Non-Convex Optimization via Real Algebraic Geometry

Constantine Caramanis*
Massachusetts Institute of Technology

November 29, 2001

The following paper represents the material from a collection of different sources, as listed in the bibliography. I would like to acknowledge Pablo Parrilo for several enlightening conversations, and email exchanges.

1 Introduction

The high level purpose of this paper is to describe some recent advances in the field of Mathematics called Real Algebraic Geometry, and discuss some of its applications to complexity theory, and non-convex optimization. In particular, one of the questions underlying the entire development, is the crucial question: What makes an optimization problem difficult or easy? Along the way, we try to point out what we see as the promising research directions, and what we deem the difficulty and potential promise of these various directions to be.

Algebraic Geometry is the branch of mathematics that studies the zero sets of multivariate polynomials over algebraically closed fields, the primary example being copies of the complex plain, $\mathbb{C}^n$. These zero sets are called varieties. Algebraic Geometry has seen various applications in both pure and applied fields of mathematics and engineering, such as number theory, control, and information theory, to name a few. In the last century, as an answer to Hilbert's 17^{th} problem (see section 3 below for more, or [26] for a more comprehensive history) a new flavor of algebraic geometry has developed, called Real Algebraic Geometry, where the requirement that the underlying field be algebraically closed is replaced by the requirement that an ordering exist in the field such that any sum of squares is nonnegative.

Non-convex optimization problems play a central role in many engineering applications, as well as being fundamental to complexity theory. Being non-convex, and often discrete, non-convex problems often defy solution by methods successful in the convex domain. Indeed, many non-convex optimization problems are considered intractable, and as such have deep connections with complexity theory. Starting with some ideas developed by N.Z. Shor in the 80s (see [29]) and then more recently followed up by Parrilo (see [19], [20], [21]) and Lasserre (see [13], [14], and [15]) in the last two years, real algebraic geometry has been used to provide a new approach, indeed a new framework, for many problems, and in particular non-convex optimization.

On a first level, this paper describes the application of these methods to the problem of optimizing an arbitrary polynomial over subsets of Euclidean space defined by polynomial equalities and inequalities. In addition, in this survey paper, we try to present a unified exposition of these recent advances in Real Algebraic Geometry, and their applications to complexity and optimization, bringing together three perspectives: the proof system perspective, where we are interested in providing polynomially verifiable certificates of membership (or lack of membership) in some particular

*cmcaram@mit.edu

family (note that this is crucial to the relation of $\mathcal{NP}$ and co–$\mathcal{NP}$); and then optimization, from the primal approach, and then the dual approach, which illustrates the duality between polynomial optimization, and moment problems.

This paper describes the connection between optimization and nonnegative polynomials. The key to understanding these polynomials is by using polynomials that can be written as sums of squares of other polynomials. This crucial relationship is developed in sections 4 and 5.

In section 2 we discuss why non-convex optimization is hard, and at the same time very important. In section 3 we provide a brief history of Hilbert's 17^{th} problem, and the study of the polynomials that are sums of squares of other polynomials. We also outline the connection to moment problems, which also dates back to before the beginning of the 20^{th} century. Section 4 describes the cone of positive polynomials, the cone of polynomials that are sums of squares of other polynomials, the relation of these two, and their respective computability. Section 5 goes on to describe the general principle of using a computationally tractable set, and some "easily" verifiable certificate, to show membership in a computationally intractable set. Furthermore, using some of the more powerful results of Real Algebraic Geometry that have emerged in the last decade, this section describes how for the case of the cone of sum of squares, and of positive polynomials, certificates as mentioned above may be obtained canonically, hence providing an algorithmically appealing and theoretically unified approach to nonconvex optimization. Section 6 develops the dual perspective, linking the theory of moments with the theory of the cone of positive polynomials. Furthermore, in this section we give some very promising randomized techniques, and connections to some much–studied, but still not particularly well–understood combinatorial optimization problems. Finally, in section 7, we give a high level overview, and discuss some possible future work.

2 Motivation

The field of optimization is concerned with solving the problem:

$$\min_{x \in \mathcal{X}} f(x),$$

(note that minimization, as opposed to maximization, presents no loss of generality). The statement of the problem as such is very general. We have as of yet placed no structural restrictions on either the function $f(x)$, or the set of values $\mathcal{X}$ over which we search for an optimizing x. Indeed, the problem of optimizing an arbitrary function over an arbitrary set, can be thought of as a very general framework, into which a wide array of problems may be cast. Certainly problems in control theory, in communications, very much including information theory, which after all explicitly features optimization in its capacity definition, as well as problems in other fields like complexity theory, and logic, to name a few, can be formulated as optimization problems.

While the unrestricted and very general framework given above is powerful in its flexibility and applicability, problems without special structure, in general, cannot be solved efficiently, if at all. The solvability of the problem, and the complexity of that solution, in any sense of the word complexity, all depend on the special properties of the function f, and the set $\mathcal{X}$, as well as how explicitly each is given. For instance, writing,

$$\min_{x \in \mathcal{X}^*} f(x),$$

where we set

$$\mathcal{X}^* := \{x \, : \, f(x) \text{ is minimal}\},$$

is not particularly helpful.

2.1 Convex Optimization

One of the most used, and fruitful techniques, is to restrict ourselves to the class of continuous sets $\mathcal{X}$, say $\mathcal{X} \subset \mathbb{R}^n$, and functions f with some smoothness properties. This is still quite general, and many problems with these properties are still computationally intractable. However many other subcategories of problems have efficient, i.e. polynomial running time algorithms. Linear optimization, convex quadratic optimization, and semidefinite optimization, are just a few examples of more restrictive classes of functions f and sets $\mathcal{X}$, that are amenable to an efficient algorithmic solution.

There are two common elements to all such methods. The first, is viewing the function to be optimized as belonging to some smooth (to some degree) class of functions, and hence seeking to exploit the differential structure of that space. Ultimately, the backbone of these methods is a technique introduced in beginning calculus, of obtaining the zeros of a function's derivative, and testing for optimality. Such derivative tests obtain local information, and hence yield solutions that are locally optimal. As this local information is indeed just that, and has no global optimality guarantees, we often restrict further our class of functions f and sets $\mathcal{X}$, to functions and sets that are convex. In this case, locally and globally optimal points in fact coincide, and the local theory of the calculus is sufficient. Semidefinite and Linear Optimization, are both examples in the above restricted class, as they call for the optimization of a linear, and hence convex, function, subject to convex constraints (for more on Semidefinite Optimization see, e.g. , [6], and for Linear Optimization see [3]).

2.2 Non-Convex Optimization

While Semidefinite and Linear Optimization are very powerful tools that can successfully formulate many problems, nevertheless, these classes, and more generally the above methods, are inherently inadequate to deal with a number of other problems that arise very often, and naturally. Removing the convexity assumption on the function f to be optimized, even in the presence of regularity assumptions such as smoothness, the calculus methods that take advantage of this smoothness may prove severely inefficient, as they cannot provide anything more than local information. For instance, for a function with peaks of height 1 at, say, a collection of 1,000 points $\mathcal{S}$, yet with a peak of height 1,000 at some single point s^*, setting the derivative to zero and solving will provide a list of 2,001 points, none of which may be distinguished, *a priori*, and thus requiring a brute force approach. A concrete example is that of general (and hence possibly nonconvex) quadratic optimization over the hypercube $[0,1]^N$, a problem known to be $\mathcal{NP}$–complete (see, e.g. [2] for a nice summary of the state of the art in quadratic optimization).

Integer, and mixed continuous–integer problems are of particular interest in this context. First, as many quantities which arise naturally in practice are integer valued, integer optimization problems, that is, when we search for optimal solutions over sets of integers, arise very naturally, and are very important. In the context of the above discussion, integer problems present difficulties, as even though the function f may have nice properties, such as smoothness and convexity, the set over which we optimize is not even continuous, let alone convex, and hence there is no corresponding notion of calculus and differentiability tests to provide local information. While many heuristic algorithms have been suggested, for various special classes of problems with particular structure, little concrete progress has been made, and often times brute force methods are the only recourse. Mixed, or hybrid, continuous and discrete problems also prove quite difficult to solve, for the same reasons of lack of convexity, and discreteness. These too are very natural, and appear in a diverse array of application areas.

2.3 Convexity not the Key: An Overview of Problems and Solutions

Convexity is often considered the "magic word" when it comes to determining tractability of an optimization problem. However, it has long been known that convexity alone is not enough: what is needed is a convex set with the separation property, i.e. the property that any point not in the set can be (easily) separated from the set with a hyperplane.

On the other hand, some problems which are, at least in their original formulations, not convex, have tractable solutions. A particularly interesting example of this is the optimization of a (possibly indefinite) quadratic, subject to a fixed number k, of (possibly indefinite) quadratic constraints. This problem was just recently (Spring 2002) shown by Pasechnik et al. (paper not yet available) to be solvable in polynomial time.

In this section, we further stress this theme that convexity alone is far from being the full answer to the fundamental question: *What makes an optimization problem difficult or easy.*

In the sequel, we show that any polynomial optimization problem, convex or not, may in fact be written as a *convex* optimization problem, of dimension polynomial in the definition of the problem. The point is merely to note that convexity may not necessarily be the crucial property which defines computational complexity. We illustrate this point, as well as the approach we take, using some of the notation to be defined and discussed in more detail later in the paper. For now, we define:

$$\mathcal{P}_+^d(K) := \{\text{degree } d \text{ polynomials nonnegative on a set } K\}.$$

Note that this set is a closed convex cone, of dimension polynomial in the degree of the objective function f. Then, if f is a degree d polynomial, the optimization problem

$$\max_{x \in K} f(x),$$

becomes,

$$\max \gamma \ s.t. \ f - \gamma \in \mathcal{P}_+^d(K).$$

This is the optimization of a linear functional subject to a convex constraint. Unless we believe that $\mathcal{P} = \mathcal{NP}$, then we must believe that while convex, the constraint renders the problem intractable, in the general case.

The point is that showing that a polynomial lies in the set $\mathcal{P}_+^d(K)$ may be very difficult. In other words, the "membership test" for $\mathcal{P}_+^d(K)$ is difficult. However, as it turns out, showing that a polynomial belongs to some smaller subset of $\mathcal{P}_+^d(K)$ may be significantly easier. Therefore, in such a case, we obtain tractable relaxations, by performing the above optimization over a smaller set: a subset of $\mathcal{P}_+^d(K)$. The remainder of this paper discusses various algebraic results, which are used to demonstrate that some polynomial f belongs to a subset of $\mathcal{P}_+^d(K)$. The sequel develops algebraic results which give a nested sequence of subsets:

$$\Lambda_+^{d,0}(K) \subseteq \Lambda_+^{d,1}(K) \subseteq \cdots \Lambda_+^{d,N-1}(K) \subseteq \Lambda_+^{d,N}(K) \subseteq \mathcal{P}_+^d(K),$$

where for some sets K (e.g. K a finite set such as $\{0,1\}^d$) we are guaranteed equality with $\mathcal{P}_+^d(K)$ for some N, in the above chain of inclusions. Each subset provides (increasingly difficult) membership tests for each subset. Thus we have relaxations:

$\mathcal{R}_k$:

$$\max \gamma \ s.t. \ f - \gamma \in \Lambda_+^{d,k}(K).$$

The *computational key* of this method is the connection of Semidefinite Optimization, and Sums of Squares of polynomials. These results, along with the *theoretical key*, are developed in sections 4 and 5 below.

3 A Historical Introduction

A nice overview and outline of the theory may be found in Reznick's paper, "Some Concrete Aspects of Hilbert's 17^{th} Problem" ([26]). Also, the book "Squares" ([25]) may be useful.

We are now interested in the subset of nonnegative polynomials, and the (perhaps strictly smaller) subset of polynomials that can be expressed as a sum of squares of other polynomials, because of computational and complexity consequences. Indeed, these two sets of polynomials, and their distinction, is at the core of this paper. Around these two sets of polynomials, are centered the theoretical key, and the computational key, which together yields the methods outlined in thie paper. Historically, however, without high powered computers to bring questions of computability and complexity to the forefront, mathematicians were interested in these subsets of the polynomial ring for other reasons, primarily in the interest of understanding the structure of this mathematical object.

The history of the problem of expressing nonnegative polynomials as a sum of squares dates back to before the beginning of the 20^{th} century. We define some notation that we use throughout this paper. Let $\mathcal{P}_+$ denote the set of polynomials (the underlying polynomial ring is understood from the context) that are globally nonnegative, $\mathcal{P}_+(K)$ the set of polynomials nonnegative on a set $K \subseteq \mathbb{R}^n$, and Σ^2 the set of polynomials that can be expressed as the sum of squares of other polynomials. Both of these subsets of the ring of polynomials are in fact closed cones, as discussed further below. Starting with the observation that any univariate polynomial that is nonnegative on all of $\mathbb{R}$ may be written as the sum of squares of other polynomials (in fact, of two other polynomials), Hilbert asked whether this fact generalizes to higher dimensions, that is, to multivariate polynomials. In other words, having seen that $\Sigma^2 = \mathcal{P}_+$ in $\mathbb{R}[x]$, Hilbert asked whether this equality is always true, or whether the inclusion is ever strict. Hilbert gave the negative answer himself with a nonconstructive proof. At the Paris Congress in 1900, he then posed his famous 17^{th} question, now known as Hilbert's 17^{th} problem, of whether a nonnegative polynomial can be expressed as the sum of squares of *rational* functions.

One reason for the interest in nonnegative polynomials, was (and continues to be) the link to the classical moment problem. This is described more fully below in section 6. The problem here is to determine if a given sequence of numbers is indeed the moment sequence of some probability distribution, with a given support. Many mathematicians have worked on this problem, including Stieljes, Haviland, Hamburger, Riesz, and others. See Akhiezer ([1]) for a more complete history, and development of the classical moment problem.

The study of polynomials has traditionally been the realm of Algebraic Geometry, which studies so–called varieties, the zero sets of multivariate polynomials over algebraically closed fields, such as $\mathbb{C}^n$. In response to Hilbert's problem, a related field developed, which studies similar questions as algebraic geometry, but over fields with the property that -1 cannot be written as a sum of squares. Such fields are called *formally real*, and certainly, $\mathbb{R}$ is an example of such. In 1927, Artin developed, and then used what is now known as the Artin–Schreier theory of real closed fields, to give an affirmative answer to Hilbert's 17^{th} problem: Any polynomial, nonnegative on all of $\mathbb{R}^n$, may be expressed as the sum of squares of quotients of polynomials.

One of the interesting corrolaries to this positive answer, is that if a polynomial g is globally nonnegative, then there exists some polynomial $f \in \Sigma^2$ for which fg may be written as a sum of squares. We give an example of this below. B. Reznick has shown that if g is strictly positive, globally, then f may be chosen to be the polynomial

$$f(x) = \left(\sum x_i^2\right)^d,$$

for some d sufficiently large.

While Hilbert proved that the inclusion $\Sigma \subseteq \mathcal{P}_+$ is in general strict, the proof was not constructive. Motzkin is the first credited to have written down a concrete example, and this did not happen until the 1960's. Motzkin's famous example is the homogeneous polynomial (or form) in three variables:

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

While this is nonnegative for any $(x,y,z) \in \mathbb{R}^3$, it cannot be written as a sum of squares of other polynomials. It is possible, as Artin's theory promises, to multiply $M(x,y,z)$ by a polynomial that is itself a sum of squares, and have the resulting product be expressible as a sum of squares. Indeed, we have,

$$\begin{aligned}(x^2+y^2+z^2)M(x,y,z) &= (x^2yz-yz^3)^2+(xy^2z-xz^3)^2+(x^2y^2-z^4)^2\\ &\quad +\frac{1}{4}(xy^3-x^3y)^2+\frac{3}{4}(xy^3+x^3y-2xyz^2)^2,\end{aligned}$$

thus illustrating Artin's result. In addition, it illustrates, since $(x^2+y^2+z^2)$ is, evidently, globally nonnegative, that so is $M(x,y,z)$ because the product is.

Consider the dehomogenizations of $M(x,y,z)$ in y and z respectively:

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

Neither of these can be written as a sum of squares. While

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

the polynomial $g(x,y)-\gamma$ cannot be written as a sum of squares for any value of $\gamma\in\mathbb{R}$.

More recently, the connection between the cone $\mathcal{P}_+$ and polynomial optimization has begun to be exploited, due to the computational tractability of the cone Σ^2, as described below. This paper contains a description of the associated methods.

4 Geometrical Cones of Polynomials

We consider now the optimization perspective, and in this section and the next, we develop the connection to the cones $\mathcal{P}_+$ and Σ^2. Suppose we aim to minimize a polynomial f over some set $\mathcal{X}$. Note that we have the equivalence,

$$\min_{x\in\mathcal{X}} f(x) \Longleftrightarrow \max\gamma\in\mathbb{R} \text{ s.t. } f(x)-\gamma\geq 0,\ \forall x\in\mathcal{X}.$$

Then we see that minimizing a polynomial is essentially equivalent to determining when it is nonnegative over some specified set.

4.1 The Cone $\mathcal{P}_+(K)$

Let us define the above more precisely. For $K\subseteq\mathbb{R}^n$ let $\mathcal{P}_+(K)$ denote the set of polynomials (we assume in all this that we are working in $\mathbb{R}[x_1,\ldots,x_n]$, where the value of n is unambiguous) that assume nonnegative values on all of K, and $\mathcal{P}_+^d(K)$ those polynomials of degree at most d:

$$\begin{aligned}\mathcal{P}_+^d(K) &:= \{g(x)=g(x_1,\ldots,x_n)\in\mathbb{R}[x_1,\ldots,x_n] : \deg(g)\leq d,\ \ g(x)\geq 0, \forall x\in K\},\\ \mathcal{P}_+(K) &:= \{g(x)=g(x_1,\ldots,x_n)\in\mathbb{R}[x_1,\ldots,x_n] : g(x)\geq 0, \forall x\in K\}\\ &= \bigcup_{d=1}^{\infty}\mathcal{P}_+^d(K).\end{aligned}$$

Then our optimization problem is reduced to computing the membership test:

$$\min_{x\in\mathcal{X}} f(x) \Longleftrightarrow \max\gamma\in\mathbb{R} \text{ s.t. } f(x)-\gamma\in\mathcal{P}_+^d(\mathcal{X}).$$

Note that for any set K, $\mathcal{P}_+^d(K)$ is a convex cone in the linear space of polynomials, spanned by the monomials of degree at most d. Then, as mentioned in section 2.3 above, any polynomial optimization problem, may be written as the optimization of a linear function subject to a convex constraint set. Indeed, if $g_1(x), g_2(x) \in \mathcal{P}_+^d(K)$, then for any $\lambda_1, \lambda_2 \geq 0$, we must also have $g(x) = \lambda_1 g_1(x) + \lambda_2 g_2(x) \in \mathcal{P}_+^d(K)$. Other examples of geometric cones are, for instance, the positive orthant, $\mathbb{R}_+^n$. Note, however, that while both of these are examples of cones, the former cone, $\mathcal{P}_+^d(K)$ is specified by the evaluation of polynomials at all the points of K, where as the latter cone of the positive orthant is described directly by its facets, namely, the usual cartesian axes. Indeed, given some vector $v \in \mathbb{R}^n$, verifying membership in $\mathbb{R}_+^n$ involves n evaluations, that is, just a componentwise analysis. On the other hand, consider some polynomial g, of degree $n-1$ in only one variable. This polynomial, like the vector v, lies in an n dimensional Euclidean space, however, verifying membership in the cone $\mathcal{P}_+^{n-1}(K)$ is *a priori* difficult to do without some exhaustive evaluation scheme.

The point of this discussion is that while geometrically, the positive polynomials and the positive orthant are both cones, and hence share similar geometric properties, computationally, the former seems to be in general intractable, while the latter is quite computationally tractable. We need then to see what we can say about membership, in terms of tests, and also verifiable certificates, in the intractable cone $\mathcal{P}_+(K)$. Before we address this problem, we develop the theory of another cone of polynomials, introduced in section 3 above, related to the cone of positive polynomials, but, it turns out, such that the membership test is efficiently computable.

4.2 The Cone Σ^2

As described in section 3 above, let $\Sigma^2 \subset \mathbb{R}[x_1, \ldots, x_n]$ denote the set of polynomials which can be written as a sum of squares of other polynomials, that is,

$$\Sigma^2 := \{g(x) \in \mathbb{R}[x_1, \ldots, x_n] \,:\, \exists h_1(x), \ldots, h_m(x) \in \mathbb{R}[x_1, \ldots, x_n], \text{some } m, \text{ s.t. } g(x) = \textstyle\sum_{i=1}^m h_i(x)^2.\}$$

While Hilbert was interested in determining the precise nature of the relatinoship of the cones Σ^2 and $\mathcal{P}_+$ in order to further understand the structure of the ring $\mathbb{R}[x_1, \ldots, x_n]$, we have an additional motivation: As claims the following proposition, and is subsequently proved below, the membership test for Σ^2 can be performed efficiently. We show in section 5 that this fact is the key computational result, which together with the algebraic results, underlie the methods described in this paper.

Proposition 1 *The membership test "$g(x) \in \Sigma^2$" can be performed in time polynomial in the size of the polynomial $g(x)$.*

This proof of this proposition is the link of sum of squares to semidefinite optimization. Recall first the following facts about semidefinite matrices.

Lemma 1 *A symmetric matrix A is positive semidefinite if and only if there exists a matrix B for which we have:*

$$A = BB^T.$$

PROOF. The proof is an immediate result of, say, the Jordan decomposition. □

Now consider any polynomial $f(x) \in \mathbb{R}[x_1, \ldots, x_n]$, of degree $2d$. Let X be the vector of all monomials in $x_1, \ldots, x_n$ of degree d and below. For example, if $n = 2, d = 2$ then,

$$X = \{1, x_1, x_2, x_1^2, x_1x_2, x_2^2\}.$$

There are $\binom{n+d}{d}$ such monomials. Any polynomial, positive or not, a sum of squares or not, may be written in matrix form as

$$f(x) = X^T Q X,$$

for some symmetric matrix Q. Now, if the matrix Q is positive semidefinite, then $Q = BB^T$ for some B, and we have,

$$\begin{array}{rcl} f(x) = X^T Q X & = & X^T B B^T X \\ & = & \langle (X^T B), (X^T B) \rangle, \end{array}$$

and hence $f(x)$ is a sum of squares. Conversely, suppose that $f(x)$ does indeed have a sum of squares decomposition:

$$f(x) = h_1(x)^2 + \cdots + h_r(x)^2,$$

for some $h_1, \ldots, h_r$. By abuse of notation, we let h_i also denote the vector in $\mathbb{R}^{\kappa(n,d)}$ of the coefficients of the polynomial $h_i(x)$, where $\kappa(n,d) = \binom{n+d}{d}$ is the number of monomials in n variables of degree d or less. Then, note that we have,

$$\begin{aligned} h_i(x)^2 &= \langle h_i, X \rangle^2 \\ &= X^T(h_i \cdot h_i^T)X \\ &= X^T Q_i X, \end{aligned}$$

where $Q_i = h_i \cdot h_i^T$, and thus is positive semidefinite. But then we have,

$$\begin{aligned} f(x) &= h_1(x)^2 + \cdots + h_r(x)^2 \\ &= X^T Q_1 X + \cdots + X^T Q_r X \\ &= X^T(Q_1 + \cdots + Q_r)X = X^T Q X, \end{aligned}$$

where Q is positive semidefinite because the Q_i are. We have proved,

Lemma 2 *The degree* $2d$ *polynomial* $f(x)$ *has a sum of squares decomposition if and only if there exists a positive semidefinite matrix* Q *for which*

$$f(x) = X^T Q X.$$

The choice of matrix Q which we use to express f is not unique. Essentially, we have mapped ourselves to a polynomial space with more variables, enabling us to write any polynomial as a quadratic form. This mapping is known as the Veronese mapping in Algebraic Geometry, and is given by defining a new variable y_i for each monomial in the vector X. Then, for the example above, we have

$$(y_1, y_2, y_3, y_4, y_5, y_6) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2).$$

Note, however, that our new variables are not independent. For instance, we have $y_1y_4 = y_2y_2$. This means then, that

$$X^T(\lambda C)X = \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & \frac{\lambda}{2} & 0 & 0 \\ 0 & -\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\lambda}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix} = 0,$$

for all values of $\lambda \in \mathbb{R}$. Therefore, given some polynomial $f(x)$, there is an affine set (a subspace translated away from the origin) of matrices Q, for which $f(x) = X^T Q X$. Searching for such a matrix is a semidefinite optimization feasibility problem, and as such can be solved efficiently, in time polynomial (cubic) in the size of the resulting matrix. Hence the membership test can indeed be performed in time polynomial in the size of the polynomial $f(x)$ in question. This concludes the proof of the proposition.

In the next section we show how the computational tractability of the cone of sum of squares polynomials, can be used to obtain information about the cone of positive polynomials.

5 Using Σ^2 to Understand $\mathcal{P}_+(K)$

In the last section we proved, given that semidefinite optimization problems may be solved efficiently, that the membership test in Σ^2 may be performed efficiently. In this section we demonsrate how this may be used in the interest of understanding $\mathcal{P}_+$, and $\mathcal{P}_+(K)$.

5.1 A Simple Optimization Problem

We start with an easy, although illustrative example. It is a fact that for the case of univariate polynomials, the two cones of nonnegative polynomials and sum of squares polynomials coincide. That is, any polynomial that is nonnegative for all $x \in \mathbb{R}$, has a sum of squares decomposition (in fact it can be written as the sum of two squares). The converse is obvious and is always true. Then in this case we have a computable membership test for $\mathcal{P}_+$. Note that the coefficients of the polynomial $f(x)$ which we express as $f(x) = X^T QX$, enter affinely into the matrix Q. Therefore if we write

$$f(x) - \gamma = X^T Q_\gamma X,$$

the matrix Q_γ depends affinely on γ. Therefore, as easily as we can check if $f(x)$ has a sum of squares decomposition, we can check if $(f(x) - \gamma)$ has a sum of squares decomposition for some value of γ. By letting γ be our objective function which we seek to maximize, we have exactly:

$$\begin{array}{rl} \max: & \gamma \\ \text{s.t.}: & f(x) - \gamma = X^T QX, \\ & Q \succeq 0. \end{array}$$

This will provide the largest value of γ for which $(f(x) - \gamma)$ belongs to Σ^2. In the univariate case where $\Sigma^2 = \mathcal{P}_+$, we will have solved for the largest value of γ for which $(f(x) - \gamma) \in \mathcal{P}_+$, i.e. the largest value of γ for which $(f(x) - \gamma)$ is nonnegative, and hence we have the true minimum value of $f(x)$ over all of $\mathbb{R}$. Note that there has been no mention of convexity of the polynomial $f(x)$ which we minimize. Indeed, we are now not even using the fact that polynomials are continuous functions. Instead, we exploit their purely algebraic properties, which are, in this sense, blind to issues of convexity, and local versus global behavior. Essentially, what we have done is to translate the pointwise property, "$\geq 0, \forall x \in \mathbb{R}$" to an algebraic property, "$\in \Sigma^2$."

For multivariate polynomials, as we have seen with Motzkin's examples, we have a strict containement, $\Sigma^2 \subsetneq \mathcal{P}_+$. Therefore the largest γ for which $(f(x) - \gamma) \in \Sigma^2$, will only be a lower bound, in general, to the true minimum of the polynomial $f(x)$. In some cases, as in the dehomogenization of the Motzkin polynomial,

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

the gap can be ∞.

5.2 Polynomial Certificates

The simple example above gives an easy instance of how one might use the tractable cone of sums of squares polynomials, in order to show that some polynomial is in fact in the cone of globally nonnegative polynomials. More generally, we may think of the sum of squares decomposition as a *certificate* verifiable in polynomial time, that the polynomial in question is nonnegative. Since in general, $\Sigma^2 \subsetneq \mathcal{P}_+$, certificates of nonnegativity do not always take the simple form of a sum of squares decomposition. There are other possible certificates, however. For example, suppose we have, for some polynomial $f(x)$,

$$\begin{aligned} f(x) &= h_1(x) + g_1(x)u(x) \\ &= h_2(x) - g_2(x)u(x), \end{aligned}$$

where $h_1, h_2, g_1, g_2 \in \Sigma^2$, and $u(x)$ is some polynomial. It is not difficult to see that these five polynomials $\{h_1, h_2, g_1, g_2, u\}$, are a certificate that demonstrates that $f(x) \in \mathcal{P}_+$.

In section 3 we gave a certificate for the nonnegativity of the Motzkin polynomial, by demonstrating that

$$(x^2 + y^2 + z^2)M(x, y, z) \in \Sigma^2.$$

It is a fact that for any nonnegative polynomial f, there exists a polynomial $g \in \Sigma^2$, such that

$$g(x) \cdot f(x) \in \Sigma^2.$$

This result says that there is always a certificate of nonnegativity using sums of squares. The immediate question then is what is the complexity of determining this polynomial g, and the sum of squares decomposition. There is a strong relation here to various proof systems in the theory of logic and complexity (see, e.g. [23], or [10]). Specifically, what rules do we allow ourselves for derivation, and what is the complexity of this system of rules?

In the next section we describe a very powerful proof system called the Positivstellensatz calculus, which provides polynomial certificates not only for global nonnegativity, but more generally for local nonnegativity over more general sets as well, which is really what we are after for non-convex optimization.

5.3 The Positivstellensatz

In Algebraic Geometry, it is of interest when some set of polynomial equations has no set of common solutions, i.e. there are no points for which all the polynomial equations are satisfied. In other words, if we have a family of equations:

$$\{f_i(x) = 0, i = 1, \ldots, m\},$$

then Hilbert's Nullstellensatz tells us that the following are equivalent:

1. $K = \{x \in \mathbb{C}^n \mid f_i(x) = 0, \forall i\} = \emptyset$,
2. $\exists g_1, \ldots, g_m, \text{s.t. } f_1 g_1 + \cdots + f_m g_m \equiv 0\}$.

Given a set of polynomials $\{f_i\}$ as above, the equation

$$f_1 g_1 + \cdots + f_m g_m \equiv 0,$$

is a polynomial certificate of the emptiness of the set K defined as above. In the section above, we saw that polynomial certificates may take various forms. Hilbert's theorem, guarantees us that if a set K is empty, then there is a polynomial certificate of that fact, that assumes the above form. In 1974, G. Stengle (see [30]) proved an analog of the Nullstellensatz for Real Algebraic Geometry, called the Positivstellensatz. Just as Hilbert's Nullstellensatz gives a form that polynomial certificates of unsolvability take, Stengle's Positivstellensatz provides the form that polynomial certificates take, that guarantee a function's nonnegativity over a set. In other words, if a function f is nonnegative over a set K, Stengle's theorem guarantees that there exists a polynomial certificate of a specific form (see below) that demonstrates that nonnegativity.

Before we give the statement of the theorem, we introduce three algebraic subsets of a ring.

Definition 1 (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 2 *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 3 (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 4 (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. Similar ideas have been considered in the field of combinatorial optimization, and specifically zero–one problems, where various methods have been proposed to obtain valid inequalities (see, e.g. Lovasz–Schrijver [17], and Sherali–Adams [28]).

Restating the result of the Positivstellensatz, we see, as claimed at the beginning of this section, that the theorem gives a polynomial certificate that a polynomial f_0, is nonnegative over a set K of the form above.

Certificate of Nonnegativity: A polynomial f_0 is nonnegative over a set K (and hence belongs to $\mathcal{P}_+(K)$) if and only if there exist polynomials

1. $F \in \mathcal{A}(-f_0, f_1, \ldots, f_{n_1})$,
2. $G \in \mathcal{M}(-f_0, g_1, \ldots, g_{n_1})$,
3. $H \in I(h_1, \ldots, h_{n_3})$,

such that the identity

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

holds.

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

Example: Quadratic Optimization over Two Ellipses

Suppose we want to find the mininimum of an arbitrary (not necessarily convex) quadratic function, over the intersection of two ellipses. This problem can be written as:

$$\begin{array}{rl} \text{minimize}: & x^T Q x + c^T x \\ \text{s.t.}: & x^T Q_1 x \leq b_1 \\ & x^T Q_2 x \leq b_2 \\ & Q_1, Q_2 \succeq 0. \end{array}$$

Using the Positivstellensatz, we have, equivalently,

$$\begin{array}{rl} \text{maximize}: & \gamma \\ \text{s.t.}: & K = \{\gamma - x^T Q x + c^T x \geq 0, x^T Q_i x \geq 0, i = 1, 2, \gamma - x^T Q x + c^T x \neq 0\} = \emptyset. \end{array}$$

Writing $f_\gamma = \gamma - x^T Q x + c^T x$, and $f_i = x^T Q_i x$, $i = 1, 2$, we have the equivalent statement in terms of the certificates:

$$\begin{array}{rl} \text{maximize}: & \gamma \\ \text{s.t.}: & s_0 + s_1 f_\gamma + s_2 f_1 + s_3 f_2 + s_{12} f_\gamma f_1 + s_{13} f_\gamma f_2 + s_{23} f_1 f_2 + s_{123} f_\gamma f_1 f_2 + f_\gamma^{2d} \equiv 0, \\ & s_i, s_{ij}, s_{ijk} \in \Sigma^2, \forall i, j, k, \\ & d \in \mathbb{N} \cup \{0\}. \end{array}$$

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. We can rewrite the last step in the sequence of equivalences, as follows:

$$\begin{array}{rl} \text{maximize :} & \gamma \\ \text{s.t. :} & -(s_1 f_\gamma + s_2 f_1 + s_3 f_2 + s_{12} f_\gamma f_1 + s_{13} f_\gamma f_2 + s_{23} f_1 f_2 + s_{123} f_\gamma f_1 f_2 + f_\gamma^{2d}) \in \Sigma^2, \\ & s_i, s_{ij}, s_{ijk} \in \Sigma^2, \forall i, j, k, \\ & d \in \mathbb{N} \cup \{0\}. \end{array}$$

In other words, the problem is to find the largest γ, such that there exist elements in the cone Σ^2, which in turn are such that the expression above is itself in that cone. The key point is that via the reduction to semidefinite optimization described above, performing the membership test of some polynomial in Σ^2, and moreover finding the actual sum of squares decomposition, is "easy," that is, easy with respect to the size of the polynomial. There are two obstacles that keep 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 that $\mathcal{P} = \mathcal{NP}$). First and foremost, 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. In fact, little is known in terms of a tight upper bound. In [20], Parrilo and Sturmfels report an unpublished result of Lombardi and Roy, who have announced a triple exponential upper bound on the sufficient degree. As is often true, however, the worst case scenario does not seem to be an appropriate measure for the practical complexity of the problem. That is to say, the problems typically encountered do not seem to require triple exponential degree polynomial certificates. In addition to this there is another technical problem. Recall in the reduction to the semidefinite problem, that we need the variables to appear affinely in the matrix. This is not true in the above certificate, as if we have $d \geq 1$, then γ no longer appears affinely. Note, however, that the product of f_γ with the s_I does not concern us, as it is still what is known as quasi-convex, and thus can be solved efficiently.

5.4 Convex Relaxations

By limiting the degree of the polynomial certificates obtained via the positivstellensatz, we obtain a sequence of relaxations, each of which is a convex optimization problem, and the last of which solves our original optimization problem exactly. We noted above, that for a semi–algebraic set K defined as above, a degree d polynomial f is nonnegative on K, i.e. $f \in \mathcal{P}_+^d(K)$ if and only if there exist polynomials

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

such that the identity

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

holds. As we have stressed, this is the certificate of nonnegativity on K. Therefore, the above statement says,

$$\begin{array}{rcl} \mathcal{P}_+^d(K) & = & \{\text{degree } d \text{ polynomials } f \text{ such that } f(x) \geq 0 \text{ for all } x \in K\} \\ & = & \{\text{degree } d \text{ polynomials that have a certificate of nonnegativity as above.}\} \end{array}$$

We next define the subset of polynomials, that have a *polynomial certificate of bounded degree*. We define,

$$\Lambda_+^{d,N}(K) = \{\text{degree } d \text{ polynomials that have a certificate of nonnegativity of degree at most } N\}.$$

Then, we have the obvious inclusion,

$$\Lambda_+^{d,N}(K) \subseteq \mathcal{P}_+^d(K).$$

For certain sets K, e.g. finite sets, then for N sufficiently large, we have equality.

We have discussed that the general optimization problem, for f an objective function of degree d, may be rewritten as,

$$\begin{aligned} \max: \quad & \gamma \\ \text{s.t.}: \quad & f - \gamma \in \mathcal{P}_+^d(K). \end{aligned}$$

Then we obtain a sequence of relaxations,

$\mathcal{R}_k$:

$$\begin{aligned} \max: \quad & \gamma \\ \text{s.t.}: \quad & f - \gamma_k \in \Lambda_+^{d,k}(K). \end{aligned}$$

Each relaxation $\mathcal{R}_k$ provides a lower bound γ_k on the true minimum γ^*. By solving successive relaxations, we obtain a sequence of values, $\{\gamma_k\}$ such that $\gamma_k \leq \gamma_{k+1}$, and $\gamma_k \to \gamma^*$.

Algebraic Complexity:

For special classes of sets K we are guaranteed that for some (perhaps extremely large) integer M, $\Lambda_+^{d,M}(K) = \mathcal{P}_+^d(K)$, and hence the relaxation $\mathcal{R}_M$ is exact. The smallest such integer M, is a function of the set K, and the degree d. For instance, we know that if $d = 2$ and the set K is the nonnegativity set of an arbitrary quadratic in $\mathbb{R}^n$ (or, in particular, an ellipse in $\mathbb{R}^n$) then $M = 2$. We can consider this integer M to be the algebraic complexity of the problem defined by d and K. Note that a particular function $f \in \mathcal{P}_+^d(K)$, may belong to $\Lambda_+^{d,k}(K)$ for $k < M$. This leads us to define the algebraic complexity associated to a degree d function f, or a family of degree d functions $\mathcal{F}$, as the smallest integer $M_{\mathcal{F}}$ such that for every $f \in \mathcal{F}$, $f \in \Lambda_+^{d,M_{\mathcal{F}}}(K)$. As discussed further in section 5.6 below, calculating such numbers $M_{\mathcal{F}}$ for interesting classes of K, d, and $\mathcal{F}$, is of primary importance (and, unfortunately, difficulty), as they provide concrete complexity results.

We now give a more interesting example of a polynomial certificate.

Example: Suppose we want to find the smallest distance from a given point $(x_0, y_0) = (1, 1)$ to an algebraic curve,

$$C(x, y) := x^3 - 8x - 2y = 0.$$

This is the minimization of a quadratic polynomial subject to the cubic constraints of the curve:

$$\begin{aligned} \text{minimize}: \quad & f(x, y) = (x-1)^2 + (y-1)^2 \\ \text{s.t.}: \quad & C(x, y) = 0. \end{aligned}$$

Using the Positivstellensatz, we want the largest γ for which

$$C(x, y) = 0 \implies f(x, y) - \gamma \geq 0.$$

This will happen exactly when the set

$$K := \{(x, y) \in \mathbb{R}^2 \, : \, \gamma - f(x, y) \geq 0, f(x, y) - \gamma \neq 0, C(x, y) = 0\},$$

is empty. By the Positivstellensatz, this set is empty if and only if there exists a polynomial certificate of the form

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

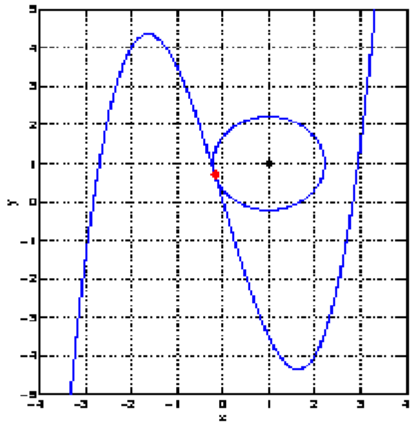

Figure 1: Example and Figure from Parrilo, [19]

where $F \in \mathcal{A}(\gamma - f(x,y))$, $G \in \mathcal{M}(f(x,y)-\gamma)$, and $H \in I(C(x,y))$. Equivalently, the set K is empty if and only if there exist polynomials $s(x,y) \in \Sigma^2$, and $r(x,y) \in \mathbb{R}[x,y]$, and nonnegative integer d, such that

$$\big(s(x,y)(f(x,y)-\gamma) - (\gamma - f(x,y))^{2d} + r(x,y)C(x,y)\big) \in \Sigma^2.$$

Restricting ourselves to linear auxiliary polynomials, we can compute a lower bound on γ and thus the minimization, by computing the maximum γ for which

$$(x-1)^2 + (y-1)^2 - \gamma^2 + (a+bx)(x^3 - 8x - 2y) \in \Sigma^2.$$

The optimal solution yields $\gamma \approx 1.4722165$, and this is indeed the correct value as it is achieved at

$$x \approx -0.176299246, \quad y \approx 0.70257168.$$

5.5 Schmüdgen and Putinar's Positivstellensatz

Under certain restrictions on the semialgebraic set K, we can formulate a stronger, in some sense, form of the Positivstellensatz. We now state these stronger versions of the Positivstellensatz, and then discuss their applicability, and their relative advantages and disadvantages when compared to the Positivstellensatz as stated above. Consider again the Positivstellensatz, as providing a polynomial certificate of nonnegativity of a polynomial f, over a semialgebraic set of the form

$$K = \{x \in \mathbb{R}^n \;:\; f_1(x) \geq 0, \dots, f_m(x) \geq 0\}.$$

Then the Positivstellensatz guarantees the existence of polynomials $s(x), s_I(x) \in \Sigma^2$, for $I \subset \{0, 1 \dots, m\}$ such that, setting $f_0 = -f$,

$$s(x) + \sum_I s_I(x) \prod_{i \in I} f_i(x) + f_0^{2k} \equiv 0,$$

for some $k \in \mathbb{N}$. Now, suppose we take,

$$s(x) \equiv s_I(x) \equiv 0, \;\; \forall I \text{ such that } 0 \notin I,$$

and $k = 1$. Then the above factors as

$$f(x)(-f(x) + s_0(x) + \sum_I \prod_{i \in I} f_i(x)) \equiv 0,$$

which is satisfied only if

$$f(x) = s_0(x) + \sum_I \prod_{i \in I} f_i(x).$$

If, furthermore, we were to set $s_I = 0$ for every $I \neq \{0, i\}$, then we would have

$$f(x)(-f(x) + s_0(x) + \sum_i f_i(x)) \equiv 0,$$

which can happen only if

$$f(x) = s_0(x) + \sum_i f_i(x).$$

Since we obtained both these expressions from forcing restrictive choices on the form of the Positivstellensatz certificate, it is clear that both are sufficient conditions to ensure nonnegativity of the polynomial f, but we have no *a priori* guarantee that there exists a certificate of that restricted form. The theorems of Schmüdgen and Putinar show that if K satisfies some technical conditions, then in fact we are guaranteed that a certificate of the above forms (respectively) exists.

The first is a theorem proved by Schmüdgen in 1991 (see [27]), under the assumption that the semialgebraic set K is defined only by polynomial inequalities, and, moreover, it is compact. Many problems satisfy this additional compactness constraint, for instance any nondegenerate ellipse or hypercube constrained problems, and all zero–one combinatorial optimization problems, to name a few. Schmüdgen's theorem says the following:

Theorem 2 (Schmüdgen) *Suppose we have a subset $K \subseteq \mathbb{R}^n$ described as*

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

and suppose further that K is compact. Then, if g is positive on all of K, g must be in the algebraic cone generated by the $\{f_i\}$. More explicitly, this means that there exist polynomials

$$\{s\}, \{s_i\}, \{s_{i_1 i_2}\}, \ldots, \{s_{1,2,\ldots,m}\},$$

all sums of squares, such that we have

$$\begin{aligned} g(x) &\equiv s(x) + \sum_i s_i(x) f_i(x) + \sum_{i_1, i_2} s_{i_1 i_2}(x) \cdot f_{i_1}(x) \cdot f_{i_2}(x) + \cdots + s_{1,2,\ldots,m}(x) \cdot f_1(x) \cdots f_m(x) \\ &= s(x) + \sum_{I \subseteq \{1,\ldots,m\}} s_I(x) \prod_{i \in I} f_i(x). \end{aligned}$$

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 exists a polynomial $F \in \mathcal{A}(f_1, \ldots, f_m)$ such that*

   $$g - F \equiv 0.$$

Note that, as with the Positivstellensatz above, Schmüdgen's theorem has an exponential number of terms in the polynomial certificate, where the number of terms may be exponential in the number of constraints, m. However, if we consider a sequence of relaxations, as in section 5.4 above, then the exponentially many constraints do not appear until the last stages of the relaxation.

In 1993, Putinar proved (see [24]) that under an additional technical condition, the linear terms in Schmüdgen's representation suffice. Thus, Putinar's strengthened version adds one more condition to the domain set K, but in exchange, requires only $m+1$ terms in the expression of the positive polynomial, rather than the 2^m terms required in Schmüdgen's original statement above.

Theorem 4 (Putinar) *Suppose we are given a set,*

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

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 $s_0, s_1, \ldots, s_m$ sums of squares, such that

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

Putinar's theorem as well, may be written as an equivalence of a set K being empty, and a polynomial certificate of this emptyness, of the specified form. Next we merely note that for a large host of applications, the additional constraint required for Putinar's theorem is easily satisfied by the corresponding sets K. The following four cases fall into this category.

1. Suppose some f_i in the definition of K satisfies, on its own, the condition $\{f_i(x) \geq 0\}$ compact. Then Putinar's theorem applies. This includes any instance where we are taking intersections with ellipses, or circles, among others.

2. Perhaps very importantly, any $0-1$ program falls easily into this framework. The integer constraint is given by polynomial relations $x_i^2 - x_i = 0$. Consider now the polynomial $u(x) = \sum_i (x_i - x_i^2)$. This is of the correct form, and indeed satisfies $\{u(x) \geq 0\}$ compact.

3. If K is compact, and is defined only by linear functions, then we can directly apply Putinar's theorem. Note that this includes all polytopes.

4. If we know that the compact set K lies inside some ball of radius M, we can simply add the interior of the ball: $\sum_i x_i^2 \leq M^2$ as a redundant constraint, thus not changing K, but automatically satisfying Putinar's theorem, without appreciably changing the size of the definition of the problem (especially if we already have a large number of functions defining K).

The stronger versions of Schmüdgen and Putinar given above, have some considerable advantages, both aesthetic and practical. We focus on Putinar's theorem, and in what follows we always assume that the set K satisfies the hypotheses of the theorem. Optimization using Putinar's theorem is formulated in a parallel manner as for the Positivstellensatz, namely, we translate the pointwise property "$g(x) > 0, \forall x \in K$" to the algebraic property,

$$\text{“}\exists\, s_0, s_1, \ldots, s_m \in \Sigma^2 \quad \text{such that} \quad g(x) \equiv s_0 + \sum_{i=1}^{m} s_i(x) f_i(x)\text{”},$$

or, equivalently,

$$\text{“}\exists\, s_1, \ldots, s_m \in \Sigma^2 \quad \text{such that} \quad \left(g(x) - \sum_{i=1}^{m} s_i(x) f_i(x) \right) \in \Sigma^2\text{”}.$$

Optimization problems then become:

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

equivalently written using Putinar, as

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

Similarly to the Positivstellensatz, under the assumptions on K, Putinar's theorem guarantees the existence of a polynomial certificate of positivity, however it makes no statement as to the degree of the certificate. In fact, from the point of view of proof systems, while Putinar's theorem is appealing because of its simpler and more explicit expression of the certificate, we might also expect it to require higher degree certificates, as the form of the representation is more restrictive, hence corresponding to more strict derivation rules.

Example: The following example is due to Stengle ([31]). Consider the simple univariate problem,

$$\begin{aligned} \text{minimize :} \quad & f(x) = (1 - x^2) \\ \text{s.t. :} \quad & x \in K := \{x \in R \ : \ (1 - x^2)^3 \geq 0\}. \end{aligned}$$

This problem is easy, as $K = [-1, 1]$, and hence $f(x)$ has its minima at $x = 1$ and $x = -1$, with value 0. Note that $f(x)$ is not strictly positive on all of K, and hence we are not guaranteed that it has an expression in terms of sums of squares as dictated by Putinar's theorem. In fact, consider the equation,

$$1 - x^2 = s_0(x) + s_1(x) \cdot (1 - x^2)^3, \ \ s_0, s_1 \in \Sigma^2.$$

Since the left hand side has a zero at $x = 1$, the right hand side must also vanish at 1. There are two terms on the right hand side. The second term, $s_1(x) \cdot (1 - x^2)^3$ vanishes at least to 3^{rd} order at $x = 1$. Since the second term vanishes at 1, the first term must also vanish there. Since the first term is a sum of squares, it must vanish at least to second order at 1. Therefore the entire right hand side vanishes at least to second order at $x = 1$, which is absurd, as the left hand side only vanishes to first order. Therefore we conclude that there do not exist sums of squares $s_0(x), s_1(x)$ such that the above expression holds. This illustrates that the strict positivity in Putinar's theorem is in fact necessary. Now consider the family of functions $\{f_\varepsilon(x)\}$, for $f_\varepsilon(x) = f(x) + \varepsilon$. For all $\varepsilon > 0$, $f_\varepsilon(x) > 0$ for all $x \in K$, and therefore by Putinar's theorem,

$$f_\varepsilon(x) = s_0^\varepsilon(x) + s_1^\varepsilon(x) \cdot (1 - x^2)^3,$$

for some $s_0^\varepsilon, s_1^\varepsilon \in \Sigma^2$.

Claim 1 *The value of* $\max\{\mathit{degree}(s_0^\varepsilon), \mathit{degree}(s_1^\varepsilon)\}$ *goes to* ∞ *as* $\varepsilon \to 0$.

PROOF. Suppose not. Then let M be the degree bound. Now, by virtue of the fact that s_0^ε and s_1^ε are both sums of squares, and therefore nonnegative, and by their definition, the sequence of polynomials $s_0^\varepsilon, s_1^\varepsilon$ must lie inside a finite dimensional (because of their bounded degree) ball of radius, say, 1, using the supremum norm. The supremum ball in finite dimensional vector space is compact, and therefore there must exist a convergent subsequence. Since convergence in any norm in finite dimensions, and pointwise convergence are equivalent, the limit as $\varepsilon \to 0$, yields sums of squares polynomials s_0, s_1 for which $f(x) = s_0(x) + s_1(x) \cdot (1 - x^2)^3$, which we have already seen to be impossible. The contradiction concludes the proof. ☐

Therefore the degree of the polynomials $s_0^\varepsilon, s_1^\varepsilon$ in the polynomial certificate of positivity must become unbounded, as $\varepsilon \to 0$. Next we consider the same optimization problem using the Positivstellensatz. Minimizing the polynomial $f(x) = 1 - x^2$ over K as given above, is equivalent, using the Positivstellensatz, to the problem:

$$\begin{array}{rl} \text{maximize :} & \gamma \\ \text{s.t. :} & \exists\ s_0, s_1, s_2, s_3 \in \Sigma^2,\ k \in \mathbb{N},\quad \text{such that} \\ & s_0(x) + s_1(x)(\gamma - f(x)) + s_2(x)(1-x^2)^3 + s_3(x)(\gamma - f(x))(1-x^2)^3 \\ & \quad +(\gamma - f(x))^{2k} \equiv 0. \end{array}$$

The values

$$\gamma = 0, s_0(x) = (1-x^2)^4, s_1(x) = s_2(x) = 0, s_3(x) = 1, k = 2,$$

provide a certificate, and thus a lower bound on γ. This is indeed the correct value of γ. Note that the defining complexity of this solution, in terms of degree, is 4, and we obtain the answer exactly.

This example demonstrates that there may be potentially significant advantages to using the Positivstellensatz, rather than Schmüdgen, or Putinar's theorems. Again the analogy to proof systems is that the richer the derivation rules, the shorter the resulting proofs in that proof system tend to be.

Nevertheless, there are many advantages to Putinar's theorem, and its considerably more simple form of polynomial certificate. From the optimization point of view, it offers a canonical sequence of convex relaxations to any polynomial optimization problem, where the optimization is performed over a domain K satisfying the theorem's assumptions. Indeed, as with the Positivstellensatz, using Putinar's theorem, once we decide on a degree bound on the sums of squares $\{s_i(x)\}$ in the representation, then we can obtain these polynomials by solving a semidefinite problem of size polynomial in the degree bound. We have then a canonical sequence of relaxations, where the d^{th} relaxation is given by the restriction that we limit the individual degree of the terms of the right hand side to $d + D$, and $D = \max\{\text{degree}(f_i)\}$, where the f_i are the polynomials that define the semialgebraic set K:

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

Then $\gamma_d \le \gamma_{d+1}$, and $\gamma_d \to \gamma^*$, where γ^* is the exact value for the minimization. This corresponds to considering the subsets of $\mathcal{P}_+^d(K)$ given by the collection of degree d polynomials that have a representation, as in Putinar's theorem, with various degree bounds on the sums of square polynomials $s_i(x)$. Indeed, we note further that if we are interested in the ε–approximation, for any strictly positive value of ε, we are guaranteed to obtain this approximation in a finite number of steps. Lasserre, in [15], proves that for zero–one combinatorial optimization problems, the exact solution, i.e. $\varepsilon = 0$, is obtained at the n^{th} relaxation, where n is the dimension of the problem. Note, however, that at this stage the size of the problem is nevertheless exponential, thus not promising that $\mathcal{P} = \mathcal{NP}$.

5.6 Research Directions

The Positivstellensatz and the Schmüdgen and Putinar Theorems stated above, immediately suggest a number of interesting research directions. We have seen in the examples above that while these algebraic methods provide a unifying theory to a wide class of problems in non–convex optimization, nevertheless the general problem remains intractable. Indeed, as long as we believe that $\mathcal{P} \subsetneq \mathcal{NP}$, then we expect that there will always be instances of positive polynomials which have representations that require exponentially high degree Putinar, and also Positivstellensatz nonnegativity or positivity certificates. Nevertheless, this method gives a new notion of complexity, that is compelling, although is not yet well understood. Certainly, it seems related to our traditional notion of complexity in some way, for there are cases, such as convex quadratic optimization, which is well known to be efficiently solvable, and at the same time can be seen to have short certificates of nonnegativity in its algebraic formulation. Lasserre, in [13] and [14], suggests a connection with a generalization of the notion of duality and the theory of optimization and Lagrange multipliers. On the other hand, it is unclear (to this author, at least) how the algebraic complexity is affected by what we ordinarily consider as the complexity of the constraint set.

Perhaps one of the biggest areas remaining to be investigated, in the interest of providing concrete complexity results, is the relative geometry of polynomials with short and long certificates of positivity, or nonnegativity. Within this question lie embedded at least two other important questions: How "common" or "dense" are the polynomials with very long certificates, in the space of all polynomials, for some appropriate notion of "common," or "dense"? How "close" are these long–certificate polynomials to some short–certificate polynomial?

The first question is a common one in complexity theory, especially in light of the frequently observed phenomenon of problems arising in practice being significantly better behaved than the worst case problem in some problem class. The second question also relates to approximability. For instance, for some relaxation d strictly smaller than the first d^* for which $\gamma_d = \gamma^*$, what can we say about γ_d? How fast does γ_d converge to γ^*? While these questions may have no specific answers in general, it would be interesting to understand their answer for restricted classes of polynomials that have some special structure.

6 The Moment Approach

There is another significant advantage to the Schmüdgen and Putinar versions of the Positivstellensatz, that relates to the 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. Given a full, or partial sequence of moments $\{m_i\}$, there is a distribution with support in $\mathbb{R}$, matching the given moments if and only if the matrices

$$M_{2n} = \begin{pmatrix} m_0 & m_1 & \cdots & m_n \\ m_1 & m_2 & \cdots & m_{n+1} \\ \vdots & & \ddots & \vdots \\ m_n & & & m_{2n} \end{pmatrix}$$

are positive semidefinite for all n up the the highest order of the given moments. Similar semidefinite constraints make up the necessary and sufficient conditions for $K = \mathbb{R}_+$ and $K = [a, b]$. The generalization to arbitrary sets K is not known, and in fact there are various negative results in this direction (see Powers [22]).

The moment problem is, in some sense, dual to the problem of determining positivity of polynomials. A given moment sequence can be thought to specify a linear operator on the linear space spanned by the monomials corresponding to the given moments. Note that if a polynomial $p(x)$ is nonnegative on the support K of any distribution ν, then, evidently,

$$\int_K p(x)\, d\nu \geq 0.$$

We make the following definition.

Definition 5 *We say that a linear functional $\mathcal{L}$ has the K–* **positivity property**, *if for any function f that is nonnegative on K, we also have $\mathcal{L}(f) \geq 0$.*

Near the beginning of the 20^{th} century, Haviland showed (see [11]) that a moment sequence is K–valid if and only if the linear operator $\mathcal{L}$ defined by the given moments, has an extension with the K–positivity property, as defined above, to the space of all polynomials.

Schmüdgen and Putinar's theorems, under the requred restrictions on the support set K, provide a representation for the positive polynomials. This gives us necessary and sufficient conditions for the K–validity of a moment sequence, for

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

and satisfying the required hypotheses of the respective theorems. Then, a moment sequence is K–valid if and only if the operator $\mathcal{L}$ satisfies

$$\mathcal{L}(h_i(x)^2 \cdot f_i(x)) \geq 0,$$

for every polynomial $h_i(x)$, and every $i = 1, \ldots, m$. Letting X denote the vector of monomials, as before, the condition above reduces to the matrix semidefinite condition,

$$\mathcal{L}\left[XX^T \cdot f_i\right] \succeq 0,$$

for every $i = 1, \ldots, m$. For the truncated moment problem, that is, when only moments up to some degree N are given, the K–validity is not computationally tractable, because of the fact that, as we have seen, the individual terms in Putinar's representation may have degree much larger than the actual polynomial being represented. For further discussion on this see [8].

There are many applications of the moment problem that arise when partial information about the moments is known. For instance, see [7] for an application of these methods to bounding linear functionals of solutions to partial differential equations. Additional examples of natural applications of moment problems, as well as a historical overview, may be found in [4].

6.1 Lasserre's Approach

This approach can be applied directly to polynomial optimization problems. This moment approach to optimization is described in detail by Lasserre, in [13], [14], and [15]. 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 [8], [13], the membership test for $V(K)$ may be intractable, as the size of the semidefinite sufficient conditions for K–validity may be exponential in the size of the problem. However, analogously to the canonical sequence of relaxations to the optimization approach using Schmüdgen and Putinar's theorems, we obtain relaxations by using an increasingly tight exterior approximation to $V(K)$, by limiting the size of the semidefinite constraints. We define the d^{th} relaxation to be given by the above minimization performed over the set $V_d(K) \supseteq V(K)$, where we define $V_d(K)$ to be the set of moment sequences (possibly not valid) that satisfy the semidefinite constraints,

$$\mathcal{L}(XX^T \cdot f_i(x)),$$

for X the vector of monomials of degree D_i such that $D_i + \text{degree}(f_i) \leq d + \max_j\{\text{degree}(f_j)\}$.

For zero–one combinatorial optimization problems, Lasserre demonstrates in [15], that $V_n(K) = V(K)$ where n is the dimension of the problem. Laurent, in [16], casts Lasserre's moment approach into a lift–and–project setting, comparing it to the well-known methods of Lovasz–Schijver, and Sherali–Adams. Furthermore, Laurent shows that Lasserre's relaxations are refinements of both the previous methods.

6.2 Applications of Randomization

Lasserre and Parrilo, in their respective papers, develop the machinery described above. Furthermore, they apply these methods to various benchmark discrete, continuous, and mixed optimization problems, with very promising results. However, to the extent that we believe that $\mathcal{P} \neq \mathcal{NP}$, there are many instances of problems that will have poor behavior under these algebraic approaches. The methods and techniques having already been outlined, our aim is to say something specific about the complexity and quality of the intermediate relaxations, for certain classes of problems. In a seminal paper, Goemans and Williamson ([9]) give a randomized algorithm that approximates MAXCUT to at least 0.878. This result may be recast in the context of the above developments, and thus suggest possible improvements for the Goemans–Williamson method, and then generalizations to a much wider class of problems.

6.2.1 Goemans, Williamson, and MAXCUT

Given a graph $G = (V, E)$, with nonnegative weights c_{ij} on the edges, the MAXCUT problem is to find a subset of vertices $S \subseteq V$ such that the weight of the cut, i.e. the weight of the vertices crossing from S to S^c, is maximized. This is an $\mathcal{NP}$–complete problem. It can be formulated as a zero–one integer programming problem. Labeling the nodes $V = \{x_1, \ldots, x_n\}$, we have,

$$\begin{aligned} \text{maximize :} \quad & \frac{1}{2}\sum_{i<j}(1 - x_i x_j)c_{ij}, \\ \text{s.t. :} \quad & x_i^2 = 1. \end{aligned}$$

Goemans and Williamson ([9]) then consider the semidefinite relaxation:

$$\begin{aligned} \text{maximize :} \quad & \frac{1}{2}\sum_{i<j}(1 - z_{ij})c_{ij}, \\ \text{s.t. :} \quad & z_{00} = 1, \\ & Z = (z_{ij}) \succeq 0. \end{aligned}$$

They then define a multivariate normal random variable Y with zero mean and covariance matrix Z, for Z the solution to the semidefinite relaxation above (which, being semidefinite, is a valid covariance matrix). Finally, they obtain a solution to MAXCUT by sampling Y, and rounding up to $+1$ or down to -1, by using zero as a threshold. The expected value of this solution is at least a 0.878–approximation of the true maximum value of the MAXCUT problem. We can recast this in terms of our framework developed above, in a manner that lends itself to generalizations, and potential improvements to the method.

In Lasserre's framework, note that any solution to the convexified problem, i.e. the measure space optimization, yields a measure on the space of feasible solutions. We perform the optimization over the space of moment sequences. A valid moment sequence corresponds to (at least one) a measure on the space of solutions. However, except in the case when one of the early relaxations is exact, since our relaxations define outer approximations of the space of valid moment sequence space, $V(K)$, the optimization procedure will yield some optimizing moment sequence that lies in $V_d(K)$ but generally not in $V(K)$. This means that there does not correspond a measure with support in K with this moment sequence.

The Goemans and Williams procedure can be seen as performing the moment–sequence optimization with the first relaxation, namely, when $d = 0$ in our notation, therefore only enforcing the semidefinite constraint. Since they consider only quadratic moments, this yields a moment sequence of degree 2 that lies in $V_0(K)$. It is straightforward to see that the set of quadratic moments in $V_0(K)$ is a subset of the set of quadratic moment sequences valid for some normal random variable. Goemans and Williamson then choose random samples from this normal random variable, and round to either $+1$ or -1.

There are two ways that immediately suggest themselves for improving this method. First, we may run the optimization to obtain higher order moment sequences, and then attempt to sample from a distribution that matches those higher order moments. This may be a fruitful idea, however there are several difficulties. First, there is no easy way to find a distribution that matches higher order moments. This is further complicated, because unlike the Gaussian case, we have no guarantee even of existence of any distribution matching these "moments." A more promising direction, perhaps, is to consider running the optimization to higher order moments, and then projecting down to the second order moments, and sampling from an appropriate normal distribution. The idea is that this projection will result in a tighter outer approximation of the true valid set $V(K)$.

Karloff's Negative Result:

Promising as the above sounds, a result of Karloff (see [12]) indicates that there is more of a roadblock to improved results, than the above discussion might suggest. Karloff, while he does not seem to consider any of the results contained in this survey paper, nevertheless demonstrates that for the MAXCUT problem in particular, improving the relaxation cannot improve the worst–case bound, as this may be entirely caused by the randomization technique. Karloff provides an instance of MAXCUT where a convex combination of the optimal points gives a feasible optimal solution to the semidefinite optimization, but then the suboptimal binary sampling approximation, of Gaussian sampling followed by rounding, causes a deterioration in the final result to 87.8%. Since this feasible point is in the true moment region, it is feasible for all tighter relaxations. His example shows that the particular method of binary sampling can also contribute significantly to the error of the randomization.

It is not clear whether this problem is inherent to the MAXCUT problem, which naturally has symmetries, and thus multiple optimal solutions, and whether the approaches suggested above may lead to improved results in other problems. Nevertheless, the above does suggest that improved binary sampling and moment matching techniques would lead to an improved performance of these methods in a host of combinatorial optimization approximation problems.

6.2.2 Quadric Polytope

For zero–one random variables, it can be seen that the space $V(K)$ of valid quadratic moment sequences, is the same as what is known as the Boolean Quadric Polytope in combinatorial optimization. This is a polytope that is intractable, in the sense that it has exponentially many facets, and even more, all of those facets are not even known at the current time. Note that the results of Lasserre show that for any particular dimension, eventually the semidefinite program exactly yields the correct quadratic moment sequences, and hence the boolean quadric polytope. Yet this is nevertheless different from knowledge of all the facets, or an explicit description of all the facets of the general boolean quadric polytope, since the semidefinite constraints give successively tighter non–linear outer approximations.

The current work on the Boolean Quadric Polytope (see, for instance, [18]) may have a potentially fruitfully collaboration with the semidefinite and algebraic techniques outlined above. It would be interesting to see if some of the known facet–defining inequalities that are known for the boolean quadric polytope, provide some concrete improvement to the semidefinite outer approximations $V_d(K)$ to $V(K)$, with quality–improving results on the subsequent sampling.

7 Future Directions

These algebraic methods are fresh 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 $co-\mathcal{NP}$.

7.1 Problem and Complexity Classification

As discussed above, one of the most interesting problems remaining wide open in this area, is to gain any understanding of the relative geometry of the hard and easy problems. It is a fact, for instance, that in the infinite dimensional space of polynomials, for any polynomial $f(x) \in \mathcal{P}_+ - \Sigma^2$, and for any $\varepsilon > 0$,

$$B_\varepsilon(f) \not\subseteq \mathcal{P}_+ - \Sigma^2,$$

i.e. the ε–ball about $f(x)$ is never contained in the set of polynomials that are globally nonnegative but cannot be written as a sum of squares (using a norm equivalent to one on the coefficients). This means that, in some appropriate sense, any polynomial that is globally nonnegative may be perturbed to one that has a sum of squares representation. If we are interested in optimization over a compact domain, this is a particularly promising result, as the minimum value is a continuous function of the ε–perturbations, and thus can be precisely controlled. In addition to this, in many instances, particularly continuous problems, the data is only given to the precision of the measurement or storage apparatus, and hence we can justify a slight perturbation of the data, in order to move to a tractable problem. However, the catch is that the perturbation occurs in an infinite dimensional ball. In other words, the perturbation of a 3^{rd} degree polynomial may result in a 2014^{th} degree polynomial, with small, yet nonzero highest order coefficient. In the framework developed here, this would cause a huge increase in the complexity of the problem. Again, if we believe that $\mathcal{P} \neq \mathcal{NP}$, then we must also believe that there exist polynomials that must be perturbed in very high orders in order to make them into a sum of squares. Nevertheless, it is a promising avenue to explore.

Other questions relating to the relative geometry of polynomials, is how ε–closeness of a positive polynomial to a sum of squares polynomial, may translate into approximability, in the sense of the canonical, or other, relaxations, as discussed above.

In summary then, this method seems (to this author) aesthetically appealing, and computationally and theoretically promising. The results are still few, and the questions many, which, for the time being, makes this an exciting field, with many possibilities.

References

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

[2] Bellare, M.; Rogaway, P. "The Complexity of Approximating a Nonlinear Program," J. of Math. Prog. B, (69) 3: 429–441, (1995)

[3] Bertsimas, D.; Tsitsiklis, J. "Introduction to Linear Optimization," Athena Scientific: Belmont MA, 1997.

[4] Bertsimas, D.; Popescu, I. "Optimal Inequalities in Probability Theory: A Convex Optimization Approach" Sloan Working Paper, MIT, (1999)

[5] Bochnak, J.; Coste, M.; Roy, M. "Real Algebraic Geometry," Springer–Verlag: Berlin, 1998.

[6] Vandenberghe, L.; Boyd, S. "Semidefinite Programming," SIAM Review, 38(1):49-95, March 1996.

[7] Caramanis, C. "Solving Linear Partial Differential Equations via Semidefinite Optimization," Master's Thesis, MIT, 2001.

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

[9] Goemans, M.X.; Williamson, D. "Improved Approximation Algorithms for Maximum Cut and Satisfiabiity Problems Using Semidefinite Programming," Proc. 26th An. ACM Symposium on Theory of Comp., Montreal, 422-431, 1994.

[10] Grigoriev, D.; Hirsch, E.; Pasechnik, D. "Inequalities–Based Proof Systems," available at: http://www.maths.univ-rennes1.fr/ dima/ (2001).

[11] Haviland, E. K., "On the momentum problem for distribution functions in more than one dimension II," *Amer. J. Math.*, 58, 164-168, 1936.

[12] Karloff, H. "How Good is the Goemans–Williamson MAX CUT Algorithm," SIAM J. Comput., 29, 336-350, 1999.

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

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

[15] 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.

[16] Laurent, M. "A Comparison of the Sherali–Adams, Lovasz–Schrijver and Lasserre Relaxations for 0–1 Programming," Technical Paper, 2001.

[17] Lovasz, L.; Schrijver, A. "Cones of Matrices and Set-functions and 0–1 Optimization," SIAM J. Opt., 1(2):166-190, (1991)

[18] Macambira, E.; de Souza, C. "The Edge–Weighted Clique Problem: Valid Inequalities, Facets, and Polyhedral Computations," European J. of Op. Res. 123 (2000) 346-371.

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

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

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

[22] Powers, V., Scheiderer, C.; "The Moment Problem for Non-Compact Semialgebraic Sets," Advances in Geometry **1** (2001), 71-88.

[23] Pudlak, P. "On the Complexity of the Propositional Calculus," Sets and Proofs: Invited papers from Logic Colloquium, 1997, 197-218, Cambridge University Press, 1999.

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

[25] Rajwade, A. "Squares," volume 171 of London Math. Society Lecture Note Series. Cambridge University Press, 1993.

[26] 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).

[27] Schmüdgen, K. "The K–Moment Problem for Compact Semialgebraic Sets," Math. Ann. 289, 203-206 (1991).

[28] Sherali, H.; Adams, W. "A Hierarchy of Relaxations Between the Continuous and Convex Hull Representations for 0–1 Programming Problems," SIAM J. Disc. Math., 3(3):411-439, (1990)

[29] Shor, N.Z. "Class of Global Minimum Bounds of Polynomial Functions," Cybernetics, 23(6):731-734, 1987.

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

[31] Stengle, G. "Complexity Estimates for the Schmüdgen Positivstellensatz," J. of Comp., 12, 167-174 (1996).