



# EE210A: Adaptation and Learning

## Professor Ali H. Sayed



# MAIN RESULT FROM PREVIOUS LECTURE

**Theorem 3.1 (Optimal linear estimator)** Given zero-mean random variables  $x$  and  $y$ , the linear least-mean-squares estimator (l.l.m.s.e.) of  $x$  given  $y$  is

$$\hat{x} = K_o y$$

where  $K_o$  is any solution to the linear system of equations  $K_o R_y = R_{xy}$ . This estimator minimizes the following two error measures:

$$\min_K \mathbb{E} \tilde{x}^* \tilde{x} \quad \text{and} \quad \min_K \mathbb{E} \tilde{x} \tilde{x}^*$$

The scalar cost on the left is the trace of the matrix cost on the right. The resulting minimum mean-square errors, as defined by (3.4) and (3.22), are given by

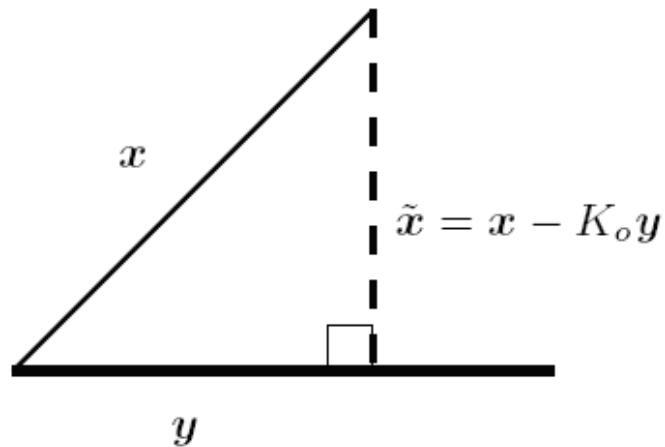
$$\min_K \mathbb{E} \tilde{x}^* \tilde{x} = \text{Tr}(R_x - K_o R_y K_o^*)$$

$$\min_K \mathbb{E} \tilde{x} \tilde{x}^* = R_x - K_o R_y K_o^*$$

# MAIN RESULT FROM PREVIOUS LECTURE

$$\tilde{x} \perp \hat{x}$$

(4.5)



**FIGURE 4.4** The orthogonality condition for linear estimation:  $\tilde{x} \perp y$ .

# LECTURE #04

## NORMAL EQUATIONS

Sections in order: B.2, 4.1, 4.3, 4.4

# BACKGROUND: LINEAR ALGEBRA

## B.2 RANGE SPACES AND NULLSPACES OF MATRICES

Let  $A$  denote an  $m \times n$  matrix without any constraint on the relative sizes of  $n$  and  $m$ .

**Range spaces.** The *column span* or the *range space* of  $A$  is defined as the set of all  $m \times 1$  vectors  $q$  that can be generated by  $Ap$ , for all  $n \times 1$  vectors  $p$ . We denote the column span of  $A$  by

$$\mathcal{R}(A) \triangleq \{\text{set of all } q \text{ such that } q = Ap \text{ for some } p\}$$

**Nullspaces.** The nullspace of  $A$  is the set of all  $n \times 1$  vectors  $p$  that are annihilated by  $A$ , namely, that satisfy  $Ap = 0$ . We denote the nullspace of  $A$  by

$$\mathcal{N}(A) \triangleq \{\text{set of all } p \text{ such that } Ap = 0\}$$

# BACKGROUND: LINEAR ALGEBRA

**Properties.** A useful property that follows from the definitions of range spaces and nullspaces is that any vector  $z$  from the nullspace of  $A^*$  (not  $A$ ) is orthogonal to any vector  $p$  in the range space of  $A$ , i.e.,

$$z \in \mathcal{N}(A^*), \quad q \in \mathcal{R}(A) \implies z^* q = 0$$

Indeed,  $z \in \mathcal{N}(A^*)$  implies that  $A^* z = 0$  or, equivalently,  $z^* A = 0$ . Now write  $q = Ap$  for some  $p$ . Then  $z^* q = z^* Ap = 0$ , as desired. Another useful property is that the matrices  $A^* A$  and  $A^*$  have the *same* range space (i.e., they span the same space). Also,  $A$  and  $A^* A$  have the same nullspace.

**Lemma B.3 (Range and nullspaces)** For any  $m \times n$  matrix  $A$ , it holds that  $\mathcal{R}(A^*) = \mathcal{R}(A^* A)$  and  $\mathcal{N}(A) = \mathcal{N}(A^* A)$ .

# BACKGROUND: LINEAR ALGEBRA

**Proof:** One direction is immediate. Take a vector  $q \in \mathcal{R}(A^*A)$ , i.e.,  $q = A^*Ap$  for some  $p$ . Define  $r = Ap$ , then  $q = A^*r$ . This shows that  $q \in \mathcal{R}(A^*)$  and we conclude that  $\mathcal{R}(A^*A) \subset \mathcal{R}(A^*)$ . The proof of the converse statement requires more effort.

Take a vector  $q \in \mathcal{R}(A^*)$  and let us show that  $q \in \mathcal{R}(A^*A)$ . Assume, to the contrary, that  $q$  does not lie in  $\mathcal{R}(A^*A)$ . This implies that there exists a vector  $z$  in the nullspace of  $A^*A$  that is not orthogonal to  $q$ , i.e.,  $A^*Az = 0$  and  $z^*q \neq 0$ . Now, if we multiply the equality  $A^*Az = 0$  by  $z^*$  from the left we obtain that  $z^*A^*Az = 0$  or, equivalently,  $\|Az\|^2 = 0$ , where  $\|\cdot\|$  denotes the Euclidean norm of its vector argument. Therefore,  $Az$  is necessarily the zero vector,  $Az = 0$ . But from  $q \in \mathcal{R}(A^*)$  we have that  $q = A^*p$  for some  $p$ . Then  $z^*q = z^*A^*p = 0$ , which contradicts  $z^*q \neq 0$ . Therefore, we must have  $q \in \mathcal{R}(A^*A)$  and we conclude that  $\mathcal{R}(A^*) \subset \mathcal{R}(A^*A)$ .

The second assertion in the lemma is more immediate. If  $Ap = 0$  then  $A^*Ap = 0$  so that  $\mathcal{N}(A) \subset \mathcal{N}(A^*A)$ . Conversely, if  $A^*Ap = 0$  then  $p^*A^*Ap = 0$  and we must have  $Ap = 0$ . That is,  $\mathcal{N}(A^*A) \subset \mathcal{N}(A)$ . Combining both facts we conclude that  $\mathcal{N}(A) = \mathcal{N}(A^*A)$ . ◇

# CONSISTENT EQUATIONS

A useful consequence of Lemma B.3 is that linear systems of equations of the form  $A^*Ax = A^*b$  always have a solution  $x$  for any vector  $b$ . This is because  $A^*b$  belongs to  $\mathcal{R}(A^*)$  and, therefore, also belongs to  $\mathcal{R}(A^*A)$ .

**Rank.** The rank of a matrix  $A$  is defined as the number of linearly independent columns (or rows) of  $A$ . It holds that

$$\boxed{\text{rank}(A) \leq \min\{m, n\}}$$

That is, the rank of a matrix never exceeds its smallest dimension. A matrix is said to have *full rank* if

$$\boxed{\text{rank}(A) = \min\{m, n\}}$$

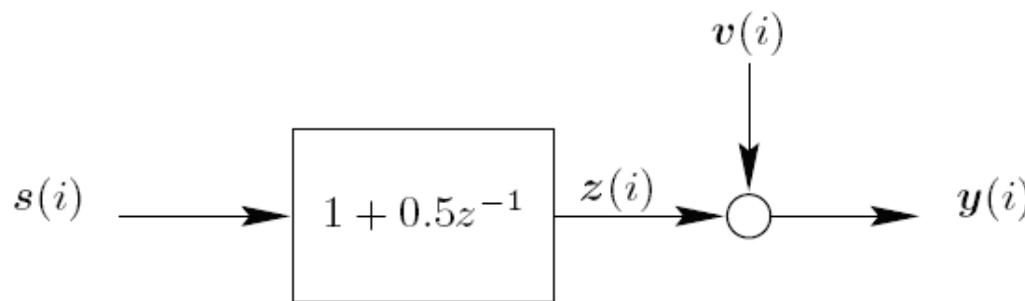
Otherwise, the matrix is said to be *rank deficient*.

If  $A$  is a square matrix (i.e.,  $m = n$ ), then rank deficiency is also equivalent to a zero determinant,  $\det A = 0$ . Indeed, if  $A$  is rank deficient, then there exists a nonzero  $p$  such that  $Ap = 0$ . This means that 0 is an eigenvalue of  $A$  so that its determinant must be zero — recall that the determinant of a square matrix is equal to the product of its eigenvalues (see Prob. II.2).

# EXAMPLE: DATA ESTIMATION

## Example 4.3 (Transmissions over a noisy channel)

Consider again the setting of Exs. A.4 and 2.1, where independent and identically distributed (i.i.d.) symbols  $\{s(i)\}$  are transmitted over the FIR channel  $C(z) = 1 + 0.5z^{-1}$ . Each symbol is either  $+1$  with probability  $p$  or  $-1$  with probability  $1 - p$ , and the output of the channel is corrupted by zero-mean additive white Gaussian noise  $v(i)$  of unit variance. The noise and the symbols are independent of each other (see Fig. 4.2). We also assume, for simplicity, that  $p = 1/2$ .



**FIGURE 4.2** Data transmissions through an additive Gaussian-noise channel.

# EXAMPLE: DATA ESTIMATION

We want to

estimate the vector  $\mathbf{x} = \text{col}\{s(0), s(1)\}$  from the observation vector  $\mathbf{y} = \text{col}\{\mathbf{y}(0), \mathbf{y}(1)\}$ , where

$$\mathbf{y}(0) = s(0) + \mathbf{v}(0), \quad \mathbf{y}(1) = s(1) + 0.5s(0) + \mathbf{v}(1) \quad (4.1)$$

We are assuming that transmissions start at time 0 so that  $s(-1) = 0$ . According to Thm. 3.1, the optimal linear estimator of  $\mathbf{x}$  is  $\hat{\mathbf{x}} = K_o \mathbf{y}$ , where  $K_o R_y = R_{xy}$ .

It follows from the relations (4.1) that

$$R_{xy} = \mathbb{E} \begin{bmatrix} s(0) \\ s(1) \end{bmatrix} \begin{bmatrix} \mathbf{y}^*(0) & \mathbf{y}^*(1) \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$$

Moreover,

$$R_y = \mathbb{E} \begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \end{bmatrix} \begin{bmatrix} \mathbf{y}^*(0) & \mathbf{y}^*(1) \end{bmatrix} = \begin{bmatrix} 2 & 1/2 \\ 1/2 & 9/4 \end{bmatrix}$$

so that

$$K_o = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1/2 \\ 1/2 & 9/4 \end{bmatrix}^{-1} = \frac{4}{17} \begin{bmatrix} 2 & 1/2 \\ -1/2 & 2 \end{bmatrix}$$

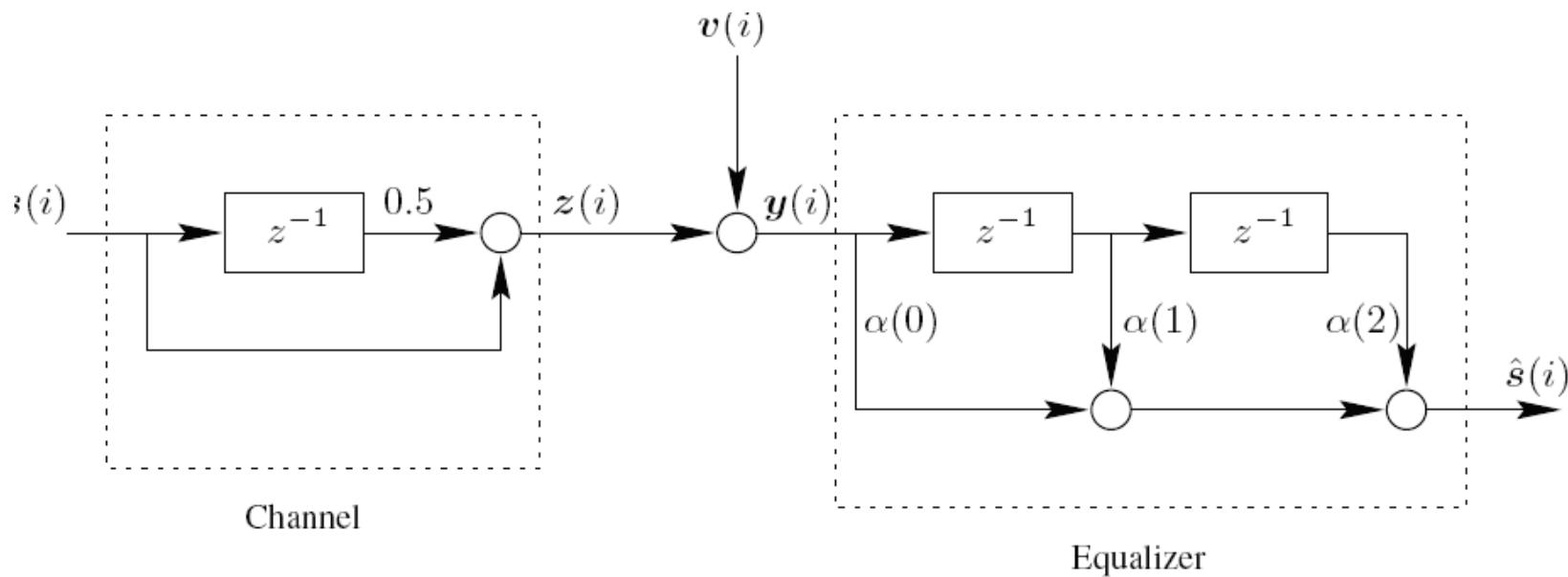
That is,

$$\hat{s}(0) = 8\mathbf{y}(0)/17 + 2\mathbf{y}(1)/17 \quad \text{and} \quad \hat{s}(1) = -2\mathbf{y}(0)/17 + 8\mathbf{y}(1)/17$$

# EXAMPLE: CHANNEL EQUALIZATION

## Example 4.4 (Linear channel equalization)

Consider again the setting of Ex. 4.1, and assume now that the transmissions started in the remote past (rather than at time 0, i.e.,  $i > -\infty$ ) so that all random processes  $\{s(i), v(i), y(i)\}$  can be assumed to be wide-sense stationary — the need for this assumption will become evident soon. By a wide-sense stationary process  $s(\cdot)$ , we mean one with a constant mean and whose auto-correlation sequence is only a function of the time lag, i.e.,  $r_s(k) = \mathbb{E} s(i)s^*(i - k)$  is only a function of  $k$ .



**FIGURE 4.3** Linear channel equalization.

# EXAMPLE: CHANNEL EQUALIZATION

Now referring to Fig. 4.3, we see that the output of the channel at any time instant  $i$  is a linear combination of the current symbol  $s(i)$  and the previous symbol  $s(i - 1)$ , i.e.,

$$z(i) = s(i) + 0.5s(i - 1)$$

We therefore say that the channel introduces *inter-symbol interference* or ISI, since a symbol transmitted at a prior time,  $s(i - 1)$ , interferes with the output at the time of the current symbol,  $s(i)$ . The measurement  $y(i)$  that is available at the receiver is a noisy version of  $z(i)$ , namely,

$$y(i) = s(i) + 0.5s(i - 1) + v(i) \quad (4.2)$$

The purpose of this example is to show how to design an equalizer for the channel. The function of the equalizer is to process the received signal  $\{y(i)\}$  in order to recover the transmitted symbol  $\{s(i)\}$ , or a delayed version of it, say  $\{s(i - \Delta)\}$  for some  $\Delta$ . There are many different structures that can be used for equalization purposes. In Fig. 4.3 we show an FIR equalizer structure that consists of three taps  $\{\alpha(0), \alpha(1), \alpha(2)\}$ . Its output at any time instant  $i$  is given by the linear combination

$$\alpha(0)y(i) + \alpha(1)y(i - 1) + \alpha(2)y(i - 2)$$

We wish to determine the taps  $\{\alpha(0), \alpha(1), \alpha(2)\}$  so that the output of the equalizer is the optimal linear estimator for  $s(i)$  (we choose  $\Delta = 0$  in this example — see Prob. II.17 for nonzero values of  $\Delta$ ).

# EXAMPLE: CHANNEL EQUALIZATION

Observe from Fig. 4.3 that at any time instant  $i$ , the equalizer uses three observations  $\{y(i), y(i-1), y(i-2)\}$  in order to estimate  $s(i)$ . Therefore, the observation vector is  $y = \text{col}\{y(i), y(i-1), y(i-2)\}$  and the variable we wish to estimate is  $x = s(i)$ . We then know from Thm.3.1 that the optimal linear estimator for  $x$  is given by

$$\hat{x} \triangleq \hat{s}(i) = k_o^* y$$

where the row vector  $k_o^*$  is found from solving the normal equations  $k_o^* R_y = R_{xy}$ . Once  $k_o^*$  is found, its entries give the desired tap coefficients  $\{\alpha(0), \alpha(1), \alpha(2)\}$ , i.e.,

$$k_o^* = \begin{bmatrix} \alpha(0) & \alpha(1) & \alpha(2) \end{bmatrix}$$

In order to find  $k_o^*$ , we need to determine  $\{R_{xy}, R_y\}$ . Thus let  $r_y(k)$  denote the auto-correlation sequence of the stationary process  $\{y(i)\}$ , i.e.,  $r_y(k) = \mathbb{E} y(i)y^*(i-k)$ . Then

$$R_y = \mathbb{E} yy^* = \begin{bmatrix} r_y(0) & r_y(1) & r_y(2) \\ r_y^*(1) & r_y(0) & r_y(1) \\ r_y^*(2) & r_y^*(1) & r_y(0) \end{bmatrix}$$

# EXAMPLE: CHANNEL EQUALIZATION

To determine  $\{r_y(0), r_y(1), r_y(2)\}$ , we use the output equation (4.2). Multiplying it from the right by  $\mathbf{y}^*(i)$  we get

$$\mathbf{y}(i)\mathbf{y}^*(i) = s(i)\mathbf{y}^*(i) + 0.5s(i-1)\mathbf{y}^*(i) + \mathbf{v}(i)\mathbf{y}^*(i)$$

Taking expectations of both sides, and recalling that the variables  $\{s(i), s(i-1), \mathbf{v}(i)\}$  are independent of each other, we find that

$$\begin{aligned}\mathbb{E} \mathbf{y}(i)\mathbf{y}^*(i) &= r_y(0) \\ \mathbb{E} s(i)\mathbf{y}^*(i) &= \mathbb{E} s(i)[s(i) + 0.5s(i-1) + \mathbf{v}(i)]^* = 1 \\ \mathbb{E} s(i-1)\mathbf{y}^*(i) &= \mathbb{E} s(i-1)[s(i) + 0.5s(i-1) + \mathbf{v}(i)]^* = 1/2 \\ \mathbb{E} \mathbf{v}(i)\mathbf{y}^*(i) &= \mathbb{E} \mathbf{v}(i)[s(i) + 0.5s(i-1) + \mathbf{v}(i)]^* = 1\end{aligned}$$

so that

$$r_y(0) = 1 + 1/4 + 1 = 9/4$$

# EXAMPLE: CHANNEL EQUALIZATION

Likewise, multiplying (4.2) from the right by  $\mathbf{y}^*(i-1)$  and taking expectations of both sides, we get  $r_y(1) = 1/2$ . Finally, multiplying (4.2) from the right by  $\mathbf{y}^*(i-2)$  and taking expectations we get  $r_y(2) = 0$ . In summary, we find that

$$R_y = \begin{bmatrix} 9/4 & 1/2 & 0 \\ 1/2 & 9/4 & 1/2 \\ 0 & 1/2 & 9/4 \end{bmatrix}$$

In a similar fashion, we can evaluate  $R_{xy}$ ,

$$R_{xy} = \mathbb{E} \mathbf{x} \mathbf{y}^* = \begin{bmatrix} \mathbb{E} s(i) \mathbf{y}^*(i) & \mathbb{E} s(i) \mathbf{y}^*(i-1) & \mathbb{E} s(i) \mathbf{y}^*(i-2) \end{bmatrix}$$

Thus multiplying (4.2) from the right by  $s^*(i)$  and taking expectations we get  $\mathbb{E} \mathbf{y}(i) s^*(i) = 1$ . Likewise, multiplying (4.2) from the right by  $s^*(i+1)$  and taking expectations we get  $\mathbb{E} \mathbf{y}(i) s^*(i+1) = 0$ . Similarly,  $\mathbb{E} \mathbf{y}(i) s^*(i+2) = 0$ . It further follows from the assumed stationarity of the processes  $\{s(i), \mathbf{v}(i)\}$  that

$$\mathbb{E} \mathbf{y}(i) s^*(i+1) = \mathbb{E} \mathbf{y}(i-1) s^*(i) = [\mathbb{E} s(i) \mathbf{y}^*(i-1)]^*$$

$$\mathbb{E} \mathbf{y}(i) s^*(i+2) = \mathbb{E} \mathbf{y}(i-2) s^*(i) = [\mathbb{E} s(i) \mathbf{y}^*(i-2)]^*$$

Hence,

$$R_{xy} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

# EXAMPLE: CHANNEL EQUALIZATION

Using the just derived values for  $\{R_{xy}, R_y\}$ , we are led to

$$k_o^* = R_{xy}R_y^{-1} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 9/4 & 1/2 & 0 \\ 1/2 & 9/4 & 1/2 \\ 0 & 1/2 & 9/4 \end{bmatrix}^{-1} = \begin{bmatrix} 0.4688 & -0.1096 & 0.0244 \end{bmatrix}$$

That is,  $\alpha(0) = 0.4688$ ,  $\alpha(1) = -0.1096$ , and  $\alpha(2) = 0.0244$ . Moreover, the resulting m.m.s.e. is  $m.m.s.e. = \sigma_x^2 - R_{xy}k_o = 0.5312$ , where  $\sigma_x^2 = \sigma_s^2 = 1$ . A computer project at the end of this part illustrates the operation of optimal linear equalizers designed in this manner.



## 4.3 EXISTENCE OF SOLUTIONS

Consider again the normal equations (3.9), namely,

$$K_o R_y = R_{xy} \quad (4.7)$$

In general, such linear systems of equations can have a unique solution, no solution at all, or an infinite number of solutions. This depends on the rank of the coefficient matrix  $R_y$  and on how the right-hand side matrix  $R_{xy}$  relates to  $R_y$ . We shall explain these facts here and, in the process, provide the reader with an opportunity to get acquainted with some basic concepts from matrix theory and linear algebra.

# NORMAL EQUATIONS

$$K_o R_y = R_{xy} \quad (4.7)$$

**Unique solution.** The argument will show that a unique solution  $K_o$  exists if, and only if, the covariance matrix  $R_y$  is positive-definite. Indeed, if  $R_y > 0$  then all its eigenvalues are positive and, consequently,  $R_y$  is nonsingular. In this case, equation (4.7) will have a unique solution  $K_o$  given by  $K_o = R_{xy}R_y^{-1}$ . Conversely, assume a unique solution  $K_o$  exists to the normal equations (4.7) and let us establish that it must hold that  $R_y > 0$ . Assume, to the contrary, that  $R_y$  is singular. Then there should exist a nonzero row vector  $c^*$  such that  $c^* R_y = 0$  or, equivalently,  $R_y c = 0$ . The vector  $c$  belongs to the *nullspace* of  $R_y$ , written as  $c \in \mathcal{N}(R_y)$ . It is now easy to see that by adding to the rows of  $K_o$  any such vector  $c^*$ , we obtain a matrix  $K'_o$  that satisfies the same equations (4.7), i.e.,  $K'_o R_y = R_{xy}$ . This contradicts the fact that  $K_o$  is unique so that  $R_y$  must be positive-definite.

# NORMAL EQUATIONS

**Infinitely many solutions.** We now show that the normal equations (4.7) will have infinitely many solutions  $K_o$  if, and only if,  $R_y$  is singular. One direction of the proof is obvious. Assume  $K_o$  is one solution and that  $R_y$  is singular. Then by adding to the rows of  $K_o$  any vector  $c^*$  (or any combination of vectors) from the nullspace of  $R_y$ , we obtain another solution  $K'_o$  (as explained above). Hence, infinitely many solutions exist in this case. Conversely assume that many solutions exist. Let  $K_o$  and  $K'_o$  denote any two of these solutions. Then subtracting the equations  $K_o R_y = R_{xy}$  and  $K'_o R_y = R_{xy}$  we obtain  $(K_o - K'_o)R_y = 0$ , which means that  $R_y$  is singular, since there is at least one nonzero row in  $K_o - K'_o$  and this row annihilates  $R_y$ .

# NORMAL EQUATIONS

**Existence of solutions.** The only question that remains regarding  $K_oR_y = R_{xy}$  is whether solutions always exist. The answer is affirmative. That is, the normal equations (4.7) are always *consistent*.

To establish this fact for the normal equations  $K_oR_y = R_{xy}$ , we need to show that for any two random variables  $\{x, y\}$ , it always holds that the columns of  $R_{yx}$  lie in the column span of  $R_y$ , i.e., that there exists at least one matrix  $K_o^*$  that satisfies  $R_y K_o^* = R_{yx}$ , which by transposition is equivalent to  $K_o R_y = R_{xy}$ . This statement is obviously true when  $R_y$  is nonsingular since then  $K_o = R_{xy} R_y^{-1}$ . The argument is more involved when  $R_y$  is singular. We first establish a preliminary result that provides an equivalent characterization for checking whether the normal equations  $K_o R_y = R_{xy}$  are consistent or not.

# NORMAL EQUATIONS

**Lemma 4.1 (Consistent equations)** The equations  $K_o R_y = R_{xy}$  are consistent (i.e., there exists at least one solution  $K_o$ ) if, and only if,

$$c^* R_{yx} = 0 \quad \text{for any } c \in \mathcal{N}(R_y)$$

that is,  $c^* R_{yx} = 0$  for any column vector  $c$  satisfying  $R_y c = 0$ .

**Proof:** We first verify that the condition  $c^* R_{yx} = 0$  for all  $c \in \mathcal{N}(R_y)$  implies the existence of a matrix  $K_o$  satisfying  $K_o R_y = R_{xy}$ . Let us assume, to the contrary, that the equations  $K_o R_y = R_{xy}$  are not consistent. This means that  $R_{yx}$  does not lie in the column span of  $R_y$ , which in turn means that there should exist a vector  $c \in \mathcal{N}(R_y)$  that is not orthogonal to  $R_{yx}$ . This conclusion contradicts the assumption that  $c^* R_{yx} = 0$  for all  $c \in \mathcal{N}(R_y)$ .

We now establish the converse statement, namely, that the existence of a  $K_o$  satisfying  $K_o R_y = R_{xy}$  implies  $c^* R_{yx} = 0$  for all  $c \in \mathcal{N}(R_y)$ . This claim is obvious. For any such  $c$ , we have  $R_y c = 0$  and, hence,  $K_o R_y c = 0$ . This implies that  $R_{xy} c = 0$  or, equivalently,  $c^* R_{yx} = 0$ .  $\diamond$

# NORMAL EQUATIONS

Therefore, in order to prove that the equations  $K_o R_y = R_{xy}$  are consistent, it is enough to prove that  $c^* R_{yx} = 0$  for all  $c \in \mathcal{N}(R_y)$ . So let us assume that this latter condition does not hold. This means that there exists at least one nonzero vector  $c$  such that  $c^* R_y = 0$  and  $c^* R_{yx} \neq 0$ . This statement leads to a contradiction. Indeed,  $c^* R_y = 0$  implies that  $c^* R_y c = c^*(\mathbb{E} \mathbf{y} \mathbf{y}^*)c = 0$  or, equivalently,  $\mathbb{E} |c^* \mathbf{y}|^2 = 0$ . We therefore have a zero-mean random variable  $c^* \mathbf{y}$  (because  $\mathbf{y}$  is zero-mean itself) with zero variance. Using Remark A.1, we conclude that  $c^* \mathbf{y}$  is the zero random variable with probability one. It follows that

$$c^* R_{yx} \stackrel{\Delta}{=} c^*(\mathbb{E} \mathbf{y} \mathbf{x}^*) = \mathbb{E} (c^* \mathbf{y}) \mathbf{x}^* = 0$$

This contradicts the assumption  $c^* R_{yx} \neq 0$  and we conclude that the normal equations (4.7) are always consistent.

# NORMAL EQUATIONS

**Uniqueness of estimator.** Interesting enough, regardless of which solution  $K_o$  we pick (in the case when a multitude of solutions exist), the resulting estimator,  $\hat{x} = K_o \mathbf{y}$ , and the resulting m.m.s.e.,  $R_x - K_o R_y K_o^*$ , will always assume the same values. In other words, their values are independent of the specific choice of  $K_o$ .

To see this, let us first establish the result for the m.m.s.e. Thus let  $K_o$  and  $K'_o$  be two possible solutions of (4.7), i.e.,  $K_o R_y = R_{xy}$  and  $K'_o R_y = R_{xy}$ . Then the difference  $K_o - K'_o$  satisfies

$$[K_o - K'_o] R_y = 0 \quad (4.8)$$

Now denote the minimum mean-square errors by

$$\Delta_1 = R_x - K_o R_y K_o^* = R_x - R_{xy} K_o^* \quad \text{and} \quad \Delta_2 = R_x - K'_o R_y K_o'^* = R_x - R_{xy} K_o'^*$$

Subtracting the expressions for  $\{\Delta_1, \Delta_2\}$  we obtain

$$\Delta_2 - \Delta_1 = R_{xy} [K_o - K'_o]^* = K_o R_y [K_o - K'_o]^* = 0$$

where in the second equality we substituted  $R_{xy}$  by  $K_o R_y$ , and in the third equality we used (4.8). Therefore,  $\Delta_1 = \Delta_2$ . This means, as desired, that the value of the m.m.s.e. is independent of  $K_o$ .

# NORMAL EQUATIONS

Let us now verify that no matter which  $K_o$  we pick, the corresponding estimator  $\hat{\mathbf{x}}$  will be the same. So let again  $K_o$  and  $K'_o$  be two possible solutions and define  $C = K'_o - K_o$ . Then from (4.8) we have  $CR_y = 0$ . Let further  $\hat{\mathbf{x}} = K_o \mathbf{y}$  and  $\hat{\mathbf{x}}' = K'_o \mathbf{y}$ . We want to verify that  $\hat{\mathbf{x}} = \hat{\mathbf{x}}'$  with probability one. For this purpose, note that the condition  $CR_y = 0$  implies  $CR_y C^* = 0$  or, equivalently,  $E(C\mathbf{y})(\mathbf{y}^* C^*) = 0$ . We therefore have a zero-mean random variable  $C\mathbf{y}$  (because  $\mathbf{y}$  is zero-mean itself) with a zero covariance matrix. Using Remark A.1, we again conclude that  $C\mathbf{y}$  is the zero random variable with probability one. It then follows from

$$\hat{\mathbf{x}}' = K'_o \mathbf{y} = (K_o + C)\mathbf{y} = \hat{\mathbf{x}} + C\mathbf{y}$$

that  $\hat{\mathbf{x}} = \hat{\mathbf{x}}'$  with probability one, as desired. We summarize our discussions in the following statement.

# NORMAL EQUATIONS

**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,  $\hat{x} = K_o y$ , and the m.m.s.e.,  $(R_x - K_o R_y K_o^*)$ , remain invariant.

## 4.4 NONZERO-MEAN VARIABLES

Starting from Sec. 3.1, the discussion has focused so far on zero-mean random variables  $x$  and  $y$ . When the means are nonzero, we should seek an unbiased estimator for  $x$  of the form

$$\hat{x} = Ky + b \quad (4.9)$$

for some matrix  $K$  and some vector  $b$ . As before, the optimal values for  $\{K, b\}$  are determined through the minimization of the mean-square error,

$$\min_{K, b} E \tilde{x}^* \tilde{x} \quad (4.10)$$

where  $\tilde{x} = x - \hat{x}$ .

# NONZERO MEANS

To solve this problem, we start by noting that since the estimator should be unbiased we must enforce  $\mathbb{E}\hat{\mathbf{x}} = \bar{\mathbf{x}}$ . Taking expectations of both sides of (4.9) shows that the vector  $\mathbf{b}$  must satisfy  $\bar{\mathbf{x}} = K\bar{\mathbf{y}} + \mathbf{b}$ . Using this expression for  $\mathbf{b}$ , we can eliminate it from the expression for  $\hat{\mathbf{x}}$ , which becomes  $\hat{\mathbf{x}} = K\mathbf{y} + (\bar{\mathbf{x}} - K\bar{\mathbf{y}})$  or, equivalently,

$$(\hat{\mathbf{x}} - \bar{\mathbf{x}}) = K(\mathbf{y} - \bar{\mathbf{y}}) \quad (4.11)$$

This expression shows that the desired gain matrix  $K$  should map the now zero-mean variable  $(\mathbf{y} - \bar{\mathbf{y}})$  to another zero-mean variable  $(\hat{\mathbf{x}} - \bar{\mathbf{x}})$ . In other words, we are reduced to solving the problem of estimating the zero-mean random variable  $\mathbf{x} - \bar{\mathbf{x}}$  from the also zero-mean random variable  $\mathbf{y} - \bar{\mathbf{y}}$ .

We already know that the solution  $K_o$  is found by solving

$$K_o R_y = R_{xy} \quad (4.12)$$

in terms of the covariance and cross-covariance matrices  $\{R_y, R_{xy}\}$  of the zero-mean variables  $\{\mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}}\}$ , i.e.,

$$R_y \triangleq \mathbb{E}(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^*, \quad R_{xy} \triangleq \mathbb{E}(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^*$$

# NONZERO MEANS

Therefore, the optimal solution in the nonzero-mean case is given by

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + K_o[\mathbf{y} - \bar{\mathbf{y}}] \quad (4.13)$$

with  $K_o$  obtained from solving (4.12).

$$K_o \mathsf{E}(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^* = \mathsf{E}(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^*$$

or, equivalently,

$$\mathsf{E}[(\mathbf{x} - \bar{\mathbf{x}}) - K_o(\mathbf{y} - \bar{\mathbf{y}})](\mathbf{y} - \bar{\mathbf{y}})^* = 0$$

so that the orthogonality condition (4.4) in the nonzero-mean case becomes

$$\mathsf{E}\tilde{\mathbf{x}}(\mathbf{y} - \bar{\mathbf{y}})^* = 0 \quad \text{or, equivalently,} \quad \tilde{\mathbf{x}} \perp (\mathbf{y} - \bar{\mathbf{y}})$$

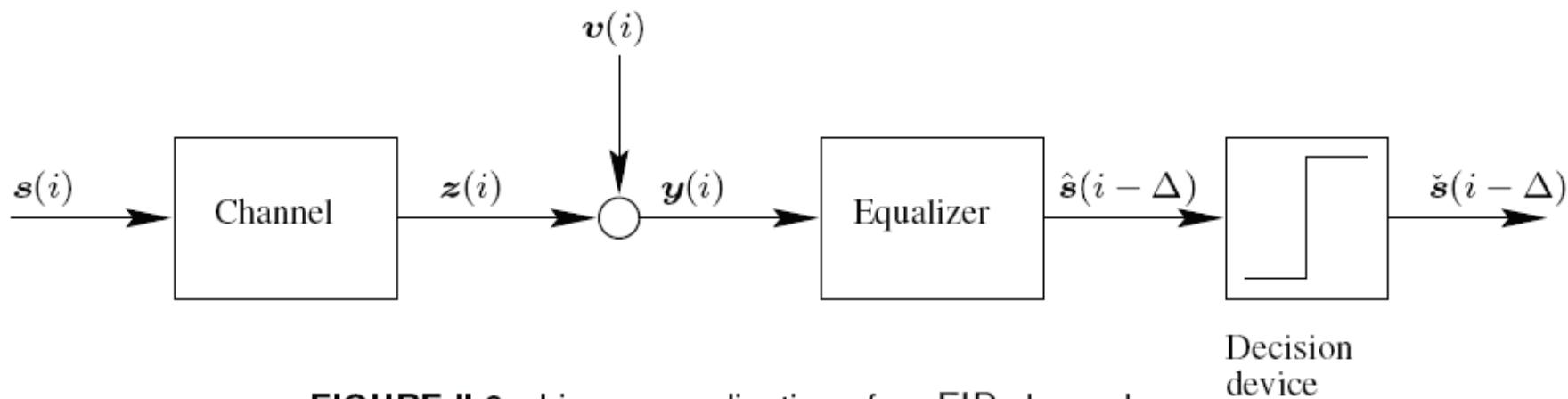
where  $\tilde{\mathbf{x}} = (\mathbf{x} - \bar{\mathbf{x}}) - (\hat{\mathbf{x}} - \bar{\mathbf{x}}) = \mathbf{x} - \hat{\mathbf{x}}$ . Moreover, the resulting m.m.s.e. matrix  $\mathsf{E}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*$  is equal to

$$\text{m.m.s.e.} = R_x - K_o R_y K_o^*$$

with  $R_x = \mathsf{E}(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^*$ .

# COMPUTE RPROJECT

**Project II.1 (Linear equalization and decision devices)** Consider the three-tap linear equalizer discussed in Ex. 4.1, and studied further in Prob. II.17 where the optimal equalizer coefficients were determined for several values of  $\Delta$ . The equalizer structure is shown again in Fig. II.6. Symbols  $\{s(i)\}$  are transmitted through an FIR channel and corrupted by additive white noise  $\{v(i)\}$ . The received signal  $\{y(i)\}$  is processed by a linear equalizer to generate estimators  $\{\hat{s}(i - \Delta)\}$ , which are further fed into a decision device, as explained in parts (c) and (d) below. The purpose of this device is to map each  $\hat{s}(i - \Delta)$  to the closest symbol in the constellation; the result is denoted by  $\check{s}(i - \Delta)$ . Choose initially  $\sigma_v^2 = 0.004$  so that, according to part (f) of Prob. II.17, the signal-to-noise ratio at the input of the equalizer is approximately 25 dB (the dB value is obtained by computing  $10 \log(\cdot)$ ). The purpose of this project is to examine the performance and operation of this three-tap equalizer.



**FIGURE II.6** Linear equalization of an FIR channel.

# COMPUTER PROJECT

- (a) Write a program that evaluates the optimal equalizer coefficients for  $\Delta = 0, 1, 2, 3$  and for arbitrary noise variance  $\sigma_v^2$ . Use the program to feed 2000 BPSK symbols  $\{s(i)\}$  into the channel  $1 + 0.5z^{-1}$  and generate the corresponding equalizer outputs  $\{\hat{s}(i - \Delta)\}$ , for  $\Delta = 0, 1, 2, 3$ . Plot the scatter diagrams of  $\{s(i), y(i), \hat{s}(i - \Delta)\}$ . For each  $\Delta$ , 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 those obtained from theory (cf. Prob. II.17). Plot also the scatter diagrams of  $\{y(i), \hat{s}(i - \Delta)\}$  for  $\Delta = 0$  when  $\sigma_v^2 = 0.05$ , which corresponds to SNR= 14 dB. Compare with the scatter diagrams at 25 dB.

- (b) How would the equalizer coefficients  $\{\alpha(0), \alpha(1), \alpha(2)\}$  change if the input signal  $\{s(i)\}$  were instead chosen uniformly from a QPSK constellation, i.e.,  $s(i) \in \{\sqrt{2}(\pm 1 \pm j)/2\}$ ? Repeat the simulations of part (a) for QPSK data with  $\sigma_v^2 = 0.004$  (and also  $\sigma_v^2 = 0.05$ ). Now, however, the white noise sequence  $\{v(i)\}$  needs to be complex-valued. In order to generate such a sequence with variance  $\sigma_v^2$ , simply generate two separate real-valued white noise sequences  $\{a(i), b(i)\}$  with variance  $\sigma_v^2/2$  each. Then set  $v(i) = a(i) + jb(i)$ , where  $j = \sqrt{-1}$ .

# COMPUTER PROJECT

- (c) For all cases in part (a), assume now that the output of the equalizer is applied to the nonlinear decision device:

$$\text{sign}[\hat{s}(i - \Delta)] = \begin{cases} +1 & \text{if } \hat{s}(i - \Delta) \geq 0 \\ -1 & \text{if } \hat{s}(i - \Delta) < 0 \end{cases}$$

Determine the number of erroneous decisions for each  $\Delta$ . Which  $\Delta$  results in the smallest number of errors?

- (d) For all cases in part (b), assume that the output of the equalizer is applied to the nonlinear decision device:

$$\text{dec}[\hat{s}(i - \Delta)] = \frac{\sqrt{2}}{2} \{ \text{sign}[ \text{Re}(\hat{s}(i - \Delta)) ] + j \text{sign}[ \text{Im}(\hat{s}(i - \Delta)) ] \}$$

which assigns  $\hat{s}(i - \Delta)$  to the closest symbol in the QPSK constellation. Determine the number of erroneous decisions for each  $\Delta$ . Which  $\Delta$  results in the smallest number of errors?

- (e) Fix  $\Delta = 1$  and vary the value of  $\sigma_v^2$  between 0.004 and 0.2 in increments of 0.001, so that the SNR is varied between 8 and 25 dB. Write a program that generates a plot showing how the symbol error rate (SER) varies with SNR. [The SER is defined as the number of erroneous symbols relative to the total number of transmissions.]

# COMPUTER PROJECT

- (f) Repeat part (e) for QPSK data, by repeating the simulations of parts (b) and (d).
- (g) In order to visualize the improvement that is provided by the presence of the linear equalizer, assume that the received signal  $\{\mathbf{y}(i)\}$  is applied directly to the nonlinear decision device (i.e., let us remove the equalizer). The output of the decision device is then taken as  $\hat{s}(i - \Delta)$ . For both cases of BPSK and QPSK data, generate plots that show how the symbol error rate varies with the SNR. Compare these plots with the ones obtained in parts (e) and (f) for small and large values of  $\sigma_v^2$ .

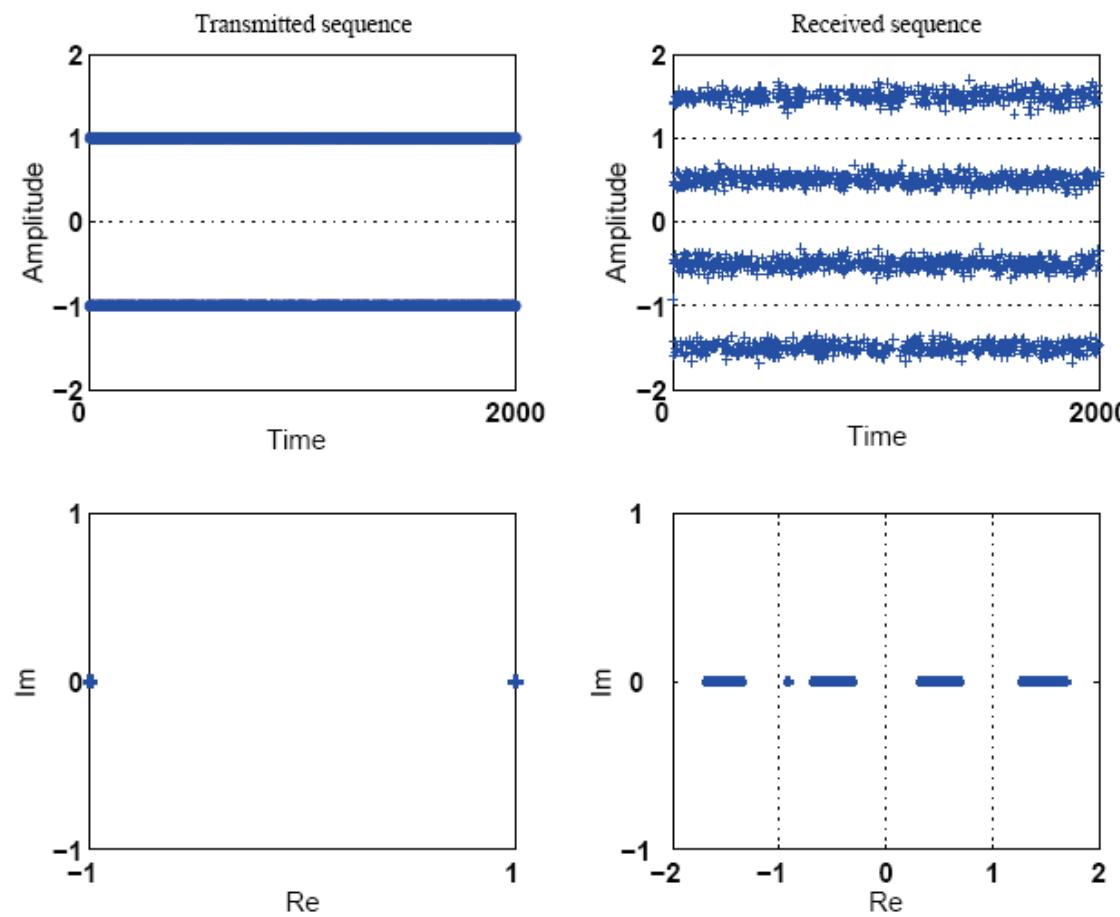
# COMPUTER PROJECT

**Project II.1 (Linear equalization and decision devices)** The programs that solve this project are the following.

1. partA.m This program solves parts (a) and (c), which deal with BPSK data. The program generates 5 plots that show the time and scattering diagrams of the transmitted sequence  $\{s(i)\}$ , the received sequence  $\{y(i)\}$  (which is the input to the equalizer), and the equalizer output  $\{\hat{s}(i - \Delta)\}$  for several values of  $\Delta$ . A final plot shows the values of the estimated m.m.s.e. and the number of erroneous decisions as a function of  $\Delta$ . The program allows the user to change  $\sigma_v^2$ . Typical outputs of this program are shown in the sequel.

Figure 1 shows four graphs displayed in a  $2 \times 2$  matrix grid. The graphs on the left column pertain to the transmitted sequence  $s(i)$ , while the graphs on the right column pertain to the received sequence  $y(i)$ . The graphs on the bottom row are scatter diagrams; they show where the symbols  $\{s(i), y(i)\}$  are located in the complex plane for all 2000 transmissions. We thus see that the  $s(i)$  are concentrated at  $\pm 1$ , as expected, while the  $y(i)$  appear to be concentrated around the four values  $\{-1.5, -0.5, 0.5, 1.5\}$ . The graphs on the top row are time diagrams; they show what values the symbols  $\{s(i), y(i)\}$  assume over the 2000 transmissions. Clearly, the  $s(i)$  are either  $\pm 1$ , while the  $y(i)$  assume values around  $\{\pm 0.5, \pm 1.5\}$ .

# COMPUTER PROJECT



**Figure II.1.** The plots show the time (*top row*) and scatter (*bottom row*) diagrams of the transmitted and received sequences  $\{\mathbf{s}(i), \mathbf{y}(i)\}$  for BPSK transmissions. Observe how the  $\mathbf{s}(i)$  are concentrated at  $\pm 1$  while the  $\mathbf{y}(i)$  are concentrated around  $\{\pm 0.5, \pm 1.5\}$ .

# COMPUTER PROJECT

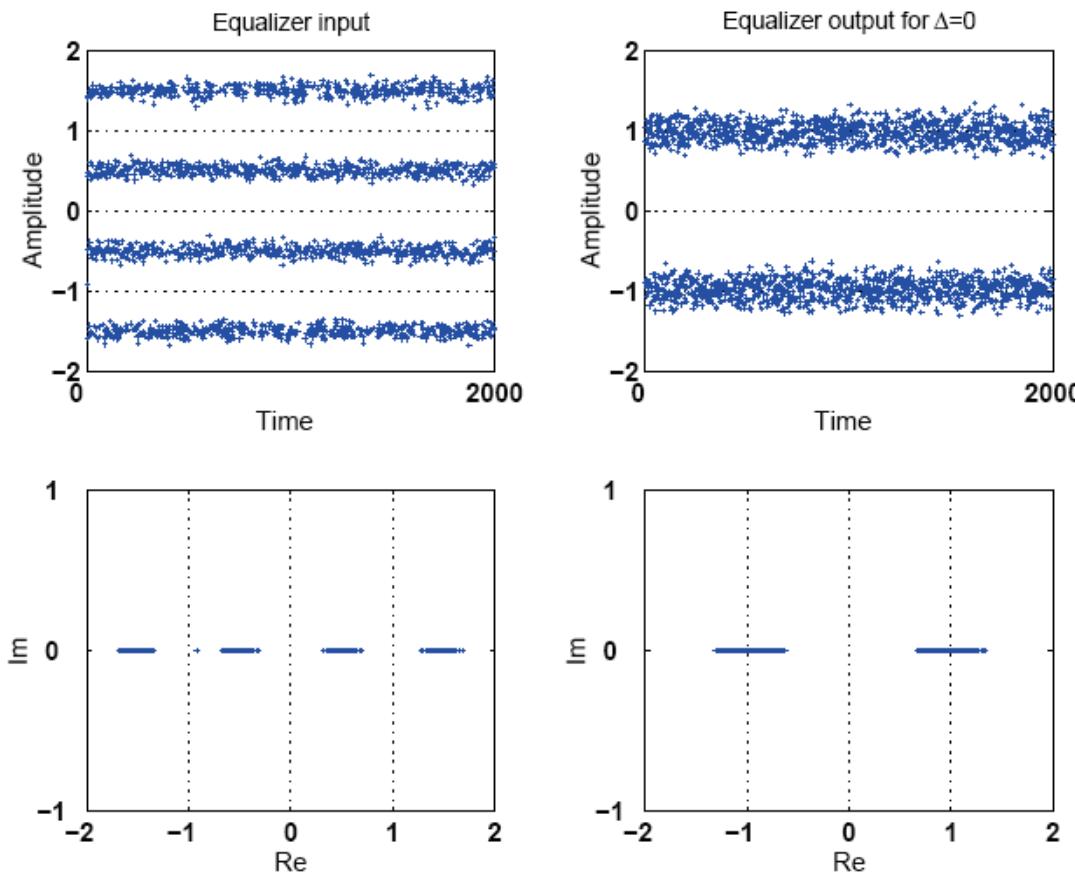
Figure 2 again shows four graphs displayed in a  $2 \times 2$  matrix grid. The graphs on the left column pertain to the received sequence  $\mathbf{y}(i)$ , while the graphs on the right column pertain to the output of the equalizer,  $\hat{\mathbf{s}}(i - \Delta)$ , for  $\Delta = 0$ , i.e.,  $\hat{\mathbf{s}}(i)$ . The graphs on the bottom row are the scatter diagrams while the graphs on the top row are the time diagrams. Observe how the symbols at the output of the equalizer are concentrated around  $\pm 1$ .

Figure 3 is similar to Fig. 2, except that it relates to  $\Delta = 1$ . Comparing Figs. 2 and 3, observe how the scatter diagram of the output of the equalizer is more spread in the latter case. Figure 4 is also similar to Fig. 2, except that it now relates to  $\Delta = 3$ . Observe how the performance of the equalizer is poor in this case.

Figure 5 shows two plots: one pertains to the m.m.s.e as a function of  $\Delta$ , while the other pertains to the number of erroneous decisions as a function of  $\Delta$ . It is seen that although the choice  $\Delta = 0$  leads to the lowest m.m.s.e. at the assumed SNR level of 25 dB, all three choices  $\Delta = 0, 1, 2$  lead to zero errors.

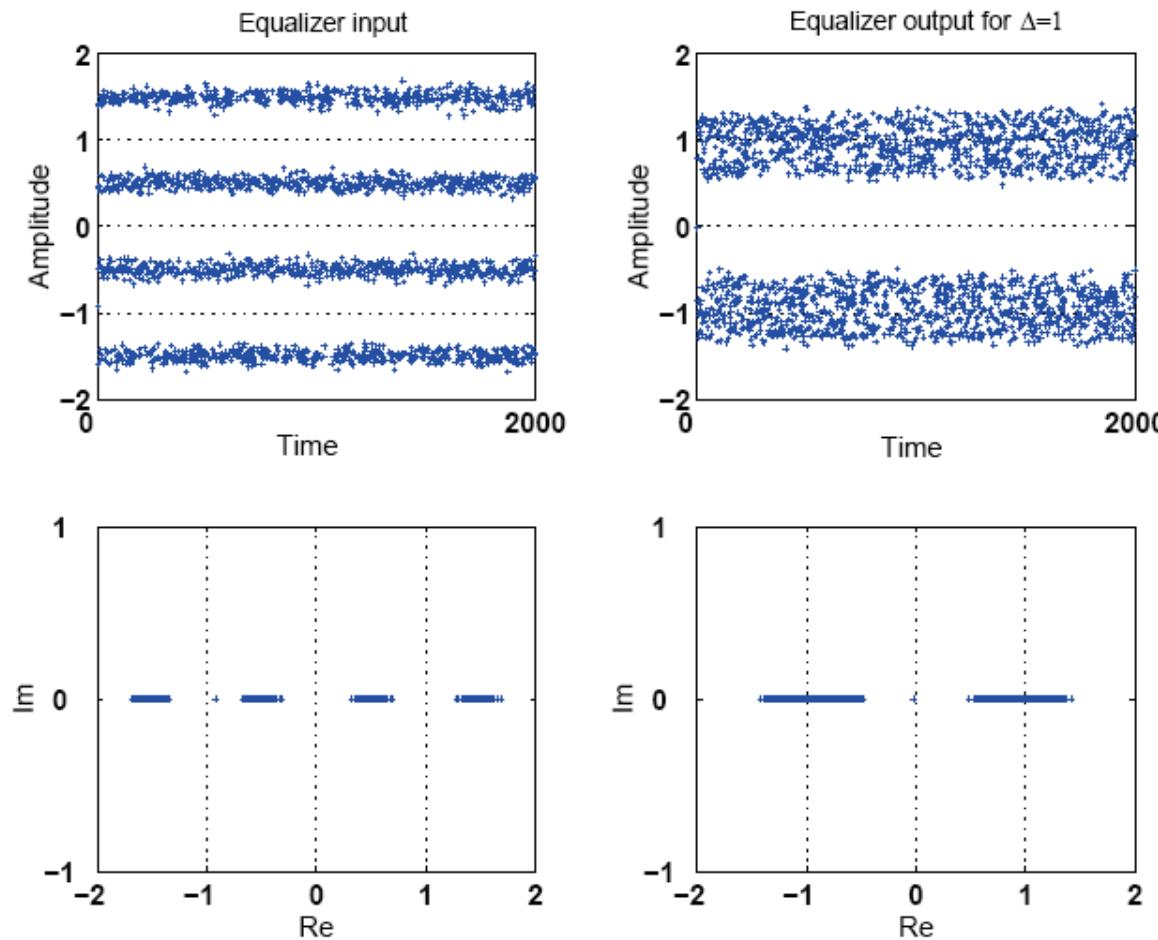
Figure 6 again shows the time and scatter diagrams of the equalizer input (*left*) and output (*right*) using  $\Delta = 0$  and  $\sigma_v^2 = 0.05$ , i.e., SNR= 14 dB. Observe how the equalizer performance is more challenging at this SNR level. Still, the number of erroneous decisions came out as zero in this run.

# COMPUTER PROJECT



**Figure II.2.** The plots show the time (*top row*) and scatter (*bottom row*) diagrams of the input and output of the equalizer, i.e.,  $\{\mathbf{y}(i), \hat{\mathbf{s}}(i)\}$ , for  $\Delta = 0$  and BPSK transmissions. Observe how the equalizer output is concentrated around  $\pm 1$ .

# COMPUTER PROJECT



**Figure II.3.** The plots show the time (*top row*) and scatter (*bottom row*) diagrams of the input and output of the equalizer, i.e.,  $\{\mathbf{y}(i), \hat{\mathbf{s}}(i)\}$ , for  $\Delta = 1$  and BPSK transmissions. Observe how the equalizer output is again concentrated around  $\pm 1$ .

# COMPUTER PROJECT

2. partB.m This program solves parts (b) and (d), which deal with QPSK data. The program generates 3 plots that show the scatter diagrams of the transmitted sequence  $\{\mathbf{s}(i)\}$ , the received sequence  $\{\mathbf{y}(i)\}$ , and the equalizer output  $\{\hat{\mathbf{s}}(i - \Delta)\}$  for several values of  $\Delta$ . A final plot shows the values of the estimated m.m.s.e. and the number of erroneous decisions as a function of  $\Delta$ . The program allows the designer to change the value of the noise variance  $\sigma_v^2$ . Typical outputs of this program are shown in the sequel.

Figure 7 shows the scatter diagrams of the transmitted (*left*) and received (*right*) sequences,  $\{\mathbf{s}(i), \mathbf{y}(i)\}$ .

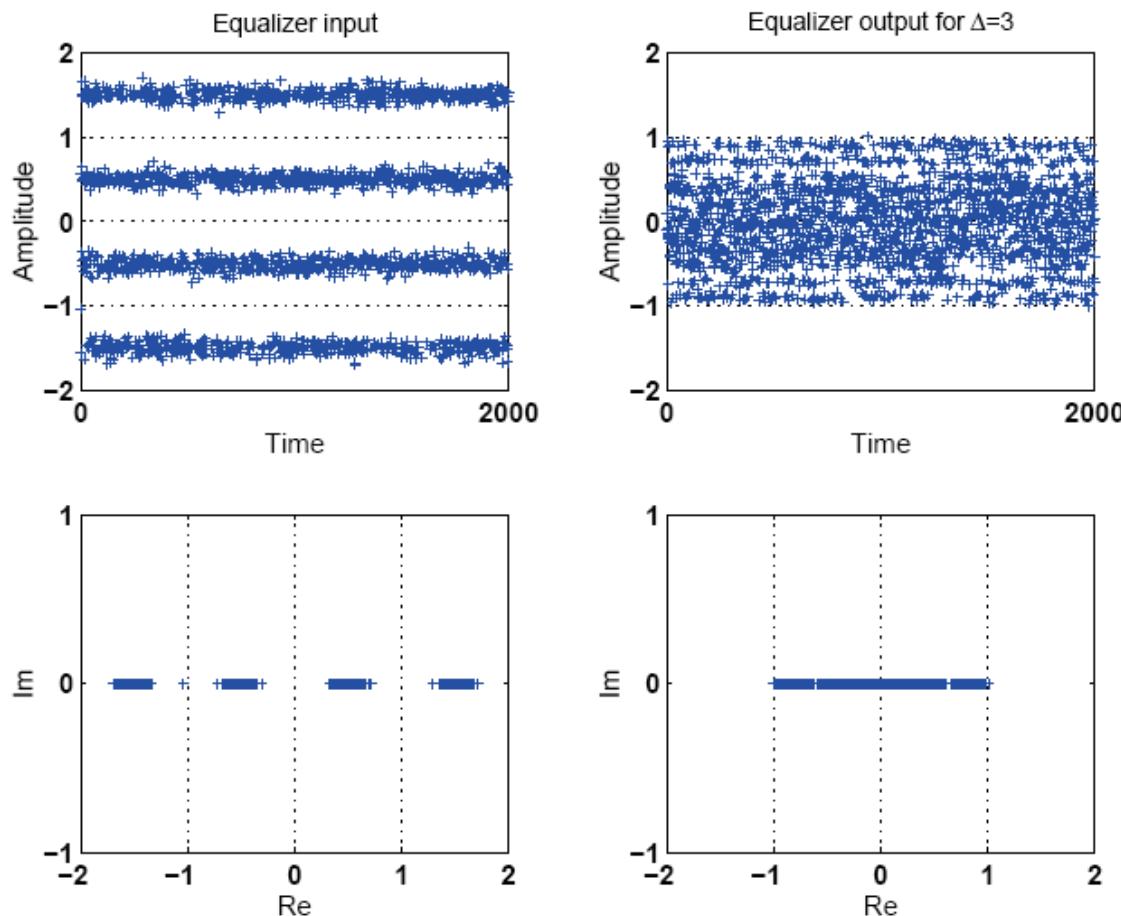
Observe how the  $\mathbf{s}(\cdot)$  are concentrated at  $\sqrt{2}(\pm 1 \pm j)/2$ , as expected, while the  $\mathbf{y}(i)$  appear to be concentrated around 16 values.

Figure 8 shows the scatter diagrams of the equalizer input (*left*) and output (*right*) sequences for the cases  $\Delta = 0$  and  $\Delta = 1$ . Observe how the equalizer output is now concentrated around the QPSK constellation symbols.

Figure 9 repeats the same scatter diagrams for the cases  $\Delta = 2$  and  $\Delta = 3$ . Observe how the equalizer fails completely for  $\Delta = 3$ .

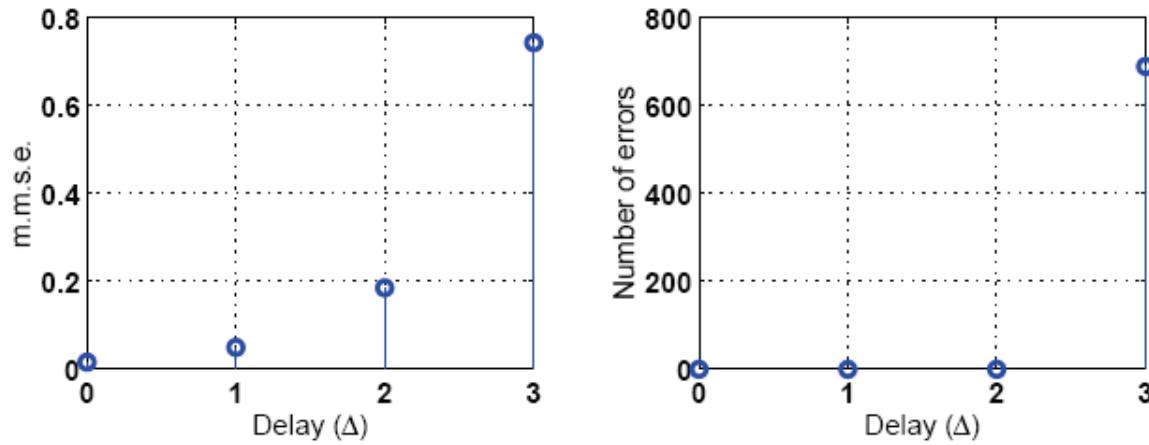
The plot showing the m.m.s.e. and the number of erroneous decisions as a function of  $\Delta$  is similar to Fig. 5. Figure 10, on the other hand, shows the time and scatter diagrams of the equalizer input (*left*) and output (*right*) using  $\Delta = 0$  and  $\Delta = 1$  for  $\sigma_v^2 = 0.05$ , i.e., SNR= 14 dB. The equalizer performance is more challenging at this SNR level. Still, the number of erroneous decisions came out as zero in this run.

# COMPUTER PROJECT



**Figure II.4.** The plots show the time (*top row*) and scatter (*bottom row*) diagrams of the input and output of the equalizer i.e.,  $\{\mathbf{y}(i), \hat{\mathbf{s}}(i)\}$ , for  $\Delta = 3$  and BPSK transmissions. Observe how the performance of the equalizer is poor in this case.

# COMPUTER PROJECT

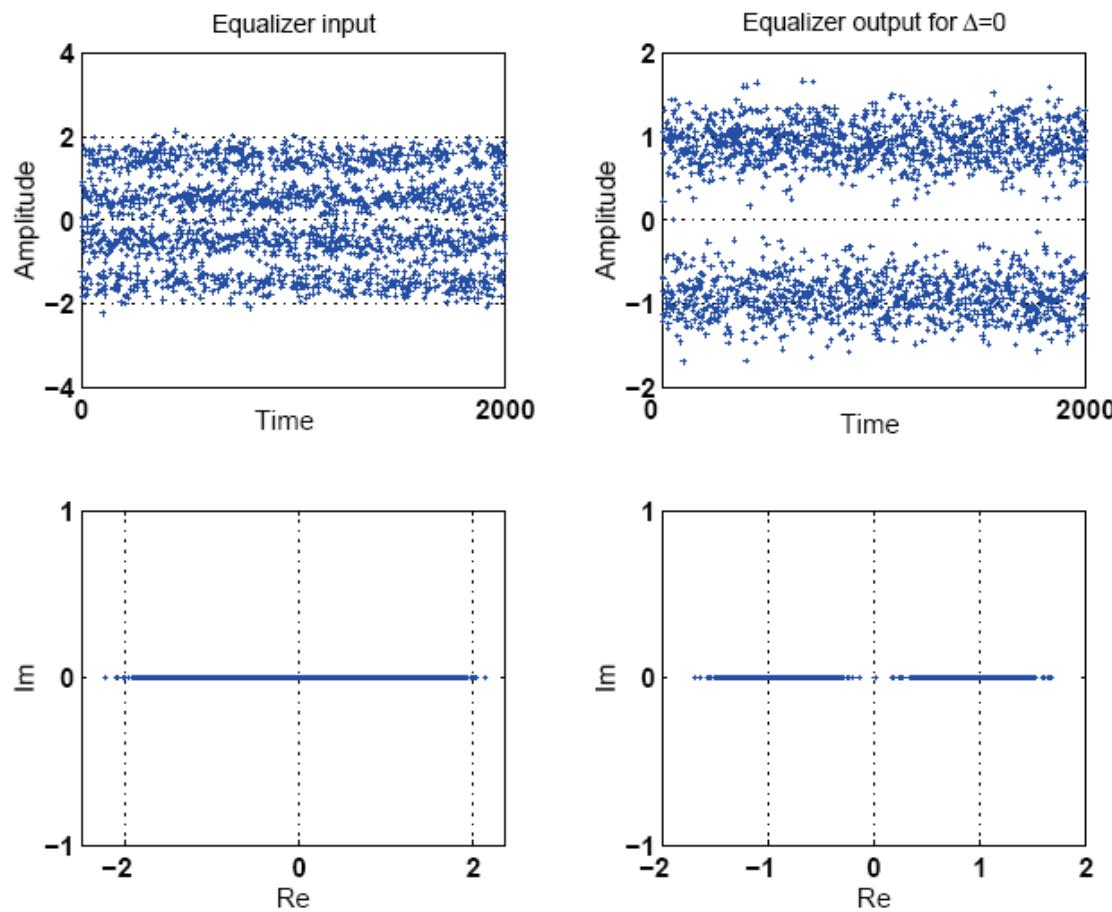


**Figure II.5.** The plots show the m.m.s.e (left) and the number of erroneous decisions (right) as a function of  $\Delta$  for BPSK transmissions.

# COMPUTER PROJECT

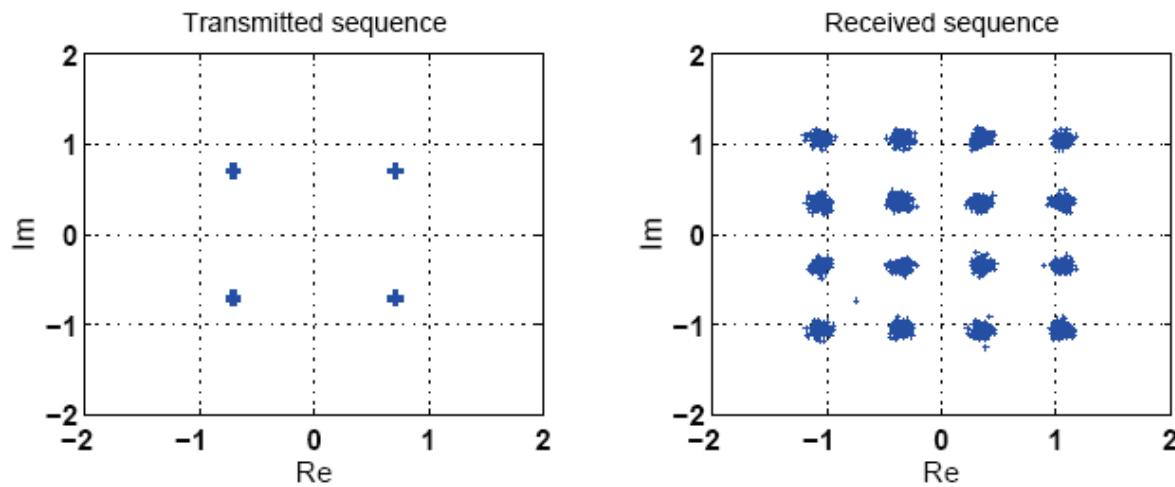
3. partE.m This program solves parts (e) and (g) for BPSK data and for various noise levels. It generates a plot of the symbol error rate, measured as the ratio of the number of erroneous decisions to the total number of transmissions (adjusted to 20000) as a function of the SNR at the input of the equalizer. The plot shows the SER curves with and without equalization. A typical output is shown in the left plot of Fig. 11.
4. partF.m This program solves parts (f) and (g) for QPSK data and for various noise levels. It also generates a plot of the symbol error rate. The plot shows the SER curves with and without equalization. A typical output is shown in the right plot of Fig. 11.

# COMPUTER PROJECT



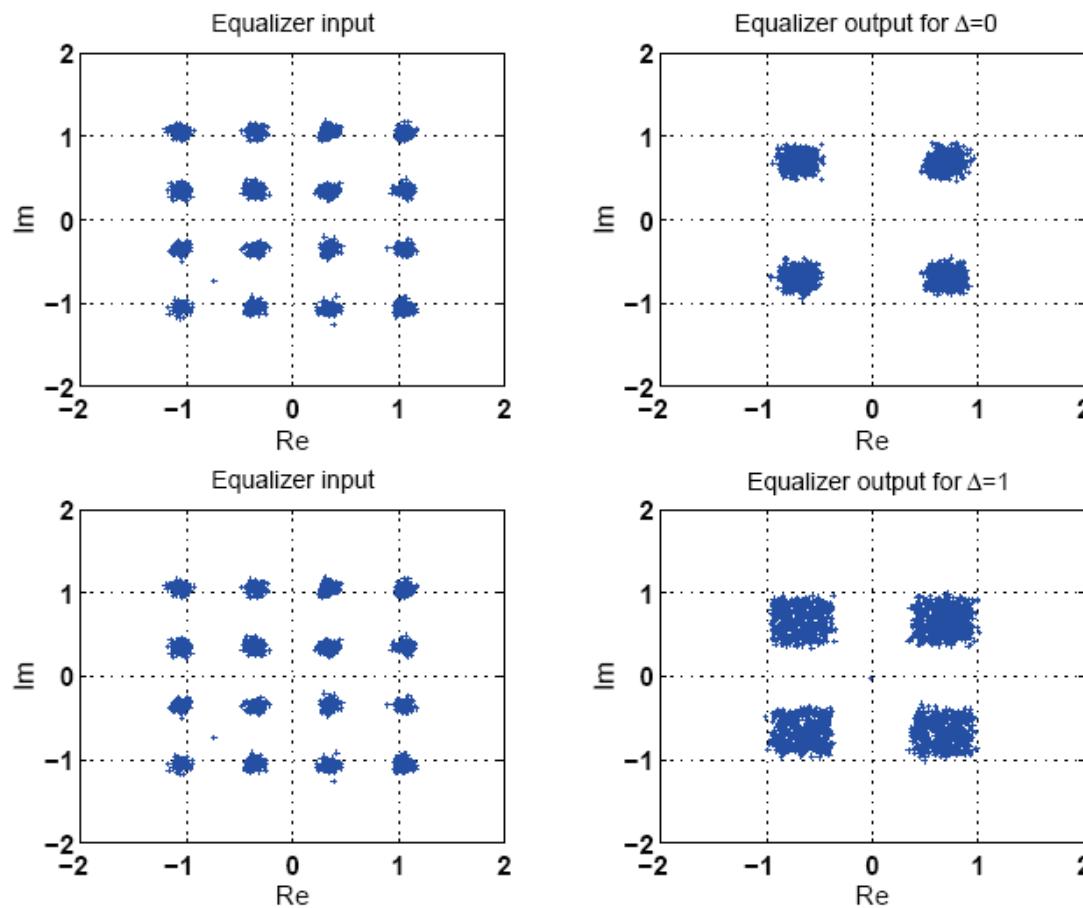
**Figure II.6.** The plots show the time (*top row*) and scatter (*bottom row*) diagrams of the input and output of the equalizer, i.e.,  $\{\mathbf{y}(i), \hat{\mathbf{s}}(i)\}$ , for  $\Delta = 0$ , BPSK transmissions, and  $\sigma_v^2 = 0.05$  (i.e., SNR= 14 dB).

# COMPUTER PROJECT



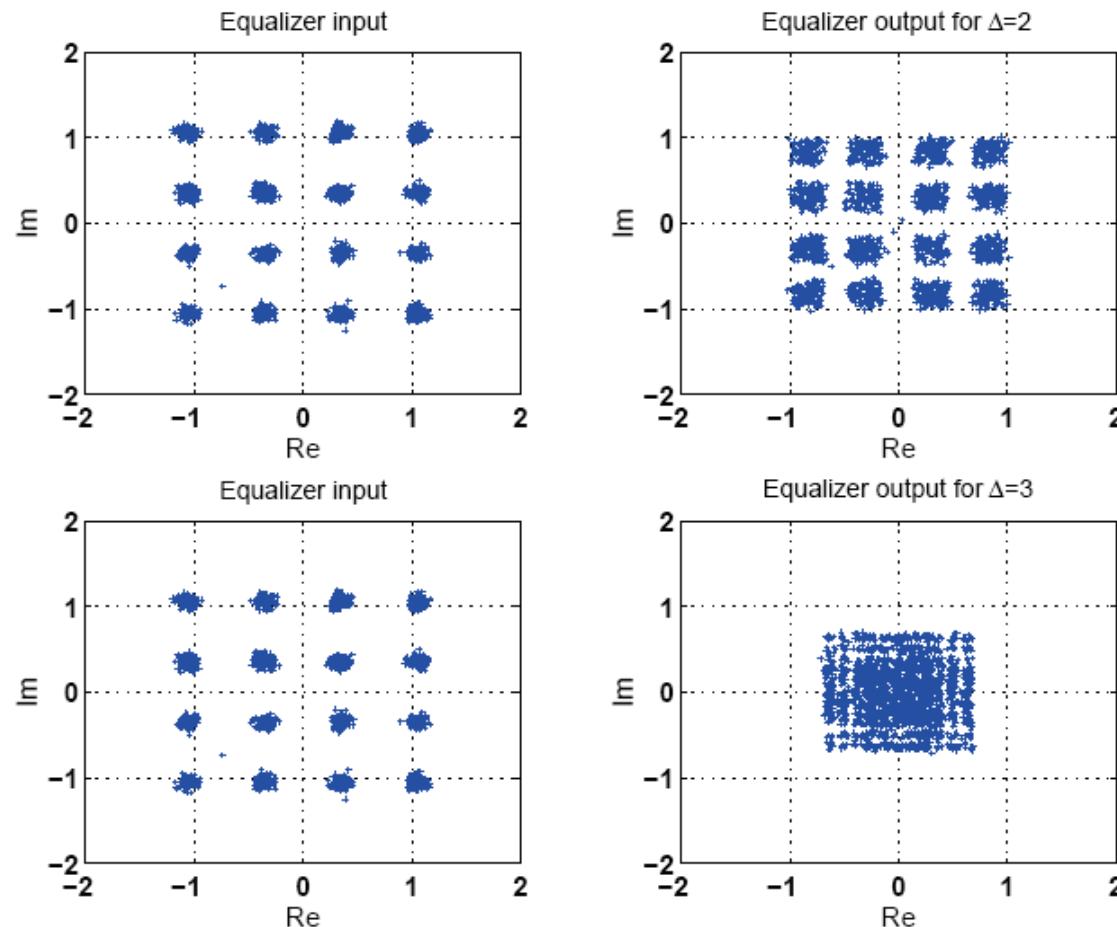
**Figure II.7.** The plots show the scatter diagrams of the transmitted (*left*) and received (*right*) sequences  $\{s(i), y(i)\}$  for QPSK transmissions.

# COMPUTER PROJECT



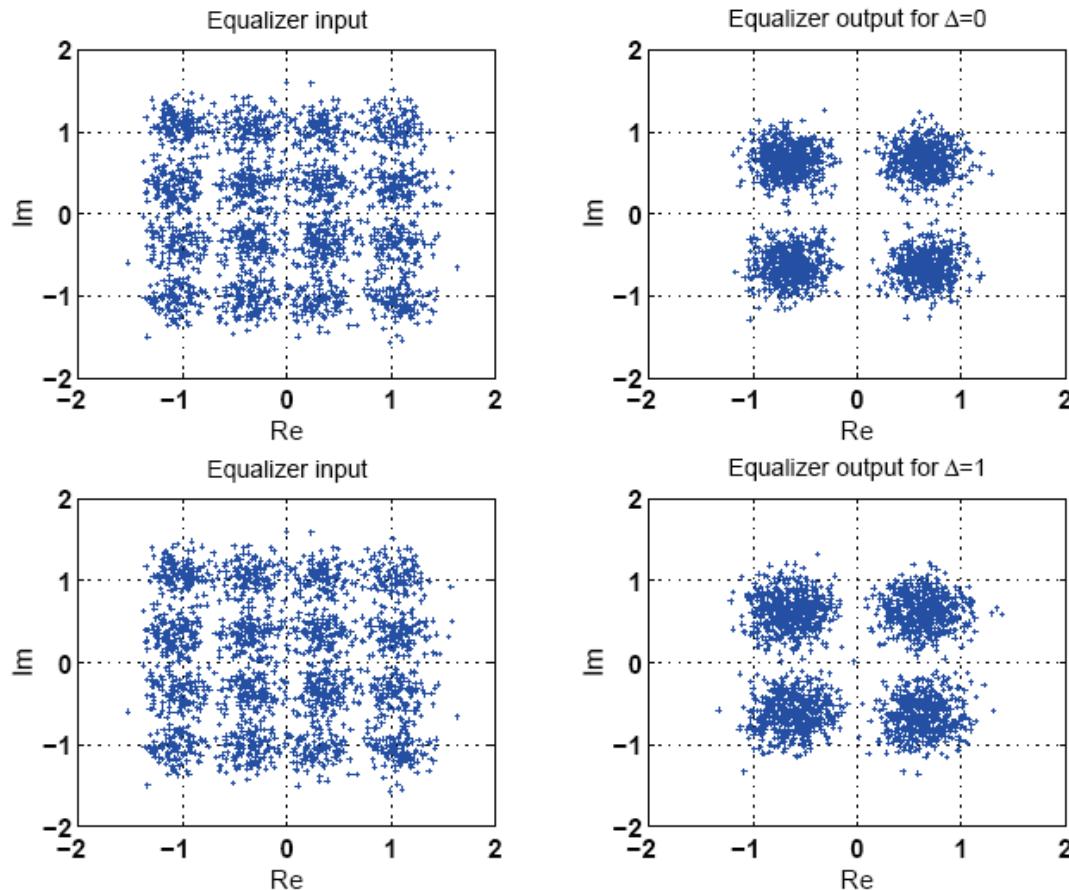
**Figure II.8.** The plots show the scatter diagrams of the equalizer input (*left*) and output (*right*) for the cases  $\Delta = 0$  and  $\Delta = 1$  for QPSK transmissions.

# COMPUTER PROJECT



**Figure II.9.** The plots show the scatter diagrams of the equalizer input (*left*) and output (*right*) for the cases  $\Delta = 2$  and  $\Delta = 3$  for QPSK transmissions.

# COMPUTER PROJECT



**Figure II.10.** The plots show the scatter diagrams of the equalizer input (*left*) and output (*right*) for QPSK data with  $\Delta = 0$  (*top row*) and  $\Delta = 1$  (*bottom row*) and at SNR= 14 dB.

# COMPUTER PROJECT

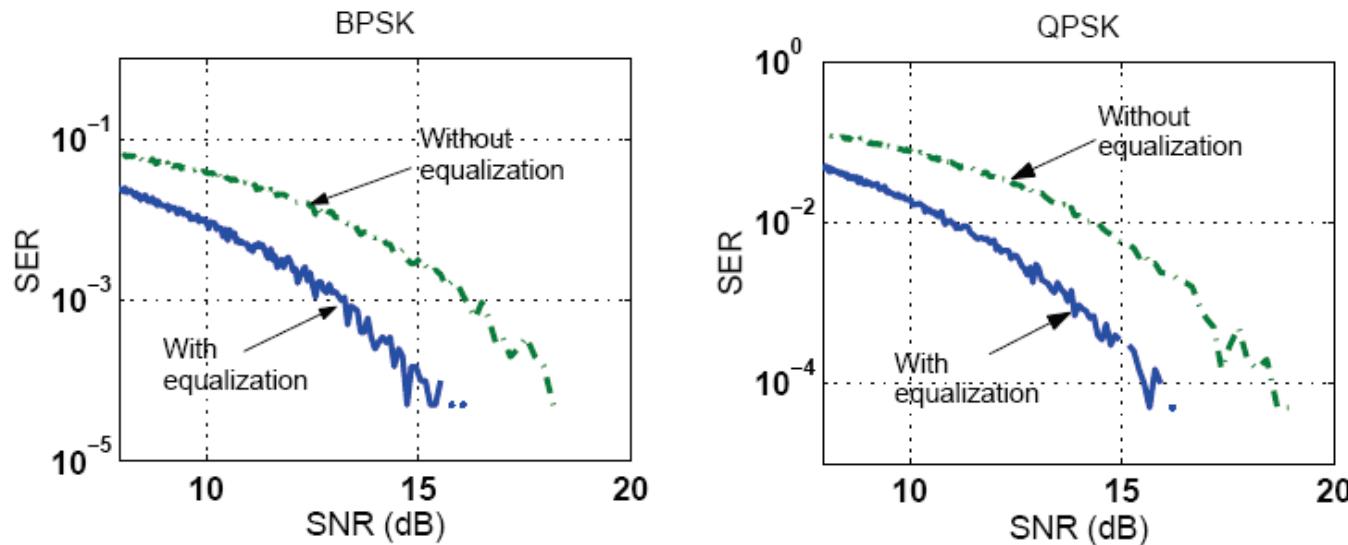


Figure II.11. Plots of the SER over 20000 transmitted BPSK symbols (*left*) and QPSK symbols (*right*), with and without equalization, for SNR values between 8 and 20 dB and using  $\Delta = 1$ .