Detection, Estimation, and Modulation Theory, Part I:
Detection, Estimation, and Linear Modulation Theory. Harry L. Van Trees

ISBNs: 0-471-09517-6 (Paperback); 0-471-22108-2 (Electronic)

5

Estimation of Continuous Waveforms

5.1 INTRODUCTION

Up to this point we have considered the problems of detection and parameter estimation. We now consider the problem of estimating a *continuous* waveform. Just as in the parameter estimation problem, we shall find it convenient to discuss both nonrandom waveforms and waveforms that are sample functions from a random process. We shall find that the estimation procedure for nonrandom waveforms is straightforward. By contrast, when the waveform is a sample function from a random process, the formulation is straightforward but the solution is more complex.

Before solving the estimation problem it will be worthwhile to investigate some of the physical problems in which we want to estimate a continuous waveform. We consider the random waveform case first.

An important situation in which we want to estimate a random waveform is in analog modulation systems. In the simplest case the message $a(t)$ is the input to a no-memory modulator whose output is $s(t, a(t))$ which is then transmitted as shown in Fig. 5.1. The transmitted signal is deterministic in the sense that a given sample function $a(t)$ causes a unique output $s(t, a(t))$. Some common examples are the following:

$$s(t, a(t)) = \sqrt{2P}\, a(t) \sin \omega_c t. \tag{1}$$

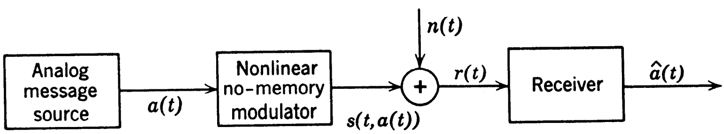

Fig. 5.1 A continuous no-memory modulation system.

This is double sideband, suppressed carrier, amplitude modulation (DSB-SC-AM).

$$s(t, a(t)) = \sqrt{2P}\,[1 + ma(t)] \sin \omega_c t. \tag{2}$$

This is conventional DSB-AM with a residual carrier component.

$$s(t, a(t)) = \sqrt{2P} \sin [\omega_c t + \beta a(t)]. \tag{3}$$

This is phase modulation (PM).

The transmitted waveform is corrupted by a sample function of zero-mean Gaussian noise process which is independent of the message process. The noise is completely characterized by its covariance function $K_n(t, u)$. Thus, for the system shown in Fig.5.1, the received signal is

$$r(t) = s(t, a(t)) + n(t), \qquad T_i \le t \le T_f. \tag{4}$$

The simple system illustrated is not adequate to describe many problems of interest. The first step is to remove the no-memory restriction. A modulation system with memory is shown in Fig. 5.2. Here, $h(t, u)$ represents the impulse response of a linear, not necessarily time-invariant, filter. Examples are the following:

1. The linear system is an integrator and the no-memory device is a phase modulator. In this case the transmitted signal is

$$s(t, x(t)) = \sqrt{2P} \sin \left[\omega_c t + \int_{T_i}^{t} a(u)\, du\right]. \tag{5}$$

This is frequency modulation (FM).

2. The linear system is a realizable time-invariant network and the no-memory device is a phase modulator. The transmitted signal is

$$s(t, x(t)) = \sqrt{2P} \sin \left[\omega_c t + \int_{T_i}^{t} h(t - u)\, a(u)\, du\right]. \tag{6}$$

This is pre-emphasized angle modulation.

Figures 5.1 and 5.2 describe a broad class of analog modulation systems which we shall study in some detail. We denote the waveform of interest, $a(t)$, as the *message*. The message may come from a variety of sources. In

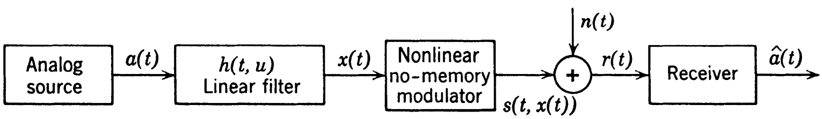

Fig. 5.2 A modulation system with memory.

commercial FM it corresponds to music or speech. In a satellite telemetry system it might correspond to analog data from a sensor (e.g., temperature or attitude).

The waveform estimation problem occurs in a number of other areas. If we remove the modulator in Fig. 5.1,

$$r(t) = a(t) + n(t), \qquad T_i \leq t \leq T_f. \tag{7}$$

If $a(t)$ represents the position of some object we are trying to track in the presence of measurement noise, we have the simplest form of the control problem.

Many more complicated systems also fit into the model. Three are shown in Fig. 5.3. The system in Fig. 5.3*a* is an FM/FM system. This type of system is commonly used when we have a number of messages to transmit. Each message is modulated onto a subcarrier at a different frequency, the modulated subcarriers are summed, and modulated onto the main carrier. In the Fig. 5.3*a* we show the operations for a single message. The system in Fig. 5.3*b* represents an FM modulation system

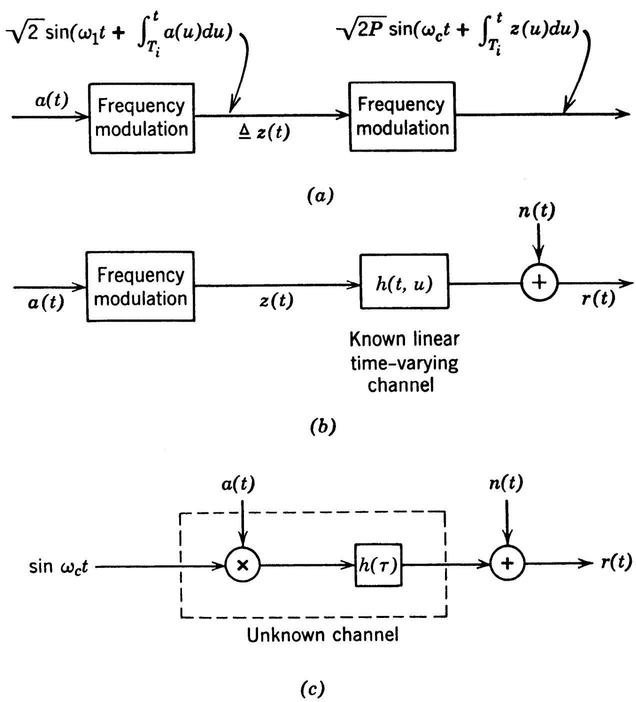

Fig. 5.3 Typical systems: (*a*) an FM/FM system; (*b*) transmission through varying channel; (*c*) channel measurement.

transmitting through a *known* linear time-varying channel. In Fig. 5.3c the channel has an impulse response that depends on the random process $a(t)$. The input is a deterministic signal and we want to estimate the channel impulse response. Measurement problems of this type arise frequently in digital communication systems. A simple example was encountered when we studied the Rayleigh channel in Section 4.4. Other examples will arise in Chapters II.2 and II.3. Note that the channel process is the "message" in this class of problems.

We see that all the problems we have described correspond to the first level in the hierarchy described in Chapter 1. We referred to it as the known signal-in-noise problem. It is important to understand the meaning of this description in the context of continuous waveform estimation. *If* $a(t)$ were known, then $s(t, a(t))$ would be known. In other words, *except* for the additive noise, the mapping from $a(t)$ to $r(t)$ is deterministic.

We shall find that in order to proceed it is expedient to assume that $a(t)$ is a sample function from a Gaussian random process. In many cases this is a valid assumption. In others, such as music or speech, it is *not*. Fortunately, we shall find experimentally that if we use the Gaussian assumption in system design, the system will work well for many non-Gaussian inputs.

The chapter proceeds in the following manner. In Section 5.2 we derive the equations that specify the optimum estimate $\hat{a}(t)$. In Section 5.3 we derive bounds on the mean-square estimation error. In Section 5.4 we extend the results to vector messages and vector received signals. In Section 5.5 we solve the nonrandom waveform estimation problem.

The purpose of the chapter is to develop the necessary equations and to look at some of the properties that can be deduced without solving them. A far more useful end result is the solutions of these equations and the resulting receiver structures. In Chapter 6 we shall study the linear modulation problem in detail. In Chapter II.2 we shall study the nonlinear modulation problem.

5.2 DERIVATION OF ESTIMATOR EQUATIONS

In this section we want to solve the estimation problem for the type of system shown in Fig. 5.1. The general category of interest is defined by the property that the mapping from $a(t)$ to $s(t, a(t))$ is a no-memory transformation.

The received signal is

$$r(t) = s(t, a(t)) + n(t), \qquad T_i \le t \le T_f. \tag{8}$$

By a no-memory transformation we mean that the transmitted signal at some time t_0 depends only on $a(t_0)$ and not on the past of $a(t)$.

5.2.1 No-Memory Modulation Systems. Our specific assumptions are the following:

1. The message $a(t)$ and the noise $n(t)$ are sample functions from independent, continuous, zero-mean Gaussian processes with covariance functions $K_a(t, u)$ and $K_n(t, u)$, respectively.

2. The signal $s(t, a(t))$ has a derivative with respect to $a(t)$. As an example, for the DSB-SC-AM signal in (1) the derivative is

$$\frac{\partial s(t, a(t))}{\partial a(t)} = \sqrt{2P} \sin \omega_c t. \tag{9}$$

Clearly, whenever the transformation $s(t, a(t))$ is a linear transformation, the derivative will not be a function of $a(t)$. We refer to these cases as linear modulation schemes. For PM

$$\frac{\partial s(t, a(t))}{\partial a(t)} = \sqrt{2P}\, \beta \cos(\omega_c t + \beta a(t)). \tag{10}$$

The derivative is a function of $a(t)$. This is an example of a nonlinear modulation scheme. These ideas are directly analogous to the linear signaling and nonlinear signaling schemes in the parameter estimation problem.

As in the parameter estimation case, we must select a suitable criterion. The mean-square error criterion and the maximum a posteriori probability criterion are the two logical choices. Both are conceptually straightforward and lead to identical answers for linear modulation schemes.

For nonlinear modulation schemes both criteria have advantages and disadvantages. In the minimum mean-square error case, if we formulate the a posteriori probability density of $a(t)$ over the interval $[T_i, T_f]$ as a Gaussian-type quadratic form, it is difficult to find an explicit expression for the conditional mean. On the other hand, if we model $a(t)$ as a component of a vector Markov process, we shall see that we can find a differential equation for the conditional mean that represents a formal explicit solution to the problem. This particular approach requires background we have not developed, and we defer it until Chapter II.2. In the maximum a posteriori probability criterion case we are led to an integral equation whose solution is the MAP estimate. This equation provides a simple physical interpretation of the receiver. The MAP estimate will turn out to be asymptotically efficient. Because the MAP formulation is more closely related to our previous work, we shall emphasize it.†

† After we have studied the problem in detail we shall find that in the region in which we get an estimator of practical value the MMSE and MAP estimates coincide.

To help us in solving the waveform estimation problem let us recall some useful facts from Chapter 4 regarding parameter estimation.

In (4.464) and (4.465) we obtained the integral equations that specified the optimum estimates of a set of parameters. We repeat the result. If $a_1, a_2, \ldots, a_K$ are independent zero-mean Gaussian random variables, which we denote by the vector **a**, the MAP estimates $\hat{a}_i$ are given by the simultaneous solution of the equations,

$$\hat{a}_i = \sigma_i^2 \int_{T_i}^{T_f} \frac{\partial s(z, \mathbf{A})}{\partial A_i}\bigg|_{\mathbf{A}=\hat{\mathbf{a}}} [r_g(z) - g(z)]\, dz, \qquad i = 1, 2, \ldots, K, \tag{11}$$

where

$$\sigma_i^2 \triangleq \mathrm{Var}\,(a_i), \tag{12}$$

$$r_g(z) \triangleq \int_{T_i}^{T_f} Q_n(z, u)\, r(u)\, du, \qquad T_i \le z \le T_f, \qquad (13)\dagger$$

$$g(z) \triangleq \int_{T_i}^{T_f} Q_n(z, u)\, s(u, \hat{\mathbf{a}})\, du, \qquad T_i \le z \le T_f, \tag{14}$$

and the received waveform is

$$r(t) = s(t, \mathbf{A}) + n(t). \qquad T_i \le t \le T_f. \tag{15}$$

Now we want to apply this result to our problem. From our work in Chapter 3 we know that we can represent the message, $a(t)$, in terms of an orthonormal expansion:

$$a(t) = \underset{K\to\infty}{\mathrm{l.i.m.}} \sum_{i=1}^{K} a_i\, \psi_i(t), \qquad T_i \le t \le T_f, \tag{16}$$

where the $\psi_i(t)$ are solutions to the integral equation

$$\mu_i\, \psi_i(t) = \int_{T_i}^{T_f} K_a(t, u)\, \psi_i(u)\, du, \qquad T_i \le t \le T_f \tag{17}$$

and

$$a_i = \int_{T_i}^{T_f} a(t)\, \psi_i(t)\, dt. \tag{18}$$

The a_i are independent Gaussian variables:

$$E(a_i) = 0 \tag{19}$$

and

$$\mathrm{E}\,(a_i a_j) = \mu_i\, \delta_{ij}. \tag{20}$$

Now we consider a subclass of processes, those that can be represented by the first K terms in the orthonormal expansion. Thus

$$a_K(t) = \sum_{i=1}^{K} a_i\, \psi_i(t), \qquad T_i \le t \le T_f. \tag{21}$$

† Just as in the colored noise discussions of Chapter 4, the end-points must be treated carefully. Throughout Chapter 5 we shall include the end-points in the interval.

Our logic is to show how the problem of estimating $a_K(t)$ in (21) is identical to the problem we have already solved of estimating a set of K independent parameters. We then let $K \to \infty$ to obtain the desired result. An easy way to see this problem is identical is given in Fig. 5.4*a*. If we look only at the modulator, we may logically write the transmitted signal as

$$s(t, a_K(t)) = s\left(t, \sum_{i=1}^{K} A_i\, \psi_i(t)\right). \tag{22}$$

By grouping the elements as shown in Fig. 5.4*b*, however, we may logically write the output as $s(t, \mathbf{A})$. Clearly the two forms are equivalent:

$$s(t, \mathbf{A}) = s\left(t, \sum_{i=1}^{K} A_i\, \psi_i(t)\right). \tag{23}$$

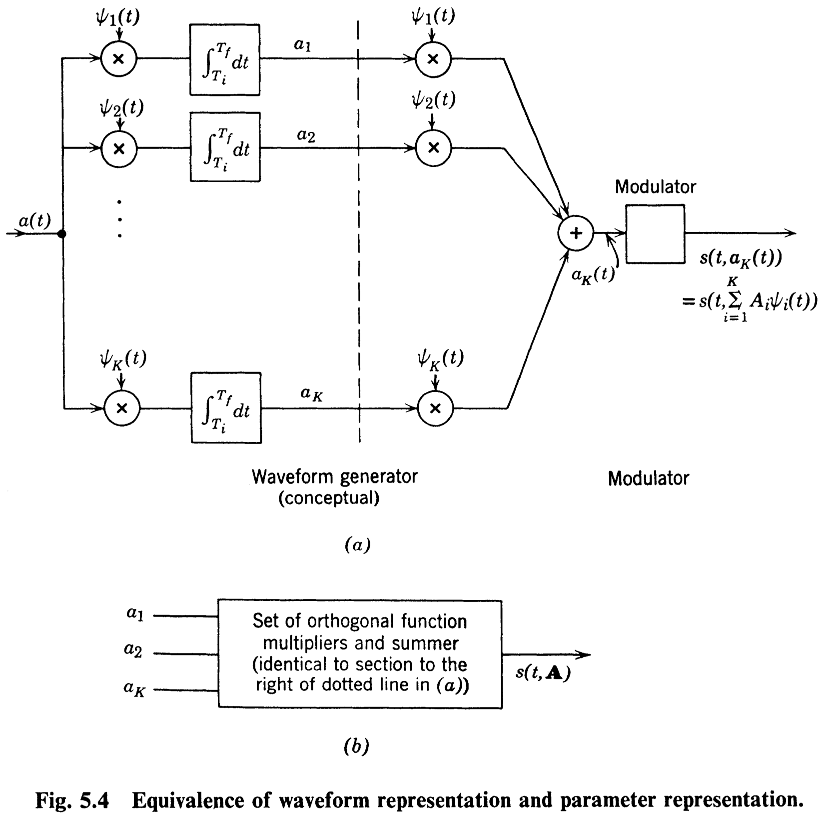

Fig. 5.4 Equivalence of waveform representation and parameter representation.

We *define* the MAP estimate of $a_K(t)$ as

$$\hat{a}_K(t) = \sum_{r=1}^{K} \hat{a}_r\, \psi_r(t), \qquad T_i \le t \le T_f. \tag{24}$$

We see that $\hat{a}_K(t)$ is an *interval estimate*. In other words, we are estimating the *waveform* $a_K(t)$ over the entire interval $T_i \le t \le T_f$ rather than the value at a single instant of time in the interval. To find the estimates of the coefficients we can use (11).

Looking at (22) and (23), we see that

$$\begin{aligned}\frac{\partial s(z, \mathbf{A})}{\partial A_r} &= \frac{\partial s(z, a_K(z))}{\partial A_r} \\ &= \frac{\partial s(z, a_K(z)}{\partial a_K(z)} \cdot \frac{\partial a_K(z)}{\partial A_r} \\ &= \left[\frac{\partial s(z, a_K(z))}{\partial a_K(z)}\right] \psi_r(z),\end{aligned} \tag{25}$$

where the last equality follows from (21). From (11)

$$\hat{a}_r = \mu_r \int_{T_i}^{T_f} \left.\frac{\partial s(z, a_K(z))}{\partial a_K(z)}\right|_{a_K(z) = \hat{a}_K(z)} \psi_r(z)[r_g(z) - g(z)]\, dz, \qquad r = 1, 2, \ldots, K. \tag{26}$$

Substituting (26) into (24) we see that

$$\hat{a}_K(t) = \sum_{r=1}^{K} \psi_r(t)\mu_r \int_{T_i}^{T_f} \left.\frac{\partial s(z, a_K(z))}{\partial a_K(z)}\right|_{a_K(z) = \hat{a}_K(z)} \psi_r(z)[r_g(z) - g(z)]\, dz \tag{27}$$

or

$$\hat{a}_K(t) = \int_{T_i}^{T_f} \left.\frac{\partial s(z, a_K(z))}{\partial a_K(z)}\right|_{a_K(z) = \hat{a}_K(z)} \left[\sum_{r=1}^{K} \mu_r\, \psi_r(t)\, \psi_r(z)\right] [r_g(z) - g(z)]\, dz. \tag{28}$$

In this form it is now easy to let $K \to \infty$. From Mercer's theorem in Chapter 3

$$\lim_{K \to \infty} \sum_{r=1}^{K} \mu_r\, \psi_r(t)\, \psi_r(z) = K_a(t, z). \tag{29}$$

Now define

$$\hat{a}(t) = \underset{K \to \infty}{\text{l.i.m.}}\ \hat{a}_K(t), \qquad T_i \le t \le T_f. \tag{30}$$

The resulting equation is

$$\hat{a}(t) = \int_{T_i}^{T_f} \frac{\partial s(z, \hat{a}(z))}{\partial \hat{a}(z)} K_a(t, z)[r_g(z) - g(z)]\, dz, \qquad T_i \le t \le T_f, \quad (31)\dagger$$

where

$$r_g(z) = \int_{T_i}^{T_f} Q_n(z, u)\, r(u)\, du, \qquad T_i \le z \le T_f, \quad (32)$$

and

$$g(z) = \int_{T_i}^{T_f} Q_n(z, u)\, s(u, \hat{a}(u))\, du, \qquad T_i \le z \le T_f. \quad (33)$$

Equations 31, 32, and 33 specify the MAP estimate of the waveform $a(t)$. These equations (and their generalizations) form the basis of our study of analog modulation theory. For the special case in which the additive noise is white, a much simpler result is obtained. If

$$K_n(t, u) = \frac{N_0}{2}\, \delta(t - u), \tag{34}$$

then

$$Q_n(t, u) = \frac{2}{N_0}\, \delta(t - u). \tag{35}$$

Substituting (35) into (32) and (33), we obtain

$$r_g(z) = \frac{2}{N_0}\, r(z) \tag{36}$$

and

$$g(z) = \frac{2}{N_0}\, s(z, \hat{a}(z)). \tag{37}$$

Substituting (36) and (37) into (31), we have

$$\hat{a}(t) = \frac{2}{N_0} \int_{T_i}^{T_f} K_a(t, z) \frac{\partial s(z, \hat{a}(z))}{\partial \hat{a}(z)} [r(z) - s(z, \hat{a}(z))]\, dz, \qquad T_i \le t \le T_f. \tag{38}$$

Now the estimate is specified by a single nonlinear integral equation.

In the parameter estimation case we saw that it was useful to interpret the integral equation specifying the MAP estimate as a block diagram. This interpretation is even more valuable here. As an illustration, we consider two simple examples.

† The results in (31)–(33) were first obtained by Youla [1]. In order to simplify the notation we have made the substitution

$$\frac{\partial s(z, \hat{a}(z))}{\partial \hat{a}(z)} \triangleq \left.\frac{\partial s(z, a(z))}{\partial a(z)}\right|_{a(z) = \hat{a}(z)}.$$

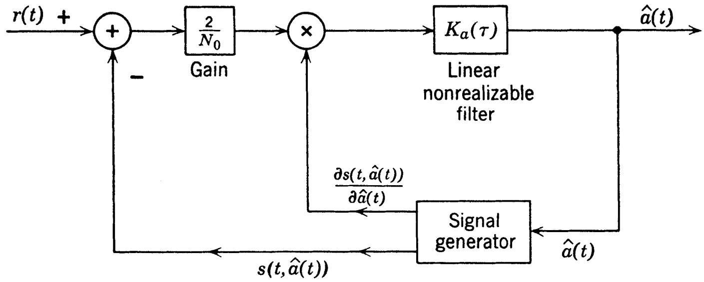

Fig. 5.5 A block diagram of an unrealizable system: white noise.

Example 1. Assume that

$$T_i = -\infty,\ T_f = \infty, \tag{39}$$

$$K_a(t, u) = K_a(t - u), \tag{40}$$

$$K_n(t, u) = \frac{N_0}{2}\,\delta(t - u). \tag{41}$$

In this case (38) is appropriate.† Substituting into (38), we have

$$\hat{a}(t) = \frac{2}{N_0}\int_{-\infty}^{\infty} K_a(t - z)\left\{\frac{\partial s(z, \hat{a}(z))}{\partial \hat{a}(z)}\,[r(z) - s(z, \hat{a}(z))]\right\} dz, \qquad -\infty < t < \infty. \tag{42}$$

We observe that this is simply a convolution of the term inside the braces with a linear filter whose impulse response is $K_a(\tau)$. Thus we can visualize (42) as the block diagram in Fig. 5.5. Observe that the linear filter is *unrealizable*. It is important to emphasize that the block diagram is only a conceptual aid in interpreting (42). It is clearly not a practical solution (in its present form) to the nonlinear integral equation because we cannot build the unrealizable filter. One of the problems to which we shall devote our attention in succeeding chapters is finding a practical approximation to the block diagram.

A second easy example is the nonwhite noise case.

Example 2. Assume that

$$T_i = -\infty,\ T_f = \infty, \tag{43}$$

$$K_a(t, u) = K_a(t - u), \tag{44}$$

$$K_n(t, u) = K_n(t - u). \tag{45}$$

Now (43) and (45) imply that

$$Q_n(t, u) = Q_n(t - u) \tag{46}$$

As in Example 1, we can interpret the integrals (31), (32), and (33) as the block diagram shown in Fig. 5.6. Here, $Q_n(\tau)$ is an unrealizable time-invariant filter.

† For this case we should derive the integral equation by using a spectral representation based on the integrated transform of the process instead of a Karhunen-Loève expansion representation. The modifications in the derivation are straightforward and the result is identical; therefore we relegate the derivation to the problems (see Problem 5.2.6).

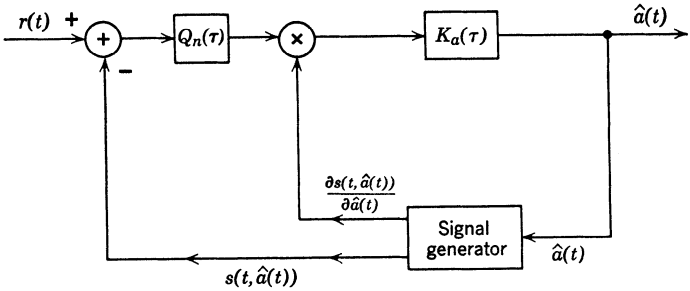

Fig. 5.6 A block diagram of an unrealizable system: colored noise.

Before proceeding we recall that we assumed that the modulator was a no-memory device. This assumption is too restrictive. As we pointed out in Section 1, this assumption excludes such common schemes as FM. While the derivation is still fresh in our minds we can modify it to eliminate this restriction.

5.2.2 Modulation Systems with Memory†

A more general modulation system is shown in Fig. 5.7. The linear system is described by a deterministic impulse response $h(t, u)$. It may be time-varying or unrealizable. Thus we may write

$$x(t) = \int_{T_i}^{T_f} h(t, u)\, a(u)\, du, \qquad T_i \le t \le T_f. \tag{47}$$

The modulator performs a no-memory operation on $x(t)$,

$$s(t, x(t)) = s\left(t, \int_{T_i}^{T_f} h(t, u)\, a(u)\, du\right). \tag{48}$$

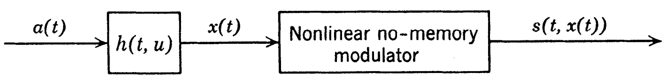

Fig. 5.7 Modulator with memory.

† The extension to include FM is due to Lawton [2], [3]. The extension to arbitrary linear operations is due to Van Trees [4]. Similar results were derived independently by Rauch in two unpublished papers [5], [6].

As an example, for FM,

$$x(t) = d_f \int_{T_i}^{t} a(u)\, du, \tag{49}$$

where d_f is the frequency deviation. The transmitted signal is

$$s(t, x(t)) = \sqrt{2P} \sin \left(\omega_c t + d_f \int_{T_i}^{t} a(u)\, du\right). \tag{50}$$

Looking back at our derivation we see that everything proceeds identically until we want to take the partial derivative with respect to A_r in (25). Picking up the derivation at this point, we have

$$\begin{aligned} \frac{\partial s(z, \mathbf{A})}{\partial A_r} &= \frac{\partial s(z, x_K(z))}{\partial A_r} \\ &= \frac{\partial s(z, x_K(z))}{\partial x_K(z)} \frac{\partial x_K(z)}{\partial A_r}; \end{aligned} \tag{51}$$

but

$$\begin{aligned} \frac{\partial x_K(z)}{\partial A_r} &= \frac{\partial}{\partial A_r} \int_{T_i}^{T_f} h(z, y)\, a_K(y)\, dy \\ &= \frac{\partial}{\partial A_r} \int_{T_i}^{T_f} h(z, y) \sum_{i=1}^{K} A_i\, \psi_i(y)\, dy \\ &= \int_{T_i}^{T_f} h(z, y)\, \psi_r(y)\, dy. \end{aligned} \tag{52}$$

It is convenient to give a label to the output of the linear operation when the input is $\hat{a}(t)$. We define

$$\tilde{x}_K(t) = \int_{T_i}^{T_f} h(t, y)\, \hat{a}_K(y)\, dy. \tag{53}$$

It should be observed that $\tilde{x}_K(t)$ is not defined to be the MAP estimate of $x_K(t)$. In view of our results with finite sets of parameters, we suspect that it is. For the present, however, it is simply a function defined by (53) From (11),

$$\hat{a}_r = \mu_r \int_{T_i}^{T_f} \frac{\partial s(z, \tilde{x}_K(z))}{\partial \tilde{x}_K(z)} \left(\int_{T_i}^{T_f} h(z, y)\, \psi_r(y)\, dy\right)(r_g(z) - g(z))\, dz. \tag{54}$$

As before,

$$\hat{a}_K(t) = \sum_{r=1}^{K} \hat{a}_r\, \psi_r(t), \tag{55}$$

and from (54),

$$\hat{a}_K(t) = \int_{T_i}^{T_f} \frac{\partial s(z, \tilde{x}_K(z))}{\partial \tilde{x}_K(z)} \left\{ \int_{T_i}^{T_f} h(z, y) \left[\sum_{r=1}^{K} \mu_r \, \psi_r(t) \, \psi_r(y) \right] dy \right\} \times [r_g(z) - g(z)] \, dz. \tag{56}$$

Letting $K \rightarrow \infty$, we obtain

$$\boxed{\hat{a}(t) = \iint_{T_i}^{T_f} dy \, dz \frac{\partial s(z, \tilde{x}(z))}{\partial \tilde{x}(z)} h(z, y) \, K_a(t, y)[r_g(z) - g(z)], \qquad T_i \leq t \leq T_f,} \tag{57}$$

where $r_g(z)$ and $g(z)$ were defined in (32) and (33) [replace $\hat{a}(u)$ by $\tilde{x}(u)$]. Equation 57 is similar in form to the no-memory equation (31). If we care to, we can make it identical by performing the integration with respect to y in (57)

$$\int_{T_i}^{T_f} h(z, y) \, K_a(t, y) \, dy \triangleq h_a(z, t) \tag{58}$$

so that

$$\hat{a}(t) = \int_{T_i}^{T_f} dz \frac{\partial s(z, \tilde{x}(z))}{\partial \tilde{x}(z)} h_a(z, t)(r_g(z) - g(z)) \, dz, \qquad T_i \leq t \leq T_f. \tag{59}$$

Thus the block diagram we use to represent the equation is identical in structure to the no-memory diagram given in Fig. 5.8. This similarity in structure will prove to be useful as we proceed in our study of modulation systems.

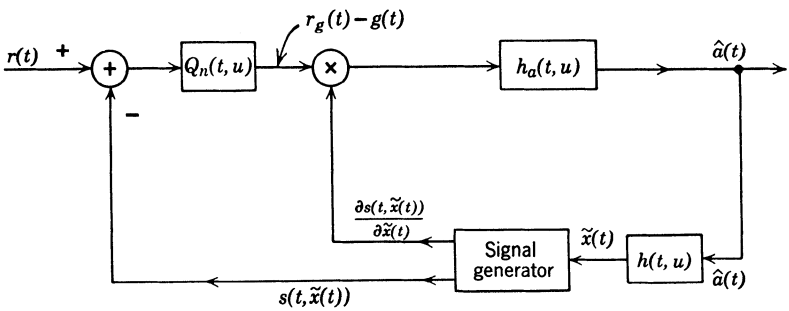

Fig. 5.8 A block diagram.

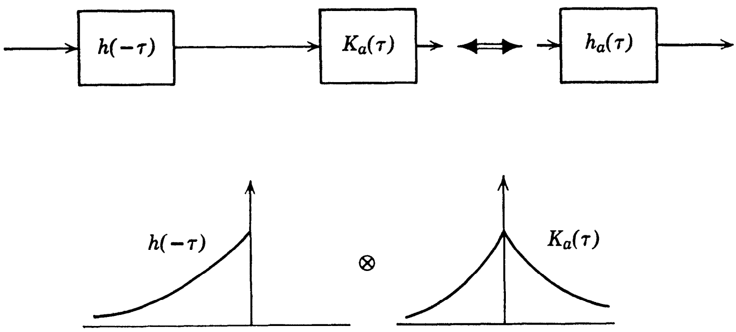

Fig. 5.9 Filter interpretation.

An interesting interpretation of the filter $h_a(z, t)$ can be made for the case in which $T_i = -\infty$, $T_f = \infty$, $h(z, y)$ is time-invariant, and $a(t)$ is stationary. Then

$$h_a(z, t) = \int_{-\infty}^{\infty} h(z - y)\, K_a(t - y)\, dy = h_a(z - t). \tag{60}$$

We see that $h_a(z - t)$ is a cascade of two filters, as shown in Fig. 5.9. The first has an impulse response corresponding to that of the filter in the modulator *reversed in time*. This is familiar in the context of a matched filter. The second filter is the correlation function.

A final question of interest with respect to modulation systems with memory is: Does

$$\tilde{x}(t) = \hat{x}(t)? \tag{61}$$

In other words, is a linear operation on a MAP estimate of a continuous random process equal to the MAP estimate of the output of the linear operation? We shall prove that (61) is true for the case we have just studied. More general cases follow easily. From (53) we have

$$\tilde{x}(\tau) = \int_{T_i}^{T_f} h(\tau, t)\, \hat{a}(t)\, dt. \tag{62}$$

Substituting (57) into (62), we obtain

$$\tilde{x}(\tau) = \int_{T_i}^{T_f} \frac{\partial s(z, \tilde{x}(z))}{\partial \tilde{x}(z)} [r_g(z) - g(z)] \times \left[\iint_{T_i}^{T_f} h(\tau, t)\, h(z, y)\, K_a(t, y)\, dt\, dy \right] dz. \tag{63}$$

We now want to write the integral equation that specifies $\hat{x}(\tau)$ and compare it with the right-hand side of (63). The desired equation is identical to (31), with $a(t)$ replaced by $x(t)$. Thus

$$\hat{x}(\tau) = \int_{T_i}^{T_f} \frac{\partial s(z, \hat{x}(z))}{\partial \hat{x}(z)} K_x(\tau, z)[r_g(z) - g(z)]\, dz. \tag{64}$$

We see that $\tilde{x}(\tau) = \hat{x}(\tau)$ if

$$\iint_{T_i}^{T_f} h(\tau, t)\, h(z, y)\, K_a(t, y)\, dt\, dy = K_x(\tau, z), \qquad T_i \le \tau, z \le T_f; \tag{65}$$

but

$$K_x(\tau, z) \triangleq E[x(\tau)\, x(z)] = E\left[\int_{T_i}^{T_f} h(\tau, t)\, a(t)\, dt \int_{T_i}^{T_f} h(z, y)\, a(y)\, dy\right]$$

$$= \iint_{T_i}^{T_f} h(\tau, t)\, h(z, y)\, K_a(t, y)\, dt\, dy, \qquad T_i \le \tau, z \le T_f, \tag{66}$$

which is the desired result. Thus we see that the operations of maximum a posteriori interval estimation and linear filtering commute. This result is one that we might have anticipated from the analogous results in Chapter 4 for parameter estimation.

We can now proceed with our study of the characteristics of MAP estimates of waveforms and the structure of the optimum estimators.

The important results of this section are contained in (31), (38), and (57). An alternate derivation of these equations with a variational approach is given in Problem 5.2.1.

5.3 A LOWER BOUND ON THE MEAN-SQUARE ESTIMATION ERROR†

In our work with estimating finite sets of variables we found that an extremely useful result was the lower bound on the mean-square error that any estimate could have. We shall see that in waveform estimation such a bound is equally useful. In this section we derive a lower bound on the mean-square error that any estimate of a random process can have.

First, define the error waveform

$$a_\epsilon(t) = a(t) - \hat{a}(t) \tag{67}$$

and

$$e_I = \frac{1}{T_f - T_i} \int_{T_i}^{T_f} [a(t) - \hat{a}(t)]^2\, dt = \frac{1}{T} \int_{T_i}^{T_f} a_\epsilon^2(t)\, dt, \tag{68}$$

† This section is based on Van Trees [7].

where $T \triangleq T_f - T_i$. The subscript I emphasizes that we are making an interval estimate. Now e_I is a random variable. We are concerned with its expectation,

$$T\xi_I \triangleq TE(e_I) = E\left\{\int_{T_i}^{T_f} \sum_{i=1}^{\infty} (a_i - \hat{a}_i)\,\psi_i(t) \sum_{j=1}^{\infty} (a_j - \hat{a}_j)\,\psi_j(t)\,dt\right\}. \quad (69)$$

Using the orthogonality of the eigenfunctions, we have

$$\xi_I T = \sum_{i=1}^{\infty} E(a_i - \hat{a}_i)^2. \quad (70)$$

We want to find a lower bound on the sum on the right-hand side. We first consider the sum $\sum_{i=1}^{K} E(a_i - \hat{a}_i)^2$ and then let $K \to \infty$. The problem of bounding the mean-square error in estimating K random variables is familiar to us from Chapter 4.

From Chapter 4 we know that the first step is to find the information matrix $\mathbf{J}_T$, where

$$\mathbf{J}_T = \mathbf{J}_D + \mathbf{J}_P, \quad (71a)$$

$$J_{D_{ij}} = -E\left[\frac{\partial^2 \ln \Lambda(\mathbf{A})}{\partial A_i\, \partial A_j}\right], \quad (71b)$$

and

$$J_{P_{ij}} = -E\left[\frac{\partial^2 \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_i\, \partial A_j}\right]. \quad (71c)$$

After finding $\mathbf{J}_T$, we invert it to obtain $\mathbf{J}_T^{-1}$. Throughout the rest of this chapter we shall always be interested in $\mathbf{J}_T$ so we suppress the subscript T for convenience. The expression for $\ln \Lambda(\mathbf{A})$ is the vector analog to (4.217)

$$\ln \Lambda(\mathbf{A}) = \iint_{T_i}^{T_f} [r(t) - \tfrac{1}{2} s(t, \mathbf{A})]\, Q_n(t, u)\, s(u, \mathbf{A})\, dt\, du \quad (72a)$$

or, in terms of $a_K(t)$,

$$\ln \Lambda(a_K(t)) = \iint_{T_i}^{T_f} [r(t) - \tfrac{1}{2} s(t, a_K(t))]\, Q_n(t, u)\, s(u, a_K(u))\, dt\, du. \quad (72b)$$

From (19) and (20)

$$\ln p_{\mathbf{a}}(\mathbf{A}) = \sum_{i=1}^{K} \left[-\frac{A_i^2}{2\mu_i} - \frac{1}{2} \ln (2\pi)\right]. \quad (72c)$$

Adding (72*b*) and (72*c*) and differentiating with respect to A_i, we obtain

$$\frac{\partial[\ln p_{\mathbf{a}}(\mathbf{A}) + \ln \Lambda(a_K(t))]}{\partial A_i} =$$

$$-\frac{A_i}{\mu_i} + \int_{T_i}^{T_f} dt\, \psi_i(t) \frac{\partial s(t, a_K(t))}{\partial a_K(t)} \int_{T_i}^{T_f} Q_n(t, u)[r(u) - s(u, a_K(u))]\, du. \quad (72d)$$

Differentiating with respect to A_j and including the minus signs, we have

$$J_{ij} = \frac{\delta_{ij}}{\mu_i} + E \int_{T_i}^{T_f}\!\!\int dt\, du\, \psi_i(t)\, \psi_j(u) \frac{\partial s(t, a_K(t))}{\partial a_K(t)} Q_n(t, u) \frac{\partial s(u, a_K(u))}{\partial a_K(u)}$$

$$+ \text{ terms with zero expectation.} \tag{73}$$

Looking at (73), we see that an efficient estimate will exist only when the modulation is linear (see p. 84).

To interpret the first term recall that

$$K_a(t, u) = \sum_{i=1}^{\infty} \mu_i\, \psi_i(t)\, \psi_i(u), \qquad T_i \le t, u \le T_f. \tag{74}$$

Because we are using only K terms, we define

$$K_{a_K}(t, u) \triangleq \sum_{i=1}^{K} \mu_i\, \psi_i(t)\, \psi_i(u), \qquad T_i \le t, u \le T_f. \tag{75}$$

The form of the first term in (73) suggests defining

$$Q_{a_K}(t, u) = \sum_{i=1}^{K} \frac{1}{\mu_i} \psi_i(t)\, \psi_i(u), \qquad T_i \le t, u \le T_f. \tag{76}$$

We observe that

$$\int_{T_i}^{T_f} Q_{a_K}(t, u)\, K_{a_K}(u, z)\, du = \sum_{i=1}^{K} \psi_i(t)\, \psi_i(z), \qquad T_i \le t, z \le T_f. \tag{77}$$

Once again, $Q_{a_K}(t, u)$ is an inverse kernel, but because the message $a(t)$ does not contain a white noise component, the limit of the sum in (76) as $K \to \infty$ will not exist in general. Thus we must eliminate $Q_{a_K}(t, u)$ from our solution before letting $K \to \infty$. Observe that we may write the first term as

$$\frac{\delta_{ij}}{\mu_i} = \int_{T_i}^{T_f}\!\!\int Q_{a_K}(t, u)\, \psi_i(t)\, \psi_j(u)\, dt\, du, \tag{78}$$

so that if we define

$$J_K(t, u) \triangleq Q_{a_K}(t, u) + E\left[\frac{\partial s(t, a_K(t))}{\partial a_K(t)} Q_n(t, u) \frac{\partial s(u, a_K(u))}{\partial a_K(u)}\right] \tag{79}$$

we can write the elements of the $\mathbf{J}$ matrix as

$$J_{ij} = \int_{T_i}^{T_f}\!\!\int J_K(t, u)\, \psi_i(t)\, \psi_j(u)\, dt\, du, \qquad i, j = 1, 2, \ldots, K. \tag{80}$$

Now we find the inverse matrix $\mathbf{J}^{-1}$. We can show (see Problem 5.3.6) that

$$J^{ij} = \iint_{T_i}^{T_f} J_K^{-1}(t, u)\, \psi_i(t)\, \psi_j(u)\, dt\, du \tag{81}$$

where the function $J_K^{-1}(t, u)$ satisfies the equation

$$\int_{T_i}^{T_f} J_K^{-1}(t, u) J_K(u, z)\, du = \sum_{i=1}^{K} \psi_i(t)\, \psi_i(z). \tag{82}$$

(Recall that the superscript ij denotes an element in $\mathbf{J}^{-1}$.) We now want to put (82) into a more usable form.

If we denote the derivative of $s(t, a_K(t))$ with respect to $a_K(t)$ as $d_s(t, a_K(t))$, then

$$E\left[\frac{\partial s(t, a_K(t))}{\partial a_K(t)} \frac{\partial s(u, a_K(u))}{\partial a_K(u)}\right] = E[d_s(t, a_K(t))\, d_s(u, a_K(u))]$$
$$\triangleq R_{d_s K}(t, u). \tag{83a}$$

Similarly

$$E\left[\frac{\partial s(t, a(t))}{\partial a(t)} \frac{\partial s(u, a(u))}{\partial a(u)}\right] = E[d_s(t, a(t))\, d_s(u, a(u))] \triangleq R_{d_s}(t, u). \tag{83b}$$

Therefore

$$J_K(u, z) \triangleq Q_{a_K}(u, z) + R_{d_s K}(u, z)\, Q_n(u, z). \tag{84}$$

Substituting (84) into (82), multiplying by $K_{a_K}(z, x)$, integrating with respect to z, and letting $K \to \infty$, we obtain the following integral equation for $J^{-1}(t, x)$,

$$J^{-1}(t, x) + \int_{T_i}^{T_f} du \int_{T_i}^{T_f} dz\, J^{-1}(t, u)\, R_{d_s}(u, z)\, Q_n(u, z)\, K_a(z, x) = K_a(t, x), \qquad T_i \le t, x \le T_f. \tag{85}$$

From (2.292) we know that the diagonal elements of $\mathbf{J}^{-1}$ are lower bounds on the mean-square errors. Thus

$$E[(a_i - \hat{a}_i)^2] \ge J^{ii}. \tag{86a}$$

Using (81) in (86*a*) and the result in (70), we have

$$T\xi_I \ge \lim_{K \to \infty} \sum_{i=1}^{K} \iint_{T_i}^{T_f} J_K^{-1}(t, u)\, \psi_i(t)\, \psi_i(u)\, dt\, du \tag{86b}$$

or, using (3.128),

$$\xi_I \ge \frac{1}{T} \int_{T_i}^{T_f} J^{-1}(t, t)\, dt. \tag{87}$$

Therefore to evaluate the lower bound we must solve (85) for $J^{-1}(t, x)$ and evaluate its trace. By analogy with the classical case, we refer to $J(t, x)$ as the *information kernel.*

We now want to interpret (85). First consider the case in which there is only a white noise component so that,

$$Q_n(t, u) = \frac{2}{N_0}\,\delta(t - u). \tag{88}$$

Then (85) becomes

$$J^{-1}(t, x) + \int_{T_i}^{T_f} du\,\frac{2}{N_0}\,J^{-1}(t, u)\,R_{d_s}(u, u)\,K_a(u, x) = K_a(t, x),$$
$$T_i \leq t, x \leq T_f. \tag{89}$$

The succeeding work will be simplified if $R_{d_s}(t, t)$ is a constant. A sufficient, but not necessary, condition for this to be true is that $d_s(t, a(t))$ be a sample function from a stationary process. We frequently encounter estimation problems in which we can approximate $R_{d_s}(t, t)$ with a constant without requiring $d_s(t, a(t))$ to be stationary. A case of this type arises when the transmitted signal is a bandpass waveform having a spectrum centered around a carrier frequency ω_c; for example, in PM,

$$s(t, a(t)) = \sqrt{2P}\sin\left[\omega_c t + \beta a(t)\right]. \tag{90}$$

Then

$$d_s(t, a(t)) = \frac{\partial s(t, a(t))}{\partial a(t)} = \sqrt{2P}\,\beta\cos\left[\omega_c t + \beta\, a(t)\right] \tag{91}$$

and

$$R_{d_s}(t, u) = \beta^2 P E_a\{\cos\left[\omega_c(t - u) + \beta\, a(t) - \beta\, a(u)\right]$$
$$+ \cos\left[\omega_c(t + u) + \beta\, a(t) + \beta\, a(u)\right]\}. \tag{92}$$

Letting $u = t$, we observe that

$$R_{d_s}(t, t) = \beta^2 P(1 + E_a\{\cos\left[2\omega_c t + 2\beta\, a(t)\right]\}). \tag{93}$$

We assume that the frequencies contained in $a(t)$ are low relative to ω_c. To develop the approximation we fix t in (89). Then (89) can be represented as the linear time-varying system shown in Fig. 5.10. The input is a function of u, $J^{-1}(t, u)$. Because $K_a(u, x)$ corresponds to a low-pass filter and $K_a(t, x)$, a low-pass function, we see that $J^{-1}(t, x)$ must be low-pass and the double-frequency term in $R_{d_s}(u, u)$ may be neglected. Thus we can make the approximation

$$R_{d_s}(t, t) \simeq \beta^2 P \simeq R_{d_s}^*(0) \tag{94}$$

to solve the integral equation. The function $R_{d_s}^*(0)$ is simply the stationary component of $R_{d_s}(t, t)$. In this example it is the low-frequency component.

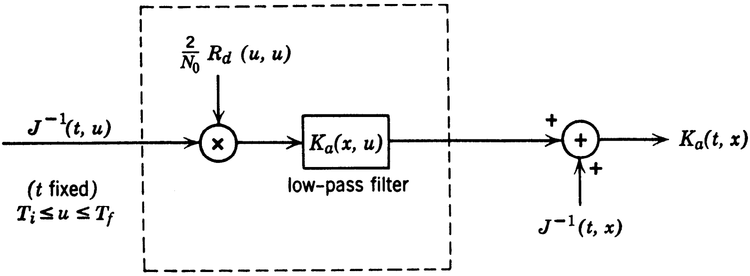

Fig. 5.10 Linear system interpretation.

(Note that $R_{d_s}^*(0) = R_{d_s}(0)$ when $d_s(t, a(t))$ is stationary.) For the cases in which (94) is valid, (89) becomes

$$J^{-1}(t, x) + \frac{2R_{d_s}^*(0)}{N_0} \int_{T_i}^{T_f} du\, J^{-1}(t, u)\, K_a(u, x) = K_a(t, x), \quad T_i \leq t, x \leq T_f, \tag{95}$$

which is an integral equation whose solution is the desired function. From the mathematical viewpoint, (95) is an adequate final result.

We can, however, obtain a very useful physical interpretation of our result by observing that (95) is familiar in a different context. Recall the following *linear* filter problem (see Chapter 3, p. 198).

$$r(t) = a(t) + n_1(t), \qquad T_i \leq t \leq T_f, \tag{96}$$

where $a(t)$ is the same as our message and $n_1(t)$ is a sample function from a white noise process ($N_1/2$, double-sided). We want to design a linear filter $h(t, x)$ whose output is the estimate of $a(t)$ which minimizes the mean-square error. This is the problem that we solved in Chapter 3. The equation that specifies the filter $h_o(t, x)$ is

$$\frac{N_1}{2} h_o(t, x) + \int_{T_i}^{T_f} h_o(t, u)\, K_a(u, x)\, du = K_a(t, x), \qquad T_i \leq t, x \leq T_f, \tag{97}$$

where $N_1/2$ is the height of the white noise. We see that if we let

$$N_1 = \frac{N_0}{R_{d_s}^*(0)} \tag{98}$$

then

$$J^{-1}(t, x) = \frac{N_0}{2R_{d_s}^*(0)} h_o(t, x). \tag{99}$$

The error in the linear filtering problem is

$$\xi_I = \frac{1}{T}\int_{T_i}^{T_f} \frac{N_1}{2} h_o(x, x)\, dx = \frac{1}{T}\int_{T_i}^{T_f} J^{-1}(x, x)\, dx, \tag{100}$$

Our bound for the nonlinear modulation problem corresponds to the mean-square error in the linear filter problem except that the noise level is *reduced* by a factor $R_{d_s}^*(0)$.

The quantity $R_{d_s}^*(0)$ may be greater or less than one. We shall see in Examples 1 and 2 that in the case of linear modulation we can increase $R_{d_s}(0)$ only by *increasing* the transmitted power. In Example 3 we shall see that in a nonlinear modulation scheme such as phase modulation we can increase $R_{d_s}^*(0)$ by increasing the modulation index. This result corresponds to the familiar PM improvement.

It is easy to show a similar interpretation for colored noise. First define an effective noise whose inverse kernel is,

$$Q_{ne}(t, u) = R_{d_s}(t, u)\, Q_n(t, u). \tag{101}$$

Its covariance function satisfies the equation

$$\int_{T_i}^{T_f} K_{ne}(t, u)\, Q_{ne}(u, z)\, du = \delta(t - z), \qquad T_i < t, z < T_f. \tag{102}$$

Then we can show that

$$J^{-1}(u, z) = \int_{T_i}^{T_f} K_{ne}(u, x)\, h_o(x, z)\, dx, \qquad T_i \le u, z \le T_f \tag{103}$$

where $h_o(x, z)$ is the solution to,

$$\int_{T_i}^{T_f} [K_a(x, t) + K_{ne}(x, t)]\, h_o(t, z)\, dt = K_a(x, z), \quad T_i \le x, z \le T_f. \tag{104}$$

This is the colored noise analog to (95).

Two special but important cases lead to simpler expressions.

Case 1. $\mathbf{J^{-1}(t, u) = J^{-1}(t - u)}$. Observe that when $J^{-1}(t, u)$ is a function only of the difference of its two arguments

$$J^{-1}(t, t) = J^{-1}(t - t) = J^{-1}(0). \tag{105}$$

Then (87) becomes,

$$\xi_I \ge J^{-1}(0). \tag{106}$$

If we define

$$\mathfrak{J}^{-1}(\omega) = \int_{-\infty}^{\infty} J^{-1}(\tau) e^{-j\omega\tau}\, d\tau, \tag{107}$$

then

$$\xi_I = \int_{-\infty}^{\infty} \mathfrak{J}^{-1}(\omega) \frac{d\omega}{2\pi}. \tag{108}$$

A further simplification develops when the observation interval includes the infinite past and future.

Case 2. Stationary Processes, Infinite Interval.† Here, we assume

$$T_i = -\infty, \tag{109}$$

$$T_f = \infty, \tag{110}$$

$$K_a(t, u) = K_a(t - u), \tag{111}$$

$$K_n(t, u) = K_n(t - u), \tag{112}$$

$$R_{d_s}(t, u) = R_{d_s}(t - u). \tag{113}$$

Then

$$J^{-1}(t, u) = J^{-1}(t - u). \tag{114}$$

The transform of $J(\tau)$ is

$$\mathfrak{J}(\omega) = \int_{-\infty}^{\infty} J(\tau)e^{-j\omega\tau}\,d\tau. \tag{115}$$

Then, from (82) and (85),

$$\mathfrak{J}^{-1}(\omega) = \frac{1}{\mathfrak{J}(\omega)} = \left[\frac{1}{S_a(\omega)} + S_{d_s}(\omega) \otimes \frac{1}{S_n(\omega)}\right]^{-1} \tag{116}$$

(where $\otimes$ denotes convolution‡) and the resulting error is

$$\xi_I \geq \int_{-\infty}^{\infty} \left[\frac{1}{S_a(\omega)} + S_{d_s}(\omega) \otimes \frac{1}{S_n(\omega)}\right]^{-1} \frac{d\omega}{2\pi}. \tag{117}$$

Several simple examples illustrate the application of the bound.

Example 1. We assume that Case 2 applies. In addition, we assume that

$$s(t, a(t)) = a(t). \tag{118}$$

Because the modulation is linear, an efficient estimate exists. There is no carrier so $\partial s(t, a(t))/\partial a(t) = 1$ and

$$S_{d_s}(\omega) = 2\pi\delta(\omega). \tag{119}$$

Substituting into (117), we obtain

$$\xi_I = \int_{-\infty}^{\infty} \frac{S_a(\omega)\, S_n(\omega)}{S_a(\omega) + S_n(\omega)} \frac{d\omega}{2\pi}. \tag{120}$$

The expression on the right-hand side of (120) will turn out to be the minimum mean-square error with an unrealizable linear filter (Chapter 6). Thus, as we would expect, the *efficient* estimate is obtained by processing $r(t)$ with a linear filter.

A second example is linear modulation onto a sinusoid.

Example 2. We assume that Case 2 applies and that the carrier is amplitude-modulated by the message,

$$s(t, a(t)) = \sqrt{2P}\, a(t) \sin \omega_c t, \tag{121}$$

† See footnote on p. 432 and Problem 5.3.3.

‡ We include $1/2\pi$ in the convolution operation when ω is the variable.

where $a(t)$ is low-pass compared with ω_c. The derivative is,

$$\frac{\partial s(t, a(t))}{\partial a(t)} = \sqrt{2P} \sin \omega_c t. \tag{122}$$

For simplicity, we assume that the noise has a flat spectrum whose bandwidth is much larger than that of $a(t)$. It follows easily that

$$\xi_I = \int_{-\infty}^{\infty} \frac{S_a(\omega)}{1 + S_a(\omega)(2P/N_0)} \frac{d\omega}{2\pi}. \tag{123}$$

We can verify that an estimate with this error can be obtained by multiplying $r(t)$ by $\sqrt{2/P} \sin \omega_c t$ and passing the output through the same linear filter as in Example 1. Thus once again an *efficient* estimate exists and is obtained by using a linear system at the receiver.

Example 3. Consider a phase-modulated sine wave in additive white noise. Assume that $a(t)$ is stationary. Thus

$$s(t, a(t)) = \sqrt{2P} \sin [\omega_c t + \beta\, a(t)], \tag{124}$$

$$\frac{\partial s(t, a(t))}{\partial a(t)} = \sqrt{2P}\, \beta \cos [\omega_c t + \beta\, a(t)] \tag{125}$$

and

$$K_n(t, u) = \frac{N_0}{2} \delta(t-u). \tag{126}$$

Then, using (92), we see that

$$R_{d_s}^*(0) = P\beta^2. \tag{127}$$

* By analogy with Example 2 we have

$$\xi_I \geq \int_{-\infty}^{\infty} \frac{S_a(\omega)}{1 + S_a(\omega)(2P\beta^2/N_0)} \frac{d\omega}{2\pi}. \tag{128}$$

In linear modulation the error was only a function of the spectrum of the message, the transmitted power and the white noise level. For a given spectrum and noise level the only way to decrease the mean-square error is to increase the transmitted power. In the nonlinear case we see that by increasing β, the modulation index, we can decrease the bound on the mean-square error. We shall show that as P/N_0 is increased the mean-square error of a MAP estimate approaches the bound given by (128). Thus the MAP estimate is asymptotically efficient. On the other hand, if β is large and P/N_0 is decreased, any estimation scheme will exhibit a "threshold." At this point the estimation error will increase rapidly and the bound will no longer be useful. This result is directly analogous to that obtained for parameter estimation (Example 2, Section 4.2.3). We recall that if we tried to make β too large the result obtained by considering the local estimation problem was meaningless. In Chapter II.2, in which we discuss nonlinear modulation in more detail, we shall see that an analogous phenomenon occurs. We shall also see that for large signal-to-noise ratios the mean-square error approaches the value given by the bound.

The principal results of this section are (85), (95), and (97). The first equation specifies $J^{-1}(t, x)$, the inverse of the information kernel. The trace of this inverse kernel provides a lower bound on the mean-square interval error in continuous waveform estimation. This is a generalization of the classical Cramér-Rao inequality to random processes. The second equation is a special case of (85) which is valid when the additive noise is white and the component of $d_s(t, a(t))$ which affects the integral equation is stationary. The third equation (97) shows how the bound on the mean-square interval estimation error in a nonlinear system is identical to the actual mean-square interval estimation error in a linear system whose white noise level is divided by $R_{\dot{d}_s}^*(0)$.

In our discussion of detection and estimation we saw that the receiver often had to process multiple inputs. Similar situations arise in the waveform estimation problem.

5.4 MULTIDIMENSIONAL WAVEFORM ESTIMATION

In Section 4.5 we extended the detection problem to M received signals. In Problem 4.5.4 of Chapter 4 it was demonstrated that an analogous extension could be obtained for linear and nonlinear estimation of a single parameter. In Problem 4.6.7 of Chapter 4 a similar extension was obtained for multiple parameters. In this section we shall estimate N continuous messages by using M received waveforms. As we would expect, the derivation is a simple combination of those in Problems 4.6.7 and Section 5.2.

It is worthwhile to point out that all one-dimensional *concepts* carry over directly to the multidimensional case. We can almost guess the form of the particular results. Thus most of the interest in the multidimensional case is based on the solution of these equations for actual physical problems. It turns out that many issues not encountered in the scalar case must be examined. We shall study these issues and their implications in detail in Chapter II.5. For the present we simply derive the equations that specify the MAP estimates and indicate a bound on the mean-square errors.

Before deriving these equations, we shall find it useful to discuss several physical situations in which this kind of problem occurs.

5.4.1 Examples of Multidimensional Problems

Case 1. Multilevel Modulation Systems. In many communication systems a number of messages must be transmitted simultaneously. In one common method we perform the modulation in two steps. First, each of the

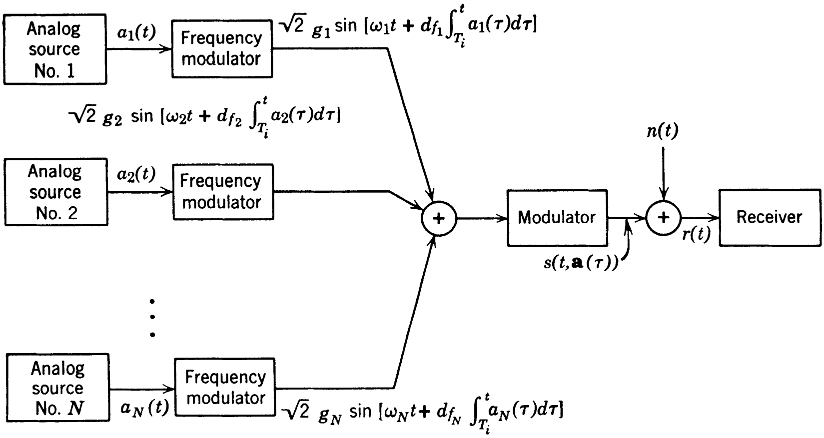

Fig. 5.11 An FM/FM system.

messages is modulated onto individual subcarriers. The modulated subcarriers are then summed and the result is modulated onto a main carrier and transmitted. A typical system is the FM/FM system shown in Fig. 5.11, in which each message $a_i(t)$, $(i = 1, 2, \ldots, N)$, is frequency-modulated onto a sine wave of frequency ω_i. The ω_i are chosen so that the modulated subcarriers are in disjoint frequency bands. The modulated subcarriers are amplified, summed and the result is frequency-modulated onto a main carrier and transmitted.

Notationally, it is convenient to denote the N messages by a column matrix,

$$\mathbf{a}(\tau) \triangleq \begin{bmatrix} a_1(\tau) \\ a_2(\tau) \\ \vdots \\ a_N(\tau) \end{bmatrix}. \tag{129}$$

Using this notation, the transmitted signal is

$$s(t, \mathbf{a}(\tau)) = \sqrt{2P} \sin \left[\omega_c t + d_{f_c} \int_{T_i}^{t} z(u)\, du\right], \tag{130†}$$

where

$$z(u) = \sum_{j=1}^{N} \sqrt{2}\, g_j \sin \left[\omega_j u + d_{f_j} \int_{T_i}^{u} a_j(\tau)\, d\tau\right]. \tag{131}$$

† The notation $s(t, \mathbf{a}(\tau))$ is an abbreviation for $s(t; \mathbf{a}(\tau), T_i \le \tau \le t)$. The second variable emphasizes that the modulation process has memory.

The channel adds noise to the transmitted signal so that the received waveform is

$$r(t) = s(t, \mathbf{a}(\tau)) + n(t). \tag{132}$$

Here we want to estimate the N messages simultaneously. Because there are N messages and one received waveform, we refer to this as an $N \times 1$-dimensional problem.

FM/FM is typical of many possible multilevel modulation systems such as SSB/FM, AM/FM, and PM/PM. The possible combinations are essentially unlimited. A discussion of schemes currently in use is available in [8].

Case 2. Multiple-Channel Systems. In Section 4.5 we discussed the use of diversity systems for digital communication systems. Similar systems can be used for analog communication. Figure 5.12 in which the message $a(t)$ is frequency-modulated onto a set of carriers at different frequencies is typical. The modulated signals are transmitted over separate channels, each of which attenuates the signal and adds noise. We see that there are M received waveforms,

$$r_i(t) = s_i(t, a(\tau)) + n_i(t), \qquad (i = 1, 2, \dots, M), \tag{133}$$

where

$$s_i(t, a(\tau)) = g_i\sqrt{2P_i}\sin\left(\omega_c t + d_{f_i}\int_{T_i}^{t} a(u)\,du\right). \tag{134}$$

Once again matrix notation is convenient. We define

$$\mathbf{s}(t, a(\tau)) = \begin{bmatrix} s_1(t, a(\tau)) \\ s_2(t, a(\tau)) \\ \vdots \\ s_M(t, a(\tau)) \end{bmatrix} \tag{135}$$

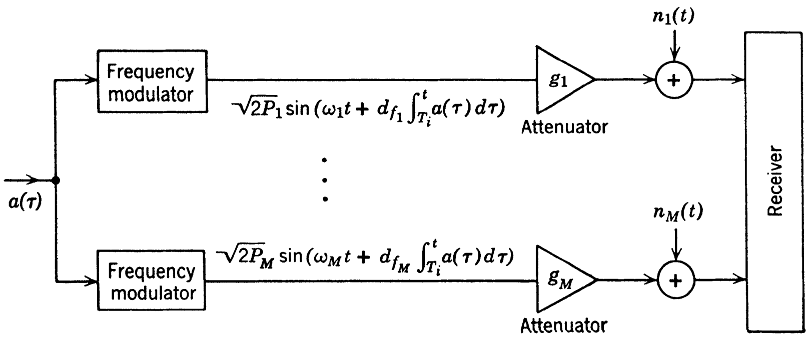

Fig. 5.12 Multiple channel system.

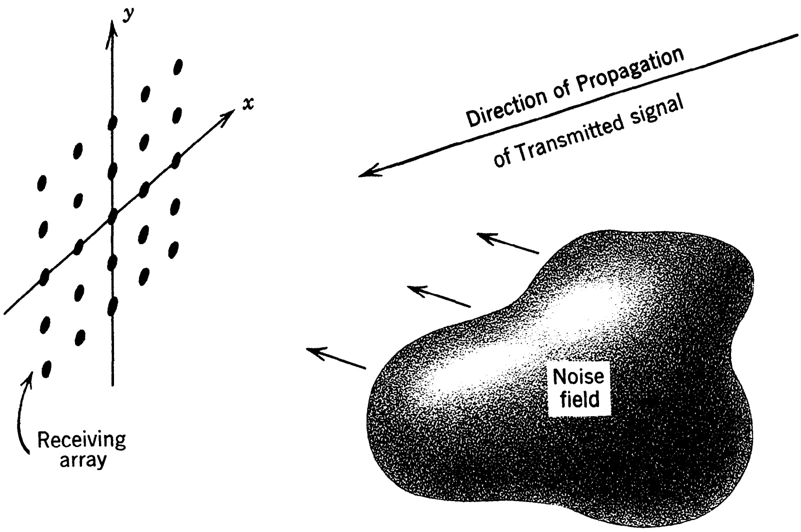

Fig. 5.13 A space-time system.

and

$$\mathbf{n}(t) = \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix}. \tag{136}$$

Then

$$\mathbf{r}(t) = \mathbf{s}(t, a(\tau)) + \mathbf{n}(t). \tag{137}$$

Here there is one message $a(t)$ to estimate and M waveforms are available to perform the estimation. We refer to this as a $1 \times M$-dimensional problem. The system we have shown is a frequency diversity system. Other obvious forms of diversity are space and polarization diversity.

A physical problem that is essentially a diversity system is discussed in the next case.

Case 3. A Space-Time System. In many sonar and radar problems the receiving system consists of an array of elements (Fig. 5.13). The received signal at the ith element consists of a signal component, $s_i(t, a(\tau))$, an external noise term $n_{Ei}(t)$, and a term due to the noise in the receiver element, $n_{Ri}(t)$. Thus the total received signal at the ith element is

$$r_i(t) = s_i(t, a(\tau)) + n_{Ri}(t) + n_{Ei}(t). \tag{138}$$

We define

$$n_i(t) = n_{Ri}(t) + n_{Ei}(t). \tag{139}$$

We see that this is simply a different physical situation in which we have M waveforms available to estimate a single message. Thus once again we have a $1 \times M$-dimension problem.

Case 4. $N \times M$-Dimensional Problems. If we take any of the multilevel modulation schemes of Case 1 and transmit them over a diversity channel, it is clear that we will have an $N \times M$-dimensional estimation problem. In this case the ith received signal, $r_i(t)$, has a component that depends on N messages, $a_j(t)$, $(j = 1, 2, \dots, N)$. Thus

$$r_i(t) = s_i(t, \mathbf{a}(\tau)) + n_i(t), \qquad i = 1, 2, \dots, M. \tag{140}$$

In matrix notation

$$\mathbf{r}(t) = \mathbf{s}(t, \mathbf{a}(\tau)) + \mathbf{n}(t). \tag{141}$$

These cases serve to illustrate the types of physical situations in which multidimensional estimation problems appear. We now formulate the model in general terms.

5.4.2 Problem Formulation†

Our first assumption is that the messages $a_i(t)$, $(i = 1, 2, \dots, N)$, are sample functions from continuous, jointly Gaussian random processes. It is convenient to denote this set of processes by a single vector process $\mathbf{a}(t)$. (As before, we use the term vector and column matrix interchangeably.) We assume that the vector process has a zero mean. Thus it is completely characterized by an $N \times N$ covariance matrix,

$$\mathbf{K_a}(t, u) \triangleq E[(\mathbf{a}(t)\, \mathbf{a}^T(u))]$$

$$= \begin{bmatrix} K_{a_1 a_1}(t, u) & K_{a_1 a_2}(t, u) & \cdots & K_{a_1 a_N}(t, u) \\ \vdots & & & \vdots \\ K_{a_N a_1}(t, u) & & \cdots & K_{a_N a_N}(t, u) \end{bmatrix}. \tag{142}$$

Thus the ijth element represents the cross-covariance function between the ith and jth messages.

The transmitted signal can be represented as a vector $\mathbf{s}(t, \mathbf{a}(\tau))$. This vector signal is deterministic in the sense that if a particular vector sample

† The multidimensional problem for no-memory signaling schemes and additive channels was first done in [9]. (See also [10].)

function $\mathbf{a}(\tau)$ is given, $\mathbf{s}(t, \mathbf{a}(\tau))$ will be uniquely determined. The transmitted signal is corrupted by an additive Gaussian noise $\mathbf{n}(t)$. The signal available to the receiver is an M-dimensional vector signal $\mathbf{r}(t)$,

$$\mathbf{r}(t) = \mathbf{s}(t, \mathbf{a}(\tau)) + \mathbf{n}(t), \qquad T_i \leq t \leq T_f, \tag{143}$$

or

$$\begin{bmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_M(t) \end{bmatrix} = \begin{bmatrix} s_1(t, \mathbf{a}(\tau)) \\ s_2(t, \mathbf{a}(\tau)) \\ \vdots \\ s_M(t, \mathbf{a}(\tau)) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix}, \qquad T_i \leq t \leq T_f. \tag{144}$$

The general model is shown in Fig. 5.14.

We assume that the M noise waveforms are sample functions from zero-mean jointly Gaussian random processes and that the messages and noises are statistically independent. (Dependent messages and noises can easily be included, cf. Problem 5.4.1.) We denote the M noises by a vector noise process $\mathbf{n}(t)$ which is completely characterized by an $M \times M$ covariance matrix $\mathbf{K_n}(t, u)$.

5.4.3 Derivation of Estimator Equations.

We now derive the equations for estimating a vector process. For simplicity we shall do only the no-memory modulation case here. Other cases are outlined in the problems.

The received signal is

$$\mathbf{r}(t) = \mathbf{s}(t, \mathbf{a}(t)) + \mathbf{n}(t), \qquad T_i \leq t \leq T_f, \tag{145}$$

where $\mathbf{s}(t, \mathbf{a}(t))$ is obtained by a no-memory transformation on the vector $\mathbf{a}(t)$. We also assume that $\mathbf{s}(t, \mathbf{a}(t))$ is differentiable with respect to each $a_i(t)$.

The first step is to expand $\mathbf{a}(t)$ in a vector orthogonal expansion.

$$\mathbf{a}(t) = \underset{K \to \infty}{\text{l.i.m.}} \sum_{r=1}^{K} a_r \boldsymbol{\psi}_r(t), \qquad T_i \leq t \leq T_f, \tag{146}$$

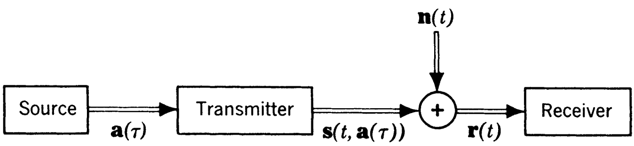

Fig. 5.14 The vector estimation model.

or

$$a_i(t) = \underset{K\to\infty}{\text{l.i.m.}} \sum_{r=1}^{K} a_r\, \psi_r^{(i)}(t), \qquad T_i \le t \le T_f, \tag{147}$$

where the $\boldsymbol{\psi}_r(t)$ are the vector eigenfunctions corresponding to the integral equation

$$\mu_k \boldsymbol{\psi}_k(t) = \int_{T_i}^{T_f} \mathbf{K_a}(t, u)\boldsymbol{\psi}_k(u)\, du, \qquad T_i \le t \le T_f. \tag{148}$$

This expansion was developed in detail in Section 3.7. Then we find $\hat{a}_r$ and define

$$\hat{\mathbf{a}}(t) = \underset{K\to\infty}{\text{l.i.m.}} \sum_{r=1}^{K} \hat{a}_r \boldsymbol{\psi}_r(t). \tag{149}$$

We have, however, already solved the problem of estimating a parameter a_r. By analogy to (72d) we have

$$\frac{\partial\,[\ln \Lambda(\mathbf{A}) + \ln p_{\mathbf{a}}(\mathbf{A})]}{\partial A_r} = \int_{T_i}^{T_f} \frac{\partial \mathbf{s}^T(z, \mathbf{a}(z))}{\partial A_r}\, [\mathbf{r_g}(z) - \mathbf{g}(z)]\, dz - \frac{A_r}{\mu_r}, \qquad (r = 1, 2, \dots), \tag{150}$$

where

$$\mathbf{r_g}(z) \triangleq \int_{T_i}^{T_f} \mathbf{Q_n}(z, u)\, \mathbf{r}(u)\, du \tag{151}$$

and

$$\mathbf{g}(z) \triangleq \int_{T_i}^{T_f} \mathbf{Q_n}(z, u)\, \mathbf{s}(u, \mathbf{a}(u))\, du. \tag{152}$$

The left matrix in the integral in (150) is

$$\frac{\partial \mathbf{s}^T(t, \mathbf{a}(t))}{\partial A_r} = \left[\frac{\partial s_1(t, \mathbf{a}(t))}{\partial A_r} \,\vdots \cdots \vdots\, \frac{\partial s_M(t, \mathbf{a}(t))}{\partial A_r}\right]. \tag{153}$$

The first element in the matrix can be written as

$$\frac{\partial s_1(t, \mathbf{a}(t))}{\partial A_r} = \frac{\partial s_1(t, \mathbf{a}(t))}{\partial a_1(t)}\, \psi_r^{(1)}(t) + \frac{\partial s_1(t, \mathbf{a}(t))}{\partial a_2(t)}\, \psi_r^{(2)}(t) + \cdots + \frac{\partial s_1(t, \mathbf{a}(t))}{\partial a_N(t)}\, \psi_r^{(N)}(t). \tag{154}$$

Looking at the other elements, we see that if we define a derivative matrix

$$\mathbf{D}(t, \mathbf{a}(t)) \triangleq \nabla_{\mathbf{a}(t)}\{\mathbf{s}^T(t, \mathbf{a}(t))\} = \begin{bmatrix} \dfrac{\partial s_1(t, \mathbf{a}(t))}{\partial a_1(t)} & \cdots & \dfrac{\partial s_M(t, \mathbf{a}(t))}{\partial a_1(t)} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial s_1(t, \mathbf{a}(t))}{\partial a_N(t)} & \cdots & \dfrac{\partial s_M(t, \mathbf{a}(t))}{\partial a_N(t)} \end{bmatrix} \tag{155}$$

we may write

$$\frac{\partial \mathbf{s}^T(t, \mathbf{a}(t))}{\partial A_r} = \boldsymbol{\psi}_r^T(t)\, \mathbf{D}(t, \mathbf{a}(t)), \tag{156}$$

so that

$$\frac{\partial\,[\ln \Lambda(\mathbf{A}) + \ln p_{\mathbf{a}}(\mathbf{A})]}{\partial A_r} = \int_{T_i}^{T_f} \boldsymbol{\psi}_r^T(z)\, \mathbf{D}(z, \mathbf{a}(z))[\mathbf{r}_g(z) - \mathbf{g}(z)]\, dz - \frac{A_r}{\mu_r}. \tag{157}$$

Equating the right-hand side to zero, we obtain a necessary condition on the MAP estimate of A_r. Thus

$$\hat{a}_r = \mu_r \int_{T_i}^{T_f} \boldsymbol{\psi}_r^T(z)\, \mathbf{D}(z, \hat{\mathbf{a}}(z))[\mathbf{r}_g(z) - \mathbf{g}(z)]\, dz. \tag{158}$$

Substituting (158) into (149), we obtain

$$\hat{\mathbf{a}}(t) = \int_{T_i}^{T_f} \left[\sum_{r=1}^{\infty} \mu_r \boldsymbol{\psi}_r(t) \boldsymbol{\psi}_r^T(z)\right] \mathbf{D}(z, \hat{\mathbf{a}}(z))[\mathbf{r}_g(z) - \mathbf{g}(z)]\, dz. \tag{159}$$

We recognize the term in the bracket as the covariance matrix. Thus

$$\boxed{\hat{\mathbf{a}}(t) = \int_{T_i}^{T_f} \mathbf{K}_{\mathbf{a}}(t, z)\, \mathbf{D}(z, \hat{\mathbf{a}}(z))[\mathbf{r}_g(z) - \mathbf{g}(z)]\, dz, \qquad T_i \le t \le T_f.} \tag{160}$$

As we would expect, the form of these equations is directly analogous to the one-dimensional case.

The next step is to find a lower bound on the mean-square error in estimating a vector random process.

5.4.4 Lower Bound on the Error Matrix

In the multidimensional case we are concerned with estimating the vector $\mathbf{a}(t)$. We can define an error vector

$$\hat{\mathbf{a}}(t) - \mathbf{a}(t) = \mathbf{a}_\epsilon(t), \tag{161}$$

which consists of N elements: $a_{\epsilon_1}(t), a_{\epsilon_2}(t), \ldots, a_{\epsilon_N}(t)$. We want to find the error correlation matrix.

Now,

$$\hat{\mathbf{a}}(t) - \mathbf{a}(t) = \sum_{i=1}^{\infty} (\hat{a}_i - a_i)\boldsymbol{\psi}_i(t) \triangleq \sum_{i=1}^{\infty} a_{\epsilon_i} \boldsymbol{\psi}_i(t). \tag{162}$$

Then, using the same approach as in Section 5.3, (68),

$$\mathbf{R}_{e_I} \triangleq \frac{1}{T} E\left[\int_{T_i}^{T_f} \mathbf{a}_\epsilon(t)\, \mathbf{a}_\epsilon^T(t)\, dt\right]$$
$$= \frac{1}{T} \lim_{K \to \infty} \int_{T_i}^{T_f} dt \sum_{i=1}^{K} \sum_{j=1}^{K} \boldsymbol{\psi}_i(t)\, E(a_{\epsilon_i} a_{\epsilon_j}) \boldsymbol{\psi}_j^T(t). \tag{163}$$

We can lower bound the error matrix by a bound matrix $\mathbf{R}_B$ in the sense that the matrix $\mathbf{R}_{e_I} - \mathbf{R}_B$ is nonnegative definite. The diagonal terms in $\mathbf{R}_B$ represent lower bounds on the mean-square error in estimating the $a_i(t)$. Proceeding in a manner analogous to the one-dimensional case, we obtain

$$\mathbf{R}_B = \frac{1}{T}\int_{T_i}^{T_f} \mathbf{J}^{-1}(t, t)\, dt. \tag{164}$$

The kernel $\mathbf{J}^{-1}(t, x)$ is the inverse of the *information matrix kernel* $\mathbf{J}(t, x)$ and is defined by the matrix integral equation

$$\mathbf{J}^{-1}(t, x) + \int_{T_i}^{T_f} du \int_{T_i}^{T_f} dz\, \mathbf{J}^{-1}(t, u)\{E[\mathbf{D}(u, \mathbf{a}(u))\, \mathbf{Q}_\mathbf{n}(u, z)\, \mathbf{D}^T(z, \mathbf{a}(z))]\}\mathbf{K}_\mathbf{a}(z, x)$$
$$= \mathbf{K}_\mathbf{a}(t, x), \qquad T_i \le t, x \le T_f. \tag{165}$$

The derivation of (164) and (165) is quite tedious and does not add any insight to the problem. Therefore we omit it. The details are contained in [11].

As a final topic in our current discussion of multiple waveform estimation we develop an interesting interpretation of nonlinear estimation in the presence of noise that contains both a colored component and a white component.

5.4.5 Colored Noise Estimation

Consider the following problem:

$$r(t) = s(t, a(t)) + w(t) + n_c(t), \qquad T_i \le t \le T_f. \tag{166}$$

Here $w(t)$ is a white noise component ($N_0/2$, spectral height) and $n_c(t)$ is an independent colored noise component with covariance function $K_c(t, u)$. Then

$$K_n(t, u) = \frac{N_0}{2}\,\delta(t - u) + K_c(t, u). \tag{167}$$

The MAP estimate of $a(t)$ can be found from (31), (32), and (33)

$$\hat{a}(t) = \int_{T_i}^{T_f} K_a(t, u)\,\frac{\partial s(u, \hat{a}(u))}{\partial \hat{a}(u)}\,[r_g(u) - g(u)]\, du, \qquad T_i \le t \le T_f, \tag{168}$$

where

$$r(t) - s(t, \hat{a}(t)) = \int_{T_i}^{T_f} K_n(t, u)[r_g(u) - g(u)]\, du, \qquad T_i \le t \le T_f. \tag{169}$$

Substituting (167) into (169), we obtain

$$r(t) - s(t, \hat{a}(t)) = \int_{T_i}^{T_f} \left[\frac{N_0}{2}\,\delta(t - u) + K_c(t, u)\right][r_g(u) - g(u)]\, du. \tag{170}$$

We now want to demonstrate that the same estimate, $\hat{a}(t)$, is obtained if we estimate $a(t)$ and $n_c(t)$ *jointly*. In this case

$$\mathbf{a}(t) \triangleq \begin{bmatrix} a(t) \\ n_c(t) \end{bmatrix} \tag{171}$$

and

$$\mathbf{K_a}(t, u) = \begin{bmatrix} K_a(t, u) & 0 \\ 0 & K_c(t, u) \end{bmatrix}. \tag{172}$$

Because we are including the colored noise in the message vector, the only additive noise is the white component

$$K_n(t, u) = \frac{N_0}{2}\,\delta(t - u). \tag{173}$$

To use (160) we need the derivative matrix

$$\mathbf{D}(t, \mathbf{a}(t)) = \begin{bmatrix} \dfrac{\partial s(t, a(t))}{\partial a(t)} \\ 1 \end{bmatrix}. \tag{174}$$

Substituting into (160), we obtain two scalar equations,

$$\hat{a}(t) = \int_{T_i}^{T_f} \frac{2}{N_0} K_a(t, u) \frac{\partial s(u, \hat{a}(u))}{\partial \hat{a}(u)} [r(u) - s(u, \hat{a}(u)) - \hat{n}_c(u)]\, du, \qquad T_i \le t \le T_f, \tag{175}$$

$$\hat{n}_c(t) = \int_{T_i}^{T_f} \frac{2}{N_0} K_c(t, u)[r(u) - s(u, \hat{a}(u)) - \hat{n}_c(u)]\, du, \quad T_i \le t \le T_f. \tag{176}$$

Looking at (168), (170), and (175), we see that $\hat{a}(t)$ will be the same in both cases if

$$\int_{T_i}^{T_f} \frac{2}{N_0} [r(u) - s(u, \hat{a}(u)) - \hat{n}_c(u)] \left[\frac{N_0}{2}\,\delta(t - u) + K_c(t, u)\right] du = r(t) - s(t, \hat{a}(t)); \tag{177}$$

but (177) is identical to (176).

This leads to the following conclusion. Whenever there are independent white and colored noise components, we may always consider the colored noise as a message and jointly estimate it. The reason for this result is that the message and colored noise are independent and the noise enters into $r(t)$ in a linear manner. Thus the $N \times 1$ vector white noise problem includes all scalar colored noise problems in which there is a white noise component.

Before summarizing our results in this chapter we discuss the problem of estimating nonrandom waveforms briefly.

5.5 NONRANDOM WAVEFORM ESTIMATION

It is sometimes unrealistic to consider the signal that we are trying to estimate as a random waveform. For example, we may know that each time a particular event occurs the transmitted message will have certain distinctive features. If the message is modeled as a sample function of a random process then, in the processing of designing the optimum receiver, we may average out the features that are important. Situations of this type arise in sonar and seismic classification problems. Here it is more useful to model the message as an unknown, but nonrandom, waveform. To design the optimum processor we extend the maximum-likelihood estimation procedure for nonrandom variables to the waveform case. An appropriate model for the received signal is

$$r(t) = s(t, a(t)) + n(t), \qquad T_i \le t \le T_f, \tag{178}$$

where $n(t)$ is a zero-mean Gaussian process.

To find the maximum-likelihood estimate, we write the ln likelihood function and then choose the waveform $a(t)$ which maximizes it. The ln likelihood function is the limit of (72*b*) as $K \to \infty$.

$$\ln \Lambda(a(t)) = \int_{T_i}^{T_f} dt \int_{T_i}^{T_f} du\, s(t, a(t))\, Q_n(t, u)[r(u) - \tfrac{1}{2} s(u, a(u))], \tag{179}$$

where $Q_n(t, u)$ is the inverse kernel of the noise. For arbitrary $s(t, a(t))$ the minimization of $\ln \Lambda(a(t))$ is difficult. Fortunately, in the case of most interest to us the procedure is straightforward. This is the case in which the range of the function $s(t, a(t))$ includes all possible values of $r(t)$. An important example in which this is true is

$$r(t) = a(t) + n(t). \tag{180}$$

An example in which it is not true is

$$r(t) = \sin(\omega_c t + a(t)) + n(t). \tag{181}$$

Here all functions in the range of $s(t, a(t))$ have amplitudes less than one while the possible amplitudes of $r(t)$ are not bounded.

We confine our discussion to the case in which the range includes all possible values of $r(t)$. A necessary condition to minimize $\ln \Lambda(a(t))$ follows easily with variational techniques:

$$\int_{T_i}^{T_f} a_\epsilon(t)\, dt \left\{ \frac{\partial s(t, \hat{a}_{ml}(t))}{\partial \hat{a}_{ml}(t)} \int_{T_i}^{T_f} Q_n(t, u)[r(u) - s(u, \hat{a}_{ml}(u))]\, du \right\} = 0 \tag{182}$$

for every $a_\epsilon(t)$. A solution is

$$r(u) = s(u, \hat{a}_{ml}(u)), \qquad T_i \le u \le T_f, \tag{183}$$

because for every $r(u)$ there exists at least one $a(u)$ that could have mapped into it. There is no guarantee, however, that a unique inverse exists. Once again we can obtain a useful answer by narrowing our discussion. Specifically, we shall consider the problem given by (180). Then

$$\hat{a}_{ml}(u) = r(u). \tag{184}$$

Thus the maximum-likelihood estimate is simply the received waveform. It is an unbiased estimate of $a(t)$. It is easy to demonstrate that the maximum-likelihood estimate is efficient. Its variance can be obtained from a generalization of the Cramér-Rao bound or by direct computation. Using the latter procedure, we obtain

$$\sigma_l^2 \triangleq E\left\{\int_{T_i}^{T_f} [\hat{a}_{ml}(t) - a(t)]^2 \, dt\right\}. \tag{185}$$

It is frequently convenient to normalize the variance by the length of the interval. We denote this normalized variance (which is just the average mean-square estimation error) as ξ_{ml}:

$$\xi_{ml} \triangleq E\left\{\frac{1}{T}\int_{T_i}^{T_f} [\hat{a}_{ml}(t) - a(t)]^2 \, dt\right\}. \tag{186}$$

Using (180) and (184), we have

$$\xi_{ml} = \frac{1}{T}\int_{T_i}^{T_f} K_n(t, t) \, dt. \tag{187}$$

Several observations follow easily.

If the noise is white, the error is infinite. This is intuitively logical if we think of a series expansion of $a(t)$. We are trying to estimate an infinite number of components and because we have assumed no a priori information about their contribution in making up the signal we weight them equally. Because the mean-square errors are equal on each component, the equal weighting leads to an infinite mean-square error. Therefore to make the problem meaningful we must assume that the noise has finite energy over any finite interval. This can be justified physically in at least two ways:

1. The receiving elements (antenna, hydrophone, or seismometer) will have a finite bandwidth;

2. If we assume that we know an approximate frequency band that contains the signal, we can insert a filter that passes these frequencies without distortion and rejects other frequencies.†

† Up to this point we have argued that a white noise model had a good physical basis. The essential point of the argument was that if the noise was wideband compared with the bandwidth of the processors then we could consider it to be white. We tacitly assumed that the receiving elements mentioned in (1) had a much larger bandwidth than the signal processor. Now the mathematical model of the signal does not have enough structure and we must impose the bandwidth limitation to obtain meaningful results.

If the noise process is stationary, then

$$\xi_{ml} = \frac{1}{T}\int_{T_i}^{T_f} K_n(0)\,dt = K_n(0) = \int_{-\infty}^{\infty} S_n(\omega)\frac{d\omega}{2\pi}. \tag{188}$$

Our first reaction might be that such a crude procedure cannot be efficient. From the parameter estimation problem, however, we recall that a priori knowledge was not important when the measurement noise was small. The same result holds in the waveform case. We can illustrate this result with a simple example.

Example. Let $n'(t)$ be a white process with spectral height $N_0/2$ and assume that $T_i = -\infty$, $T_f = \infty$. We know that $a(t)$ does not have frequency components above W cps. We pass $r(t)$ through a filter with unity gain from $-W$ to $+W$ and zero gain outside this band. The output is the message $a(t)$ plus a noise $n(t)$ which is a sample function from a flat bandlimited process. The ML estimate of $a(t)$ is the output of this filter and

$$\xi_{ml} = N_0 W. \tag{189}$$

Now suppose that $a(t)$ is actually a sample function from a bandlimited random process $(-W, W)$ and spectral height P. If we used a MAP or an MMSE estimate, it would be efficient and the error would be given by (120),

$$\xi_{ms} = \xi_{map} = \frac{PN_0W}{P + N_0/2}. \tag{190}$$

The normalized errors in the two cases are

$$\xi_{ml:n} = \frac{N_0W}{P} \tag{191}$$

and

$$\xi_{ms:n} = \xi_{map:n} = \frac{N_0W}{P}\left(1 + \frac{N_0}{2P}\right)^{-1}. \tag{192}$$

Thus the difference in the errors is negligible for $N_0/2P < 0.1$.

We see that in this example both estimation procedures assumed a knowledge of the signal bandwidth to design the processor. The MAP and MMSE estimates, however, also required a knowledge of the spectral heights. Another basic difference in the two procedures is not brought out by the example because of the simple spectra that were chosen. The MAP and MMSE estimates are formed by attenuating the various frequencies,

$$H_o(j\omega) = \frac{S_a(\omega)}{N_0/2 + S_a(\omega)}. \tag{193}$$

Therefore, unless the message spectrum is uniform over a fixed bandwidth, the message will be *distorted*. This distortion is introduced to reduce the total mean-square error, which is the sum of message and noise distortion.

On the other hand, the ML estimator never introduces any message distortion; the error is due solely to the noise. (For this reason ML estimators are also referred to as distortionless filters.)

In the sequel we concentrate on MAP estimation (an important exception is Section II.5.3); it is important to remember, however, that in many cases ML estimation serves a useful purpose (see Problem 5.5.2 for a further example.)

5.6 SUMMARY

In this chapter we formulated the problem of estimating a continuous waveform. The primary goal was to develop the equations that specify the estimates.

In the case of a single random waveform the MAP estimate was specified by two equations,

$$\hat{a}(t) = \int_{T_i}^{T_f} K_a(t, u) \frac{\partial s(u, \hat{a}(u))}{\partial \hat{a}(u)} [r_g(u) - g(u)]\, du, \qquad T_i \le t \le T_f, \quad (194)$$

where $[r_g(u) - g(u)]$ was specified by the equation

$$r(t) - s(t, \hat{a}(t)) = \int_{T_i}^{T_f} K_n(t, u)[r_g(u) - g(u)]\, du, \qquad T_i \le t \le T_f. \quad (195)$$

For the special case of white noise this reduced to

$$\hat{a}(t) = \frac{2}{N_0} \int_{T_i}^{T_f} K_a(t, u)[r(u) - s(u, \hat{a}(u))]\, du, \qquad T_i \le t \le T_f. \quad (196)$$

We then derived a bound on the mean-square error in terms of the trace of the information kernel,

$$\xi_I \ge \frac{1}{T} \int_{T_i}^{T_f} J^{-1}(t, t)\, dt. \quad (197)$$

For white noise this had a particularly simple interpretation.

$$\xi_I \ge \frac{N_0}{2R_{\dot{d}_s}^*(0)} \frac{1}{T} \int_{T_i}^{T_f} h_o(t, t)\, dt, \quad (198)$$

where $h_o(t, u)$ satisfied the integral equation

$$K_a(t, u) = \frac{N_0}{2R_{\dot{d}_s}^*(0)} h_o(t, u) + \int_{T_i}^{T_f} h_o(t, z)\, K_a(z, u)\, dz, \qquad T_i \le t, u \le T_f. \quad (199)$$

The function $h_o(t, u)$ was precisely the optimum processor for a *related* linear estimation problem. We showed that for linear modulation $\hat{a}_{\text{map}}(t)$ was efficient. For nonlinear modulation we shall find that $\hat{a}_{\text{map}}(t)$ is asymptotically efficient.

We then extended these results to the multidimensional case. The basic integral equations were logical extensions of those obtained in the scalar case. We also observed that a colored noise component could always be treated as an additional message and simultaneously estimated.

Finally, we looked at the problem of estimating a nonrandom waveform. The result for the problem of interest was straightforward. A simple example demonstrated a case in which it was essentially as good as a MAP estimate.

In subsequent chapters we shall study the estimator equations and the receiver structures that they suggest in detail. In Chapter 6 we study linear modulation and in Chapter II.2, nonlinear modulation.

5.7 PROBLEMS

Section P5.2 Derivation of Equations

Problem 5.2.1. If we approximate $a(t)$ by a K-term approximation $a_K(t)$, the inverse kernel $Q_{a_K}(t, u)$ is well-behaved. The logarithm of the likelihood function is

$$\ln \Lambda(a_K(t)) + \ln p_{\mathbf{a}}(\mathbf{A}) = \int\int_{T_i}^{T_f} [s(t, a_K(t))]\, Q_n(t, u)[r(u) - \tfrac{1}{2}s(u, a_K(u))]\, du$$

$$-\tfrac{1}{2}\int\int_{T_i}^{T_f} a_K(t)\, Q_{a_K}(t, u)\, a_K(u)\, dt\, du + \text{constant terms.}$$

1. Use an approach analogous to that in Section 3.4.5 to find $\hat{a}_K(t)$ [i.e., let $a_K(t) = \hat{a}_K(t) + \epsilon a_\epsilon(t)$].

2. Eliminate $Q_{a_K}(t, u)$ from the result and let $K \to \infty$ to get a final answer.

Problem 5.2.2. Let

$$r(t) = s(t, a(t)) + n(t), \qquad T_i \le t \le T_f,$$

where the processes are the same as in the text. Assume that

$$E[a(t)] = m_a(t).$$

1. Find the integral equation specifying $\hat{a}(t)$, the MAP estimate of $a(t)$.
2. Consider the special case in which

$$s(t, a(t)) = a(t).$$

Write the equations for this case.

Problem 5.2.3. Consider the case of the model in Section 5.2.1 in which

$$K_a(t, u) = \sigma_a^2, \qquad 0 \le t, u \le T.$$

1. What does this imply about $a(t)$.
2. Verify that (33) reduces to a previous result under this condition.

Problem 5.2.4. Consider the amplitude modulation system shown in Fig. P5.1. The Gaussian processes $a(t)$ and $n(t)$ are stationary with spectra $S_a(\omega)$ and $S_n(\omega)$, respectively. Let $T_i = -\infty$ and $T_f = \infty$.

1. Draw a block diagram of the optimum receiver.
2. Find $E[a_\epsilon^2(t)]$ as a function of $H(j\omega)$, $S_a(\omega)$, and $S_n(\omega)$.

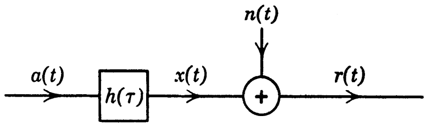

Fig. P5.1

Problem 5.2.5. Consider the communication system shown in Fig. P5.2. Draw a block diagram of the optimum nonrealizable receiver to estimate $a(t)$. Assume that a MAP interval estimate is required.

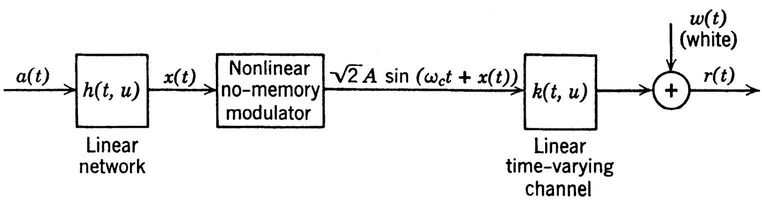

Fig. P5.2

Problem 5.2.6. Let

$$r(t) = s(t, a(t)) + n(t), \qquad -\infty < t < \infty,$$

where $a(t)$ and $n(t)$ are sample functions from zero-mean independent stationary Gaussian processes. Use the integrated Fourier transform method to derive the infinite time analog of (31) to (33) in the text.

Problem 5.2.7. In Chapter 4 (pp. 299–301) we derived the integral equation for the colored noise detection problem by using the idea of a sufficient statistic. Derive (31) to (33) by a suitable extension of this technique.

Section P5.3 Lower Bounds

Problem 5.3.1. In Chapter 2 we considered the case in which we wanted to estimate a linear function of the vector **A**,

$$d \triangleq g_d(\mathbf{A}) = \mathbf{G}_d\mathbf{A},$$

and proved that if $\hat{d}$ was unbiased then

$$E[(\hat{d} - d)^2] \geq \mathbf{G}_d \mathbf{J}^{-1} \mathbf{G}_d^T$$

[see (2.287) and (2.288)]. A similar result can be derived for random variables. Use the random variable results to derive (87).

Problem 5.3.2. Let

$$r(t) = s(t, a(t)) + n(t), \qquad T_i \leq t \leq T_f.$$

Assume that we want to estimate $a(T_f)$. Can you modify the results of Problem 5.3.1 to derive a bound on the mean-square *point* estimation error?

$$\xi_P \triangleq E\{[\hat{a}(T_f) - a(T_f)]^2\}.$$

What difficulties arise in nonlinear *point* estimation?

Problem 5.3.3. Let

$$r(t) = s(t, a(t)) + n(t), \qquad -\infty < t < \infty.$$

The processes $a(t)$ and $n(t)$ are statistically independent, stationary, zero-mean, Gaussian random processes with spectra $S_a(\omega)$ and $S_n(\omega)$ respectively. Derive a bound on the mean-square estimation error by using the integrated Fourier transform approach.

Problem 5.3.4. Let

$$r(t) = s(t, x(t)) + n(t), \qquad T_i \leq t \leq T_f,$$

where

$$x(t) = \int_{T_i}^{T_f} h(t, u)\, a(u)\, du, \qquad T_i \leq t \leq T_f,$$

and $a(t)$ and $n(t)$ are statistically independent, zero-mean Gaussian random processes.

1. Derive a bound on the mean-square interval error in estimating $a(t)$.
2. Consider the special case in which

$$s(t, x(t)) = x(t), \quad h(t, u) = h(t - u), \quad T_i = -\infty, \quad T_f = \infty,$$

and the processes are stationary. Verify that the estimate is efficient and express the error in terms of the various spectra.

Problem 5.3.5. Explain why a necessary and sufficient condition for an efficient estimate to exist in the waveform estimation case is that the modulation be linear [see (73)].

Problem 5.3.6. Prove the result given in (81) by starting with the definition of $\mathbf{J}^{-1}$ and using (80) and (82).

Section P5.4 Multidimensional Waveforms

Problem 5.4.1. The received waveform is

$$r(t) = s(t, a(t)) + n(t), \qquad 0 \leq t \leq T.$$

The message $a(t)$ and the noise $n(t)$ are sample functions from zero-mean, jointly Gaussian random processes.

$$E[a(t)\, a(u)] \triangleq K_{aa}(t, u),$$

$$E[a(t)\, n(u)] \triangleq K_{an}(t, u),$$

$$E[n(t)\, n(u)] \triangleq \frac{N_0}{2}\, \delta(t - u) + K_c(t, u).$$

Derive the integral equations that specify the MAP estimate $\hat{a}(t)$. *Hint.* Write a matrix covariance function $\mathbf{K_x}(t, u)$ for the vector $\mathbf{x}(t)$, where

$$\mathbf{x}(t) \triangleq \begin{bmatrix} a_K(t) \\ n(t) \end{bmatrix}.$$

Define an inverse matrix kernel,

$$\int_0^T \mathbf{Q_x}(t, u)\, \mathbf{K_x}(u, z)\, du = \mathbf{I}\, \delta(t - z).$$

Write

$$\ln \Lambda(\mathbf{x}(t)) = -\tfrac{1}{2} \iint_0^T [a_K(t) \,\vdots\, r(t) - s(t, a_K(t))] \mathbf{Q_x}(t, u) \begin{bmatrix} a_K(u) \\ r(u) - s(u, a_K(u)) \end{bmatrix} dt\, du,$$

$$-\tfrac{1}{2} \iint_0^T r(t) Q_{\mathbf{x},22}(t, u)\, r(u)\, dt\, du$$

$$-\tfrac{1}{2} \iint_0^T a_K(t)\, Q_a(t, u) a_K(u)\, dt\, du.$$

Use the variational approach of Problem 5.2.1.

Problem 5.4.2. Let

$$r(t) = s(t, a(t), \mathbf{B}) + n(t), \qquad T_i \le t \le T_f,$$

where $a(t)$ and $n(t)$ are statistically independent, zero-mean, Gaussian random processes. The vector $\mathbf{B}$ is Gaussian, $N(0, \mathbf{\Lambda_B})$, and is independent of $a(t)$ and $n(t)$. Find an equation that specifies the joint MAP estimates of $a(t)$ and $\mathbf{B}$.

Problem 5.4.3. In a PM/PM scheme the messages are phase-modulated onto subcarriers and added:

$$z(t) = \sum_{j=1}^{N} \sqrt{2}\, g_j \sin [\omega_j t + \beta_j\, a_j(t)].$$

The sum $z(t)$ is then phase-modulated onto a main carrier.

$$s(t, \mathbf{a}(t)) = \sqrt{2P} \sin [\omega_c t + \beta_c z(t)].$$

The received signal is

$$r(t) = s(t, \mathbf{a}(t)) + w(t), \qquad -\infty < t < \infty.$$

The messages $a_i(t)$ are statistically independent with spectrum $S_a(\omega)$. The noise is independent of $\mathbf{a}(t)$ and is white ($N_0/2$). Find the integral equation that specifies $\hat{\mathbf{a}}(t)$ and draw the block diagram of an unrealizable receiver. Simplify the diagram by exploiting the frequency difference between the messages and the carriers.

Problem 5.4.4. Let

$$\mathbf{r}(t) = \mathbf{a}(t) + \mathbf{n}(t), \qquad -\infty < t < \infty,$$

where $\mathbf{a}(t)$ and $\mathbf{n}(t)$ are independent Gaussian processes with spectral matrices $\mathbf{S_a}(\omega)$ and $\mathbf{S_n}(\omega)$, respectively.

1. Write (151), (152), and (160) in the frequency domain, using integrated transforms.
2. Verify that $\mathscr{F}[\mathbf{Q_n}(\tau)] = \mathbf{S_n}^{-1}(\omega)$.
3. Draw a block diagram of the optimum receiver. Reduce it to a single matrix filter.
4. Derive the frequency domain analogs to (164) and (165) and use them to write an error expression for this case.
5. Verify that exactly the same results (parts 1, 3, and 4) can be obtained *heuristically* by using ordinary Fourier transforms.

Problem 5.4.5. Let

$$\mathbf{r}(t) = \int_{-\infty}^{\infty} \mathbf{h}(t - \tau)\, a(\tau)\, d\tau + \mathbf{n}(t), \qquad -\infty < t < \infty,$$

where $\mathbf{h}(\tau)$ is a matrix filter with one input and N outputs. Repeat Problem 5.4.4.

Problem 5.4.6. Consider a simple five-element linear array with uniform spacing Δ. (Fig. P5.3). The message is a plane wave whose angle of arrival is θ and velocity of propagation is c. The output at the first element is

$$r_1(t) = a(t) + n_1(t), \qquad -\infty < t < \infty.$$

The output at the second element is

$$r_2(t) = a(t - \tau_\Delta) + n_2(t), \qquad -\infty < t < \infty,$$

where $\tau_\Delta = \Delta \sin \theta / c$. The other outputs follow in an obvious manner. The noises are statistically independent and white ($N_0/2$). The message spectrum is $S_a(\omega)$.

1. Show that this is a special case of Problem 5.4.5.
2. Give an intuitive interpretation of the optimum processor.
3. Write an expression for the minimum mean-square interval estimation error.

$$\xi_I = E[a_\epsilon^2(t)].$$

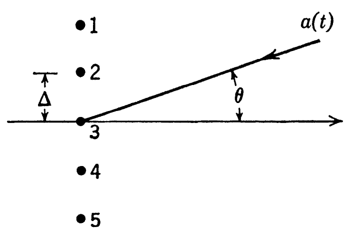

Fig. P5.3

Problem 5.4.7. Consider a zero-mean stationary Gaussian random process $a(t)$ with covariance function $K_a(\tau)$. We observe $a(t)$ in the interval (T_i, T_f) and want to estimate $a(t)$ in the interval (T_α, T_β), where $T_\alpha > T_f$.

1. Find the integral equation specifying $\hat{a}_{\text{map}}(t)$, $T_\alpha \le t \le T_\beta$.
2. Consider the special case in which

$$K_a(\tau) = \sigma_a^2 e^{-k|\tau|}, \qquad -\infty < \tau < \infty.$$

Verify that

$$\hat{a}_{\text{map}}(t_1) = a(T_f)e^{-k(t_1 - T_f)} \quad \text{for} \quad T_\alpha \le t_1 \le T_\beta.$$

Hint. Modify the procedure in Problem 5.4.1.

Section P5.5 Nonrandom Waveforms

Problem 5.5.1. Let

$$\mathbf{r}(t) = \mathbf{x}(t) + \mathbf{n}(t), \qquad -\infty < t < \infty,$$

where $\mathbf{n}(t)$ is a stationary, zero-mean, Gaussian process with spectral matrix $\mathbf{S_n}(\omega)$ and $\mathbf{x}(t)$ is a vector signal with finite energy. Denote the vector integrated Fourier transforms of the function as $\mathbf{Z_r}(\omega)$, $\mathbf{Z_x}(\omega)$, and $\mathbf{Z_n}(\omega)$, respectively [see (2.222) and (2.223)]. Denote the Fourier transform of $\mathbf{x}(t)$ as $\mathbf{X}(j\omega)$.

1. Write $\ln \Lambda(\mathbf{x}(t))$ in terms of these quantities.
2. Find $\hat{\mathbf{X}}_{ml}(j\omega)$.
3. Derive $\hat{\mathbf{X}}_{ml}(j\omega)$ heuristically using ordinary Fourier transforms for the processes.

Problem 5.5.2. Let

$$\mathbf{r}(t) = \int_{-\infty}^{\infty} \mathbf{h}(t - \tau)\, a(\tau)\, d\tau + \mathbf{n}(t), \qquad -\infty < t < \infty,$$

where $\mathbf{h}(\tau)$ is the impulse response of a matrix filter with one input and N outputs and transfer function $\mathbf{H}(j\omega)$.

1. Modify the results of Problem 5.5.1 to include this case.
2. Find $\hat{a}_{ml}(t)$.
3. Verify that $\hat{a}_{ml}(t)$ is unbiased.
4. Evaluate $\text{Var}\,[\hat{a}_{ml}(t) - a(t)]$.

REFERENCES

[1] D. C. Youla, "The Use of Maximum Likelihood in Estimating Continuously Modulated Intelligence Which Has Been Corrupted by Noise," *IRE Trans. Inform. Theory*, **IT-3**, 90–105 (March 1954).
[2] J. G. Lawton, private communication, dated August 20, 1962.
[3] J. G. Lawton, T. T. Chang, and C. J. Henrich, "Maximum Likelihood Reception of FM Signals," Detection Memo No. 25, Contract No. AF 30(602)-2210, Cornell Aeronautical Laboratory Inc., January 4, 1963.
[4] H. L. Van Trees, "Analog Communication Over Randomly Time-Varying Channels," presented at WESCON, August, 1964.

[5] L. L. Rauch, "Some Considerations on Demodulation of FM Signals," unpublished memorandum, 1961.
[6] L. L. Rauch, "Improved Demodulation and Predetection Recording," Research Report, Instrumentation Engineering Program, University of Michigan, February 28, 1961.
[7] H. L. Van Trees, "Bounds on the Accuracy Attainable in the Estimation of Random Processes," *IEEE Trans. Inform. Theory*, **IT-12**, No. 3, 298–304 (April 1966).
[8] N. H. Nichols and L. Rauch, *Radio Telemetry*, Wiley, New York, 1956.
[9] J. B. Thomas and E. Wong, "On the Statistical Theory of Optimum Demodulation," *IRE Trans. Inform. Theory*, **IT-6**, 420–425 (September 1960).
[10] J. K. Wolf, "On the Detection and Estimation Problem for Multiple Nonstationary Random Processes," Ph.D. Thesis, Princeton University, Princeton, New Jersey, 1959.
[11] H. L. Van Trees, "Accuracy Bounds in Multiple Process Estimation," Internal Memorandum, DET Group, M.I.T., April 1965.