OPTIMAL INEQUALITIES IN PROBABILITY THEORY: A CONVEX OPTIMIZATION APPROACH[*]

DIMITRIS BERTSIMAS[†] AND IOANA POPESCU[‡]

Abstract. We propose a semidefinite optimization approach to the problem of deriving tight moment inequalities for $P(X \in S)$, for a set S defined by polynomial inequalities and a random vector X defined on $\Omega \subseteq \mathcal{R}^n$ that has a given collection of up to kth-order moments. In the univariate case, we provide optimal bounds on $P(X \in S)$, when the first k moments of X are given, as the solution of a semidefinite optimization problem in $k + 1$ dimensions. In the multivariate case, if the sets S and Ω are given by polynomial inequalities, we obtain an improving sequence of bounds by solving semidefinite optimization problems of polynomial size in n, for fixed k.

We characterize the complexity of the problem of deriving tight moment inequalities. We show that it is NP-hard to find tight bounds for $k \geq 4$ and $\Omega = \mathcal{R}^n$ and for $k \geq 2$ and $\Omega = \mathcal{R}^n_+$, when the data in the problem is rational. For $k = 1$ and $\Omega = \mathcal{R}^n_+$ we show that we can find tight upper bounds by solving n convex optimization problems when the set S is convex, and we provide a polynomial time algorithm when S and Ω are unions of convex sets, over which linear functions can be optimized efficiently. For the case $k = 2$ and $\Omega = \mathcal{R}^n$, we present an efficient algorithm for finding tight bounds when S is a union of convex sets, over which convex quadratic functions can be optimized efficiently.

Key words. probability bounds, Chebyshev inequalities, semidefinite optimization, convex optimization

AMS subject classifications. 60E15, 90C22, 90C25

DOI. 10.1137/S1052623401399903

1. Introduction. The problem of deriving bounds on the probability that a certain random variable belongs in a given set, given information on some of its moments, has a very rich and interesting history, which is very much connected with the development of probability theory in the twentieth century. The inequalities due to Markov, Chebyshev, and Chernoff are some of the classical and widely used results of modern probability theory. Natural questions, however, that arise are the following:

1. *Are such bounds "best possible"; i.e., do there exist distributions that match them?*

2. *Can such bounds be generalized in multivariate settings, and in what circumstances can they be explicitly and/or algorithmically computed?*

3. *Is there a general theory based on optimization methods to address in a unified manner moment-inequality problems in probability theory?*

In this paper, we formulate the problem of obtaining best possible bounds as an optimization problem and use modern optimization theory, in particular convex and semidefinite programming, to give concrete answers to the questions above. We first introduce some notation. Let $\kappa = (k_1, \ldots, k_n)'$ with $k_j \in \mathcal{Z}_+$ nonnegative integers. We use the notation $\sigma_\kappa = \sigma_{k_1, \ldots, k_n}$ and $J_k = \{\kappa = (k_1, \ldots, k_n)' \mid k_1 + \cdots + k_n \leq k, \ k_j \in \mathcal{Z}_+, \ j = 1, \ldots, n\}$. We next introduce the notion of a *feasible moment sequence.*

DEFINITION 1.1. *A sequence $\overline{\sigma} = (\sigma_\kappa, \ \kappa \in J_k)$ is a feasible (n, k, Ω)-moment vector (or sequence), if there is a random vector $X = (X_1, \ldots, X_n)'$ defined on $\Omega \subseteq \mathcal{R}^n$ endowed with its Borel sigma algebra of events, whose moments are given by $\overline{\sigma}$, that is, $\sigma_\kappa = \sigma_{k_1, \ldots, k_n} = E[X_1^{k_1} \cdots X_n^{k_n}]$ for all $\kappa \in J_k$. We say that any such random variable X has a $\overline{\sigma}$-feasible distribution and denote this as $X \sim \overline{\sigma}$.*

Throughout the paper, the underlying probability space is implicitly assumed to be $\Omega \subseteq \mathcal{R}^n$ endowed with its Borel sigma algebra of events. We denote by $\mathcal{M} = \mathcal{M}(n, k, \Omega)$ the set of feasible (n, k, Ω)-moment vectors. For the univariate case ($n = 1$), the problem of deciding if $\overline{\sigma} = (M_1, M_2, \ldots, M_k)'$ is a feasible $(1, k, \Omega)$-moment vector is the classical moment problem. This problem has been completely characterized by necessary and sufficient conditions by Stieltjes [43], [44] in 1894-95, who adopts the "moment" terminology from mechanics (see also Karlin and Shapley [18], Akhiezer [1], Siu, Sengupta, and Lind [38], and Kemperman [21]). For univariate, nonnegative random variables ($\Omega = \mathcal{R}_+$), these conditions can be written as $R_k \succeq 0$ and $R_{k-1} \succeq 0$, where for any integer $l \geq 0$ we define

$$R_{2l} = \begin{pmatrix} 1 & M_1 & \ldots & M_l \\ M_1 & M_2 & \ldots & M_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ M_l & M_{l+1} & \ldots & M_{2l} \end{pmatrix}, \qquad R_{2l+1} = \begin{pmatrix} M_1 & M_2 & \ldots & M_{l+1} \\ M_2 & M_3 & \ldots & M_{l+2} \\ \vdots & \vdots & \ddots & \vdots \\ M_{l+1} & M_{l+2} & \ldots & M_{2l+1} \end{pmatrix}.$$

For univariate random variables with $\Omega = \mathcal{R}$, the necessary and sufficient condition (given by Hamburger [12], [13] in 1920-21) for a vector $\overline{\sigma} = (M_1, M_2, \ldots, M_k)'$ to be a feasible $(1, k, \mathcal{R})$-moment sequence is that $R_{2\lfloor \frac{k}{2} \rfloor} \succeq 0$. In the multivariate case, the formulation of the problem can be traced back to Haviland [14], [15] in 1935-36 (see also Godwin [10]). To date, the sufficiency part of the moment problem has not been completely resolved in the multivariate case, although substantial progress has been made in the last decade by Schmudgen [37] and Putinar [34].

Suppose that $\overline{\sigma}$ is a feasible moment sequence and X has a $\overline{\sigma}$-feasible distribution. We now define the central problem that this paper addresses.

The (n, k, Ω)-bound problem. Given a sequence $\sigma_\kappa, \kappa = (k_1, k_2, \ldots, k_n)' \in J_k$ of up to kth-order moments

$$\sigma_\kappa = E[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}], \quad \kappa \in J_k$$

of a random vector $X = (X_1, X_2, \ldots, X_n)'$ on $\Omega \subseteq \mathcal{R}^n$ endowed with its Borel sigma algebra, find the "best possible" or "tight" upper and lower bounds on $P(X \in S)$, for measurable events $S \subseteq \Omega$.

The term "best possible" or "tight" upper (and by analogy lower) bound above is defined as follows.

DEFINITION 1.2. *We say that γ is a tight upper bound on $P(X \in S)$ if $\gamma = \sup_{X \sim \overline{\sigma}} P(X \in S)$.*

Note that a bound can be tight without necessarily being *exactly achievable* (i.e., there is a random variable $X^{(0)} \sim \overline{\sigma}$ for which $P(X^{(0)} \in S) = \gamma$), but only *asymptotically* (i.e., there exists a sequence of random variables $\{X^{(i)}\}_{i \geq 1}$ such that $X^{(i)} \sim \overline{\sigma}$ and $\lim_{i \to \infty} P(X^{(i)} \in S) = \gamma$).

The well-known inequalities due to Markov, Chebyshev, and Chernoff, which are widely used if we know the first moment, the first two moments, and all moments (i.e., the generating function) of a random variable, respectively, are feasible but not necessarily *optimal* solutions to the (n, k, Ω)-bound problem; i.e., they are not necessarily tight bounds.

Literature and historical perspective. The history of the developments in the area of (n, k, Ω)-bound problems, sometimes referred to as generalized Chebyshev inequalities, can be traced back to the work of Gauss, Cauchy, Chebyshev, Markov, etc., and has witnessed an unexpected evolution. The problem of finding bounds on univariate distributions under moment constraints was proposed and formulated without proof initially by Chebyshev [6] in 1874 and resolved ten years later by his student Markov [25] in his Ph.D. thesis, using continued fractions techniques. In the 1950s and 1960s there was a revival of the interest in this area, that resulted in a large literature on Chebyshev systems and inequalities. Surveys of early literature can be found in Shohat and Tamarkin [40] and Godwin [9], [10]. A detailed, unified account of the evolution of Chebyshev systems is given by Karlin and Studden [19] in their 1966 monograph (see in particular Chapters 12 and 13 that deal with (n, k, Ω)-type bounds). In particular, they summarize the results of Marshall and Olkin [27], [28] who computed tight, explicit bounds on probabilities given first- and second-order moments (the $(n, 2, \Omega)$ problem in our context), thus generalizing Chebyshev's inequality to a multivariate setting.

Not for the first time in its history, and in the words of Shohat and Tamarkin in 1943 [40, p. 10], "the problem of moments lay dormant for more than 20 years." It revived briefly in the 1980s, with the book on probability inequalities and multivariate distributions of Tong [45] in 1980, who also published a monograph on probability inequalities in 1984. The latter notably contains, among others, a generalization of Markov's inequality for multivariate tails, due to Marshall [26]. A volume in *Moments in Mathematics* edited by Landau in 1987 includes a background survey by the same author [23], as well as relevant papers of Kemperman [22] and Diaconis [7].

The idea that optimization methods and duality theory can be used to address moment-type inequalities in probability first appeared in 1960, and is due independently and simultaneously to Isii [16] and Karlin (lecture notes at Stanford, see [19, p. 472]), who showed via strong duality results that certain types of Chebyshev inequalities for univariate random variables are sharp. Isii [17] extended these duality results for random vectors on complete regular spaces. Thirty two years after Isii's [17] original multivariate proof, Smith [42] rederived the duality result and proposed new interesting applications in decision analysis, dynamic programming, statistics, and finance. Shapiro [39] provided a rigorous discussion of necessary topological conditions for strong duality to hold, in that sense relaxing the compactness assumptions underlying Isii's proof. For an in-depth account of strong duality and sensitivity analysis for a general class of semi-infinite programming problems see Bonnans and Shapiro [3, Section 5.4].

For a broader investigation of the optimization framework underlying this type of problem, we refer the interested reader to Borwein and Lewis [4], [5] who provide an in-depth analysis of partially finite convex programming.

Despite its long and scattered history, the common belief among researchers is still that "the theory [of moment problems] is not up to the demands of applications" (Diaconis [7, p. 129]). The same author suggested that one of the reasons could be the high complexity of the problem: "numerical determination . . . is feasible for a small number of moments, but appears to be quite difficult in general cases." Another reason, as identified by Kemperman (see [22, p. 20]), is the lack of a general algorithmic approach:

> "... a deep study of algorithms has been rare so far in the theory
> of moments, except for certain very specific practical applications, for

> instance, to crystallography, chemistry and tomography. No doubt,
> there is a considerable need for developing reasonably good numer-
> ical procedures for handling the great variety of moment problems
> which do arise in pure and applied mathematics and in the sciences
> in general . . . ."

In an attempt to address Kemperman's criticism, Smith [42] actually introduced a computational procedure for the $(n, k, \mathcal{R}^n)$-bound problem, although he does not refer to it in this way. Unfortunately, the procedure is far from a formal algorithm, as there is no proof of convergence, and no investigation (theoretical or experimental) of its efficiency. It is fair to say that a thorough understanding of the algorithmic aspects and the complexity of the (n, k, Ω)-bound problem is still lacking.

Yet another criticism brought by Smith is the lack of simple, closed-form solutions for the $(n, k, \mathcal{R}^n)$-bound problem: "the bounds given by Chebyshev's inequalities . . . are quite loose. The more general versions are rarely used because of the lack of simple closed-form expressions for the bounds" (see [42, p. 808]).

Goals and contributions. The previous discussion motivates our desire in the present paper to understand the complexity of deriving tight moment inequalities, search for efficient algorithms in a general framework, and, when possible, derive simple closed-form bounds. In particular, we characterize which classes of (n, k, Ω)-bound problems are efficiently solvable and which are NP-hard.

Let us remark that the theory of polynomial solvability and NP-hardness assume a rational computational model. For this reason, as far as complexity results are concerned, we work under the assumption that the problem data is rational; i.e., moments are rational numbers and the sets Ω and S are specified in terms of rational numbers. Specifically, they are defined by semialgebraic sets (i.e., sets that are defined in terms of polynomial inequalities), whose parameters are rational numbers. In addition, throughout the paper we refer to an efficient, or polynomial time algorithm when it takes polynomial time in the problem data and $\log(1/\epsilon)$, and it computes a bound within ϵ of the tight bound for all $\epsilon > 0$.

More concretely, the contributions and structure of the paper are as follows.

1. In section 3, we investigate in detail the univariate case, i.e., the $(1, k, \Omega)$-bound problem for $\Omega = \mathcal{R}, \mathcal{R}_+$. We show that tight bounds can be computed efficiently by solving a single semidefinite optimization problem. We also derive tight bounds for tail probability events in closed form, when up to three moments are given. For $k = 1$, we recover the Markov inequality, which also shows that the Markov bound is tight. For $k = 2$, we recover a strict improvement of the Chebyshev inequality that retains the simplicity of the bound. This inequality dates back at least to Uspensky's book (see [46, p. 198]) from 1937, who proposed it as an exercise. Despite its simplicity, the bound has been strangely ignored in the recent literature and textbooks. For $k = 3$, we present closed-form tight bounds that appear to be new.

2. In section 4, we generalize the results of the previous section to the multivariate (n, k, Ω)-bound problem. For a fairly general class of semialgebraic sets S and Ω (not necessarily convex), we propose a sequence of increasingly stronger, asymptotically exact upper bounds by solving semidefinite optimization problems of polynomial size in n. This includes the case when Ω is a bounded polyhedron, or a semialgebraic set such that $\Omega \subseteq \{x \in \mathcal{R}^n \mid x'x \leq M^2\}$ and M is known a priori. The proposed approach gives rise to a family of semidefinite relaxations whose size is polynomial in n. We also show that the $(n, k, \mathcal{R}^n_+)$, $(n, k, \mathcal{R}^n)$-bound problems for $k \geq 2$, respectively $k \geq 4$, are NP-hard.

	$n = 1$	n	
		S convex set	S union of convex sets
$k = 1$	Theorem 3.3	Theorem 5.1	Algorithm A, section 5.2
	Markov [25]	Marshall [26], BP	BP
$k = 2$	Theorem 3.3	Theorem 6.1	Algorithm B, section 6.2
	Chebyshev [6]	Marshall and Olkin [27]	BP
k	Theorem 3.2: SDP	Theorem 4.3: SDP; Propositions 4.5, 4.6: NP-hardness	
	BP	BP	

3. In section 5, we address the $(n, 1, \Omega)$ problem. We show that if the set S in the definition of the $(n, 1, \mathcal{R}^n)$-bound problem is convex, we find best possible bounds for $P(X \in S)$ explicitly as a solution of n convex optimization problems. We also provide a polynomial time algorithm to solve the $(n, 1, \Omega)$ problem, when the sets S and Ω are the union of disjoint convex sets, over which a linear function can be efficiently optimized. These bounds represent natural extensions of the inequalities due to Markov in the univariate case, and Marshall [26] for multivariate tails.

4. In section 6, we first review the work of Marshall and Olkin [27] who showed that the $(n, 2, \mathcal{R}^n)$-bound problem can be solved as a single convex optimization problem. We provide an efficient algorithm for the case when S is a union of disjoint convex sets, over which a convex quadratic function can be optimized efficiently.

Table 1 summarizes the contributions of the paper to the (n, k, Ω)-bound problem.

2. Primal and dual formulations of the (n, k, Ω)-bound problem. In this section, we formulate the (n, k, Ω)-upper bound problem as an optimization problem, where $\Omega \in \mathcal{R}^n$ is the domain of the random variables under consideration. We examine the corresponding dual problem and present weak and strong duality results that permit us to develop algorithms for the problem. The same approach and results apply to the (n, k, Ω)-lower bound problem.

We use the notation $\bar{z}^\kappa = z_1^{k_1} \cdots z_n^{k_n}$, where $\bar{z} = (z_1, \ldots, z_n)'$ and $\kappa = (k_1, \ldots, k_n)'$, $k_j \in \mathcal{Z}_+$. Recall that $J_k = \{\kappa = (k_1, \ldots, k_n)' \mid k_1 + \cdots + k_n \leq k, \; k_j \in \mathcal{Z}_+, \; j = 1, \ldots, n\}$. The (n, k, Ω)-upper bound problem can be formulated as the following optimization problem:

$$(2.1) \qquad \begin{aligned} Z_P = \max_{\mu} \quad & \int_S \mathbf{1} \, d\mu \\ \text{subject to} \quad & \int_\Omega \bar{z}^\kappa d\mu = \sigma_\kappa \quad \forall \kappa \in J_k. \end{aligned}$$

The variable of the infinite-dimensional optimization problem (2.1) is the probability measure μ (so implicitly $d\mu(\bar{z}) \geq 0$ a.e.). We assume that $\sigma_{0,\ldots,0} = 1$, corresponding to the probability mass constraint. If problem (2.1) is feasible, then $\bar{\sigma}$ is a feasible moment sequence, and any feasible distribution μ is a $\bar{\sigma}$-feasible distribution. The feasibility problem is exactly the classical multidimensional moment problem.

In the spirit of linear programming duality theory, we associate a dual variable $y_\kappa = y_{k_1,\ldots,k_n}$ with each equality constraint of the primal and we obtain

$$(2.2) \qquad \begin{aligned} Z_D = \min_{y \in \mathcal{R}^{|J_k|}} \quad & \sum_{\kappa \in J_k} y_\kappa \sigma_\kappa \\ \text{subject to} \quad & g(\bar{z}) = \sum_{\kappa \in J_k} y_\kappa \bar{z}^\kappa \geq 1 \quad \forall \bar{z} \in S, \end{aligned}$$

$$g(\bar{z}) = \sum_{\kappa \in J_k} y_\kappa \bar{z}^\kappa \geq 0 \forall \bar{z} \in \Omega.$$

While in principle the dual constraints may only need to hold a.e., this is equivalent to solving (2.2) since polynomial positivity holds a.e. if and only if it holds everywhere.

In general, the optimum in problem (2.1) may not be achievable. Whenever the primal optimum is achieved, we call the corresponding distribution *an extremal distribution*. In the case when only (say upper) bounds on moments σ_κ are known, then in the dual problem we add the inequalities $y_\kappa \geq 0$. Thus, our results (specifically Theorems 3.2 and 4.3 below) regarding efficient solvability of the underlying problem generalize to the case when only bounds on moments are known. We next establish weak duality.

THEOREM 2.1 (weak duality). $Z_P \leq Z_D$.

Proof. Let μ be a primal feasible solution and let y_κ, $\kappa \in J_k$, be a dual feasible solution. From dual feasibility we have that for all $\bar{z} \in \Omega$, $g(\bar{z}) = \sum_{\kappa \in J_k} y_\kappa \bar{z}^\kappa \geq \chi_S(\bar{z})$, where $\chi_S(\bar{z}) = 1$ if $\bar{z} \in S$, and 0, otherwise. Then

$$Z_P = \int_S \mathbf{1}\, d\mu = \int_\Omega \chi_S(\bar{z})\, d\mu \leq \int_\Omega g(\bar{z})\, d\mu = \sum_{\kappa \in J_k} y_\kappa \int_\Omega \bar{z}^\kappa d\mu = \sum_{\kappa \in J_k} y_\kappa \sigma_\kappa = Z_D. \qquad \square$$

Theorem 2.1 indicates that by solving the dual problem (2.2) we obtain an upper bound on the primal objective, and hence on the probability we are trying to estimate. If the moment vector $\bar{\sigma}$ is an interior point of the set of feasible moment vectors, the dual bound turns out to be tight. This strong duality result follows from a univariate result due to Karlin and Isii in 1960 (see Karlin and Studden [19, p. 472]), and generalized by Isii in 1963 [17] for random vectors defined on complete regular spaces. Shapiro [39, Proposition 3.4] provides a modern proof based on conic linear duality, under a general topological structure. The following theorem is a consequence of their work, and holds for general distributions of X (discrete, continuous, with or without atoms, etc.) over a metric space Ω endowed with its Borel sigma algebra.

THEOREM 2.2 (strong duality). *If the moment vector $\bar{\sigma}$ is an interior point of the set $\mathcal{M}$ of feasible moment vectors, then $Z_P = Z_D$.*

If the dual is unbounded, it is immediate from weak duality that the multidimensional moment problem is infeasible. On the other hand, if the common optimal value is finite, then the set of dual optimal solutions is nonempty and bounded (see Isii [17], Kemperman [20], or Shapiro [39]). Furthermore, if $\bar{\sigma}$ is a boundary point of $\mathcal{M}$, then it can be shown that the $\bar{\sigma}$-feasible distributions are concentrated on a subset Ω_0 of Ω, and strong duality holds provided we relax the dual to Ω_0 (see Isii [17, p. 190] or Smith [42, p. 824]). These authors also prove that it is equivalent to optimize only over discrete distributions that are concentrated on $m+2$ points, where m is the number of moment constraints.

In the univariate case, Isii [16] proves that if $\bar{\sigma}$ is a boundary point of $\mathcal{M}$, then exactly one $\bar{\sigma}$-feasible distribution exists. Kemperman [20] proves that for almost every $\bar{\sigma}$ (with respect to the Lebesgue measure) such that the dual is finite, the dual solution is unique (see also Shapiro [39, Proposition 3.5]). In Chapter 5.4 of their recent book, Bonnans and Shapiro [3] provide necessary and sufficient conditions for the uniqueness of the optimal distribution, as well as sensitivity results.

If strong duality holds, then by optimizing over problem (2.2) we obtain a *tight* bound on $P(X \in S)$. On the other hand, it is worthwhile to remark that under

certain technical conditions (see Grötschel, Lovász, and Schrijver [11]) solving problem (2.2) is equivalent to solving the corresponding separation problem, which amounts to verifying polynomial positivity conditions over S and Ω.

3. The $(1, k, \Omega)$-bound problem. In this section, we restrict our attention to univariate random variables. Given the first k moments $M_1, \ldots, M_k$ (we let $M_0 = 1$) of a real random variable X with domain $\Omega \subseteq \mathcal{R}$, we are interested in deriving tight bounds on $P(X \in S)$. This is the $(1, k, \Omega)$-bound problem. Our main result in this section is that tight bounds can be derived as a solution to a single semidefinite optimization problem. We also derive closed-form tight bounds when up to the first three moments are given.

3.1. Tight bounds as semidefinite optimization problems. In the one-dimensional case, problem (2.2) becomes

$$
\text{minimize} \quad \sum_{r=0}^{k} y_r M_r
$$

(3.1)
$$
\text{subject to} \quad \sum_{r=0}^{k} y_r x^r \geq 1 \quad \forall x \in S,
$$

$$
\sum_{r=0}^{k} y_r x^r \geq 0 \quad \forall x \in \Omega.
$$

Problem (3.1) naturally leads us to investigate conditions for polynomials to be non-negative. When S and Ω are intervals on the real line, we show in the next proposition that the feasible region of problem (3.1) can be expressed using semidefinite constraints. Semidefinite optimization problems are efficiently solvable using interior point methods. For a review of semidefinite optimization see Wolkowicz, Saigal, and Vandenberghe [47]. The results and the proofs in the following proposition are inspired by Nesterov [31] (see also Ben-Tal and Nemirovski [2, pp. 140–142]).

PROPOSITION 3.1. (a) *The polynomial* $g(x) = \sum_{r=0}^{2k} y_r x^r$ *satisfies* $g(x) \geq 0$ *for all* $x \in \mathcal{R}$ *if and only if there exists a positive semidefinite matrix* $X = [x_{ij}]_{i,j=0,\ldots,k}$, *such that*

(3.2)
$$
y_r = \sum_{i,j:i+j=r} x_{ij}, \quad r = 0, \ldots, 2k, \qquad X \succeq 0.
$$

(b) *The polynomial* $g(x) = \sum_{r=0}^{k} y_r x^r$ *satisfies* $g(x) \geq 0$ *for all* $x \geq 0$ *if and only if there exists a positive semidefinite matrix* $X = [x_{ij}]_{i,j=0,\ldots,k}$, *such that*

$$
0 = \sum_{i,j:i+j=2l-1} x_{ij}, \quad l = 1, \ldots, k,
$$

(3.3)
$$
y_l = \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,
$$

$$
X \succeq 0.
$$

(c) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [0, a]$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\ldots,k}$, such that*

$$0 = \sum_{i,j:i+j=2l-1} x_{ij}, \quad l = 1, \ldots, k,$$

(3.4)
$$\sum_{r=0}^{l} y_r \binom{k-r}{l-r} a^r = \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,$$

$$X \succeq 0.$$

(d) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [a, \infty)$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\ldots,k}$, such that*

$$0 = \sum_{i,j:i+j=2l-1} x_{ij}, \quad l = 1, \ldots, k,$$

(3.5)
$$\sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,$$

$$X \succeq 0.$$

(e) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in (-\infty, a]$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\ldots,k}$, such that*

$$0 = \sum_{i,j:i+j=2l-1} x_{ij}, \quad l = 1, \ldots, k,$$

(3.6)
$$(-1)^l \sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,$$

$$X \succeq 0.$$

(f) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [a, b]$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\ldots,k}$, such that*

$$0 = \sum_{i,j:i+j=2l-1} x_{ij}, \quad l = 1, \ldots, k,$$

(3.7)
$$\sum_{m=0}^{l} \sum_{r=m}^{k+m-l} y_r \binom{r}{m} \binom{k-r}{l-m} a^{r-m} b^m = \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,$$

$$X \succeq 0.$$

Proof.
(a) Suppose (3.2) holds. Let $x_{(k)} = (1, x, x^2, \ldots, x^k)'$. Then

$$g(x) = \sum_{r=0}^{2k} \sum_{i+j=r} x_{ij} x^r$$
$$= \sum_{i=0}^{k} \sum_{j=0}^{k} x_{ij} x^i x^j$$
$$= x'_{(k)} X x_{(k)}$$
$$\geq 0,$$

since $X \succeq 0$.

Conversely, suppose that the polynomial $g(x)$ of degree $2k$ is nonnegative for all x. Then, the real roots of $g(x)$ should have even multiplicity, otherwise $g(x)$ would alter its sign in a neighborhood of a root. Let λ_i, $i = 1, \ldots, r$, be its real roots with corresponding multiplicity $2m_i$. Its complex roots can be arranged in conjugate pairs, $a_j + ib_j$, $a_j - ib_j$, $j = 1 \ldots, h$. Then,

$$g(x) = y_{2k} \prod_{i=1}^{r} (x - \lambda_i)^{2m_i} \prod_{j=1}^{h} ((x - a_j)^2 + b_j^2).$$

The leading coefficient y_{2k} needs to be positive. Thus, by expanding the terms in the products, we see (cf. Proposition 4.2 below) that $g(x)$ can be written as a sum of squares of polynomials of the form

$$
\begin{aligned}
g(x) &= \sum_{i=1}^{p} \left(\sum_{j=0}^{k} q_{ij} x^j \right)^2 \\
&= x'_{(k)} Q' Q x_{(k)} = x'_{(k)} X x_{(k)},
\end{aligned}
$$

with $X = Q'Q$ positive semidefinite, from which (3.2) follows.

(b) We observe that $g(x) \geq 0$ for $x \geq 0$ if and only if $g(t^2) \geq 0$ for all t. Since

$$g(t^2) = y_0 + 0 \cdot t + y_1 t^2 + 0 \cdot t^3 + y_2 t^4 + \cdots + y_k t^{2k},$$

we obtain (3.3) by applying part (a).

(c) We observe that $g(x) \geq 0$ for $x \in [0, a]$ if and only if

$$(1 + t^2)^k g \left(\frac{at^2}{1 + t^2} \right) \geq 0 \quad \forall t.$$

Since

$$
\begin{aligned}
(1 + t^2)^k g \left(\frac{at^2}{1 + t^2} \right) &= \sum_{r=0}^{k} y_r a^r t^{2r} (1 + t^2)^{k-r} \\
&= \sum_{r=0}^{k} y_r a^r \sum_{l=0}^{k-r} \binom{k-r}{l} t^{2(l+r)} \\
&= \sum_{j=0}^{k} t^{2j} \left(\sum_{r=0}^{j} y_r \binom{k-r}{j-r} a^r \right),
\end{aligned}
$$

by applying part (a) we obtain (3.4).

(d) We observe that $g(x) \geq 0$ for $x \in [a, \infty)$ if and only if

$$g(a + t^2) \geq 0 \quad \forall t.$$

Since

$$
\begin{aligned}
g(a + t^2) &= \sum_{r=0}^{k} y_r (a + t^2)^r \\
&= \sum_{r=0}^{k} y_r \sum_{l=0}^{r} \binom{r}{l} a^{r-l} t^{2l} \\
&= \sum_{l=0}^{k} t^{2l} \left(\sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} \right),
\end{aligned}
$$

by applying part (a) we obtain (3.5).

(e) We observe that $g(x) \geq 0$ for $x \in (-\infty, a]$ if and only if

$$g(a - t^2) \geq 0 \quad \forall t.$$

Since

$$
\begin{aligned}
g(a - t^2) &= \sum_{r=0}^{k} y_r (a - t^2)^r \\
&= \sum_{r=0}^{k} y_r \sum_{l=0}^{r} \binom{r}{l} a^{r-l} (-t^2)^l \\
&= \sum_{l=0}^{k} t^{2l} \left((-1)^l \sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} \right),
\end{aligned}
$$

by applying part (a) we obtain (3.6).

(f) We observe that $g(x) \geq 0$ for $x \in [a, b]$ if and only if

$$(1 + t^2)^k g\left(a + (b - a) \frac{t^2}{1 + t^2} \right) \geq 0 \quad \forall t.$$

Since

$$
\begin{aligned}
(1 + t^2)^k g\left(a + (b - a) \frac{t^2}{1 + t^2} \right) &= \sum_{r=0}^{k} y_r (a + bt^2)^r (1 + t^2)^{k-r} \\
&= \sum_{r=0}^{k} y_r \sum_{m=0}^{r} \binom{r}{m} a^{r-m} b^m t^{2m} \sum_{j=0}^{k-r} \binom{k-r}{j} t^{2j} \\
&= \sum_{l=0}^{k} t^{2l} \left(\sum_{m=0}^{l} \sum_{r=m}^{k+m-l} y_r \binom{r}{m} \binom{k-r}{l-m} a^{r-m} b^m \right),
\end{aligned}
$$

by applying part (a) we obtain (3.7).   □

We next show that problem (3.1) can be written as a semidefinite optimization problem. Corresponding formulations, which we omit here for the sake of conciseness, can be obtained similarly for the problem of establishing lower bounds, i.e., problem (2.1) in which we replace maximization by minimization.

THEOREM 3.2. *Given the first k moments $(M_1, \ldots, M_k)$ (we let $M_0 = 1$) of a random variable X defined on Ω we obtain the following tight upper bounds.*

(a) *If $\Omega = \mathcal{R}^+$, the tight upper bound on $P(X \geq a)$ is given as the solution of the*

semidefinite optimization problem

$$\text{minimize} \quad \sum_{r=0}^{k} y_r M_r$$

$$\text{subject to} \quad 0 = \sum_{i,j:i+j=2l-1} x_{ij}, \qquad\qquad l = 1, \ldots, k,$$

$$(y_0 - 1) + \sum_{r=1}^{k} y_r a^r = x_{00},$$

(3.8)

$$\sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} x_{ij}, \qquad l = 1, \ldots, k,$$

$$0 = \sum_{i,j:i+j=2l-1} z_{ij}, \qquad\qquad l = 1, \ldots, k,$$

$$\sum_{r=0}^{l} y_r \binom{k-r}{l-r} a^r = \sum_{i,j:i+j=2l} z_{ij}, \qquad l = 0, \ldots, k,$$

$$X, Z \succeq 0.$$

If $\Omega = \mathcal{R}$, then the last equation in (3.8) should be replaced by

$$(-1)^l \sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} z_{ij}, \quad l = 0, \ldots, k.$$

(b) *If $\Omega = \mathcal{R}^+$, the tight upper bound on $P(a \leq X \leq b)$ is given as the solution of the semidefinite optimization problem*

(3.9)

$$\text{minimize} \sum_{r=0}^{k} y_r M_r$$

$$\text{subject to } 0 = \sum_{i,j:i+j=2l-1} x_{ij}, \qquad\qquad\qquad l = 1, \ldots, k,$$

$$\sum_{m=0}^{l} \sum_{r=m}^{k+m-l} y_r \binom{r}{m}\binom{k-r}{l-m} a^{r-m} b^m = \binom{k}{l} + \sum_{i,j:i+j=2l} x_{ij}, \quad l = 0, \ldots, k,$$

$$0 = \sum_{i,j:i+j=2l-1} z_{ij}, \qquad\qquad\qquad l = 1, \ldots, k,$$

$$y_l = \sum_{i,j:i+j=2l} z_{ij}, \qquad\qquad\qquad l = 0, \ldots, k,$$

$$X, Z \succeq 0.$$

If $\Omega = \mathcal{R}$, then the last equation in (3.9) should be replaced by

$$(-1)^l \sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} z_{ij}, \quad l = 0, \ldots, k,$$

and the following equations added:

$$0 = \sum_{i,j:i+j=2l-1} u_{ij}, \quad l = 1, \ldots, k,$$

$$\sum_{r=l}^{k} y_r \binom{r}{l} b^{r-l} = \sum_{i,j:i+j=2l} u_{ij}, \quad l = 0, \ldots, k,$$

$$U \succeq 0.$$

Proof.

(a) The feasible region of problem (3.1) for $S = [a, \infty)$ and $\Omega = \mathcal{R}_+$ becomes

$$g(x) = \sum_{r=0}^{k} y_r x^r \geq 1 \quad \forall x \in [a, \infty) \quad \text{and} \quad g(x) \geq 0 \quad \forall x \in [0, a).$$

By applying Proposition 3.1(c), (d) we obtain (3.8). If $\Omega = \mathcal{R}$, we apply Proposition 3.1(d), (e).

(b) The feasible region of problem (3.1) for $S = [a, b]$ and $\Omega = \mathcal{R}_+$ becomes

$$g(x) = \sum_{r=0}^{k} y_r x^r \geq 1 \quad \forall x \in [a, b] \quad \text{and} \quad g(x) \geq 0 \quad \forall x \in [0, \infty).$$

By applying Proposition 3.1(b), (f) we obtain (3.9). If $\Omega = \mathcal{R}$, we apply Proposition 3.1(d), (e), (f). □

3.2. Closed-form bounds. In this section, we present closed-form bounds when moments up to third order are given. We define the squared coefficient of variation $C_M^2 = \frac{M_2 - M_1^2}{M_1^2}$, and the third-order coefficient of variation $D_M^2 = \frac{M_1 M_3 - M_2^2}{M_1^4}$. Let $\delta > 0$.

THEOREM 3.3. *The following bounds in Table 2 are tight for $k = 1, 2, 3$.*

The bounds marked with an asterisk (*) assume that $\delta < 1$ (the other case is trivial). The following definitions are used:

$$f_1(C_M^2, D_M^2, \delta) = \begin{cases} \min\left(\dfrac{C_M^2}{C_M^2 + \delta^2}, \dfrac{1}{1+\delta} \cdot \dfrac{D_M^2}{D_M^2 + (C_M^2 - \delta)^2}\right), & \text{if } \delta > C_M^2, \\[2ex] \dfrac{1}{1+\delta} \cdot \dfrac{D_M^2 + (1+\delta)(C_M^2 - \delta)}{D_M^2 + (1+C_M^2)(C_M^2 - \delta)}, & \text{if } \delta \leq C_M^2, \end{cases}$$

$$f_2(C_M^2, D_M^2, \delta) = 1 - \dfrac{(C_M^2 + \delta)^3}{(D_M^2 + (C_M^2 + 1)(C_M^2 + \delta))(D_M^2 + (C_M^2 + \delta)^2)},$$

$$f_3(C_M^2, D_M^2, \delta) = \min\left(1, 1 + 3^3 \dfrac{D_M^2 + C_M^4 - \delta^2}{4 + 3(1 + 3\delta^2) + 2(1 + 3\delta^2)^{\frac{3}{2}}}\right).$$

The proof of the theorem is given in Popescu [33].

4. The (n, k, Ω)-bound problem: Semidefinite formulations and complexity. In this section, we generalize the results of the previous section to the multivariate case by investigating the (n, k, Ω)-bound problem. For a fairly general class

TABLE 2
Tight bounds for the $(1, k, \Omega)$-problem for $k \leq 3$.

(k, Ω)	$P(X > (1 + \delta)M_1)$	$P(X < (1 - \delta)M_1)$	$P(|X - M_1| > \delta M_1)$
$(1, \mathcal{R}_+)$	$\frac{1}{1+\delta}$	1^*	1^*
$(2, \mathcal{R})$	$\frac{C_M^2}{C_M^2 + \delta^2}$	$\frac{C_M^2}{C_M^2 + \delta^2}$	$\min\left(1, \frac{C_M^2}{\delta^2}\right)$
$(3, \mathcal{R}_+)$	$f_1(C_M^2, D_M^2, \delta)$	$f_2(C_M^2, D_M^2, \delta)^*$	$f_3(C_M^2, D_M^2, \delta)^*$

of semialgebraic sets S and Ω, i.e., sets that are given as intersections of inequalities involving polynomials, we propose a sequence of increasingly stronger, asymptotically exact upper bounds by solving semidefinite optimization problems. This includes the case when Ω is a bounded polyhedron, or a semialgebraic set such that $\Omega \subseteq \{x \in \mathcal{R}^n \mid x'x \leq M^2\}$ and M is known a priori. Note that it is not necessary for the sets S and Ω to be convex. We expect that in general an exact semidefinite formulation, if it exists, may be exponential in n even for fixed k. This fact should not be surprising, given that in section 4.2 we prove that it is NP-hard to find best possible bounds for the $(n, k, \mathcal{R}^n)$-bound problem for $k \geq 4$. We also show that it is NP-hard to find best possible bounds for the $(n, k, \mathcal{R}_+^n)$-bound problem for $k \geq 2$.

4.1. Semidefinite programming formulations. The approach we follow in this section has its origin in the work of Shor [41]. Recently it has been used by Lasserre [24] and Parrilo [32] to provide semidefinite relaxations for discrete optimization and nonconvex optimization problems.

In dimension $n = 1$, we have seen in the proof of Proposition 3.1 that a polynomial in one dimension is nonnegative if and only if it can be written as a sum of squares of polynomials. Clearly, in multiple dimensions if a polynomial can be written as a sum of squares of other polynomials, then it is nonnegative. Hilbert, in a nonconstructive way, has shown that it is possible for a polynomial in higher dimensions to be nonnegative without being a sum of squares. Motzkin (see Reznick [35]) has shown that the polynomial

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

is nonnegative without being a sum of squares of polynomials. The connection between nonnegative polynomials and sum of squares representations has a long history that is nicely outlined in Reznick [35]. In this paper we will rely on the following theorem due to Putinar [34].

THEOREM 4.1 (Putinar [34]). *Suppose that the set*

$$K = \{x \in \mathcal{R}^n \mid g_i(x) \geq 0, \ i \in I\}$$

is compact and there exists a polynomial $h(x)$ of the form

$$h(x) = h_0(x) + \sum_{i \in I} h_i(x) g_i(x),$$

such that $\{x \in \mathcal{R}^n \mid h(x) \geq 0\}$ is compact and $h_i(x)$, $i \in I \cup \{0\}$ are polynomials that have a sum of squares representation. Then, for any polynomial $g(x)$ that is strictly

positive for all $x \in K$, there exist $p_i(x)$, $i \in I \cup \{0\}$, that are sums of squares such that

$$(4.1) \qquad g(x) = p_0(x) + \sum_{i \in I} p_i(x) g_i(x).$$

Putinar [34] shows that the conditions for Theorem 4.1 to hold are relatively mild. In particular, examples of sets K that satisfy the conditions of Theorem 4.1 include the following.

(a) One of the defining inequalities of K represents a compact set $\{x \in \mathcal{R}^n \mid g_l(x) \geq 0\}$. This is the case, for example, when $g_l(x) = c - (x - x_0)'Q(x - x_0)$, and $Q \succ 0$.

(b) K is a bounded polyhedron.

(c) K is compact, and there is a bound M a priori known such that $K \subseteq \{x \in \mathcal{R}^n \mid x'x \leq M^2\}$.

While Theorem 4.1 guarantees the existence of the representation in (4.1), it does not give any information on the degree of the polynomials $p_i(x)$. From Theorem 4.1 it is natural to investigate when a polynomial is a sum of squares. Let $\mathrm{x}_{(d)}$ be the vector of all monomials in the variables $x_1, \ldots, x_n$ of degree d and below. For example, for $n = 2$, $d = 2$,

$$\mathrm{x}_{(2)} = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)'.$$

There are $a = \binom{n+d}{d}$ such monomials.

PROPOSITION 4.2. *The polynomial $f(x)$ of degree $2d$ has a sum of squares decomposition if and only if there exists a positive semidefinite matrix Q for which $f(x) = \mathrm{x}'_{(d)} Q \mathrm{x}_{(d)}$.*

Proof. To simplify notation, let $\mathrm{x} = \mathrm{x}_{(d)}$ throughout the proof. An arbitrary polynomial of degree $2d$ can be written as $f(x) = \mathrm{x}'Q\mathrm{x}$ for some $a \times a$ matrix Q (the matrix Q is not unique, however). If $Q \succeq 0$, then $Q = HH'$ for some H, and thus,

$$f(x) = \mathrm{x}'HH'\mathrm{x} = \sum_{i=1}^{a} (H'\mathrm{x})_i^2.$$

Since $(H'\mathrm{x})_i$ is a polynomial, we have expressed $f(x)$ as a sum of squares of the polynomials $(H'\mathrm{x})_i$.

Conversely, suppose that $f(x)$ has a sum of squares decomposition

$$f(x) = \sum_{i=1}^{l} h_i(x)^2 = \sum_{i=1}^{l} (h_i'\mathrm{x})^2 = \sum_{i=1}^{l} \mathrm{x}'(h_i h_i')\mathrm{x} = \mathrm{x}' \left(\sum_{i=1}^{l} (h_i h_i') \right) \mathrm{x} = \mathrm{x}'Q\mathrm{x},$$

where h_i is the vector of coefficients of the polynomial $h_i(x)$. Since $Q = \sum_{i=1}^{l} (h_i h_i') \succeq 0$, the proposition follows. $\square$

Therefore, if the sets Ω and S satisfy the conditions of Theorem 4.1, by applying Proposition 4.2, we can express problem (2.2) as a semidefinite optimization problem. Suppose that the sets Ω and S are semialgebraic sets, given by

$$\Omega = \left\{ x \in \mathcal{R}^n \mid \omega_i(x) = \sum_{\kappa \in J_l} \omega_\kappa^i x^\kappa \geq 0, \ i = 1, \ldots, r \right\},$$

$$S = \left\{ x \in \mathcal{R}^n \mid s_i(x) = \sum_{\kappa \in J_t} s_\kappa^i x^\kappa \geq 0, \ i = 1, \ldots, m \right\},$$

where ω_κ^i, $s_\kappa^i \in \mathcal{R}$, that is Ω and S are defined by polynomial inequalities. We use the notation $\delta_{\kappa,0} = 1$ if $\kappa = 0$, and zero, otherwise.

THEOREM 4.3. *If the sets Ω and S satisfy the conditions of Theorem 4.1, then for every $\epsilon > 0$ there exists a nonnegative integer $d \in \mathcal{Z}_+$ (representing the degree of the polynomials in (4.1)), such that the objective function value Z_D in problem (2.2) satisfies $|Z_D - Z_D(d)| \le \epsilon$, where $Z_D(d)$ is the value of the following semidefinite program:*

(4.2)
$$Z_D(d) = \min \quad \sum_{\kappa \in J_k} y_\kappa \sigma_\kappa$$

$$\text{subject to} \quad y_\kappa - \delta_{\kappa,0} = q_\kappa^0 + \sum_{i=1}^m \sum_{\substack{\eta \in J_d,\, \theta \in J_t \\ \eta + \theta = \kappa}} q_\eta^i s_\theta^i \quad \forall \kappa \in J_k,$$

$$0 = q_\kappa^0 + \sum_{i=1}^m \sum_{\substack{\eta \in J_d,\, \theta \in J_t \\ \eta + \theta = \kappa}} q_\eta^i s_\theta^i \qquad \forall \kappa \in J_{s+d} \setminus J_k,$$

$$y_\kappa = p_\kappa^0 + \sum_{i=1}^r \sum_{\substack{\eta \in J_d,\, \theta \in J_l \\ \eta + \theta = \kappa}} p_\eta^i \omega_\theta^i \qquad \forall \kappa \in J_k,$$

$$0 = p_\kappa^0 + \sum_{i=1}^r \sum_{\substack{\eta \in J_d,\, \theta \in J_l \\ \eta + \theta = \kappa}} p_\eta^i \omega_\theta^i \qquad \forall \kappa \in J_{l+d} \setminus J_k,$$

$$q_\kappa^i = \sum_{\eta,\theta \in J_d,\, \eta+\theta=\kappa} q_{\eta,\theta}^i \qquad \forall \kappa \in J_d,\ i = 0,1,\ldots,m,$$

$$p_\kappa^i = \sum_{\eta,\theta \in J_d,\, \eta+\theta=\kappa} p_{\eta,\theta}^i \qquad \forall \kappa \in J_d,\ i = 0,1,\ldots,m,$$

$$Q^i = [q_{\eta,\theta}^i]_{\eta,\theta \in J_d} \succeq 0, \qquad i = 0,1,\ldots,r,$$

$$P^i = [p_{\eta,\theta}^i]_{\eta,\theta \in J_d} \succeq 0, \qquad i = 0,1,\ldots,m.$$

Proof. We first remark that the value of the dual problem (2.2) equals that of the following strict inequality formulation:

(4.3)
$$Z_D = \inf_{y \in \mathcal{R}^{|J_k|}} \sum_{\kappa \in J_k} y_\kappa \sigma_\kappa$$
$$\text{subject to} \quad g(x) = \sum_{\kappa \in J_k} y_\kappa x^\kappa > 1 \quad \forall x \in S,$$
$$g(x) = \sum_{\kappa \in J_k} y_\kappa x^\kappa > 0 \quad \forall x \in \Omega.$$

This problem may not admit an optimal solution. However, for any $\epsilon > 0$ there exists a feasible polynomial g_ϵ resulting in an objective value that is less than $Z_D + \epsilon$. Fix $\epsilon > 0$ and let $g = g_\epsilon$. For this particular g, from Theorem 4.1 and Proposition 4.2, the above feasibility constraint $g(x) > 0$ for all $x \in \Omega$ is equivalent to

$$g(x) = p_0(x) + \sum_{i=1}^m p_i(x)\omega_i(x),$$

where $p_i(x) = \mathrm{x}'P^i\mathrm{x}$, $P^i \succeq 0$, and x includes monomials up to a certain degree d_0. Writing

$$p_i(x) = \sum_{\eta \in J_d} p^i_\eta x^\eta, \qquad \omega_i(x) = \sum_{\theta \in J_l} \omega^i_\theta x^\theta,$$

we have that $g(x) > 0$ for $x \in \Omega$ if and only if

$$g(x) := \sum_{\kappa \in I} y_\kappa x^\kappa = \sum_{\eta \in J_d} p^0_\eta x^\eta + \sum_{i=1}^r \left(\sum_{\eta \in J_d} p^i_\eta x^\eta \right) \left(\sum_{\theta \in J_l} \omega^i_\theta x^\theta \right).$$

Equating terms, we obtain the third and fourth sets of linear constraints in problem (4.2), corresponding to the degree d_0.

Similarly, one can translate the feasibility constraint $g(x) - 1 > 0$ for all $x \in S$ into the first two sets of constraints in problem (4.2), corresponding to a certain degree d_1. It follows that the vector of coefficients y of the polynomial g is feasible for problem (4.2) with degree $d = \max(d_0, d_1)$, so $Z_D(d) \leq Z_D + \epsilon$.

On the other hand, for any d, problem (4.2) is a restriction of problem (4.3) for feasible polynomials of degree d, so $Z_D(d) \geq Z_D$, and the desired result follows. $\square$

Theorem 4.3 gives an asymptotically exact sequence of semidefinite formulations (4.2) of problem (2.2). Unfortunately, the sizes of these formulations are not bounded by a polynomial in n, even for fixed k, as they depend on the degree d of the polynomials appearing in (4.1). This is not surprising, given our NP-hardness results in the next section. Nevertheless, we can obtain increasingly better upper bounds on Z_P by solving semidefinite problems of size polynomially bounded in n for fixed k. Indeed, since any g that can be represented using sums of squares of polynomials of degree d can clearly also be represented using sums of squares of polynomials of degree $d + 1$, we obtain a family of semidefinite relaxations as follows.

COROLLARY 4.4. $Z_P = Z_D \leq Z_D(d) \leq \cdots \leq Z_D(2) \leq Z_D(1)$.

4.2. On the complexity of the (n, k, Ω)-bound problem. In this section, we show that the separation problem associated with problem (2.2) for the cases $(n, k, \mathcal{R}^n_+)$, $(n, 2k, \mathcal{R}^n)$ are NP-hard for $k \geq 2$. By the equivalence of optimization and separation (see Grötschel, Lovász, and Schrijver [11]), solving problem (2.2) is NP-hard as well in these cases. Finally, since strong duality (Theorem 2.2) holds in these instances, solving the $(n, k, \mathcal{R}^n_+)$-bound problems for $k \geq 2$ and solving the $(n, k, \mathcal{R}^n)$-bound problems with $k \geq 4$ is NP-hard.

The $(n, 2k + 1, \mathcal{R}^n)$-bound problem does not make a case on its own, since an odd degree polynomial cannot be nonnegative over all of $\mathcal{R}^n$. This means that in the corresponding dual problem, the variables y corresponding to $(2k + 1)$-degree coefficients must be zero for the problem to be feasible. Thus the $(n, 2k + 1, \mathcal{R}^n)$-bound problem reduces to the $(n, 2k, \mathcal{R}^n)$-bound problem. In other words, if the highest-order moments that are known are of odd order, they can be disregarded as they will not improve the bound.

4.2.1. The complexity of the $(n, k, \mathcal{R}^n_+)$-bound problem. The separation problem is equivalent to the following problem.

Problem k-SEP$^+$: Given a multivariate polynomial $g(x)$ with rational coefficients and degree k, and a (nonempty) set $S \subseteq \mathcal{R}^n_+$, does there exist $x \in \mathcal{R}^n_+$ such that $g(x) < \chi_S(x)$?

PROPOSITION 4.5. *Problem k-SEP$^+$ is NP-hard for $k \geq 2$, even for polyhedral sets S.*

Proof. Consider the following problem of checking if a given matrix is not coposi-itive.

Problem COPOS: Given a matrix H with rational entries, does there exist $x \in \mathcal{R}_+^n$ such that $x'Hx < 0$?

This problem is NP-hard (see Murty and Kabadi [29]). We will prove that the problem COPOS reduces to Problem k-SEP$^+$. Since COPOS is NP-hard, so is k-SEP$^+$.

Consider first the case $k = 2$, $g(x) = x'Hx + 1$, where H has rational entries, and $S = \mathcal{R}_+^n$. Problem 2-SEP$^+$ is equivalent to the question of whether the matrix H is not copositive. Now, for arbitrary $k \geq 2$, consider the k-degree polynomial $g(x) = x_n^k + (x_1, \dots, x_{n-1})H(x_1, \dots, x_{n-1})' + 1$, where H has rational entries, and $S = \mathcal{R}_+^n$. The separation problem k-SEP$^+$ amounts in this case to checking $g(x) < 1$ for some nonnegative x, which is equivalent to checking if matrix H is not copositive, therefore it is NP-hard.   □

4.2.2. The complexity of the $(n, 2k, \mathcal{R}^n)$-bound problem for $k \geq 2$. For $k \geq 2$, the separation problem can be formulated as follows.

Problem $2k$-SEP: Given a multivariate polynomial $g(\cdot)$ with rational coefficients and degree $2k \geq 4$, and a (nonempty) set $S \subseteq \mathcal{R}^n$, does there exist $x \in \mathcal{R}^n$ such that $g(x) < \chi_S(x)$?

PROPOSITION 4.6. *Problem $2k$-SEP is NP-hard for $k \geq 2$ even for polyhedral sets S.*

Proof. As in the previous proof, we show that the problem COPOS polynomial time reduces to Problem $2k$-SEP for $k \geq 2$. Since COPOS is NP-hard (see Murty and Kabadi [29]), it follows that $2k$-SEP is also NP-hard.

Consider first the case $k = 2$, $g(x) = (x_1^2, \dots, x_n^2)'H(x_1^2, \dots, x_n^2) + 1$, where H has rational entries, and $S = \mathcal{R}^n$. Problem 4-SEP is equivalent to the question of whether the matrix H is not copositive. Now, for arbitrary $k \geq 2$, consider the $2k$-degree polynomial $g(x) = x_n^{2k} + (x_1^2, \dots, x_{n-1}^2)H(x_1^2, \dots, x_{n-1}^2)' + 1$, where H has rational entries, and $S = \mathcal{R}^n$. The separation problem $2k$-SEP amounts in this case to checking $g(x) < 1$ for some $x \in \mathcal{R}^n$, which is equivalent to checking if matrix H is not copositive; therefore it is NP-hard.   □

5. The $(n, 1, \Omega)$-bound problem. In this section, we address the $(n, 1, \Omega)$-bound problem. For convex sets S, the tight bound for the $(n, 1, \mathcal{R}_+^n)$-bound problem is computed in (5.2) as the solution of n convex optimization problems. We present a polynomial time algorithm for more general sets.

Given a vector $M > 0$ representing the means of a random vector X defined in $\mathcal{R}_+^n$, we would like to find tight bounds on $P(X \in S)$ for a convex set S. Marshall [26] derived a tight bound for the case that $S = \{x_i \geq (1 + \delta_i)M_i, \ i = 1, \dots, n\}$ (see (5.3) below).

THEOREM 5.1. *The tight $(n, 1, \mathcal{R}_+^n)$-upper bound for a convex event S is given by*

$$(5.1) \qquad \sup_{X \sim M} P(X \in S) = \min\left(1, \frac{1}{\inf_{x \in S} \max_{i=1,\dots,n} \frac{x_i}{M_i}}\right).$$

Proof. Since only first moments are given, the condition that the vector of moments is in the interior is trivially satisfied, and thus Theorem 2.2 applies. The

corresponding dual problem (2.2) is

$$
\begin{aligned}
Z_D = \min \quad & a'M + b \\
\text{subject to} \quad & a'x + b \geq 1 \quad \forall x \in S, \\
& a'x + b \geq 0 \quad \forall x \in \mathcal{R}_+^n.
\end{aligned}
$$

If the optimal solution (a_0, b_0) satisfies $\min_{x \in S} a_0'x + b_0 = \kappa > 1$, then the solution $\left(\frac{a_0}{\kappa}, \frac{b_0}{\kappa}\right)$ has value $Z_D/\kappa < Z_D$. Therefore, $\inf_{x \in S} a_0'x + b_0 = 1$. By a similar argument we have that $b_0 \leq 1$. Moreover, since $a'x + b \geq 0$ for all $x \in \mathcal{R}_+^n$, $a \geq 0$, and $b \geq 0$. We thus obtain

$$
\begin{aligned}
Z_D = \min \quad & a'M + b \\
\text{subject to} \quad & \inf_{x \in S} a'x = 1 - b, \\
& a \geq 0, \quad 0 \leq b \leq 1.
\end{aligned}
$$

Let $a = \lambda v$, where $\lambda \geq 0$ is a scalar, and $v \geq 0$ is a vector with $\|v\| = 1$. We obtain

$$
\begin{aligned}
Z_D = \min \quad & (1-b)\frac{v'M}{\inf_{x \in S} v'x} + b \\
\text{subject to} \quad & v \geq 0, \quad \|v\| = 1, \ 0 \leq b \leq 1.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Z_D &= \min\left(1, \ \min_{\|v\|=1, v \geq 0} \frac{v'M}{\inf_{x \in S} v'x}\right) \\
&= \min\left(1, \ \frac{1}{\max_{\|v\|=1, v \geq 0} \inf_{x \in S} \frac{v'x}{v'M}}\right) \\
&= \min\left(1, \ \frac{1}{\inf_{x \in S} \max_{\|v\|=1, v \geq 0} \frac{v'x}{v'M}}\right) \\
&= \min\left(1, \ \frac{1}{\inf_{x \in S} \max_{i=1,\ldots,n} \frac{x_i}{M_i}}\right).
\end{aligned}
$$

We exchanged the order of max and inf (see [36, p. 382]); then we relied on the fact that $\max_{\|v\|=1, v \geq 0} \frac{v'x}{v'M}$ is attained at $v = e_j$, where $\frac{x_j}{M_j} = \max_{i=1,\ldots,n} \frac{x_i}{M_i}$. $\quad\square$

Denote $\phi(x) = \max_{i=1,\ldots,n} \frac{x_i}{M_i}$, so $\phi(x) = \frac{x_i}{M_i}$ whenever $x \in S_i$. The last term in (5.1) can be written as

$$
\frac{1}{\inf_{x \in S} \phi(x)} = \max_{i=1,\ldots,n} \sup_{x \in S_i} \frac{1}{\phi(x)} = \max_{i=1,\ldots,n} \sup_{x \in S_i} \frac{M_i}{x_i} = \max_{i=1,\ldots,n} \frac{M_i}{\inf_{x \in S_i} x_i}.
$$

This shows that the bound in Theorem 5.1 can be obtained by solving n convex optimization problems:

$$
(5.2) \qquad \sup_{X \sim M} P(X \in S) = \min\left(1, \ \max_{i=1,\ldots,n} \frac{M_i}{\inf_{x \in S_i} x_i}\right),
$$

where $S_i = \{x \in S \mid \frac{x_i}{M_i} \geq \frac{x_j}{M_j} \ \forall j \neq i\}$ is the convex subset of S for which the mean-rescaled ith coordinate is largest. Remark that $\inf_{x \in S_i} x_i = 0$ for some i if and only if 0 is a boundary point of S, in which case the bound is trivially 1.

When we specialize bound (5.2) for the set $S = \{x \mid x_i \geq (1 + \delta_i)M_i, \ \forall i = 1, \ldots, n\}$, we obtain

$$(5.3) \qquad \sup_{X \sim M} P(X_1 \geq (1 + \delta_1)M_1, \ldots, X_n \geq (1 + \delta_n)M_n) = \min_{i=1,\ldots,n} \frac{1}{1 + \delta_i},$$

which represents a multidimensional generalization of Markov's inequality due to Marshall [26]. In particular, for a univariate random variable, in the case that $S = [(1 + \delta)M, \infty)$, this is exactly Markov's inequality:

$$\sup_{X \sim M} P(X \geq (1 + \delta)M) = \frac{1}{1 + \delta}.$$

5.1. Extremal distributions for the $(n, 1, \mathcal{R}_+^n)$-bound problem. In this section, we construct a distribution that achieves bound (5.1). We will say that the bound (5.1) is achievable, when there exists an $x^* \in S$ such that

$$\min\left(1, \frac{1}{\inf_{x \in S} \max_{i=1,\ldots,n} \frac{x_i}{M_i}}\right) = \frac{M_i}{x_i^*} < 1.$$

In particular, the bound is achievable when the set S is closed and $M \notin S$.

THEOREM 5.2. (a) *If $M \in S$ or if the bound (5.1) is achievable, then there is an extremal distribution that exactly achieves it.*

(b) *Otherwise, there is a sequence of distributions defined on $\mathcal{R}_+^n$ with mean M, that asymptotically achieves it.*

Proof. (a) If $M \in S$, then the extremal distribution is simply $P(X = M) = 1$. Now suppose that $M \notin S$ and the bound (5.1) is achievable. We assume without loss of generality that the bound equals $\frac{M_1}{x_1^*} < 1$, and it is achieved at $x^* \in S$ with $\frac{x_1^*}{M_1} = \max_{i=1,\ldots,n} \frac{x_i^*}{M_i}$, i.e., $x^* \in S_1$. The random variable

$$X = \begin{cases} x^*, & \text{with probability } p = \dfrac{M_1}{x_1^*}, \\[2mm] v = \dfrac{x_1^* M - M_1 x^*}{x_1^* - M_1}, & \text{with probability } 1 - p = 1 - \dfrac{M_1}{x_1^*}, \end{cases}$$

has mean $E[X] = M$ and is nonnegative: $v_i = \frac{M_i x_1^* - M_1 x_i^*}{x_1^* - M_1} \geq 0$ for all i since $\frac{x_1^*}{M_1} = \max_{i=1,\ldots,n} \frac{x_i^*}{M_i}$. Moreover, $v \notin S$, or else convexity of S implies $M = px^* + (1-p)v \in S$, a contradiction. Therefore,

$$P(X \in S) = P(X = x^*) = \frac{M_1}{x_1^*}.$$

(b) If $M \notin S$ and the bound (5.1) is not achievable, then we construct a sequence of nonnegative distributions with mean M that approach it. Suppose without loss of generality that $\inf_{x \in S} \max_{i=1,\ldots,n} \frac{x_i}{M_i} = \inf_{x \in S} \frac{x_1}{M_1} = \frac{x_1^*}{M_1}$. Thus, bound (5.1) equals $\min(1, \frac{M_1}{x_1^*})$. Consider a sequence $x^k \in S_1$ with $x_1^k \to x_1^*$, so $\lim_{k \to \infty} \max_{i=1,\ldots,n} \frac{x_i^k}{M_i} = \lim_{k \to \infty} \frac{x_1^k}{M_1} = \frac{x_1^*}{M_1}$, and a sequence p_k, $0 < p_k < \min(1, \frac{M_1}{x_1^k})$ so that $p_k \to \min(1, \frac{M_1}{x_1^*})$. Therefore the random variables

$$X_k = \begin{cases} x^k, & \text{with probability } p_k, \\[2mm] v^k = \dfrac{M - p_k x^k}{1 - p_k}, & \text{with probability } 1 - p_k, \end{cases}$$

are nonnegative with mean $E[X_k] = M$. Also $v^k \notin S$ or else $M \in S$, so $P(X_k \in S) = P(X_k = x^k) = p_k \to \min(1, \frac{M_1}{x_1^*})$. This shows that the sequence of nonnegative distributions X_k with mean M asymptotically achieves the bound (5.1). □

5.2. A polynomial time algorithm for unions of convex sets. In this section, we present a polynomial time algorithm that computes a tight $(n, 1, \Omega)$-bound for events S and Ω that can be decomposed as a disjoint union of a polynomial (in n) number of convex sets such that we can solve linear optimization problems over them in polynomial time. Examples include polyhedra and sets defined by semidefinite constraints. Our overall strategy is to formulate the problem as an optimization problem, consider its dual, and exhibit an algorithm that solves the corresponding separation problem in polynomial time.

In this case we are given the mean vector $M = (M_1, \ldots, M_n)$ of an n-dimensional random variable X with domain Ω that can be decomposed in a polynomial (in n) number of convex sets, and we want to derive tight bounds on $P(X \in S)$. Problem (2.2) can be written as follows:

(5.4)
$$Z_D = \min \quad y'M + y_0$$
$$\text{subject to} \quad g(x) = y'x + y_0 \geq \chi_S(x) \quad \forall x \in \Omega.$$

The separation problem associated with problem (5.4) is defined as follows: given a vector a and a scalar b we want to check whether $g(x) = a'x + b \geq \chi_S(x)$, for all $x \in \Omega$, and if not, we want to exhibit a violated inequality. The following algorithm achieves this goal.

ALGORITHM A.
1. Solve the problem $\inf_{x \in \Omega} g(x)$ (note that the problem involves a polynomial number of convex optimization problems; in particular if Ω is polyhedral, this is a linear optimization problem). Let z_0 be the optimal solution value and let $x_0 \in \Omega$ be an optimal solution.
2. If $z_0 < 0$, then we have $g(x_0) = z_0 < 0$; this constitutes a violated inequality.
3. Otherwise, we solve $\inf_{x \in S} g(x)$ (again, the problem involves a polynomial number of convex optimization problems, while if S is polyhedral, this is a linear optimization problem). Let z_1 be the optimal solution value and let $x_1 \in S$ be an optimal solution.
   (a) If $z_1 < 1$, then for $x_1 \in S$ we have $g(x_1) = z_1 < 1$; this constitutes a violated inequality.
   (b) If $z_1 \geq 1$, then a, b are feasible.

The above algorithm solves the separation problem in polynomial time, and thus the $(n, 1, \Omega)$-upper bound problem is polynomially solvable (see Nesterov and Nemirovskii [30] and Grötschel, Lovász, and Schrijver [11]).

6. The $(n, 2, \mathcal{R}^n)$-bound problem. In this section, we address the $(n, 2, \mathcal{R}^n)$-bound problem. Rather than assuming that $E[X]$ and $E[XX']$ are known, we assume equivalently that the vector $M = E[X]$ and the covariance matrix $\Gamma = E[(X - M)(X - M)']$ are known and that Γ is invertible. Given a set $S \subseteq \mathcal{R}^n$, we find tight upper bounds, denoted by $\sup_{X \sim (M, \Gamma)} P(X \in S)$, on the probability $P(X \in S)$ for all random vectors X defined on $\mathcal{R}^n$ with mean $M = E[X]$ and covariance matrix $\Gamma = E[(X - M)(X - M)']$. If the set S is convex, we review the work of Marshall and Olkin [27], who solved the problem as a convex optimization problem. If the set S is a union of convex sets, over which we can optimize a convex quadratic function efficiently, we provide a polynomial time algorithm.

6.1. The $(n, 2, \mathcal{R}^n)$-bound problem for a convex set. The following result is due to Marshall and Olkin [27], who give a constructive proof. An alternative, optimization based proof is provided in Popescu [33].

THEOREM 6.1 (Marshall and Olkin [27]). *The tight $(n, 2, \mathcal{R}^n)$-upper bound for a convex event S is given by*

$$(6.1) \qquad \sup_{X \sim (M, \Gamma)} P(X \in S) = \frac{1}{1 + d^2},$$

where $d^2 = \inf_{x \in S}(x - M)'\Gamma^{-1}(x - M)$ is the squared distance from M to the set S, under the norm induced by the matrix Γ^{-1}.

The actual formulation provided by Marshall and Olkin [27] is the following:

$$(6.2) \qquad \sup_{X \sim (0, \Gamma)} P(X \in S) = \inf_{a \in S^\perp} \frac{1}{1 + (a'\Gamma a)^{-1}},$$

where $S^\perp = \{a \in \mathcal{R}^n \mid a'x \geq 1 \; \forall x \in S\}$, is the so-called antipolar of S (a.k.a "blocker", or "upper-dual"). The above result is with zero mean, but can be easily extended for nonzero mean by a simple transformation (see [27, p. 1013, eqs. (7.8)–(7.9)], or the first part of the proof of Theorem 6.2). Given that $(a'\Gamma a)(x'\Gamma^{-1}x) \geq (a'x)^2 \geq 1$ for all $x \in S$, $a \in S^\perp$, one can easily see that the bound (6.1) is at least as tight as (6.2). Equality follows from nonlinear gauge duality principles (see Freund [8]).

Marshall and Olkin [27] construct an extremal distribution of a random variable $X \sim (M, \Gamma)$, so that $P(X \in S) = 1/(1 + d^2)$ with $d^2 = \inf_{x \in S}(x - M)'\Gamma^{-1}(x - M)$. We will say that the bound d is achievable, when there exists an $x^* \in S$ such that $d^2 = (x^* - M)'\Gamma^{-1}(x^* - M)$. In particular, d is achievable if the set S is closed.

THEOREM 6.2 (see [27]).

(a) *If $M \notin S$ and if $d^2 = \inf_{x \in S}(x - M)'\Gamma^{-1}(x - M)$ is achievable, then there is an extremal distribution that exactly achieves the bound* (6.1).

(b) *Otherwise, if $M \in S$ or if d^2 is not achievable, then there is a sequence of (M, Γ)-feasible distributions that asymptotically approach the bound* (6.1).

Bounds on tail probabilities. Given a vector $M = (M_1, \ldots, M_n)'$, and an $n \times n$ positive definite, full rank matrix Γ, we derive next tight bounds on the following upper, lower, and two-sided tail probabilities of a random vector $X = (X_1, \ldots, X_n)'$ with mean $M = E[X]$ and covariance matrix $\Gamma = E[(X - M)(X - M)']$:

$$P(X > M_{e+\delta}) = P(X_i > (1 + \delta_i)M_i \; \forall i = 1, \ldots, n),$$
$$P(X < M_{e-\delta}) = P(X_i < (1 - \delta_i)M_i \; \forall i = 1, \ldots, n),$$
$$P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) = P(X_i - M_i > \delta_i M_i \; \forall i, \text{ or } X_i - M_i < -\delta_i M_i \; \forall i),$$

where $\delta = (\delta_1, \ldots, \delta_n)'$, and we denote $M_\delta = (\delta_1 M_1, \ldots, \delta_n M_n)'$. In order to obtain nontrivial bounds we require that not all $\delta_i M_i \leq 0$, which expresses the fact that the tail event does not include the mean vector.

The one-sided Chebyshev inequality. In the following, we find a tight bound for $P(X > M_{e+\delta})$. The bound immediately extends to $P(X < M_{e-\delta})$.

PROPOSITION 6.3. (a) *The tight multivariate one-sided $(n, 2, \mathcal{R}^n)$-Chebyshev bound is*

$$(6.3) \qquad \sup_{X \sim (M, \Gamma)} P(X > M_{e+\delta}) = \frac{1}{1 + d^2},$$

where d^2 is given by

$$(6.4) \qquad \begin{aligned} d^2 = \min \ \ & x'\Gamma^{-1}x \\ \text{subject to} \ \ & x \geq M_\delta, \end{aligned}$$

or alternatively d^2 is given by the gauge dual problem of (6.4)

$$(6.5) \qquad \begin{aligned} \frac{1}{d^2} = \min \ \ & x'\Gamma x \\ \text{subject to} \ \ & x'M_\delta = 1, \\ & x \geq 0. \end{aligned}$$

(b) *If $\Gamma^{-1}M_\delta \geq 0$, then the tight bound is expressible in closed form:*

$$(6.6) \qquad \sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta}) = \frac{1}{1 + M_\delta'\Gamma^{-1}M_\delta}.$$

Proof. (a) Applying the bound (6.1) for $S = \{x \mid x_i > (1+\delta_i)M_i \ \forall i = 1,\ldots,n\}$, and changing variables we obtain (6.3). The alternative expression (6.5) for d^2 follows from elementary gauge duality theory (see Freund [8]).

(b) The Kuhn–Tucker conditions for problem (6.4) are as follows:

$$2\Gamma^{-1}x - \lambda = 0, \quad \lambda \geq 0, \quad x \geq M_\delta, \qquad \lambda_i(x_i - (M_\delta)_i) = 0 \quad \forall i.$$

The choice $x = M_\delta$, $\lambda = 2\Gamma^{-1}M_\delta \geq 0$ (by assumption) satisfies the Kuhn–Tucker conditions, which are sufficient (this is a convex quadratic optimization problem). Thus, $d^2 = M_\delta'\Gamma^{-1}M_\delta$, and hence, (6.6) follows. □

The two-sided Chebyshev inequality. The following result provides a tight bound for $P(X > M_{e+\delta} \text{ or } X < M_{e-\delta})$.

PROPOSITION 6.4 (see [27]).

(a) *The tight multivariate two-sided $(n, 2, \mathcal{R}^n)$-Chebyshev bound is*

$$(6.7) \qquad \sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) = \min(1, t^2),$$

where

$$(6.8) \qquad \begin{aligned} t^2 = \min \ \ & x'\Gamma x \\ \text{subject to} \ \ & x'M_\delta = 1, \\ & x \geq 0. \end{aligned}$$

(b) *If $\Gamma^{-1}M_\delta \geq 0$, then the tight bound is expressible in closed-form:*

$$(6.9) \qquad \sup_{X \sim (M,\Gamma)} P(X > M_{e+\delta} \text{ or } X < M_{e-\delta}) = \min\left(1, \frac{1}{M_\delta'\Gamma^{-1}M_\delta}\right).$$

This formulation follows via gauge duality principles from the original zero-mean result of Marshall and Olkin [28].

In the univariate case $M_\delta = \delta M$ and $\Gamma = \sigma^2$. Therefore, $\Gamma^{-1}M_\delta = \frac{\delta M}{\sigma^2} \geq 0$, and the closed-form bound applies, i.e.,

$$(6.10) \qquad P(X > (1+\delta)M) \leq \frac{C_M^2}{\delta^2 + C_M^2},$$

where $C_M^2 = \frac{\sigma^2}{M^2}$ is the coefficient of variation of the random variable X. The usual Chebyshev inequality is given by $P(X > (1+\delta)M) \leq \frac{C_M^2}{\delta^2}$. Inequality (6.10) is always stronger. Moreover, there exist extremal distributions that satisfy it with equality (see Theorem 6.2). The original univariate result can be traced back to the 1937 book of Uspensky [46], and is mentioned later by Marshall and Olkin (1960) [27], [28], but has not received much attention in modern probability textbooks.

6.2. The tight $(n, 2, \mathcal{R}^n)$-bound for unions of convex sets. We are given first- and second-order moment information (M, Γ) on the n-dimensional random variable X, and we would like to compute $\sup_{X \sim (M,\Gamma)} P(X \in S)$. Recall that the corresponding dual problem can be written as

(6.11)
$$Z_D = \min_{Y, y, y_0} \quad Y \bullet \Gamma + y'M + y_0$$
$$\text{subject to} \quad g(x) = x'Yx + y'x + y_0 \geq \chi_S(x) \quad \forall x \in \mathcal{R}^n.$$

We consider convex sets for which we can solve a convex quadratic optimization problem over them in polynomial time. Examples include polyhedra and sets defined by semidefinite constraints. The separation problem corresponding to problem (6.11) can be stated as follows: Given a matrix H, a vector c, and a scalar d, we need to check whether $g(x) = x'Hx + c'x + d \geq \chi_S(x)$ for all $x \in \mathcal{R}^n$, and, if not, we need to find a violated inequality. Notice that we can assume without loss of generality that the matrix H is symmetric.

The following algorithm solves the separation problem in polynomial time.

ALGORITHM B.
1. If H is not positive semidefinite, then one can exhibit a polynomial size $x_0 \in \mathcal{Q}^n$ so that $x_0'Hx_0 < -1$. Let $c_0 = c'x_0$, so $g(\lambda x_0) < -\lambda^2 + \lambda c_0 + d$ for any scalar λ. Choosing λ large enough (e.g., $\lambda > |c_0| + \sqrt{|d|}$) so that $g(\lambda x_0) < 0$, produces a violated inequality.
2. Otherwise, if H is positive semidefinite, then we do the following:
   (a) We test if $g(x) \geq 0$ for all $x \in \mathcal{R}^n$ by solving the convex optimization problem

   $$\inf_{x \in \mathcal{R}^n} g(x).$$

   Let z_0 be the optimal value. If $z_0 < 0$, we find x_0 such that $g(x_0) < 0$, which represents a violated inequality.
   (b) Otherwise, we test if $g(x) \geq 1$ for all $x \in S$ by solving a polynomial collection of convex optimization problems

   $$\inf_{x \in S} g(x).$$

   Let z_1 be the optimal value. If $z_1 \geq 1$, then $g(x) \geq 1$ for all $x \in S$, and thus (H, c, d) is feasible. If not, we exhibit an x_1 such that $g(x_1) < 1$, and thus we identify a violated inequality.

Since we can solve the separation problem in polynomial time, we can also solve (within ϵ) the $(n, 2, \mathcal{R}^n)$-bound problem in polynomial time (in the problem data and $\log \frac{1}{\epsilon}$).

7. Concluding remarks. In this paper we characterized the (n, k, Ω)-bound problem. We provided polynomial time algorithms via semidefinite programming algorithms for the $(1, k, \Omega)$-bound problem, and via convex optimization methods, for the

$(n, 1, \mathcal{R}_+^n)$, $(n, 2, \mathcal{R}^n)$-bound problems. We showed that the $(n, k, \mathcal{R}_+^n)$ and $(n, k, \mathcal{R}^n)$-bound problems are NP-hard for $k \geq 2$, respectively $k \geq 4$. We also provided a family of semidefinite relaxations of polynomial size for the general (n, k, Ω)-bound problem that provide increasingly stronger bounds.

REFERENCES

[1] N. I. Akhiezer, *The Classical Moment Problem*, Hafner, New York, 1965.

[2] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS/SIAM Ser. Optim. 2, SIAM, Philadelphia, 2001.

[3] F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[4] J. M. Borwein and A. S. Lewis, *Partially finite convex programming, part* I*: Quasi relative interiors and duality theory*, Math. Program., 57 (1992), pp. 15–48.

[5] J. M. Borwein and A. S. Lewis, *Partially finite convex programming, part* II*: Explicit lattice models*, Math. Program., 57 (1992), pp. 49–83.

[6] P. Chebyshev, *Sur les valeurs limites des intégrales*, J. Math. Pure. Appl., 19 (1874), pp. 157–160.

[7] P. Diaconis, *Application of the method of moments in probability and statistics*, in Moments in Mathematics, Proc. Sympos. Appl. Math. 37, H. J. Landau, ed., AMS, Providence, RI, 1987, pp. 125–139.

[8] R. Freund, *Dual gauge programs, with applications to quadratic programming and the minimum-norm problem*, Math. Program., 38 (1987), pp. 47–67.

[9] H. J. Godwin, *On generalizations of Tchebycheff's inequality*, J. Amer. Statist. Assoc., 50 (1955), pp. 923–945.

[10] H. J. Godwin, *Inequalities on distribution functions*, in Griffin's Statistical Monographs and Courses 16, M. G. Kendall, ed., Charles Griffin and Co., London, 1964.

[11] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Algorithms and Combinatorics, Springer, Berlin, Heidelberg, 1988.

[12] H. Hamburger, *Üeber eine erweiterung der Stieltjes'schen momentenproblems*, Math. Ann., 81 (1920), pp. 235–319.

[13] H. Hamburger, *Üeber eine erweiterung der Stieltjes'schen momentenproblems*, Math. Ann., 82 (1921), pp. 120–164, 168–187.

[14] E. K. Haviland, *On the momentum problem for distributions in more than one dimension*, Amer. J. Math., 57 (1935), pp. 562–568.

[15] E. K. Haviland, *On the momentum problem for distribution functions in more than one dimension*, Amer. J. Math., 58 (1936), pp. 164–168.

[16] K. Isii, *The extrema of probability determined by generalized moments.* I. *Bounded random variables*, Ann. Inst. Statist. Math., 12 (1960), pp. 119–133.

[17] K. Isii, *On the sharpness of Chebyshev-type inequalities*, Ann. Inst. Statist. Math., 14 (1963), pp. 185–197.

[18] S. Karlin and L. S. Shapley, *Geometry of moment spaces*, Mem. Amer. Math. Soc., 12 (1953).

[19] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Pure Appl. Math. 15, Interscience, John Wiley and Sons, New York, 1966.

[20] J. H. B. Kemperman, *The general moment problem, a geometric approach*, Ann. Math. Statist., 39 (1968), pp. 93–122.

[21] J. H. B. Kemperman, *On the role of duality in the theory of moments*, in Semi-infinite Programming and Applications, Lecture Notes in Econom. and Math. Systems 215, Springer, Berlin, New York, 1983, pp. 63–92.

[22] J. H. B. Kemperman, *Geometry of the moment problem*, in Moments in Mathematics, Proc. Sympos. Appl. Math. 37, H. J. Landau, ed., AMS, Providence, RI, 1987, pp. 16–53.

[23] H. J. Landau, *Classical background on the moment problem*, in Moments in Mathematics, Proc. Sympos. Appl. Math. 37, H. J. Landau, ed., AMS, Providence, RI, 1987, pp. 1–15.

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

[25] A. Markov, *On Certain Applications of Algebraic Continued Fractions*, Ph.D. thesis, St. Petersburg, 1884 (in Russian).

[26] A. Marshall, *Markov's inequality for random variables taking values in a linear topological space*, in Inequalities in Statistics and Probability, IMS Lecture Notes Monogr. Ser. 5, Y. L. Tong, ed., Institute of Mathematical Statistics, Hayward, CA, 1984, pp. 104–108.

[27] A. Marshall and I. Olkin, *Multivariate Chebyshev inequalities*, Ann. Math. Statist., 31 (1960), pp. 1001–1014.

[28] A. Marshall and I. Olkin, *A one-sided inequality of the Chebyshev type*, Ann. Math. Statist., 31 (1960), pp. 488–491.

[29] K. G. Murty and S. N. Kabadi, *Some NP-complete problems in quadratic and nonlinear programming*, Math. Program., 39 (1987), pp. 117–129.

[30] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, 1994.

[31] Yu. Nesterov, *Structure of Non-Negative Polynomial and Optimization Problems*, Preprint DP 9749, Louvain-la-Neuve, Belgium, 1997.

[32] P. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, Caltech, 2000.

[33] I. Popescu, *Applications of Optimization in Probability, Finance and Revenue Management*, Ph.D. Dissertation, Applied Mathematics Department and Operations Research Center, MIT, Cambridge, MA, 1999.

[34] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J., 42 (1993), pp. 969–984.

[35] B. Reznick, *Some concrete aspects of Hilbert's 17th problem*, in Real Algebraic Geometry and Ordered Structures, Contemp. Math. 253, C. N. Delzell and J. J. Madden, eds., AMS, Providence, RI, 2000, pp. 251–272.

[36] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[37] K. Schmudgen, *The k-moment problem for compact semialgebraic sets*, Math. Ann., 289 (1991), pp. 203–206.

[38] S. Sengupta, W. Siu, and N. C. Lind, *A linear programming formulation of the problem of moments*, Z. Angew. Math. Mech., 10 (1979), pp. 533–537.

[39] A. Shapiro, *On duality theory of conic linear problems*, in Semi-Infinite Programming: Recent Advances, Miguel A. Goberna and Marco A. López, eds., Kluwer, Massachusetts, 2001, pp. 135–165.

[40] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, AMS, New York, 1943.

[41] N. Z. Shor, *Class of global minimum bounds of polynomial functions*, Cybernetics, 23 (1987), pp. 731–734.

[42] J. Smith, *Generalized Chebyshev inequalities: Theory and applications in decision analysis*, Oper. Res., 43 (1995), pp. 807–825.

[43] T. J. Stieltjes, *Recherches sur les fractions continues*, Ann. Fac. Sci. Toulouse Math. (6), 8 (1894), pp. 1–122.

[44] T. J. Stieltjes, *Recherches sur les fractions continues*, Ann. Fac. Sci. Toulouse Math. (6), 9 (1895), pp. 5–47.

[45] Y. L. Tong, *Probability Inequalities in Multivariate Distributions*, Academic Press, New York, 1980.

[46] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw–Hill, New York, 1937.

[47] H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*, Kluwer, Massachusetts, 2000.