efficiently by several ad hoc methods of convex optimization. Therefore a highly desirable feature of a general purpose approach is the ability somehow to *recognize* easier convex problems and behave accordingly (even if this may not be as efficient as specific methods tailored to the convex case).

A third message of this book is that this is indeed the case for the moment-SOS approach based on semidefinite relaxations, which uses representation results of the form (1.7) based on SOS. This is not the case for the moment-LP approach based on LP-relaxations, which uses other representation results.

In our mind this is an important and remarkable feature of the moment-SOS approach. For instance, and as already mentioned, a Boolean variable x_i is not treated with any particular attention and is modeled via the quadratic equality constraint $x_i^2 - x_i = 0$, just one among the many other polynomial equality or inequality constraints in the definition of the feasible set $\mathbf{K}$ in (1.2). Running a local minimization algorithm of continuous optimization with such a modeling of a Boolean constraint would not be considered wise (to say the least)! Hence this might justify some doubts concerning the efficiency of the moment-SOS approach by lack of specialization. Yet, and remarkably, the resulting semidefinite relaxations

- still provide the strongest relaxation algorithms for hard combinatorial optimization problems, and
- *recognize* easy convex problems as in this latter case convergence is even finite (and sometimes at the first semidefinite relaxation of the hierarchy)!

Of course, and especially in view of the present status of semidefinite solvers, the moment-SOS approach is still limited to problems of modest size; however, if symmetries or some structured sparsity in the problem data are detected and taken into account, then problems of much larger size can be solved.

1.2.3 Global optimality conditions

Another message of this book is that in the moment-SOS approach, *generically* the convergence $\rho_\ell \to f^*$ as $\ell \to \infty$ is finite (and genericity will be given a precise meaning)!

And so in particular, generically, solving a polynomial optimization problem (1.1) on a compact set $\mathbf{K}$ as in (1.2) reduces to solving a *single* semidefinite program.

But of course and as expected, the size of the resulting semidefinite program is not known in advance and can be potentially large.

Moreover, the powerful Putinar representation (1.7) of polynomials that are (strictly) positive on $\mathbf{K}$ also holds generically for polynomials that are only nonnegative on $\mathbf{K}$. And this translates into *global optimality* conditions that must be satisfied by global minimizers, provided that each global minimizer satisfies standard well-known constraint qualification, strict complementarity and second-order sufficiency conditions in nonlinear programming. Again, such conditions hold generically.

Remarkably, these optimality conditions are the perfect analogues in (nonconvex) polynomial optimization of the Karush–Kuhn–Tucker (KKT) optimality conditions in convex programming. In particular, and in contrast to the KKT conditions in nonconvex programming, the constraints that are important but not active at a global minimizer still play a role in the optimality conditions.

1.2.4 Extensions

The final message is that the above methodology can also be applied in the following situations.

- To handle *semi-algebraic* functions, a class of functions much larger than the class of polynomials. For instance one may handle functions like

$$f(\mathbf{x}) := \sqrt{\min[q_1(\mathbf{x}), q_2(\mathbf{x})] - \max[q_3(\mathbf{x}), q_4(\mathbf{x})] + (q_5(\mathbf{x}) + q_6(\mathbf{x}))^{1/3}},$$

  where the q_i are given polynomials.
- To handle extensions like *parametric* and *inverse* optimization problems.
- To build up polynomial *convex underestimators* of a given nonconvex polynomial on a box $\mathbf{B} \subset \mathbb{R}^n$.
- To approximate as closely as desired, sets defined with quantifiers, for example the set

$$\{\mathbf{x} \in \mathbf{B} : \; f(\mathbf{x}, \mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x}, \mathbf{y}) \in \mathbf{K}\},$$

  where $\mathbf{K} \subset \mathbb{R}^{n+p}$ is a set of the form (1.2), and $\mathbf{B} \subset \mathbb{R}^n$ is a simple set.

1.3 Plan of the book

The book is divided into three main parts.

Part I is introductory and Chapter 2 is dedicated to presenting basic and important results on the representation of polynomials that are positive on a subset $\mathbf{K}$ of $\mathbb{R}^n$. This problem of real algebraic geometry has a nice *dual facet*, the so-called $\mathbf{K}$-moment problem in Functional Analysis and Probability. And so results on the algebraic side are complemented with their dual counterparts on the moment side. Interestingly, convex duality in optimization (applied to appropriate convex cones) nicely captures this duality. Chapter 3 describes another characterization of polynomials nonnegative on a closed set $\mathbf{K}$ which is of independent interest and is *dual* to the characterizations of Chapter 2. Now knowledge on $\mathbf{K}$ is from moments of a measure supported on $\mathbf{K}$ rather than from polynomials that describe the boundary of $\mathbf{K}$. These two dual points of view are exploited in Chapter 4 to provide explicit outer and inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$.

Part II is dedicated to polynomial and semi-algebraic optimization. It describes how to use results of Part I to define hierarchies of convex relaxations whose optimal values provide monotone sequences of lower bounds which converge to the global optimum. Depending on the type of representation (or positivity certificate) used, one obtains a hierarchy of linear programs or semidefinite programs. Their respective merits and drawbacks are analyzed, especially in the light of global optimality conditions. In particular, we describe a global optimality condition which is the exact analogue in nonconvex polynomial optimization of the celebrated KKT optimality conditions in convex optimization.

Using the representation results described in Chapter 3, one also obtains a hierarchy of eigenvalue problems which provide a monotone sequence of upper bounds which converges to the global optimum. Notice that most (primal-type) minimization algorithms provide sequences of upper bounds on the global minimum (as they move from a feasible point to another feasible point) but in general their convergence (if it eventually takes place) is guaranteed to a local minimum only. We also (briefly) describe how to use sparsity or symmetry to reduce the computational burden associated with the hierarchy of relaxations. It is worth noticing that the extension from polynomials to semi-algebraic functions (in both the objective function and the description of the feasible set) enlarges significantly the range of potential applications that can be treated.

Part III describes some specializations and extensions.

- *Convex polynomial optimization* to show that the moment-SOS approach somehow "recognizes" some classes of easy convex problems; in particular the hierarchy of semidefinite relaxations has finite convergence. Some properties of convex polynomials and convex basic semi-algebraic sets are also described and analyzed.
- *Parametric polynomial optimization*, that is, optimization problems where the criterion to minimize as well as the constraints that describe the feasible set, may depend on some parameters that belong to a given set. The ultimate and difficult goal is to compute or at least provide some information and/or approximations on the global optimum and the global minimizers, viewed as *functions* of the parameters. Hence there is a qualitative jump in difficulty as one now searches for *functions* on some domain (an infinite-dimensional object) rather than a vector $\mathbf{x} \in \mathbb{R}^n$ (a finite-dimensional object). With this in mind we describe what we call the "joint+marginal" approach to parametric optimization and show that in this context the moment-SOS approach is well suited for providing good approximations (theoretically as closely as

desired) when the parametric optimization problem is described via polynomials and basic semi-algebraic sets.

- *Inverse polynomial optimization* where given a point $\mathbf{y} \in \mathbf{K}$ (think of an iterate of some local minimization algorithm) and a polynomial criterion f, one tries to find a polynomial $\tilde{f}$ as close as possible to f and for which $\mathbf{y}$ is a global minimizer on $\mathbf{K}$. This problem has interesting potential theoretical and practical applications. For instance, suppose that $\mathbf{y} \in \mathbf{K}$ is a current iterate of a local optimization algorithm to minimize f on $\mathbf{K}$, and suppose that solving the inverse problem provides a new criterion $\tilde{f}$ relatively close to f. Should we spend (expensive) additional effort to obtain a better iterate or should we stop (as $\mathbf{y}$ solves an optimization problem close to the original one)? In addition, the inverse problem is also a way to measure how ill-conditioned is the (direct) optimization problem.
- *Convex underestimators*. For difficult large scale nonlinear problems and particularly for mixed-integer nonlinear programs (MINLP), the only practical way to approximate the global minimum is to explore an appropriate Branch and Bound search tree in which exploration is guided by lower bounds computed at each node of the tree. The quality of lower bounds is crucial for the overall efficiency of the approach. In general and for obvious reasons, efficient computation of lower bounds is possible only on some appropriate convex relaxation of the problem described at the current node. A standard way to obtain a convex relaxation is to replace the nonconvex objective function with a convex underestimator; similarly, an inequality constraint $g(\mathbf{x}) \leq 0$ is replaced with $\tilde{g}(\mathbf{x}) \leq 0$ for some convex underestimator $\tilde{g}$ of g. Therefore deriving *tight* convex underestimators is of crucial importance for the quality of the resulting lower bounds. We show that the moment-SOS approach is particularly well suited to obtaining tight convex underestimators.
- *Polynomial optimization on sets defined with quantifiers*. In this context one is given a set $\mathbf{K} := \{(\mathbf{x},\mathbf{y}) : g_j(\mathbf{x},\mathbf{y}) \geq 0,\ j = 1,\dots,m\}$ for some polynomials g_j, $j = 1,\dots,m$. Given a set $\mathbf{B} \subset \mathbb{R}^n$ (typically a box or an ellipsoid) the goal is to approximate as closely as desired the set

$$\mathbf{K}_{\mathbf{x}} := \{\mathbf{x} \in \mathbf{B} : f(\mathbf{x},\mathbf{y}) \leq 0 \text{ for all } \mathbf{y} \text{ such that } (\mathbf{x},\mathbf{y}) \in \mathbf{K}\},$$

by a set $\mathbf{K}^{\ell}_{\mathbf{x}} \subset \mathbf{K}_{\mathbf{x}}$ simply defined by $\mathbf{K}^{\ell}_{\mathbf{x}} = \{\mathbf{x} : h_\ell(\mathbf{x}) \leq 0\}$ for some polynomial h_ℓ. We show how to obtain a sequence of polynomials (h_ℓ) of increasing degree $\ell \in \mathbb{N}$, such that the Lebesgue volume $\mathrm{vol}(\mathbf{K}_{\mathbf{x}} \setminus \mathbf{K}^{\ell}_{\mathbf{x}})$ tends to zero as $\ell \to \infty$. And so any optimization problem involving the set $\mathbf{K}_{\mathbf{x}}$ (difficult to handle) can be approximated by substituting $\mathbf{K}_{\mathbf{x}}$ with $\mathbf{K}^{\ell}_{\mathbf{x}}$ for some sufficiently large ℓ.