# EE210A: Adaptation and Learning

Professor Ali H. Sayed

# MAIN RESULT FROM LAST LECTURE

**Theorem 5.1 (Linear estimator for linear models)** Let $\{\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}\}$ be zero-mean random variables that are related via the linear model $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{v}$, for some data matrix $H$ of compatible dimensions. Both $\boldsymbol{x}$ and $\boldsymbol{v}$ are assumed uncorrelated with invertible covariance matrices, $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^*$ and $R_x = \mathsf{E}\,\boldsymbol{x}\boldsymbol{x}^*$. The linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ can be evaluated by either expression:

$$\hat{\boldsymbol{x}} \;=\; R_x H^* \left[R_v + H R_x H^*\right]^{-1} \boldsymbol{y} \;=\; \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}$$

and the resulting minimum mean-square error matrix is

$$\text{m.m.s.e.} \;=\; \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1}$$

# CONSTRAINED LINEAR ESTIMATION

Sections in order: 6.1-6.3

# A MATRIX RESULT

**Lemma B.4 (Invertible product)** Let $A$ be $N \times n$, with $N \geq n$. Then

$$A \text{ has full rank} \iff A^*A \text{ is positive-definite}$$

That is, every tall full rank matrix is such that the square matrix $A^*A$ is invertible (actually, positive-definite).

**Proof:** Let us first show that $A$ has full rank only if $A^*A$ is invertible. Thus assume $A$ has full rank but that $A^*A$ is not invertible. This means that there exists a nonzero vector $p$ such that $A^*Ap = 0$, which implies $p^*A^*Ap = 0$ or, equivalently, $\|Ap\|^2 = 0$. That is, $Ap = 0$. This shows that the nullspace of $A$ is nontrivial so that $A$ cannot have full rank, which is a contradiction. Therefore a full rank $A$ implies an invertible matrix $A^*A$.

Conversely, assume $A^*A$ is invertible and let us show that $A$ has to have full rank. Assume not. Then there exists a nonzero vector $p$ such that $Ap = 0$. It follows that $A^*Ap = 0$, which contradicts the invertibility of $A^*A$. This is because $A^*Ap = 0$ implies that $p$ is an eigenvector of $A^*A$ corresponding to the zero eigenvalue. Hence, the determinant of $A^*A$ is necessarily zero.

Finally, let us show that $A^*A$ is positive-definite. For this purpose, take any nonzero vector $x$ and consider the product $x^*A^*Ax$, which evaluates to $\|Ax\|^2$. Then, the product $x^*A^*Ax$ is necessarily positive; it cannot be zero since the nullspace of $A$, in view of $A$ being full rank, contains only the zero vector.

$\diamond$

# A MATRIX RESULT

In fact, when $A$ has full rank, not only $A^*A$ is positive-definite, but any product of the form $A^*BA$ for any Hermitian positive-definite matrix $B$:

$$\boxed{A: \; N \times n, \; N \geq n, \; \text{full-rank} \quad \Longleftrightarrow \quad A^*BA > 0} \tag{B.1}$$

To see this, recall from App. B.1 that every Hermitian matrix $B$ admits an eigen-decomposition of the form

$$B = U\Lambda U^* \tag{B.2}$$

where $\Lambda$ is diagonal with the eigenvalues of $B$, and $U$ is a unitary matrix with the orthonormal eigenvectors of $B$. Define the matrices

$$\Lambda^{1/2} \;\stackrel{\Delta}{=}\; \text{diag}\left\{\sqrt{\lambda_1}, \; \sqrt{\lambda_2}, \; \ldots, \; \sqrt{\lambda_n}\right\}, \quad \bar{A} \;\stackrel{\Delta}{=}\; \Lambda^{1/2}U^*A$$

Then $A^*BA = \bar{A}^*\bar{A}$. Now the matrix $\bar{A}$ has full rank if, and only if, $A$ has full rank and, in view of the result of the previous lemma, the full rank property of $\bar{A}$ is equivalent to the positive-definiteness of $\bar{A}^*\bar{A}$, as desired.

# MOTIVATION

In Sec. 5.1 we studied the problem of estimating a random variable $\boldsymbol{x}$ from a noisy observation $\boldsymbol{y}$ that is related to $\boldsymbol{x}$ via the linear model

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{v} \tag{6.1}$$

where $H$ is a known data matrix and $\boldsymbol{v}$ is some disturbance, with $\boldsymbol{x}$ and $\boldsymbol{v}$ satisfying

$$\mathsf{E}\,\boldsymbol{x} = 0, \quad \mathsf{E}\,\boldsymbol{v} = 0, \quad \mathsf{E}\,\boldsymbol{x}\boldsymbol{x}^* = R_x, \quad \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^* = R_v, \quad \mathsf{E}\,\boldsymbol{x}\boldsymbol{v}^* = 0 \tag{6.2}$$

The linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ was found to be given by either expression

$$\hat{\boldsymbol{x}} = R_x H^* \left[R_v + H R_x H^*\right]^{-1} \boldsymbol{y} \;=\; \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y} \tag{6.3}$$

with the right-most expression valid whenever $R_x > 0$ and $R_v > 0$.

# EXAMPLE

In Sec. 5.5, we applied these results to a simple, yet revealing example. Given $N$ measurements $\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(N-1)\}$ of a random variable $\boldsymbol{x}$ with variance $\sigma_x^2$,

$$\boldsymbol{y}(i) = \boldsymbol{x} + \boldsymbol{v}(i), \qquad i = 0, 1, \ldots, N-1$$

i.e., given

$$\begin{bmatrix} \boldsymbol{y}(0) \\ \boldsymbol{y}(1) \\ \vdots \\ \boldsymbol{y}(N-1) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \boldsymbol{v}(0) \\ \boldsymbol{v}(1) \\ \vdots \\ \boldsymbol{v}(N-1) \end{bmatrix} \tag{6.4}$$

we estimated $\boldsymbol{x}$ from the $\{\boldsymbol{y}(i)\}$ and found that

$$\hat{\boldsymbol{x}} = \frac{1}{N + 1/\mathsf{SNR}} \sum_{i=0}^{N-1} \boldsymbol{y}(i) \tag{6.5}$$

where $\mathsf{SNR} = \sigma_x^2/\sigma_v^2$. In (6.4), the variable $\boldsymbol{x}$ is assumed to have been initially selected at random and then $N$ noisy measurements of this same value are made — see Fig. 2.9. The observations are subsequently used to estimate $\boldsymbol{x}$ according to (6.5).

# UNKOWN CONSTANT

But what if we consider a different model for $\boldsymbol{x}$, whereby it is assumed to be a *constant* of unknown value, say $x$, rather than a random quantity? How will the expression for $\hat{\boldsymbol{x}}$ change? The purpose of this chapter is to study such estimators. Specifically, we shall now consider linear models of the form

$$\boldsymbol{y} = Hx + \boldsymbol{v} \tag{6.6}$$

where, compared with (6.1), we are replacing the boldface letter $\boldsymbol{x}$ by the normal letter $x$ (remember that we reserve the boldface notation to random variables throughout this book).

## 6.1 MINIMUM-VARIANCE UNBIASED ESTIMATION

Thus consider a zero-mean random noise variable $\boldsymbol{v}$ with a positive-definite covariance matrix $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^* > 0$, and let $\boldsymbol{y}$ be a noisy measurement of $Hx$,

$$\boxed{\boldsymbol{y} = Hx + \boldsymbol{v}} \tag{6.7}$$

where $x$ is the unknown constant vector that we wish to estimate. The dimensions of the data matrix $H$ are denoted by $N \times n$ and it is further assumed that $N \geq n$,

$$\boxed{H : N \times n, \quad N \geq n} \tag{6.8}$$

That is, $H$ is assumed to be a tall matrix so that the number of available entries in $\boldsymbol{y}$ is at least as many as the number of unknown entries in $x$. Note that we use the capital letter $N$ for the larger dimension and the small letter $n$ for the smaller dimension.

We also assume that the matrix $H$ in (6.7) has *full rank*, i.e., that all its columns are linearly independent and, hence,

$$\boxed{\text{rank}(H) \;=\; n} \tag{6.9}$$

This condition guarantees that the matrix product $H^*H$ is invertible (in fact, positive-definite — recall Lemma B.4). It also guarantees that the product $H^*R_v^{-1}H$ is positive-definite — see expression (B.1). For the benefit of the reader, Sec. B.2 reviews several basic concepts regarding range spaces, nullspaces, and ranks of matrices.

# MVUE PROBLEM FORMULATION

We are interested in determining a linear estimator for $x$ of the form $\hat{\boldsymbol{x}} = K\boldsymbol{y}$, for some $n \times N$ matrix $K$. The choice of $K$ should satisfy two conditions:

1. **Unbiasedness**. First, the estimator $\hat{\boldsymbol{x}}$ should be unbiased. That is, the choice of $K$ should guarantee $\mathsf{E}\,\hat{\boldsymbol{x}} = x$, which is the same as $K\mathsf{E}\,\boldsymbol{y} = x$. But from (6.7) we have $\mathsf{E}\,\boldsymbol{y} = Hx$ so that $K$ should satisfy $KHx = x$, no matter what the value of $x$ is. This condition means that $K$ should satisfy

$$\boxed{KH = \mathrm{I}} \tag{6.10}$$

   Note that $KH$ is $n \times n$ and is therefore a square matrix.

2. **Optimality**. Second, the choice of $K$ should minimize the covariance matrix of the estimation error, $\tilde{\boldsymbol{x}} = x - \hat{\boldsymbol{x}}$. Using the condition $KH = \mathrm{I}$, we find that

$$\hat{\boldsymbol{x}} = K\boldsymbol{y} = K(Hx + \boldsymbol{v}) = KHx + K\boldsymbol{v} = x + K\boldsymbol{v}$$

   so that $\tilde{\boldsymbol{x}} = -K\boldsymbol{v}$. This means that the error covariance matrix, as a function of $K$, is given by

$$\boxed{\mathsf{E}\,\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^* = \mathsf{E}\,(K\boldsymbol{v}\boldsymbol{v}^*K^*) \;=\; KR_vK^*} \tag{6.11}$$

# MVUE OR BLUE PROBLEM

Combining (6.10) and (6.11), we conclude that the desired $K$ is found by solving the following constrained optimization problem:

$$\boxed{\min_{K} \quad K R_v K^* \qquad \text{subject to} \quad KH = \mathrm{I}} \tag{6.12}$$

The estimator $\hat{\boldsymbol{x}} = K_o \boldsymbol{y}$ that results from the solution of (6.12) is known as the *minimum-variance-unbiased estimator*, or m.v.u.e. for short. It is also sometimes called the best linear unbiased estimator (BLUE).

# INTERPRETATION

Let $\mathcal{J}(K)$ denote the cost function that appears in (6.12), i.e.,

$$\mathcal{J}(K) \stackrel{\Delta}{=} KR_vK^*$$

Then problem (6.12) means the following. We seek a matrix $K_o$ satisfying $K_oH = \mathrm{I}$ such that

$$\mathcal{J}(K) - \mathcal{J}(K_o) \geq 0 \quad \text{for all} \quad K \quad \text{satisfying} \quad KH = \mathrm{I}$$

There are several ways of determining $K_o$. We choose to use the already known solution of the linear estimation problem (cf. Sec. 5.1) in order to guess what the solution $K_o$ for (6.12) should be. Once this is done, we shall then provide an independent verification of the result.

# INTUITION

Thus recall, as mentioned in the introduction of this chapter, that for two zero-mean random variables $\{\boldsymbol{x}, \boldsymbol{y}\}$ that are related as in (6.1), the linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ is (cf. the second expression in (6.3)):

$$\hat{\boldsymbol{x}} = (R_x^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y}$$

Now assume that the covariance matrix of $\boldsymbol{x}$ has the particular form $R_x = \alpha \mathrm{I}$, with a sufficiently large positive scalar $\alpha$ (i.e., $\alpha \to \infty$). That is, assume that the variance of each of the entries of $\boldsymbol{x}$ is infinitely large. In this way, $\boldsymbol{x}$ can be "interpreted" as playing the role of some unknown constant vector, $x$. Then the above expression for $\hat{\boldsymbol{x}}$ reduces to

$$\hat{\boldsymbol{x}} = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y}$$

This conclusion suggests that the choice $K_o = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1}$ solves the problem of estimating the unknown vector $x$ from model (6.7). We shall now establish this result more directly; the result is known as the Gauss-Markov theorem.

# GAUSS-MARKOV THEOREM

Carl F. Gauss (1777-1855)

Andrey A. Markov (1856-1922)

**Theorem 6.1 (Gauss-Markov Theorem)** Consider the linear model $\boldsymbol{y} = H x + \boldsymbol{v}$, where $\boldsymbol{v}$ is a zero-mean random variable with positive-definite covariance matrix $R_v$, and $x$ is an unknown constant vector. Assume further that $H$ is a full-rank $N \times n$ matrix with $N \geq n$. Then the minimum-variance-unbiased linear estimator of $x$ given $\boldsymbol{y}$ is $\hat{\boldsymbol{x}} = K_o \boldsymbol{y}$, where

$$K_o = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1}$$

Moreover, the resulting cost is m.m.s.e. $= (H^* R_v^{-1} H)^{-1}$.

# ARGUMENT

**Proof:** For any matrix $K$ that satisfies $KH = \mathrm{I}$, it is easy to verify that

$$\mathcal{J}(K) = KR_vK^* \;=\; (K - K_o)R_v(K - K_o)^* \;+\; K_oR_vK_o^* \tag{6.13}$$

This is because

$$KR_vK_o^* \;=\; KR_v[R_v^{-1}H(H^*R_v^{-1}H)^{-1}] \;=\; KH(H^*R_v^{-1}H)^{-1} = (H^*R_v^{-1}H)^{-1}$$

Likewise, $K_oR_vK_o^* \;=\; (H^*R_v^{-1}H)^{-1}$. Relation (6.13) expresses the cost $\mathcal{J}(K)$ as the sum of two nonnegative-definite terms: one is independent of $K$ and is equal to $K_oR_vK_o^*$, while the other is dependent on $K$. It is then clear, since $R_v > 0$, that the cost is minimized by choosing $K = K_o$, and that the resulting minimum cost is $K_oR_vK_o^* = (H^*R_v^{-1}H)^{-1}$. Note further that the matrix $K_o$ in the statement of the theorem satisfies the constraint $K_oH = \mathrm{I}$.

$\diamond$

**Remark 6.1 (Constrained optimization)** Sometimes in applications (see Secs. 6.4 and 6.5), optimization problems of the form (6.12) arise without being explicitly related to a minimum-variance-unbiased estimation problem (as in the statement of Thm. 6.1). For this reason, we also state the following conclusion here for later reference. The solution of a generic constrained optimization problem of the form

$$\min_{K} \quad K R_v K^* \qquad \text{subject to } KH = \mathrm{I} \text{ and } R_v > 0 \tag{6.14}$$

is given by

$$K_o = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1}$$

with the resulting minimum cost equal to

$$\text{minimum cost} = (H^* R_v^{-1} H)^{-1} \tag{6.15}$$

◇

## 6.2 EXAMPLE: MEAN ESTIMATION

Let us reconsider the example of Sec. 5.5, where we assumed that we are given $N$ measurements

$$\boldsymbol{y}(i) = \boldsymbol{x} + \boldsymbol{v}(i), \qquad i = 0, 1, \ldots, N-1$$

of the same random variable $\boldsymbol{x}$ with variance $\sigma_x^2$. The noise sequence $\boldsymbol{v}(i)$ was further assumed to be white with zero mean and variance $\sigma_v^2$. The linear least-mean-squares estimator (l.l.m.s.e.) of $\boldsymbol{x}$ given the $\{\boldsymbol{y}(i)\}$ was found to be (cf. (6.5)):

$$\boxed{\hat{x}_{\text{llmse}} = \frac{1}{N + 1/\text{SNR}} \sum_{i=0}^{N-1} \boldsymbol{y}(i)}$$

where $\text{SNR} = \sigma_x^2/\sigma_v^2$.

# EXAMPLE

Now assume instead that we model $x$ as an unknown constant, rather than a random variable, say

$$\boldsymbol{y}(i) = x + \boldsymbol{v}(i), \qquad i = 0, 1, \ldots, N-1 \tag{6.16}$$

In this case, the value of $x$ can be regarded as the mean value of each $\boldsymbol{y}(i)$. If we collect the measurements and the noises into vector form,

$$\boldsymbol{y} \triangleq \text{col}\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(N-1)\}, \qquad \boldsymbol{v} \triangleq \text{col}\{\boldsymbol{v}(0), \boldsymbol{v}(1), \ldots, \boldsymbol{v}(N-1)\}$$

and define the data vector $h = \text{col}\{1, 1, \ldots, 1\}$. Then $\boldsymbol{y} = hx + \boldsymbol{v}$, with $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^* = \sigma_v^2 \mathrm{I}$. Invoking the result of Thm. 6.1 with $H = h$, we conclude that the optimal linear estimator, or the m.v.u.e., of $x$ is $\hat{\boldsymbol{x}}_{\text{mvue}} = (H^*H)^{-1}H^*\boldsymbol{y}$, i.e.,

$$\boxed{\hat{\boldsymbol{x}}_{\text{mvue}} = \frac{1}{N}\sum_{i=0}^{N-1} \boldsymbol{y}(i)} \tag{6.17}$$

## 6.3 APPLICATION: CHANNEL AND NOISE ESTIMATION

We reconsider the channel estimation problem of Sec. 5.2, except that now the channel tap vector is modeled as an unknown *constant* vector, $c$, rather than a random vector, $\boldsymbol{c}$, as shown in Fig. 6.1.

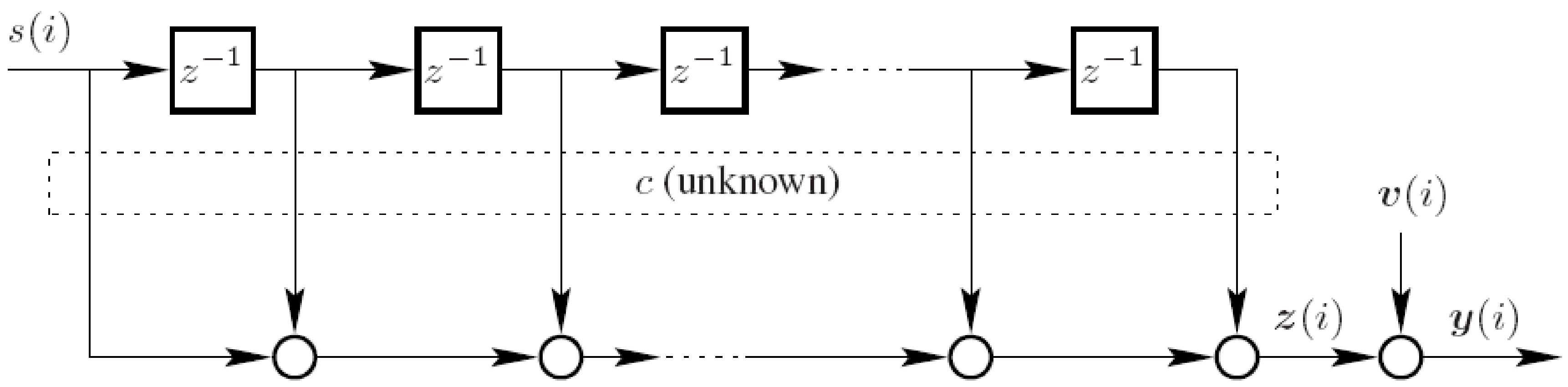


**FIGURE 6.1** Channel and noise estimation.

# COLLECTING DATA

By repeating the construction of Sec. 5.2 we again obtain (cf. (5.7)):

$$\underbrace{\begin{bmatrix} \boldsymbol{y}(0) \\ \boldsymbol{y}(1) \\ \boldsymbol{y}(2) \\ \boldsymbol{y}(3) \\ \boldsymbol{y}(4) \\ \boldsymbol{y}(5) \\ \boldsymbol{y}(6) \end{bmatrix}}_{\boldsymbol{y}:(N+1)\times 1} = \underbrace{\begin{bmatrix} s(0) & & \\ s(1) & s(0) & \\ s(2) & s(1) & s(0) \\ s(3) & s(2) & s(1) \\ s(4) & s(3) & s(2) \\ s(5) & s(4) & s(3) \\ s(6) & s(5) & s(4) \end{bmatrix}}_{H:(N+1)\times M} c + \underbrace{\begin{bmatrix} \boldsymbol{v}(0) \\ \boldsymbol{v}(1) \\ \boldsymbol{v}(2) \\ \boldsymbol{v}(3) \\ \boldsymbol{v}(4) \\ \boldsymbol{v}(5) \\ \boldsymbol{v}(6) \end{bmatrix}}_{\boldsymbol{v}:(N+1)\times 1}$$

where we are defining the quantities $\{\boldsymbol{y}, H, \boldsymbol{v}\}$ and where $H$ is $(N+1) \times M$. Using the result of Thm. 6.1, we find that the optimal estimator of $c$ is now given by

$$\boxed{\hat{c}_{\text{mvue}} = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y}} \qquad (6.18)$$

where $R_v$ is the covariance matrix of $\boldsymbol{v}$.

# COMPARING ESTIMATORS

$$\boxed{\hat{c}_{\text{mvue}} = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y}} \tag{6.18}$$

where $R_v$ is the covariance matrix of $\boldsymbol{v}$. This result is different from the linear least-mean-squares estimator (l.l.m.s.e.) found in Sec. 5.2 (see (5.8)), namely,

$$\hat{c}_{\text{llmse}} = \left[R_c^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}$$

which requires knowledge of the covariance matrix $R_c = \mathsf{E}\,\boldsymbol{c}\boldsymbol{c}^*$ when $\boldsymbol{c}$ is modeled as a random variable. The estimator (6.18) requires knowledge of only $\{H, R_v, \boldsymbol{y}\}$. Actually, if the noise sequence $\{\boldsymbol{v}(i)\}$ is modeled as white with variance $\sigma_v^2$, then $R_v = \sigma_v^2 \mathrm{I}$ and $R_v$ would end up disappearing from the expression for $\hat{c}_{\text{mvue}}$. Specifically, (6.18) would become

$$\boxed{\hat{c}_{\text{mvue}} = (H^* H)^{-1} H^* \boldsymbol{y}} \tag{6.19}$$

It is worth remarking that expression (6.19) has the form of a least-squares solution;

# NOISE POWER ESTIMATION

Note from (6.19) that we do not need to know $\sigma_v^2$; the estimator is now only dependent on the available data (namely, the measurements $\{\boldsymbol{y}(i)\}$ and the data matrix $H$). If desired, we can estimate $\sigma_v^2$ itself as follows. Since

$$\boldsymbol{v}(i) = \boldsymbol{y}(i) - \left[\begin{array}{cccc} s(i) & s(i-1) & \ldots & s(i-M+1) \end{array}\right] c$$

an estimator for $\sigma_v^2$ would be

$$\widehat{\boldsymbol{\sigma}_v^2} \;=\; \frac{1}{N+1}\sum_{i=0}^{N} \left| \boldsymbol{y}_i - \left[\begin{array}{cccc} s(i) & s(i-1) & \ldots & s(i-M+1) \end{array}\right] \hat{\boldsymbol{c}}_{\sf mvue} \right|^2$$

i.e.,

$$\boxed{\widehat{\boldsymbol{\sigma}_v^2} \;=\; \frac{1}{N+1}\, \|\boldsymbol{y} - H\hat{\boldsymbol{c}}_{\sf mvue}\|^2} \tag{6.20}$$

# DECISION-FEEDBACK EQUALIZATION

consider an FIR channel with a known column tap vector $c$ of length $M$ (i.e., with $M$ taps), say with transfer function

$$C(z) = c(0) + c(1)z^{-1} + \ldots + c(M-1)z^{-(M-1)}$$

Data symbols $\{\boldsymbol{s}(\cdot)\}$ are transmitted through the channel and the output sequence is measured in the presence of additive noise, $\boldsymbol{v}(i)$. The signals $\{\boldsymbol{v}(\cdot), \boldsymbol{s}(\cdot)\}$ are assumed uncorrelated. Due to the channel memory, each measurement $\boldsymbol{y}(i)$ contains contributions not only from $\boldsymbol{s}(i)$ but also from prior symbols, since

$$\boldsymbol{y}(i) = c(0)\boldsymbol{s}(i) + \underbrace{\sum_{k=1}^{M-1} c(k)\boldsymbol{s}(i-k)}_{\text{ISI}} + \boldsymbol{v}(i)$$

The second term on the right-hand side describes the *inter-symbol-interference* (ISI); it refers to the interference that is caused by prior symbols. The purpose of an equalizer is to combat ISI and to recover $\boldsymbol{s}(i)$ from measurements of the output sequence.

# CHANNEL AND EQUALIZER MODELS

As was discussed in Sec. 5.4, in order to achieve this task, a linear equalizer employs current and prior measurements $\{\boldsymbol{y}(i-k)\}$, say for $k = 0, 1, \ldots, L-1$. This is because prior measurements contain information that is correlated with the ISI term in $\boldsymbol{y}(i)$ and, therefore, they can help in estimating the interference term and removing its effect. Of course, if possible, it would be preferable to use the prior symbols $\{\boldsymbol{s}(i-1), \boldsymbol{s}(i-2), \ldots\}$ themselves in order to cancel their effect from $\boldsymbol{y}(i)$ rather than rely on the prior measurements $\{\boldsymbol{y}(i-1), \boldsymbol{y}(i-2), \ldots\}$.

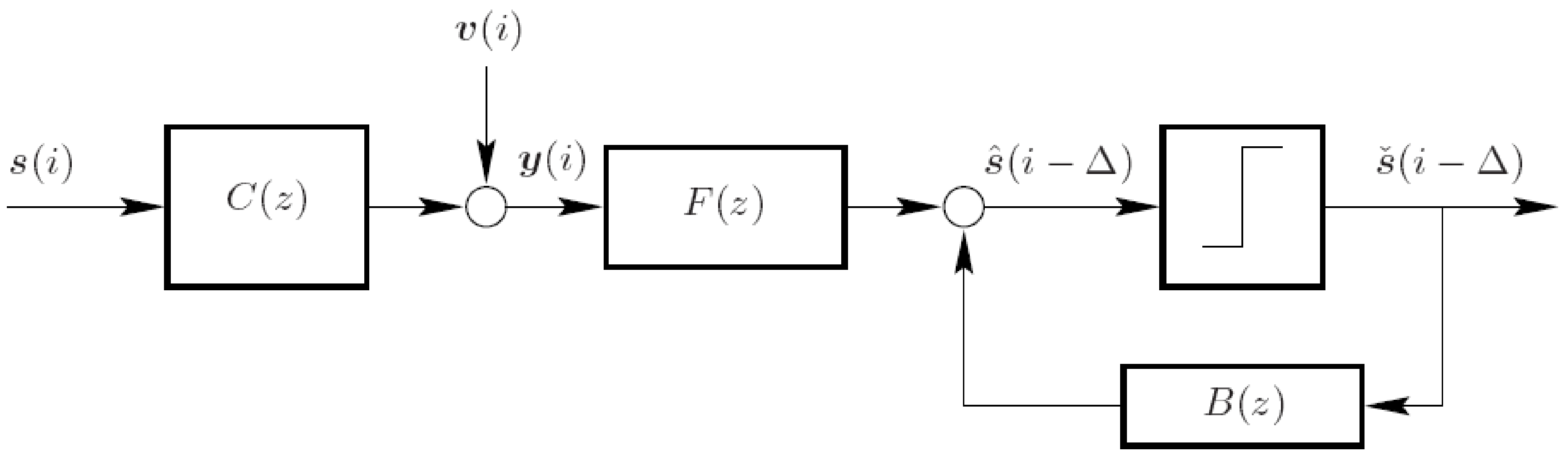


**FIGURE 6.2** A decision-feedback equalizer. It consists of a feedforward filter, a feedback filter, and a decision device.

# DFEs

Decision-feedback equalizers (DFE) attempt to implement this strategy and are therefore better suited for channels with pronounced ISI. In addition to using an FIR filter in the feedforward path, as in linear equalization, a DFE employs a *feedback* filter in order to feed back previous decisions and use them to reduce ISI. The DFE structure is shown in Fig. 6.2 for estimating a delayed version of $\boldsymbol{s}(i)$, with the transfer functions of the feedforward and feedback filters denoted by $\{F(z), B(z)\}$, respectively. It is seen from the figure that the input to the feedback filter comes from the output of the decision device, denoted by $\check{\boldsymbol{s}}(i-\Delta)$. The purpose of this device is to map the estimator $\hat{\boldsymbol{s}}(i-\Delta)$, which is obtained by combining the outputs of the feedforward and feedback filters, to the closest point in the symbol constellation. Now in linear equalization, the feedforward filter reduces ISI by attempting to force the combined system $C(z)F(z)$ to be close to

$$C(z)F(z) \approx z^{-\Delta}$$

In general, this objective is difficult to attain, especially for channels with pronounced ISI, and $C(z)F(z)$ will have a nontrivial impulse response sequence (we say that $C(z)F(z)$ will have trailing inter-symbol interference). The purpose of the feedback filter in a DFE implementation is to use prior decisions in order to cancel this trailing ISI.

# EQUALIZER DESIGN

Assume the feedforward filter has $L$ taps and denote its transfer function by

$$F(z) = f(0) + f(1)z^{-1} + \ldots + f(L-1)z^{-(L-1)}$$

with coefficients $\{f(i)\}$. Likewise, assume the feedback filter has $Q$ taps with a transfer function of the form

$$B(z) = -b(1)z^{-1} - b(2)z^{-2} - \ldots - b(Q)z^{-Q}$$

with coefficients denoted by $\{-b(i)\}$ for convenience. Note that this filter is strictly causal in that it does not have a direct path from its input to its output (i.e., $b(0) = 0$). This is because *previous* decisions are being fed back through $B(z)$.

# EQUALIZER DESIGN

The criterion for designing the equalizer coefficients $\{f(i), b(i)\}$ is, as usual, to minimize the variance of the error signal, $\tilde{\boldsymbol{s}}(i-\Delta) = \boldsymbol{s}(i-\Delta) - \hat{\boldsymbol{s}}(i-\Delta)$. In so doing, the designer expects that $\hat{\boldsymbol{s}}(i-\Delta)$ will be sufficiently close to $\boldsymbol{s}(i-\Delta)$ so that the decision device would be able to map $\hat{\boldsymbol{s}}(i-\Delta)$ to the correct symbol in the signal constellation. Therefore, the $\{f(i), b(i)\}$ will be determined by solving

$$\boxed{\min_{\left\{\begin{array}{c} f(0), f(1), \ldots, f(L-1) \\ b(1), b(2), \ldots, b(Q) \end{array}\right\}} \mathsf{E}\, |\tilde{\boldsymbol{s}}(i-\Delta)|^2} \qquad (6.21)$$

The presence of the decision device makes (6.21) a *nonlinear* optimization problem. This is because $\hat{\boldsymbol{s}}(i-\Delta)$ will be a nonlinear function of the measured data $\{\boldsymbol{y}(i)\}$. In order to facilitate the design of $\{F(z), B(z)\}$, it is customary to assume that

$$\boxed{\text{The decisions } \{\check{\boldsymbol{s}}(i-\Delta)\} \text{ are correct and equal to } \{\boldsymbol{s}(i-\Delta)\}} \qquad (6.22)$$

That is, we assume that the decision device gives correct decisions.

# EQUALIZER DESIGN

To solve (6.21) we first examine the dependence of the error variance on the unknown coefficients $\{f(i), b(i)\}$. From Fig. 6.2 we have

$$\begin{aligned}\hat{\boldsymbol{s}}(i-\Delta) \;=\;& [\, f(0)\boldsymbol{y}(i) + f(1)\boldsymbol{y}(i-1) \ldots + f(L-1)\boldsymbol{y}(i-L+1)\,] \\ & -[\, b(1)\boldsymbol{s}(i-\Delta-1) + b(2)\boldsymbol{s}(i-\Delta-2) + \ldots + b(Q)\boldsymbol{s}(i-\Delta-Q)\,]\end{aligned}$$

where we used assumption (6.22) to replace $\check{\boldsymbol{s}}(j)$ by $\boldsymbol{s}(j)$. We can rewrite this expression more compactly in vector form as follows. We collect the coefficients of $F(z)$ into a row vector:

$$f^* \triangleq [\; f(0) \quad f(1) \quad \ldots \quad f(L-1) \;]$$

and the coefficients of $-B(z)$ into another row vector with a leading entry that is equal to one,

$$b^* \triangleq [\; 1 \quad b(1) \quad b(2) \quad \ldots \quad b(Q) \;]$$

# EQUALIZER DESIGN

We also define the following column vectors of observations and data symbols:

$$\boldsymbol{y}_i \triangleq \underbrace{\begin{bmatrix} \boldsymbol{y}(i) \\ \boldsymbol{y}(i-1) \\ \boldsymbol{y}(i-2) \\ \vdots \\ \boldsymbol{y}(i-L+1) \end{bmatrix}}_{L\times 1}, \qquad \boldsymbol{s}_\Delta = \underbrace{\begin{bmatrix} \boldsymbol{s}(i-\Delta) \\ \boldsymbol{s}(i-\Delta-1) \\ \boldsymbol{s}(i-\Delta-2) \\ \vdots \\ \boldsymbol{s}(i-\Delta-Q) \end{bmatrix}}_{(Q+1)\times 1} \qquad (6.23)$$

and denote their covariances and cross-covariances by

$$\boxed{R \triangleq \mathsf{E}\begin{bmatrix} \boldsymbol{s}_\Delta \\ \boldsymbol{y}_i \end{bmatrix}\begin{bmatrix} \boldsymbol{s}_\Delta \\ \boldsymbol{y}_i \end{bmatrix}^* \triangleq \begin{bmatrix} R_s & R_{sy} \\ R_{ys} & R_y \end{bmatrix}}$$

where $R_s$ is $(Q+1)\times(Q+1)$ and $R_{sy}$ is $(Q+1)\times L$.

# ASSUMPTIONS

where $R_s$ is $(Q+1) \times (Q+1)$ and $R_{sy}$ is $(Q+1) \times L$. We assume that the processes $\{\boldsymbol{s}(\cdot), \boldsymbol{y}(\cdot)\}$ are jointly wide-sense stationary so that the quantities $\{R_{sy}, R_y, R_s\}$ are independent of $i$. We further assume that the covariance matrix $R$ is positive-definite and, hence, invertible. The positive-definiteness of $R$ guarantees that both $R_y$ and the Schur complement of $R$ with respect to $R_y$ are positive-definite matrices as well (see Sec. B.3), i.e.,

$$\boxed{R_y > 0 \quad \text{and} \quad R_\delta \triangleq R_s - R_{sy} R_y^{-1} R_{ys} > 0} \tag{6.24}$$

where we are denoting the Schur complement by $R_\delta$. Hence, $\{R_y, R_\delta\}$ are also invertible.

# PROBLEM FORMULATION

With these definitions, the error signal $\tilde{\boldsymbol{s}}(i-\Delta)$ can be written as

$$\tilde{\boldsymbol{s}}(i-\Delta) \;=\; \boldsymbol{s}(i-\Delta) - \hat{\boldsymbol{s}}(i-\Delta) \;=\; b^*\boldsymbol{s}_\Delta - f^*\boldsymbol{y}_i$$

so that the optimization problem (6.21) becomes

$$\min_{f,b} \; \mathsf{E}\,|b^*\boldsymbol{s}_\Delta - f^*\boldsymbol{y}_i|^2 \tag{6.25}$$

We shall denote the optimal vector solutions by $f^*_{\rm opt}$ and $b^*_{\rm opt}$. Rather than minimize the variance of $b^*\boldsymbol{s}_\Delta - f^*\boldsymbol{y}_i$ simultaneously over $\{f, b\}$, we shall minimize it over one vector at a time. Thus assume that we *fix* the vector $b$ and let us minimize the error variance over $f$. To do so, we introduce the scalar $\boldsymbol{\alpha} = b^*\boldsymbol{s}_\Delta$ so that the error signal becomes $\tilde{\boldsymbol{s}}(i-\Delta) = \boldsymbol{\alpha} - f^*\boldsymbol{y}_i$. In this way, $\tilde{\boldsymbol{s}}(i-\Delta)$ can be interpreted as the error that results from estimating $\boldsymbol{\alpha}$ from $\boldsymbol{y}_i$ through the choice of $f$. In other words, we are reduced to solving

$$\boxed{\min_{f} \quad \mathsf{E}\,|\boldsymbol{\alpha} - f^*\boldsymbol{y}_i|^2 \qquad \text{where} \qquad \boldsymbol{\alpha} = b^*\boldsymbol{s}_\Delta}$$

which is a standard linear least-mean-squares estimation problem.

# FORWARD FILTER

which is a standard linear least-mean-squares estimation problem. From Thm. 3.1, we know that the optimal choice for $f$ is

$$f^*_{\text{opt}} = R_{\alpha y} R_y^{-1} \; = \; b^* R_{sy} R_y^{-1} \tag{6.26}$$

where we used the fact that

$$R_{\alpha y} \; \stackrel{\Delta}{=} \; \mathsf{E}\, \alpha \boldsymbol{y}_i^* = \mathsf{E}\, b^* \boldsymbol{s}_\Delta \boldsymbol{y}_i^* = b^* R_{sy}$$

The resulting minimum mean-square error is, again from Thm. 3.1,

$$\begin{aligned} \text{m.m.s.e.} \;\; &\stackrel{\Delta}{=} \;\; \mathsf{E}\,|\alpha - f^*_{\text{opt}} \boldsymbol{y}_i|^2 \\ &= \;\; R_\alpha - R_{\alpha y} R_y^{-1} R_{y\alpha} \\ &= \;\; b^* R_s b - b^* R_{sy} R_y^{-1} R_{ys} b \\ &= \;\; b^* [R_s - R_{sy} R_y^{-1} R_{ys}] b \\ &= \;\; b^* R_\delta b \end{aligned} \tag{6.27}$$

# BACKWARD FILTER

Substituting this expression into (6.25), we find that we now need to solve

$$\min_b \ b^* R_\delta b \tag{6.28}$$

But recall that the leading entry of $b$ is unity, so that (6.28) is actually a constrained problem of the form

$$\boxed{\min_b \quad b^* R_\delta b \qquad \text{subject to} \ \ b^* e_0 = 1}$$

where $e_0$ is the first basis vector, of dimension $(Q+1) \times 1$,

$$e_0 \ \stackrel{\Delta}{=}\ \text{col}\{1, 0, 0 \ldots, 0\}$$

Using the result stated in Remark 6.1, we find that the optimal choice of $b$ is

$$b^*_{\text{opt}} = \frac{e_0^{\mathsf{T}} R_\delta^{-1}}{e_0^{\mathsf{T}} R_\delta^{-1} e_0}$$

# BACKWARD FILTER

The term that appears in the denominator is the $(0,0)$ entry of $R_\delta^{-1}$, while the term in the numerator is the first row of $R_\delta^{-1}$. This means that the optimal vector $b^*_{\rm opt}$ is obtained by normalizing the first row of $R_\delta^{-1}$ to have a unit leading entry. Substituting the above expression for $b^*_{\rm opt}$ into $b^* R_\delta b$ we find that the resulting m.m.s.e. of the original optimization problem (6.21) is

$$\boxed{\text{m.m.s.e.} \;=\; \frac{1}{e_0^{\sf T} R_\delta^{-1} e_0}} \qquad (6.29)$$

In summary, under assumption (6.22) that the decisions $\{\check{\boldsymbol{s}}(i-\Delta)\}$ are correct, the optimal coefficients $\{f(i), b(i)\}$ of the DFE can be found as follows:

$$\boxed{b^*_{\rm opt} \;=\; \frac{e_0^{\sf T} R_\delta^{-1}}{e_0^{\sf T} R_\delta^{-1} e_0} \qquad \text{and} \qquad f^*_{\rm opt} = b^*_{\rm opt} R_{sy} R_y^{-1}} \qquad (6.30)$$

The entries of $\{f^*_{\rm opt}, b^*_{\rm opt}\}$ provide the desired tap coefficients $\{b(i), f(i)\}$.

# CHANNEL MODEL

$$\boxed{\text{m.m.s.e.} \;=\; \frac{1}{e_0^{\mathsf{T}} R_\delta^{-1} e_0}} \qquad (6.29)$$

$$\boxed{b^*_{\text{opt}} \;=\; \frac{e_0^{\mathsf{T}} R_\delta^{-1}}{e_0^{\mathsf{T}} R_\delta^{-1} e_0} \qquad \text{and} \qquad f^*_{\text{opt}} \;=\; b^*_{\text{opt}} R_{sy} R_y^{-1}} \qquad (6.30)$$

The expressions (6.29)–(6.30) for $\{b^*_{\text{opt}}, f^*_{\text{opt}}, \text{m.m.s.e.}\}$ are in terms of the covariance and cross-covariance matrices $\{R_s, R_{sy}, R_y\}$, which can be evaluated from the channel model $C(z)$ and from the given statistical information about $\{\boldsymbol{s}(\cdot), \boldsymbol{v}(\cdot)\}$. To do so, we proceed as in Sec. 5.4.

We first express the observation vector $\boldsymbol{y}_i$ in terms of the transmitted data. Assume for the sake of illustration that $L = 5$ (i.e., a feedforward filter with 5 taps) and $M = 4$ (a channel with four taps). Then we can write

# CHANNEL MODEL

We first express the observation vector $\boldsymbol{y}_i$ in terms of the transmitted data. Assume for the sake of illustration that $L = 5$ (i.e., a feedforward filter with 5 taps) and $M = 4$ (a channel with four taps). Then we can write

$$\underbrace{\begin{bmatrix} \boldsymbol{y}(i) \\ \boldsymbol{y}(i-1) \\ \boldsymbol{y}(i-2) \\ \boldsymbol{y}(i-3) \\ \boldsymbol{y}(i-4) \end{bmatrix}}_{\boldsymbol{y}_i : L\times 1} = \underbrace{\begin{bmatrix} c(0) & c(1) & c(2) & c(3) & & & & \\ & c(0) & c(1) & c(2) & c(3) & & & \\ & & c(0) & c(1) & c(2) & c(3) & & \\ & & & c(0) & c(1) & c(2) & c(3) & \\ & & & & c(0) & c(1) & c(2) & c(3) \end{bmatrix}}_{H : L\times(L+M-1)} \underbrace{\begin{bmatrix} \boldsymbol{s}(i) \\ \boldsymbol{s}(i-1) \\ \boldsymbol{s}(i-2) \\ \boldsymbol{s}(i-3) \\ \boldsymbol{s}(i-4) \\ \boldsymbol{s}(i-5) \\ \boldsymbol{s}(i-6) \\ \boldsymbol{s}(i-7) \end{bmatrix}}_{\underline{\boldsymbol{s}}_i : (L+M)-1\times 1}$$

$$+ \underbrace{\begin{bmatrix} \boldsymbol{v}(i) \\ \boldsymbol{v}(i-1) \\ \boldsymbol{v}(i-2) \\ \boldsymbol{v}(i-3) \\ \boldsymbol{v}(i-4) \end{bmatrix}}_{\boldsymbol{v}_i : L\times 1}$$

# CHANNEL MODEL

That is,

$$\boldsymbol{y}_i = H\underline{\boldsymbol{s}}_i + \boldsymbol{v}_i$$

where, for general $\{L, M\}$,

$$\underline{\boldsymbol{s}}_i \triangleq \begin{bmatrix} \boldsymbol{s}(i) \\ \boldsymbol{s}(i-1) \\ \boldsymbol{s}(i-2) \\ \vdots \\ \boldsymbol{s}(i-L-M+2) \end{bmatrix}, \qquad \boldsymbol{v}_i \triangleq \begin{bmatrix} \boldsymbol{v}(i) \\ \boldsymbol{v}(i-1) \\ \boldsymbol{v}(i-2) \\ \vdots \\ \boldsymbol{v}(i-L+1) \end{bmatrix} \tag{6.31}$$

and $H$ is the $L \times (L+M-1)$ channel matrix. We therefore find that there is a linear relation between the vectors $\{\boldsymbol{y}_i, \underline{\boldsymbol{s}}_i\}$ and this relation can be used to evaluate $R_y$ as

$$R_y = \mathsf{E}\,(H\underline{\boldsymbol{s}}_i + \boldsymbol{v}_i)(H\underline{\boldsymbol{s}}_i + \boldsymbol{v}_i)^* = HR_{\underline{s}}H^* + R_v$$

where

$$R_{\underline{s}} \triangleq \mathsf{E}\,\underline{\boldsymbol{s}}_i\underline{\boldsymbol{s}}_i^* \quad ((L+M-1)\times(L+M-1)) \quad \text{and} \quad R_v \triangleq \mathsf{E}\,\boldsymbol{v}_i\boldsymbol{v}_i^* \quad (L \times L)$$

# CHANNEL MODEL

Likewise,

$$R_{sy} = \mathsf{E}\, s_\Delta (H \underline{s}_i + v_i)^* \;=\; (\mathsf{E}\, s_\Delta \underline{s}_i^*)\, H^*$$

since $\{v(\cdot), s(\cdot)\}$ are uncorrelated. We still need to evaluate $\mathsf{E}\, s_\Delta \underline{s}_i^*$, where $\{s_\Delta, \underline{s}_i\}$ are defined by (6.23) and (6.31) in terms of the transmitted symbols. Of course, the value of $\mathsf{E}\, s_\Delta \underline{s}_i^*$ depends on the assumed correlation between the transmitted symbols.

It is common that $\Delta$ be chosen such that all the entries of $s_\Delta$ fall within the entries of $\underline{s}_i$. This condition requires the channel and filter lengths, as well as the delay $\Delta$, to satisfy

$$\boxed{\Delta + Q \leq L + M - 2} \tag{6.32}$$

With this condition, if the $\{s(\cdot)\}$ are assumed to be independent and identically distributed with variance $\sigma_s^2$, then it can be seen that

$$\mathsf{E}\, s_\Delta \underline{s}_i^* \;=\; \begin{bmatrix} 0 & \ldots & 0 & \sigma_s^2 \mathrm{I}_{Q+1} & 0 & \ldots & 0 \end{bmatrix} \quad ((Q+1) \times (L+M-1))$$

with $\Delta$ leading zero columns, followed by a $(Q+1) \times (Q+1)$ identity matrix scaled by $\sigma_s^2$, followed (or not) by zero columns.

# CHANNEL MODEL

We can express the above $\mathsf{E}\, s_\Delta \underline{s}_i^*$ more compactly as

$$\mathsf{E}\, s_\Delta \underline{s}_i^* \;=\; \left[\begin{array}{ccc} 0_{(Q+1)\times\Delta} & \sigma_s^2 \mathrm{I}_{Q+1} & 0 \end{array}\right]$$

Likewise,

$$R_s = \sigma_s^2 \mathrm{I}_{Q+1} \qquad \text{and} \qquad R_{\underline{s}} = \sigma_s^2 \mathrm{I}_{L+M-1}$$

We continue with the assumption of i.i.d. symbols $\{s(\cdot)\}$ for simplicity of presentation. But it should be noted that the derivation applies even for correlated (but stationary) data. In a similar vein, we assume that the noise sequence $\{v(\cdot)\}$ is white with variance $\sigma_v^2$ so that $R_v = \sigma_v^2 \mathrm{I}$. Again, the development applies even for correlated (but stationary) noise. We thus find that

$$\boxed{R_y = \sigma_s^2 H H^* + \sigma_v^2 \mathrm{I}_L \qquad \text{and} \qquad R_{sy} = \left[\begin{array}{ccc} 0_{(Q+1)\times\Delta} & \sigma_s^2 \mathrm{I}_{Q+1} & 0 \end{array}\right] H^*} \tag{6.33}$$

and

$$R_\delta \;=\; \sigma_s^2 \mathrm{I} - R_{sy}(\sigma_s^2 H H^* + \sigma_v^2 \mathrm{I}_L)^{-1} R_{ys}$$

# CHANNEL MODEL

This latter expression can be rewritten, by virtue of the matrix inversion lemma (5.4), as

$$\boxed{R_\delta = \Phi \left( \frac{1}{\sigma_s^2} \mathrm{I} + \frac{1}{\sigma_v^2} H^* H \right)^{-1} \Phi^*} \tag{6.34}$$

where

$$\Phi = \left[ \begin{array}{ccc} 0_{(Q+1)\times\Delta} & \mathrm{I}_{Q+1} & 0 \end{array} \right]$$

Expressions (6.33) and (6.34) can now be used with (6.29)–(6.30) to determine the optimal equalizer coefficients and the resulting m.m.s.e.

## Example 6.1 (Numerical illustration)

Let us reconsider Ex. 4.1 and design a DFE equalizer rather than a linear equalizer for the channel $C(z) = 1 + 0.5z^{-1}$, for which $M = 2$. We select a feedforward filter with 3 taps (i.e., $L = 3$) and a feedback filter with one tap (i.e., $Q = 1$). We also select $\Delta = 1$. The resulting structure is shown in Fig. 6.3.

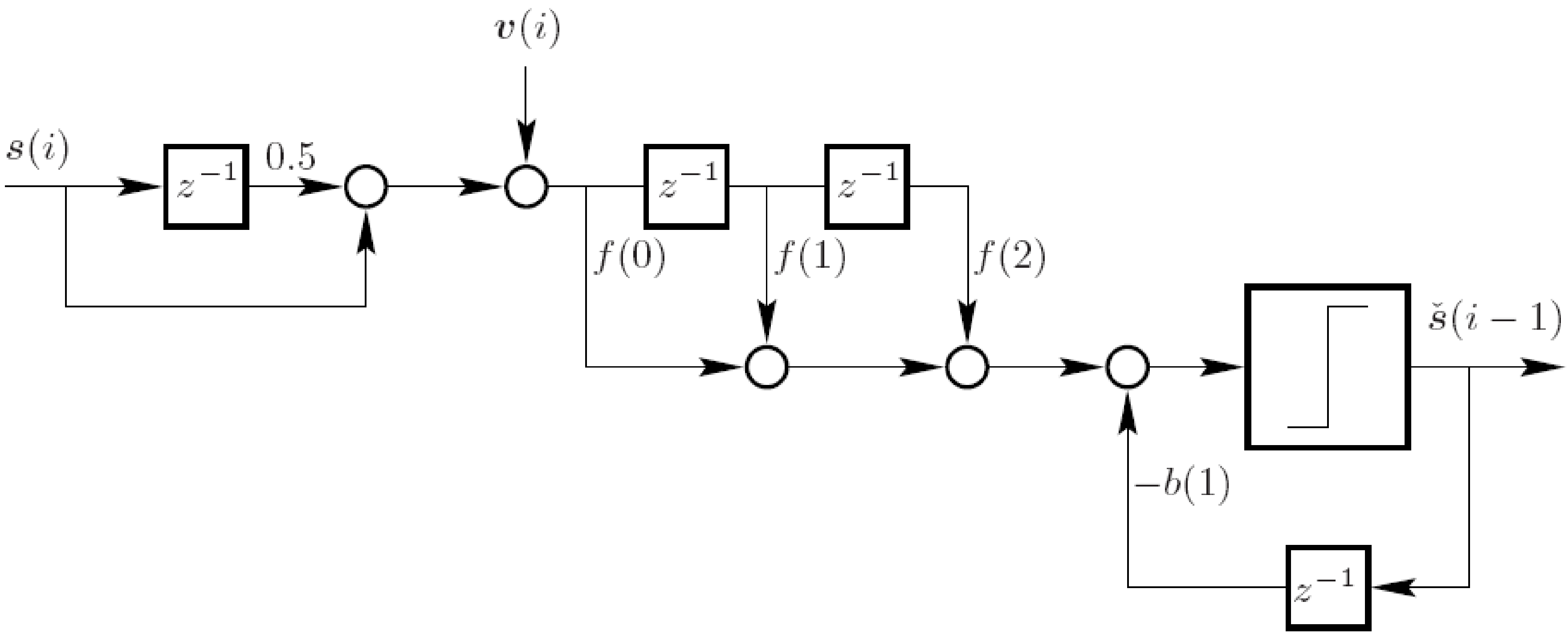


**FIGURE 6.3** A DFE structure for the channel $1 + 0.5z^{-1}$.

# EXAMPLE

For this example, $\sigma_s^2 = 1$, $\sigma_v^2 = 1$, and

$$H = \begin{bmatrix} 1 & 0.5 & & \\ & 1 & 0.5 & \\ & & 1 & 0.5 \end{bmatrix}$$

so that from (6.33) and (6.34)

$$R_y = \begin{bmatrix} 9/4 & 1/2 & 0 \\ 1/2 & 9/4 & 1/2 \\ 0 & 1/2 & 9/4 \end{bmatrix}, \quad R_{sy} = \begin{bmatrix} 0.5 & 1 & 0 \\ 0 & 0.5 & 1 \end{bmatrix}, \quad R_\delta = \begin{bmatrix} 0.4992 & -0.1218 \\ -0.1218 & 0.5175 \end{bmatrix}$$

Using (6.30) we obtain

$$b_{\text{opt}}^* = \begin{bmatrix} 1.0000 & 0.2354 \end{bmatrix}, \qquad f_{\text{opt}}^* = \begin{bmatrix} 0.1176 & 0.4706 & 0.0000 \end{bmatrix}$$

and the resulting m.m.s.e. is

$$\text{m.m.s.e.} \;=\; b_{\text{opt}} R_\delta b_{\text{opt}}^* \;=\; 0.4705$$

# EXAMPLE

That is, $B_{\rm opt}(z) = -0.2354z^{-1}$ and $F_{\rm opt}(z) = 0.1176 + 0.4706z^{-1}$. Had we selected instead $\Delta = 0$, as we did in Ex. 4.1, then the only quantity that changes is the cross-covariance $R_{sy}$, which becomes

$$R_{sy} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \end{bmatrix} \quad \text{so that} \quad R_\delta = \begin{bmatrix} 0.5312 & -0.1248 \\ -0.1248 & 0.4992 \end{bmatrix}$$

and, therefore,

$$b_{\rm opt}^* = \begin{bmatrix} 1 & 0.2500 \end{bmatrix}, \qquad f_{\rm opt}^* = \begin{bmatrix} 0.5000 & 0.0000 & 0.0000 \end{bmatrix}$$

That is, $B_{\rm opt}(z) = -0.25z^{-1}$ and $F_{\rm opt}(z) = 0.5$. The resulting m.m.s.e. in this case is

$$\text{m.m.s.e.} = b_{\rm opt}^* R_\delta b_{\rm opt} = 0.5000$$

We see that for this example with $\Delta = 0$, the DFE results in a smaller mean-square error than the linear equalizer designed in Ex. 4.1, which resulted in m.m.s.e. $= 0.5312$. A more noticeable difference in performance between decision-feedback equalizers and linear equalizers can be observed for channels with more pronounced inter-symbol interference. The performance of DFEs is examined in greater detail in a computer project at the end of this part.

◇

**Project II.3 (Decision feedback equalization)** In this project we study the performance of decision feedback equalization for the channel

$$C(z) = 0.5 + 1.2z^{-1} + 1.5z^{-2} - z^{-3}$$

The symbols $\{s(i)\}$ that are transmitted through $C(z)$ are i.i.d. and chosen from a QPSK constellation, i.e., each $s(i)$ is selected randomly from the set

$$\left\{ \pm\frac{\sqrt{2}}{2} \pm j\frac{\sqrt{2}}{2} \right\}, \qquad j = \sqrt{-1}$$

The noise sequence $\{v(i)\}$ is assumed i.i.d. and complex-valued; its real and imaginary parts are uncorrelated Gaussian random variables with variances $0.039$ each, so that $\sigma_v^2 = 0.078$ and the SNR ratio at the input of the equalizer is approximately $18$ dB. We start with $L = 13$ and $Q = 2$.

(a) Plot the impulse response, as well as the magnitude of the frequency response, of the channel. Is the channel minimum phase?

(b) Generate $N = 2000$ QPSK data points $\{s(i)\}$ and transmit them through the channel. Plot the scatter diagrams of the transmitted sequence $\{\mathrm{s}(i)\}$ and the received sequence $\{y(i)\}$.

# COMPUTER PROJECT

(c) Compute the optimal filters $\{f^*_{\text{opt}}, b^*_{\text{opt}}\}$ for values of $\Delta$ in the interval $0 \leq \Delta \leq 15$ and generate the sequences $\{\hat{s}(i-\Delta)\}$ and $\{\check{s}(i-\Delta)\}$ at the input and output of the decision device, which is defined by the equation

$$\text{dec}[x] = \frac{\sqrt{2}}{2}\{\text{sign}[\,\text{Re}(x)\,] + j\text{sign}[\,\text{Im}(x)\,]\}$$

Plot the number of erroneous decisions as a function of $\Delta$. For $\Delta = 5$, plot the scatter diagrams of the received sequence $\{\boldsymbol{y}(i)\}$ and of the input to the decision device, $\{\hat{s}(i-5)\}$.

(d) For each $\Delta$, compute the theoretical m.m.s.e. by using m.m.s.e. $= 1/e_1^{\mathsf{T}} R_\delta^{-1} e_1$, and plot its value as a function of $\Delta$. Using the actual data, estimate the m.m.s.e. by computing

$$\frac{1}{N-\Delta}\sum_{i=\Delta+1}^{N}|s(i)-\hat{s}(i)|^2$$

Compare the resulting values with the theoretical values. Can you explain why there is a bad fit between theory and practice for smaller values of $\Delta$? Plot also for $\Delta = 5$, the following sequences on three separate subplots:

(i) The channel impulse response sequence.

(ii) The impulse response sequence of the cascade combination of the channel and the feed-forward filter.

(iii) The impulse response sequence of the feedback filter delayed by the value of $\Delta$.

You will observe that the sequence in part (ii) has an almost unit-magnitude sample at time instant 5, followed by two nonzero samples that correspond to what we call *post inter-symbol interference*. This interference should be cancelled by the feedback filter. Any residual ISI prior to the peak sample at time instant 5 will not be equalized. Compare the coefficients of the feedback filter in (iii) to the values of the post ISI.

(e) Fix $\Delta = 5$ and let us now examine the effect of changing the length of the feedforward filter. Generate a plot showing the number of erroneous decisions as a function of $L$, for $L$ varying between 1 and 15. Keep $Q$ fixed at $Q = 2$. Which value of $L$ results in the smallest number of errors?

(f) Now fix $\Delta = 5$ and $L = 6$, and let us vary $Q$. Generate a plot showing the number of erroneous decisions as a function of $Q$, for $Q$ varying between 1 and 6. Which value of $Q$ results in the smallest number of errors?

(g) Now fix $\Delta = 5$, $L = 6$, and $Q = 1$. That is, the feedforward filter has 6 taps and the feedback filter has a single tap. In all derivations and simulations so far we assumed $\sigma_v^2 = 0.078$. Now let $\sigma_v^2$ vary between $0.12$ and $0.78$, say in increments of $0.001$. Write a program that generates a plot showing how the symbol error rate (SER) varies with SNR.

# COMPUTER PROJECT

(h) Let us now compare the performance of the DFE with that of a linear equalizer for the same channel. Recall that we studied linear equalizers in Computer Project II.1. Write a program that determines the optimal linear equalizer for $L$ varying between 1 and 10. The output of the equalizer is fed into the decision device. Generate a plot that shows the number of erroneous decisions as a function of $L$. Use $\sigma_v^2 = 0.078$ and $\Delta = 4$ for the linear equalizer. Fix $L = 4$ for the linear equalizer and plot the scatter diagrams of the received sequence $\{\boldsymbol{y}(i)\}$ and of the input to the decision device, $\{\hat{\boldsymbol{s}}(i-4)\}$. For this particular channel, do you see any advantage in using the DFE structure over the linear structure?

(i) Now assume the channel $C(z)$ and the noise variance $\sigma_v^2$ are not known beforehand but that we know the first 200 transmitted symbols $\{s(i)\}$, in addition to the entire received data record $\{\boldsymbol{y}(i)\}$. Use the initial 200 data $\{s(i), \boldsymbol{y}(i)\}$ to estimate $C(z)$ and $\sigma_v^2$, as explained in Sec. 6.3. Note that while the coefficients of the actual channel $C(z)$ are real-valued, the estimated coefficients will in general be complex-valued. You may use the complex-valued estimates, or you may keep only their real parts. If the estimates are good enough, their imaginary parts should be small compared to the real parts. Plot the impulse and frequency responses of the estimated channel and compare them with that of the actual channel. Repeat the design of the DFE equalizer by using the estimates of $C(z)$ and $\sigma_v^2$ instead. Use $\sigma_v^2 = 0.078$, $L = 6, Q = 1$, and $\Delta = 5$. Compare the number of errors in this case with the one obtained in part (g) using the exact channel model and the exact noise variance.

(j) Now repeat part (i) using a linear equalizer of length 4, followed by the nonlinear decision device. Compare the number of errors you get in this case with that obtained in part (h) and also with the DFE.

**Project II.3 (Decision feedback equalization)** The programs that solve this project are the following.

1. partA.m This program solves part (a) and generates a plot showing the impulse response sequence and the magnitude of the frequency response of the channel, $|C(e^{j\omega})|$ over $[0, \pi]$. Its output is shown in Fig. 21.

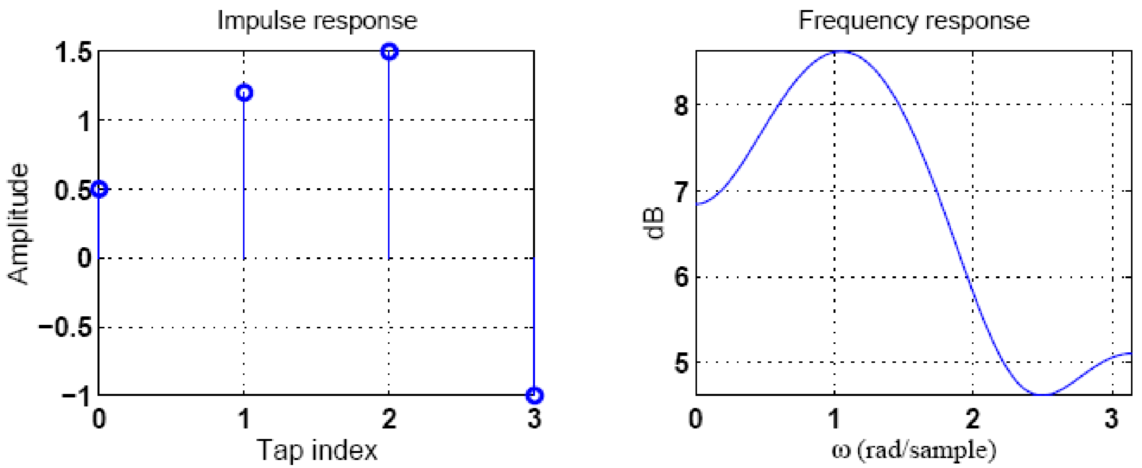


**Figure II.21.** Impulse and frequency response of the channel $C(z) = 0.5 + 1.2z^{-1} + 1.5z^{-2} - z^{-3}$.

2. partB.m This program solves part (b) and generates a plot showing a scatter diagram of the transmitted and received sequences $\{s(i), y(i)\}$. A typical output is shown in Fig. 22.

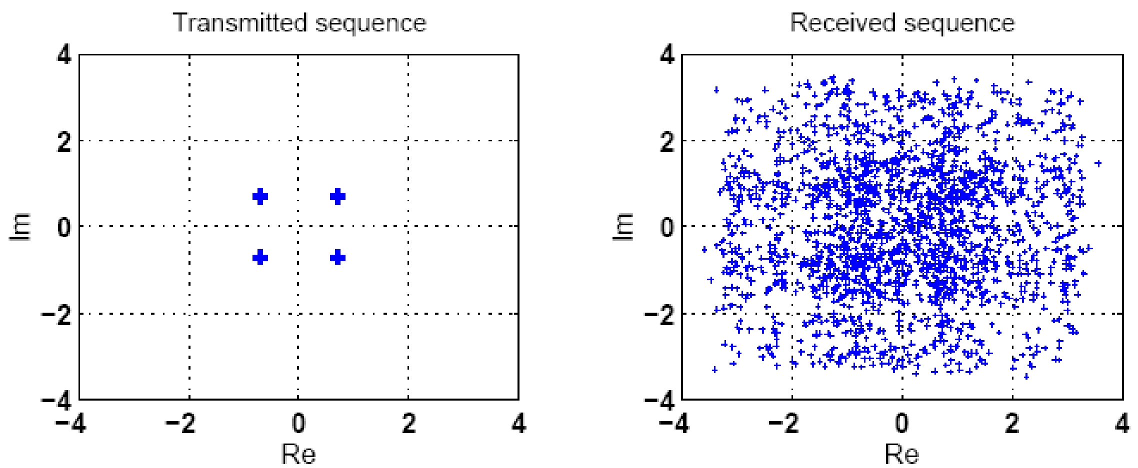


**Figure II.22.** Scatter diagrams of the transmitted (*left*) and received sequences (*right*), $\{s(i), y(i)\}$, for QPSK transmissions.

3. partC.m This program solves parts (c) and (d) and generates three plots. One plot shows scatter diagrams for the received sequence $\{y(i)\}$ and the sequence at the input of the decision device (for the case $\Delta = 5$) — see Fig. 23. A second plot shows the number of erroneous decisions as a function of $\Delta$, as well as the m.m.s.e. (both theoretical and measured) as a function of $\Delta$ — see Fig. 24. A third plot shows the impulse response sequences of the channel, the combination channel-feedforward filter, and the feedback filter delayed by $\Delta$ — see Fig. 25. The impulse response of the feedback filter has also been extended by some zeros in the plot in order to match its length with the convolution of the channel and the feedforward filter for ease of comparison. Observe how the taps of the feedback filter cancel the post ISI. The number of errors for this simulation was zero.

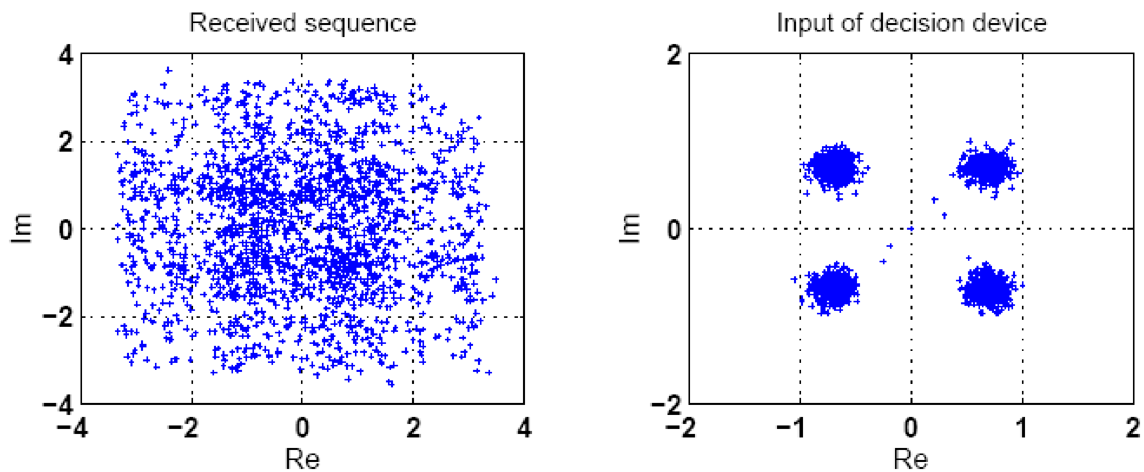


**Figure II.23.** Scatter diagrams of the received sequence (input of the equalizer) and the sequence at the input of the decision device for $\Delta = 5$ and QPSK transmissions.

# COMPUTER PROJECT

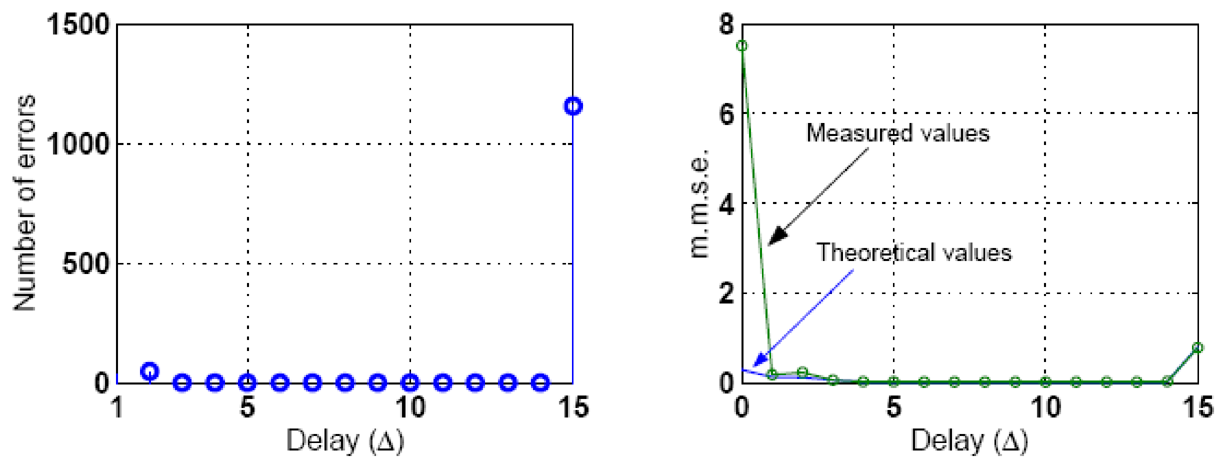


**Figure II.24.** The plot on the left shows the number of erroneous decisions as a function of $\Delta$, while the plot on the right shows the m.m.s.e. as a function of $\Delta$ as deduced from both theory and measurements.

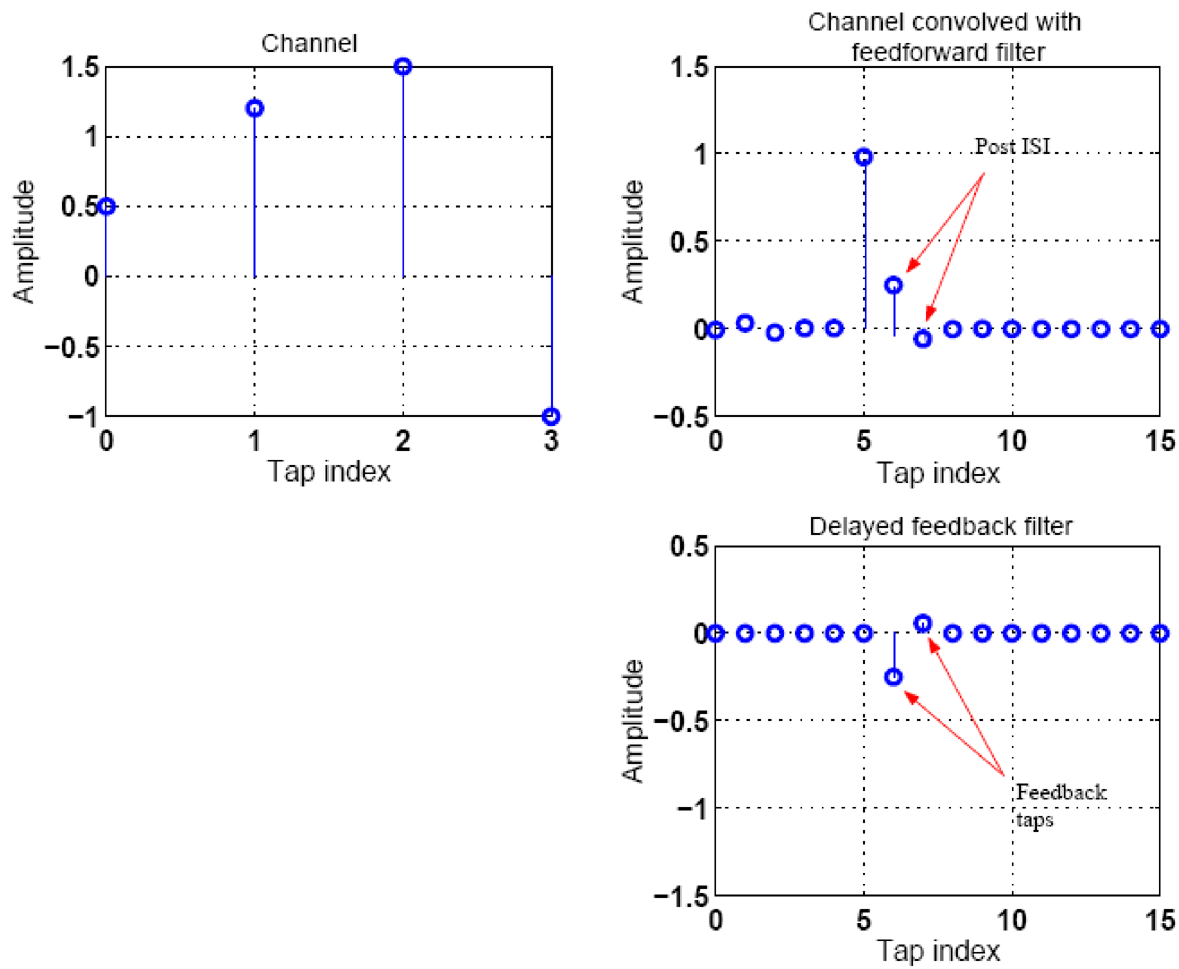


**Figure II.25.** The top left plot shows the impulse response sequence of the channel. The two plots on the right show the convolution of the channel and feedforward filter impulse responses (*top*) and the feedback filter response delayed by $\Delta$ (*bottom*).

4. partE.m This program generates a plot showing the number of erroneous decisions as a function of the length of the feedforward filter, $L$ — see the plot on the left in Fig. 26.
5. partF.m This program generates a plot showing the number of erroneous decisions as a function of the length of the feedback filter, $Q$ — see the plot on the right in Fig. 26.
6. partG.m This program solves part (g) and generates a plot showing the symbol error rate as a function of the SNR level at the input of the equalizer. Each point in the plot is measured using $2 \times 10^4$ transmissions. A typical output is shown in Fig. 27.

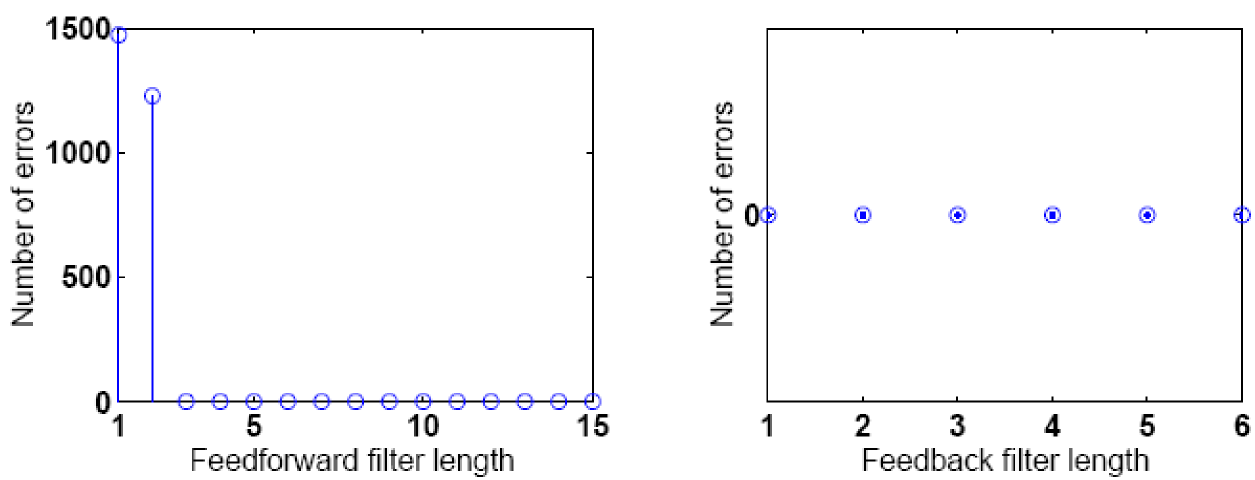


**Figure II.26.** Number of erroneous decisions as a function of the feedforward filter length using $\Delta = 5$ and $Q = 2$ (*left*), and as a function of the feedback filter length using $\Delta = 5$ and $L = 6$ (*right*).

# COMPUTER PROJECT

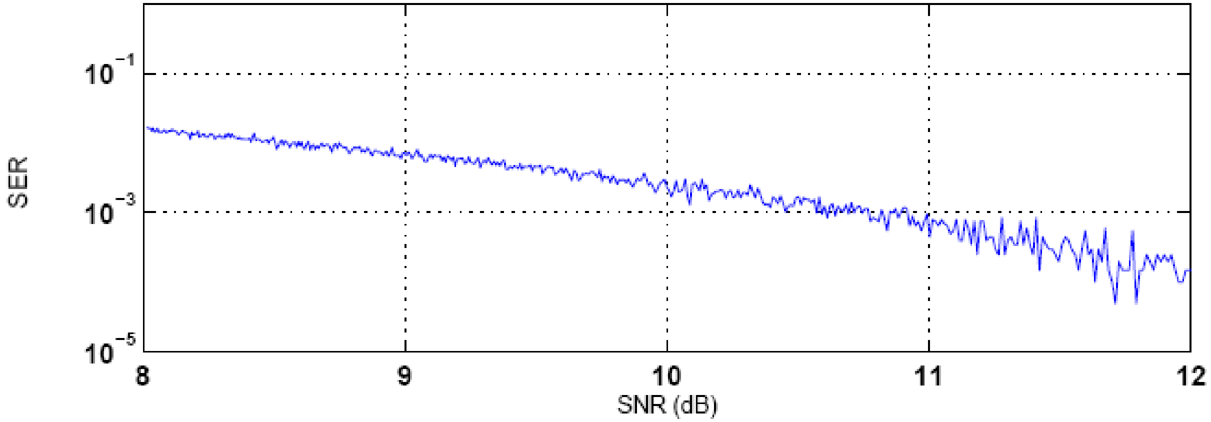


**Figure II.27.** The plot shows the symbol error rate as a function of the SNR level at the input of the equalizer. The simulation assumes $\Delta = 5$, $L = 6$ and $Q = 1$.

# COMPUTER PROJECT

7. partH.m This program solves part (h) and generates a plot that shows the scatter diagrams of the received sequence and of the sequence at the input to the decision device (for $L = 4$ and $\Delta = 4$). The plot also shows the number of erroneous decisions as a function of $L$, as well as the theoretical and estimated values of the m.m.s.e. for various $L$. A typical output is shown in Fig. 28. The number of errors in this simulation was zero.

8. partI.m This program solves part (i) and generates a plot that shows the impulse and frequency responses of both the actual channel and its estimate, as well as scatter diagrams of the received sequence $\{\boldsymbol{y}(i)\}$ and the sequence at the input of the decision device — see Fig. 29.

9. partJ.m This program solves part (j) and generates a plot that shows the impulse and frequency responses of the actual channel and its estimate, as well as scatter diagrams of the received sequence $\{\boldsymbol{y}(i)\}$ and the sequence at the input of the decision device — see Fig. 30.

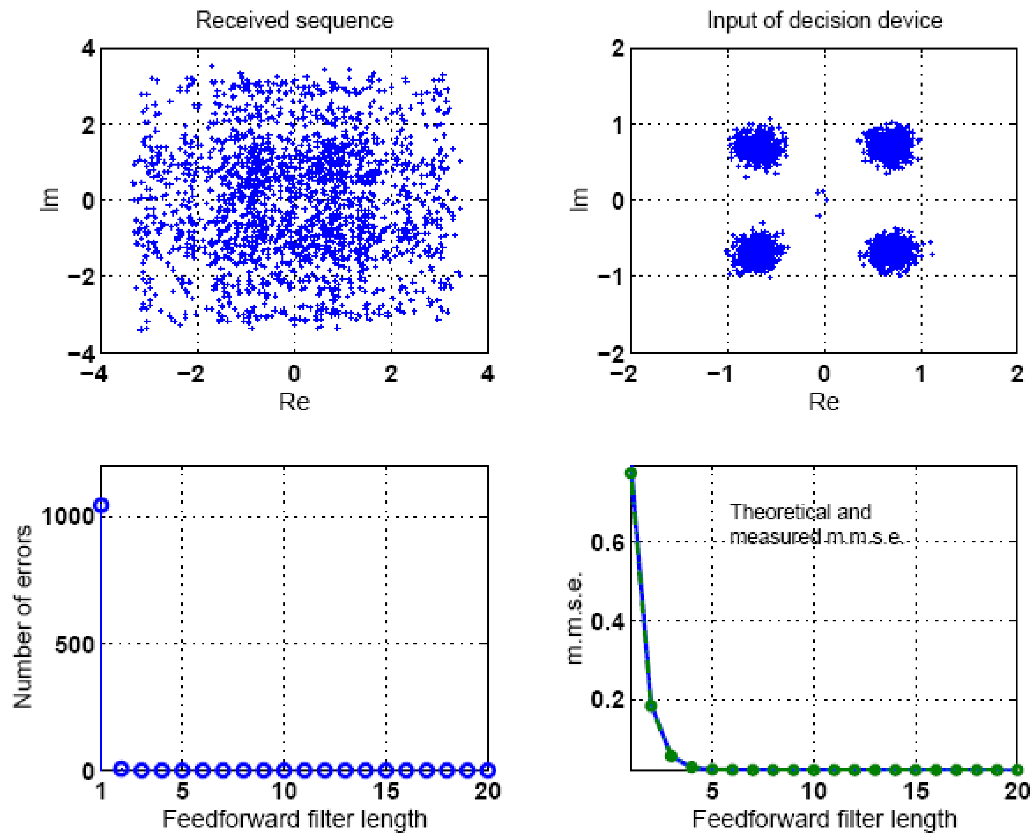


**Figure II.28.** The plots in the first row show the scatter diagrams of the received sequence and of the sequence at the input of the decision device for $L = 4$ and $\Delta = 4$, using a linear equalizer structure. The plots in the second row show the number of erroneous decisions as a function of the feedforward filter length ($L$), as well as the theoretical and estimated values of m.m.s.e. for various values of $L$.

# COMPUTER PROJECT

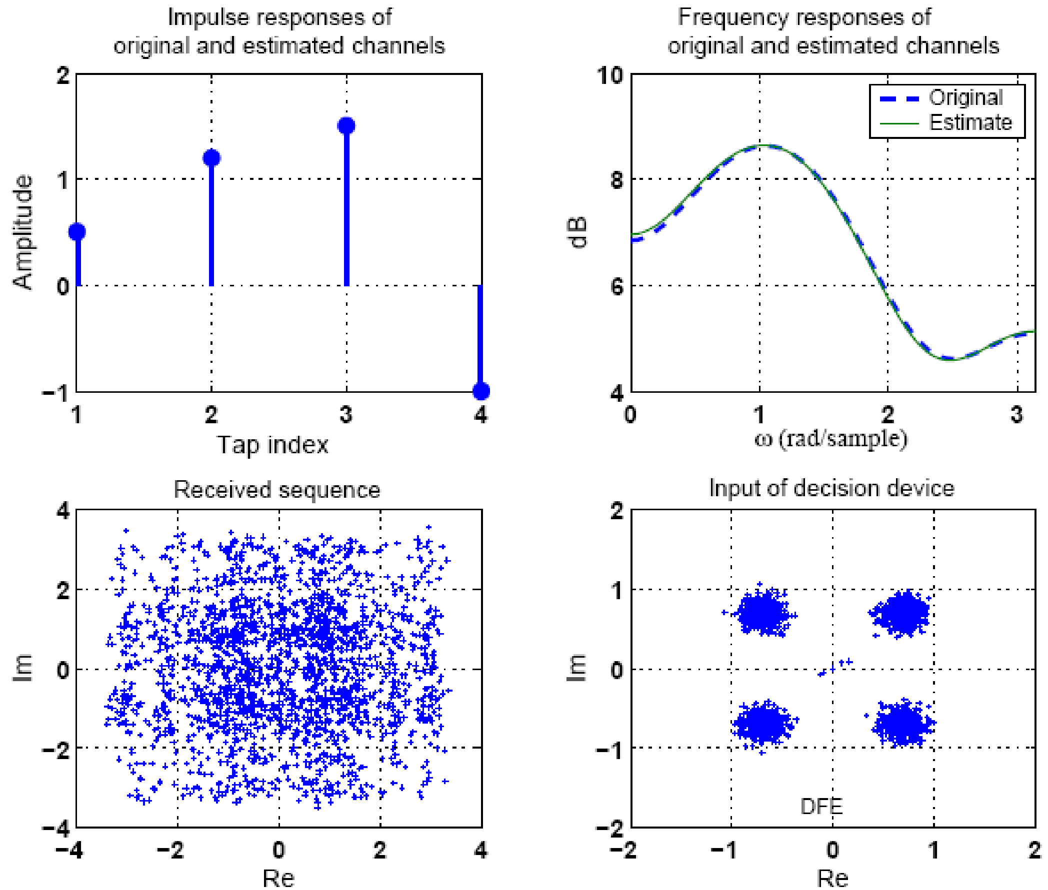


**Figure II.29.** The plots in the first row show the impulse and frequency responses of the channel and its estimate. The plots in the second row show the scatter diagrams of the received sequence and the sequence at the input of the decision device. This simulation pertains to a DFE implementation with $L = 6$, $Q = 1$, and $\Delta = 5$.

# COMPUTER PROJECT

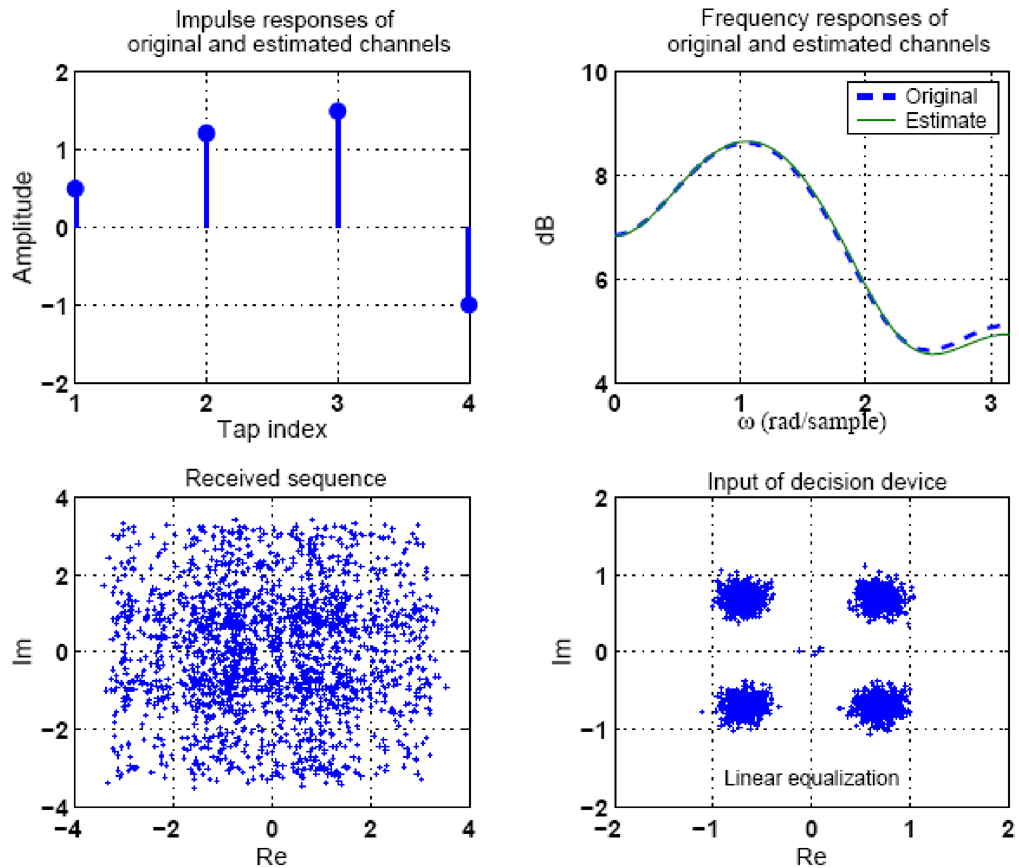


**Figure II.30.** The plots in the first row show the impulse and frequency responses of the channel and its estimate. The plots in the second row show the scatter diagrams of the received sequence and the sequence at the input of the decision device. This simulation pertains to a linear equalizer with $L = 4$ and $\Delta = 4$.