Lectures 11-12
Applications in control and communication

Mikael Johansson
S3/Automatic Control
mikaelj@s3.kth.se

Anders Forsgren
Math/OptSyst
andersf@kth.se

Aims

After this lecture, you should be able to

- perform parametric estimation via convex optimization
- determine optimal detectors
- compute maximum margin linear classifiers
- perform optimal transceiver design

Parametric distribution estimation

Distribution estimation: estimate probability density function p(y) of a random variable based on observations y

Parametric distribution estimation: choose from familiy of densities $p_x(y)$ indexed by a parameter x

Maximum likelihood estimation

$$\text{maximize (over } x\text{)} \quad \log p_x(y)$$

convex if log-likelihood function $l(x) = \log p_x(y)$ is concave in x

Can add constraints $x \in C$ explicitly, or let $p_x(y)=0$ for $x \notin C$

Linear measurement with IID noise

Linear measurement model

$$y_i = a_i^T x + v_i, \qquad i = 1, \ldots, m$$

x is vector of unknown parameters, v_i is IID noise with density p(z)

measurements y_i have density $p_x(y) = \prod_{i=1}^{m} p(y_i - a_i^T x)$

maximum likelihood estimate: any solution to

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

Examples

If noise is zero-mean Gaussian,

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

ML estimate is the least-squares solution

For Laplacian noise

$$p(z) = 1/(2a) e^{-|z|/a}$$

ML estimate is the l_1-norm solution

For uniform noise on –[a,a], ML estimate is any x with $|a_i^T x - y_i| \le a$

Proof:

Link with penalty function approximation

Any penalty function approximation problem

$$\text{minimize} \quad \sum_{i=1}^{m} \phi(a_i^T x - b_i)$$

can be interpreted as a maximum likelihood problem

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

with noise density

$$p(z) = \frac{e^{-\phi(z)}}{\int e^{-\phi(u)}\, du}$$

Logistic regression

Random variable $y \in \{0,1\}$ with distribution

$$p = \mathbf{Prob}(y=1) = \frac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$$

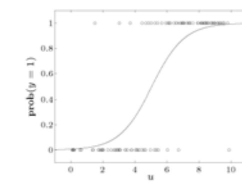
where a,b are parameters and $u \in \mathbb{R}^m$ are (observable) explanatory variables

Claim: The log-likelihood function is concave in (a,b)

Proof:

Logistic regression example

Example: n=1, m=50 measurements



dots are measured values (u_i, y_i)
solid line is estimated probability p=**Prob**(y=1)

Disposition

- parametric estimation
- optimal detection
- maximum margin linear classifiers
- optimal tranceiver design

(Binary) hypothesis testing

Detection (hypothesis testing) problem:

given observation of a random variable $X \in \{1, \ldots, N\}$, choose between

- Hypothesis 1: X was generated by distribution $p=\{p_1, \ldots, p_N\}$
- Hypothesis 2: X was generated by distribution $q=\{q_1, \ldots, q_N\}$

Randomized detector

- a nonnegative matrix $T \in \mathbb{R}^{2\times n}$ with $\mathbf{1}^T T = \mathbf{1}$
- if we observe X=k, we choose distribution p with probability t_{1k} and distribution q wit probability t_{2k}

If all elements of T are in {0,1} detector is deterministic

Detection probability and optimal design

Detection probability matrix

$$D = \begin{bmatrix} Tp & Tq \end{bmatrix} = \begin{bmatrix} 1 - P_{\mathrm{fp}} & P_{\mathrm{fn}} \\ P_{\mathrm{fp}} & 1 - P_{\mathrm{fn}} \end{bmatrix}$$

- P_{fp} is probability of selecting hypothesis 2 when X is generated by distribution p (false positive)
- P_{fn} is probability of selecting hypothesis 1 when X is generated by distribution q (false negative)

Multicriterion formulation of optimal design:

$$\begin{array}{lll} \text{minimize (w.r.t } \mathbb{R}^2_+) & (P_{\mathrm{fp}}, P_{\mathrm{fn}}) = (Tp_2, Tq_1) & \\ \text{subject to} & t_{1k} + t_{2k} = 1 & k = 1, \ldots, n \\ & t_{ik} \geq 0, & i = 1, 2, \; k = 1, \ldots, n \end{array}$$

Optimal detector design

Scalarization (with weight λ)

$$\begin{array}{ll} \text{minimize} & Tp_2 + \lambda(Tq_1) \\ \text{subject to} & t_{1k} + t_{2k} = 1, \quad t_{ik} \geq 0, \quad i = 1, 2, \quad k = 1, \ldots, n \end{array}$$

an LP with a simple analytical solution

$$(t_{1k}, t_{2k}) = \begin{cases} (1, 0) & p_k \geq \lambda q_k \\ (0, 1) & p_k < \lambda q_k \end{cases}$$

Optimal detector is deterministic, and relies on likelihood ratio test

Minimax design

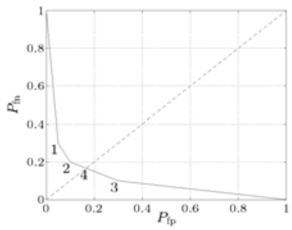
$$\begin{array}{ll} \text{minimize} & \max(Tp_2, Tq_1) \\ \text{subject to} & t_{1k} + t_{2k} = 1, \quad t_{ik} \geq 0, \quad i = 1, 2, \quad k = 1, \ldots, n \end{array}$$

An LP; solution is usually not deterministic

Example: binary hypothesis testing

Example:

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$



solutions 1,2,3 (and endpoints) deterministic; 4 is minimax detector

Linear discrimination

Separate two sets of points $\{x_1, \ldots, x_N\}$, $\{y_1, \ldots, y_M\}$ by a hyperplane

$$a^T x_i + b > 0, \quad i = 1, \ldots, N \qquad a^T y_i + b < 0, \quad i = 1, \ldots, M$$

Homogeneous in a,b; equivalent to

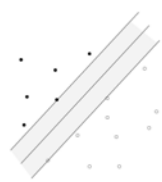
$$a^T x_i + b \geq 1, \quad i = 1, \ldots, N \qquad a^T y_i + b \leq -1, \quad i = 1, \ldots, M$$

a set of linear inequalities in a,b

Robust linear discrimination

Euclidean distance between hyperplanes

$$\mathcal{H}_1 = \{z \mid a^T z + b = 1\}$$
$$\mathcal{H}_2 = \{z \mid a^T z + b = -1\}$$



is $\mathrm{dist}(\mathcal{H}_1, \mathcal{H}_2) = 2/\|a\|_2$

To separate two sets of points with maximum margin

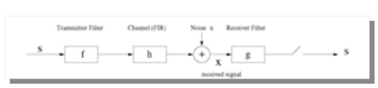
$$\begin{array}{ll} \text{minimize} & \|a\|_2^2 \\ \text{subject to} & a^T x_i + b \geq 1, \quad i = 1, \ldots, N \\ & a^T y_i + b \leq -1, \quad i = 1, \ldots, M \end{array}$$

a quadratic program in a,b.

Disposition

- parametric estimation
- optimal detection
- maximum margin linear classifiers
- optimal transceiver design

Single user communication scheme



Here,

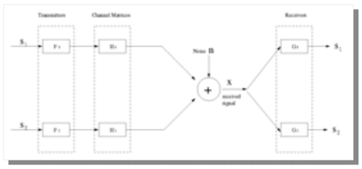
- s is input signal (assumed to be statistically white)
- $h \in \mathbb{R}^k$ is a linear time-invariant (FIR) channel (assumed known)
- f is a transmitter filter
- g is an equilizer/receiver filter
- n is additive (Gaussian) noise

Tranceiver design

From a course perspective: the use of four key techniques

- Variable elimination
- A variable transformation
- A monotonicity argument
- The Schur complement

Two-user multiple access scheme



We assume that $\mathbf{E}\{s_i s_i^H\} = I$ Then, the average transmit powers are

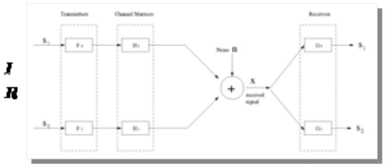
$$\mathbf{E}\{\sum_j |y_{ij}|^2\} = \mathbf{E}\{y_i^H y_i\} = \mathbf{E}\{\mathbf{Tr}\, y_i y_i^H\}$$

and the transmit power constraints can be written as

$$\mathbf{E}\{\mathbf{Tr}\, y_i y_i^H\} = \mathbf{E}\{\mathbf{Tr}\{F_i s_i s_i^H F_i^H\} = \mathbf{Tr} F_i F_i^H \leq p_{i,\text{tot}}$$

Two-user multiple access scheme

$$\mathbf{E}\{s_i s_i^*\} = I$$
$$\mathbf{E}\{nn^*\} = R$$



Assuming s_1, s_2, n mutually uncorrelated, the received signal is

$$x = H_1 F_1 s_1 + H_2 F_2 s_2 + n$$

with covariance matrix

$$\begin{aligned}
&\mathbf{E}\{xx^H\} = \\
&= \mathbf{E}\{H_1 F_1 s_1 s_1^H F_1^H H_1^H + H_2 F_2 s_2 s_2^H F_2^H H_2^H + nn^H\} = \\
&= \underbrace{H_1 F_1 F_1^H H_1^H + H_2 F_2 F_2^H H_2^H + R}_{W^{-1}}
\end{aligned}$$

Optimal tranceiver design

With

$$e_i = \hat{s}_i - s_i = G_i x - s_i$$

Optimal (minimal mean-square error) tranceiver design problem

$$\begin{array}{ll}
\text{minimize} & f_0(F_1, G_1, F_2, G_2) = \mathbf{E}\{e_1^H e_1 + e_2^H e_2\} = \mathbf{Tr}\,\mathbf{E}\{e_1 e_1^H\} + \mathbf{E}\{e_2 e_2^H\} \\
\text{subject to} & \mathbf{E}\{y_1^H y_1\} \leq p_{1,\text{tot}} \\
& \mathbf{E}\{y_2^H y_2\} \leq p_{2,\text{tot}}
\end{array}$$

Optimal tranceiver design

Study first term of objective function

$$\begin{aligned}
\mathbf{E}\{e_1 e_1^H\} &= \mathbf{E}\{(\hat{s}_1 - s_1)(\hat{s}_1 - s_1)^H\} \\
&= \mathbf{E}\{(G_1 x - s_1)(G_1 x - s_1)^H\} \\
&= \mathbf{E}\{G_1 x x^H G_1^H - s_1 x^H G_1^H - G_1 x s_1^H + s_1 s_1^H\} \\
&= G_1 W^{-1} G_1^H - G_1 H_1 F_1 - (G_1 H_1 F_1)^H + I
\end{aligned}$$

where we have used that

$$\begin{aligned}
\mathbf{E}\{x s_1^H\} &= \mathbf{E}\{(H_1 F_1 s_1 + H_2 F_2 s_2 + n) s_1^H\} \\
&= \mathbf{E}\{H_1 F_1 s_1 s_1^H\} = H_1 F_1
\end{aligned}$$

Optimal tranceiver design

Optimization formulation

$$\begin{array}{ll}
\text{minimize} & f_0(F_1, G_1, F_2, G_2) = \mathbf{Tr}\{\mathbf{E}\, e_1 e_1^H + \mathbf{E}\, e_2 e_2^H\} \\
\text{subject to} & \mathbf{Tr}\, F_1 F_1^H \leq p_{1,\text{tot}} \\
& \mathbf{Tr}\, F_2 F_2^H \leq p_{2,\text{tot}}
\end{array}$$

Non-convex due to coupling between F_i and G_i.

Observation: G_i are unconstrained. Rewrite objective as

$$\tilde{f}_0(F_1, F_2) = \min_{G_1, G_2} f_0(F_1, G_1, F_2, G_2)$$

First key step

Re-write objective as

$$\tilde{f}_0(F_1, F_2) = \min_{G_1, G_2} f_0(F_1, G_1, F_2, G_2)$$

Only first (second) term depends on G_1 (G_2):

$$\mathbf{E}\{e_1 e_1^H\} = G_1 W^{-1} G_1^H - G_1 H_1 F_1 - (G_1 H_1 F_1)^H + I$$

Minimum attained for

$$G_1^\star = F_1^H H_1^H W$$

resulting in

$$\mathbf{E}\{e_1 e_1^H\} = I - F_1^H H_1^H W H_1 F_1$$

First key step cont'd

Can simplify objective function

$$\begin{aligned}&\mathbf{Tr}\, e_1 e_1^H + \mathbf{Tr}\, e_2 e_2^H \\ &= \mathbf{Tr}\, I - F_1^H H_1^H W H_1 F_1 + \mathbf{Tr}\, I - F_2^H H_2^H W H_2 F_2 \\ &= 2n - \mathbf{Tr}\, F_1^H H_1^H W H_1 F_1 - \mathbf{Tr}\, F_2^H H_2^H W H_2 F_2 \\ &= 2n - \mathbf{Tr}\, W H_1 F_1 F_1^H H_1^H - \mathbf{Tr}\, W H_2 F_2 F_2^H H_2^H \\ &= 2n - \mathbf{Tr}\, W (H_1 F_1 F_1^H H_1^H + H_2 F_2 F_2^H H_2^H) \\ &= 2n - \mathbf{Tr}\, W (W^{-1} - R) \\ &= n + \mathbf{Tr}\, WR\end{aligned}$$

First key step cont'd

Current formulation

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}\, WR \\ \text{subject to} & \mathbf{Tr}\, F_1 F_1^H \le p_{1,\text{tot}} \\ & \mathbf{Tr}\, F_2 F_2^H \le p_{2,\text{tot}} \\ \\ \text{where} & W = (H_1 F_1 F_1^H H_1^H + H_2 F_2 F_2^H H_2^H + R)^{-1}\end{array}$$

Second key step

A change-of variables:

$$U_1 = F_1 F_1^H$$
$$U_2 = F_2 F_2^H$$

leads to

$$\begin{array}{lll}\text{minimize} & \mathbf{Tr}\, WR & \\ \text{subject to} & \mathbf{Tr}\, U_1 \le p_{1,\text{tot}}, & U_1 \succeq 0 \\ & \mathbf{Tr}\, U_2 \le p_{2,\text{tot}}, & U_2 \succeq 0 \\ \\ \text{where} & W = (H_1 U_1 H_1^H + H_2 U_2 H_2^H + R)^{-1} & \end{array}$$

Nonlinear equality constraint problematic?

Third key step

Can consider inequality for W rather than equality, e.g.

$$\begin{array}{lll}\text{minimize} & \mathbf{Tr}\, WR & \\ \text{subject to} & \mathbf{Tr}\, U_1 \le p_{1,\text{tot}}, & U_1 \succeq 0 \\ & \mathbf{Tr}\, U_2 \le p_{2,\text{tot}}, & U_2 \succeq 0 \\ & W \succeq (H_1 U_1 H_1^H + H_2 U_2 H_2^H + R)^{-1} & \end{array}$$

Why?

Third key step cont'd

Recall that

$$\mathbf{Tr}\, ZY \ge 0 \text{ whenever } Z, Y \succeq 0$$

Consider the formulation

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}\, WR \\ & W \succeq X, \quad X \in \mathcal{X}\end{array}$$

If the inequality is not tight at optimum, there exists $S \succeq 0$:

$$W^\star = X^\star + S$$

Note, however, that $W = W^\star - S$ achieves a lower objective value since

$$\mathbf{Tr}\, W^\star R - \mathbf{Tr}\, WR = \mathbf{Tr}\, SR \ge 0$$

which contradicts that $W^\star$ is optimal

Fourth key step

Transform nonlinear equality into LMI using Schur complement

$$\begin{array}{lll}\text{minimize} & \mathbf{Tr}\, WR & \\ \text{subject to} & \mathbf{Tr}\, U_1 \le p_{1,\text{tot}}, & U_1 \succeq 0 \\ & \mathbf{Tr}\, U_2 \le p_{2,\text{tot}}, & U_2 \succeq 0 \\ & \begin{bmatrix} W & I \\ I & H_1 U_1 H_1^H + H_2 U_2 H_2^H + R \end{bmatrix} \succeq 0 & \end{array}$$

A semidefinite programming problem in W, U_1 and U_2.

Optimal filter via Cholesky factorization of U_1, U_2.

Much much more...

- Other performance objectives
- Other channel models (MIMO, Broadcast)
- Further simpliciations and efficient algorithms

Convex optimization applications abound in communications
- Filter design
- Beamforming
- Coding
- Resource allocation problems

⋮

Many more applications were presented at research paper day!

Summary

- parametric estimation via convex optimization
- optimal detectors
- maximum margin linear classifiers
- optimal transceiver design