Convex Optimization

Lecture 7 - Applications in Signal Processimg

Instructor: Yuanzhang Xiao

University of Hawaii at Manoa

Fall 2017

Today's Lecture

1. Statistical Estimation
2. Hypothesis Testing
3. Filter Design

Outline

1. Statistical Estimation

2. Hypothesis Testing

3. Filter Design

Parametric Distribution Estimation

$p_x(y)$: $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

- as a function of y: probability distribution of y, indexed by x
- as a function of x: likelihood function for fixed y

parametric distribution estimation:

- given y, choose the "most likely" distribution $p_x(\cdot)$ (i.e., choose x)

Maximum Likelihood Estimation

maximum likelihood estimation:

$$\text{maximize} \quad p_x(y)$$

with optimization variable x

log-likelihood function: (more like to be concave)

$$\ell(x) = \log p_x(y)$$

equivalent formulation:

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

with optimization variable x

Linear Measurements With IID Noise

linear measurement model:

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

- $x_i \in \mathbb{R}^n$: parameters to be estimated
- $y_i \in \mathbb{R}$: measurement
- $v_i \in \mathbb{R}$: noise

we assume

- v_i are independent, identically distributed (IID)
- v_i has probability density function $p(\cdot)$

log-likelihood function:

$$\ell(x) = \log p_x(y) = \log \prod_{i=1}^{m} p\left(y_i - a_i^T x\right) = \sum_{i=1}^{m} \log p\left(y_i - a_i^T x\right)$$

Linear Measurements With IID Gaussian Noise

Gaussian noise with zero mean and variance σ^2:

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

log-likelihood function:

$$\begin{aligned} \ell(x) &= -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y_i - a_i^T x\right)^2 \\ &= -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Ax - y\|_2^2 \end{aligned}$$

maximum likelihood estimation:

$$\text{minimize} \quad \|Ax - y\|_2^2$$

Linear Measurements With IID Laplacian Noise

Laplacian noise with zero mean and variance σ^2:

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

with $a > 0$

log-likelihood function:

$$\begin{aligned} \ell(x) &= -m \log(2a) - \frac{1}{a} \sum_{i=1}^{m} \left| y_i - a_i^T x \right| \\ &= -m \log(2a) - \frac{1}{a} \|Ax - y\|_1 \end{aligned}$$

maximum likelihood estimation:

$$\text{minimize} \quad \|Ax - y\|_1$$

Linear Measurements With IID Uniform Noise

uniformly distributed noise on $[-a, a]$:

$$p(z) = \begin{cases} \frac{1}{2a} & \text{if } z \in [-a, a] \\ 0 & \text{otherwise} \end{cases}$$

log-likelihood function:

$$\begin{aligned} \ell(x) &= \begin{cases} -m \log(2a) & \text{if } \left|y_i - a_i^T x\right| \leq a, \ \ i = 1, \ldots, m \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} -m \log(2a) & \text{if } \|Ax - y\|_\infty \leq a \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

maximum likelihood estimation:

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \|Ax - y\|_\infty \leq a \end{array}$$

a feasibility problem

Counting Problems With Poisson Distribution

measurement y is nonnegative integer with Poisson distribution:

$$\text{prob}(y = k) = \frac{e^{-\mu}\mu^k}{k!}$$

- e.g., # of cars passing an intersection, # of traffic accidents
- μ is the average number per unit time

assume $\mu = a^T u + b$

- $u \in \mathbb{R}^n$: explanatory variables
  - e.g., $u = (\text{total traffic flow}, \text{rainfall}, \text{peak hours or not}) \in \mathbb{R}^3$
- $a \in \mathbb{R}^n, b \in \mathbb{R}$: model parameters to be estimated

try to estimate how traffic accidents depend on various variables

Counting Problems With Poisson Distribution

m measurements: $(u_i, y_i),\ i = 1, \ldots, m$

likelihood function:

$$\prod_{i=1}^{m} \frac{(a^T u_i + b)^{y_i} e^{-(a^T u_i + b)}}{y_i!}$$

log-likelihood function:

$$\ell(x) = \sum_{i=1}^{m} \left[y_i \log(a^T u_i + b) - (a^T u_i + b) - \log(y_i!) \right]$$

maximum likelihood estimation:

$$\text{maximize} \quad \sum_{i=1}^{m} \left[y_i \log(a^T u_i + b) - (a^T u_i + b) \right]$$

with optimization variables $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$

Nonparametric Distribution Estimation

a random variable X with values in finite set $\{\alpha_1, \ldots, \alpha_n\}$

- distribution of X characterized by $p \in \mathbb{R}^n$,
  where $\mathrm{prob}(X = \alpha_k) = p_k, \;\; k = 1, \ldots, n$

nonparametric distribution estimation:

- estimate the distribution p
- base on prior information, observations, measurements, etc.

Prior Information

prior information $\Rightarrow$ constraints on p

examples of prior information that result in linear constraints:

- mean: $\mathbf{E}(X) = \alpha \Rightarrow \sum_{i=1}^{n} \alpha_i p_i = \alpha$
- second moment: $\mathbf{E}(X^2) = \beta \Rightarrow \sum_{i=1}^{n} \alpha_i^2 p_i = \beta$
- $\text{prob}(X \geq 0) \leq 0.2 \Rightarrow \sum_{i:\alpha_i \geq 0} p_i \leq 0.2$

examples of prior information that result in nonlinear constraints:

- variance: (concave in p)

$$\mathbf{var}(X) = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2 = \sum_{i=1}^{n} \alpha_i^2 p_i - \sum_{i=1}^{n} (\alpha_i p_i)^2$$

- Kullback-Leiber divergence from distribution q: (convex in p)

$$\sum_{i=1}^{n} p_i \log(p_i / q_i)$$

Maximum Likelihood Estimation

prior information represented by $p \in \mathcal{P}$

- $\mathcal{P}$ results from the prior information discussed earlier
- $\mathcal{P} \subseteq \{p : 1^T p = 1, p \geq 0\}$

observations:

- N independent samples
- the number of samples with value α_i is k_i

log-likelihood function

$$\ell(p) = \log \prod_{i=1}^{n} p_i^{k_i} = \sum_{i=1}^{n} k_i \log(p_i)$$

maximum likelihood estimation:

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

Other Problems

the distribution with the minimum expected value of a function:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f(\alpha_i) p_i \\ \text{subject to} & p \in \mathcal{P} \end{array}$$

where $f(\cdot)$ can be any function (even nonconvex)

the distribution with minimum K-L divergence from q:

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

- $q = \frac{1}{n}\mathbf{1}$: find the most random distribution

Outline

Hypothesis Testing

a random variable X with values in finite set $\{\alpha_1, \ldots, \alpha_n\}$

- distribution of X parameterized by $\theta \in \{\theta_1, \ldots, \theta_m\}$
- matrix $P \in \mathbb{R}^{n \times m}$ with

$$p_{ij} = \mathrm{prob}\,(X = \alpha_i | \theta = \theta_j)$$

observe samples of X, then estimate θ

- θ: hypothesis $\Rightarrow$ hypothesis testing
- θ: events $\Rightarrow$ event detection

note the difference from nonparametric distribution estimation:
P is known in hypothesis testing

Detector

detector: a random variable $\hat{\theta}$ with distribution depending on X

$T \in \mathbb{R}^{m\times n}$ with

$$t_{ji} = \text{prob}\left(\hat{\theta} = \theta_j | X = \alpha_i\right)$$

- when observe α_i, the detector give $\hat{\theta} = \theta_j$ with probability t_{ji}
- ith column of T, t_i: probability distribution of $\hat{\theta}$ given α_i

T must satisfy

$$t_i \succeq 0, \quad \mathbf{1}^T t_i = 1, \; i = 1, \ldots, n$$

Detection Probability Matrix

detection probability matrix: $D = TP \in \mathbb{R}^{m \times m}$, where

$$\begin{aligned} D_{ji} &= (TP)_{ji} \\ &= \sum_{k=1}^{n} \text{prob}\left(\hat{\theta} = \theta_j | X = \alpha_k\right) \text{prob}\left(X = \alpha_k | \theta = \theta_i\right) \\ &= \text{prob}\left(\hat{\theta} = \theta_j | \theta = \theta_i\right) \end{aligned}$$

- detection probabilities denoted by $P^d \in \mathbb{R}^m$ with

$$P_i^d = D_{ii} = \text{prob}\left(\hat{\theta} = \theta_i | \theta = \theta_i\right)$$

- error probabilities denoted by $P^e \in \mathbb{R}^m$ with

$$P_i^e = \sum_{j \neq i} D_{ji} = \text{prob}\left(\hat{\theta} \neq \theta_i | \theta = \theta_i\right)$$

P^d and P^e are linear in detector T

Optimal Detector Design – Minimax Detector

minimax detector design:

$$\begin{array}{ll} \text{minimize} & \max_j P_j^e \\ \text{subject to} & t_i \geq 0, \mathbf{1}^T t_i = 1, i = 1, \ldots, n \end{array}$$

minimize the worst-case error probability

can add many constraints:

- lower bounds on detection probabilities:

$$P_i^d = D_{ii} \geq L_i$$

- upper bounds on error probabilities:

$$D_{ji} \leq U_{ji}$$

Optimal Detector Design – Bayes Detector

suppose that there is a prior distribution for the hypotheses $q \in \mathbb{R}^m$

$$q_i = \text{prob}\,(\theta = \theta_i)$$

Bayes detector design:

$$\begin{array}{ll} \text{minimize} & q^T P^e \\ \text{subject to} & t_i \geq 0, 1^T t_i = 1, i = 1, \ldots, n \end{array}$$

minimize the expected error probability

Robust Detector Design – Robust Minimax Detector

suppose that P is not known exactly but $P \in \mathcal{P}$

robust (worst-case) minimax detector design:

$$\begin{array}{ll} \text{minimize} & \max_j \ \sup_{p \in P} P_j^e \\ \text{subject to} & t_i \geq 0, \mathbf{1}^T t_i = 1, i = 1, \ldots, n \end{array}$$

where $P_j^e = 1 - D_{jj}$

Outline

FIR Filters

finite impulse response (FIR) filters:

$$y(t) = \sum_{\tau=0}^{n-1} h_\tau u(t-\tau), \quad t \in \mathbb{Z}$$

- u: input signal/sequence
- y: output signal/sequence
- h_i: filter coefficients
- n: filter length

Frequency Response

frequency response:

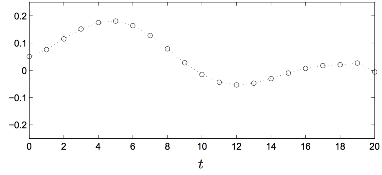
$$\begin{aligned} H(\omega) &= h_0 + h_1 e^{-j\omega} + \cdots + h_{n-1} e^{-j(n-1)\omega} \\ &= \sum_{t=0}^{n-1} h_t e^{-j(t-1)\omega} \end{aligned}$$

- $H(\omega + 2\pi) = H(\omega)$, $H(\omega + \pi) = -H(\omega)$
- need to specify $H(\cdot)$ only in $[0, \pi]$

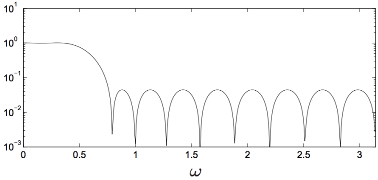
filter design: choose h so that h and H satisfy certain specifications

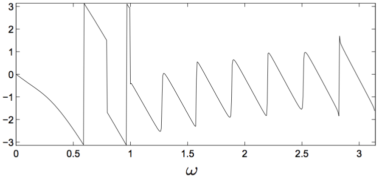
Frequency Response - Example

FIR filter $h(t)$ of length $n = 21$:



frequency response $H(\omega)$:





magnitude $|H(\omega)|$ phase $\angle H(\omega)|$

Chebychev Design

Chebychev design:

$$\text{minimize} \quad \max_{\omega \in [0,\pi]} |H(\omega) - H_{\text{des}}(\omega)|$$

with optimization variables $h(0), \ldots, h(n-1)$

- H_{des}: desired frequency response
- convex optimization (may not be easy to solve)

relaxation:

$$\text{minimize} \quad \max_{k=1,\ldots,m} |H(\omega_k) - H_{\text{des}}(\omega_k)|$$

with optimization variables $h(0), \ldots, h(n-1)$

- close to the desired frequency response at m sample points

Chebychev Design

Chebychev design as SOCP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \|A_k h - b_k\| \leq t, \;\; k = 1, \ldots, m \end{array}$$

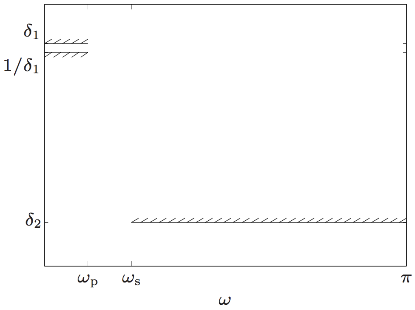
where

$$A_k = \begin{bmatrix} 1 & \cos\omega_k & \cdots & \cos(n-1)\omega_k \\ 0 & \sin\omega_k & \cdots & \sin(n-1)\omega_k \end{bmatrix}$$

$$b_k = \begin{bmatrix} \mathrm{Re}\left(H_{\mathrm{des}}(\omega_k)\right) \\ \mathrm{Im}\left(H_{\mathrm{des}}(\omega_k)\right) \end{bmatrix}$$

Lowpass Filter Design

lowpass filter design:



- low frequency $[0, \omega_p]$: magnitude within $[1/\delta_1, \delta_1]$

$$1/\delta_1 \leq |H(\omega)| \leq \delta_1, \; 0 \leq \omega \leq \omega_p$$

- high frequency $[\omega_s, \pi]$: magnitude below δ_2

$$|H(\omega)| \leq \delta_2, \; \omega_s \leq \omega \leq \pi$$

Lowpass Filter Design

samples at frequencies $\omega_1, \ldots, \omega_m$

minimum stopband magnitude:

$$
\begin{array}{ll}
\text{minimize} & \delta_2 \\
\text{subject to} & 1/\delta_1 \leq |H(\omega_k)| \leq \delta_1, \ 0 \leq \omega_k \leq \omega_p \\
& |H(\omega_k)| \leq \delta_2, \ \omega_s \leq \omega_k \leq \pi
\end{array}
$$

nonconvex due to $1/\delta_1 \leq |H(\omega_k)|$

Lowpass Filter Design – Linear Phase Filters

consider filter h such that

- $n = 2N + 1$ is odd
- h is symmetric: $h_t = h_{n-1-t}, \; t = 0, \ldots, n-1$

$$\begin{aligned} H(\omega) &= h_0 + h_1 e^{-j\omega} + \cdots + h_{n-1} e^{-j(n-1)\omega} \\ &= h_N e^{-jN\omega} + \sum_{t=0}^{N-1} h_t e^{-jt\omega} + h_{n-1-t} e^{-j(n-1-t)\omega} \\ &= h_N e^{-jN\omega} + \sum_{t=0}^{N-1} h_t \left(e^{-jt\omega} + e^{-j(2N-t)\omega} \right) \\ &= e^{-jN\omega} \left[h_N + \sum_{t=0}^{N-1} h_t \left(e^{-j(t-N)\omega} + e^{-j(N-t)\omega} \right) \right] \\ &= e^{-jN\omega} \left(h_N + \sum_{t=0}^{N-1} 2 h_t \cos(t-N)\omega \right) \end{aligned}$$

Lowpass Filter Design – Linear Phase Filters

linear phase filter:

- $H(\omega) = e^{-jN\omega}\tilde{H}(\omega)$
- phase $N\omega$ is linear in frequency
- magnitude is $\tilde{H}(\omega)$
- $\tilde{H}(\omega)$ is real and linear in h

minimum stopband magnitude:

$$\begin{array}{ll} \text{minimize} & \delta_2 \\ \text{subject to} & 1/\delta_1 \leq \tilde{H}(\omega_k) \leq \delta_1, \; 0 \leq \omega_k \leq \omega_p \\ & \tilde{H}(\omega_k) \leq \delta_2, \; \omega_s \leq \omega_k \leq \pi \end{array}$$

linear program

Lowpass Filter Design – Linear Phase Filters

minimum passband fluctuation:

$$\begin{array}{ll} \text{minimize} & \delta_1 \\ \text{subject to} & \delta_1 \geq 1 \\ & 1/\delta_1 \leq \tilde{H}(\omega_k) \leq \delta_1, \; 0 \leq \omega_k \leq \omega_p \\ & \tilde{H}(\omega_k) \leq \delta_2, \; \omega_s \leq \omega_k \leq \pi \end{array}$$

convex optimization (but not LP)

Lowpass Filter Design – Convex Reformulation

suppose that we consider general filters (not linear phase)

change of variables: autocorrelation coefficients:

$$r_t = \sum_{\tau} h_\tau h_{\tau+t}$$

where $h_t = 0$ for $t < 0$ and $t \geq n$

- $r_t = r_{-t}$ and $r_t = 0$ for $|t| \geq n$
- need to specify $r = (r_0, \ldots, r_{n-1}) \in \mathbb{R}^n$
- Fourier transform

$$R(\omega) = \sum_{\tau} r_\tau e^{-j\tau\omega} = r_0 + \sum_{t=1}^{n-1} 2 r_t \cos \omega t = |H(\omega)|^2$$

convex constraints: $L(\omega)^2 \leq R(\omega) \leq U(\omega)^2$