

# Adaptive Systems

Winter Term 2018/19

Instructor:  
Thomas Wilding

Assistants:  
—

Signal Processing and Speech Communication Laboratory, Inffeldgasse 16c/EG  
written by Bernhard Geiger, Christian Stetco and Thomas Wilding  
last update October 1, 2018 from Thomas Wilding

# Organizational Information

## Course Webpage

You can find a pdf version of the handout, homeworks and additional material on our webpage.  
<http://www.spssc.tugraz.at/courses/adaptive/>

## Newsgroup

There is a newsgroup for the discussion of all course-relevant topics at:  
`news:tu-graz.lv.adaptive`

## Schedule

Seven or eight meetings ( $\approx 90$  minutes each) on Fridays from 10:15 to 12:00 in lecture hall HS i12. Please refer to TUGraz-online or our website to get the actual schedule.

## Grading

Three homework assignments consisting of analytical problems as well as simulations using Matlab or Octave (30 to 35 points each, 100 points in total without bonus problems). Solving bonus problems gives additional points. Work should be done in pairs.


achieved points	grade
$\geq 88$	1
75...87	2
62...74	3
49...61	4
$\leq 48$	5

A delayed submission results in a penalty of **10 points per day**. Submitting your work as a L<sup>A</sup>T<sub>E</sub>X-document can earn you up to 3 (additional) points.

## Prerequisites

- (Discrete-time) Signal Processing (FIR/IIR Filters,  $z$ -Transform, DTFT, ...)
- Stochastic Signal Processing (Expectation Operator, Correlation, ...)
- Linear Algebra (Matrix Calculus, Eigenvector/-value Decomposition, Gradient, ...)
- Matlab or Octave

## Additional Information

Note that we do not cover all problems in class. Some of them are additional problems which you can solve by your own, e.g. during exam preparation or personal interest. All the problems which are additional are marked with a house icon .

If you detect any typos or errors in the handout, just write an e-Mail to [thomas.wilding@tugraz.at](mailto:thomas.wilding@tugraz.at).

# Contents

<b>1. The Optimum Linear Filtering Problem—LS and Wiener Filters</b>	<b>4</b>
1.1. Transversal Filter . . . . .	4
1.2. The Linear Filtering Problem . . . . .	4
1.3. Least-Squares Filters . . . . .	5
1.4. The Wiener Filter . . . . .	5
1.5. System Identification . . . . .	7
1.6. System Identification in a Noisy Environment . . . . .	7
1.7. Iterative Solution without Matrix Inversion—Gradient Search . . . . .	8
<b>2. Adaptive Transversal Filter Using The LMS Algorithm</b>	<b>9</b>
2.1. The LMS Adaptation Algorithm . . . . .	9
2.2. Normalized LMS Adaption Algorithm . . . . .	10
2.3. System Identification Using an Adaptive Filter . . . . .	11
<b>3. Interference Cancelation</b>	<b>14</b>
<b>4. Adaptive Linear Prediction</b>	<b>16</b>
4.1. Autoregressive spectrum analysis . . . . .	16
4.2. Linear prediction . . . . .	16
4.3. Yule-Walker Equations . . . . .	17
4.4. Periodic Interference Cancelation without an External Reference Source . . . . .	18
<b>5. Adaptive Equalization</b>	<b>19</b>
5.1. Principle . . . . .	19
5.2. Decision-Directed Learning . . . . .	20
5.3. Alternative Equalizer Structures . . . . .	21
<b>A. Moving Average (MA) Process</b>	<b>23</b>
<b>B. Autoregressive (AR) Process</b>	<b>23</b>

# 1. The Optimum Linear Filtering Problem—Least-Squares and Wiener Filters

## 1.1. Transversal Filter

We write the convolution sum as an inner vector product

$$y[n] = \sum_{k=0}^{N-1} c_k^*[n]x[n-k] = \mathbf{c}^H \mathbf{x}[n].$$

where

	$n$	...	time index, $n \in \mathbb{Z}$
	$x[n]$	...	input sample at time $n$
	$y[n]$	...	output sample at time $n$
	$(\cdot)^H$	...	Hermitian transpose
$\mathbf{x}[n]$	$= [x[n], x[n-1], \dots, x[n-N+1]]^T$	...	tap-input vector at time $n$
$\mathbf{c}^H[n]$	$= [c_0^*[n], c_1^*[n], \dots, c_{N-1}^*[n]]$	...	Hermitian transpose of coefficient vector at time $n$ (time-varying system)
	$N$	...	number of coefficients, length of $\mathbf{x}[n]$
	$N-1$	...	number of delay elements, filter order

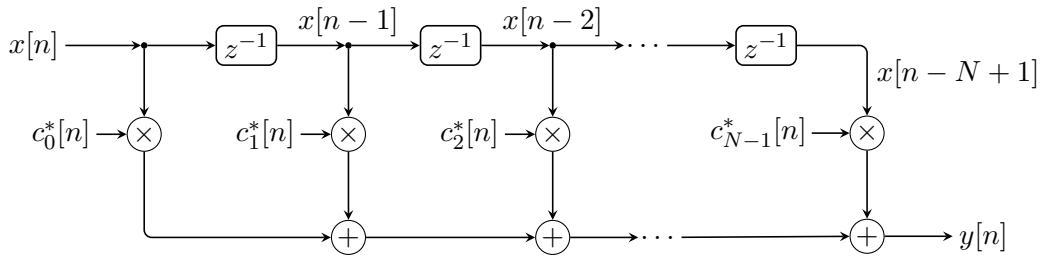


Figure 1: Transversal filter structure.

Special case:  $\mathbf{c}[n] = \mathbf{c} \rightarrow$  time-invariant FIR filter of order  $N-1$

## 1.2. The Linear Filtering Problem

The problem is to approximate a desired signal  $d[n]$  by filtering the input signal  $x[n]$ . For simplicity, we first consider a fixed (i.e., non-adaptive) filter  $\mathbf{c}[n] = \mathbf{c}$ .

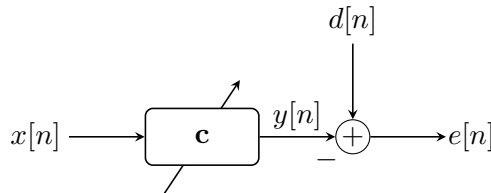


Figure 2: The linear filtering problem.

The goal is to find the optimum filter coefficients  $\mathbf{c}$ . But what does *optimum* mean?

### 1.3. Least-Squares Filters

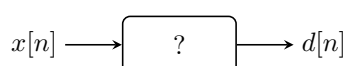
Consider a (finite) set of observations of  $\{x[n]\}$  and of  $\{d[n]\}$  is given (e.g., all past samples from  $n = 0$  to now). We define a *deterministic cost function* as

$$J_{LS}(\mathbf{c}) = \sum_{k=0}^n |e[k]|^2,$$

and the problem is to find those filter coefficients that minimize this cost function:

$$\mathbf{c}_{LS} = \underset{\mathbf{c}}{\operatorname{argmin}} J_{LS}(\mathbf{c}).$$

**Problem 1.1.** The following input/output measurements performed on a black box are given:



$n$	$x[n]$	$d[n]$
0	-1	-3
1	-1	-5
2	1	0

Find the optimum Least-Squares Filter with  $N = 2$  coefficients. Use matrix/vector notation for the general solution. Note that the input signal  $x[n]$  is applied to the system at time  $n = 0$ , i.e.,  $x[-1] = 0$ .

**Problem 1.2.** The previous problem has demonstrated that gradient calculus is important. To practice this calculus, determine  $\nabla_{\mathbf{c}} J(\mathbf{c})$  for the following cost functions:

- (i)  $J(\mathbf{c}) = K$
- (ii)  $J(\mathbf{c}) = \mathbf{c}^T \mathbf{v} = \mathbf{v}^T \mathbf{c} = \langle \mathbf{c}, \mathbf{v} \rangle$
- (iii)  $J(\mathbf{c}) = \mathbf{c}^T \mathbf{c} = \|\mathbf{c}\|^2 = \langle \mathbf{c}, \mathbf{c} \rangle$
- (iv)  $J(\mathbf{c}) = \mathbf{c}^T \mathbf{A} \mathbf{c}$ , where  $\mathbf{A}^T = \mathbf{A}$ .

### 1.4. The Wiener Filter

We consider  $x[n]$  and  $d[n]$  as (jointly) stationary stochastic processes<sup>1</sup>. The cost function is now stochastic:

$$J_{MSE}(\mathbf{c}) = \mathbb{E}\{|e[n]|^2\} \dots \text{Mean Squared Error (MSE)}$$

and the optimum solution in the MSE sense is obtained as:

$$\mathbf{c}_{MSE} = \underset{\mathbf{c}}{\operatorname{argmin}} J_{MSE}(\mathbf{c}).$$

**Problem 1.3.** The autocorrelation sequence of a stochastic process  $x[n]$  is defined as

$$r_{xx}[n, k] := \mathbb{E}\{x[n+k]x^*[n]\}.$$

If  $x[n]$  is *stationary*, then the autocorrelation sequence does not depend on time  $n$ , i.e.,

$$r_{xx}[k] = \mathbb{E}\{x[n+k]x^*[n]\}.$$

Calculate the autocorrelation sequence for the following signals ( $A$  and  $\theta$  are constant and  $\varphi$  is uniformly distributed over  $(-\pi, \pi]$ ):

<sup>1</sup>Note that if two processes are jointly WSS, they are WSS. The converse, however, is not necessarily true (i.e., two WSS processes need not be jointly WSS).

- (i)  $x[n] = A \sin(\theta n)$
- (ii)  $x[n] = A \sin(\theta n + \varphi)$
- (iii)  $x[n] = Ae^{j(\theta n + \varphi)}$

**Problem 1.4.** For the optimum linear filtering problem, find  $\mathbf{c}_{MSE}$  (i.e., the *Wiener-Hopf equation*). What statistical measurements must be known to get the solution?

**Problem 1.5.** Assume that  $x[n]$  and  $d[n]$  are a jointly wide-sense stationary, zero-mean processes.

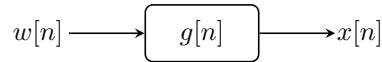
- (i) Specify the autocorrelation matrix  $\mathbf{R}_{xx} = \mathbb{E}\{\mathbf{x}[n]\mathbf{x}^T[n]\}$ .
- (ii) Specify the cross-correlation vector  $\mathbf{p} = \mathbb{E}\{d[n]\mathbf{x}[n]\}$ .
- (iii) Assume that  $d[n]$  is the output of a linear FIR filter to the input  $x[n]$ , i.e.,  $d[n] = \mathbf{h}^T \mathbf{x}[n]$ . Furthermore,  $\dim(\mathbf{h}) = \dim(\mathbf{c})$ . What is the optimal solution in the MSE sense?

**Problem 1.6.** In order to get the MSE-optimal coefficients, the first  $N$  samples of  $r_{xx}[k]$ , the auto-correlation sequence of  $x[n]$ , and the cross-correlation between the tap-input vector  $\mathbf{x}[n]$  and  $d[n]$  need to be known. This and the next problem are to practice the computation of correlations.

Let the input signal be  $x[n] = A \sin(\theta n + \varphi)$ , where  $\varphi$  is a random variable, uniformly distributed over  $(-\pi, \pi]$ , usually written as  $\varphi \sim \mathcal{U}(-\pi, \pi]$ .

- (i) Calculate the auto-correlation sequence  $r_{xx}[k]$ .
- (ii) Write the auto-correlation matrix  $\mathbf{R}_{xx}$  for a Wiener-filtering problem with  $N = 1$ ,  $N = 2$ , and  $N = 3$  coefficients.
- (iii) Answer for these 3 cases, whether the Wiener-Hopf equation can be solved or not?
- (iv) Repeat the last tasks for the following input signal:  $x[n] = Ae^{j(\theta n + \varphi)}$ .

**Problem 1.7.** The input signal  $x[n]$  is now zero-mean white noise  $w[n]$  filtered by an FIR filter with impulse response  $g[n]$ .



Find the auto-correlation sequence  $r_{xx}[k]$ .

Filtering with the Wiener filter allows us to make a few statements about the statistics of the error signal:

**Theorem 1** (Principle of Orthogonality). *The estimate  $y[n]$  of the desired signal  $d[n]$  (stationary process) is optimal in the sense of a minimum mean squared error if, and only if, the error  $e[n]$  is orthogonal to the input  $x[n - m]$  for  $m = 0 \dots N - 1$ .*

*Proof.* Left as an exercise. □

**Corollary 1.** *When the filter operates in its optimum condition, also the error  $e[n]$  and the estimate  $y[n]$  are orthogonal to each other.*

*Proof.* Left as an exercise. □

**Problem 1.8.** Show that the minimum mean squared error equals

$$J_{MMSE} = J(\mathbf{c}_{MSE}) = \mathbb{E}\{|d[n]|^2\} - \mathbf{p}^H \mathbf{R}_{xx}^{-1} \mathbf{p}.$$

## 1.5. System Identification

We now apply the solution of the linear filtering problem to system identification. Let  $d[n] = \mathbf{h}^H \mathbf{x}[n]$ , where  $\mathbf{h}$  is the impulse response of the system to be identified.

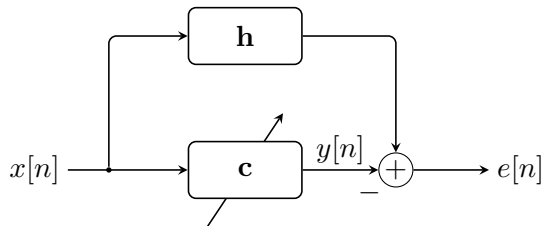


Figure 3: The system identification problem in a noise-free environment.

**Problem 1.9.** Let the order of the unknown system be  $M - 1$ , and let the order of the Wiener filter be  $N - 1$ , where  $N \geq M$ . Determine the MSE-optimal solution under the assumption that the autocorrelation sequence  $r_{xx}[k]$  of the input signal is known.

**Problem 1.10.** Repeat the previous problem when the order of the unknown system is  $M - 1$  and the order of the Wiener filter is  $N - 1$  with  $N < M$ . Use vector/matrix notation!

**Problem 1.11.** Let the order of the unknown system be 1 ( $d[n] = h_0x[n] + h_1x[n-1]$ ) but the Wiener filter is just a simple gain factor ( $y[n] = c_0x[n]$ ). Determine the optimum value for this gain factor. The autocorrelation sequence  $r_{xx}[k]$  of the input signal is known. Consider the cases when  $x[n]$  is white noise and also when  $x[n]$  is not white.

## 1.6. System Identification in a Noisy Environment

In contrary to the previous scenario, we now consider the case where the output signal of the system we want to identify is superimposed by a noise signal  $w[n]$ , as depicted in Fig. 4.

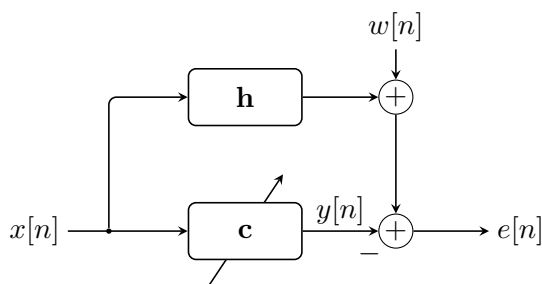


Figure 4: The system identification problem in a noisy environment.

The desired signal is now given as

$$d[n] = \mathbf{h}^H \mathbf{x}[n] + w[n]$$

where  $\mathbf{h}$  is the impulse response of the system to be identified and  $w[n]$  is stationary, additive noise.

**Problem 1.12.** Show that the optimal coefficient vector of the Wiener filter equals the impulse response of the system, i.e.,  $\mathbf{c}_{MSE} = \mathbf{h}$  if, and only if,  $w[n]$  is orthogonal to  $x[n - m]$  for  $m = 0 \dots N - 1$ .

**Problem 1.13.** Under what condition is the minimum mean squared error equal to  $J_{MSE}(\mathbf{c}_{MSE}) = \mathbb{E}\{|w[n]|^2\}$ ?

## 1.7. Iterative Solution without Matrix Inversion—Gradient Search

Recall that for the optimal filtering problem the cost function evaluates to

$$J_{MSE}(\mathbf{c}) = \mathbb{E}\{|d[n] - \mathbf{c}^H \mathbf{x}[n]|^2\} = \mathbb{E}\{|d[n]|^2\} - 2\mathbf{p}^H \mathbf{c} + \mathbf{c}^H \mathbf{R}_{xx} \mathbf{c}$$

and that the gradient of this cost function with respect to the coefficient vector  $\mathbf{c}$  equals

$$\nabla_{\mathbf{c}} J_{MSE}(\mathbf{c}) = 2(\mathbf{R}_{xx} \mathbf{c} - \mathbf{p}).$$

In Problem 1.4 we used these expressions to derive the Wiener-Hopf solution, which required the inversion of the autocorrelation matrix  $\mathbf{R}_{xx}$ .

In contrary to that, the *Gradient Search Method* is an iterative method which updates the coefficient vector  $\mathbf{c}[n]$  depending on the gradient of the cost function in a direction minimizing the MSE. Thus, this iterative algorithm is also called the *Method of Steepest Descent*. Mathematically, the coefficient update rule is given by

$$\mathbf{c}[n] = \mathbf{c}[n - 1] + \mu (\mathbf{p} - \mathbf{R}_{xx} \mathbf{c}[n - 1])$$

where  $\mu$  is a stepsize parameter and where the term in parentheses is the negative gradient, i.e.,

$$\mathbf{p} - \mathbf{R}_{xx} \mathbf{c}[n - 1] = -\nabla_{\mathbf{c}} J_{MSE}(\mathbf{c}) \Big|_{\mathbf{c}=\mathbf{c}[n-1]}.$$

**Problem 1.14.** Assuming convergence, show that this algorithm converges toward the MSE-optimal coefficient vector  $\mathbf{c}_{MSE}$ . Derive the range for the step size parameter  $\mu$  for which the algorithm is stable.

**Problem 1.15.** Calculate the convergence time constant(s) of the decay of the misalignment vector  $\mathbf{v}[n] = \mathbf{c}[n] - \mathbf{c}_{MSE}$  (coefficient deviation) when the gradient search method is applied.



### 🏠 Problem 1.16.

- (i) Express the MSE as a function of the misalignment vector  $\mathbf{v}[n] = \mathbf{c}[n] - \mathbf{c}_{MSE}$ .
- (ii) Find an expression for the learning curve  $J_{MSE}[n] = J_{MSE}(\mathbf{c}[n])$  when the gradient search is applied.
- (iii) Determine the time constant(s) of the learning curve.

**Problem 1.17.** Consider a noise-free system identification problem where both the adaptive and the unknown transversal filter have 2 coefficients. The statistics of the input signal are known as

$$\mathbf{R}_{xx} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}.$$

The Gradient Method with  $\mu = 1/2$  is used to solve for the filter coefficients. The coefficients of the unknown system are  $\mathbf{h} = [2, 1]^T$ .

- (i) Simplify the adaptation algorithm by substitution for  $\mathbf{p} = \mathbb{E}\{\mathbf{x}[n]d[n]\}$  according to the given system identification problem. Additionally, introduce the misalignment vector  $\mathbf{v}[n]$  and rewrite the adaptation algorithm such that  $\mathbf{v}[n]$  is adapted.
- (ii) The coefficients of the adaptive filter are initialized with  $\mathbf{c}[0] = [0, -1]^T$ . Find an expression for  $\mathbf{v}[n]$  either analytically or by calculation of some (three should be enough) iteration steps. Do the components of  $\mathbf{v}[n]$  show an exponential decay? If yes, determine the corresponding time constant.
- (iii) Repeat the previous task when the coefficients are initialized with  $\mathbf{c}[0] = [0, 3]^T$ . Do the components of  $\mathbf{v}[n]$  show an exponential decay? If yes, determine the corresponding time constant.
- (iv) Repeat the previous task when the coefficients are initialized with  $\mathbf{c}[0] = [0, 0]^T$ . Do the components of  $\mathbf{v}[n]$  show an exponential decay? If yes, determine the corresponding time constant.
- (v) Explain, why an exponential decay of the components of  $\mathbf{v}[n]$  can be observed although the input signal  $x[n]$  is not white.

## 2. Adaptive Transversal Filter Using The LMS Algorithm

### 2.1. The LMS Adaptation Algorithm

We now analyze the LMS adaptation algorithm, whose update rule is given as:

$$\mathbf{c}[n] = \mathbf{c}[n-1] + \mu e^*[n] \mathbf{x}[n]$$

where

$$e[n] = d[n] - y[n] = d[n] - \mathbf{c}^H[n-1] \mathbf{x}[n]$$

$\mathbf{c}[n]$	...	new coefficient vector
$\mathbf{c}[n-1]$	...	old coefficient vector
$\mu$	...	step-size parameter
$e[n]$	...	error at time $n$
$d[n]$	...	desired output at time $n$
$\mathbf{x}[n]$	...	tap-input vector at time $n$

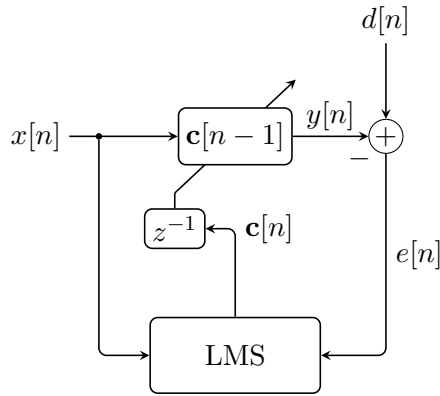


Figure 5: Adaptive transversal filter.

**How to choose  $\mu$ ?** As it was shown in the lecture course, a sufficient (deterministic) stability condition is given by

$$0 < \mu < \frac{2}{\|\mathbf{x}[n]\|^2} \quad \forall n$$

where  $\|\mathbf{x}[n]\|^2$  is the tap-input energy at time  $n$ .

Note that the stochastic stability conditions in related literature, i.e.,

$$0 < \mu < \frac{2}{\mathbb{E}\{\|\mathbf{x}[n]\|^2\}} \quad \forall n$$

or

$$0 < \mu < \frac{2}{N\sigma_x^2} \quad \text{for stationary input,}$$

only ensure ‘stability on average’.

## 2.2. Normalized LMS Adaption Algorithm

To make the step size parameter independent of the energy of the input signal, the normalized LMS can be used:

$$\mathbf{c}[n] = \mathbf{c}[n-1] + \frac{\tilde{\mu}}{\alpha + \mathbf{x}^H[n] \mathbf{x}[n]} e^*[n] \mathbf{x}[n]$$

where  $\alpha$  is a small positive constant to avoid division by zero.

**How to choose  $\tilde{\mu}$ ?**

Here the algorithm can be shown to be stable if (sufficient stability condition)

$$0 < \tilde{\mu} < 2.$$

## 2.3. System Identification Using an Adaptive Filter

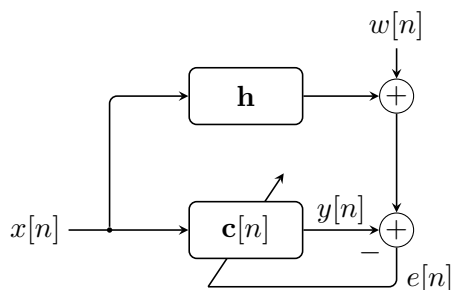


Figure 6: System identification using an adaptive filter and LMS.

**Minimum Error, Excess Error, and Misadjustment** If we can get only noisy measurements from the unknown system

$$d[n] = \sum_{k=0}^{M-1} h_k^* x[n-k] + w[n],$$

the MSE  $J_{MSE}[n] = \mathbb{E}\{|e[n]|^2\}$  does not vanish completely if the time goes toward infinity.

We assume  $x[n]$  and  $w[n]$  are jointly stationary, uncorrelated processes. The remaining error can be written as

$$\lim_{n \rightarrow \infty} J_{MSE}(\mathbf{c}[n]) = J_{\text{excess}} + J_{MSE}(\mathbf{c}_{MSE})$$

where  $J_{MSE}(\mathbf{c}_{MSE}) = \sigma_w^2$  is the minimum MSE (MMSE), which would be achieved by the Wiener-Hopf solution. The *excess error* ( $J_{MSE}(\mathbf{c}_{MSE})$ ) is caused by a remaining misalignment between the Wiener-Hopf solution and the coefficient vector at time  $n$ , i.e., it relates to  $\mathbf{v}[n] \neq 0$  (as it happens all the time for the LMS).

Finally, we define the *ratio* between the excess error and the MMSE as the *misadjustment*

$$M = \frac{J_{\text{excess}}}{J_{MSE}(\mathbf{c}_{MSE})} \approx \frac{\mu N \sigma_x^2}{2}.$$

From the stability bounds on  $\mu$  follows that  $0 < M < 1$ .

**Problem 2.1.** Let  $\bar{\tau}$  be the average convergence time constant defined by  $\bar{\tau} = 1/\mu\bar{\lambda}$ , where  $\bar{\lambda}$  is the average of all  $N$  eigenvalues of  $\mathbf{R}_{xx}$ . Show that the following trade-off between convergence time, filter order, and misadjustment exists:

$$\bar{\tau}M = \frac{N}{2}$$

As a consequence, a large filter order leads either to slow convergence or to a large misadjustment.

**Problem 2.2.** Convergence analysis of the LMS algorithm.

(i) Assuming convergence of the LMS algorithm

$$\mathbf{c}[n] = \mathbf{c}[n-1] + \mu e^*[n] \mathbf{x}[n],$$

find the limit  $\mathbf{c}_\infty$  of the sequence of coefficient vectors.

- (ii) Show that the following expression for a sufficient condition for convergence is true in the case of a noise-free system identification task

$$\|\mathbf{v}[n]\|^2 < \|\mathbf{v}[n-1]\|^2 \quad \forall n$$

where  $\mathbf{v}[n]$  is the misalignment vector  $\mathbf{v}[n] = \mathbf{c}[n] - \mathbf{c}_\infty$ . Which requirements on  $\mu$  can be derived from this expression?

- (iii) What happens when the environment is noisy?

**🏠 Problem 2.3. Joint Recursive Optimality.** For a wide class of adaptation algorithms (see [3] for more information), the underlying cost function  $J(\mathbf{c}, n)$  can be written as

$$J(\mathbf{c}, n) = (\mathbf{c} - \mathbf{c}[n-1])^T \mathbf{G}^{-1}[n] (\mathbf{c} - \mathbf{c}[n-1]) + \gamma^{-1}[n] (d[n] - \mathbf{c}^T \mathbf{x}[n])^2$$

where  $\mathbf{c}$  is the new coefficient vector of the adaptive transversal filter and  $\mathbf{c}[n-1]$  is the previous one,  $\mathbf{x}[n]$  is the tap-input vector, and  $d[n]$  is the desired output. Note that the expression  $d[n] - \mathbf{c}^T \mathbf{x}[n]$  is the *a-posteriori* error  $\epsilon[n]$ . The weights  $\mathbf{G}[n]$  and  $\gamma[n]$  are normalized such that

$$\gamma[n] + \mathbf{x}^T[n] \mathbf{G}[n] \mathbf{x}[n] = 1,$$

and  $\mathbf{G}[n]$  is symmetric ( $\mathbf{G}[n] = \mathbf{G}^T[n]$ ).

- (i) Find a recursive expression to adapt  $\mathbf{c}[n]$  given  $\mathbf{c}[n-1]$  such that the cost function  $J(\mathbf{c}, n)$  is minimized:

$$\mathbf{c}[n] = \arg \min_{\mathbf{c}} J(\mathbf{c}, n).$$

Note, this expression should contain the *a-priori* error  $e[n]$

$$e[n] = d[n] - \mathbf{c}^T[n-1] \mathbf{x}[n]$$

and not the *a-posteriori* error (Hint: find the ratio between the *a-posteriori* and the *a-priori* error first.).

- (ii) Determine the weights  $\mathbf{G}[n]$  and  $\gamma[n]$  for the case of the LMS algorithm and for the case of the Normalized LMS algorithm.

- (iii) Show that

$$\min_{\mathbf{c}} J(\mathbf{c}, n) = e^2[n]$$

for all  $n > 0$ .

**Problem 2.4.** For a noisy system identification problem, the following two measurements could be accomplished at the plant:  $\sigma_x^2 = 1$  and  $\sigma_d^2 = 2$ .  $x[n]$  and  $w[n]$  can be assumed to be stationary, zero-mean, white noise processes and orthogonal to each other. The unknown system can be assumed to be linear and time-invariant. The adaptive filter has been specified to have  $N = 100$  coefficients, and we can assume that no undermodeling occurs. The LMS algorithm is used to adapt the coefficients, and a maximum misadjustment of  $-10$ dB should be reached.

- (i) How should you choose the step size  $\mu$  and what convergence time constant will be obtained?
- (ii) After applying the adaptive filter another measurement has been performed and an error variance of  $\sigma_e^2 = 0.011$  has been obtained. Calculate  $\sigma_w^2$  and the ‘coefficient-to-deviation ratio’  $10 \log_{10} \left( \frac{\|\mathbf{h}\|^2}{\|\mathbf{v}\|^2} \right)$  in dB (i.e., a kind of signal-to-noise ratio).

**Problem 2.5. The Coefficient Leakage LMS Algorithm.** We will investigate the leaky LMS algorithm which is given by

$$\mathbf{c}[n] = (1 - \mu\alpha)\mathbf{c}[n-1] + \mu e^*[n]\mathbf{x}[n]$$

with the leakage parameter  $0 < \alpha \ll 1$ .

Consider a noisy system identification task (proper number of coefficients) and assume the input signal  $x[n]$  and the additive noise  $w[n]$  to be orthogonal. The input signal  $x[n]$  is zero-mean white noise with variance  $\sigma_x^2$ . Assume  $\mu$  and  $\alpha$  have been chosen to assure convergence. Determine where this algorithm converges to (on average). Compare your solution with the Wiener-Hopf solution. Also, calculate the average bias.

### 3. Interference Cancellation

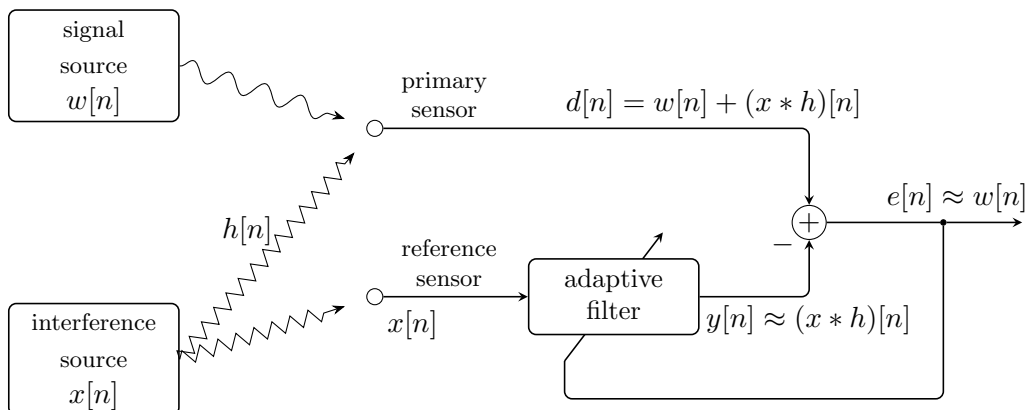


Figure 7: Adaptive noise canceler.

**The primary sensor** (i.e., the sensor for the desired signal  $d[n]$ ) receives the signal of interest  $w[n]$  corrupted by an interference that went through the so-called interference path. When the isolated interference is denoted by  $x[n]$  and the impulse response of the interference path by  $h[n]$ , the sensor receives

$$d[n] = w[n] + \mathbf{h}^T \mathbf{x}[n].$$

For simplification we assume  $\mathbb{E}\{w[n]x[n-k]\} = 0$  for all  $k$ .

**The reference sensor** receives the isolated interference  $x[n]$ .

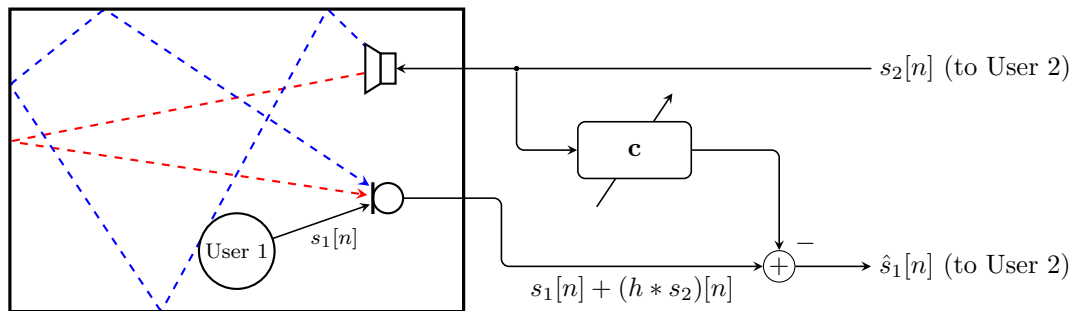
**The error signal** is  $e[n] = d[n] - y[n] = w[n] + \mathbf{h}^T \mathbf{x}[n] - y[n]$ . The adaptive filtering operation is perfect if  $y[n] = \mathbf{h}^T \mathbf{x}[n]$ . In this case the system output is  $e[n] = w[n]$ , i.e., the isolated signal of interest.

**FIR model for the interference path:** If we assume that  $\mathbf{h}$  is the impulse response of an FIR system (i.e.,  $\dim \mathbf{h} = N$ ) the interference cancellation problem is equal to the system identification problem.

**Problem 3.1.** We use a first-order transversal filter to eliminate an interference of the  $f_{AC} = 50$  Hz AC power supply from an ECG signal. The sampling frequency is  $f_s$  ( $f_s > 100$  Hz).

- (i) Using the method of *equating the coefficients*, express the optimum coefficients  $c_0$  and  $c_1$  that fully suppress the interference in terms of  $A := |H(e^{j\theta_{AC}})|$  and  $\vartheta := \arg H(e^{j\theta_{AC}})$ , i.e., in terms of the magnitude and the phase of the frequency response of the interference path at the frequency of the interference.
- (ii) Calculate the auto-correlation sequence  $r_{xx}[k]$  of the reference input signal as a function of the sampling frequency and build the auto-correlation matrix  $\mathbf{R}_{xx}$ .
- (iii) Determine the cross-correlation vector  $\mathbf{p}$  and solve the Wiener-Hopf equation to obtain the MSE-optimal coefficients. Show that these coefficients are equal to those found in (i).
- (iv) Determine the condition number  $\kappa = \frac{\lambda_{max}}{\lambda_{min}}$  of the auto-correlation matrix for the given problem as a function of the sampling frequency. Is the unit delay in the transversal filter a clever choice?

**Problem 3.2.** Consider the acoustic echo cancellation scenario below:



- (i) Assuming that all  $s_1$  and  $s_2$  are jointly stationary and uncorrelated, derive the coefficient vector  $\mathbf{c}$  optimal in the MSE sense.
- (ii) Given that the room has a dimension of 3 times 4 meters, and assuming that the speech signals are sampled with  $f_s = 8$  kHz, what order should  $\mathbf{c}$  have such that at least first-order reflections can be canceled? Note, that *physically* the impulse response of the room is infinite!
- (iii) Assume the filter coefficients are updated using the LMS algorithm to track changes in the room impulse response. What problems can occur?

## 4. Adaptive Linear Prediction

### 4.1. Autoregressive spectrum analysis

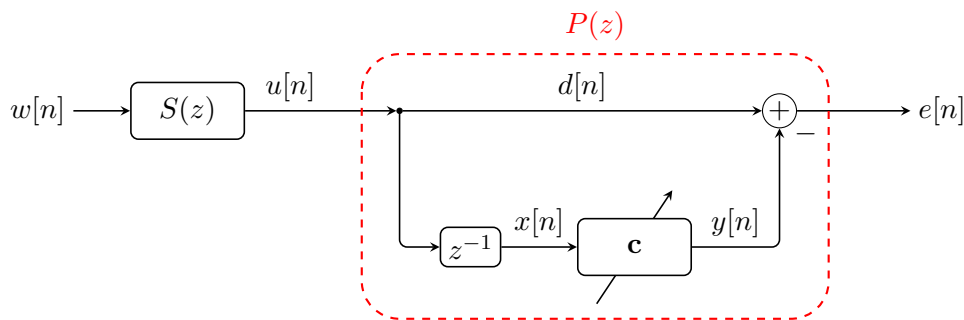


Figure 8: Linear prediction using an adaptive filter.

Let  $w[n]$  be a white input sequence, and let  $S(z)$  be an all-pole synthesis filter with difference equation

$$u[n] = w[n] - \sum_{k=1}^L a_k^* u[n-k].$$

In this case,  $u[n]$  is called an autoregressive (AR) process (see appendix B). We can estimate the AR coefficients  $a_1, \dots, a_L$  by finding the MSE-optimal coefficients of a linear predictor. Once the AR coefficients have been obtained, the squared-magnitude frequency response of the recursive process-generator filter can be used as an estimate of the power-spectral density (PSD) of the process  $u[n]$  (sometimes called *AR Modeling*).

**Problem 4.1. Autocorrelation Sequence of an AR Process** Consider the following difference equation

$$u[n] = w[n] + 0.5 u[n-1],$$

i.e., a purely recursive linear system with input  $w[n]$  and output  $u[n]$ .  $w[n]$  are samples of a stationary white noise process with zero mean and  $\sigma_w^2 = 1$ . Calculate the auto-correlation sequence  $r_{uu}[k]$  of the output.

### 4.2. Linear prediction

A linear predictor tries to predict the present sample  $u[n]$  from the  $N$  preceding samples  $u[n-1], \dots, u[n-N]$  using a linear combination:

$$\hat{u}[n] = \sum_{k=1}^N c_k^* u[n-k].$$

The prediction error is given by  $e[n] = u[n] - \hat{u}[n] = w[n] - \sum_{k=1}^L a_k^* u[n-k] - \sum_{k=1}^N c_k^* u[n-k]$ . Minimizing the mean-squared prediction error yields the proper predictor coefficients  $c_k$  for  $k = 1, \dots, N$ . In the ideal case ( $N = L$ ), the error is a minimum when only the non-predictable white noise excitation  $w[n]$  remains as  $e[n]$ . In this case, we obtain the (negative) AR coefficients:  $a_k = -c_k$  for  $k = 1, \dots, L$ .

For adaptive linear prediction (see Fig. 8), the adaptive transversal filter is the linear combiner, and an adaptation algorithm (e.g., the LMS algorithm) is used to optimize the coefficients and to minimize the prediction error.

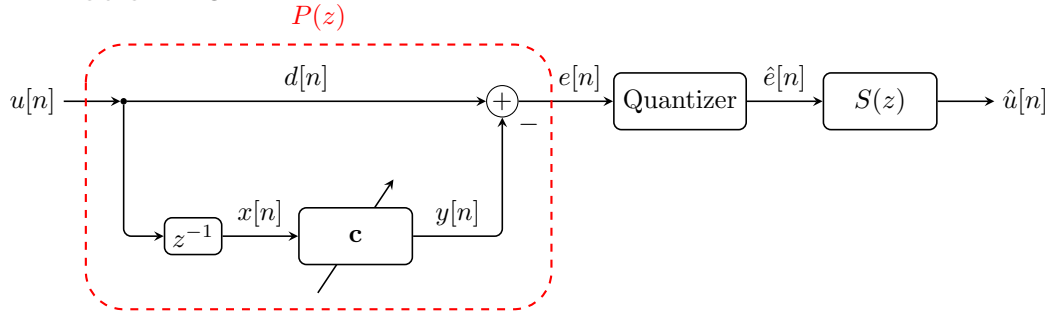


**Problem 4.2. AR Modeling.** Consider a linear prediction scenario as shown in Fig. 8. The mean squared error should be used as the underlying cost function. The auto-correlation sequence of  $u[n]$  is given by

$$r_{uu}[k] = 4/3 \cdot (1/2)^{|k|}.$$

Compute the AR coefficients  $a_1, a_2, \dots$  and the variance of the white-noise excitation  $\sigma_w^2$ . Start the calculation for an adaptive filter with 1 coefficient. Then, repeat the calculation for 2 and 3 coefficients.

### 🏠 Problem 4.3.



Consider the scenario depicted above, where  $u[n]$  is an AR process. The quantizer depicted shall have a resolution of  $B$  bits, and is modeled by an additive source with zero mean and variance  $\gamma\sigma_e^2$ , where  $\gamma$  is a constant depending on  $B$  and where  $\sigma_e^2$  is the variance of the prediction error  $e[n]$ .  $S(z)$  is the synthesis filter, which is the inverse of the prediction filter.

- (i) For a ideal prediction (i.e., the prediction error  $e[n]$  is white), compute the output SNR, which is given as

$$\frac{\mathbb{E}\{u^2[n]\}}{\mathbb{E}\{(u[n] - \hat{u}[n])^2\}} = \frac{\sigma_u^2}{\sigma_r^2}$$

where  $r[n] = u[n] - \hat{u}[n]$ .

- (ii) Repeat the previous task for the case where no prediction filter was used. What can you observe?

## 4.3. Yule-Walker Equations

In order to get the Wiener-Hopf solution for the MSE-optimal coefficient vector  $\mathbf{c}_{MSE}$

$$\mathbf{R}_{xx}\mathbf{c}_{MSE} = \mathbf{p},$$

we have to substitute  $x[n] = u[n-1]$  and  $d[n] = u[n]$  and get

$$\mathbf{R}_{uu}\mathbf{c}_{MSE} = \mathbf{r}_{uu+1}.$$

In non-vector notation this reads

$$\begin{bmatrix} r_{uu}[0] & r_{uu}[1] & \dots & r_{uu}[L-1] \\ r_{uu}^*[1] & r_{uu}[0] & \dots & r_{uu}[L-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{uu}^*[L-1] & r_{uu}^*[L-2] & \dots & r_{uu}[0] \end{bmatrix} \mathbf{c}_{MSE} = \begin{bmatrix} r_{uu}^*[1] \\ r_{uu}^*[2] \\ \vdots \\ r_{uu}^*[L] \end{bmatrix}.$$

These equations are termed the *Yule-Walker equations*. Note that in the ideal case  $\mathbf{c}_{MSE} = [-a_1, -a_2, \dots, -a_L]^T$  (given that the order of the transversal filter matches the order of the AR process).

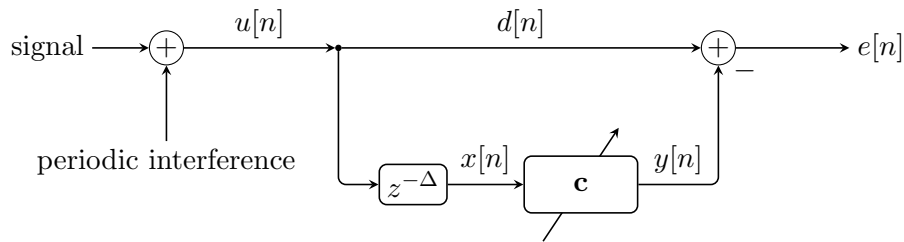


Figure 9: Canceling a periodic interference using a linear predictor.

#### 4.4. Periodic Interference Cancellation without an External Reference Source

Fig. 9 shows the usage of an adaptive linear predictor to remove a periodic interference of a broadband signal. The output is simply the whitened prediction error.

Things to be aware of:

- The delay length  $\Delta$  must be longer than the correlation time of the broadband signal (but not too long to avoid echoes).
- More coefficients yield a sharper filter and therefore less distortion of the broadband signal. But more coefficients increase also the convergence time.

**Problem 4.4. Periodic Interference Cancellation without an External Reference Source** Consider a measured signal  $u[n]$  that is the sum of a white-noise signal  $w[n]$  with variance  $\sigma_w^2 = 1$  and a sinusoidal:  $u[n] = w[n] + \cos(\pi/2 \cdot n + \varphi)$  (with  $\varphi$  a random phase offset).

- Calculate the auto-correlation sequence  $r_{uu}[k]$  of  $u[n]$ .
- Let's attenuate the sinusoidal using the setup of Fig. 9. A delay of 1 should be enough for the white  $v[n]$ . Compute the optimum coefficients  $c_0, c_1$  of the first-order adaptive filter in the sense of a minimum mean squared error.
- Determine the transfer function of the prediction-error filter, compute its poles and zeros, and plot the pole/zero diagram. Sketch its frequency response.

## 5. Adaptive Equalization

### 5.1. Principle

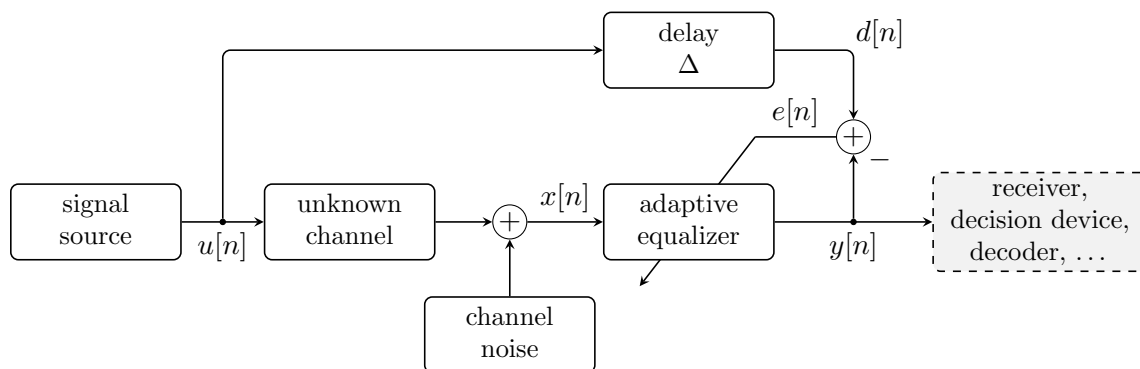


Figure 10: Adaptive equalization (or inverse modeling).

Fig. 10 shows the principle of adaptive equalization. The goal is to adapt the transversal filter to obtain

$$H(z)C(z) = z^{-\Delta},$$

i.e., to find the inverse (except for a delay) of the transfer function of the unknown channel. In the case of a communication channel, this eliminates the intersymbol interference (ISI) introduced by the temporal dispersion of the channel.

#### Difficulties:

1. Assume  $H(z)$  has a finite impulse response (FIR)  $\Rightarrow$  the inverse system  $H^{-1}(z)$  is an IIR filter. Using a finite-length adaptive transversal filter only yields an approximation of the inverse system.
2. Assume  $H(z)$  is a non-minimum phase system (FIR or IIR)  $\Rightarrow$  the inverse system  $H^{-1}(z)$  is not stable.
3. We typically have to introduce the extra delay  $\Delta$  (i.e., the group delay of the cascade of both the channel and the equalizer).

Situations where a reference, i.e., the original signal, is available to the adaptive filter:

- Audio: adaptive concert hall equalization, car HiFi, airplanes, ... (equalizer = *pre-emphasis* or *pre-distortion*; microphone where optimum quality should be received)
- Modem: transmission of an initial *training sequence* to adapt the filter and/or periodic interruption of the transmission to re-transmit a known sequence to re-adapt.

Often there is no possibility to access the original signal. In this case we have to 'guess' the reference: *Blind Adaptation*. Examples are *Decision-Directed Learning* or the *Constant Modulus Algorithm*, which exploit a-priori knowledge of the source.

**Problem 5.1. ISI and Open-Eye Condition** For the following equivalent discrete-time channel impulse responses

(i)  $\mathbf{h} = [0.8, 1, 0.8]^T$

(ii)  $\mathbf{h} = [0.4, 1, 0.4]^T$

(iii)  $\mathbf{h} = [0.5, 1, 0.5]^T$

calculate the worst-case ISI for binary data  $u[n] \in \{+1, -1\}$ . Is the channel's eye opened or closed?

**Problem 5.2. Least-Squares and MinMSE Equalizer** For a noise-free channel with given impulse response  $\mathbf{h} = [1, 2/3, 1/3]^T$ , compute the optimum coefficients of the equal-length, zero-delay equalizer in the least-squares sense. Can the equalizer open the channel's eye? Is the least-squares solution equivalent to the minimum-MSE solution for white data  $u[n]$ ?

**Problem 5.3. MinMSE Equalizer for a Noisy Channel** Consider a channel with impulse response  $\mathbf{h} = [1, 2/3, 1/3]^T$  and additive white noise  $\eta[n]$  with zero mean and variance  $\sigma_\eta^2$ . Compute the optimum coefficients of the equal-length equalizer in the sense of a minimum mean squared error.

## 5.2. Decision-Directed Learning

Let us now assume that we know the modulation alphabet of the digital transmission system (e.g., binary antipodal modulation, PSK, etc.). The demodulator chooses the output symbol as the element of the modulation alphabet with the minimum distance to the received signal. (For binary antipodal modulation this can be accomplished by a simple threshold device.)

If we now assume that the distortion by the channel is moderate, one can use the distance between the chosen output and the received signal as the error signal for adapting the equalizer (see Fig. 11).

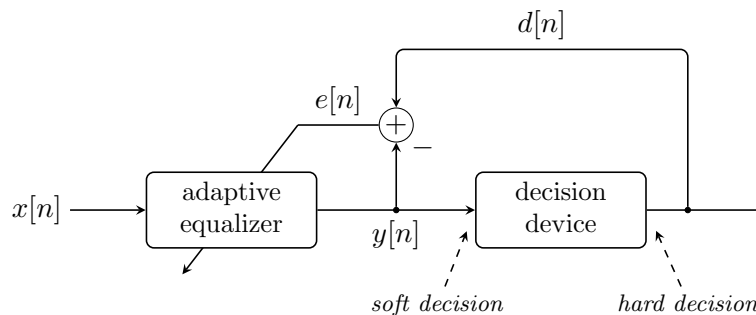


Figure 11: Decision-directed adaptive channel equalizer.

**Problem 5.4. Decision-Feedback Equalizer (DFE).** Consider the *feedback-only equalizer* in Fig. 12. Assume that the transmitted data  $u[n]$  is white and has zero mean.

- (i) For a general channel impulse response  $\mathbf{h}$  and a given delay  $\Delta$ , calculate the optimum (min. MSE) coefficients  $\mathbf{c}_b$  of the feedback equalizer.
- (ii) What is the resulting impulse response of the overall system when the equalizer operates at its optimum?

- (iii) Do the MSE-optimum coefficients  $\mathbf{c}_b$  of the feedback equalizer change for a noisy channel?

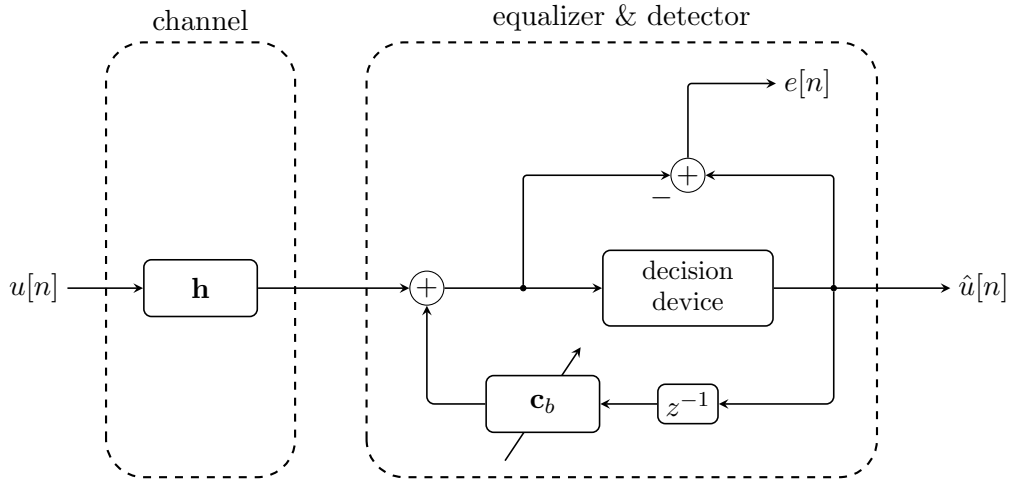


Figure 12: Decision-feedback equalizer.

### 5.3. Alternative Equalizer Structures

**Problem 5.5.** Extend the decision-feedback equalizer structure in Fig. 12 by an additional forward (or transversal) equalizer filter with coefficients  $\mathbf{c}_f$  right after the channel. Derive the design equation for both MSE-optimum  $\mathbf{c}_f$  and  $\mathbf{c}_b$  (use an overall coefficient vector  $\mathbf{c}^T = [\mathbf{c}_f^T \ \mathbf{c}_b^T]$ ).

**Problem 5.6. Fractionally-Spaced Equalizer.** A fractionally-spaced equalizer runs at a sampling rate that is higher than the symbol rate. Consider the  $T/2$ -fractionally-spaced equalizer (i.e., it runs at the double rate) in Fig. 13 where  $T$  is the symbol duration. The decision device is synchronized with the transmitted symbols, which correspond to the even-indexed samples at the double rate.

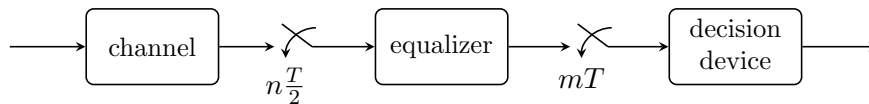


Figure 13: Fractionally-spaced equalizer.

The discrete-time description of the channel for the high sampling rate is

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + h_3 z^{-3} = 1/2 + z^{-1} + 1/2 z^{-2} + 1/4 z^{-3},$$

i.e., the unit delay  $z^{-1}$  corresponds to  $\frac{T}{2}$ .

- (i) Calculate the coefficients of the equal-length equalizer

$$C(z) = c_0 + c_1 z^{-1} + c_2 z^{-2} + c_3 z^{-3}$$

such that the cascade of the given channel and the equalizer  $H(z)C(z) = 1$ , i.e., it enables a delay-free and ISI-free transmission.

- (ii) Calculate the coefficients of the equalizer such that the cascade is a pure delay of 1 symbol, i.e.,  $H(z)C(z) = z^{-2}$ .

(iii) Consider the channel to be noisy (additive white noise). Compute the noise gains of the two equalizers of the previous tasks. Which one should be chosen?

(iv) Let the channel be

$$H(z) = 1 + 1/2 z^{-1} + 1/4 z^{-2} + 1/8 z^{-3}.$$

Compute again the coefficients of an equal-length equalizer.

## A. Moving Average (MA) Process

A stationary MA process  $u[n]$  satisfies the difference equation

$$u[n] = v[n] + \sum_{k=1}^K g^*[k]v[n-k],$$

where  $K$  is the order and  $v[n]$  is white noise with zero mean and variance  $\sigma_v^2$ , i.e.,  $u[n]$  is white noise filtered by an FIR filter with impulse response  $g[n]$  where  $g[0] = 1$  (as defined in [6, 7]).

The *auto-correlation sequence* of the output  $u[n]$  is given by (see Problem 1.7)

$$r_{uu}[k] = \sigma_v^2 \sum_i g[i]g^*[i-k].$$

The variance of  $u[n]$  can be obtained by setting  $k = 0$ :

$$\sigma_u^2 = \sigma_v^2 \sum_i |g[i]|^2.$$

The factor  $\sum_i |g[i]|^2$  is termed the *Noise Gain*.

## B. Autoregressive (AR) Process

A stationary AR process  $u[n]$  satisfies the recursive linear difference equation

$$u[n] = v[n] - \sum_{k=1}^L a_k u[n-k],$$

where  $L$  is the order, and  $v[n]$  is white noise with zero mean and variance  $\sigma_v^2$ , i.e.,  $u[n]$  is white noise filtered by an all-pole IIR filter. The process is fully specified by the AR coefficients  $a_k$ ,  $k = 1 \dots L$  and the white-noise variance  $\sigma_v^2$ .

The *auto-correlation sequence*  $r_{uu}[n]$  can be expressed by a zero-input version of the above recursive difference equation (see Problem 4.1)

$$r_{uu}[n] = - \sum_{k=1}^L a_k r_{uu}[n-k] \quad \text{for } n > 0.$$

For instance, knowing the first  $L$  samples of the auto-correlation sequence  $r_{uu}[0] \dots r_{uu}[L-1]$  is sufficient to calculate  $r_{uu}[n] \forall n \in \mathbb{Z}$  by recursion (when the AR coefficients  $a_k$ ,  $k = 1 \dots L$  are known). Considering the symmetry of  $r_{uu}[n]$  and evaluating the difference equation for  $n = 1 \dots L$  yields the Yule-Walker equations (see Problem 4.2) that allow the computation of the AR coefficients from the first  $L+1$  samples of the auto-correlation sequence  $r_{uu}[0] \dots r_{uu}[L]$ . For  $n = 0$ , the following equation is obtained

$$r_{uu}[0] + \sum_{k=1}^L a_k r_{uu}[k] = \sigma_v^2,$$

which shows the relation between the variances  $\sigma_v^2$  and  $\sigma_u^2$ . Using this equation, the *noise gain* of the AR process-generator filter can be calculated as

$$\frac{\sigma_u^2}{\sigma_v^2} = \frac{1}{1 + \sum_{k=1}^L a_k \frac{r_{uu}[k]}{r_{uu}[0]}}.$$

🏠 **Problem B.1.** Assume the process generator difference equation is given as

$$u[n] = v[n] + au[n - 1]$$

where  $v[n]$  is white noise with variance  $\sigma_v^2 = r_{vv}[0]$  and  $|a| < 1$ . We know that for  $k > 0$ ,

$$r_{uu}[k] = ar_{uu}[k - 1] = a^k r_{uu}[0].$$

To fully specify the autocorrelation function  $r_{uu}$  we therefore only need to determine  $r_{uu}[0] = \sigma_u^2$ . To this end, observe that the impulse response of above system is given as  $h[n] = (-a)^n u[n]$ . To a white noise input, the variance of the output can be computed using the noise gain of the system, i.e.,

$$r_{uu}[0] = \sigma_u^2 = \sigma_v^2 \sum_{n=-\infty}^{\infty} |h[n]|^2 = \sigma_v^2 \frac{1}{1 - a^2}.$$

Thus, and with the symmetry of  $r_{uu}$ ,

$$r_{uu}[k] = \sigma_v^2 \frac{a^{|k|}}{1 - a^2}.$$

## References

- [1] Simon Haykin: “Adaptive Filter Theory,” Fourth Edition, Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.
- [2] George Moschytz and Markus Hofbauer: “Adaptive Filter,” Springer-Verlag, Berlin Heidelberg, 2000.
- [3] Gernot Kubin: “Joint Recursive Optimality—A Non-Probabilistic Approach to Joint Recursive Optimality—A Non-Probabilistic Approach to,” *Journal Computers and Electrical Engineering*, Vol. 18, No. 3/4, pp. 277–289, 1992.
- [4] Bernard Widrow and Samuel D. Stearns: “Adaptive Signal Processing,” Prentice-Hall, Inc., Upper Saddle River, NJ, 1985.
- [5] Edward A. Lee and David G. Messerschmitt: “Digital Communication,” Third Edition, Kluwer Academic Publishers, 2004.
- [6] Steven M. Kay: “Fundamentals of Statistical Signal Processing—Estimation Theory,” Volume 1, Prentice-Hall, Inc., 1993.
- [7] Steven M. Kay: “Fundamentals of Statistical Signal Processing—Detection Theory,” Volume 2, Prentice-Hall, Inc., 1998.