Generalized Chebyshev Inequalities and Semidefinite Programming

Lieven Vandenberghe

Electrical Engineering Department
University of California Los Angeles

joint work with Stephen Boyd, Stanford University,
Katherine Comanor, UCLA

MSRI 10/9/2002

Outline

- probability bounds via SDP
- proof from SDP duality
- geometrical interpretation
- examples and applications

Generalized Chebyshev inequalities

lower bounds on

$$\mathbf{Prob}(X \in C)$$

- $X \in \mathbf{R}^n$ is a random variable with $\mathbf{E}\,X = a$, $\mathbf{E}\,XX^T = S$
- $C \subseteq \mathbf{R}^n$ is defined by quadratic inequalities

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

cf. the classical Chebyshev inequality on $\mathbf{R}$

$$\mathbf{Prob}(X < 1) \geq \frac{1}{1 + \sigma^2}$$

if $\mathbf{E}\,X = 0$, $\mathbf{E}\,X^2 = \sigma^2$

Probability bound via SDP

$$
\begin{array}{ll}
\text{minimize} & 1 - \sum_{i=1}^m \lambda_i \\
\text{subject to} & \mathbf{Tr}\, A_i Z_i + 2b_i^T z_i + c_i \lambda_i \geq 0, \quad i = 1, \ldots, m \\
& \sum_{i=1}^m \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array} \right] \preceq \left[\begin{array}{cc} S & a \\ a^T & 1 \end{array} \right] \\
& \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array} \right] \succeq 0, \quad i = 1, \ldots, m
\end{array}
$$

- an SDP with variables $Z_i \in \mathbf{S}^n$, $z_i \in \mathbf{R}^n$, $\lambda_i \in \mathbf{R}$
- optimal value is a sharp lower bound on $\mathbf{Prob}(X \in C)$
- can construct a distribution with $\mathbf{E}\, X = a$, $\mathbf{E}\, XX^T = S$ that attains the lower bound

The dual SDP

$$\begin{array}{ll} \text{maximize} & 1 - \mathbf{Tr}\, SP - 2a^T q - r \\ \text{subject to} & \left[\begin{array}{cc} P & q \\ q^T & r-1 \end{array}\right] \succeq \tau_i \left[\begin{array}{cc} A_i & b_i \\ b_i^T & c_i \end{array}\right], \quad i = 1, \ldots, m \\ & \tau_i \geq 0, \quad i = 1, \ldots, m \\ & \left[\begin{array}{cc} P & q \\ q^T & r \end{array}\right] \succeq 0 \end{array}$$

- variables $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$, $\tau \in \mathbf{R}^m$
- optimal value is (the same) sharp lower bound on $\mathbf{Prob}(X \in C)$

Proof

classical proof: combine results derived in the 60s (by Isii, Marshall & Olkin, Karlin & Studden) with the S-procedure

SDP duality based proof

- dual SDP: maximizes a lower bound on $\mathbf{Prob}(X \in C)$, valid for all distributions with $\mathbf{E}\, X = a$, $\mathbf{E}\, XX^T = S$
- primal SDP: minimizes $\mathbf{Prob}(X \in C)$ over a set of discrete distributions with $\mathbf{E}\, X = a$, $\mathbf{E}\, XX^T = S$
- by strong duality, the optimal values are equal

Interpretation of dual SDP

dual feasibility: $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$, $\tau \in \mathbf{R}^m$ satisfy

$$\begin{bmatrix} P & q \\ q^T & r \end{bmatrix} \succeq 0, \quad \begin{bmatrix} P & q \\ q^T & r-1 \end{bmatrix} \succeq \tau_i \begin{bmatrix} A_i & b_i \\ b_i^T & c_i \end{bmatrix}, \quad \tau_i \geq 0, \quad i = 1, \ldots, m$$

interpretation: $f(x) = x^T P x + 2q^T x + r$ satisfies $f(x) \geq 0$ and

$$x \notin C \Longrightarrow x^T A_i x + 2b_i^T x + c_i \geq 0 \text{ for some } i \;\Longrightarrow f(x) \geq 1$$

therefore $\mathbf{Prob}(X \notin C) \leq \mathbf{E}\, f(X) = \mathbf{Tr}\, SP + 2a^T q + r$

$$\mathbf{Prob}(X \in C) \geq 1 - \mathbf{Tr}\, SP - 2a^T q - r$$

the dual SDP maximizes this lower bound

A result from linear algebra

if $Z \in \mathbf{S}^n$, $z \in \mathbf{R}^n$ satisfy

$$Z \succeq zz^T, \qquad \mathbf{Tr}\, AZ + 2b^T z + c \geq 0$$

then there exist $v_1, \ldots, v_K \in \mathbf{R}^n$, $\alpha_1, \ldots, \alpha_K \geq 0$ such that

$$v_i^T A v_i + 2b_i^T z + c_i \geq 0, \quad \sum_{i=1}^K \alpha_i = 1, \quad \sum_{i=1}^K \alpha_i v_i = z, \quad \sum_{i=1}^K \alpha_i v_i v_i^T \preceq Z$$

interpretation (with $z = \mathbf{E}\, X$, $Z = \mathbf{E}\, XX^T$): if

$$\mathbf{E}(X^T A X + 2b^T X + c) \geq 0$$

then there is a discrete random variable Y with

$$Y^T A Y + 2b^T Y + c \geq 0, \qquad \mathbf{E}\, Y = \mathbf{E}\, X, \qquad \mathbf{E}\, YY^T \preceq \mathbf{E}\, XX^T$$

constructive proof

- if $z^TAz + 2b^Tz + c \geq 0$, choose $K = 1$, $v_1 = z$, $\alpha_1 = 1$

- if $\lambda = z^TAz + 2b^Tz + c < 0$, define w_i, μ_i as

$$\sum_{i=1}^{n} w_i w_i^T = Z - zz^T, \qquad \mu_i = w_i^T A w_i$$

  with $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0 \geq \mu_{r+1} \geq \cdots \geq \mu_n$

  choose $K = 2r$, and for $i = 1, \ldots, r$,

$$\begin{array}{rclrcl} v_i & = & z + \beta_i w_i & \alpha_i & = & \mu_i / ((1 - \beta_i/\beta_{i+r})(\sum_{i=1}^{r} \mu_i)) \\ v_{i+r} & = & z + \beta_{i+r} w_i & \alpha_{i+r} & = & -\alpha_i \beta_i / \beta_{i+r} \end{array}$$

  where β_i, β_{i+r} are the two roots of

$$\mu_i \beta^2 + 2w_i^T(Az + b)\beta + \lambda = 0$$

Interpretation of primal feasibility

$Z_i \in \mathbf{S}^n$, $z_i \in \mathbf{R}^n$, $\lambda_i \in \mathbf{R}$ satisfy

$$\mathbf{Tr}\, A_i Z_i + 2b_i^T z_i + c_i \lambda_i \geq 0, \quad \begin{bmatrix} Z_i & z_i \\ z_i^T & \lambda_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m \qquad (1)$$

$$\sum_{i=1}^{m} \begin{bmatrix} Z_i & z_i \\ z_i^T & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} S & a \\ a^T & 1 \end{bmatrix} \qquad (2)$$

- from (1): if $\lambda_i > 0$, can construct a random variable Y_i with

$$\mathbf{E}\, Y_i = z_i/\lambda_i, \qquad \mathbf{E}\, Y_i Y_i^T \preceq Z_i/\lambda_i, \qquad Y_i^T A_i Y_i + 2b_i^T Y_i + c_i \geq 0$$

- from (2): define X with $\mathbf{E}\, X = a$, $\mathbf{E}\, X X^T = S$

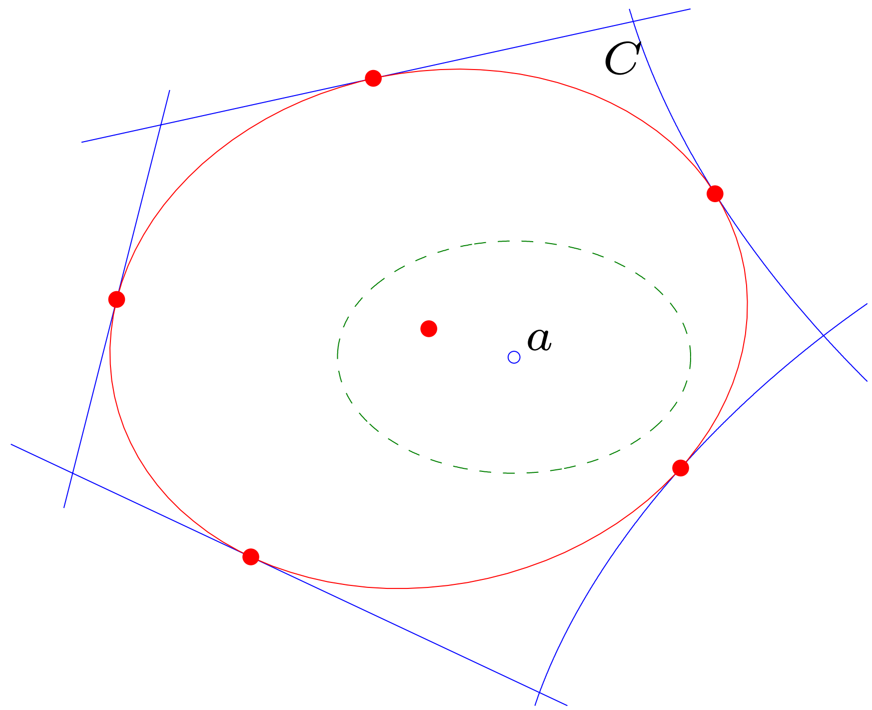
$$X = Y_i \text{ with probability } \lambda_i \quad \Longrightarrow \quad \mathbf{Prob}(X \notin C) \geq \sum_{i=1}^{m} \lambda_i$$

Interpretation of primal SDP

$$
\begin{array}{ll}
\text{minimize} & 1 - \sum_{i=1}^m \lambda_i \\
\text{subject to} & \mathbf{Tr}\, A_i Z_i + 2 b_i^T z_i + c_i \lambda_i \geq 0, \quad i = 1, \ldots, m \\
& \sum_{i=1}^m \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array} \right] \preceq \left[\begin{array}{cc} S & a \\ a^T & 1 \end{array} \right] \\
& \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array} \right] \succeq 0, \quad i = 1, \ldots, m
\end{array}
$$

interpretation: minimize $\mathbf{Prob}(X \in C)$ over discrete distributions that satisfy $\mathbf{E}\, X = a$, $\mathbf{E}\, X X^T = S$

Complementary slackness



- $a = \mathbf{E}\, X$; dashed line shows $\{x \mid (x-a)^T(S-aa^T)^{-1}(x-a)=1\}$
- lower bound on $\mathbf{Prob}(X \in C)$ is 0.3992, achieved by distribution shown in red
- ellipse is defined by $x^TPx + 2q^Tx + r = 1$

Geometrical interpretation of dual problem

for $a = 0$, $S = I$, dual problem is equivalent to

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, P + x_c^T P x_c \\ \text{subject to} & \mathcal{E} \subseteq \mathbf{cl}\, C \\ & P \succeq 0 \end{array}$$

where $\mathcal{E}$ is the ellipsoid

$$\mathcal{E} = \{x \mid (x - x_c)^T P (x - x_c) \leq 1\}$$

an extremal ellipsoid enclosed in a (possibly nonconvex) set C

Two-sided Chebyshev inequality

$$C = (-1, 1) = \{x \in \mathbf{R} \mid x^2 < 1\}, \qquad \mathbf{E}\, X = a, \qquad \mathbf{E}\, X^2 = s$$

primal SDP (variables λ, Z, $z \in \mathbf{R}$)

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

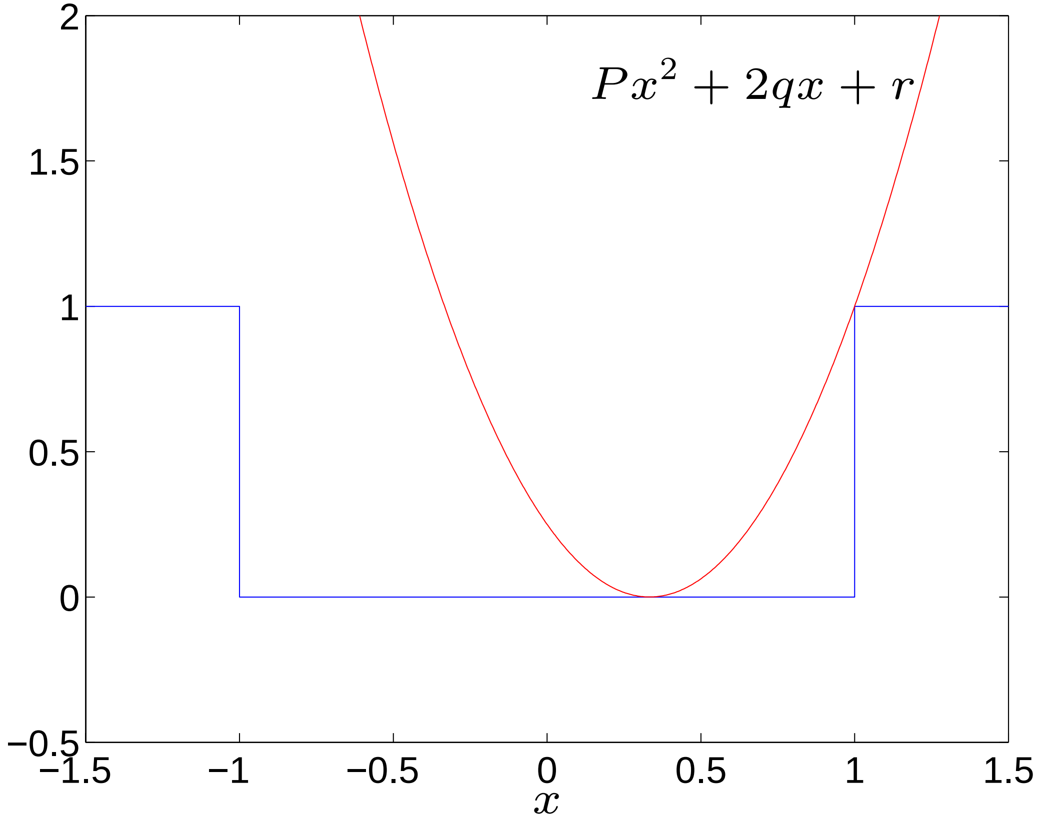
optimal value

$$\inf \mathbf{Prob}(X^2 < 1) = \begin{cases} 0 & 1 \leq s \\ 1 - s & |a| \leq s < 1 \\ (1 - |a|)^2/(s - 2|a| + 1) & s < |a| \end{cases}$$

reduces to two-sided Chebyshev inequality if $a = 0$

example: $a = \mathbf{E}\,X = 0.4$, $s = \mathbf{E}\,X^2 = 0.2$: $\mathbf{Prob}(X^2 < 1) \geq 0.9$

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



Extension to $\mathbf{R}^n$

$$C = \{x \in \mathbf{R}^n \mid x^T x < 1\}, \qquad a = \mathbf{E}\, X, \qquad S = \mathbf{E}\, X X^T$$

primal SDP (variables λ, $Z \in \mathbf{S}^n$, $z \in \mathbf{R}^n$)

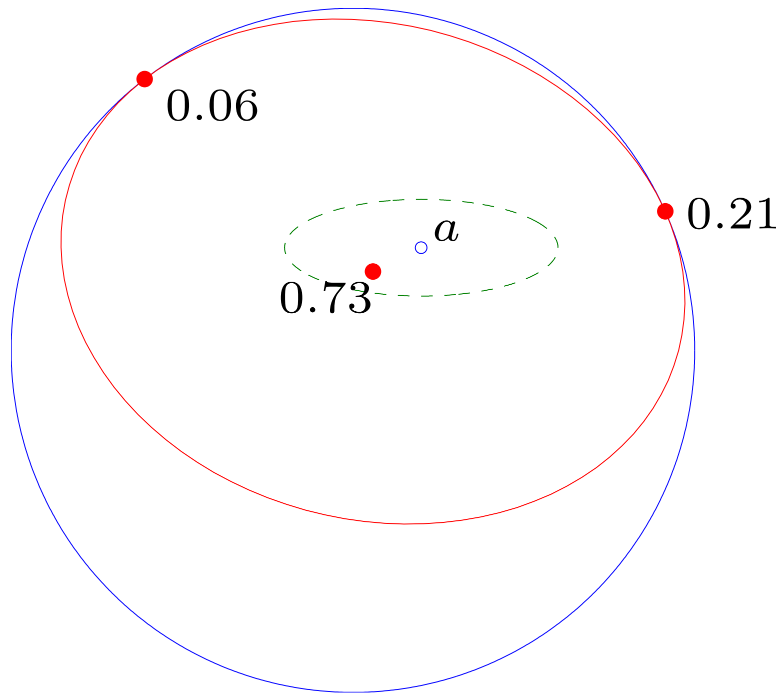
$$\begin{array}{ll} \text{minimize} & 1 - \lambda \\ \text{subject to} & \mathbf{Tr}\, Z \geq \lambda \\ & 0 \preceq \left[\begin{array}{cc} Z & z \\ z & \lambda \end{array}\right] \preceq \left[\begin{array}{cc} S & a \\ a^T & 1 \end{array}\right] \end{array}$$

dual SDP (variables $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$, $\tau \in \mathbf{R}$)

$$\begin{array}{ll} \text{maximize} & 1 - \mathbf{Tr}\, SP - 2a^T q - r \\ \text{subject to} & \left[\begin{array}{cc} P - \tau I & q \\ q^T & r + \tau - 1 \end{array}\right] \succeq 0, \qquad \left[\begin{array}{cc} P & q \\ q^T & r \end{array}\right] \succeq 0, \qquad \tau \geq 0 \end{array}$$

example

$$C = \{x \in \mathbf{R}^n \mid x^T x < 1\}, \qquad a = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix}, \qquad S = \begin{bmatrix} 0.20 & 0.06 \\ 0.06 & 0.11 \end{bmatrix}$$



distribution achieves lower bound $\mathbf{Prob}(X^T X < 1) \geq 0.73$

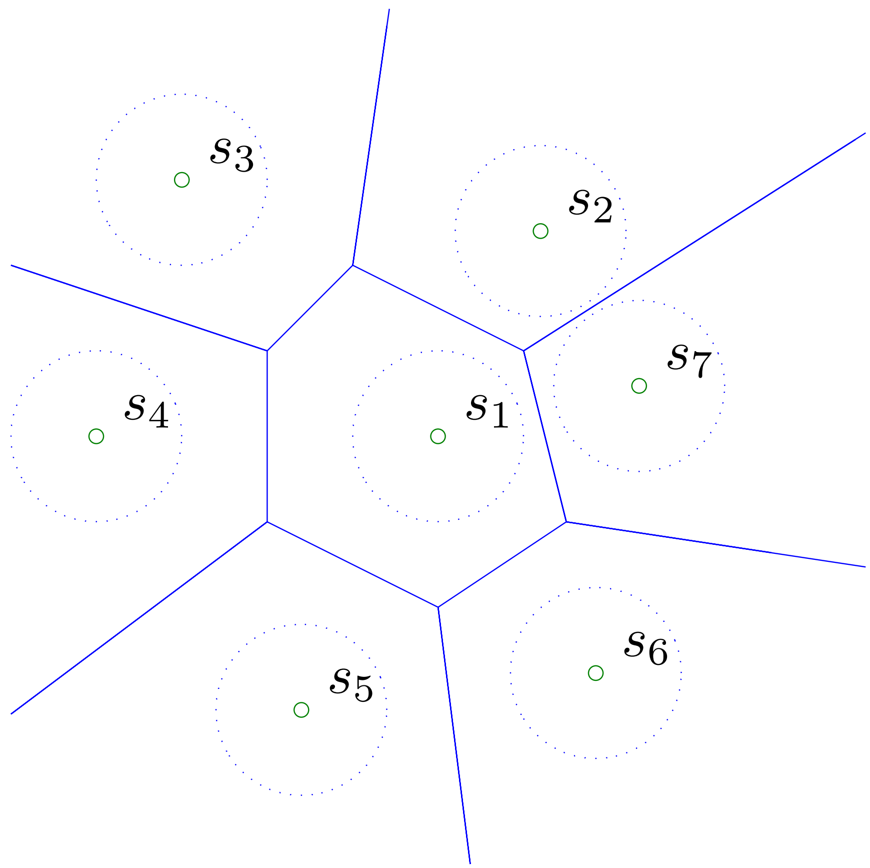
Detection example

$$x = s + v$$

- $x \in \mathbf{R}^n$: received signal
- s: transmitted signal $s \in \{s_1, s_2, \ldots, s_N\}$ (one of N possible symbols)
- v: noise with $\mathbf{E}\, v = 0$, $\mathbf{E}\, vv^T = \sigma^2 I$

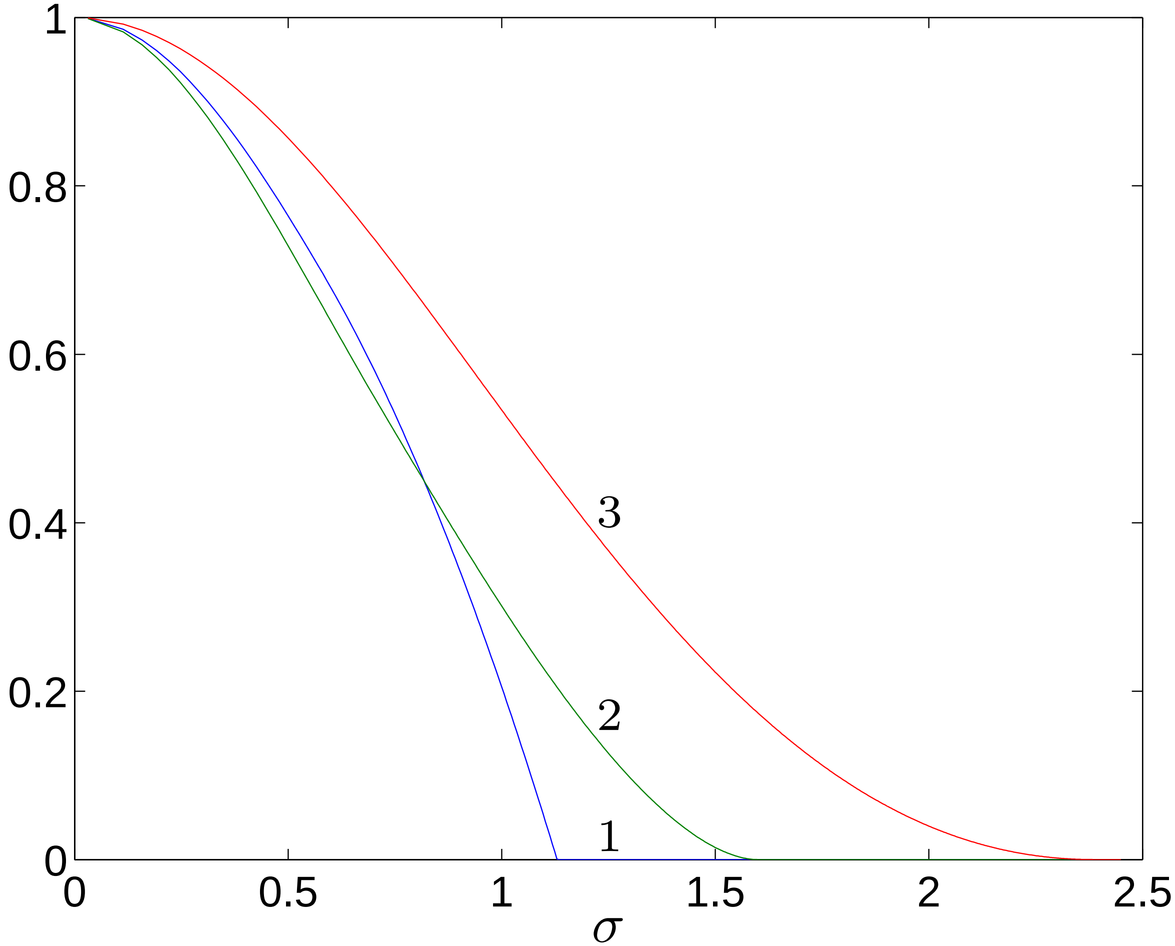
detection problem: given observed value of x, estimate s

example $(n = 2,\ N = 7)$

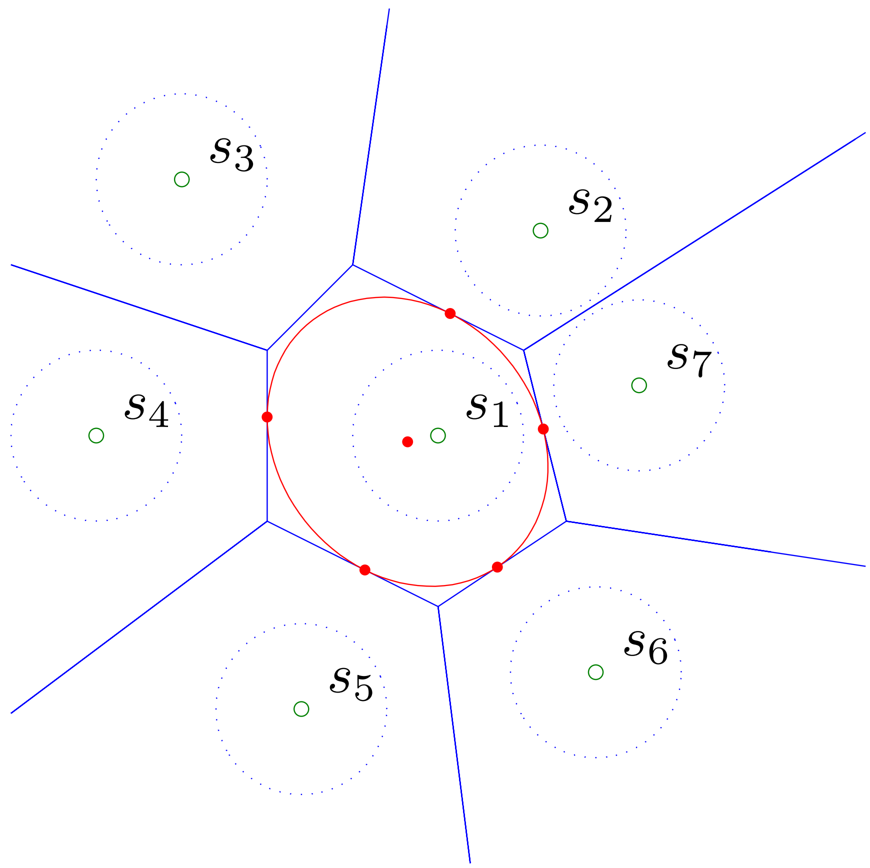


- detector selects symbol s_k closest to received signal x
- correct detection if $s_k + v$ lies in the Voronoi region around s_k

SDP lower bounds on probability of correct detection of s_1, s_2, s_3



example ($\sigma = 1$): bound on probability of correct detection of s_1 is 0.205



- solid circles: distribution with probability of correct detection 0.205
- ellipse is defined by $x^T P x + 2q^T x + r = 1$

Detection with unequal noise covariances

$$x = s + v$$

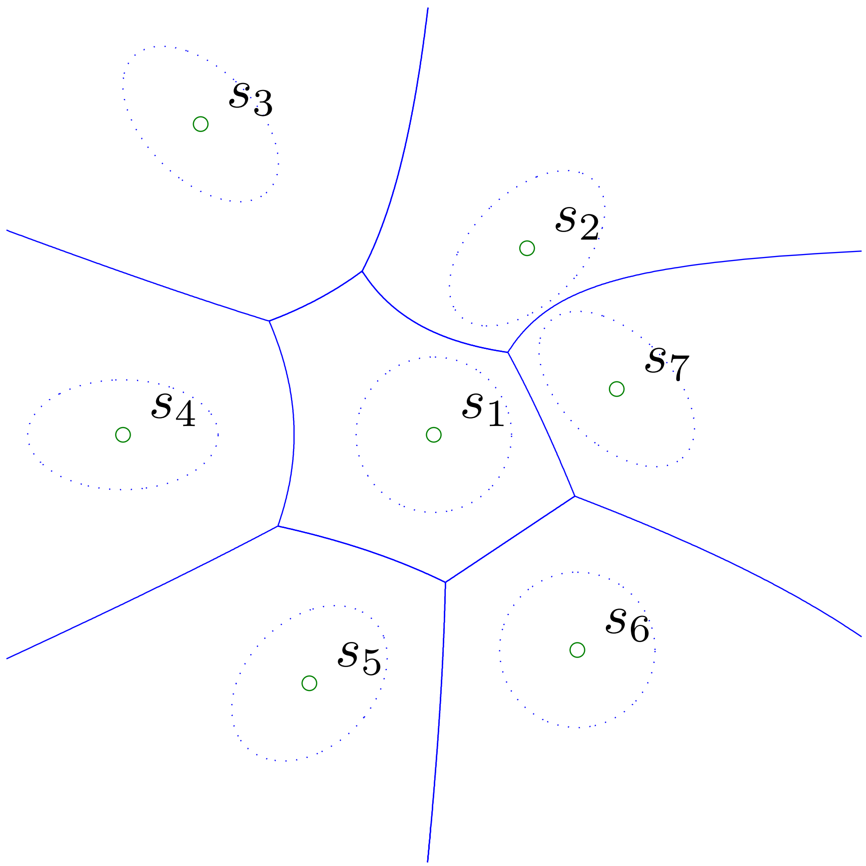
- $x \in \mathbf{R}^n$: received signal
- transmitted signal $s \in \{s_1, s_2, \ldots, s_N\}$
- v: noise with $\mathbf{E}\, v = 0$, $\mathbf{E}\, vv^T = \Sigma_k$ if symbol s_k was sent

detector: given observed value of x, choose s_k if

$$(x - s_k)^T \Sigma_k^{-1} (x - s_k) < (x - s_j)^T \Sigma_j^{-1} (x - s_j), \qquad j \neq k$$

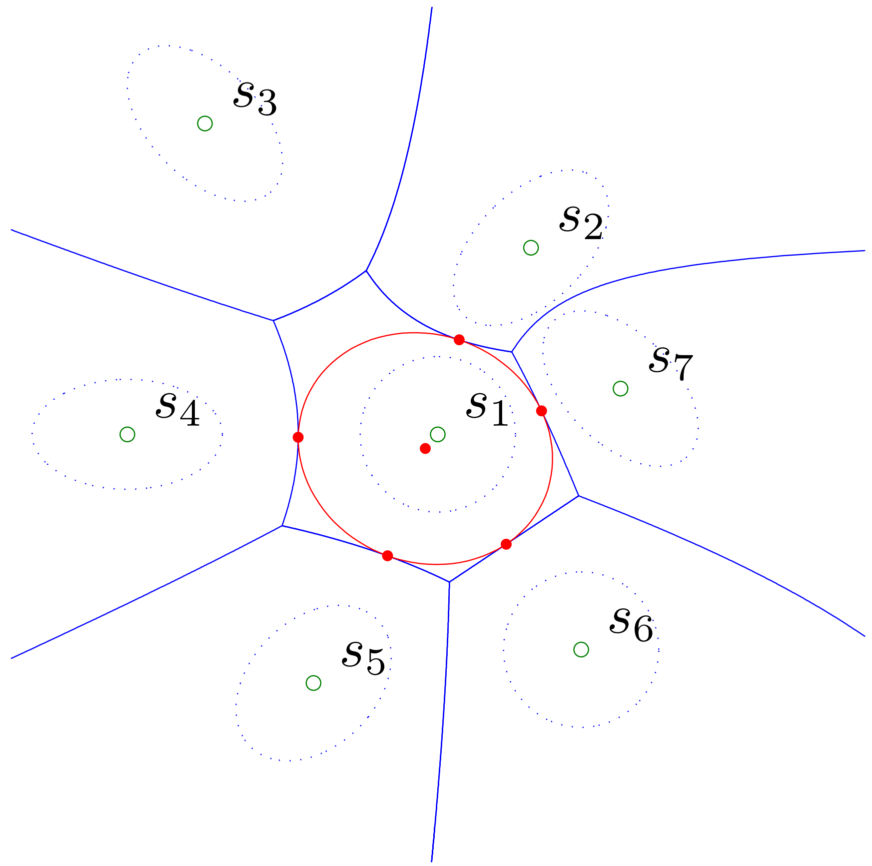
- a set of $N - 1$ indefinite quadratic inequalities
- maximum-likelihood detector if v is Gaussian

example ($n = 2$, $m = 7$)



dashed ellipses are the sets $\{x \mid (x - s_k)^T \Sigma_k^{-1} (x - s_k) = 1\}$

lower bound on probability of correct detection of s_1 is 0.145



- solid circles: distribution with probability of correct detection 0.145
- ellipse is defined by $x^T P x + 2q^T x + r = 1$

Hypothesis testing based on moments

based on observed value of $X \in \mathbf{R}^n$, choose one of two hypotheses:

1. $\mathbf{E}\,X = a_1$, $\mathbf{E}\,XX^T = S_1$
2. $\mathbf{E}\,X = a_2$, $\mathbf{E}\,XX^T = S_2$

randomized detector: a function $t : \mathbf{R}^n \to [0,1]$; if we observe x, we choose hypothesis 1 with probability $t(x)$, hypothesis 2 with probability $1 - t(x)$

worst-case probability of error

1. false positive: $P_{\mathrm{fp}} = \sup\{\mathbf{E}\,t(X) \mid \mathbf{E}\,X = a_2, \ \mathbf{E}\,XX^T = S_2\}$
2. false negative: $P_{\mathrm{fn}} = \sup\{1 - \mathbf{E}\,t(X) \mid \mathbf{E}\,X = a_1, \ \mathbf{E}\,XX^T = S_1\}$

minimax detector: t that minimizes $\max\{P_{\mathrm{fp}}, P_{\mathrm{fn}}\}$

upper bounds on $P_{\rm fp}$, $P_{\rm fn}$: suppose

$$f_1(x) = x^T P_1 x + 2q_1^T x + r_1, \qquad f_2(x) = x^T P_2 x + 2q_2^T x + r_2$$

satisfy $f_1(x) \leq t(x) \leq f_2(x)$

$$\begin{aligned} P_{\rm fp} &= \sup\{\mathbf{E}\, t(X) \mid \mathbf{E}\, X = a_2,\ \mathbf{E}\, XX^T = S_2\} \\ &\leq \mathbf{Tr}\, S_2 P_2 + 2a_2^T q_2 + r_2 \\ P_{\rm fn} &= \sup\{1 - \mathbf{E}\, t(X) \mid \mathbf{E}\, X = a_1,\ \mathbf{E}\, XX^T = S_1\} \\ &\leq 1 - \mathbf{Tr}\, S_1 P_1 - 2a_1^T q_1 - r_1 \end{aligned}$$

minimax detector design (variables $t(x)$, P_1, P_2, q_1, q_2, r_1, r_2)

$$\begin{array}{ll} \text{minimize} & \max\{\mathbf{Tr}\, S_2 P_2 + 2a_2^T q_2 + r_2, 1 - \mathbf{Tr}\, S_1 P_1 - 2a_1^T q_1 - r_1\} \\ \text{subject to} & x^T P_1 x + 2q_1^T x + r_1 \leq t(x) \leq x^T P_2 x + 2q_2^T x + r_2 \\ & 0 \leq t(x) \leq 1 \end{array}$$

after eliminating t:

$$\begin{array}{ll} \text{minimize} & \max\{\mathbf{Tr}\, S_2P_2 + 2a_2^T q_2 + r_2, 1 - \mathbf{Tr}\, S_1P_1 - 2a_1^T q_1 - r_1\} \\ \text{subject to} & x^T P_1 x + 2q_1^T x + r_1 \le x^T P_2 x + 2q_2^T x + r_2 \\ & x^T P_1 x + 2q_1^T x + r_1 \le 1 \\ & x^T P_2 x + 2q_2^T x + r_2 \ge 0 \end{array}$$

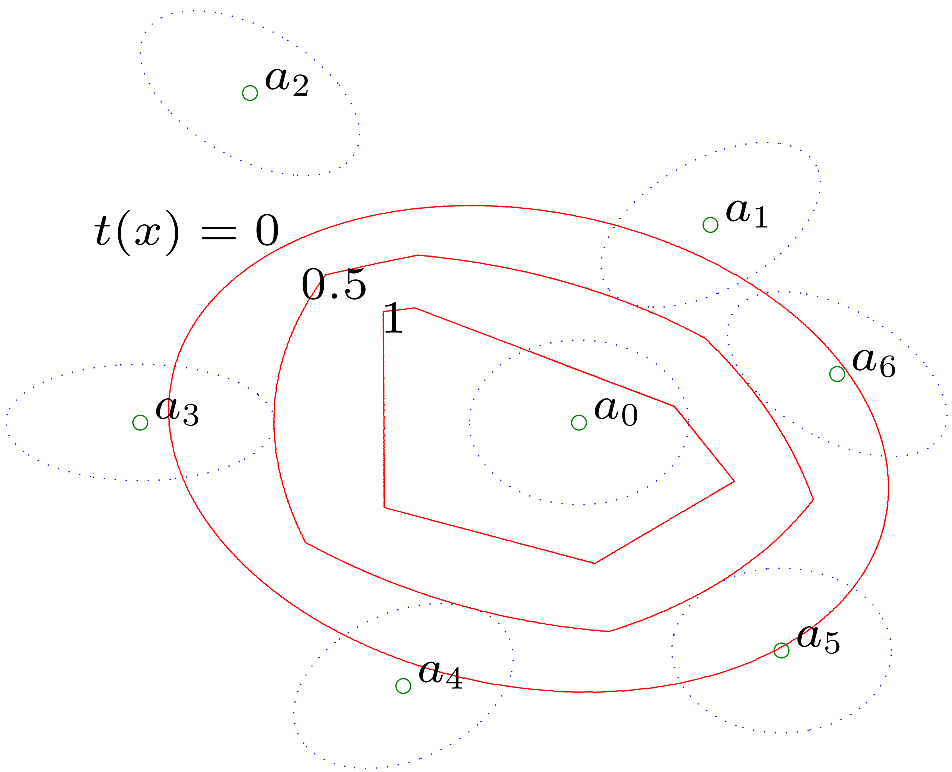
and choose t such that $\max\{0, f_1(x)\} \le t(x) \le \min\{1, f_2(x)\}$

an SDP with variables γ, P_1, P_2, q_1, q_2, r_1, r_2:

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \mathbf{Tr}\, S_2P_2 + 2a_2^T q_2 + r_2 \le \gamma \\ & 1 - \mathbf{Tr}\, S_1P_1 - 2a_1^T q_1 - r_1 \le \gamma \\ & \left[\begin{array}{cc} P_2 - P_1 & q_2 - q_1 \\ (q_2 - q_1)^T & r_2 - r_1 \end{array}\right] \succeq 0 \\ & \left[\begin{array}{cc} P_1 & q_1 \\ q_1^T & r_1 - 1 \end{array}\right] \preceq 0, \qquad \left[\begin{array}{cc} P_2 & q_2 \\ q_2^T & r_2 \end{array}\right] \succeq 0 \end{array}$$

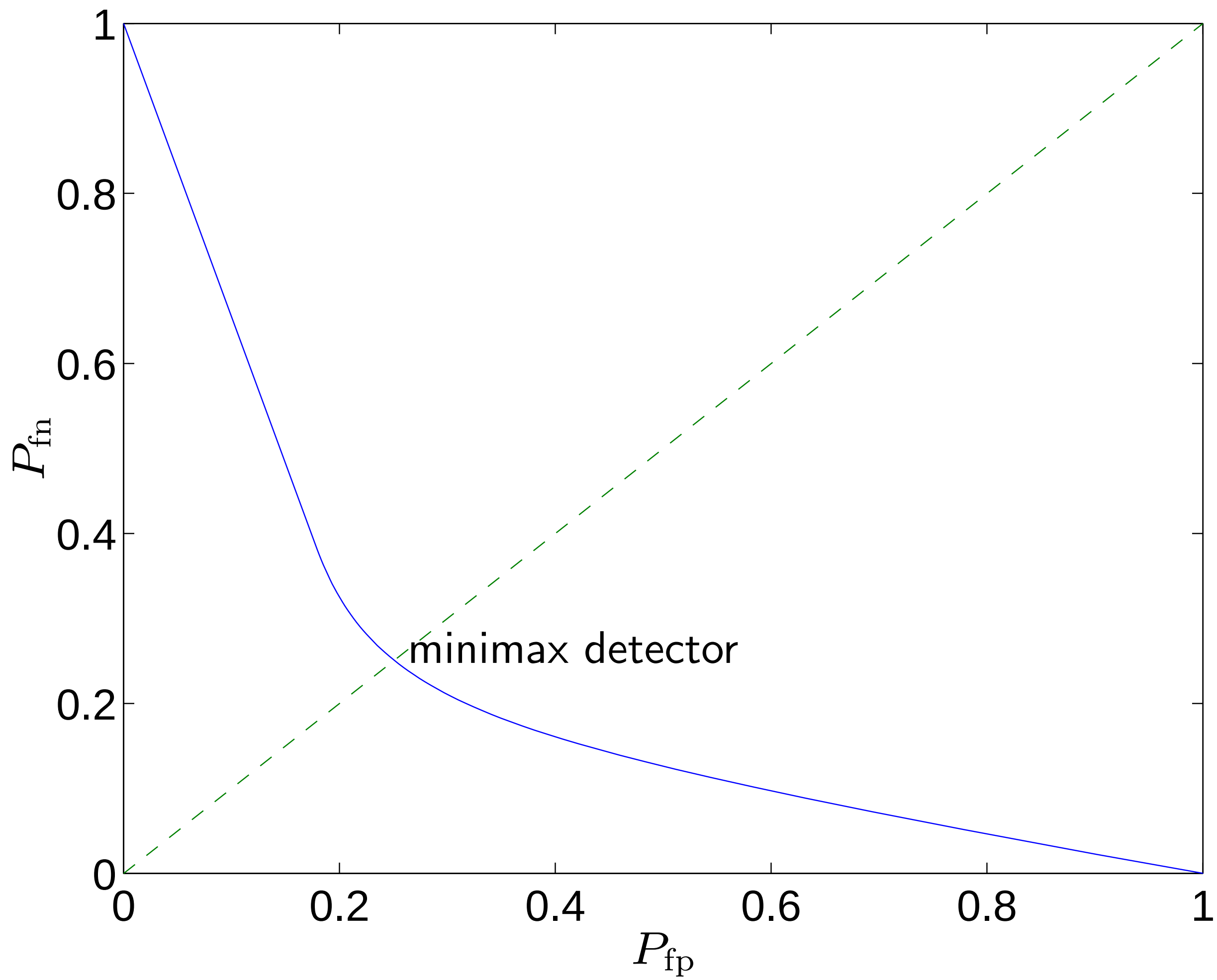
example: two hypotheses

1. $\mathbf{E}\,X = a_0$, $\mathbf{E}\,XX^T = S_0$
2. $(\mathbf{E}\,X, \mathbf{E}\,XX^T) \in \{(a_1, S_1), \ldots, (a_6, S_6)\}$



contour lines of a minimax detector $t(x)$ ($P_{\rm fn} = P_{\rm fp} = 0.251$)

trade-off curve between P_{fp} and P_{fn}



Bounding manufacturing yield

manufacturing yield

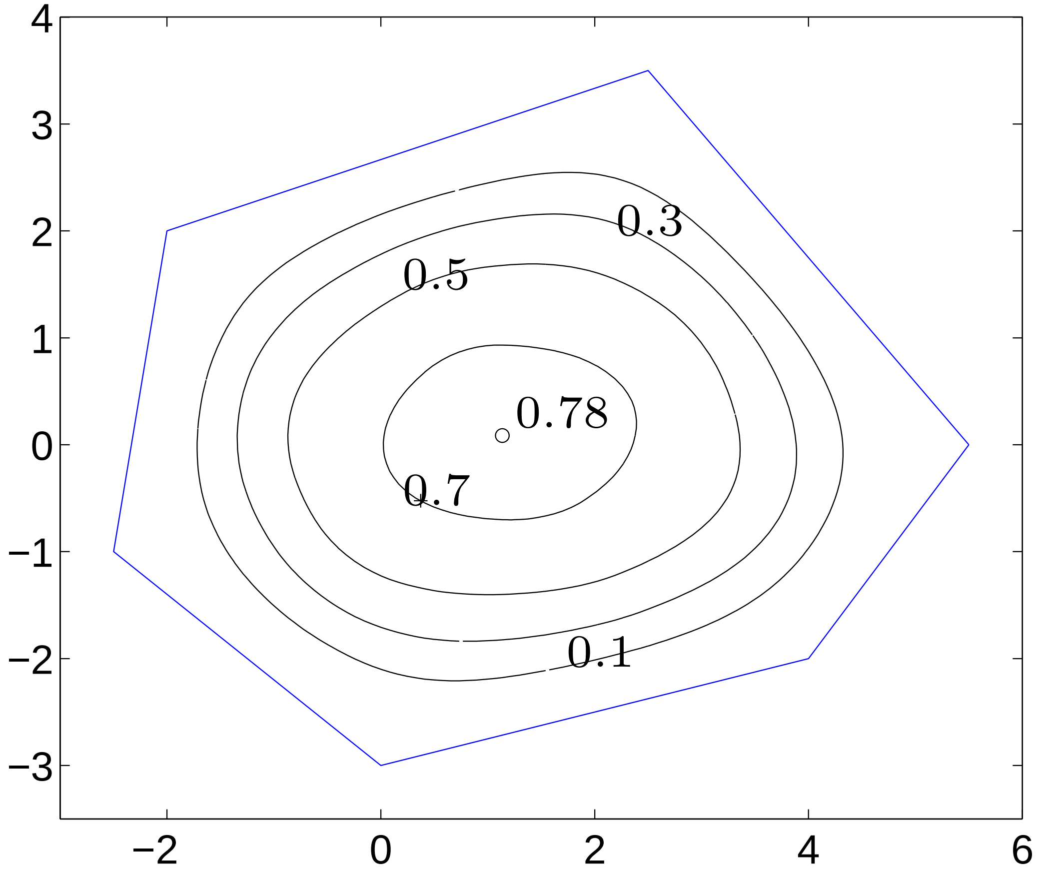
$$Y(a) = \mathbf{Prob}(a + w \in C)$$

- $a \in \mathbf{R}^n$: nominal or target value of design parameters
- $w \in \mathbf{R}^n$: manufacturing errors; zero mean random variable
- $C \subseteq \mathbf{R}^n$: specifications; set of acceptable values

lower bound on yield via SDP

- given $\mathbf{E}\, ww^T = \Sigma$
- C described by (possibly non-convex) quadratic inequalities

example $(\mathbf{E}\, ww^T = I)$



plot shows contour lines of lower bound on $Y(a) = \mathbf{Prob}(a + w \in C)$

Design centering

lower bound on yield $Y(a)$,

$$\inf\{\mathbf{Prob}(a+w \in C) \mid \mathbf{E}\, w = 0,\ \ \mathbf{E}\, ww^T = \Sigma\},$$

is the optimal value of

$$\begin{array}{ll} \text{maximize} & 1 - \mathbf{Tr}\,\Sigma P - a^T P a - 2a^T q - r \\ \text{subject to} & \left[\begin{array}{cc} P & q \\ q^T & r-1 \end{array}\right] \succeq \tau_i \left[\begin{array}{cc} A_i & b_i \\ b_i^T & c_i \end{array}\right], \quad i = 1, \ldots, m \\ & \tau_i \geq 0, \quad i = 1, \ldots, m \\ & \left[\begin{array}{cc} P & q \\ q^T & r \end{array}\right] \succeq 0 \end{array}$$

- for fixed a, an SDP in variables P, q, r, τ
- can alternate maximization over P, q, r, τ and maximization over a ($i.e.$, set $a = -P^{-1}q$)

Conclusion

- lower bounds on $\mathbf{Prob}(X \in C)$ where
  - $\mathbf{E}\,X$, $\mathbf{E}\,XX^T$ are given
  - C is defined by quadratic inequalities
- bounds are sharp; distribution that achieves may be unrealistic
- applications in classification and detection, design centering, . . .