Geometry, Moments, and Semidefinite Optimization

Dimitris Bertsimas * Constantine Caramanis†

May, 2002

Abstract

Optimization formulations seek to take advantage of structure and information in a particular problem, and investigate how successfully this information constrains the performance measures of interest. In this paper, we apply some recent results of algebraic geometry, to show how the underlying geometry of the problem may be incorporated in a natural way, in a semidefinite optimization formulation. In particular, we discuss two examples; the first, an application to partial differential equations, and the second, an application to deriving optimal inequalities in probability theory. In both cases, the additional "geometry–constraints" significantly improve the quality of the bounds. Furthermore, we believe this framework to be a general one, with many potential applications.

1 Introduction

Is the set over which we optimize connected? Is the set compact? Is it convex? What is the nature of its nonconvexity? Is it discrete? These are among the first questions one encounters when attempting to solve an optimization problem. They help us determine, to a first degree, whether the problem is difficult, how difficult it is, and what solution methods might prove tractable. In many instances, this information may help us properly formulate the problem, and use the correct optimization framework. However, the specific details of this geometry are rarely exploited as extensively as they can, and should be exploited. In this paper, we use the underlying geometry to propose two powerful applications of semidefinite optimization.

The first problem we consider, is obtaining bounds on linear functionals of linear partial differential equations. We do this without the usual discretizing of the state space, into a grid or mesh, as is typical. In addition, by enforcing the constraints implied by the geometry, we are able to obtain information about the nonlinear functionals **max** and **min**.

The second problem we consider is that of obtaining optimal inequalities in probability theory. This was first considered by Bertsimas and Popescu in [2]. Here, we show how the geometry–induced constraints may be strengthened, to yield tighter solutions. In both, we apply some results from the field of Real Algebraic Geometry, and show how *a priori* knowledge of the support of the solution, in the PDE problem, and the support of the distribution, in the probability inequality

*Boeing Professor of Operations Research, Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, E53-363, Cambridge, MA 02142, dbertsim@mit.edu; Research supported by the SMA-MIT alliance.

†Laboratory for Information and Decision Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307; cmcaram@mit.edu. Research partially supported by the SMA-MIT alliance and an NSF predoctoral fellowship.

problem, can be incorporated via semidefinite optimization and the theory of moments, and fruitfully exploited to obtain significantly improved bounds.

In section 2 we give the main idea behind the geometry–induced bounds, and ouline the pertinent theory of moments, polynomials, and semidefinite optimization. In section 3 we give a semidefinite approach to solving for linear functionals of linear PDEs, along with some promising numerical results in section 4. In section 5 we outline the procedure for obtaining optimal probability inequalities, first given by Bertsimas and Popescu ([2]), and then show how this may be strengthened with the algebraic ideas discussed here.

2 Exploiting Geometry via Moments

Every distribution μ, has a sequence of moments (bounded or not) given by integration:

$$m_{\boldsymbol{\alpha}} := \int x^{\boldsymbol{\alpha}}\, d\mu,$$

where $\boldsymbol{\alpha}$ is a multiindex. Of course, not every sequence of numbers is a moment sequence. For instance, we know that the second moment must be nonnegative. The problem of classifying valid moment sequences has a long history in mathematics, and is intimately connected with the analytic, geometric, and algebraic branches of mathematics. See [1], [2], and the references therein for more on the history and developments. Let $V(K)$ denote all the possible moment sequences that correspond to a distribution with support on the set $K \subseteq \mathbb{R}^n$. Next, let P denote a polyhedral set, defined by some inequalities and equalities. Suppose that we wish to maximize some function $f(\boldsymbol{m})$ of the moments of an unknown distribution, and suppose we know that these moments must lie in the polyhedron P. Then, we could solve:

$$\begin{aligned} \max: \quad & f(\boldsymbol{m}) \\ s.t.: \quad & \boldsymbol{m} \in P. \end{aligned}$$

In this case, however, we have not used two important pieces of information specified by the problem: First, the sequence $\boldsymbol{m}$ is not just a vector, it must correspond to the moments of some distribution. In addition, that distribution has known support K. In this paper, we show that solving, instead, the strengthened problem

$$\begin{aligned} \max: \quad & f(\boldsymbol{m}) \\ s.t.: \quad & \boldsymbol{m} \in P \cap V(K), \end{aligned}$$

can result in a significantly tighter formulation. We use this tighter formulation to give two novel applications of semidefinite optimization. Furthermore, we provide a discussion of the complexity tradeoff of capturing this geometry constraint.

2.1 The Moment Problem

The key then is in computing the set $V(K)$ above, as it is there where the geometry constraints are encoded. As mentioned above, computing $V(K)$, i.e. characterizing all moment sequences with support on the set K, is known as the classical moment problem. The problem is to determine, given some sequence of numbers, whether it is a valid moment sequence, that is to say, whether the numbers given are indeed the moments of a nonnegative function or distribution, with the given support. In one dimension, a sequence of moments $\{m_i\}$ comes from a distribution with support on $\mathbb{R}$, if and only if the matrix

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

is positive semidefinite for every n. If we demand that the support be $[0, \infty)$, we need to add the additional constraint, that the matrix

$$M_{2n+1} = \begin{pmatrix} m_1 & m_2 & \cdots & m_{n+1} \\ m_2 & m_3 & \cdots & m_{n+2} \\ \vdots & & \ddots & \vdots \\ m_{n+1} & & & m_{2n+1} \end{pmatrix}$$

also be positive semidefinite.

In multiple dimensions, it is generally unknown which are the exact necessary and sufficient conditions for $M_{\boldsymbol{\alpha}}$ to be a valid moment sequence, when we are working over a general domain. For a wide class of domains, however, Schmüdgen [10] and Putinar [8] find such conditions. We review this work briefly, and use it to derive the necessary and sufficient conditions for $M_{\boldsymbol{\alpha}}$ to be a moment sequence.

An Operator Approach

Given a closed subset Ω of $\mathbb{R}^d$, a sequence of numbers $M_{\boldsymbol{\alpha}}$ defines a valid moment sequence if there exists a measure μ such that

$$M_{\boldsymbol{\alpha}} = \int_\Omega \boldsymbol{x}^{\boldsymbol{\alpha}} d\mu.$$

We define the linear operator

$$Hf = \int_\Omega f(\boldsymbol{x}) d\mu.$$

It is obviously necessary that $Hf \geq 0$, whenever $f \geq 0$ on Ω. A classical theorem says that it is also sufficient:

Theorem 1 (Haviland [4]). *If $\Omega \subseteq \mathbb{R}^n$ is closed, then $M_{\boldsymbol{\alpha}}$ defines a valid moment sequence if and only if the linear operator H is nonnegative on all polynomials that are nonnegative on Ω.*

Theorem 1 implies that the problem of finding necessary and sufficient conditions for $M_{\boldsymbol{\alpha}}$ to be a moment sequence, reduces to checking the nonnegativity of the image of a polynomial that is nonnegative on Ω. In one dimension, we know that any polynomial that is nonnegative may be written as the sum of squares. Since the square of a polynomial may be written as a quadratic form, the nonnegativity of the operator reduces to matrix semidefiniteness conditions. The Motzkin polynomial in $\mathbb{R}^3$,

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

is an example that shows that in higher dimensions, the sum of squares decomposition of a nonnegative polynomial is not in general possible (see Parrilo [7] or Reznick [9] for details). However, Schmüdgen [10] and Putinar [8] give a representation of all polynomials that are nonnegative over a compact finitely generated semialgebraic set K, as defined in the theorem below. This leads to necessary and sufficient conditions for a moment sequence to be valid on K.

Theorem 2 (Putinar [8]). *Suppose* $K := \{\boldsymbol{x} \in \mathbb{R}^n : f_i(\boldsymbol{x}) \geq 0,\ 1 \leq i \leq r\}$ *is closed and bounded, where* $f_i(\boldsymbol{x})$ *are polynomials. Suppose further that there exists a polynomial* $h(x)$ *with compact level set, that is of the form*

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

where the $s_i(x)$ *are sums of squares of polynomials. Then a polynomial* $g(\boldsymbol{x})$ *is positive on* K *if and only if it is expressible as a sum of terms of the form*

$$q_i^2(\boldsymbol{x}) f_i(\boldsymbol{x}),$$

for $i \in \{0, 1, \ldots, r\}$*,* $f_0 = 1$*, and* $q_i(\boldsymbol{x})$ *some polynomial.*

Theorems 1 and 2 lead to the following result.

Theorem 3. *Consider the vector* $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{\alpha}}$*, and the semidefinite matrix* $\boldsymbol{F}(\boldsymbol{x})\boldsymbol{F}(\boldsymbol{x})'$*. Given* $\boldsymbol{M} = [M_{\boldsymbol{\alpha}}]$*, there exists a distribution* $\mu(\boldsymbol{x})$ *such that*

$$M_{\boldsymbol{\alpha}} = \int_{\Omega} \boldsymbol{x}^{\boldsymbol{\alpha}}\, d\mu(\boldsymbol{x})$$

for a closed and bounded domain Ω *of the form above, if and only if the following matrices are positive semidefinite:*

$$\int_{\Omega} [\boldsymbol{F}(\boldsymbol{x})\boldsymbol{F}(\boldsymbol{x})'] f_i(x), \tag{1}$$

for $i \in \{0, 1, \ldots, r\}$*.*

Examples of domains for which the above result applies include the unit ball in $\mathbb{R}^d$, which can be written as

$$B = \{\boldsymbol{x} \in \mathbb{R}^d : 1 - x_1^2 - \cdots - x_d^2 \geq 0\},$$

the unit hypercube,

$$C = \{\boldsymbol{x} \in \mathbb{R}^d : x_i \geq 0, 1 - x_i \geq 0,\ 1 \leq i \leq d\},$$

and the discrete set,

$$\{-1, 1\}^d = \{\boldsymbol{x} \in \mathbb{R}^d : x_i^2 - 1 \geq 0, 1 - x_i^2 \geq 0, i = 1, \ldots, d\}.$$

Complexity

Note that in general, while the above semidefinite conditions are necessary and sufficient, the semidefiniteness of the truncated matrix is only a necessary condition. This is a consequence of the fact that in Putinar's representation of positive polynomials, the degree of the sum of squares polynomials is not specified. Indeed, unless $\mathcal{P} = \mathcal{NP}$, then we are guaranteed that in the worst case, the representation will require exponentially high degree polynomials. Nevertheless, the semidefiniteness of each truncation of the matrices is a necessary condition. Thus, we have obtained a hierarchy of increasingly tight semidefinite constraints on our moments, giving us a tradeoff between complexity, and tight description of the set $V(K)$ of valid moments.

3 Partial Differential Equations

In this section, we give what we believe to be a novel approach to bounding functionals of linear partial differential equations. The linear functionals we consider, are functions of the moments of the solution to the PDE. First, we exploit the linearity of the operator to obtain linear constraints on the moments. Stopping at this stage would yield an optimization problem of the form illustrated at the beginning of section 2:

$$\begin{aligned} \max / \min : \quad & f(\boldsymbol{m}) \\ s.t. : \quad & \boldsymbol{m} \in P, \end{aligned}$$

where P denotes the polytope defined by the linear constraints. In the sequel, we use the methodology developed in section 2 above, to formulate the significantly stronger optimization problem,

$$\begin{aligned} \max / \min : \quad & f(\boldsymbol{m}) \\ s.t. : \quad & \boldsymbol{m} \in P \cap V(K), \end{aligned}$$

As the numerical results in section 4 indicate, the geometry–induced constraints are critical for obtaining reasonably tight upper and lower bounds.

Linearity

Suppose we are given the partial differential operators L and G operating on some distribution space $\mathcal{A}$:

$$L, G : \mathcal{A} \longrightarrow \mathcal{A},$$

and we are interested in finding

$$\int Gu(\boldsymbol{x}),$$

where $u \in \mathcal{A}$ (note also that $f \in \mathcal{A}$) satisfies the PDE,

$$Lu(\boldsymbol{x}) = f(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_1, \ldots, x_d) \in \Omega \subset \mathbb{R}^d, \tag{2}$$

including the appropriate boundary conditions on $\partial\Omega$,

Eq. (2) is understood in the sense that both sides of the equation act in the same way on a given class of functions $\mathcal{D}$, i.e.,

$$Lu = f \Longleftrightarrow \int (Lu)\phi = \int f\phi, \quad \forall \phi \in \mathcal{D},$$

where $\mathcal{D}$ is taken to be some sufficiently nice class of test functions—typically a subset of the smooth functions $\mathcal{C}^\infty$.

We will assume that the operators L and G are linear operators with coefficients that are polynomials of the variables. In Section 3.5 we discuss extensions for a nonlinear operator G. In particular,

$$Lu(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} L_{\boldsymbol{\alpha}}(\boldsymbol{x}) \frac{\partial^{\boldsymbol{\alpha}} u(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}}, \qquad Gu(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} G_{\boldsymbol{\alpha}}(\boldsymbol{x}) \frac{\partial^{\boldsymbol{\alpha}} u(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}},$$

where $\boldsymbol{\alpha} = (i_1, \ldots, i_d)$ is a multi-index,

$$\frac{\partial^{\boldsymbol{\alpha}} u(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} = \frac{\partial^{\sum_k i_k} u(\boldsymbol{x})}{\partial x_1^{i_1} \cdots \partial x_d^{i_d}},$$

and $L_{\boldsymbol{\alpha}}(\boldsymbol{x})$ and $G_{\boldsymbol{\alpha}}(\boldsymbol{x})$ are multivariate polynomials (we discuss extensions in Section 3.6). We will restrict ourselves to the case where $\mathcal{D}$ is separable, that is, it has a countable dense subset. This restriction is not as limiting as it might first appear. In particular, if the solution u has compact support, then we may also assume without loss of generality that every element of $\mathcal{D}$ has compact support as well, and thus by the Stone–Weierstrass theorem, $\mathcal{D}$ is separable. The condition that u have compact support may also be replaced by the (slightly) weaker condition that u have exponentially decaying tails.

Let $\mathcal{F} = \{\phi_1, \phi_2, \dots\}$ generate (in the basis sense) a dense subset of $\mathcal{D}$. Then, by the linearity of integration we have

$$\begin{aligned} Lu = f \quad \Longleftrightarrow \quad & \int (Lu)\phi = \int f\phi, \quad \forall \phi \in \mathcal{D}, \\ \Longleftrightarrow \quad & \int (Lu)\phi_i = \int f\phi_i, \quad \forall \phi_i \in \mathcal{F}. \end{aligned}$$

We discuss different choices for the subset $\mathcal{F}$ in Section 3.6. One separable subspace around which this paper focuses is the subspace spanned by the monomials $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{i_1} \dots x_d^{i_d}$. Polynomials have the property that they are closed under action by polynomial coefficient differential operators.

The Adjoint Operator

The adjoint operator, L^*, is defined by the equation:

$$\int (Lu)\phi = \int u(L^*\phi), \quad \forall \phi \in \mathcal{D}.$$

Therefore, if we have both L and L^*, then equality in the original PDE becomes:

$$\begin{aligned} Lu = f \quad \Longleftrightarrow \quad & \int (Lu)\phi = \int f\phi, \quad \forall \phi \in \mathcal{D}, \\ \Longleftrightarrow \quad & \int (Lu)\phi_i = \int f\phi_i, \quad \forall \phi_i \in \mathcal{F}, \\ \Longleftrightarrow \quad & \int u(L^*\phi_i) = \int f\phi_i, \quad \forall \phi_i \in \mathcal{F}. \end{aligned} \tag{3}$$

To illustrate the computation of the adjoint operator, we consider the one dimensional case. The general term of this operator is, up to a constant multiple:

$$x^a \frac{\partial^b}{\partial x^b}.$$

Using the notation $\tilde{\phi} = x^a\phi$, this term's contribution to the adjoint operator is as follows.

$$\begin{aligned} \int_\Omega x^a(\partial^b u)\phi &= \int_\Omega (\partial^b u)(x^a\phi)\,dx = \int_\Omega (\partial^b u)\tilde{\phi}\,dx \\ &= u^{(b-1)}\tilde{\phi}\Big|_{\partial\Omega} + \cdots + (-1)^{k+1}u^{(b-k)}\tilde{\phi}^{(k-1)}\Big|_{\partial\Omega} + \cdots \\ &\quad + (-1)^{b+1}u\tilde{\phi}^{(b-1)}\Big|_{\partial\Omega} + (-1)^b \int_\Omega u\partial^b\tilde{\phi}\,dx. \end{aligned}$$

Thus, while perhaps notationally tedious in higher dimensions, computing the adjoint of a linear partial differential operator with polynomial coefficients is essentially only as difficult as performing the chain rule for differentiation on polynomials, and in particular, it may be easily automated.

3.1 Linear Constraints

Let us define variables in an optimization sense

$$M_{\boldsymbol{\alpha}} = \int_\Omega \boldsymbol{x}^{\boldsymbol{\alpha}} u(\boldsymbol{x}) = \int_\Omega {x_1}^{i_1} \cdots {x_d}^{i_d} u(\boldsymbol{x}),$$

together with variables related to the boundary $\partial\Omega$:

$$z_{\boldsymbol{\alpha}} = \int_{\partial\Omega} \boldsymbol{x}^{\boldsymbol{\alpha}} u(\boldsymbol{x}) = \int_{\partial\Omega} {x_1}^{i_1} \cdots {x_d}^{i_d} u(\boldsymbol{x}).$$

The specific form of these variables depends on the nature of the boundary conditions we are given (see Section 4 for specific examples). We refer to the quantities $M_{\boldsymbol{\alpha}}$ and $z_{\boldsymbol{\alpha}}$ as moments, even though $u(\cdot)$ is not a probability distribution. We select as ϕ_i's the family of monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$. Since, for the case we are considering, L, and thus L^*, are linear operators with coefficients that are polynomials in $\boldsymbol{x}$, then Eqs. (3) can be written as linear equations in terms of the variables $\boldsymbol{M} = (M_{\boldsymbol{\alpha}})$ and $\boldsymbol{z} = (z_{\boldsymbol{\alpha}})$.

3.2 Objective Function Value

Given that the operator G is also a linear operator with coefficients that are polynomials of the variables, then the functional $\int Gu$ can also be expressed as a linear function of the variables $\boldsymbol{M}$ and $\boldsymbol{z}$. So if we minimize or maximize this particular linear function, we obtain upper and lower bounds on the value of the functional.

3.3 Geometry–Induced Constraints

Let us assume that the solution we are looking for is bounded from below, that is $u(\boldsymbol{x}) \geq u_0$. The constant u_0 is in fact unknown. In certain cases, u_0 is naturally known; for example if $u(\boldsymbol{x})$ is a probability distribution, then $u(\boldsymbol{x}) \geq 0$, i.e., $u_0 = 0$; or if $u(\boldsymbol{x})$ represents temperature, then again $u(\boldsymbol{x}) \geq 0$. In any case, $\hat{u}(x) := u(x) - u_0 \geq 0$. Therefore all our work from section 2 applies. We can now impose semidefinite constraints on the moments that we have so–far managed to constrain with linear constraints. Indeed, we know that since $u(x) - u_0 \geq 0$ for all $x \in \Omega$, then the "moments" $\hat{\boldsymbol{m}}$, of $\hat{u}(x) = (u(x) - u_0)$ are indeed valid moments, and must therefore satisfy the constraints of section 2. Furthermore, these moments $\hat{\boldsymbol{m}}$ are related to $\boldsymbol{m}$ by the known and computable relation

$$\begin{aligned} m_{\boldsymbol{\alpha}} &= \hat{m}_{\boldsymbol{\alpha}} + u_0 \int_\Omega x^\alpha \\ &= \hat{m}_{\boldsymbol{\alpha}} + u_0 c_{\boldsymbol{\alpha}}, \end{aligned}$$

where the $\{c_{\boldsymbol{\alpha}}\}$ are known constants. Therefore, the semidefinite constraints on $\{\hat{m}_{\boldsymbol{\alpha}}\}$ become semidefinite constraints on $\{m_{\boldsymbol{\alpha}} - u_0 c_{\boldsymbol{\alpha}}\}$. Note, therefore, that we need not assume that we know u_0. If it is unknown, then it simply becomes one more decision variable in the optimization.

3.4 The Overall Formulation

As we mentioned, the variables are the triplet $(M_{\boldsymbol{\alpha}}, z_{\boldsymbol{\alpha}}, u_0)$ where we have the moments $M_{\boldsymbol{\alpha}} = \int_\Omega \boldsymbol{x}^{\boldsymbol{\alpha}} u(\boldsymbol{x})$, the boundary moments $z_{\boldsymbol{\alpha}} = \int_{\partial\Omega} \boldsymbol{x}^{\boldsymbol{\alpha}} u(\boldsymbol{x})$ and the bound u_0, which might be naturally known. The semidefinite optimization consists of linear equality constraints generated by the adjoint operator for different test functions $\boldsymbol{x}^{\boldsymbol{\alpha}}$, and of the semidefinite constraints that express the fact that the variables $M_{\boldsymbol{\alpha}}$ and $z_{\boldsymbol{\alpha}}$ are in fact moments of a distribution with a known support. Subject to these constraints, we maximize and minimize a linear function of the variables that expresses the given linear functional. The overall steps of the formulation process are then summarized as follows:

(i) Compute the adjoint operator L^*.

(ii) Generate the n^{th} equality constraint by requiring that

$$\int u(L^*\phi_n) = \int f\phi_n.$$

(iii) Generate the desired semidefinite constraints; note that these only depend on the domain Ω and not on the operator L.

(iv) Compute upper and lower bounds on the given functional by solving a semidefinite optimization problem over the intersection of the positive semidefinite cone and the equality constraints.

3.5 The Maximum and Minimum Operator

Suppose that the given functional is

$$Gu = \min_{\boldsymbol{x}\in\Omega} u(\boldsymbol{x}).$$

Then, we will formulate the objective function

$$\min u_0.$$

This approach gives a lower bound on the minimum of $u(\boldsymbol{x})$ over Ω. However, if we maximize $\max u_0$ we do not obtain a true upper bound on the minimum of $u(\boldsymbol{x})$ over Ω, only an approximation.

Similarly, if we are interested in

$$\max_{\boldsymbol{x}\in\Omega} u(\boldsymbol{x})$$

we solve $\max v_0$ such that $u(\boldsymbol{x}) \leq v_0$, which leads to semidefinite constraints involving $\boldsymbol{M}$, $\boldsymbol{z}$ and v_0. This approach gives an upper bound on v_0, while minimizing v_0 only leads to an approximation.

Note that here the semidefinite constraints are absolutely crucial. This is because the additional variable u_0 is introduced linearly, and because of the linearity of integration, cannot possibly be calculated by the family of linear constraints. Rather, the linear constraints link it to the variables of the optimization, and then it is constrained by the semidefinite constraints.

3.6 Using Trigonometric Moments

Instead of choosing polynomials as test functions, we could choose other classes of test functions. Polynomials are particularly convenient as they are closed under differentiation. While this property is not a necessary condition for the proposed method to work, it significantly limits the proliferation of variables we introduce. When the linear operator has coefficients that are not polynomials, other bases might be more appropriate.

The trigonometric functions $\{\sin(nx), \cos(nx)\}$ are also closed under differentiation (again we can form products in higher dimensions, just as with monomials). Using trigonometric functions as a basis of our test functions provides a straightforward way for us to deal with linear operators with trigonometric coefficients. This is an important point, namely, that the choice of test function basis ought to depend on the coefficients of the linear operator.

4 Examples

In this section, we illustrate our approach with two examples: (a) a simple, but interesting ODE: Bessel's equation, and (b) a two-dimensional partial differential equation known as Helmholtz's equation. We take the solution to have support on the unit interval for the ODE, and on the unit square for the PDE.

4.1 The Bessel Equation

In this section we consider Bessel's differential equation

$$x^2u'' + xu' + (x^2 - p^2)u = 0.$$

The Bessel function and its variants appear in one form or another in a wide array of engineering applications, and applied mathematics. Furthermore, while there are integral and series representations, the Bessel function is not expressible in closed form. The series representation of the Bessel function, which can be found in, e.g. Watson's "Treatise on the Theory of Bessel Functions," is:

$$J_p(x) := \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{2k+p}}{k!(k+p)!}.$$

Also, over the appropriate range, the Bessel function is neither nonnegative, nor convex. In order to avoid numerical difficulties from large constant factors, we solve a modified version of Bessel's equation:

$$x^2u'' + xu' + (49x^2 - p^2)u = 0. \tag{4}$$

The solution is $u(x) = J_p(7x)$. Assuming we are given the value of the derivatives on the boundary, using the monomials as the test functions, we obtain the adjoint equations:

$$\begin{aligned}
\phi = 1: \quad &\Rightarrow \quad -u(1) + (1 - p^2)m_0 + 49m_2 = u'(1), \\
\phi = x: \quad &\Rightarrow \quad -2u(1) + (4 - p^2)m_1 + 49m_3 = u'(1), \\
\phi = x^2: \quad &\Rightarrow \quad -3u(1) + (9 - p^2)m_2 + 49m_4 = u'(1), \\
&\vdots \\
\phi = x^n: \quad &\Rightarrow \quad -(n+1)u(1) + ((n+1)^2 - p^2)m_n + 49m_{n+2} = u'(1).
\end{aligned}$$

In what follows, we choose $p = 1$. We used SeDuMi (see [11]) to compute the moments, and also to compute the max and min. Recall from the discussion in Section 3.5 that while we are able to obtain bounds for the moments, our method can only compute *approximations* to the max and min of the solution. In the case of the Bessel function, the approximations we obtain of the minimum are greater than the actual value, and the approximations for the maximum are less than the actual value. The true values are: min $= -0.347$ and max $= 0.583$. In Table 1 we report the results from using SeDuMi. Next, we use these results to translate the function so that it is nonnegative, and so that we can compute the moments of the translated function. We use $u(x) - u_0 \geq 0$ with $u_0 = -0.4$. Again using SeDuMi, we obtain very accurate bounds to the moments. We give the first few in Table 2. We would expect by linearity, and indeed the results show, that just having a lower bound on the function is enough to find accurate results on the moments of the function.

N	Minimum	Maximum
20	-0.3087	0.4986
24	-0.3101	0.5068
30	-0.3111	0.5081
40	-0.3142	0.5046

Table 1: Approximations for the maximum and the minimum of the solution of Eq. (4) computed using SeDuMi. True Values: min $= -0.347$, max $= 0.583$. Note that these are nonlinear functionals and generally very difficult to compute.

Variable	LB	UB
m_1	0.1766	0.1766
m_2	0.0903	0.0903
m_3	0.0583	0.0583
m_4	0.0438	0.0438
m_5	0.0361	0.0361

Table 2: Upper and lower bounds for the moments of the solution to Eq. (4) for $N = 24$, computed using SeDuMi.

4.2 The Helmholtz Equation

In this section we consider the two dimensional PDE

$$\triangle u + k^2 u = f, \tag{5}$$

over $\Omega = [0, 1]^2$. To compute the adjoint operator we need to use Stokes's formula:

$$\int_\Omega d\omega = \int_{\partial\Omega} \omega.$$

Recall that in two dimensions we have:

$$\omega = f\,dx + g\,dy \Longleftrightarrow d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dxdy,$$

and thus computing the adjoint operator, we have:

$$\begin{aligned}
\int_\Omega \left(\frac{\partial^2 u}{\partial x^2}\right) \phi\,dxdy &= \int_\Omega \left(\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\cdot\phi\right) - \frac{\partial u}{\partial x}\frac{\partial \phi}{\partial x}\right) dxdy \\
&= \int_{\partial\Omega} \frac{\partial u}{\partial x}\cdot\phi\,dy - \int_\Omega \left(\frac{\partial}{\partial x}\left(u\frac{\partial\phi}{\partial x}\right) - u\frac{\partial^2\phi}{\partial x^2}\right) dxdy \\
&= \int_{\partial\Omega} \left(\frac{\partial u}{\partial x}\cdot\phi - u\frac{\partial\phi}{\partial x}\right) dy + \int_\Omega u\frac{\partial^2\phi}{\partial x^2}\,dxdy,
\end{aligned}$$

By a similar process for the $\frac{\partial^2 u}{\partial y^2}$ term, we obtain

$$\begin{aligned}
\int_\Omega \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f\right)\phi\,dxdy = \int_\Omega \left(f\phi + \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2}\right) dxdy \\
+ \int_{\partial\Omega} \left(\frac{\partial u}{\partial x}\cdot\phi - u\frac{\partial\phi}{\partial x}\right) dy - \int_{\partial\Omega} \left(\frac{\partial u}{\partial y}\cdot\phi - u\frac{\partial\phi}{\partial y}\right) dx.
\end{aligned}$$

Again we consider the family of monomials,

$$\mathcal{F} = \{x^i \cdot y^j\}, \quad \text{for } i, j \in \mathbb{N} \cup \{0\}.$$

In addition to the variables

$$m_{i,j} = \int_0^1 \int_0^1 x^i y^j u(x,y) dx dy,$$

we also introduce the boundary moment variables:

$$\begin{aligned}
b_i^{x=1} &:= \int_0^1 u(x=1,y) y^i \, dy, & d_i^{x=1} &:= \int_0^1 \partial_x u(x=1,y) y^i \, dy \\
b_i^{x=0} &:= \int_0^1 u(x=0,y) y^i \, dy, & d_i^{x=0} &:= \int_0^1 \partial_x u(x=0,y) y^i \, dy \\
b_i^{y=1} &:= \int_0^1 u(x,y=1) x^i \, dx, & d_i^{y=1} &:= \int_0^1 \partial_y u(x,y=1) x^i \, dx \\
b_i^{y=0} &:= \int_0^1 u(x,y=0) x^i \, dx, & d_i^{y=0} &:= \int_0^1 \partial_y u(x,y=0) x^i \, dx.
\end{aligned}$$

Then the adjoint relationship above yields:

$$\begin{aligned}
\phi = 1: \quad &\Rightarrow \quad d_0^{x=1} - d_0^{x=0} + d_0^{y=1} - d_0^{y=0} + m_{0,0} = \int_\Omega f \, dx dy \\
\phi = x^i: \quad &\Rightarrow \quad d_0^{x=1} + i(i-1) m_{i-2,0} + d_i^{y=1} - d_i^{y=0} + m_{i,0} = \int_\Omega f \cdot x^i \, dx dy \\
\phi = y^j: \quad &\Rightarrow \quad d_j^{x=1} - d_j^{x=0} + d_0^{y=1} + j(j-1) m_{0,j-2} + m_{0,j} = \int_\Omega f \cdot y^i \, dx dy \\
\phi = x^i y^j: \quad &\Rightarrow \quad d_j^{x=1} + i(i-1) m_{i-2,j} + d_i^{y=1} + j(j-1) m_{i,j-2} + m_{i,j} = \int_\Omega f \cdot x^i y^j \, dx dy.
\end{aligned}$$

Note that either the $\{d_i^x, d_j^y\}$, or the $\{b_i^x, b_j^y\}$, are given as boundary values. In order to compare with the exact solution, we selected the boundary conditions such that $f(x,y) = 3e^{x+y}$.

In order to illustrate the power of the semidefinite constraints, we run our optimization problem in two different stages. First, we provide the results of solving the linear optimization problem generated by the adjoint equation, using the commercial software AMPLE. Next, we enforce the semidefinite constraints. Solving a linear optimization problem, ignoring the semidefinite constraints and only imposing nonnegativity constraints on the variables, we obtain the bounds in Table 3 for $N = 5, 10, 20$.

We then add the semidefinite constraints and use SDPA (see [5]) to solve the corresponding semidefinite optimization problems. SDPA gave very tight bounds; however, there were nevertheless some numerical instabilities, which required us to change the desired accuracy in the search parameters, in order to obtain answers that made sense. In Table 4 we report upper and lower bounds for $N = 5,\ 10,\ N = 20$. Note that when we use monomials up to degree N, there are in fact N^2 such monomials.

We notice that the tightness of the bounds is nearly as dramatic as in the one dimensional case. Moreover, the bounds using semidefinite optimization are significantly tighter than the ones obtained using linear optimization. This observation is significant and emphasizes the importance of the semidefinite constraints. For example, without the semidefinite constraints, the upper bound on $m_{0,0}$ is $+\infty$, whereas we obtain very tight bounds for $N = 10$ using the semidefinite constraints.

Variable	LB, $N=5$	UB, $N=5$	LB, $N=10$	UB, $N=10$	LB, $N=20$	UB, $N=20$
$m_{0,0}$	0.0000	$+\infty$	0.0000	$+\infty$	0.0000	$+\infty$
$m_{1,0}$	0.0000	4.3142	0.0000	4.0822	0.0000	3.8694
$m_{1,1}$	0.7559	1.0881	0.8557	1.0419	0.9400	1.0120
$m_{2,0}$	0.0000	4.6790	0.0000	4.6790	0.0000	4.6790
$m_{2,1}$	0.0545	1.0563	0.1417	1.0059	0.1749	0.9753
$m_{2,2}$	0.0000	0.9447	0.0000	0.9087	0.0000	0.9087
$m_{3,0}$	0.0000	4.8743	0.0000	4.4932	0.0000	4.4414
$m_{3,1}$	0.1692	0.6806	0.4015	0.6063	0.5025	0.5758
$m_{3,2}$	0.0000	0.6383	0.0000	0.6105	0.0000	0.5989
$m_{3,3}$	0.0000	0.5291	0.1684	0.3640	0.2573	0.3296

Table 3: Upper and lower bounds from linear optimization for $N=5$, $N=10$, $N=20$. For the cases $N=5,10$ we only run up to $m_{2,2}$.

Variable	LB, $N=5$	UB, $N=5$	LB, $N=10$	UB, $N=10$	LB, $N=20$	UB, $N=20$
$m_{0,0}$	2.7687	5.5689	2.9113	2.9809	2.9377	2.9599
$m_{1,0}$	1.8377	2.2006	1.6955	1.7276	1.7106	1.7199
$m_{1,1}$	0.9679	1.0089	0.9851	1.0026	0.9988	1.0005
$m_{2,0}$	1.0860	1.8173	1.2058	1.2437	1.2261	1.2353
$m_{2,1}$	0.6089	0.7626	0.7019	0.7169	0.7177	0.7185
$m_{2,2}$	0.3192	0.5456	0.5010	0.5107	0.5148	0.5161
$m_{3,0}$	0.5263	1.8191	0.9411	1.0000	0.9601	0.9683
$m_{3,1}$	0.5296	0.6076	0.5485	0.5653	0.5623	0.5639
$m_{3,2}$	0.2021	0.4069	0.4009	0.4068	0.4036	0.4046
$m_{3,3}$	0.1025	0.3495	0.3139	0.3191	0.3163	0.3179

Table 4: Upper and lower bounds for Eq. (5) for $N=5,\ 10$ using SDPA. The computation of each bound took less than 0.5 seconds for $N=5$, 3-5 seconds for $N=10$, and 1-3 minutes for $N=20$.

5 Optimal Probability Inequalities

The following example, while trivial, illustrates our point. Suppose we know that a distribution has mean of zero. At this point, the amount of information we have about our distribution is limited. But if in addition we happen to know that the support of the distribution is $\Omega = [0,1]$, then we know the entire distribution. Bertsimas and Popescu, in [2], give a semidefinite approach to obtaining information of a distribution, based on its moments, and its support. In this section we show that the polynomial representation theorem of Schmüdgen and Putinar, and the framework developed in section 2, can be used to strengthen these results. First we briefly outline their results. We refer the interested reader to [2] for the bulk of the results, as the purpose here is to stress the importance of the geometry constraints, and show how the algebraic results outlined in section 2 aid in their impementation.

Bertsimas and Popescu consider the (n, k, Ω) problem, of obtaining as good as possible an upper bound on $P(X \in S)$, given that X is an $\mathbb{R}^n$ valued random variable, with support on Ω, and given all the moments $\{\boldsymbol{m_\alpha}\}$ up to degree k. Following them, we formulate the problem as an optimization problem, as follows:

$$\begin{aligned} \max: \quad & \int_S 1 \, d\mu \\ s.t.: \quad & \int_\Omega x^{\boldsymbol{\alpha}} \, d\mu = \boldsymbol{m}_{\boldsymbol{\alpha}}, \quad |\boldsymbol{\alpha}| \leq k. \end{aligned}$$

This is a problem with infinitely many variables, and finitely many constraints. Using standard duality theory, and associating a dual variable $\boldsymbol{y}_{\boldsymbol{\alpha}}$ to each constraint above, we can consider the dual problem, which has finitely many variables, and infinitely many constraints:

$$\begin{aligned} \min: \quad & \sum_{|\boldsymbol{\alpha}| \leq k} \boldsymbol{y}_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}} \\ s.t.: \quad & g(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq k} \boldsymbol{y}_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}} \geq 1, \quad \forall x \in S \\ & g(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq k} \boldsymbol{y}_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}} \geq 0, \quad \forall x \in \Omega. \end{aligned}$$

We see that in the dual formulation, the constraints amount to nonnegativity of the polynomial $g(\boldsymbol{x})$ over Ω, and nonnegativity of the polynomial $(g(\boldsymbol{x}) - 1)$ over the set S. Therefore we can express these constraints using semidefinite optimization, and the algebraic methods outlined in section 2.

Definition 1. *Suppose K is defined by polynomials $g_1, \ldots, g_l$, and satisfies the hypotheses of theorem 1. Then, let $\mathcal{P}_r(K)$ denote the set of polynomials positive on the set $K \subseteq \mathbb{R}^n$, that have a representation as in Putinar's theorem, in terms of polynomials of degree at most $2r$. In other words,*

$$\mathcal{P}_r(K) = \{f(x) \in \mathbb{R}[x] \, : \, f(x) = s_0(x) + \sum_i s_i(x) g_i(x)\}$$

where the polynomials $s_i(x)$ are sums of squares, and of degree at most $2r$.

As discussed in section 2, the set $\mathcal{P}_r(K)$ can be characterized in terms of a semidefinite program of size polynomial in r. Therefore, we have the following sequence of increasingly tight semidefinite programs, as r grows:

$$\begin{aligned} \min: \quad & \sum_{|\boldsymbol{\alpha}| \leq k} \boldsymbol{y}_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}} \\ s.t.: \quad & (g(x) - 1) \in \mathcal{P}_r(S) \\ & g(x) \in \mathcal{P}_r(\Omega). \end{aligned}$$

Next we give some computational results to illustrate the added power of encorporating knowledge about the precise geometry of the problem. We give a one dimensional example, and a two dimensional example. The computations were performed using SeDuMi [11], and SOSTOOLS, [6].

Examples: The Value of Geometric Knowledge

Suppose the true distribution that provides the moments, is a uniform distribution on $[-1, 3]$, and we want to upper bound the probability that lies in the set $S = [0.75, 1.2]$. We want to explore the value of knowing that the support of the unknown distribution is on $[-1, 3]$, rather than simply

Moments	Assume support is $\mathbb{R}$	Use support $[-1,3]$
0	1.0000	1.0000
1	1.0000	0.6772
2	1.0000	0.6751
3	0.4640	0.4640
4	0.4639	0.4634
5	0.4635	0.3823
6	0.4635	0.3816

Table 5: Upper bounds on $P(X \in X)$, given moments of X. In the second column, we also encorporate information about the support of X. The resulting bounds are lower, hence tighter.

assuming it to be all of $\mathbb{R}$. Given the first 6 moments, we compute the following upper bounds: The support information improves the result by 20 to 30 percent in some cases. This is significant, especially in light of the fact that higher order moment constraints come at a significantly higher computational cost, than using additional lower order moment constraints. The above example shows that there are lower order moment constraints that can be fruitfully exploited before moving to higher order moments.

A Two Dimensional Example

Now consider a uniform distribution on the circle of radius $r = 2$, in the plane, $\mathbb{R}^2$. We want to find upper bounds on the probability that the random variable takes a value in the circle of radius $r' = 1/\sqrt{2}$, centered at $(1/2, 1/2)$. Again we compare the bounds obtained when we use information about the support of the distribution, and when we do not. In this case, the first column is the order of the highest order moment used. We see in this example, that in higher di-

Moments	Assume support is $\mathbb{R}^2$	Use support knowledge
0	1.0000	1.0000
1	1.0000	0.6666
2	1.0000	0.5099
3	0.4516	0.4164
4	0.4516	0.3592

Table 6: Upper bounds on $P(X \in S)$ given moments of X. X is uniformly distributed in the radius 2 circle, centered at the origin. S is the radius $1/\sqrt{2}$ circle centered at $(1/2, 1/2)$.

mensions, as in the PDE semidefinite formulation of the previous section, capturing the geometric information becomes increasingly important. We note that in the case of probability inequalities, geometry constraints are particularly important when using only lower order moments. This should not be surprising, given that if a distribution has all its moments, then it is in fact determined by its moments. The computations illustrate this fact. Also, the effect of the geometry may be more pronounced in some set ups, than in others. Nevertheless, we believe it to be substantial.

6 Conclusion

We have given two powerful uses of semidefinite optimization, the first to solving linear partial differential equations, and the second to obtaining optimal probability inequalities, from given moments. The key ingredient in both, is encorporating our knowledge of the underlying geometry. In the PDE application, as well as in the probability application, it is very often the case that we have information about the support of the distribution. We translate this into strengthened semidefinite constraints, using a representation theorem for positive polynomials, first proved by Schmüdgen in 1991, and then strengthened by Putinar in 1993. In both examples, the information the geometry provides, appears crucial for tightening the bounds, within reasonable complexity. Furthermore, we believe this framework to be quite general, and thus we expect it to have many more applications, where using the theorems of real algebraic geometry and the connection to semidefinite optimizaiton, to encorporate geometry into the optimization, can yield significant gains.

References

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

[2] Bertsimas, D.; Popscu, I. "Optimal Inequalities in Probability Theory: A Convex Optimization Approach."

[3] Caramanis, C. "Solving Linear Partial Differential Equations via Semidefinite Optimization," M.I.T. Master's Thesis, June 2001.

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

[5] Fujisawa, K., Kojima, M. and Nakata, K., SDPA (Semidefinite Programming Algorithm) User's Manual, Version 4.10, Research Report on Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1998.

[6] Prajna, S.; Papachristodoulou, A.; Parrilo, P. "SOSTOOLS: Optimization Toolbox for MATLAB," April 11, 2002, available from: http://www.aut.ee.ethz.ch/ parrilo/sostools.

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

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

[9] Reznick, R. "Some Concrete Aspects of Hilbert's 17^{th} Problem," available at: http://www.math.uiuc.edu/Reports/reznick/98-002.html, 1999.

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

[11] Sturm, J., SeDuMi Semidefinite Software, can be downloaded from: http://www/unimaas.nl/sturm.