EE210A: Adaptation and Learning

Professor Ali H. Sayed

MAIN RESULT FROM LAST LECTURE

Theorem 4.2 (Properties of the linear estimator) Consider the same setting of Thm. 3.1. Then the normal equations $K_o R_y = R_{xy}$ that define the linear least-mean-squares estimator have the following properties:

1. They are always consistent, i.e., a solution K_o always exists.
2. The solution K_o is unique if, and only if, $R_y > 0$.
3. Infinitely many solutions K_o exist if, and only if, R_y is singular.

In case 3, regardless of which solution K_o is chosen, the values of the estimator, $\hat{x} = K_o \boldsymbol{y}$, and the m.m.s.e., $(R_x - K_o R_y K_o^*)$, remain invariant.

LECTURE #05

LINEAR MODELS

Sections in order: 5.1-5.5

LINEAR MODELS

We apply the linear estimation theory of the previous two chapters to the important special case of linear models, which arises often in applications. Specifically, we now assume that the zero-mean random vectors $\{\boldsymbol{x}, \boldsymbol{y}\}$ are related via a linear model of the form

$$\boxed{\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{v}} \tag{5.1}$$

for some $q \times p$ matrix H. Here $\boldsymbol{v}$ denotes a zero-mean random noise vector with known covariance matrix, $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^*$. The covariance matrix of $\boldsymbol{x}$ is also assumed to be known, say $\mathsf{E}\,\boldsymbol{x}\boldsymbol{x}^* = R_x$. Both $\{\boldsymbol{x}, \boldsymbol{v}\}$ are uncorrelated, i.e., $\mathsf{E}\,\boldsymbol{x}\boldsymbol{v}^* = 0$, and we further assume that $R_x > 0$ and $R_v > 0$.

5.1 ESTIMATION USING LINEAR RELATIONS

According to Thm. 3.1, when $R_y > 0$, the linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ is

$$\hat{\boldsymbol{x}} = R_{xy} R_y^{-1} \boldsymbol{y} \tag{5.2}$$

Because of (5.1), the covariances $\{R_{xy}, R_y\}$ can be determined in terms of the given matrices $\{H, R_x, R_v\}$. Indeed, the uncorrelatedness of $\{\boldsymbol{x}, \boldsymbol{v}\}$ gives

$$\begin{aligned} R_y &= \mathsf{E}\,\boldsymbol{y}\boldsymbol{y}^* = \mathsf{E}\,(H\boldsymbol{x} + \boldsymbol{v})(H\boldsymbol{x} + \boldsymbol{v})^* &= HR_xH^* + R_v \\ R_{xy} &= \mathsf{E}\,\boldsymbol{x}\boldsymbol{y}^* = \mathsf{E}\,\boldsymbol{x}(H\boldsymbol{x} + \boldsymbol{v})^* &= R_xH^* \end{aligned}$$

Moreover, since $R_v > 0$ we get $R_y > 0$. The expression (5.2) for $\hat{\boldsymbol{x}}$ then becomes

$$\boxed{\hat{\boldsymbol{x}} = R_xH^*\left[R_v + HR_xH^*\right]^{-1}\boldsymbol{y}} \tag{5.3}$$

MATRIX INVERSION LEMMA

$$\boxed{\hat{\boldsymbol{x}} = R_x H^* \left[R_v + H R_x H^*\right]^{-1} \boldsymbol{y}} \qquad (5.3)$$

This expression can be rewritten in an equivalent form by using the so-called *matrix inversion formula* or *lemma*. This formula is a very useful matrix theory result and it will be called upon several times throughout this book. The result states that for arbitrary matrices $\{A, B, C, D\}$ of compatible dimensions, if A and C are invertible, then

$$\boxed{(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}} \qquad (5.4)$$

The identity can be verified algebraically; it essentially shows how the inverse of the sum $A + BCD$ is related to the inverse of A.

Applying (5.4) to the matrix $\left[R_v + H R_x H^*\right]^{-1}$ in (5.3), with the identifications

$$A = R_v, \quad B = H, \quad C = R_x, \quad D = H^*$$

ALTERNATIVE EXPRESSION

we obtain

$$\begin{aligned}\hat{x} &= R_x H^* \left\{R_v^{-1} - R_v^{-1} H (R_x^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1}\right\} \boldsymbol{y} \\ &= \left\{R_x (R_x^{-1} + H^* R_v^{-1} H) - R_x H^* R_v^{-1} H\right\} (R_x^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y} \\ &= \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}\end{aligned}$$

where in the second equality we factored out the term $(R_x^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \boldsymbol{y}$ from the right. Hence,

$$\boxed{\hat{x} = \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}} \tag{5.5}$$

It further follows that the m.m.s.e. matrix is given by

$$\begin{aligned}\text{m.m.s.e.} = \mathsf{E}\,\tilde{x}\tilde{x}^* &= \mathsf{E}\,(x - \hat{x})x^*, \qquad \text{since } \ \tilde{x} \perp \hat{x} \\ &= R_x - \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} H R_x \\ &= \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1}\end{aligned}$$

ESTIMATION WITH LINEAR MODELS

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

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

5.2 APPLICATION: CHANNEL ESTIMATION

Consider an FIR channel whose tap vector $\boldsymbol{c}$ is unknown; it is modeled as a zero-mean random vector with a known covariance matrix, $R_c = \mathsf{E}\,\boldsymbol{c}\boldsymbol{c}^*$. The following experiment is performed with the purpose of estimating $\boldsymbol{c}$, assumed of length M. The channel is assumed initially at rest (i.e., no initial conditions in its delay elements) and a known input sequence $\{s(i)\}$, also called a *training* sequence, is applied to the channel. The resulting output sequence $\{\boldsymbol{z}(i)\}$ is measured in the presence of additive noise, $\boldsymbol{v}(i)$, as shown in Fig. 5.1. The available measurements are

$$\boldsymbol{y}(i) = \boldsymbol{z}(i) + \boldsymbol{v}(i) \tag{5.6}$$

where $\boldsymbol{v}(i)$ is a zero-mean noise sequence that is uncorrelated with $\boldsymbol{c}$.

CHANNEL MODEL

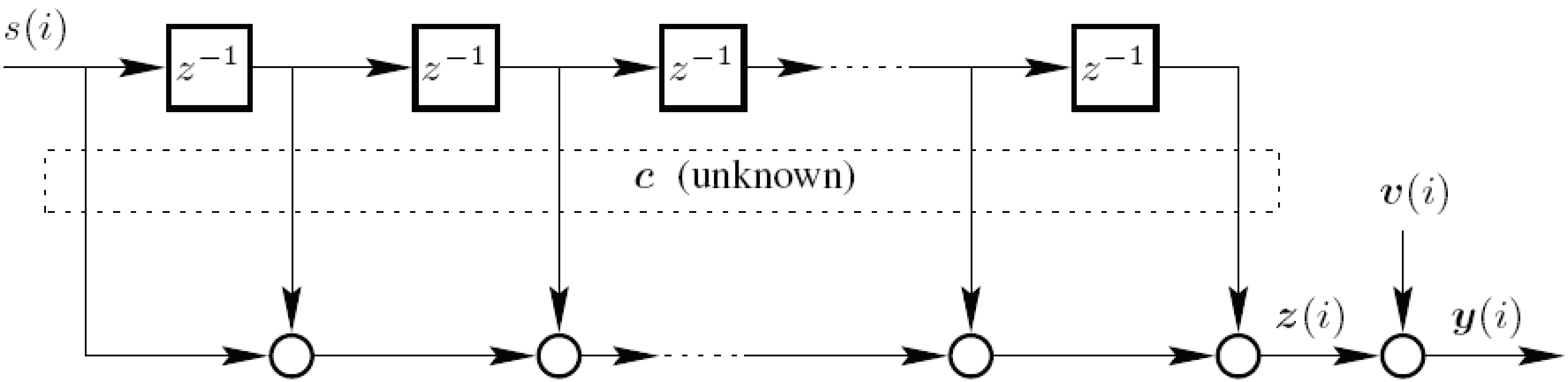

FIGURE 5.1 Channel estimation in the presence of additive noise.

COLLECTING THE DATA

Assume we collect a block of measurements $\{s(\cdot), \boldsymbol{y}(\cdot)\}$ over the interval $0 \leq i \leq N$. Then we can write in matrix form, say for $M = 3$ and $N = 6$,

$$\underbrace{\begin{bmatrix} \boldsymbol{y}(0) \\ \boldsymbol{y}(1) \\ \boldsymbol{y}(2) \\ \boldsymbol{y}(3) \\ \boldsymbol{y}(4) \\ \boldsymbol{y}(5) \\ \boldsymbol{y}(6) \end{bmatrix}}_{\boldsymbol{y}:(N+1)\times 1} = \underbrace{\begin{bmatrix} s(0) & & \\ s(1) & s(0) & \\ s(2) & s(1) & s(0) \\ s(3) & s(2) & s(1) \\ s(4) & s(3) & s(2) \\ s(5) & s(4) & s(3) \\ s(6) & s(5) & s(4) \end{bmatrix}}_{H:(N+1)\times M} \boldsymbol{c} + \underbrace{\begin{bmatrix} \boldsymbol{v}(0) \\ \boldsymbol{v}(1) \\ \boldsymbol{v}(2) \\ \boldsymbol{v}(3) \\ \boldsymbol{v}(4) \\ \boldsymbol{v}(5) \\ \boldsymbol{v}(6) \end{bmatrix}}_{\boldsymbol{v}:(N+1)\times 1} \quad (5.7)$$

where we are further defining the quantities $\{\boldsymbol{y}, H, \boldsymbol{v}\}$. Note that the *data matrix* H has a rectangular Toeplitz structure, i.e., it has constant entries along its diagonals.

CHANNEL ESTIMATION

The quantities $\{\boldsymbol{y}, H\}$ so defined are both available to the designer, in addition to the covariance matrices $\{R_c = \mathsf{E}\,\boldsymbol{c}\boldsymbol{c}^*, R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^*\}$ (by assumption). In particular, if the noise sequence $\{\boldsymbol{v}(i)\}$ is assumed white with variance σ_v^2, then $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^* = \sigma_v^2 \mathrm{I}$. With this information, we can estimate the channel as follows. Since we have a linear model relating $\boldsymbol{y}$ to $\boldsymbol{c}$, as indicated by (5.7), then according to Thm. 5.1, the optimal linear estimator for $\boldsymbol{c}$ can be obtained from either expression:

$$\boxed{\hat{\boldsymbol{c}} = R_c H^* \left[H R_c H^* + R_v\right]^{-1} \boldsymbol{y} = \left[R_c^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}} \tag{5.8}$$

5.3 APPLICATION: BLOCK DATA ESTIMATION

Our second application is in the context of data (or symbol) recovery. We consider the same FIR channel as in Fig. 5.1, except that now we assume that its tap vector is known. For example, it could have been estimated via a prior training procedure as explained in the previous section. We denote this tap vector by c, with individual entries

$$c \triangleq \text{col}\{c(0), c(1), \ldots, c(M-1)\}$$

The channel is initially at rest and its output sequence, $\{\boldsymbol{z}(i)\}$, is again measured in the presence of additive noise, $\boldsymbol{v}(i)$, as shown in Fig. 5.2. The signals $\{\boldsymbol{v}(\cdot), \boldsymbol{s}(\cdot)\}$ are assumed uncorrelated. What we would like to estimate now are the symbols $\{\boldsymbol{s}(\cdot)\}$ that are being transmitted through the channel. Observe that, to be consistent with our notation, since c is deterministic and $\boldsymbol{s}(\cdot)$ is random, we are now using a boldface letter for the latter and a normal font for the former.

DATA MODEL

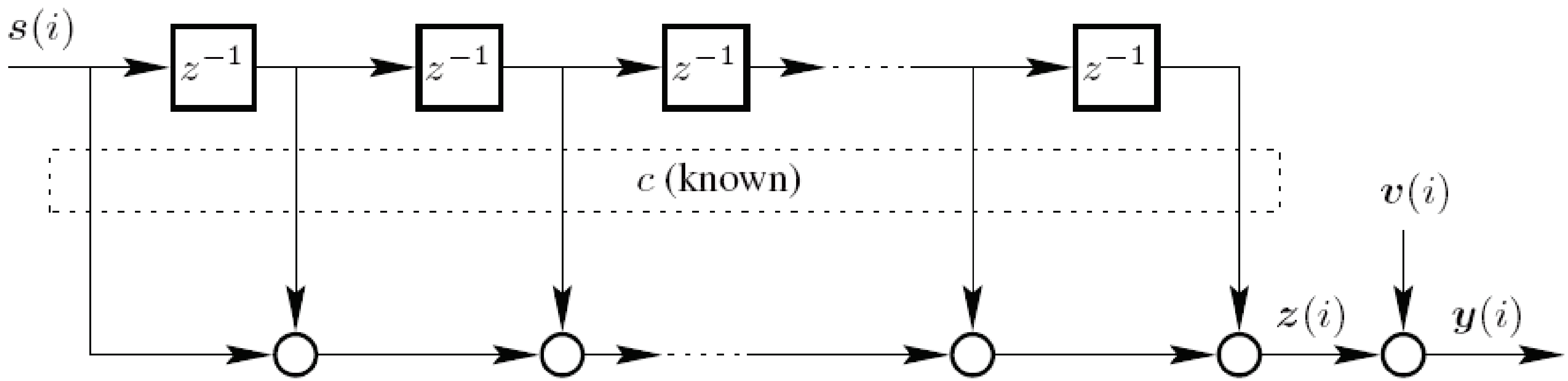

FIGURE 5.2 Block data estimation in the presence of additive noise.

COLLECTING THE DATA

Suppose we collect a block of measurements $\{\boldsymbol{y}(\cdot)\}$, say $(N+1)$ measurements over the interval $0 \leq i \leq N$. Rather than relate the $\{\boldsymbol{y}(\cdot)\}$ to the channel tap vector c through a data matrix, as we did in (5.7), we now relate the $\{\boldsymbol{y}(\cdot)\}$ to the $\{\boldsymbol{s}(\cdot)\}$ through a *channel matrix*. More specifically, assume again that $M = 3$ and $N = 6$ for illustration purposes.

Then we can write

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

BLOCK DATA ESTIMATION

Note that the channel matrix H is now square Toeplitz of size $(N+1)\times(N+1)$; it also has a banded structure. The quantities $\{\boldsymbol{y}, H\}$ so defined are available to the designer, in addition to the covariance matrices $\{R_s, R_v\}$ (by assumption). In particular, if the data and noise sequences $\{\boldsymbol{s}(\cdot), \boldsymbol{v}(\cdot)\}$ are white with variances $\{\sigma_s^2, \sigma_v^2\}$, then $R_s = \mathsf{E}\,\boldsymbol{s}\boldsymbol{s}^* = \sigma_s^2 \mathrm{I}$ and $R_v = \mathsf{E}\,\boldsymbol{v}\boldsymbol{v}^* = \sigma_v^2 \mathrm{I}$. With this information, we can estimate the symbols in the vector $\boldsymbol{s}$ as follows. Observe again that we have a linear model relating $\boldsymbol{y}$ to the unknown symbol vector $\boldsymbol{s}$. According to Thm. 5.1, the optimal linear estimator for $\boldsymbol{s}$ can then be found from either expression:

$$\boxed{\hat{\boldsymbol{s}} = R_s H^* \left[H R_s H^* + R_v\right]^{-1} \boldsymbol{y} = \left[R_s^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} \boldsymbol{y}} \tag{5.9}$$

5.4 APPLICATION: LINEAR CHANNEL EQUALIZATION

Consider again an FIR channel as shown in Fig. 5.2, with a known tap vector c of length M. Data symbols $\{\boldsymbol{s}(\cdot)\}$ are transmitted through the channel and the output sequence, $\{\boldsymbol{z}(i)\}$, is measured in the presence of additive noise, $\boldsymbol{v}(i)$. The signals $\{\boldsymbol{v}(\cdot), \boldsymbol{s}(\cdot)\}$ are assumed uncorrelated. Due to channel memory, each measurement $\boldsymbol{y}(i)$ contains contributions not only from $\boldsymbol{s}(i)$ but also from prior symbols since

$$\boldsymbol{y}(i) = c(0)\boldsymbol{s}(i) + \underbrace{\sum_{k=1}^{M-1} c(k)\boldsymbol{s}(i-k)}_{\text{ISI}} + \boldsymbol{v}(i) \tag{5.10}$$

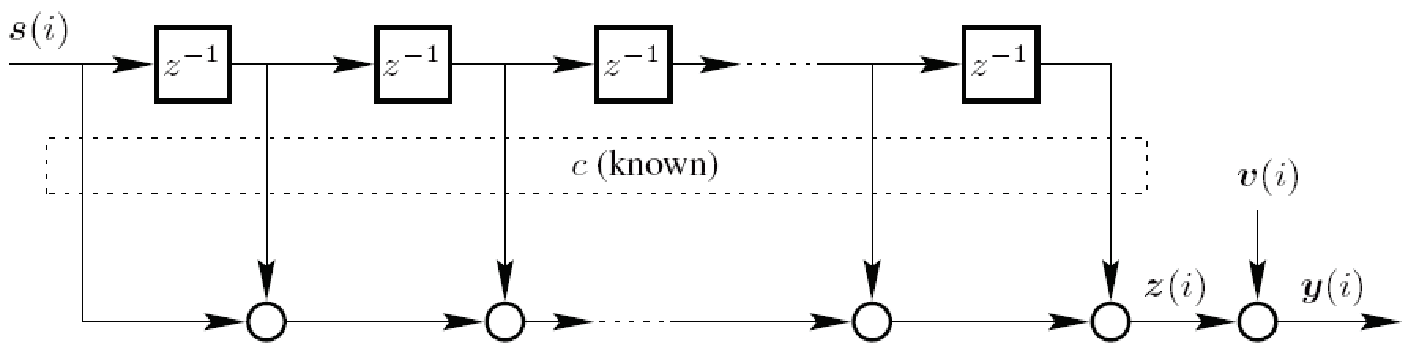

FIGURE 5.2 Block data estimation in the presence of additive noise.

INTER-SYMBOL INTERFERENCE

$$y(i) = c(0)s(i) \; + \underbrace{\sum_{k=1}^{M-1} c(k)s(i-k)}_{\text{ISI}} + v(i) \tag{5.10}$$

The second term on the right-hand side is termed *inter-symbol-interference* (ISI); it refers to the interference that is caused by prior symbols in $\boldsymbol{y}(i)$. The purpose of an equalizer is to recover $\boldsymbol{s}(i)$. To achieve this task, an equalizer does not only rely on the most recent measurement $\boldsymbol{y}(i)$, but it also employs several prior measurements $\{\boldsymbol{y}(i-k)\}$, say for $k = 1, 2, \ldots, L-1$. These prior measurements contain information that is correlated with the ISI term in $\boldsymbol{y}(i)$. For example, the expression for $\boldsymbol{y}(i-1)$ is

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

and the ISI term in it shares several data symbols with the ISI term in $\boldsymbol{y}(i)$. It is for this reason that prior measurements are useful in eliminating (or reducing) ISI.

LINEAR EQUALIZER

we are interested in an equalizer structure of the form shown in Fig. 5.3.

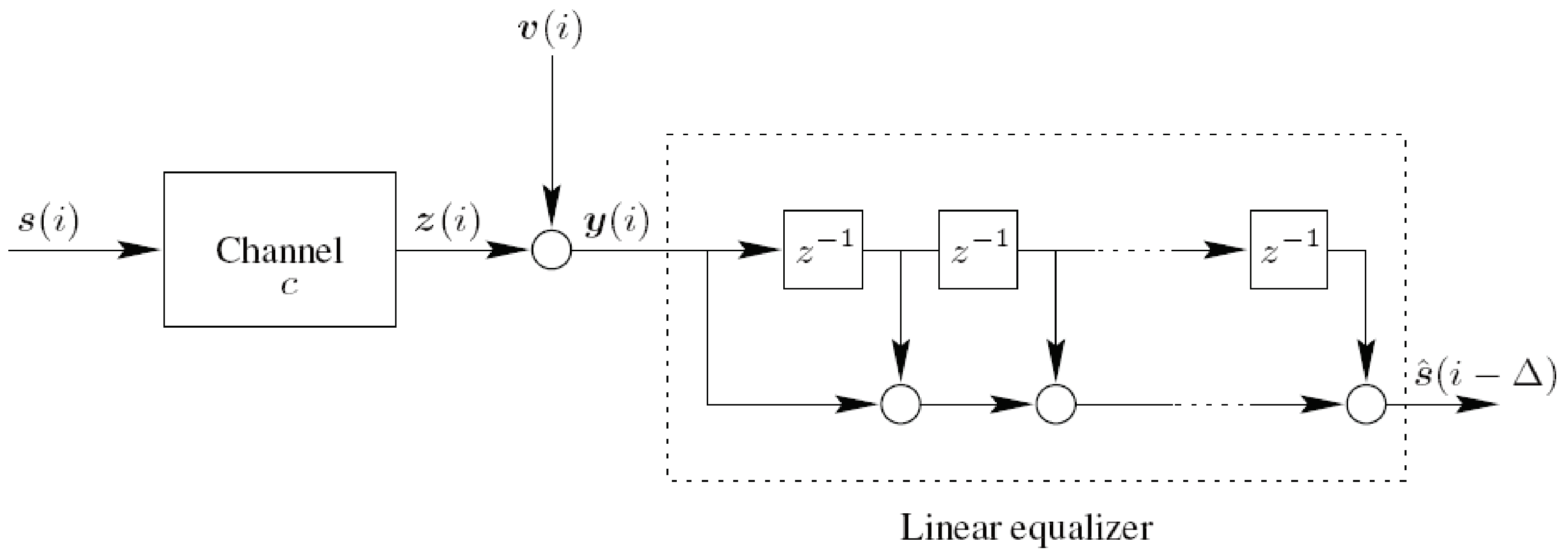

FIGURE 5.3 Linear equalization of an FIR channel in the presence of additive noise.

LINEAR EQUALIZER

The equalizer is chosen to have an FIR structure with L coefficients, so that for each time instant i, it would employ the L observations

$$\boldsymbol{y}_i \triangleq \begin{bmatrix} \boldsymbol{y}(i) \\ \boldsymbol{y}(i-1) \\ \boldsymbol{y}(i-2) \\ \vdots \\ \boldsymbol{y}(i-L+1) \end{bmatrix}$$

in order to estimate $\boldsymbol{s}(i-\Delta)$ for some integer delay $\Delta \geq 0$.

REMARK ON NOTATION

We remark that we shall frequently deal with *time sequences* in this book. So when we write $\boldsymbol{y}(i)$ we are referring to the value of the time sequence $\boldsymbol{y}(\cdot)$ at time i. Not only that, but the notation $\boldsymbol{y}(\cdot)$, with parentheses, also means that $\boldsymbol{y}(\cdot)$ is a scalar. This is because for vector-valued time-sequences, we shall instead write $\boldsymbol{y}_i$, with a subscript rather than parentheses, to refer to its value at time i. In other words, we shall use parenthesis for *time-indexing* in the scalar case, e.g., $\{\boldsymbol{y}(i)\}$, and subscripts in the vector case, e.g., $\{\boldsymbol{y}_i\}$.

RELATING THE DATA

It is useful to express the observation vector $\boldsymbol{y}_i$ in terms of the transmitted data as follows. Assume for the sake of illustration that $L = 5$ (i.e., an equalizer with 5 taps) and $M = 4$ (a channel with four taps). Then we have

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

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

LINEAR RELATION

We thus find that there is a linear relation between the vectors $\{\boldsymbol{y}_i, \boldsymbol{s}_i\}$ and this relation can be used to evaluate the covariance and cross-covariance quantities that are needed to estimate $\boldsymbol{s}(i-\Delta)$ from $\boldsymbol{y}_i$ in the linear least-mean-squares sense. Specifically, let us write

$$\hat{\boldsymbol{s}}(i-\Delta) = w^* \boldsymbol{y}_i \tag{5.11}$$

for some column vector w to be determined. We denote the optimal choice for w by w^o; the entries of w^{o*} will correspond to the optimal tap coefficients for the equalizer. According to Thm. 3.1, w^{o*} is given by

$$w^{o*} = R_{sy} R_y^{-1} \tag{5.12}$$

where

$$\boxed{R_{sy} \triangleq \mathsf{E}\, \boldsymbol{s}(i-\Delta)\boldsymbol{y}_i^* \qquad (1 \times L)} \tag{5.13}$$

denotes the cross-covariance vector between $\boldsymbol{s}(i-\Delta)$ and $\boldsymbol{y}_i$, and

$$\boxed{R_y \triangleq \mathsf{E}\, \boldsymbol{y}_i \boldsymbol{y}_i^* \qquad (L \times L)} \tag{5.14}$$

denotes the covariance matrix of the observation vector, $\boldsymbol{y}_i$.

CORRELATIONS

Observe that since the processes $\{\boldsymbol{s}(\cdot), \boldsymbol{y}(\cdot)\}$ are jointly wide-sense stationary, the quantities $\{R_{sy}, R_y\}$ are independent of i. In order to determine $\{R_{sy}, R_y\}$ we resort to the aforementioned linear model relating $\{\boldsymbol{y}_i, \boldsymbol{s}_i, \boldsymbol{v}_i, H\}$. To begin with,

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

where

$$R_s \;\stackrel{\Delta}{=}\; \mathsf{E}\,\boldsymbol{s}_i\boldsymbol{s}_i^* \quad ((L+M-1)\times(L+M-1)) \quad \text{and} \quad R_v \;\stackrel{\Delta}{=}\; \mathsf{E}\,\boldsymbol{v}_i\boldsymbol{v}_i^* \quad (L\times L)$$

and

$$R_{sy} = \mathsf{E}\,\boldsymbol{s}(i-\Delta)(H\boldsymbol{s}_i + \boldsymbol{v}_i)^* \;=\; \left(\mathsf{E}\,\boldsymbol{s}(i-\Delta)\boldsymbol{s}_i^*\right)H^*$$

since $\{\boldsymbol{v}(\cdot), \boldsymbol{s}(\cdot)\}$ are uncorrelated.

CORRELATIONS

The value of $\mathsf{E}\,\boldsymbol{s}(i-\Delta)\boldsymbol{s}_i^*$ depends on the assumed correlation between the transmitted symbols. If the $\{\boldsymbol{s}(\cdot)\}$ are independent and identically distributed with variance σ_s^2, then $R_s = \sigma_s^2 \mathrm{I}$ and

$$\begin{aligned} \mathsf{E}\,\boldsymbol{s}(i-\Delta)\boldsymbol{s}_i^* &= \mathsf{E}\,\boldsymbol{s}(i-\Delta)\left[\begin{array}{cccc} \boldsymbol{s}^*(i) & \boldsymbol{s}^*(i-1) & \ldots & \boldsymbol{s}^*(i-7) \end{array}\right] \\ &= \left[\begin{array}{ccccccc} 0 & \ldots & 0 & \sigma_s^2 & 0 & \ldots & 0 \end{array}\right] \end{aligned}$$

with Δ leading zeros.

In a similar vein, we assume that the noise sequence $\{\boldsymbol{v}(\cdot)\}$ is white with variance σ_v^2 so that $R_v = \sigma_v^2 \mathrm{I}$. Again, the development applies even for correlated (but stationary) noise. We therefore find that

$$\boxed{R_y = \sigma_s^2 H H^* + \sigma_v^2 \mathrm{I} \quad \text{and} \quad R_{sy} = \left[\begin{array}{ccccccc} 0 & \ldots & 0 & \sigma_s^2 & 0 & \ldots & 0 \end{array}\right] H^*} \tag{5.15}$$

LINEAR EQUALIZER

Observe that R_{sy} is proportional to the $(\Delta+1)-$th row of H^*. Substituting (5.15) into (5.12) we arrive at the following expression for the equalizer tap vector:

$$\boxed{w^{o*} = \sigma_s^2 e_\Delta^* H^* \left(\sigma_s^2 H H^* + \sigma_v^2 \mathrm{I}\right)^{-1}} \tag{5.16}$$

where e_Δ denotes the basis vector with Δ leading zeros,

$$e_\Delta \stackrel{\Delta}{=} \begin{bmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}^{\mathsf{T}} \qquad \text{(with } \Delta \text{ leading zeros)}$$

Moreover, according to Thm. 3.1, the resulting minimum mean-square error is given by $\text{m.m.s.e} = \sigma_s^2 - R_{sy} R_y^{-1} R_{ys}$ and, hence,

$$\boxed{\text{m.m.s.e} \;=\; \sigma_s^2 \left(1 - \sigma_s^2 e_\Delta^* H^* R_y^{-1} H e_\Delta\right)} \tag{5.17}$$

Example 5.1 (Numerical illustration)

Let us use the above results to re-examine Ex. 4.1, where $L = 3$, $M = 2$, $\Delta = 0$, $\{c(0), c(1)\} = \{1, 0.5\}$, and $\{\sigma_s^2, \sigma_v^2\} = \{1, 1\}$. Therefore, for this case, we have

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

so that from (5.15),

$$R_y = \begin{bmatrix} 9/4 & 1/2 & 0 \\ 1/2 & 9/4 & 1/2 \\ 0 & 1/2 & 9/4 \end{bmatrix} \quad \text{and} \quad R_{sy} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Using (5.16) and (5.17) we get

$$w^{o*} = \begin{bmatrix} 0.4688 & -0.1096 & 0.0244 \end{bmatrix} \quad \text{and} \quad \text{m.m.s.e} = 0.5312$$

which are the same results from Ex. 4.1.

◇

5.5 APPLICATION: MULTIPLE-ANTENNA RECEIVERS

Let us reconsider Ex. 4.1 with N noisy measurements,

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

of some zero-mean random variable $\boldsymbol{x}$. Let $\boldsymbol{y} = \text{col}\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(N-1)\}$ denote the observation vecto

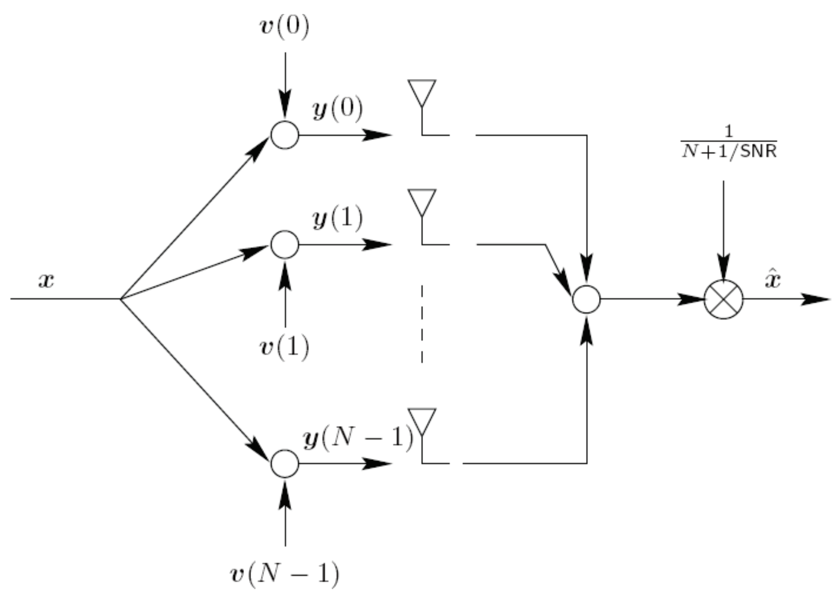

FIGURE 5.4 An optimal linear receiver for recovering a symbol x transmitted over additive-noise channels from multiple-antenna measurements.

SOLUTION

For generality, we shall assume that the variances of $\boldsymbol{x}$ and $\boldsymbol{v}(i)$ are σ_x^2 and σ_v^2, respectively. In Ex. 4.1, we used $\sigma_x^2 = \sigma_v^2 = 1$.

Introduce the $N \times 1$ column vectors:

$$\boldsymbol{v} \stackrel{\Delta}{=} \text{col}\{\boldsymbol{v}(0), \boldsymbol{v}(1), \ldots, \boldsymbol{v}(N-1)\}, \qquad h \stackrel{\Delta}{=} \text{col}\{1, 1, \ldots, 1\}$$

Then $\boldsymbol{y} = h\boldsymbol{x} + \boldsymbol{v}$ and the covariance matrix of $\boldsymbol{v}$ is $\sigma_v^2 \mathrm{I}$. We now obtain from (5.5) that

$$\hat{\boldsymbol{x}} = \left[1/\sigma_x^2 \;+\; h^* h/\sigma_v^2\right]^{-1} h^* \boldsymbol{y}/\sigma_v^2 \;=\; \frac{1}{N + 1/\mathsf{SNR}} \sum_{i=0}^{N-1} \boldsymbol{y}(i)$$

where $\mathsf{SNR} = \sigma_x^2/\sigma_v^2$. Observe that we are not dividing by the number of observations (which is N) but by $N + 1/\mathsf{SNR}$. We shall comment on the significance of this observation in the next chapter.