Non-Convex Optimization

Constantine Caramanis*
Massachusetts Institute of Technology

November 29, 2001

This is a "brief" overview of the summary on nonconvex optimization via real algebraic geometry. Many details are left out, yet hopefully this paper still gives a reasonable overview of the main ideas and flavor of the methods. This is a current research interest of the author. I would like to acknowledge Pablo Parrilo for several enlightening conversations, and email exchanges.

1 Introduction

The framework of optimization of a general function over an arbitrary set is very powerful, in the sense that many problems from applied mathematics and engineering may be formulated in that framework. Many problems in communication, information theory, control, complexity, and logic, to name a few fields, may be recast as an optimization problem. As we might expect from a framework this flexible, the general optimization problem is intractable. We seek to introduce additional structure which we may exploit. In this paper we address the optimization of polynomials subject to set constraints carved out by polynomial inequalities, equalities, and nonequalities. Examples of such set constraints are intersections of conic sections and half planes, polyhedra, but also discrete sets, like $\{0,1\}^n$. Hence we consider a wide variety of sets. Thus the main problem we consider here is,

$$\begin{aligned}
\text{minimize}: \quad & F(x_1,\ldots,x_n) \\
\text{s.t.}: \quad & x=(x_1,\ldots,x_n)\in K, \\
& K := \{x\in\mathbb{R}^n \;:\; f_i(x)\geq 0, g_j(x)\neq 0, h_k(x)=0, i=1,\ldots,m_1, \\
& \qquad\qquad j=1,\ldots,m_2, k=1,\ldots,m_3\}.
\end{aligned}$$

2 An Algebraic Approach

Most optimization methods are essentially no more than glorified versions of a method we learn in introductory calculus: take a function's derivative and solve for its zeros. Methods based on this technique take advantage of the differential structure of the function, in our case, of the polynomial. The typical problems encountered with such calculus based methods arise from the fact that the derivative is very local in nature, and hence we face the problem of global versus local optimization, which is exactly the difficult issue in nonconvex optimization. In this paper we provide an exposition of a new method that exploits the fact that a polynomial, in addition to being a continuous function, is also an element of the polynomial ring $\mathbb{R}[x_1,\ldots,x_n]$, and hence we try to take a new perspective, and exploit its algebraic structure. See Parrilo ([7], [8], [9]) for a nice development of the methods described below.

A Simple First Case

Consider a univariate polynomial, $f(x)\in\mathbb{R}[x]$, that is bounded from below. Other than that

*cmcaram@mit.edu

restriction, we do not place any restrictions on the polynomial $f(x)$, so it is free to be as nonconvex as it pleases. We will use the following fact.

Fact: A univariate polynomial $g(x)$ is nonnegative on all of $\mathbb{R}$ if and only if it can be written as a sum of squares of other polynomials. In other words,

$$g(x) \geq 0,\ \forall x \in \mathbb{R}, \iff g(x) = h_1^2(x) + \cdots + h_k^2(x),$$

for some polynomials $h_1, \ldots, h_k$. If a polynomial $g(x)$ can be written as a sum of squares, we write $g(x) \in \Sigma^2$.

Therefore, we replace the pointwise property "$g(x)$ nonnegative" with the algebraic property, "g can be written as a sum of squares of other polynomials". We use this equivalence to rephrase the problem of finding the global minimum of f over $\mathbb{R}$ as an algebraic problem:

$$\begin{array}{rl} \text{minimize}: & f(x) \\ \text{s.t.}: & x \in \mathbb{R}, \\ & \quad\Longleftrightarrow \\ \text{maximize}: & \gamma \\ \text{s.t.}: & f(x) - \gamma \geq 0,\ \ \forall x \in \mathbb{R}, \\ & \quad\Longleftrightarrow \\ \text{maximize}: & \gamma \\ \text{s.t.}: & (f(x) - \gamma) \in \Sigma^2. \end{array}$$

Thus we have translated the problem to that of finding the largest γ such that a sum of squares decomposition exists. That this equivalent problem is computationally tractable follows from the following fact. For the proof of this fact, see the full review, [2].

Fact: The membership test $f(x) \in \Sigma^2$ can be computed, and the actual sum of squares decomposition may be obtained, by performing a semidefinite optimization of size $\binom{n+d}{d}$, where d is the degree of the polynomial, and n is the number of variables.

Note then that we can solve the nonconvex problem of global minimization of a univariate polynomial, by solving a convex problem, namely, the corresponding semidefinite optimization problem.

3 The Cones Σ^2 and $\mathcal{P}_+$

We have defined the cone of sum of squares, Σ^2. Now we define the cone of polynomials, nonnegative on a set $K \subseteq \mathbb{R}^n$.

Definition 1 *We define the cone $\mathcal{P}_+(K)$, the cone of polynomials nonnegative on a set K, as (the underlying polynomial ring is always understood from the context):*

$$\mathcal{P}_+(K) = \{p(x) \in \mathbb{R}[x_1, \ldots, x_n] \,:\, p(x) \geq 0, \forall x \in K\}.$$

It is clear that in general we have $\Sigma^2 \subseteq \mathcal{P}_+(K)$, for every set $K \subseteq \mathbb{R}^n$, and in particular for $K = \mathbb{R}^n$. The fact given above states that for the ring of univariate polynomials, $\mathbb{R}[x]$, we have, $\Sigma^2 = \mathcal{P}_+ := \mathcal{P}_+(\mathbb{R})$. In general, however, this inclusion is strict, and therefore it is not in general true that $f \geq 0$ is equivalent to $f \in \Sigma^2$. Therefore the procedure outlined in the example above will in general only provide a lower bound on the minimum of the polynomial. Consider the following two examples, which are derived from the famous Motzkin polynomial,

$$M(x, y, z) = x^2y^4 + x^4y^2 + z^6 - 3x^2y^2z^2,$$

which was the first concrete example of a polynomial that is globally nonnegative, but cannot be written as a sum of squares (see [11] for a nice summary of the history and the theory relating to what is known as Hilbert's 17^{th} problem). We have,

$$\begin{aligned} f(x,z) &= M(x,1,z) = x^4 + x^2 + z^6 - 3x^2z^2, \\ g(x,y) &= M(x,y,1) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1. \end{aligned}$$

Both these polynomials are nonnegative on all $\mathbb{R}^2$, but neither can be written as a sum of squares. However,

$$f(x,z) + \frac{729}{4096} = (-\frac{9}{8}z + z^3)^2 + (\frac{27}{64} + x^2 - \frac{3}{2}z^2)^2 + \frac{5}{32}x^2.$$

Therefore asking for the maximum γ such that $f(x,z) - \gamma \in \Sigma^2$, provides a lower bound on the minimum of $f(x,z)$. The second polynomial above, however, is an example where this gap is infinite, and hence this method provides no information about the minimum.

What we have seen is that our minimization problem is equivalent to computing the membership test for the cone $\mathcal{P}_+(K)$. This, however, is in general not tractable. The membership test in Σ^2, on the other hand, is tractable (again, see [2] for the proof of this). The idea, then, is to use Σ^2 to understand $\mathcal{P}_+(K)$. When we provide a sum of squares decomposition for some polynomial $g(x)$, we can think of what we are doing as providing a polynomial *certificate* that $g(x) \in \mathcal{P}_+$. There are many such types of certificates. For instance, suppose we can write,

$$\begin{aligned} g(x) &= f_1(x) + h_1(x)u(x) \\ &= f_2(x) - h_2(x)u(x), \end{aligned}$$

where $u(x)$ is some arbitrary polynomial, and $f_1, f_2, h_1, h_2 \in \Sigma^2$. These five polynomials, and the equations above, provide a polynomial certificate that $g(x) \in \mathcal{P}_+$.

4 The Positivstellensatz

In 1974, G. Stengle proved a theorem called the Positivstellensatz (see [12]) that proves that there exist certificates of membership in $\mathcal{P}_+(K)$ of a particular form, involving elements of Σ^2. Before we give the statement of the theorem, we introduce three algebraic subsets of a ring.

Definition 2 (Algebraic Cone) *The Algebraic Cone in a commutative ring R generated by the elements $\beta_1, \ldots, \beta_n \in R$, is the set of elements,*

$$\mathcal{A}(\beta_1, \ldots, \beta_n) := \{f \in R \,:\, f = \alpha + \sum_{I \subset \{1,\ldots,n\}} \alpha_I \prod_{i \in I} \beta_i\},$$

where α, and α_I are sums of squares of elements of R.

This coincides with this alternate definition:

Definition 3 *For R as above, a subset P of R is called a cone if it satisfies the following;*

1. $\alpha, \beta \in P \Rightarrow \alpha + \beta \in P$,
2. $\alpha, \beta \in P \Rightarrow \alpha \cdot \beta \in P$,
3. $\alpha \in R \Rightarrow \alpha^2 \in P$.

In fact, it is clear that $\Sigma^2 \subseteq R$, the set of elements that are sums of squares of other elements, is the smallest cone in any ring R. Note that this is an algebraic cone, and it is different from the concept of a geometric cone, in the sense introduced above.

Definition 4 (Multiplicative Monoid) *The multiplicative monoid generated by elements $\beta_1, \ldots, \beta_n$, is the set of (finite) products of the elements β_i, including the empty product, which by convention equals 1 (the identity in R). We denote it by $\mathcal{M}(\beta_1, \ldots, \beta_n)$.*

Definition 5 (Ideal) *An ideal, $I \subseteq R$ is a set of elements closed under addition with elements of I, and closed under multiplication by elements of R.*

Now we can state Stengle's Positivstellensatz.

Theorem 1 (Positivstellensatz) *Given polynomials $\{f_1, \ldots, f_{n_1}\}$, $\{g_1, \ldots, g_{n_2}\}$, $\{h_1, \ldots, h_{n_3}\}$, elements of $\mathbb{R}[x_1, \ldots, x_n]$, the following are equivalent:*

1. *The set*

$$K := \{x \in \mathbb{R}^n \,:\, f_i(x) \geq 0, g_j(x) \neq 0, h_k(x) = 0, i = 1, \ldots, n_1, j = 1, \ldots, n_2, k = 1, \ldots n_3\},$$

is empty,

2. *There exist polynomials $F \in \mathcal{A}(f_1, \ldots, f_{n_1})$, $G \in \mathcal{M}(g_1, \ldots, g_{n_2})$, and $H \in I(h_1, \ldots, h_{n_3})$, such that we have the identity*

$$F + G^2 + H \equiv 0.$$

A set that is defined by polynomial inequalities, equalities, and nonequalities, as in the theorem above, is called a **semialgebraic** set. We can think of the Positivstellensatz as giving polynomial certificates that a semialgebraic set K is empty. The three sets we are allowed to use, namely, the cone, the multiplicative monoid, and the ideal, specify the rules of derivation we may use. For example, given an inequality, $f_1 \geq 0$, we obtain new valid inequalities by the specified rules of the cone, the monoid, and the ideal. For instance, if $f \geq 0$, then $f^3 \geq 0$, and $g^2 f \geq 0$, and so on. This is precisely what we do when we have, say, some boolean expression, and some axioms, and we combine the axioms to form new true statements, until we have proved or disproved the satisfiability of the expression.

We can formulate polynomial optimization problems using the Positivstellensatz easily. Generically, suppose we want to minimize a polynomial $f(x)$ over the set

$$K := \{x \in \mathbb{R}^n \,:\, f_i(x) \geq 0, g_j(x) \neq 0, h_k(x) = 0, i = 1, \ldots, n_1, j = 1, \ldots, n_2, k = 1, \ldots n_3\}.$$

This is equivalent to seeking the largest $\gamma \in \mathbb{R}$, for which $f(x) - \gamma \geq 0$ for all $x \in K$, or, equivalently, the largest $\gamma \in \mathbb{R}$ for which the set,

$$K' := K \cap \{x \in \mathbb{R}^n \,:\, \gamma - f(x) \geq 0, f(x) - \gamma \neq 0\},$$

is empty. Using the theorem above, we can translate the optimization of the polynomial f over the set K, to the problem:

Certificate Formulation: Find the largest $\gamma \in \mathbb{R}$, such that there exist polynomials

1. $F \in \mathcal{A}(\gamma - f, f_1, \ldots, f_{n_1})$,
2. $G \in \mathcal{M}(f - \gamma, g_1, \ldots, g_{n_1})$,
3. $H \in I(h_1, \ldots, h_{n_3})$,

such that the identity

$$F + G^2 + H \equiv 0,$$

holds. Then, analogously to our first simple example of global minimization of a univariate polynomial over $\mathbb{R}$, here we translate the pointwise property, "$\geq 0, \forall x \in K$" to the algebraic property, "$\exists\, F, G, H$ s.t. $F + G^2 + H \equiv 0$".

As it stands, however, the Positivstellensatz is nonconstructive, in the sense that it does not provide an algorithmic approach to obtain the certificates in its statement. We want to exploit the connection to semidefinite optimization, which efficiently answers the membership question, and moreover, provides the actual sum of squares decomposition in the process. Consider the example given above. There is a major obstacle that keeps us from concluding that we can solve small semidefinite optimization problems to obtain exact solutions to arbitrary nonconvex polynomial optimization problems (and hence save us from concluding the $\mathcal{P} = \mathcal{NP}$). While the Positivstellensatz does guarantee the existence of a polynomial certificate of the form given in the statement of the theorem, there are no *a priori* guarantees on the degree of the polynomials F, G, H used in the certificate. Thus, while (as described in the full summary) semidefinite optimization yields certificates in time polynomial in their degree, if this degree is, say, exponential in the size of the problem, then the computational cost is prohibitive. Therefore we use the Positivstellensatz to obtain a sequence of relaxations, indexed by the degree of certificate for which we search.

4.1 Putinar's Positivstellensatz

Under certain restrictions on the semialgebraic set K, we can formulate a stronger, in some sense, form of the Positivstellensatz. This is due to Putinar's 1993 result ([10]).

Theorem 2 (Putinar) *Suppose we are given a set,*

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

that is compact, and furthermore satisfies the condition that there exists a polynomial $h(x)$ of the form,

$$h(x) = s_0(x) + \sum_{i=1}^{m} s_i(x) \cdot f_i(x),$$

where the s_i are sums of squares, whose level set

$$\{x \in \mathbb{R}^n \, : \, h(x) \geq 0\},$$

is compact. Then, for any polynomial $g(x)$ positive on all of K, there exist $s_0, s_1, \dots, s_m$ sums of squares, such that

$$g(x) = s_0(x) + \sum_{i=1}^{m} s_i(x) \cdot f_i(x).$$

We can write this in the form of the Positivstellensatz above, namely, as a certificate of a set K being empty:

Theorem 3 *For K and g as above, the following are equivalent.*

1. *The set*

$$K' := K \cap \{x \in \mathbb{R}^n \, : \, -g(x) \geq 0, g(x) \neq 0\},$$

   is empty,

2. *There exist polynomials $s_0, s_1, \dots, s_m$ sums of squares, such that*

$$g - s_0 - \sum s_i f_i \equiv 0.$$

Optimization problems then become:

$$\begin{aligned} \text{minimize}: \quad & g(x) \\ \text{s.t.}: \quad & x \in K := \{x \in \mathbb{R}^n \, : \, f_i(x) \geq 0, i = 1, \dots, m\}, \end{aligned}$$

equivalently written using Putinar, as

$$\begin{array}{rl} \text{maximize :} & \gamma \\ \text{s.t. :} & \exists s_1, \ldots, s_m \in \Sigma^2, \text{ such that} \\ & \left((g(x) - \gamma) - \sum_{i=1}^{m} s_i(x) f_i(x) \right) \in \Sigma^2. \end{array}$$

We obtain sequential relaxations by limiting the degree of the polynomials $s_i \in \Sigma^2$ which we search for in the representation given above.

As discussed in the complete summary, [2], there are various relative advantages and disadvantages, practical, theoretical, and aesthetic, to the Positivstellensatz and Putinar's theorem.

5 The Moment Approach

There is another approach to optimization that emphasizes Putinar's theorem, and is related to the so–called classical moment problem.

The Moment Problem: Given some set $K \subseteq \mathbb{R}^n$, and a sequence of numbers, $\{m_\alpha\}, \alpha \in \mathbb{N}^n$, does there exist a probability distribution μ, with support contained in K, such that,

$$\int_K x^\alpha \, d\mu = m_\alpha,$$

for every α given. If indeed such a distribution with the specified support exists, then the given moment sequence is said to be a K–valid moment sequence.

This is an old problem in mathematics (see [1]). For the cases $K = \mathbb{R}, \mathbb{R}_+, [a, b] \subset \mathbb{R}$, the problem is known as the Hamburger moment problem, and the solution has been known since the beginning of the 20^{th} century, yet for general K the problem of obtaining necessary and sufficient conditions for validity of the moment sequence remains open. As discussed in the full summary ([2]), the moment problem is intimately tied, in fact is in some sense dual, to the problem of determining membership in $\mathcal{P}_+(K)$. For our purposes, we simply state that Putinar's theorem may be used to obtain successively tighter outer approximations to the set of moment sequences valid for some support set K that satisfies the hypotheses of Putinar's theorem. We now give the application to optimization. This moment approach to optimization is described in detail by Lasserre, in [4], [5], and [6]. The basis for this is the following equivalence, which is well known, and comes from the natural convexification of an arbitrary optimization problem by lifting it to the infinite dimensional space of measures. Let $\mathcal{P}(K)$ denote the set of probability measures with support in the set $K \subseteq \mathbb{R}^n$. Then,

$$\min_{x \in K} f(x) \iff \min_{\mu \in \mathcal{P}(K)} \int_K f(x) \, d\mu.$$

Writing the polynomial $f(x)$ explicitly: $f(x) = \sum f_\alpha x^\alpha$, where α is a multiindex, we have a next step in the above equivalence:

$$\min_{\{m_\alpha\} \in V(K)} \sum f_\alpha m_\alpha,$$

where by $V(K)$ we denote the set of K–valid moment sequences. As discussed above, and in [3], [4], the membership test for $V(K)$ may be intractable. Analogously to the direct, or primal, approach outlined above, we obtain lower bounds to the true minimum by approximating $V(K)$ from the exterior (again, see [2] for a full explanation). There are potentially interesting connections of these intermediate steps with randomized methods that have seen much success in some of their applications. Indeed, Goemans and Williamson's celebrated result on a randomized approach to MAXCUT can be understood in the context of the material described in this paper, and this reformulation suggests various possible extensions and strengthenings of the method (once again...see [2]).

6 Conclusion

These algebraic methods are recently developed tools that exploit an aspect of the mathematical structure of polynomials that previously remained unexplored, and unused, for the purposes of optimization. The preliminary results and performance of these methods on benchmark problems, seem to suggest that these methods are powerful practical tools, or have the potential to become very powerful and efficient tools.

Equally, if not more importantly, these methods are promising because of their apparently deep connections to many other areas, such as complexity theory, proof systems, moment problems, optimization, combinatorial objects, and so on. Morever, at a high level, the idea of approximating computationally intractable objects with computationally tractable objects, is very common, and useful in applications. Long and short certificates of emptyness, or membership, or non–membership, link these problems to the theory of $\mathcal{NP}$ and $co - \mathcal{NP}$.

References

[1] Akhiezer, N., The Classical Moment Problem, Eng. Ed., Oliver and Boyd, London 1965.

[2] Caramanis, C. "Nonconvex Optimization via Real Algebraic Geometry," 6.962 Full Summary, available at http://web.mit.edu/6.962/www, (2001).

[3] Caramanis, C., Unpublished notes on the Identification Problem, available upon request: cmcaram@mit.edu.

[4] Lasserre, J.B. "Optimality Conditions and LMI Relaxations for 0–1 Programs," Technical Report N. 00099, 2000.

[5] Lasserre, J.B., "Global optimization with polynomials and the problem of moments," SIAM J. Opt., 11, 2001.

[6] Lasserre, J.B., "An Explicit Exact SDP Relaxation for Nonlinear 0–1 Programs," In K. Aardal and A.M.H. Gerards, eds., Lecture Notes in Computer Science, 2081:293-303, 2001.

[7] Parrilo, P. "Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization," Ph.D. Thesis, Caltech, May 2000.

[8] Parrilo, P.; Sturmfels, B. "Minimizing Polynomial Functions," available at: http://xyz.lanl.gov/abs/math.oc/0103170, 2001.

[9] Parrilo, P. "Semidefinite Programming Relaxations for Semialgebraic Problems," Math. Prog. to appear.

[10] Putinar, M., "Positive Polynomials on Compact Semi-algebraic Sets," Indiana University Math. J., Vol 42, (3) 1993.

[11] Reznick, R. "Some Concrete Aspects of Hilbert's 17^{th} Problem," Seminaire de Structures Algébriques Ordonnées, (F. Delon, M.A. Dickmann, D. Gondard eds), Publ. Équipe de Logique, Univ. Paris VII, Jan. 1996; revised version in Real Algebraic Geometry and Ordered Structures, (C. N. Delzell, J.J. Madden eds.) Cont. Math., 253 (2000), 251-272 (MR 2001i:11042).

[12] Stengle, G. "A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry," Math. Ann. 207, 87-97 (1974).