

ELE539: Optimization of Communication Systems
Lecture 13: Estimation and Detection Problems

Professor M. Chiang
Electrical Engineering Department, Princeton University

March 3, 2004

Lecture Outline

- ML estimation by norm minimization
- Covariance estimation by SDP
- Distribution estimation by convex optimization
- Chebyshev bound by SDP
- Chernoff bound by QP
- Multiuser detection by SDP relaxation

Thanks to: Stephen Boyd (materials and graphs from Boyd and Vandenberghe's *Convex Optimization* book)

ML Estimation

Estimate $x \in \mathbf{R}^n$ from a set of measurements $y_i \in \mathbf{R}$:

$$y_i = a_i^T x + v_i, \quad i = 1, \dots, m$$

where v_i are i.i.d. noise with distribution p on \mathbf{R}

Likelihood function is

$$\prod_{i=1}^m p(y_i - a_i^T x)$$

Maximum Likelihood estimate is any optimal point for the problem in x :

$$\text{maximize } \sum_{i=1}^m \log p(y_i - a_i^T x)$$

Since many distributions are log-concave, the above problem are often convex optimization

Examples

- Gaussian noise: v_i are Gaussian with zero mean and variance σ^2 :

$$p(z) = (2\pi\sigma^2)^{-1/2} \exp(-z^2/2\sigma^2)$$

Log-likelihood function: $-(1/2) \log(2\pi\sigma) - (1/2\sigma^2) \|Ax - y\|_2^2$

$$\hat{x} = \operatorname{argmin}_x \|Ax - y\|_2$$

Solution of least-squares

- Laplacian noise: v_i are Laplacian: $p(z) = (1/2a) \exp(-|z|/a)$, $a > 0$

$$\hat{x} = \operatorname{argmin}_x \|Ax - y\|_1$$

Solution of $l - 1$ norm minimization

Covariance Estimation for Gaussian Distribution

$y \in \mathbf{R}^n$ is multivariate Gaussian with zero mean and covariance matrix $R = \mathbf{E} yy^T \in \mathbf{S}_{++}^n$ with distribution:

$$p_R(y) = (2\pi)^{-n/2} \det(R)^{-1/2} \exp(-y^T R^{-1} y/2)$$

Goal: estimate R based on N i.i.d. samples $y_1, \dots, y_N \in \mathbf{R}^n$

Sample covariance matrix

$$Y = \frac{1}{N} \sum_{k=1}^N y_k y_k^T$$

Log likelihood function $\log p_R(y_1, \dots, y_N)$

$$l(R) = -(Nn/2) \log(2\pi) - (N/2) \log \det R - (N/2) \mathbf{tr}(R^{-1} Y)$$

Information matrix $S = R^{-1}$, log likelihood function is **concave** in S :

$$l(S) = -(Nn/2) \log(2\pi) + (N/2) \log \det S - (N/2) \mathbf{tr}(SY)$$

Covariance Estimation for Gaussian Distribution

ML estimate of S found by solving convex optimization:

$$\begin{aligned} & \text{maximize} && \log \det S - \mathbf{tr}(SY) \\ & \text{subject to} && S \in \mathcal{S} \end{aligned}$$

where \mathcal{S} is a convex constraint set of S based on prior information about R

e.g., bounds on R : $L \preceq R \preceq U$ ($U^{-1} \preceq S \preceq L^{-1}$)

e.g., condition number constraint on R : $\lambda_{max}(R) \leq \kappa_{max} \lambda_{min}(R)$
($uI \preceq S \preceq \kappa_{max} uI$, variables: (S, u))

Distribution Estimation

Remove assumptions on parametrization of distribution

Random variable X with values in finite set $\{\alpha_1, \dots, \alpha_n\}$

Distribution characterized by $p \in \mathbf{R}^n$ that satisfies $p \succeq 0, \mathbf{1}^T p = 1$

Prior information about p from linear constraints, e.g.

$$\mathbf{E} X = \sum_{i=1}^n \alpha_i p_i = \alpha, \quad \mathbf{E} X^2 = \sum_{i=1}^n \alpha_i^2 p_i = \beta, \quad \sum_{i: \alpha_i \geq 0} p_i \leq 0.3$$

and from nonlinear constraints, e.g.

$$\mathbf{var}(X) = \sum_{i=1}^n \alpha_i^2 p_i - \left(\sum_{i=1}^n \alpha_i p_i \right)^2$$

which can be lower bounded and form a convex constraint

Objectives

Given convex constraints \mathcal{P} on p , can optimize for different criteria

- Maximum likelihood: Observe N independent samples, with k_i samples with value α_i

$$\begin{aligned} & \text{maximize} && \sum_{i=1}^n k_i \log p_i \\ & \text{subject to} && p \in \mathcal{P} \end{aligned}$$

- Maximum entropy:

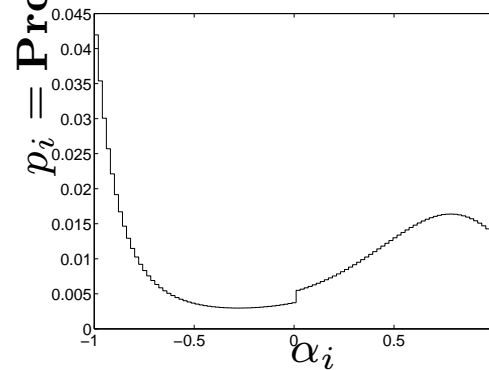
$$\begin{aligned} & \text{minimize} && \sum_{i=1}^n p_i \log p_i \\ & \text{subject to} && p \in \mathcal{P} \end{aligned}$$

- Minimum Kullback Leibler divergence with respect to another distribution q :

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^n p_i \log(p_i/q_i) \\ & \text{subject to} && p \in \mathcal{P} \end{aligned}$$

$p_i = \text{Prob}(X = \alpha_i)$

Example



Consider distribution on 100 equidistant points α_i in $[-1, 1]$ with the following prior information on p :

$$\mathbf{E} X \in [-0.1, 0.1]$$

$$\mathbf{E} X^2 \in [0.5, 0.6]$$

$$\mathbf{E}(3X^2 - 2X) \in [-0.3, -0.2]$$

$$\mathbf{Prob}(X < 0) \in [0.3, 0.4]$$

Maximal entropy distribution satisfying these prior information can be computed by convex optimization

Chebyshev Bound with Second Order Prior Information (Continue LP Lecture Note)

Given prior information $\mathbf{E} X = a \in \mathbf{R}^m$ and $\mathbf{E} X X^T = \Sigma \in \mathbf{S}^m$

Express f as a general quadratic function:

$$f(z) = z^T P z + 2q^T z + r$$

with variables $P \in \mathbf{S}^m$, $q \in \mathbf{R}^m$, $r \in \mathbf{R}$

Objective to be minimized:

$$\mathbf{E} f(X) = \mathbf{tr}(\Sigma P) + 2q^T a + r$$

Constraint $f(z) \geq 0$ expressed as LMI:

$$\begin{bmatrix} P & q \\ q^T & r \end{bmatrix} \succeq 0$$

Chebyshev Bound by SDP

Set C we are interested is complement of polyhedron \mathcal{P} :

$$C = \mathbf{R}^m - \mathcal{P}, \quad \mathcal{P} = \{z \mid a_i^T z < b_i, i = 1, \dots, k\}$$

Condition $f(z) \geq 1$ for all $z \in C$ is equivalent to

$$a_i^T z \geq b_i \Rightarrow z^T P z + 2q^T z + r \geq 1, \quad i = 1, \dots, k$$

which is equivalent to: there exist $\tau_1, \dots, \tau_k \geq 0$ such that

$$\begin{bmatrix} P & q \\ q^T & r - 1 \end{bmatrix} \succeq \tau_i \begin{bmatrix} 0 & a_i/2 \\ a_i^T/2 & -b_i \end{bmatrix}, \quad i = 1, \dots, k$$

Chebyshev bound problem as SDP in P, q, r, τ :

$$\begin{aligned}
 & \text{minimize} && \mathbf{tr}(\Sigma P) + 2q^T a + r \\
 & \text{subject to} && \begin{bmatrix} P & q \\ q^T & r - 1 \end{bmatrix} \preceq \tau_i \begin{bmatrix} 0 & a_i/2 \\ a_i^T/2 & -b_i \end{bmatrix}, \quad i = 1, \dots, k \\
 & && \begin{bmatrix} P & q \\ q^T & r \end{bmatrix} \preceq 0 \\
 & && \tau \succeq 0
 \end{aligned}$$

Optimal value is an upper bound on $\mathbf{Prob}(X \in C)$

Signal Detection Example

Set of m possible signals $s \in \{s_1, \dots, s_m\}$. One of them transmitted through a noisy channel:

$$x = s + v$$

where v is noise with zero mean and covariance matrix $\sigma^2 I$

Minimum distance detector chooses s_k closest to x in Euclidean distance (ML decoding if v is Gaussian)

Correct detection condition for detecting s_k :

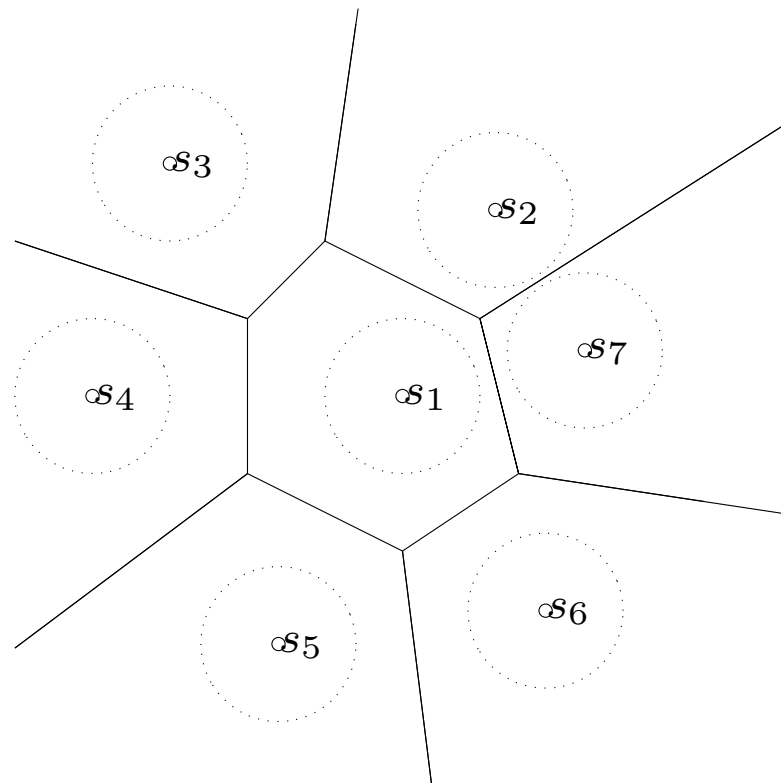
$$\|x - s_k\|_2 < \|x - s_j\|_2, \quad j \neq k$$

Voronoi region V_k of s_k :

$$2(s_j - s_k)^T (s_k + v) < \|s_j\|_2^2 - \|s_k\|_2^2, \quad j \neq k$$

Signal Detection Example

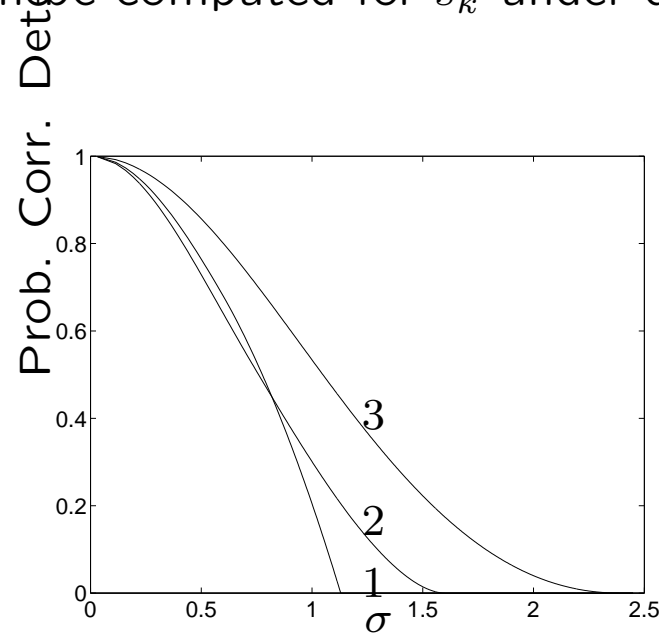
Signal constellation $s_i \in \mathbf{R}^2, i = 1, \dots, 7$, with Voronoi regions shown



Signal Detection Example

Probability of correct detection: $\mathbf{Prob}(s_k + v \in V_k)$ (a polyhedron)

Chebyshev bound can be computed for s_k under different σ^2 by SDP:



Upper bound P_e for all noise with 0 mean and $\sigma^2 I$ covariance

Chernoff Bounds

X is random variable on \mathbf{R} . Chernoff bound:

$$\mathbf{Prob}(X \geq u) \leq \inf_{\lambda \geq 0} \mathbf{E} \exp(\lambda(X - u))$$

which is equivalent to

$$\log \mathbf{Prob}(X \geq u) \leq \inf_{\lambda \geq 0} \{-\lambda u + \log \mathbf{E} \exp(\lambda X)\}$$

X : Gaussian with zero mean and unit variance, cumulant generating function is $\log \mathbf{E} \exp(\lambda X) = \frac{\lambda^2}{2}$, so Chernoff bound:

$$\mathbf{Prob}(X \geq u) \leq \exp\left(\frac{-u^2}{2}\right)$$

In general, for a set C in \mathbf{R}^m , we derive an upper bound on $\mathbf{Prob}(X \in C)$

Generalized Chernoff Bounds

Let $\lambda \in \mathbf{R}^m$ and $\mu \in \mathbf{R}$, $f : \mathbf{R}^m \rightarrow \mathbf{R}$:

$$f(z) = \exp(\lambda^T z + \mu)$$

If $f(z) \geq 1_C(z)$ for all z , then $\mathbf{Prob}(X \in C) \leq \mathbf{E} f(X)$

$f(z) \geq 0$ is always true. Need $f(z) \geq 1$ for $z \in C$:

$$-\lambda^T z \leq \mu, \forall z \in C, \Rightarrow \log \mathbf{Prob}(X \in C) \leq \mu + \log \mathbf{E} \exp(\lambda^T X)$$

Generalized Chernoff bound:

$$\log \mathbf{Prob}(X \in C) \leq \inf_{\lambda} \left(\sup_{z \in C} (-\lambda^T z) + \log \mathbf{E} \exp(\lambda^T X) \right)$$

Gaussian Variables on Polyhedron

X is Gaussian on \mathbf{R}^m with zero mean and variance I :

$$\log \mathbf{E} \exp(\lambda^T X) = \frac{\lambda^T \lambda}{2}$$

Let $C = \{x | Ax \preceq b\}$ be a polyhedron. By LP duality:

$$\sup\{y^T x | Ax \preceq b\} = \inf\{b^T u | A^T u = y, u \succeq 0\}$$

Generalized Chernoff bound becomes:

$$\log \mathbf{Prob}(X \in C) \leq \inf_{\lambda} \inf_u \{b^T u + \lambda^T \lambda / 2 | u \succeq 0, A^T u + \lambda = 0\}$$

Bound on $\mathbf{Prob}(X \in C)$ evaluated as exponential of optimal value of QP in (u, λ) :

$$\begin{array}{ll} \text{minimize} & b^T u + \lambda^T \lambda / 2 \\ \text{subject to} & A^T u + \lambda = 0, \quad u \succeq 0 \end{array}$$

Multi-user Detection

Received signal of K -user basic synchronous CDMA channel:

$$y(t) = \sum_{k=1}^K A_k b_k s_k(t) + n(t), \quad t \in [0, T]$$

Amplitude A_k , signature waveform $s_k(t)$, information bit $b_k \in \{-1, +1\}$, noise $n(t)$, period T

ML detection: find b that minimizes

$$\begin{aligned} & \int_0^T \left[y(t) - \sum_{k=1}^K A_k b_k s_k(t) \right]^2 dt \\ &= \int_0^T \left[\sum_{k=1}^K A_k b_k s_k(t) \right] dt - 2 \int_0^T \left[\sum_{k=1}^K A_k b_k s_k(t) \right] y(t) dt \\ &= b^T H b - 2b^T A y \end{aligned}$$

Integer Constrained Optimization Formulation

$A = \mathbf{diag}(A)$: amplitude matrix

$R_{ij} = \int_0^T s_i(t)s_j(t)dt$: cross-correlation matrix

$y = RA b + n$, $H = ARA$ and $p = -2Ay$

minimize $x^T H x + p^T x$

subject to $x_i \in \{-1, +1\}$, $i = 1, \dots, K$

Equivalent form: let $X = x x^T$

subject to $X_{ii} = 1$, $i = 1, \dots, K$,

$X \succeq 0$, $\text{rank}(X) = 1$

SDP Relaxation

Neglect rank constraint:

$$\begin{aligned} \text{subject to } & X_{ii} = 1, \quad i = 1, \dots, K, \\ & X \succeq 0 \end{aligned}$$

$$\text{Let } \hat{X} = \begin{bmatrix} xx^T & x \\ x^T & 1 \end{bmatrix} \text{ and } C = \begin{bmatrix} H & p/2 \\ p^T/2 & 1 \end{bmatrix}, \text{ reformulate as SDP:}$$

$$\begin{aligned} \text{minimize } & \text{tr}(C\hat{X}) \\ \text{subject to } & \hat{X}_{ii} = 1, \quad i = 1, \dots, K + 1, \\ & \hat{X} \succeq 0 \end{aligned}$$

Dual Relaxation

Relax $x^T x = K$ to $x^T x \leq K$:

$$\begin{aligned} & \text{minimize} && x^T H x + p^T x \\ & \text{subject to} && x^T x \leq K \end{aligned}$$

Lagrange dual problem:

$$\begin{aligned} & \text{minimize} && -\frac{1}{4} p^T (H + \lambda I)^{-1} p - \lambda K \\ & \text{subject to} && \lambda \geq 0 \end{aligned}$$

Gradient-descent method solves this convex optimization in **one** variable λ

Recover primal optimal solution: $x^* = (H + \lambda^* I)^{-1} A y$

Spring Break

- Feb. 2 - March 3: you've learned **a lot**

Optimization, convexity, duality

Numerical algorithms

LP, QP, QCQP, SOCP, SDP, GP

Inform. Th., Net. flow, Res. Alloc., Detection

- March 4 - March 21: spring break

Digest the first half of the course (Read the book and papers)

Homework 3

Start thinking about Final Projects

Lecture Outline

- Estimating probability distributions and detecting signals through various forms of convex optimization
- Extend Chebyshev and Chernoff bounds
- Relaxation for multi-user detection problems

Readings: Ch. 7.1-7.4 in Boyd and Vandenberghe.

W. K. Ma, T. N. Davidson, K. M. Wong, Z. Q. Luo, and P. C. Ching, "Quasi-maximum-likelihood multiuser detection using semidefinite relaxation with applications to synchronous CDMA," *IEEE Trans. Signal Proc.*, vol. 50, no. 4, pp. 912-922, April 2002.

A. Yener, R. D. Yates, and S. Ulukus, "CDMA multiuser detection: a nonlinear programming approach," *IEEE Trans. Comm.*, vol. 50, pp. 1016-1024, June 2002.