SGN 21006 Advanced Signal Processing: Lecture 12 Exam questions

Ioan Tabus

Department of Signal Processing
Tampere University of Technology
Finland

(5 points) Consider the noise cancellation problem and its solution by the fixed filter and adaptive filter schemes drawn below. First define the signals involved and then state what information is needed for implementing each of the schemes. Describe possible advantages and drawbacks for each scheme.
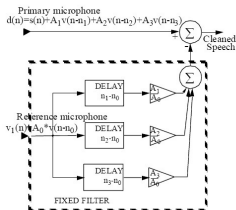
FIXED FILTER NOISE CANCELLING
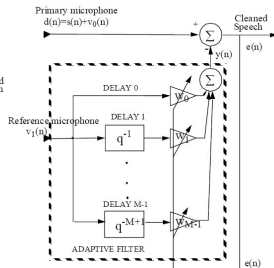

ADAPTIVE NOISE CANCELLING

- (2 points) Make a parallel between optimal filtering and adaptive filtering, considering the type of processing (batch versus sequential).
- (2 points) Which of the following methods are instances of optimal filtering and which are adaptive filtering: (a) Wiener filtering (b) LMS (c) NLMS (d) RLS (e) LS

- *Hint:*
    * In optimal filtering (or batch, or framewise) the input and desired signals are available for a given time-window, $1, \ldots, N$, and the optimal parameters of the linear filter, and subsequently the filter output for that time window, are computed only once
    * In adaptive filtering the input and desired signals are provided to the algorithm sequentially, and at every time instant a set of parameters of the linear filter are computed or updated, and they are used to compute the output of the filter for only that time instance

Lecture 1: Optimal or Adaptive Linear Filtering Modules in Applications

▶ (5 points) Draw the diagram of an adaptive filter block and specify the involved signals. Draw the block diagram of adaptive noise cancelled (or channel equalization, or echo cancellation) and specify the meaning in this application of each signal.

▶ (2 points) Draw the diagram of an adaptive filter block and specify the involved signals. For which application(s) the error signal represents the useful signal?
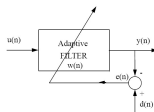
Hint:
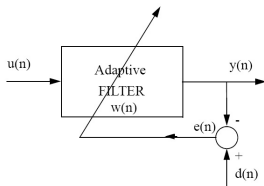
Lecture 1: A simple adaptive filter example

(3 points) Draw the diagram of an adaptive filter and write the equations for its operations, using LMS algorithm

Hint:

- Description of adaptive filtering operations, at any time instant, n:
  - * The reference noise $v_1(n)$ is processed by an adaptive filter, with time varying parameters $w_0(n), w_1(n), \ldots, w_{M-1}(n)$, to produce the output signal

  $$y(n) = \sum_{k=0}^{M-1} w_k(n) v_1(n - k)$$

  .
  - * The error signal is computed as $e(n) = d(n) - y(n)$.
  - * The parameters of the filters are modified in an adaptive manner. For example, using the LMS algorithm (the simplest adaptive algorithm)

  $$w_k(n + 1) = w_k(n) + \mu v_1(n - k) e(n) \qquad (LMS)$$

  where μ is the adaptation constant.

Lecture 2: Expectation for continuous and for discrete random variables

(3 points) For discrete random variables X taking values in the set $m, m+1, m+2, \ldots, M$ with the probability mass function $p(j) = Prob(X \leq j) - Prob(X < j) = Prob(X = j)$ the expectation of a function $g(X)$ is $E[g(X)] = \sum_{j=m}^{M} g(j)p(j)$. Define the mean, the variance and the second moment for the discrete random variable.

(3 points) A continuous random variable $x \in (-\infty, \infty)$ is fully described by the probability density function (pdf) $p(x)$. The expectation of a function $g(X)$ is $E[g(X)] = \int_{-\infty}^{\infty} g(x)p(x)dx$. Define the mean, the variance and the second moment for the random variable.

Hint:

1. mean, or first moment, or expected value for continuous r.v. X

$$\mu = E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

2. variance, or expected value of $(X - \mu)^2$, for continuous r.v. X

$$\sigma^2 = var(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$$

3. second moment, or expected value of X^2, for continuous r.v. X

$$E[X^2] = \int_{-\infty}^{\infty} x^2 p(x)dx = \sigma^2 + \mu^2$$

Lecture 2: Properties of expectation operator

(4 points) Prove the linearity property of the expectation operator: given two discrete random variables, X and Y, with joint pmf $g(x, y) = Prob(X = x; Y = y)$ and two constants a and b, then $E[aX + bY] = aE[X] + bE[Y]$.

Hint:

$$
\begin{aligned}
E[aX + bY] &= \sum_x \sum_y (ax + by) Prob(X = x; Y = y) \\
&= a \sum_x \sum_y x Prob(X = x; Y = y) + b \sum_x \sum_y y Prob(X = x; Y = y) \\
&= a \sum_x x Prob(X = x) + b \sum_y y Prob(Y = y) = aE[X] + bE[Y]
\end{aligned}
$$

(4 points) Prove that the expectation of a product of two independent random variables is equal to the product of expectations $E[XY] = E[X]E[Y]$

Hint:

If X is independent of Y, then $Prob(X = x, Y = y) = Prob(X = x)Prob(Y = y)$ and

$$
\begin{aligned}
E[XY] &= \sum_x \sum_y xy Prob(X = x; Y = y) \\
&= \sum_x x Prob(X = x) \sum_y y Prob(Y = y) = E[X]E[Y]
\end{aligned}
$$

(2 points) Is in general the expectation of a product of two random variables equal to the product of expectations, $E[XY] = E[X]E[Y]$?

(3 points) Show that the crosscorrelation function defined as $R(X, Y) = E[(X - E[X])(Y - E[Y])]$ is equal to the difference $E[XY] - E[X]E[Y]$.

Hint

$$
\begin{aligned}
R(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\
&= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]
\end{aligned}
$$

(3 points) Show that for independent variables the cross-correlation is 0.

Hint

$R(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] = 0.$

Problem statements

(3 points) State the "Least Mean squares" problem
(3 points) State the "Recursive least square" problem
(3 points) State the "Optimal Wiener filter" problem

Hint:

- Given the set of input samples $\{u(0), u(1), u(2), \ldots\}$ and the set of desired response $\{d(0), d(1), d(2), \ldots\}$

- In the family of filters computing their output according to

$$y(n) = \sum_{k=0}^{\infty} w_k u(n-k), \quad n = 0, 1, 2, \ldots$$

- Find the parameters $\{w_0, w_1, w_2, \ldots\}$ such as to minimize the mean square error defined as

$$J = E[e(n)^2]$$

where the error signal is

$$e(n) = d(n) - y(n) = d(n) - \sum_{l=0}^{\infty} w_l u(n-l)$$

(3 points) We know that for an optimum Wiener filter the errors obtained by the minimizing filter are orthogonal to the samples $u(i)$ which are used to compute the filter output, $E\left[e_o(n)u(n-k)\right] = 0$. Prove that also the output of the optimum filter is orthogonal to the estimation error $e_o(n)$.

Hint
We will index with o all the variables e.g. e_o, y_o computed using the optimal parameters $\{w_{o0}, w_{o1}, w_{o2}, \ldots\}$.
Let us compute the cross- correlation

$$E\left[e_o(n)y_o(n)\right] \quad = \quad E\left[e_o(n)\sum_{k=0}^{\infty} w_{ok}u(n-k)\right] = \sum_{k=0}^{\infty} w_{ok}E\left[u(n-k)e_o(n)\right] = 0$$

Otherwise stated, in words, we have the following **Corollary of Orthogonality Principle:**

▶ The estimation error $e_0(n)$ is orthogonal to the filter output $y_0(n)$ (when the filter has optimal parameters).

(4 points) We know that for an optimum Wiener filter the errors obtained by the minimizing filter are orthogonal to the the the samples $u(i)$ which are used to compute the filter output, $E\left[e_o(n)u(n-k)\right] = 0$. Starting from this, find the Wiener-Hopf equations

$$\sum_{i=0}^{\infty} w_{oi}r(i-k) = p(-k), \quad k = 0, 1, 2, \ldots \quad WIENER-HOPF$$

Lecture 3: AR process: autocorrelations and crosscorrelations

(6 points) Consider the AR process $x(n) + a_1 x(n-1) + a_2 x(n-2) = v(n)$. Derive the Yule Walker equations:

$$r_x(0) + a_1 r_x(1) + a_2 r_x(2) = \sigma_v^2$$
$$r_x(1) + a_1 r_x(0) + a_2 r_x(1) = 0$$
$$r_x(2) + a_1 r_x(1) + a_2 r_x(0) = 0$$

Hint:
First multiply in turn the equation with $x(n)$, $x(n-1)$ and $x(n-2)$ and then take the expectation

$$Ex(n) \times \rightarrow \quad Ex(n)x(n) + Ex(n)a_1 x(n-1) + Ex(n)a_2 x(n-2) = Ex(n)v(n)$$
$$\text{resulting in} \quad r_x(0) + a_1 r_x(1) + a_2 r_x(2) = Ex(n)v(n) = \sigma_v^2$$

$$Ex(n-1) \times \rightarrow \quad Ex(n-1)x(n) + Ex(n-1)a_1 x(n-1) + Ex(n-1)a_2 x(n-2) = Ex(n-1)v(n)$$
$$\text{resulting in} \quad r_x(1) + a_1 r_x(0) + a_2 r_x(1) = Ex(n-1)v(n) = 0$$

$$Ex(n-2) \times \rightarrow \quad Ex(n-2)x(n) + Ex(n-2)a_1 x(n-1) + Ex(n-2)a_2 x(n-2) = Ex(n-2)v(n)$$
$$\text{resulting in} \quad r_x(2) + a_1 r_x(1) + a_2 r_x(0) = Ex(n-2)v(n) = 0$$

The equality $Ex(n)v(n) = \sigma_v^2$ can be obtained multiplying the AR model difference equation with $v(n)$ and then taking expectations

$$Ev(n) \times \rightarrow \quad Ev(n)x(n) + Ev(n)a_1 x(n-1) + Ev(n)a_2 x(n-2) = Ev(n)v(n)$$
$$\text{resulting in} \quad Ev(n)x(n) = \sigma_v^2$$

since $v(n)$ is uncorrelated with older values, $x(n-\tau)$.

(3 points) State the steepest descent algorithm

Hint

Steepest descent search algorithm for finding the Wiener FIR optimal filter

Given

- the autocorrelation matrix $R = E\underline{u}(n)\underline{u}^T(n)$
- the cross-correlation vector $\underline{p}(n) = E\underline{u}(n)d(n)$

Initialize the algorithm with an arbitrary parameter vector $\underline{w}(0)$.

Iterate for $n = 0, 1, 2, 3, \ldots, n_{max}$

$$\underline{w}(n+1) = \underline{w}(n) + \mu[\underline{p} - R\underline{w}(n)]$$

Stop iterations if $\|\underline{p} - R\underline{w}(n)\| < \varepsilon$

$\square$

Designer degrees of freedom: $\mu, \varepsilon, n_{max}$

- ▶ (3 points) State the LMS algorithm
- ▶ (3 points) State the NLMS algorithm
- ▶ (2 points) Write the normalized LMS algorithm for the FIR filter with two paramteres, w_0 and w_1. How the algorithm will evolve if the input is
  $u(0) = 0$, $u(1) = 0$, $u(2) = 1$, $u(3) = 1$, $u(4) = u(5) = u(6) = \ldots = 0$ and the desired input is
  $d(0) = 0$, $d(1) = 0$, $d(2) = 0$, $d(3) = 1$, $d(4) = d(5) = d(6) = \ldots = 0$ (consider different situations for the initial weights).
- ▶ (2 points) Describe qualitatively the transient behavior of the LMS algorithm when changing the step size of the algorithm (in terms of transient time and steady state mean error).
- ▶ (2 points) Describe qualitatively the transient behavior of the LMS algorithm for different values of the eigenvalue spread (in terms of transient time and steady state mean error).

▶ (6 points) Consider a FIR(1) filter $y(n) = w(n)u(n)$ where all quantities are scalars. We intend to minimize the time varying cost function

$$J(n) = e(n)^2 + \alpha w(n)^2$$

where $e(n)$ is the estimation error

$$e(n) = d(n) - w(n)u(n)$$

$d(n)$ is the desired response, $u(n)$ is the input, and α is a constant. Show that the time update for the parameter vector $w(n)$ is defined by

$$w(n+1) = (1 - \mu\alpha)w(n) + \mu u(n)e(n)$$

What is the role of the constant α (comment the cases of very large α and very small α).

Lecture 7

- (3 points) (a) Find the least squares estimate of w_0 in the very simple model $y(t) = w_0$ when the desired data $d(t)$ is given for $t = 1, \ldots, N$ (therefore the input is assumed $u(t) = 1$ for $t = 1, \ldots, N$). What is the significance of the estimate?

  (3 points) (b) Find the recursive least squares solution $w_0(N)$ for the model at (a) as a simple equation connecting $w_0(N)$ to $w_0(N-1)$ (by elementary derivations, no need to use the general RLS equations). Try to find also the exponentially weighted solution ($\beta(n, i) = \lambda^{n-i}$).

- (3 points) State the recursive least squares estimation problem (model, data available, criterion to be minimized). Explain the role of the forgetting factor. What is the "length of memory" of the algorithm if the forgetting factor has the exponential form, $\beta(n, i) = \lambda^{n-i}$?

Overall lectures: Application descriptions

- (5 points) Draw the structure of an adaptive echo canceler. Discuss the significance of each signal.
- (5 points) Application description: Draw the structure of an adaptive noise canceller. Discuss the significance of each signal.
- (5 points) Draw the structure of an adaptive channel equalizer. Discuss the significance of each signal.

Spectrum estimation

1. (6 points)

   1.1 (2 points) The first definition of the power spectrum density (PSD) is

   $$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-i\omega k}$$

   What modification to the definition is needed for transforming it into a method for estimating the PSD?

   1.2 (2 points) What are the two main ways of estimating the autocovariance function $r(\tau)$ from a set of data $y(1), \ldots, y(N)$? Which of the two methods is preferred as an intermediate step for estimating the PSD by the correlogram method?

   1.3 (2 points) A second definition of the power spectrum density (PSD) is

   $$\phi(\omega) = \lim_{N\to\infty} E\left\{\frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\omega t}\right|^2\right\}$$

   What modifications to the definition are needed in order to transforming it into a method for estimating the PSD?

2. (10 points)

   2.1 (6 points) Explain the smearing effect and the leakage effect using the hint below. Also provide an evaluation of the resolution of the periodogram.
   Hint: The mean of the periodogram is connected to the DFT of the Bartlett window by the convolution

   $$E\left\{\hat{\phi}_p(\omega)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi(\zeta)W_B(\omega-\zeta)d\zeta$$

   The main lobe of the Bartlett window in the frequency domain has a 3dB-width proportional to $\frac{1}{N}$.

Spectrum estimation

1. (10 points)

    1.1 (6 points) Describe a method for AR spectrum estimation.
    1.2 (4 points) In what respects the AR spectrum estimate is different from the periodogram spectrum estimate? (do not specify differences of the estimation methods, but differences in the features of the estimated spectrum)