CHAPTER 8

Estimation with Minimum Mean Square Error

INTRODUCTION

A recurring theme in this text and in much of communication, control and signal processing is that of making systematic estimates, predictions or decisions about some set of quantities, based on information obtained from measurements of other quantities. This process is commonly referred to as inference. Typically, inferring the desired information from the measurements involves incorporating models that represent our prior knowledge or beliefs about how the measurements relate to the quantities of interest.

Inference about continuous random variables and ultimately about random processes is the topic of this chapter and several that follow. One key step is the introduction of an error criterion that measures, in a probabilistic sense, the error between the desired quantity and our estimate of it. Throughout our discussion in this and the related subsequent chapters, we focus primarily on choosing our estimate to minimize the expected or mean value of the square of the error, referred to as a minimum mean-square-error (MMSE) criterion. In Section 8.1 we consider the MMSE estimate without imposing any constraint on the form that the estimator takes. In Section 8.3 we restrict the estimate to be a linear combination of the measurements, a form of estimation that we refer to as linear minimum mean-square-error (LMMSE) estimation.

Later in the text we turn from inference problems for continuous random variables to inference problems for discrete random quantities, which may be numerically specified or may be non-numerical. In the latter case especially, the various possible outcomes associated with the random quantity are often termed hypotheses, and the inference task in this setting is then referred to as hypothesis testing, i.e., the task of deciding which hypothesis applies, given measurements or observations. The MMSE criterion may not be meaningful in such hypothesis testing problems, but we can for instance aim to minimize the probability of an incorrect inference regarding which hypothesis actually applies.

8.1 ESTIMATION OF A CONTINUOUS RANDOM VARIABLE

To begin the discussion, let us assume that we are interested in a random variable Y and we would like to estimate its value, knowing only its probability density function. We will then broaden the discussion to estimation when we have a measurement or observation of another random variable X, together with the joint probability density function of X and Y.

Based only on knowledge of the PDF of Y, we wish to obtain an estimate of Y — which we denote as $\widehat{y}$ — so as to minimize the mean square error between the actual outcome of the experiment and our estimate $\widehat{y}$. Specifically, we choose $\widehat{y}$ to minimize

$$E[(Y - \widehat{y})^2] = \int (y - \widehat{y})^2 f_Y(y)\, dy \; . \tag{8.1}$$

Differentiating (8.1) with respect to $\widehat{y}$ and equating the result to zero, we obtain

$$-2 \int (y - \widehat{y}) f_Y(y)\, dy = 0 \tag{8.2}$$

or

$$\int \widehat{y} f_Y(y)\, dy = \int y f_Y(y)\, dy \tag{8.3}$$

from which

$$\widehat{y} = E[Y] \; . \tag{8.4}$$

The second derivative of $E[(Y - \widehat{y})^2]$ with respect to $\widehat{y}$ is

$$2 \int f_Y(y)\, dy = 2 \; , \tag{8.5}$$

which is positive, so (8.4) does indeed define the minimizing value of $\widehat{y}$. Hence the MMSE estimate of Y in this case is simply its mean value, $E[Y]$.

The associated error — the actual MMSE — is found by evaluating the expression in (8.1) with $\widehat{y} = E[Y]$. We conclude that the MMSE is just the variance of Y, namely σ_Y^2:

$$\min E[(Y - \widehat{y})^2] = E[(Y - E[Y])^2] = \sigma_Y^2 \; . \tag{8.6}$$

In a similar manner, it is possible to show that the median of Y, which has half the probability mass of Y below it and the other half above, is the value of $\widehat{y}$ that minimizes the mean absolute deviation, $E[\,|Y - \widehat{y}|\,]$. Also, the mode of Y, which is the value of y at which the PDF $f_Y(y)$ is largest, turns out to minimize the expected value of an all-or-none cost function, i.e., a cost that is unity when the error is outside of a vanishingly small tolerance band, and is zero within the band. We will not be pursuing these alternative error metrics further, but it is important to be aware that our choice of mean square error, while convenient, is only one of many possible error metrics.

The insights from the simple problem leading to (8.4) and (8.6) carry over directly to the case in which we have additional information in the form of the measured or observed value x of a random variable X that is related somehow to Y. The only change from the previous discussion is that, given the additional measurement, we work with the conditional or *a posteriori* density $f_{Y|X}(y|x)$, rather than the unconditioned density $f_Y(y)$, and now our aim is to minimize

$$E[\{Y - \widehat{y}(x)\}^2 | X = x] = \int \{y - \widehat{y}(x)\}^2 f_{Y|X}(y|x)\, dy \; . \qquad (8.7)$$

We have introduced the notation $\widehat{y}(x)$ for our estimate to show that in general it will depend on the specific value x. Exactly the same calculations as in the case of no measurements then show that

$$\widehat{y}(x) = E[Y|X = x] \; , \qquad (8.8)$$

the conditional expectation of Y, given $X = x$. The associated MMSE is the variance $\sigma^2_{Y|X}$ of the conditional density $f_{Y|X}(y|x)$, i.e., the MMSE is the conditional variance. Thus, the only change from the case of no measurements is that we now condition on the obtained measurement.

Going a further step, if we have multiple measurements, say $X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L$, then we work with the *a posteriori* density

$$f_{Y\,|\,X_1, X_2, \cdots, X_L}(y \,|\, x_1, x_2, \cdots, x_L) \; . \qquad (8.9)$$

Apart from this modification, there is no change in the structure of the solutions. Thus, without further calculation, we can state the following:

The MMSE estimate of Y,
given $X_1 = x_1, \cdots, X_L = x_L$,
is the **conditional expectation** of Y:

$$\widehat{y}(x_1, \ldots, x_L) = E[Y \,|\, X_1 = x_1, \cdots, X_L = x_L] \qquad (8.10)$$

For notational convenience, we can arrange the measured random variables into a column vector $\mathbf{X}$, and the corresponding measurements into the column vector $\mathbf{x}$. The dependence of the MMSE estimate on the measurements can now be indicated by the notation $\widehat{y}(\mathbf{x})$, with

$$\widehat{y}(\mathbf{x}) = \int_{-\infty}^{\infty} y\, f_{Y|\mathbf{X}}(y \,|\, \mathbf{X} = \mathbf{x})\, dy = E[\, Y \,|\, \mathbf{X} = \mathbf{x}\,] \; . \qquad (8.11)$$

The minimum mean square error (or MMSE) for the given value of $\mathbf{X}$ is again the conditional variance, i.e., the variance $\sigma^2_{Y|\mathbf{X}}$ of the conditional density $f_{Y|\mathbf{X}}(y \,|\, \mathbf{x})$.

EXAMPLE 8.1 MMSE Estimate for Discrete Random Variables

A discrete-time discrete-amplitude sequence $s[n]$ is stored on a noisy medium. The retrieved sequence is $r[n]$. Suppose at some particular time instant $n = n_0$ we have

$s[n_0]$ and $r[n_0]$ modeled as random variables, which we shall simply denote by S and R respectively. From prior measurements, we have determined that S and R have the joint probability mass function (PMF) shown in Figure 8.1.

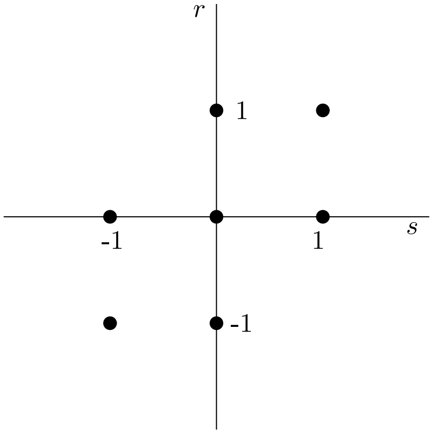

FIGURE 8.1 Joint PMF of S and R.

Based on receiving the value $R = 1$, we would like to make an MMSE estimate $\widehat{s}$ of S. From (8.10), $\widehat{s} = E(S|R = 1)$, which can be determined from the conditional PMF $P_{S|R}(s|R = 1)$, which in turn we can obtain as

$$P_{S|R}(s|R = 1) = \frac{P_{R,S}(R = 1, s)}{P_R(R = 1)} \ . \tag{8.12}$$

From Figure 8.1,

$$P_R(1) = \frac{2}{7} \tag{8.13}$$

and

$$P_{R,S}(1, s) = \begin{cases} 0 & s = -1 \\ 1/7 & s = 0 \\ 1/7 & s = +1 \end{cases}$$

Consequently,

$$P_{S|R}(s|R = 1) = \begin{cases} 1/2 & s = 0 \\ 1/2 & s = +1 \end{cases}$$

Thus, the MMSE estimate is $\widehat{s} = \frac{1}{2}$. Note that although this estimate minimizes the mean square error, we have not constrained it to take account of the fact that S can only have the discrete values of $+1$, 0 or -1. In a later chapter we will return to this example and consider it from the perspective of hypothesis testing, i.e., determining which of the three known possible values will result in minimizing a suitable error criterion.

EXAMPLE 8.2 MMSE Estimate of Signal in Additive Noise

A discrete-time sequence $s[n]$ is transmitted over a noisy channel and retrieved. The received sequence $r[n]$ is modeled as $r[n] = s[n] + w[n]$ where $w[n]$ represents the noise. At a particular time instant $n = n_0$, suppose $r[n_0]$, $s[n_0]$ and $w[n_0]$ are random variables, which we denote as R, S and W respectively. We assume that S and W are independent, that W is uniformly distributed between $+\frac{1}{2}$ and $-\frac{1}{2}$, and S is uniformly distributed between -1 and $+1$. The specific received value is $R = \frac{1}{4}$, and we want the MMSE estimate $\widehat{s}$ for S. From (8.10),

$$\widehat{s} = E(S|R = \frac{1}{4}) \tag{8.14}$$

which can be determined from $f_{S|R}(s|R = \frac{1}{4})$:

$$f_{S|R}(s|R = \frac{1}{4}) = \frac{f_{R|S}(\frac{1}{4}|s)f_S(s)}{f_R(\frac{1}{4})} \ . \tag{8.15}$$

We evaluate separately the numerator and denominator terms in (8.15). The PDF $f_{R|S}(r|s)$ is identical in shape to the PDF of W, but with the mean shifted to s, as indicated in Figure 8.2 below. Consequently, $f_{R|S}(\frac{1}{4}|s)$ is as shown in Figure 8.3, and $f_{R|S}(\frac{1}{4}|s)f_S(s)$ is shown in Figure 8.4.

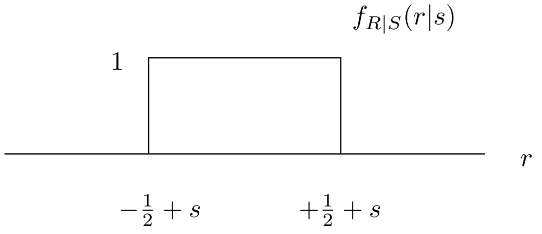

FIGURE 8.2 Conditional PDF of R given S, $f_{R|S}(r|s)$.

To obtain $f_{S|R}(s|R = \frac{1}{4})$ we divide Figure 8.4 by $f_R(\frac{1}{4})$, which can easily be obtained by evaluating the convolution of the PDF's of S and W at the argument $\frac{1}{4}$. More simply, since $f_{S|R}(s|R = \frac{1}{4})$ must have total area of unity and it is the same as Figure 8.4 but scaled by $f_R(\frac{1}{4})$, we can easily obtain it by just normalizing Figure 8.4 to have an area of 1. The resulting value for $\widehat{s}$ is the mean associated with the PDF $f_{S|R}(s|R = \frac{1}{4})$, which will be

$$\widehat{s} = \frac{1}{4} \ . \tag{8.16}$$

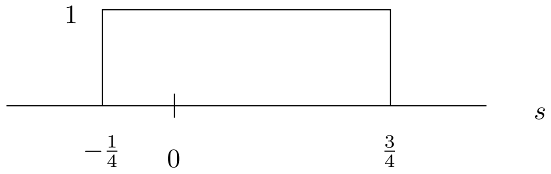

FIGURE 8.3 Plot of $f_{R|S}(\frac{1}{4}|s)$.

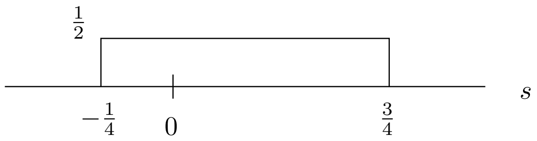

FIGURE 8.4 Plot of $f_{R|S}(\frac{1}{4}|s)f_S(s)$.

The associated MMSE is the variance of this PDF, namely $\frac{1}{12}$.

EXAMPLE 8.3 MMSE Estimate for Bivariate Gaussian Random Variables

Two random variables X and Y are said to have a bivariate Gaussian joint PDF if the joint density of the centered (i.e. zero-mean) and normalized (i.e. unit-variance) random variables

$$V = \frac{X - \mu_X}{\sigma_X} \ , \ W = \frac{Y - \mu_Y}{\sigma_Y} \tag{8.17}$$

is given by

$$f_{V,W}(v, w) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{(v^2 - 2\rho vw + w^2)}{2(1-\rho^2)}\right\} . \tag{8.18}$$

Here μ_X and μ_Y are the means of X and Y respectively, and σ_X, σ_Y are the respective standard deviations of X and Y. The number ρ is the correlation coefficient of X and Y, and is defined by

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \ , \ \text{with} \ \sigma_{XY} = E[XY] - \mu_X\mu_Y \tag{8.19}$$

where σ_{XY} is the covariance of X and Y.

Now, consider $\widehat{y}(x)$, the MMSE estimate of Y given $X = x$, when X and Y are bivariate Gaussian random variables. From (8.10),

$$\widehat{y}(x) = E[Y \mid X = x] \tag{8.20}$$

or, in terms of the zero-mean normalized random variables V and W,

$$\begin{aligned}\widehat{y}(x) &= E\left[(\sigma_Y W + \mu_Y)\,|\,V = \frac{x-\mu_X}{\sigma_X}\right] \\ &= \sigma_Y E\left[W\,|\,V = \frac{x-\mu_X}{\sigma_X}\right] + \mu_Y\ . \qquad (8.21)\end{aligned}$$

It is straightforward to show with some computation that $f_{W|V}(w\,|\,v)$ is Gaussian with mean ρv, and variance $1-\rho^2$, from which it follows that

$$E\left[W\,|\,V = \frac{x-\mu_X}{\sigma_X}\right] = \rho\left[\frac{x-\mu_X}{\sigma_X}\right]\ . \qquad (8.22)$$

Combining (8.21) and (8.22),

$$\begin{aligned}\widehat{y}(x) &= E[\,Y\,|\,\mathbf{X}=x\,] \\ &= \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) \qquad (8.23)\end{aligned}$$

The MMSE estimate in the case of bivariate Gaussian variables has a nice linear (or more correctly, affine, i.e., linear plus a constant) form.

The minimum mean square error is the variance of the conditional PDF $f_{Y|\mathbf{X}}(y|\mathbf{X}=x)$:

$$E[\,(Y-\widehat{y}(x))^2\,|\,\mathbf{X}=x\,] = \sigma_Y^2\,(1-\rho^2)\ . \qquad (8.24)$$

Note that σ_Y^2 is the mean square error in Y in the absence of any additional information. Equation (8.24) shows what the residual mean square error is after we have a measurement of X. It is evident and intuitively reasonable that the larger the magnitude of the correlation coefficient between X and Y, the smaller the residual mean square error.

8.2 FROM ESTIMATES TO AN ESTIMATOR

The MMSE estimate in (8.8) is based on knowing the specific value x that the random variable X takes. While X is a random variable, the specific value x is not, and consequently $\widehat{y}(x)$ is also not a random variable.

As we move forward in the discussion, it is important to draw a distinction between the estimate of a random variable and the procedure by which we form the estimate. This is completely analogous to the distinction between the value of a function at a point and the function itself. We will refer to the procedure or function that produces the estimate as the estimator.

For instance, in Example 8.1 we determined the MMSE estimate of S for the specific value of $R=1$. We could more generally determine an estimate of S for each of the possible values of R, i.e., $-1, 0, \text{and} + 1$. We could then have a tabulation of these results available in advance, so that when we retrieve a specific value of R

we can look up the MMSE estimate. Such a table or more generally a function of R would correspond to what we term the MMSE estimator. The input to the table or estimator would be the specific retrieved value and the output would be the estimate associated with that retrieved value.

We have already introduced the notation $\widehat{y}(x)$ to denote the estimate of Y given $X = x$. The function $\widehat{y}(\,\cdot\,)$ determines the corresponding estimator, which we will denote by $\widehat{y}(X)$, or more simply by just $\widehat{Y}$, if it is understood what random variable the estimator is operating on. Note that the estimator $\widehat{Y} = \widehat{y}(X)$ is a random variable. We have already seen that the MMSE estimate $\widehat{y}(x)$ is given by the conditional mean, $E[Y|X = x]$, which suggests yet another natural notation for the MMSE estimator:

$$\widehat{Y} = \widehat{y}(X) = E[Y|X]\ . \tag{8.25}$$

Note that $E[Y|X]$ denotes a random variable, not a number.

The preceding discussion applies essentially unchanged to the case where we observe several random variables, assembled in the vector $\mathbf{X}$. The MMSE estimator in this case is denoted by

$$\widehat{Y} = \widehat{y}(\mathbf{X}) = E[Y|\mathbf{X}]\ . \tag{8.26}$$

Perhaps not surprisingly, the MMSE estimator for Y given $\mathbf{X}$ minimizes the mean square error, averaged over all Y and $\mathbf{X}$. This is because the MMSE estimator minimizes the mean square error for each particular value $\mathbf{x}$ of $\mathbf{X}$. More formally,

$$\begin{aligned} E_{Y,\mathbf{X}}\Big(\,[Y - \widehat{y}(\mathbf{X})]^2\,\Big) &= E_{\mathbf{X}}\Big(\,E_{Y|\mathbf{X}}\Big(\,[Y - \widehat{y}(\mathbf{X})]^2\,|\,\mathbf{X}\Big)\Big) \\ &= \int_{-\infty}^{\infty}\Big(\,E_{Y|\mathbf{X}}\Big(\,[Y - \widehat{y}(\mathbf{x})]^2\,|\,\mathbf{X} = \mathbf{x}\Big)\,f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x}\ . \end{aligned} \tag{8.27}$$

(The subscripts on the expectation operators are used to indicate explicitly which densities are involved in computing the associated expectations; the densities and integration are multivariate when $\mathbf{X}$ is not a scalar.) Because the estimate $\widehat{y}(\mathbf{x})$ is chosen to minimize the inner expectation $E_{Y|\mathbf{X}}$ for each value $\mathbf{x}$ of $\mathbf{X}$, it also minimizes the outer expectation $E_{\mathbf{X}}$, since $f_{\mathbf{X}}(\mathbf{X})$ is nonnegative.

EXAMPLE 8.4 MMSE Estimator for Bivariate Gaussian Random Variables

We have already, in Example 8.3, constructed the MMSE estimate of one member of a pair of bivariate Gaussian random variables, given a measurement of the other. Using the same notation as in that example, it is evident that the MMSE estimator is simply obtained on replacing x by X in (8.23):

$$\widehat{Y} = \widehat{y}(X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)\ . \tag{8.28}$$

The conditional MMSE given $X = x$ was found in the earlier example to be $\sigma_Y^2(1-\rho^2)$, which did not depend on the value of x, so the MMSE of the estimator, averaged over all X, ends up still being $\sigma_Y^2(1-\rho^2)$.

EXAMPLE 8.5 MMSE Estimator for Signal in Additive Noise

Suppose the random variable X is a noisy measurement of the angular position Y of an antenna, so $X = Y + W$, where W denotes the additive noise. Assume the noise is independent of the angular position, i.e., Y and W are independent random variables, with Y uniformly distributed in the interval $[-1, 1]$ and W uniformly distributed in the interval $[-2, 2]$. (Note that the setup in this example is essentially the same as in Example 8.2, though the context, notation and parameters are different.)

Given that $X = x$, we would like to determine the MMSE estimate $\widehat{y}(x)$, the resulting mean square error, and the overall mean square error averaged over all possible values x that the random variable X can take. Since $\widehat{y}(x)$ is the conditional expectation of Y given $X = x$, we need to determine $f_{Y|X}(y|x)$. For this, we first determine the joint density of Y and W, and from this the required conditional density.

From the independence of Y and W:

$$f_{Y,W}(y,w) = f_Y(y)f_W(w) = \begin{cases} \frac{1}{8} & -2 \le w \le 2, -1 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

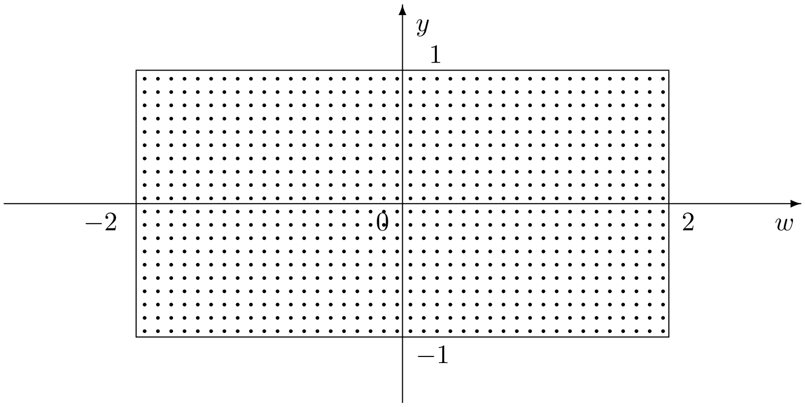

FIGURE 8.5 Joint PDF of Y and W for Example 8.5.

Conditioned on $Y = y$, X is the same as $y + W$, uniformly distributed over the interval $[y-2, y+2]$. Now

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = (\frac{1}{4})(\frac{1}{2}) = \frac{1}{8}$$

for $-1 \leq y \leq 1$, $y - 2 \leq x \leq y + 2$, and zero otherwise. The joint PDF is therefore uniform over the parallelogram shown in the Figure 8.6.

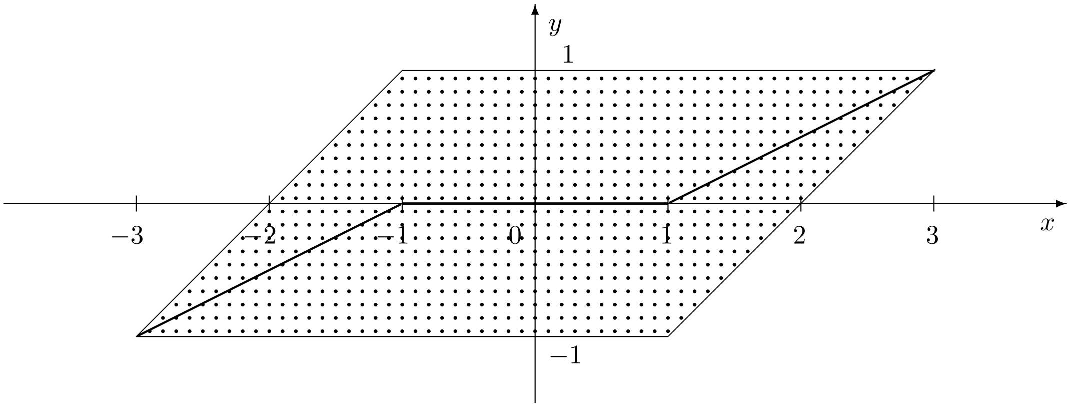

FIGURE 8.6 Joint PDF of X and Y and plot of the MMSE estimator of Y from X for Example 8.5.

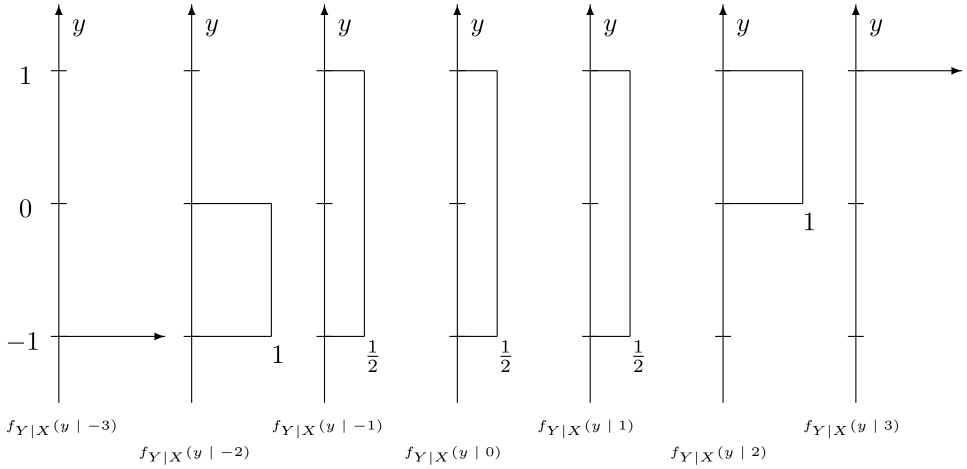

FIGURE 8.7 Conditional PDF $f_{Y|X}$ for various realizations of X for Example 8.5.

Given $X = x$, the conditional PDF $f_{Y|X}$ is uniform on the corresponding vertical section of the parallelogram:

$$f_{Y|X}(y,x) = \begin{cases} \dfrac{1}{3+x} & -3 \leq x \leq -1\,, -1 \leq y \leq x+2 \\ \dfrac{1}{2} & -1 \leq x \leq 1\,, -1 \leq y \leq 1 \\ \dfrac{1}{3-x} & 1 \leq x \leq 3\,, x-2 \leq y \leq 1 \end{cases} \tag{8.29}$$

The MMSE estimate $\widehat{y}(x)$ is the conditional mean of Y given $X = x$, and the conditional mean is the midpoint of the corresponding vertical section of the parallelogram. The conditional mean is displayed as the heavy line on the parallelogram in the second plot. In analytical form,

$$\widehat{y}(x) = E[Y \mid X = x] = \begin{cases} \frac{1}{2} + \frac{1}{2}x & -3 \leq x < -1 \\ 0 & -1 \leq x < 1 \\ -\frac{1}{2} + \frac{1}{2}x & 1 \leq x \leq 3 \end{cases} \qquad (8.30)$$

The minimum mean square error associated with this estimate is the variance of the uniform distribution in eq. (8.29), specifically:

$$E[\{Y - \widehat{y}(x)\}^2 \mid X = x] \quad \begin{cases} \frac{(3+x)^2}{12} & -3 \leq x < -1 \\ \frac{1}{3} & -1 \leq x < 1 \\ \frac{(3-x)^2}{12} & 1 \leq x \leq 3 \end{cases} \qquad (8.31)$$

Equation (8.31) specifies the mean square error that results for any specific value x of the measurement of X. Since the measurement is a random variable, it is also of interest to know what the mean square error is, averaged over all possible values of the measurement, i.e. over the random variable X. To determine this, we first determine the marginal PDF of X:

$$f_X(x) = \frac{f_{X,Y}(x,y)}{f_{Y|X}(y \mid x)} = \begin{cases} \frac{3+x}{8} & -3 \leq x < -1 \\ \frac{1}{4} & -1 \leq x < 1 \\ \frac{3-x}{8} & 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

This could also be found by convolution, $f_X = f_Y * f_W$, since Y and W are statistically independent. Then,

$$\begin{aligned} E_X[E_{Y|X}\{(Y - \widehat{y}(x)\}^2 \mid X = x]] &= \int_{-\infty}^{\infty} E[(Y - \widehat{y}(x))^2 \mid X = x] f_X(x) \mathrm{d}x \\ &= \int_{-3}^{-1} (\frac{(3+x)^2}{12})(\frac{3+x}{8})\mathrm{d}x + \int_{-1}^{1} (\frac{1}{3})(\frac{1}{4})\mathrm{d}x + \int_{1}^{3} (\frac{(3-x)^2}{12})(\frac{3-x}{8})\mathrm{d}x \\ &= \frac{1}{4} \end{aligned}$$

Compare this with the mean square error if we just estimated Y by its mean, namely 0. The mean square error would then be the variance σ_Y^2:

$$\sigma_Y^2 = \frac{[1-(-1)]^2}{12} = \frac{1}{3} ,$$

so the mean square error is indeed reduced by allowing ourselves to use knowledge of X and of the probabilistic relation between Y and X.

8.2.1 Orthogonality

A further important property of the MMSE estimator is that the residual error $Y - \widehat{y}(\mathbf{X})$ is orthogonal to any function $h(\mathbf{X})$ of the measured random variables:

$$E_{Y,X}[\{Y - \widehat{y}(\mathbf{X})\}h(\mathbf{X})] = 0 , \qquad (8.32)$$

where the expectation is computed over the joint density of Y and $\mathbf{X}$. Rearranging this, we have the equivalent condition

$$E_{Y,X}[\widehat{y}(\mathbf{X})h(\mathbf{X})] = E_{Y,X}[Yh(\mathbf{X})] , \qquad (8.33)$$

i.e., the MMSE estimator has the same correlation as Y does with any function of X. In particular, choosing $h(\mathbf{X}) = 1$, we find that

$$E_{Y,X}[\widehat{y}(\mathbf{X})] = E_Y[Y] . \qquad (8.34)$$

The latter property results in the estimator being referred to as unbiased: its expected value equals the expected value of the random variable being estimated. We can invoked the unbiasedness property to interpret (8.32) as stating that the estimation error of the MMSE estimator is uncorrelated with any function of the random variables used to construct the estimator.

The proof of the correlation matching property in (8.33) is in the following sequence of equalities:

$$\begin{aligned} E_{Y,X}[\widehat{y}(\mathbf{X})h(\mathbf{X})] &= E_X[E_{Y|X}[Y|\mathbf{X}]h(\mathbf{X})] && (8.35)\\ &= E_X[E_{Y|X}[Yh(\mathbf{X})|\mathbf{X}]] && (8.36)\\ &= E_{Y,X}[Yh(\mathbf{X})] . && (8.37)\end{aligned}$$

Rearranging the final result here, we obtain the orthogonality condition in (8.32).

8.3 LINEAR MINIMUM MEAN SQUARE ERROR ESTIMATION

In general, the conditional expectation $E(Y|\mathbf{X})$ required for the MMSE estimator developed in the preceding sections is difficult to determine, because the conditional density $f_{Y|\mathbf{X}}(y|\mathbf{x})$ is not easily determined. A useful and widely used compromise is to restrict the estimator to be a fixed linear (or actually affine, i.e., linear plus a constant) function of the measured random variables, and to choose the linear relationship so as to minimize the mean square error. The resulting estimator is called the linear minimum mean square error (LMMSE) estimator. We begin with the simplest case.

Suppose we wish to construct an estimator for the random variable Y in terms of another random variable X, restricting our estimator to be of the form

$$\widehat{Y}_\ell = \widehat{y}_\ell(X) = aX + b \ , \tag{8.38}$$

where a and b are to be determined so as to minimize the mean square error

$$E_{Y,X}[(Y - \widehat{Y}_\ell)^2] = E_{Y,X}[\{Y - (aX + b)\}^2] \ . \tag{8.39}$$

Note that the expectation is taken over the joint density of Y and X; the linear estimator is picked to be optimum when averaged over all possible combinations of Y and X that may occur. We have accordingly used subscripts on the expectation operations in (8.39) to make explicit for now the variables whose joint density the expectation is being computed over; we shall eventually drop the subscripts.

Once the optimum a and b have been chosen in this manner, the estimate of Y, given a particular x, is just $\widehat{y}_\ell(x) = ax + b$, computed with the already designed values of a and b. Thus, in the LMMSE case we construct an optimal linear estimator, and for any particular x this estimator generates an estimate that is not claimed to have any individual optimality property. This is in contrast to the MMSE case considered in the previous sections, where we obtained an optimal MMSE estimate for each x, namely $E[Y|X = x]$, that minimized the mean square error conditioned on $X = x$. The distinction can be summarized as follows: in the unrestricted MMSE case, the optimal estimator is obtained by joining together all the individual optimal estimates, whereas in the LMMSE case the (generally non-optimal) individual estimates are obtained by simply evaluating the optimal linear estimator.

We turn now to minimizing the expression in (8.39), by differentiating it with respect to the parameters a and b, and setting each of the derivatives to 0. (Consideration of the second derivatives will show that we do indeed find minimizing values in this fashion, but we omit the demonstration.) First differentiating (8.39) with respect to b, taking the derivative inside the integral that corresponds to the expectation operation, and then setting the result to 0, we conclude that

$$E_{Y,X}[Y - (aX + b)] = 0 \ , \tag{8.40}$$

or equivalently

$$E[Y] = E[aX + b] = E[\widehat{Y}_\ell] \ , \tag{8.41}$$

from which we deduce that

$$b = \mu_Y - a\mu_X \ , \tag{8.42}$$

where $\mu_Y = E[Y] = E_{Y,X}[Y]$ and $\mu_X = E[X] = E_{Y,X}[X]$. The optimum value of b specified in (8.42) in effect serves to make the linear estimator unbiased, i.e., the expected value of the estimator becomes equal to the expected value of the random variable we are trying to estimate, as (8.41) shows.

Using (8.42) to substitute for b in (8.38), it follows that

$$\widehat{Y}_\ell = \mu_Y + a(X - \mu_X) \ . \tag{8.43}$$

In other words, to the expected value μ_Y of the random variable Y that we are estimating, the optimal linear estimator adds a suitable multiple of the difference $X - \mu_X$ between the measured random variable and its expected value. We turn now to finding the optimum value of this multiple, a.

First rewrite the error criterion (8.39) as

$$E[\{(Y - \mu_Y) - (\widehat{Y}_\ell - \mu_Y)\}^2] = E[(\widetilde{Y} - a\widetilde{X})^2] \ , \tag{8.44}$$

where

$$\widetilde{Y} = Y - \mu_Y \quad \text{and} \quad \widetilde{X} = X - \mu_X \ , \tag{8.45}$$

and where we have invoked (8.43) to obtain the second equality in (8.44). Now taking the derivative of the error criterion in (8.44) with respect to a, and setting the result to 0, we find

$$E[(\widetilde{Y} - a\widetilde{X})\widetilde{X}] = 0 \ . \tag{8.46}$$

Rearranging this, and recalling that $E[\widetilde{Y}\widetilde{X}] = \sigma_{YX}$, i.e., the covariance of Y and X, and that $E[\widetilde{X}^2] = \sigma_X^2$, we obtain

$$a = \frac{\sigma_{YX}}{\sigma_X^2} = \rho_{YX}\frac{\sigma_Y}{\sigma_X} \ , \tag{8.47}$$

where ρ_{YX} — which we shall simply write as ρ when it is clear from context what variables are involved — denotes the correlation coefficient between Y and X.

It is also enlightening to understand the above expression for a in terms of the vector-space picture for random variables developed in the previous chapter.

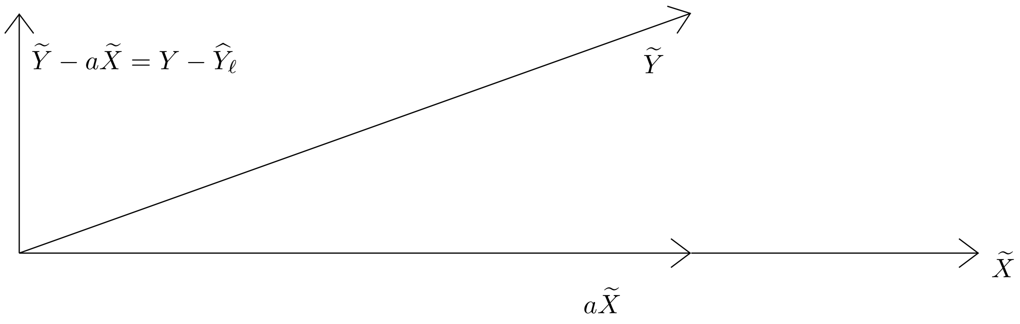

FIGURE 8.8 Expression for a from Eq. (8.47) illustrated in vector space.

The expression (8.44) for the error criterion shows that we are looking for a vector $a\widetilde{X}$, which lies along the vector $\widetilde{X}$, such that the squared length of the error vector

$\widetilde{Y} - a\widetilde{X}$ is minimum. It follows from familiar geometric reasoning that the optimum choice of $a\widetilde{X}$ must be the orthogonal projection of $\widetilde{Y}$ on $\widetilde{X}$, and that this projection is

$$a\widetilde{X} = \frac{< \widetilde{Y}, \widetilde{X} >}{< \widetilde{X}, \widetilde{X} >} \widetilde{X} . \tag{8.48}$$

Here, as in the previous chapter, $< U, V >$ denotes the inner product of the vectors U and V, and in the case where the "vectors" are random variables, denotes $E[UV]$. Our expression for a in (8.47) follows immediately. Figure 8.8 shows the construction associated with the requisite calculations. Recall from the previous chapter that the correlation coefficient ρ denotes the cosine of the angle between the vectors $\widetilde{Y}$ and $\widetilde{X}$.

The preceding projection operation implies that the error $\widetilde{Y} - a\widetilde{X}$, which can also be written as $Y - \widehat{Y}_\ell$, must be orthogonal to $\widetilde{X} = X - \mu_X$. This is precisely what (8.46) says. In addition, invoking the unbiasedness of $\widehat{Y}_\ell$ shows that $Y - \widehat{Y}_\ell$ must be orthogonal to μ_X (or any other constant), so $Y - \widehat{Y}_\ell$ is therefore orthogonal to X itself:

$$E[(Y - \widehat{Y}_\ell)X] = 0 . \tag{8.49}$$

In other words, the optimal LMMSE estimator is unbiased and such that the estimation error is orthogonal to the random variable on which the estimator is based. (Note that the statement in the case of the MMSE estimator in the previous section was considerably stronger, namely that the error was orthogonal to any function $h(X)$ of the measured random variable, not just to the random variable itself.)

The preceding development shows that the properties of (i) unbiasedness of the estimator, and (ii) orthogonality of the error to the measured random variable, completely characterize the LMMSE estimator. Invoking these properties yields the LMMSE estimator.

Going a step further with the geometric reasoning, we find from Pythagoras's theorem applied to the triangle in Figure 8.8 that the minimum mean square error (MMSE) obtained through use of the LMMSE estimator is

$$\text{MMSE} = E[(\widetilde{Y} - a\widetilde{X})^2] = E[\widetilde{Y}^2](1 - \rho^2) = \sigma_Y^2(1 - \rho^2) . \tag{8.50}$$

This result could also be obtained purely analytically, of course, without recourse to the geometric interpretation. The result shows that the mean square error σ_Y^2 that we had prior to estimation in terms of X is reduced by the factor $1 - \rho^2$ when we use X in an LMMSE estimator. The closer that ρ is to $+1$ or -1 (corresponding to strong positive or negative correlation respectively), the more our uncertainty about Y is reduced by using an LMMSE estimator to extract information that X carries about Y.

Our results on the LMMSE estimator can now be summarized in the following expressions for the estimator, with the associated minimum mean square error being given by (8.50):

$$\widehat{Y}_\ell = \widehat{y}_\ell(X) = \mu_Y + \frac{\sigma_{YX}}{\sigma_X^2}(X - \mu_X) = \mu_Y + \rho\,\frac{\sigma_Y}{\sigma_X}(X - \mu_X) , \tag{8.51}$$

or the equivalent but perhaps more suggestive form

$$\frac{\widehat{Y}_\ell - \mu_Y}{\sigma_Y} = \rho\,\frac{X - \mu_X}{\sigma_X}\ . \tag{8.52}$$

The latter expression states that the normalized deviation of the estimator from its mean is ρ times the normalized deviation of the observed variable from its mean; the more highly correlated Y and X are, the more closely we match the two normalized deviations.

Note that our expressions for the LMMSE estimator and its mean square error are the same as those obtained in Example 8.4 for the MMSE estimator in the bivariate Gaussian case. The reason is that the MMSE estimator in that case turned out to be linear (actually, affine), as already noted in the example.

EXAMPLE 8.6 LMMSE Estimator for Signal in Additive Noise

We return to Example 8.5, for which we have already computed the MMSE estimator, and we now design an LMMSE estimator. Recall that the random variable X denotes a noisy measurement of the angular position Y of an antenna, so $X = Y + W$, where W denotes the additive noise. We assume the noise is independent of the angular position, i.e., Y and W are independent random variables, with Y uniformly distributed in the interval $[-1, 1]$ and W uniformly distributed in the interval $[-2, 2]$.

For the LMMSE estimator of Y in terms of X, we need to determine the respective means and variances, as well as the covariance, of these random variables. It is easy to see that

$$\mu_Y = 0\ , \quad \mu_W = 0\ , \quad \mu_X = 0\ , \quad \sigma_Y^2 = \frac{1}{3}\ , \quad \sigma_W^2 = \frac{4}{3}\ ,$$

$$\sigma_X^2 = \sigma_Y^2 + \sigma_W^2 = \frac{5}{3}\ , \quad \sigma_{YX} = \sigma_Y^2 = \frac{1}{3}\ , \quad \rho_{YX} = \frac{1}{\sqrt{5}}\ .$$

The LMMSE estimator is accordingly

$$\widehat{Y}_\ell = \frac{1}{5}X\ ,$$

and the associated MMSE is

$$\sigma_Y^2(1-\rho^2) = \frac{4}{15}\ .$$

This MMSE should be compared with the (larger) mean square error of $\frac{1}{3}$ obtained if we simply use $\mu_Y = 0$ as our estimator for Y, and the (smaller) value $\frac{1}{4}$ obtained using the MMSE estimator in Example 8.5.

EXAMPLE 8.7 Single-Point LMMSE Estimator for Sinusoidal Random Process

Consider a sinusoidal signal of the form

$$X(t) = A\cos(\omega_0 t + \Theta) \tag{8.53}$$

where ω_0 is assumed known, while A and Θ are statistically independent random variables, with the PDF of Θ being uniform in the interval $[0, 2\pi]$. Thus $X(t)$ is a random signal, or equivalently a set or "ensemble" of signals corresponding to the various possible outcomes for A and Θ in the underlying probabilistic experiment. We will discuss such signals in more detail in the next chapter, where we will refer to them as random processes. The value that $X(t)$ takes at some particular time $t = t_0$ is simply a random variable, whose specific value will depend on which outcomes for A and Θ are produced by the underlying probabilistic experiment.

Suppose we are interested in determining the LMMSE estimator for $X(t_1)$ based on a measurement of $X(t_0)$, where t_0 and t_1 are specified sampling times. In other words, we want to choose a and b in

$$\widehat{X}(t_1) = aX(t_0) + b \tag{8.54}$$

so as to minimize the mean square error between $X(t_1)$ and $\widehat{X}(t_1)$.

We have established that b must be chosen to ensure the estimator is unbiased:

$$E[\widehat{X}(t_1)] = aE[X(t_0)] + b = E[X(t_1)]\ .$$

Since A and Θ are independent,

$$E[X(t_0)] = E\{A\}\int_0^{2\pi} \frac{1}{2\pi}\cos(\omega_0 t_0 + \theta)\, d\theta = 0$$

and similarly $E[X(t_1)] = 0$, so we choose $b = 0$.

Next we use the fact that the error of the LMMSE estimator is orthogonal to the data:

$$E[(\widehat{X}(t_1) - X(t_1))X(t_0)] = 0$$

and consequently

$$aE[X^2(t_0)] = E[X(t_1)X(t_0)]$$

or

$$a = \frac{E[X(t_1)X(t_0)]}{E[X^2(t_0)]}\ . \tag{8.55}$$

The numerator and denominator in (8.55) are respectively

$$\begin{aligned} E[X(t_1)X(t_0)] &= E[A^2]\int_0^{2\pi} \frac{1}{2\pi}\cos(\omega_0 t_1 + \theta)\cos(\omega_0 t_0 + \theta)\, d\theta \\ &= \frac{E[A^2]}{2}\cos\{\omega_0(t_1 - t_0)\} \end{aligned}$$

and $E[X^2(t_0)] = \frac{E[A^2]}{2}$. Thus $a = \cos\{\omega_0(t_1 - t_0)\}$, so the LMMSE estimator is

$$\widehat{X}(t_1) = X(t_0)\cos\{\omega_0(t_1 - t_0)\}\ . \qquad (8.56)$$

It is interesting to observe that the distribution of A doesn't play a role in this equation.

To evaluate the mean square error associated with the LMMSE estimator, we compute the correlation coefficient between the samples of the random signal at t_0 and t_1. It is easily seen that$\rho = a = \cos\{\omega_0(t_1 - t_0)\}$, so the mean square error is

$$\frac{E[A^2]}{2}\Big(1 - \cos^2\{\omega_0(t_1 - t_0)\}\Big) = \frac{E[A^2]}{2}\sin^2\{\omega_0(t_1 - t_0)\}\ . \qquad (8.57)$$

We now extend the LMMSE estimator to the case where our estimation of a random variable Y is based on observations of multiple random variables, say $X_1, \ldots, X_L$, gathered in the vector $\mathbf{X}$. The affine estimator may then be written in the form

$$\widehat{Y}_\ell = \widehat{y}_\ell(\mathbf{X}) = a_0 + \sum_{j=1}^{L} a_j X_j\ . \qquad (8.58)$$

As we shall see, the coefficient a_i of this LMMSE estimator can be found by solving a linear system of equations that is completely defined by the first and second moments (i.e., means, variances and covariances) of the random variables Y and X_j. The fact that the model (8.58) is linear in the parameters a_i is what results in a linear system of equations; the fact that the model is affine in the random variables is what makes the solution only depend on their first and second moments. Linear equations are easy to solve, and first and second moments are generally easy to determine, hence the popularity of LMMSE estimation.

The development below follows along the same lines as that done earlier in this section for the case where we just had a single observed random variable X, but we use the opportunity to review the logic of the development and to provide a few additional insights.

We want to minimize the mean square error

$$E\Big[\Big(Y - (a_0 + \sum_{j=1}^{L} a_j X_j)\Big)^2\Big]\ , \qquad (8.59)$$

where the expectation is computed using the joint density of Y and $\mathbf{X}$. We use the joint density rather than the conditional because the parameters are not going to be picked to be best for a particular set of measured values $\mathbf{x}$ — otherwise we could do as well as the nonlinear estimate in this case, by setting $a_0 = E[Y \,|\, \mathbf{X} = \mathbf{x}]$ and setting all the other a_i to zero. Instead, we are picking the parameters to be the best averaged over all possible $\mathbf{X}$. The linear estimator will in general not be as good as the unconstrained estimator, except in special cases (some of them important, as in the case of bivariate Gaussian random variables) but this estimator has the advantage that it is easy to solve for, as we now show.

To minimize the expression in (8.59), we differentiate it with respect to a_i for $i = 0, 1, \cdots, L$, and set each of the derivatives to 0. (Again, calculations involving second derivatives establish that we do indeed obtain minimizing values, but we omit these calculation here.) First differentiating with respect to a_0 and setting the result to 0, we conclude that

$$E[Y] = E[\,a_0 + \sum_{j=1}^{L} a_j X_j\,] = E[\widehat{Y}_\ell] \tag{8.60}$$

or

$$a_0 = \mu_Y - \sum_{j=1}^{L} a_j\, \mu_{X_j}\ , \tag{8.61}$$

where $\mu_Y = E[Y]$ and $\mu_{X_j} = E[X_j]$. This optimum value of a_0 serves to make the linear estimator unbiased, in the sense that (8.60) holds, i.e., the expected value of the estimator is the expected value of the random variable we are trying to estimate.

Using (8.61) to substitute for a_0 in (8.58), it follows that

$$\widehat{Y}_\ell = \mu_Y + \sum_{j=1}^{L} a_j (X_j - \mu_{X_j})\ . \tag{8.62}$$

In other words, the estimator corrects the expected value μ_Y of the variable we are estimating, by a linear combination of the deviations $X_j - \mu_{X_j}$ between the measured random variables and their respective expected values.

Taking account of (8.62), we can rewrite our mean square error criterion (8.59) as

$$E[\{(Y - \mu_Y) - (\widehat{Y}_\ell - \mu_Y)\}^2] = E\Big[\Big(\widetilde{Y} - \sum_{j=1}^{L} a_j \widetilde{X}_j)\Big)^2\Big]\ , \tag{8.63}$$

where

$$\widetilde{Y} = Y - \mu_Y \quad \text{and} \quad \widetilde{X}_j = X_j - \mu_{X_j}\ . \tag{8.64}$$

Differentiating this with respect to each of the remaining coefficients $a_i, i = 1, 2, ...L$, and setting the result to zero produces the equations

$$E[(\widetilde{Y} - \sum_{j=1}^{L} a_j \widetilde{X}_j)\widetilde{X}_i] = 0 \quad i = 1, 2, ..., L\ . \tag{8.65}$$

or equivalently, if we again take account of (8.62),

$$E[(Y - \widehat{Y}_\ell)\widetilde{X}_i] = 0 \quad i = 1, 2, ..., L\ . \tag{8.66}$$

Yet another version follows on noting from (8.60) that $Y - \widehat{Y}_\ell$ is orthogonal to all constants, in particular to μ_{X_i}, so

$$E[(Y - \widehat{Y}_\ell)X_i] = 0 \quad i = 1, 2, ..., L \, . \tag{8.67}$$

All three of the preceding sets of equations express, in slightly different forms, the orthogonality of the estimation error to the random variables used in the estimator. One moves between these forms by invoking the unbiasedness of the estimator. The last of these, (8.67), is the usual statement of the orthogonality condition that governs the LMMSE estimator. (Note once more that the statement in the case of the MMSE estimator in the previous section was considerably stronger, namely that the error was orthogonal to any function $h(\mathbf{X})$ of the measured random variables, not just to the random variables themselves.) Rewriting this last equation as

$$E[YX_i] = E[\widehat{Y}_\ell X_i] \quad i = 1, 2, ..., L \tag{8.68}$$

yields an equivalent statement of the orthogonality condition, namely that the LMMSE estimator $\widehat{Y}_\ell$ has the same correlations as Y with the measured variables X_i.

The orthogonality and unbiasedness conditions together determine the LMMSE estimator completely. Also, the preceding developments shows that the first and second moments of Y and the X_i are exactly matched by the corresponding first and second moments of $\widehat{Y}_\ell$ and the X_i. It follows that Y and $\widehat{Y}_\ell$ cannot be told apart on the basis of only first and second moments with the measured variables X_i.

We focus now on (8.65), because it provides the best route to a solution for the coefficients a_j, $j = 1, \ldots, L$. This set of equations can be expressed as

$$\sum_{j=1}^{L} \sigma_{X_i X_j} a_j = \sigma_{X_i Y} \ , \tag{8.69}$$

where $\sigma_{X_i X_j}$ is the covariance of X_i and X_j (so $\sigma_{X_i X_i}$ is just the variance $\sigma^2_{X_i}$), and $\sigma_{X_i Y}$ is the covariance of X_i and Y. Collecting these equations in matrix form, we obtain

$$\begin{bmatrix} \sigma_{X_1X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_L} \\ \sigma_{X_2X_1} & \sigma_{X_2X_2} & \cdots & \sigma_{X_2X_L} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_LX_1} & \sigma_{X_LX_2} & \cdots & \sigma_{X_LX_L} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{bmatrix} = \begin{bmatrix} \sigma_{X_1Y} \\ \sigma_{X_2Y} \\ \vdots \\ \sigma_{X_LY} \end{bmatrix} . \tag{8.70}$$

This set of equations is referred to as the normal equations. We can rewrite the normal equations in more compact matrix notation:

$$(\mathbf{C_{XX}})\, \mathbf{a} = \mathbf{C}_{\mathbf{X}Y} \tag{8.71}$$

where the definitions of $\mathbf{C_{XX}}$, $\mathbf{a}$, and $\mathbf{C}_{\mathbf{X}Y}$ should be evident on comparing the last two equations. The solution of this set of L equations in L unknowns yields the

$\{a_j\}$ for $j = 1, \cdots, L$, and these values may be substituted in (8.62) to completely specify the estimator. In matrix notation, the solution is

$$\mathbf{a} = (\mathbf{C_{XX}})^{-1}\mathbf{C_{X}}_Y \; . \tag{8.72}$$

It can be shown quite straightforwardly (though we omit the demonstration) that the minimum mean square error obtained with the LMMSE estimator is

$$\sigma_Y^2 - \mathbf{C}_{Y\mathbf{X}}(\mathbf{C_{XX}})^{-1}\mathbf{C}_{\mathbf{X}Y} = \sigma_Y^2 - \mathbf{C}_{Y\mathbf{X}}\mathbf{a} \; , \tag{8.73}$$

where $\mathbf{C}_{Y\mathbf{X}}$ is the transpose of $\mathbf{C}_{\mathbf{X}Y}$.

EXAMPLE 8.8 Estimation from Two Noisy Measurements

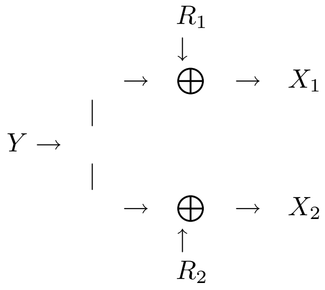

FIGURE 8.9 Illustration of relationship between random variables from Eq. (8.75) for Example 8.8.

Assume that Y, R_1 and R_2 are mutually uncorrelated, and that R_1 and R_2 have zero means and equal variances. We wish to find the linear MMSE estimator for Y, given measurements of X_1 and X_2. This estimator takes the form $\widehat{Y}_\ell = a_0 + a_1X_1 + a_2X_2$. Our requirement that $\widehat{Y}_\ell$ be unbiased results in the constraint

$$a_0 = \mu_Y - a_1\mu_{X_1} - a_2\mu_{X_2} = \mu_Y(1 - a_1 - a_2) \tag{8.74}$$

Next, we need to write down the normal equations, for which some preliminary calculations are required. Since

$$\begin{aligned} X_1 &= Y + R_1 \\ X_2 &= Y + R_2 \end{aligned} \tag{8.75}$$

and Y, R_1 and R_2 are mutually uncorrelated, we find

$$\begin{aligned} E[X_i^2] &= E[Y^2] + E[R_i^2] \; , \\ E[X_1X_2] &= E[Y^2] \; , \\ E[X_iY] &= E[Y^2] \; . \end{aligned} \tag{8.76}$$

The normal equations for this case thus become

$$\begin{bmatrix} \sigma_Y^2 + \sigma_R^2 & \sigma_Y^2 \\ \sigma_Y^2 & \sigma_Y^2 + \sigma_R^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sigma_Y^2 \\ \sigma_Y^2 \end{bmatrix} \tag{8.77}$$

from which we conclude that

$$\begin{aligned} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \frac{1}{(\sigma_Y^2 + \sigma_R^2)^2 - \sigma_Y^4} \begin{bmatrix} \sigma_Y^2 + \sigma_R^2 & -\sigma_Y^2 \\ -\sigma_Y^2 & \sigma_Y^2 + \sigma_R^2 \end{bmatrix} \begin{bmatrix} \sigma_Y^2 \\ \sigma_Y^2 \end{bmatrix} \\ &= \frac{\sigma_Y^2}{2\sigma_Y^2 + \sigma_R^2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} . \end{aligned} \tag{8.78}$$

Finally, therefore,

$$\widehat{Y}_\ell = \frac{1}{2\sigma_Y^2 + \sigma_R^2} (\sigma_R^2 \mu_Y + \sigma_Y^2 X_1 + \sigma_Y^2 X_2) \tag{8.79}$$

and applying (8.73) we get that the associated minimum mean square error (MMSE) is

$$\frac{\sigma_Y^2 \sigma_R^2}{2\sigma_Y^2 + \sigma_R^2} . \tag{8.80}$$

One can easily check that both the estimator and the associated MMSE take reasonable values at extreme ranges of the signal-to-noise ratio σ_Y^2/σ_R^2.

6.011 Introduction to Communication, Control, and Signal Processing
Spring 2010