SEMIDEFINITE PROGRAMMING AND MULTIVARIATE CHEBYSHEV BOUNDS

Katherine Comanor * **Lieven Vandenberghe** ** **Stephen Boyd** ***

* *The RAND Corporation, Santa Monica, California*
** *Electrical Engineering Department, UCLA*
*** *Electrical Engineering Department, Stanford University*

Abstract: Chebyshev inequalities provide bounds on the probability of a set based on known expected values of certain functions, for example, known power moments. In some important cases these bounds can be efficiently computed via convex optimization. We discuss one particular type of generalized Chebyshev bound, a lower bound on the probability of a set defined by strict quadratic inequalities, given the mean and the covariance of the distribution. We present a semidefinite programming formulation, give an interpretation of the dual problem, and describe some applications.

1. INTRODUCTION

The classical Chebyshev inequality states that

$$\mathbf{Prob}(|X| \geq 1) \leq \sigma^2$$

for a zero-mean random variable $X \in \mathbf{R}$ with variance σ^2. Generalized Chebyshev bounds provide similar upper or lower bounds on the probability of a set in $\mathbf{R}^n$ based on known expected values of certain functions, for example, the mean and covariance. Several such multivariate generalizations of Chebyshev's inequality appeared during the 1950s and 1960s, see Isii (1959), Isii (1963), Isii (1964), Marshall and Olkin (1960), Karlin and Studden (1966). Isii (1964), for example, considers the problem of computing upper and lower bounds on $\mathbf{E}\, f_0(X)$, where X is a random variable on $\mathbf{R}^n$ that satisfies the moment constraints

$$\mathbf{E}\, f_i(X) = a_i, \quad i = 1, \ldots, m.$$

The best lower bound on $\mathbf{E}\, f_0(X)$ is given by the optimal value of the linear optimization problem

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\, f_0(X) \\ \text{subject to} & \mathbf{E}\, f_i(X) = a_i, \quad i = 1, \ldots, m, \end{array} \tag{1}$$

where we optimize over all probability distributions on $\mathbf{R}^n$. The dual of this problem is

$$\begin{array}{ll} \text{maximize} & z_0 + \sum_{i=1}^{m} a_i z_i \\ \text{subject to} & z_0 + \sum_{i=1}^{m} z_i f_i(x) \leq f_0(x) \text{ for all } x, \end{array} \tag{2}$$

and has a finite number of variables z_i, $i = 0, \ldots, m$, but infinitely many constraints. Isii shows that strong duality holds under appropriate constraint qualifications, so we can find a sharp lower bound on $\mathbf{E}\, f_0(X)$ by solving (2). Note that the constraints in (2) can be written as a single constraint $g(z_0, \ldots, z_m) \leq 0$, where

$$g(z_0, \ldots, z_m) = \sup_x (z_0 + \sum_{i=1}^m z_i f_i(x) - f_0(x)).$$

This is a convex function of z, but in general difficult to evaluate. Hence, (2) is usually an intractable optimization problem.

Bertsimas and Sethuraman (2000), Lasserre (2002), Bertsimas and Popescu (2005), and Popescu (2005) have recently shown that various types of generalized Chebyshev bounds can be computed via semidefinite programming. In this paper we discuss one important example: a lower bound on the probability of a set defined by strict quadratic inequalities, given the mean and the covariance of the distribution. The main purpose of the paper is to outline a proof of the main result from semidefinite programming duality (see Vandenberghe et al. (2006)), give a geometric interpretation, and describe some applications.

2. PROBABILITY OF A SET DEFINED BY QUADRATIC INEQUALITIES

Let $C \subseteq \mathbf{R}^n$ be defined by m strict quadratic inequalities

$$x^T A_i x + 2b_i^T x + c_i < 0, \quad i = 1, \ldots, m, \tag{3}$$

with $A_i \in \mathbf{S}^n$ (not necessarily positive semidefinite), $b_i \in \mathbf{R}^n$, $c_i \in \mathbf{R}$. It is easily verified that the optimal value of the following semidefinite program (SDP) is a lower bound on $\mathbf{Prob}(X \in C)$ for distributions with $\mathbf{E}\, X = \bar{x}$ and $\mathbf{E}\, XX^T = S$:

$$\begin{array}{ll} \text{maximize} & 1 - \mathbf{tr}(SP) - 2q^T\bar{x} - r \\ \text{subject to} & \begin{bmatrix} P & q \\ q^T & r \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} P - \tau_i A_i & q - \tau_i b_i \\ (q - \tau_i b_i)^T & r - 1 - \tau_i c_i \end{bmatrix} \succeq 0, \\ & \qquad\qquad i = 1, \ldots, m \\ & \tau_i \geq 0, \quad i = 1, \ldots, m. \end{array} \tag{4}$$

The variables in this SDP are $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$, and $\tau_i \in \mathbf{R}$, for $i = 1, \ldots, m$.

To verify this, we first note that the constraints imply that

$$x^T P x + 2q^T x + r \geq 1 + \tau_i (x^T A_i x + 2b_i^T x + c_i)$$

and $x^T P x + 2q^T x + r \geq 0$ for all x. This means that

$$x^T P x + 2q^T x + r \geq 1 - \mathbf{1}_C(x)$$

where $\mathbf{1}_C$ is the 0-1 indicator function of C (*i.e.*, $\mathbf{1}_C(x) = 1$ if $x \in C$ and $\mathbf{1}_C(x) = 0$ otherwise). Hence, if $\mathbf{E}\, X = \bar{x}$, and $\mathbf{E}\, XX^T = S$, then

$$\begin{aligned} \mathbf{tr}(SP) + 2q^T\bar{x} + r &= \mathbf{E}(X^T P X + 2q^T X + r) \\ &\geq 1 - \mathbf{E}\,\mathbf{1}_C(X) \\ &= 1 - \mathbf{Prob}(X \in C). \end{aligned}$$

This simple argument shows that $1 - \mathbf{tr}(SP) - 2q^T\bar{x} - r$ is a lower bound on $\mathbf{Prob}(X \in C)$. In the SDP (4) we compute the best lower bound that can be constructed this way.

The proof does not establish that the bound is actually achieved by a distribution with the specified moments. This stronger result can be proved by combining Isii's semi-infinite linear programming bounds with the S-procedure (Boyd et al., 1994, page 23). It can also be proved directly from semidefnite programming duality. The dual problem of (4) is

$$\begin{array}{ll} \text{minimize} & 1 - \sum_{i=1}^m \lambda_i \\ \text{subject to} & \begin{bmatrix} Z_i & z_i \\ z_i^T & \lambda_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m \\ & \sum_{i=1}^m \begin{bmatrix} Z_i & z_i \\ z_i^T & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} S & \bar{x} \\ \bar{x}^T & 1 \end{bmatrix} \\ & \mathbf{tr}(A_i Z_i) + 2b_i^T z_i + c_i \lambda_i \geq 0, \\ & \qquad\qquad i = 1, \ldots, m, \end{array} \tag{5}$$

which is an SDP with variables $Z_i \in \mathbf{S}^n$, $z_i \in \mathbf{R}^n$, and $\lambda_i \in \mathbf{R}$. It can be shown that from every feasible solution in (5) a discrete distribution can be constructed with $\mathbf{E}\, X = \bar{x}$, $\mathbf{E}\, XX^T = S$ and $\mathbf{Prob}(X \in C) \leq 1 - \sum_i \lambda_i$ (Vandenberghe et al. (2006)). Tightness of the generalized Chebyshev inequality (4) then follows from the fact that the two SDPs are duals and have the same optimal value.

3. GEOMETRIC INTERPRETATION

Figure 1 shows an example in $\mathbf{R}^2$. The set C is defined by three linear inequalities and two concave quadratic inequalities. The mean $\bar{x} = \mathbf{E}\, X$ is shown as a small circle, and the set

$$\{x \mid (x - \bar{x})^T (S - \bar{x}\bar{x}^T)^{-1} (x - \bar{x}) = 1\}$$

is shown as the dashed ellipse. The optimal Chebyshev lower bound on $\mathbf{Prob}(X \in C)$ for this example is 0.3992. The six solid dots are the possible values of the discrete distribution computed from

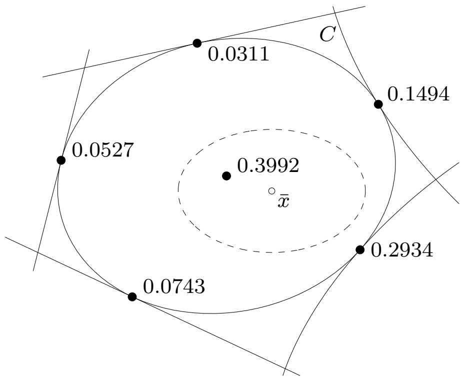

Fig. 1. The distribution that achieves the lower bound on $\mathbf{Prob}(X \in C)$ for a given mean and covariance.

the optimal solution of the SDP (5). The numbers next to the dots give the probability of the six values. (Since C is defined as an open set, the five points on the boundary are not in C itself, so $\mathbf{Prob}(X \in C) = 0.3992$ for this distribution.)

The solid ellipse is the level curve

$$\{x \mid x^T P x + 2q^T x + r = 1\}$$

where P, q, and r are the optimal solution of the lower bound SDP (4). We notice that the optimal distribution allocates nonzero probability to the points where the ellipse touches the boundary of C, and to its center. This relation between the solutions of the upper and lower bound SDPs holds in general, and is a consequence of the optimality conditions of semidefinite programming.

4. EXAMPLES

In the simplest cases, the two SDPs can be solved analytically. As an example, we derive an extension of the Chebyshev inequality known as Selberg's inequality (Karlin and Studden, 1966, page 475). We take $C = (-1, 1) = \{x \in \mathbf{R} \mid x^2 < 1\}$. We show that if $\mathbf{E}\, X = \bar{x}$ and $\mathbf{E}\, X^2 = s$, then

$$\mathbf{Prob}(|X| < 1) \geq \begin{cases} 0 & 1 \leq s \\ 1 - s & |\bar{x}| \leq s < 1 \\ \dfrac{(1 - |\bar{x}|)^2}{s - 2|\bar{x}| + 1} & s < |\bar{x}| \end{cases} \quad (6)$$

and that there is a distribution that achieves the bound. We will assume that $\bar{x} \geq 0$.

The SDP (4) is

$$\begin{array}{ll} \text{maximize} & 1 - sP - 2\bar{x}q - r \\ \text{subject to} & \begin{bmatrix} P & q \\ q & r-1 \end{bmatrix} \succeq \tau \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ & \tau \geq 0 \\ & \begin{bmatrix} P & q \\ q & r \end{bmatrix} \succeq 0, \end{array}$$

with variables $P, q, r, \tau \in \mathbf{R}$. The values $P = q = \tau = 0$, $r = 1$ are obviously feasible, with objective value zero. The values $P = \tau = 1$, $r = q = 0$ are also feasible, with objective value $1 - s$. The values

$$\begin{bmatrix} P & q \\ q & r \end{bmatrix} = \frac{1}{(s - 2\bar{x} + 1)^2} \begin{bmatrix} 1 - \bar{x} \\ s - \bar{x} \end{bmatrix} \begin{bmatrix} 1 - \bar{x} \\ s - \bar{x} \end{bmatrix}^T$$
$$\tau = \frac{1 - \bar{x}}{s - 2\bar{x} + 1}$$

are feasible if $s < \bar{x}$, with objective value

$$1 - sP - 2\bar{x}q - r = \frac{(1 - \bar{x})^2}{s - 2\bar{x} + 1}.$$

This proves (6). To show that the inequality is tight, we use the dual SDP (5):

$$\begin{array}{ll} \text{minimize} & 1 - \lambda \\ \text{subject to} & Z \geq \lambda \\ & 0 \preceq \begin{bmatrix} Z & z \\ z & \lambda \end{bmatrix} \preceq \begin{bmatrix} s & \bar{x} \\ \bar{x} & 1 \end{bmatrix} \end{array}$$

with variables $\lambda, Z, z \in \mathbf{R}$. If $\bar{x} > s$, the values

$$Z = z = \lambda = \frac{s - \bar{x}^2}{s - 2\bar{x} + 1} = \frac{s - \bar{x}^2}{s - \bar{x}^2 + (\bar{x} - 1)^2},$$

are feasible with objective function

$$1 - \lambda = \frac{(1 - \bar{x})^2}{s - 2\bar{x} + 1}.$$

If $\bar{x} \leq s < 1$, we can also take $Z = s$, $z = \bar{x}$, $\lambda = s$, which provides a feasible point with a smaller objective value, $1 - s$. Finally, if $s \geq 1$, we can take $Z = s$, $z = \bar{x}$, $\lambda = 1$, which has objective value zero. This shows that there are distributions with the specified mean and variance for which the righthand side of (6) is equal to $\mathbf{Prob}(X \in C)$.

For example, if $\bar{x} = \mathbf{E}\, X = 0.4$ and $s = \mathbf{E}\, X^2 = 0.2$ we get $\mathbf{Prob}(|X| < 1) \geq 0.9$, and the bound is achieved by the distribution

$$X = \begin{cases} 1 & \text{with probability } 0.1 \\ 1/3 & \text{with probability } 0.9. \end{cases}$$

This is illustrated in figure 2.

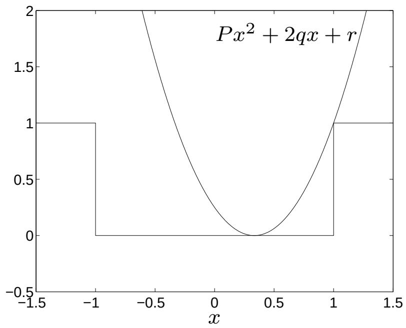

Fig. 2. Primal and dual SDP solution for $C = \{x \mid x^2 < 1\}$, $\mathbf{E}\,X = 0.4$ and $\mathbf{E}\,X^2 = 0.2$.

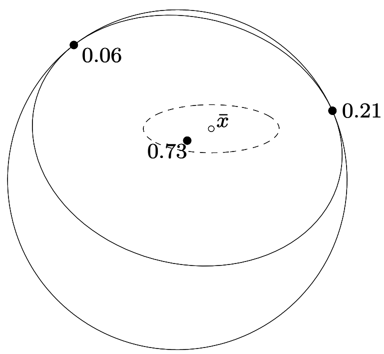

Fig. 3. Geometrical interpretation of the Chebyshev bound for $C = \{x \in \mathbf{R}^2 \mid x^T x < 1\}$.

Figure 3 illustrates the extension of the bound (6) to a unit ball $C = \{x \in \mathbf{R}^n \mid x^T x < 1\}$, for

$$\bar{x} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix}, \qquad S = \begin{bmatrix} 0.20 & 0.06 \\ 0.06 & 0.11 \end{bmatrix}$$

The optimal bound is $\mathbf{Prob}(X \in C) \geq 0.73$ and is achieved by the discrete distribution with three possible values shown in the figure.

5. BOUNDING MANUFACTURING YIELD

The yield of a manufacturing process can be expressed as

$$Y(x) = \mathbf{Prob}(x + w \in C),$$

where x denotes the nominal or target values of a set of design parameters, $w \in \mathbf{R}^n$ is a random vector that represents variations in the manufacturing process, and $C \subseteq \mathbf{R}^n$ denotes the set of acceptable parameter values for the product. If the set C is defined by quadratic inequalities, we can compute a lower bound on the yield $Y(x)$, valid for all distributions of w with given mean and covariance, by solving the SDPs (4) and (5). Figure 4 shows an example for a polyhedral set C (shown with a dashed line).

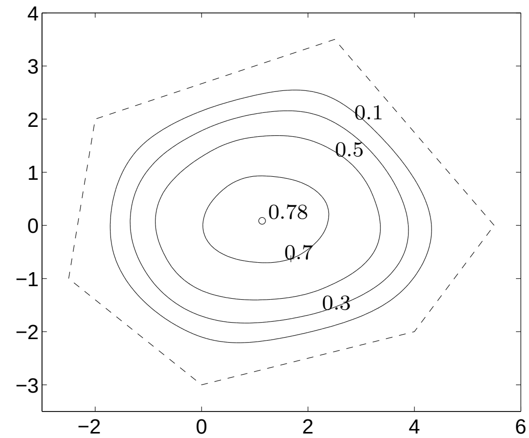

Fig. 4. A few contour lines of the Chebyshev lower bound on $Y(x) = \mathbf{Prob}(x + w) \in C$ for $\mathbf{E}\,w = 0$, $\mathbf{E}\,ww^T = I$.

6. PROBABILITY OF DETECTION ERROR

Consider a signal constellation of m possible symbols or signals $s \in \{s_1, s_2, \ldots, s_m\}$. One of the symbols is transmitted over a noisy channel. The received signal is $x = s + v$, where v is a noise vector with $\mathbf{E}\,v = 0$ and $\mathbf{E}\,vv^T = \sigma^2 I$. The receiver then estimates s based on the received x.

The minimum distance detector chooses the symbol s_k closest (in Euclidean norm) to x, so s_k is detected correctly if $x = s_k + v$ is closer to s_k than to any of the other symbols. The set of values of the random variable x for which symbol s_k is correctly detected, is therefore a polyhedron C_k (the Voronoi region of s_k in the constellation). Generalized Chebyshev bounds can be used to provide lower bounds on the probability of correct detection $\mathbf{Prob}(s_k + v \in C_k)$, valid for any noise distribution with zero mean and covariance $\sigma^2 I$. An example is given in (Boyd and Vandenberghe, 2004, page 381).

As an extension we can consider constellations with unequal noise covariances, *i.e.*, situations in which the noise covariance is $\mathbf{E}\,vv^T = \Sigma_k$ when symbol s_k is transmitted. Suppose the detector chooses the symbol s_k with the minimum Mahalanobis distance

$$\left((x - s_k)^T \Sigma_k^{-1} (x - s_k)\right)^{1/2}.$$

(This is also the maximum likelihood detector if the noise is Gaussian.) Then the region of correct detection of symbol signal s_k is a set defined by the $m - 1$ quadratic inequalities

$$\begin{aligned} &(x - s_k)^T \Sigma_k^{-1} (x - s_k) \\ &< (x - s_j)^T \Sigma_j^{-1} (x - s_j), \quad j \neq k. \end{aligned} \tag{7}$$

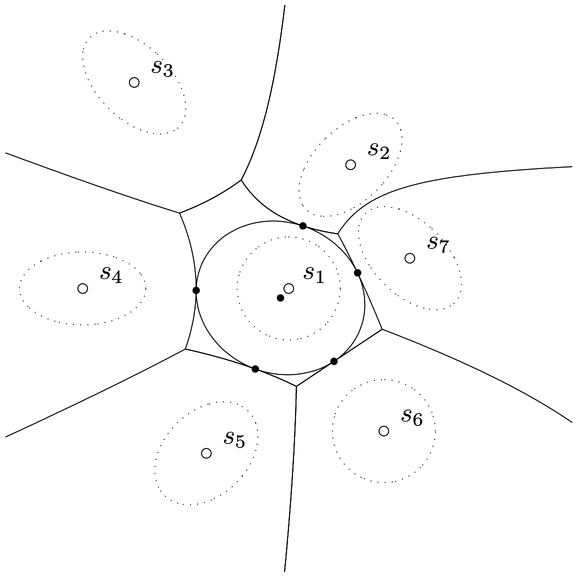

Fig. 5. Detection example.

Tight lower bounds on the probability of correct detection can be computed by solving the SDP (4).

Figure 5 shows an example with $m = 7$ symbols in $\mathbf{R}^2$, shown as circles. The dashed ellipses are defined as $\{x \mid (x - s_k)^T \Sigma_k^{-1}(x - s_k) = 1\}$. The solid lines show the boundaries of the regions of correct detection for each symbol, as defined by the quadratic inequalities (7). The figure also illustrates the optimal SDP solution for s_1. The solid ellipse is defined by $x^T Px + q^T x + r = 1$, for the optimal values P, q, r in (4). From the dual SDP we can construct the worst-case distribution (*i.e.*, with the highest probability of error) that matches the specified noise covariance. The worst-case distribution is the discrete distribution indicated by the six solid dots.

7. CONCLUSION

We have discussed a multivariate extension of Chebyshev's inequality that can be efficiently computed via semidefinite programming. The result follows from more general results of Isii (1964) and Bertsimas and Popescu (2005), and can also be proved directly from semidefinite programming duality. The bounds obtained are the best possible, over all distributions with given first and second order moments. From the optimal solution of the SDPs, the worst-case distribution can be established. We have also described some applications in detection theory and design centering.

REFERENCES

D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM J. on Optimization*, 15(3):780–804, 2005.

D. Bertsimas and J. Sethuraman. Moment problems and semidefinite optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, chapter 16, pages 469–510. Kluwer Academic Publishers, 2000.

S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994. ISBN 0-89871-334-X.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. `www.stanford.edu/~boyd/cvxbook`.

K. Isii. On a method for generalizations of Tchebycheff's inequality. *Annals of the Institute of Statistical Mathematics*, 10:65–88, 1959.

K. Isii. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14:185–197, 1963.

K. Isii. Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming. *Annals of The Institute of Statistical Mathematics*, 16:277–293, 1964.

S. Karlin and W. J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley-Interscience, 1966.

J. B. Lasserre. Bounds on measures satisfying moment conditions. *The Annals of Applied Probability*, 12(3):1114–1137, 2002.

A. W. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *Annals of the Mathematical Statistics*, 31:1001–1014, 1960.

I. Popescu. A semidefinite programming approach to optimal moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 50(3):632–657, 2005.

L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *To appear in SIAM Review*, 2006.