Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics

S-88.4212 Signal Processing in Telecommunications II

Fall 2013

Lecture 2: ML Estimation Principles

Dr. Stefan Werner
stefan.werner@aalto.fi
Office SG410, Reception Monday 10-11

Timetable

L1	Introduction; models for channels and comms. systems
L2	**ML Estimation principles**
L3	Synchronization: Overview
L4	Carrier frequency estimation I
L5	Carrier frequency estimation II
L6	Carrier phase estimation I
L7	Carrier phase estimation II
L8	Symbol timing estimation I
L9	Symbol timing estimation II
L10	Channel estimation I
L11	Channel estimation II, course review

Exam **12.12.** **Exam Thursday** 12-15 (Check!)

Contents of Lecture 2

The material in this lecture is mostly based on the book
S. Kay: *Fundamentals of Statistical Signal Processing - Estimation Theory*, Prentice-Hall 1993.

Department of Signal Processing and Acoustics

I. What is synchronization

What is synchronization

Synchronous digital transmission

- Information is carried by uniformly spaced pulses
- Signal is known except for:
  - Data symbols
  - *Reference parameters*
- **Baseband** pulse amplitude modulation:
  - Matched filtering & symbol-rate sampling
  - Optimum sampling time at pulse peaks for max eye opening and min errors
  - *Timing (clock) synchronizer*

What is synchronization...

- Coherent **passband** transmission:
  - Signal is modulated to a sinusoidal carrier frequency
  - For coherent demodulation, a local reference sinusoid is needed with the same *frequency* and *phase*
  - Requires carrier frequency and phase estimation
- Alternatives to coherent demodulation
  - *Differentially coherent:* phase *difference* between consecutive symbol samples sufficient
  - *Noncoherent:* decisions based on carrier envelope only
  - Simpler to implement, worse performance
  - Not considered in this course!

What is synchronization...

- Higher levels of synchronization
  - Block coding: *word synchronizers* needed for identification of block boundaries
  - Convolutional coding: fixed symbol segments used, *node synchronizers* needed to mark the start of each segment
  - *Frame synchronizers* for time-shared channels (like TDMA)
  - *Network synchronization*: transmitter timing adjusted in e.g. PCM networks
  - Also beyond our scope!

What is our synchronization

- Concentrate on the estimation of
  - *timing*
  - *carrier frequency* and
  - *carrier phase* parameters

  in **passband** transmission
- Review ML estimation principles first

II. Maximum Likelihood (ML) Estimation

Maximum Likelihood (ML) Estimation

- Historical development of digital synchronization algorithms:
  - heuristic methods
  - application of ML estimation methods
- We want to employ the ML method *systematically* for different synchronization tasks

ML Estimation...

- Problem of extracting values of parameters from a *discrete-time* (continuous) *waveform* or a *data-set*
- Mathematically, we have the N-point data set $\{r(0), r(1),\ldots r(N\text{-}1)\}$, which depends on our unknown synchronization parameters, $\mathbf{v}$
- Determine $\mathbf{v}$ based on the data, or define an *estimator*

$$\hat{\mathbf{v}} = g[r(0), r(1)\ldots, r(N-1)]$$

where g is some function

ML Estimation...

- To determine good estimators the first step is to *mathematically* model the data
- Data is inherently *random* ⇒ we describe it by its probability density function (PDF)

$$p(\mathbf{r};\mathbf{v}) = p[r(0), r(1)\ldots, r(N-1);\mathbf{v}]$$

- PDF is parameterized by the unknown parameter $\mathbf{v}$ (denoted by a semicolon)
  - We have a class of PDFs where each one is different due to a different value of $\mathbf{v}$

ML Estimation...

- Simple example: $N = 1$ sample, v denote the mean, the PDF might be

$$p[r(0);v] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[r(0)-v]^2\right\}$$

- Since the value of v affects the probability of $r(0)$, we should be able to guess v from $r(0)$
- In an actual problem we must choose a proper PDF
  - Should be consistent with the problem constraints
  - Mathematically tractable
- In this course the PDF will take form above (AWGN channel)

ML Estimation...

- Transmitted signal: (baseband model, linear modulation like PAM, MPSK etc.)

$$x(t) = h_T(t) * \sum_{k=-\infty}^{\infty} a_k \delta(t-kT) = \sum_{k=-\infty}^{\infty} a_k h_T(t-kT)$$

a_k = data symbols to be transmitted
$h_T(t)$ = transmitted continuous-time waveform
$\delta(t)$ = Dirac delta function

ML Estimation...

- Received continuous-time signal:

$$r(t) = x(t, \mathbf{v}) + w(t)$$

$\mathbf{v}$ = vector of unknown sync parameters

$r(t)$ = received signal waveform

w(t) = additive noise

- Consider different trial sets $\mathbf{v}_1$,$\mathbf{v}_2$ for the sync parameters and the corresponding set of realizations $\mathbf{r}$ (vector representation of $r(t)$) for the received signal

ML Estimation...

- Define probability density function (PDF) of $\mathbf{r}$ with the parameter set $\mathbf{v}_1$

$$p(\mathbf{r}; \mathbf{v}_1)$$

- The parameter set $\mathbf{v}_1$ is more likely than $\mathbf{v}_2$ if

$$p(\mathbf{r}; \mathbf{v}_1) \quad > \quad p(\mathbf{r}; \mathbf{v}_2)$$

ML Estimation...

- The maximum likelihood (ML) solution: find such parameter set **v** that the probability density function

$$p(\mathbf{r};\mathbf{v})$$

is *maximum*, or

$$\hat{\mathbf{v}}_{\mathrm{ML}}(\mathbf{r}) = \arg\left\{\max_{\mathbf{v}}\{p(\mathbf{r};\mathbf{v})\}\right\}$$

ML Estimation...

- Sometimes only a *subset* **u** of all the parameters **v** is of interest (and the remaining set $\mathbf{u}_{\mathrm{R}}$ are *unwanted*)
- Total probability theorem gives the result

$$p(\mathbf{r};\mathbf{u}) = \int_{-\infty}^{\infty} p(\mathbf{r};\mathbf{v})\,p(\mathbf{u}_{\mathrm{R}})\,d\mathbf{u}_{\mathrm{R}}$$

i.e., the unwanted parameters are *integrated away* (practical approximation: *averaging*)

- The modified ML estimate is then

$$\hat{\mathbf{u}}_{\mathrm{ML}}(\mathbf{r}) = \arg\left\{\max_{\mathbf{u}}\{p(\mathbf{r};\mathbf{u})\}\right\}$$

Scalar parameter ML Estimation

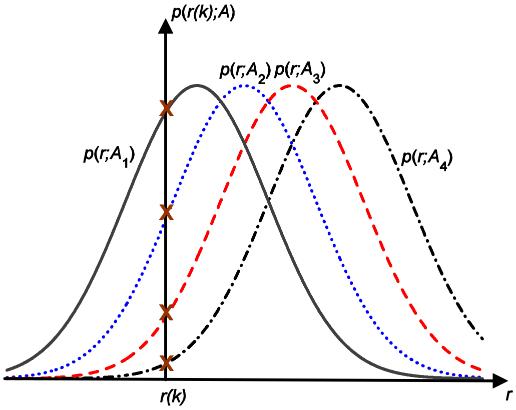
- Simple example: *Discrete-time* received signal is a DC level in zero-mean additive white Gaussian noise (AWGN):

$$r(k) = A + w(k), \qquad k = 0,1,\ldots,N-1$$

where the DC level A is *unknown* but the noise variance σ^2 is *known*. The PDF of the observation vector **r** is

$$p(\mathbf{r};A) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[\frac{-1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k)-A\right)^2\right]$$

Scalar parameter ML Estimation...



Scalar parameter ML Estimation...

- PDF is viewed as a continuous function of the unknown parameter (*likelihood function*, LF)
- For an ML estimate we need to find the A that maximizes the likelihood function
- For easier optimization, the likelihood function is usually replaced by the *log-likelihood function* (LLF)

$$\Lambda(\mathbf{r};A) = \ln p(\mathbf{r};A) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k)-A\right)^2$$

Scalar parameter ML Estimation...

- The maximum is found by setting the first derivative to zero:

$$\frac{\partial}{\partial A}\Lambda(\mathbf{r};A) = \frac{-1}{2\sigma^2}\sum_{k=0}^{N-1}(-2)\left(r(k)-A\right) = 0$$

$$\Leftrightarrow \quad \sum_{k=0}^{N-1} r(k) = NA$$

$$\Leftrightarrow \quad \hat{A} = \frac{1}{N}\sum_{k=0}^{N-1} r(k)$$

- The conventional *averager* thus gives the ML estimate for a DC level in the AWGN case.

Vector parameter ML Estimation

- Let us modify the DC-in-AWGN example so that *both A* and σ^2 are unknown. The likelihood function looks the same

$$p(\mathbf{r};\mathbf{v}) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[\frac{-1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k)-A\right)^2\right]$$

but it is now considered as a function of *two* parameters, or a function of the *vector* parameter

$$\mathbf{v} = \left[A\ \sigma^2\right]^{\mathrm{T}}$$

Vector parameter ML Estimation...

- The zero of the derivative of the LLF with respect to (w.r.t.) *A* gives the same equation as before:

$$\frac{\partial}{\partial A}\Lambda(\mathbf{r};\mathbf{v}) = \frac{-1}{2\sigma^2}\sum_{k=0}^{N-1}(-2)\left(r(k)-A\right) = 0$$

- Differentiating w.r.t. to σ^2 yields

$$\frac{\partial}{\partial \sigma^2}\Lambda(\mathbf{r};\mathbf{v}) = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=0}^{N-1}\left(r(k)-A\right)^2 = 0$$

Vector parameter ML Estimation...

- Solving these two equations simultaneously gives the ML estimates for the DC level and AWGN variance as

$$\hat{A} = \bar{r} = \frac{1}{N}\sum_{k=0}^{N-1} r(k)$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(r(k) - \bar{r}\right)^2$$

which are the familiar results for sample mean and variance.

Construction of MLE algorithms

- In principle, the design of MLE algorithms is easy:
  1) Express the LLF, preferably using AWGN model
  2) Find the maximum of the LLF
     - Differentiate w.r.t. parameters and set derivatives to zero
  3) Solve for desired parameters
- The solution always in terms of *received data samples*
- The estimator has good properties (unbiased, efficient)
- Not always in neat closed form though!
- Problems: nonlinear dependencies of parameters
  – Try linearizations & other tricks

III. ML example: sinusoid phase estimation

ML examples

- Consider sinusoidal signal

$$r(k) = A\cos(2\pi f_0 k + \theta) + w(k), \qquad k = 0,1,\ldots,N-1$$

where $w(k)$ is AWGN with known variance σ^2 , and the sine amplitude A and frequency f_0 are known as well.

- The LF is thus a function of the *unknown phase* θ:

$$p(\mathbf{r};\theta) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[\frac{-1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k) - A\cos(2\pi f_0 k + \theta)\right)^2\right]$$

ML examples...

- The log-likelihood function:

$$\Lambda(\mathbf{r};\theta) = C_1 - \frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k) - A\cos(2\pi f_0 k + \theta)\right)^2$$

- Minimization: differentiate LLF w.r.t. θ:

$$\frac{\partial\Lambda(\mathbf{r};\theta)}{\partial\theta} = \frac{-2}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k) - A\cos(2\pi f_0 k + \theta)\right)A\sin(2\pi f_0 k + \theta) = 0$$

ML examples...

- This gives equation

$$\sum_{k=0}^{N-1} r(k)\sin(2\pi f_0 k + \hat{\theta}) = A\sum_{k=0}^{N-1}\sin(2\pi f_0 k + \hat{\theta})\cos(2\pi f_0 k + \hat{\theta})$$

- Simplify:

$$\sum_{k=0}^{N-1}\sin(2\pi f_0 k + \hat{\theta})\cos(2\pi f_0 k + \hat{\theta}) = \frac{1}{2}\sum_{k=0}^{N-1}\sin(4\pi f_0 k + 2\hat{\theta}) \approx 0$$

when N is large enough and f_0 is not too close to 0 or 0.5.

ML examples...

- The approximate MLE can then be solved from

$$\sum_{k=0}^{N-1} r(k)\sin(2\pi f_0 k + \hat{\theta}) = 0$$

- Separate the phase:

$$\sum_{k=0}^{N-1} r(k)\sin(2\pi f_0 k)\cos\hat{\theta} = -\sum_{k=0}^{N-1} r(k)\cos(2\pi f_0 k)\sin\hat{\theta}$$

ML examples...

- The approximate MLE for the phase is then solved as:

$$\hat{\theta} = -\arctan\left(\frac{\sum_{k=0}^{N-1} r(k)\sin(2\pi f_0 k)}{\sum_{k=0}^{N-1} r(k)\cos(2\pi f_0 k)}\right)$$

- Closed-form (approximate) solution!

IV. Cramer-Rao lower bound

Cramer-Rao lower bound

- How good is the MLE? How good can *any* estimate be?
- The goodness of any *unbiased* estimate can be measured by the *variance* of the estimate
- The variance depends on PDF (or LF) and its *sensitivity* to the parameter in question
- If the PDF depends only weakly (or not at all!) on the parameter, we cannot expect to get a good estimate from data samples by using any technique

Cramer-Rao lower bounds...

- In the DC-in-AWGN example we obtained the LLF derivative w.r.t. *A* as

$$\frac{\partial}{\partial A}\Lambda(\mathbf{r};A)=\frac{1}{\sigma^2}\sum_{k=0}^{N-1}\big(r(k)-A\big)$$

- By setting this to zero we got the averager as the MLE:

$$\hat{A}=\bar{r}=\frac{1}{N}\sum_{k=0}^{N-1}r(k)$$

- The *second* derivative of the LLF is

$$\frac{\partial^2}{\partial A^2}\Lambda(\mathbf{r};A)=\frac{-N}{\sigma^2}$$

Cramer-Rao lower bounds...

- The second derivative is the negative inverse of the estimator variance:

$$\operatorname{var}(\hat{A})=\frac{\sigma^2}{N}=\frac{-1}{\dfrac{\partial^2}{\partial A^2}\Lambda(\mathbf{r};A)}$$

- It can be shown that this is a lower bound for *all* estimators and *all* PDFs!

Cramer-Rao lower bound

- *Cramer-Rao Lower Bound (Scalar parameter):* For any unbiased estimator of parameter *A*, the estimator variance is lower bounded by

$$\mathrm{var}(\hat{A}) \geq \frac{-1}{\mathrm{E}\left[\frac{\partial^2}{\partial A^2}\Lambda(\mathbf{r};A)\right]} = CRB$$

Cramer-Rao lower bound...

- *Cramer-Rao Lower Bound (Vector parameter):* For any unbiased estimator of parameter vector **v**, the estimator variance is lower bounded by

$$\mathrm{var}(\hat{v}_i) \geq \left[\mathbf{I}^{-1}(\mathbf{v})\right]_{ii} = CRB$$

where

$$\left[\mathbf{I}(\mathbf{v})\right]_{ij} = -\mathrm{E}\left[\frac{\partial^2 \Lambda(\mathbf{r};\mathbf{v})}{\partial v_i \partial v_j}\right]$$

- $\mathbf{I}(\mathbf{v})$ is the *Fisher information matrix*

CRB example: phase estimation

- Consider the CRB for the sine phase estimation problem:

$$r(k) = A\cos(2\pi f_0 k + \theta) + w(k), \qquad k = 0,1,\ldots,N-1$$

- The LLF:

$$\Lambda(\mathbf{r};\theta) = C_1 - \frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left(r(k) - A\cos(2\pi f_0 k + \theta)\right)^2$$

- The first derivative of LLF:

$$\frac{\partial\Lambda(\mathbf{r};\theta)}{\partial\theta} = \frac{-1}{\sigma^2}\sum_{k=0}^{N-1}\left(r(k) - A\cos(2\pi f_0 k + \theta)\right)A\sin(2\pi f_0 k + \theta)$$

$$= \frac{-A}{\sigma^2}\sum_{k=0}^{N-1}\left[r(k)\sin(2\pi f_0 k + \theta) - \tfrac{A}{2}\sin(4\pi f_0 k + 2\theta)\right]$$

CRB example: phase estimation...

- The second derivative of LLF:

$$\frac{\partial^2\Lambda(\mathbf{r};\theta)}{\partial\theta^2} = \frac{-A}{\sigma^2}\sum_{k=0}^{N-1}\left[r(k)\cos(2\pi f_0 k + \theta) - A\cos(4\pi f_0 k + 2\theta)\right]$$

- Negative expected value:

$$-\mathrm{E}\left[\frac{\partial^2\Lambda(\mathbf{r};\theta)}{\partial\theta^2}\right] = \frac{A}{\sigma^2}\sum_{k=0}^{N-1}\left[A\cos^2(2\pi f_0 k + \theta) - A\cos(4\pi f_0 k + 2\theta)\right]$$

$$= \frac{A^2}{\sigma^2}\sum_{k=0}^{N-1}\left[\tfrac{1}{2} + \tfrac{1}{2}\cos(4\pi f_0 k + 2\theta) - \cos(4\pi f_0 k + 2\theta)\right]$$

$$\approx \frac{NA^2}{2\sigma^2}$$

CRB example: phase estimation...

- Hence, the Cramer-Rao bound is obtained as:

$$\operatorname{var}(\hat{\theta}) \geq CRB = \frac{-1}{\mathrm{E}\left[\frac{\partial^2}{\partial A^2}\Lambda(\mathbf{r};\theta)\right]} = \frac{2\sigma^2}{NA^2} = \frac{1}{N \times SNR}$$

- Better phase estimates (= small variance) can be obtained with high SNR and by increasing the no. of samples

Modified CRB

- The true Cramer-Rao bound is often difficult to compute
- Requires explicit elimination of nuisance parameters
- Alternative: Replace 2nd derivative with time-domain integral:

$$MCRB(v_i) = \frac{N_0/2}{\mathrm{E}_u\left[\int_0^{T_0} \left|\partial s(t,\mathbf{v},\mathbf{u})/\partial v_i\right|^2 dt\right]}$$

where $s(t)$ is (passband) signal, $\mathbf{v}$ = sync parameters,

$\mathbf{u}$ = unwanted parameters

and v_i is the ith element in sync parameter vector $\mathbf{v}$.

- Always MCRB $\leq$ CRB ('looser bound')

Summary

Today we discussed

I What is synchronization

II ML estimation principles

III ML example: sinusoid phase estimation

IV Cramer-Rao lower bound

Next: Synchronization: overview

Department of Signal Processing and Acoustics