Several important observations should be made about this result.

1. It shows that any unbiased estimate must have a variance greater than a certain number.

2. If (187) is satisfied, the estimate $\hat{a}_{ml}(\mathbf{R})$ will satisfy the bound with an equality. We show this by combining (187) and (177). The left equality is the maximum likelihood equation. The right equality is (187):

$$0 = \frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\bigg|_{A=\hat{a}_{ml}(\mathbf{R})} = (\hat{a}(\mathbf{R}) - A)\, k(A)\bigg|_{A=\hat{a}_{ml}(\mathbf{R})}. \tag{193}$$

In order for the right-hand side to equal zero either

$$\hat{a}(\mathbf{R}) = \hat{a}_{ml}(\mathbf{R}) \tag{194}$$

or

$$k(\hat{a}_{ml}) = 0. \tag{195}$$

Because we want a solution that depends on the data, we eliminate (195) and require (194) to hold.

Thus, *if* an efficient estimate exists, it is $\hat{a}_{ml}(\mathbf{R})$ and can be obtained as a unique solution to the likelihood equation.

3. If an efficient estimate *does not* exist [i.e., $\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)/\partial A$ cannot be put into the form of (187)], we do not know how good $\hat{a}_{ml}(\mathbf{R})$ is. Further, we do not know how close the variance of any estimate will approach the bound.

4. In order to use the bound, we must verify that the estimate of concern is unbiased. Similar bounds can be derived simply for biased estimates (Problem 2.4.17).

We can illustrate the application of ML estimation and the Cramér–Rao inequality by considering Examples 2, 3, and 4. The observation model is identical. We now assume, however, that the parameters to be estimated are nonrandom variables.

***Example 2.*** From (138) we have

$$r_i = A + n_i, \qquad i = 1, 2, \ldots, N. \tag{196}$$

Taking the logarithm of (139) and differentiating, we have

$$\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} = \frac{N}{\sigma_n^2}\left(\frac{1}{N}\sum_{i=1}^{N} R_i - A\right). \tag{197}$$

Thus

$$\hat{a}_{ml}(\mathbf{R}) = \frac{1}{N}\sum_{i=1}^{N} R_i. \tag{198}$$

To find the bias we take the expectation of both sides,

$$E[\hat{a}_{ml}(\mathbf{R})] = \frac{1}{N}\sum_{i=1}^{N} E(R_i) = \frac{1}{N}\sum_{i=1}^{N} A = A, \tag{199}$$

so that $\hat{a}_{ml}(\mathbf{R})$ is unbiased.

Because the expression in (197) has the form required by (187), we know that $\hat{a}_{ml}(\mathbf{R})$ is an efficient estimate. To evaluate the variance we differentiate (197):

$$\frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2} = -\frac{N}{\sigma_n^2}. \tag{200}$$

Using (179) and the efficiency result, we have

$$\text{Var}\,[\hat{a}_{ml}(\mathbf{R}) - A] = \frac{\sigma_n^2}{N}. \tag{201}$$

Skipping Example 3 for the moment, we go to Example 4.

***Example 4.*** Differentiating the logarithm of (162), we have

$$\frac{\partial \ln \Pr(n = N|A)}{\partial A} = \frac{\partial}{\partial A}(N \ln A - A - \ln N!)$$

$$= \frac{N}{A} - 1 = \frac{1}{A}(N - A). \tag{202}$$

The ML estimate is

$$\hat{a}_{ml}(N) = N. \tag{203}$$

It is clearly unbiased and efficient. To obtain the variance we differentiate (202):

$$\frac{\partial^2 \ln \Pr(n = N|A)}{\partial A^2} = -\frac{N}{A^2}. \tag{204}$$

Thus

$$\text{Var}\,[\hat{a}_{ml}(N) - A] = \frac{A^2}{E(N)} = \frac{A^2}{A} = A. \tag{205}$$

In both Examples 2 and 4 we see that the ML estimates could have been obtained from the MAP estimates [let $\sigma_a \to \infty$ in (144) and recall that $\hat{a}_{ms}(\mathbf{R}) = \hat{a}_{map}(\mathbf{R})$ and let $\lambda \to 0$ in (169)].

We now return to Example 3.

***Example 3.*** From the first term in the exponent in (160), we have

$$\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} = \frac{1}{\sigma_n^2}\sum_{i=1}^{N}[R_i - s(A)]\frac{\partial s(A)}{\partial A}. \tag{206}$$

In general, the right-hand side cannot be written in the form required by (187), and therefore an unbiased efficient estimate does not exist.

The likelihood equation is

$$\left[\frac{\partial s(A)}{\partial A}\frac{1}{\sigma_n^2}\right]\left[\frac{1}{N}\sum_{i=1}^{N} R_i - s(A)\right]\Bigg|_{A=\hat{a}_{ml}(\mathbf{R})} = 0. \tag{207}$$

If the range of $s(A)$ includes $(1/N)\sum_{i=1}^{N} R_i$, a solution exists:

$$s[\hat{a}_{ml}(\mathbf{R})] = \frac{1}{N}\sum_{i=1}^{N} R_i. \tag{208}$$

If (208) can be satisfied, then

$$\hat{a}_{ml}(\mathbf{R}) = s^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right). \tag{209}$$

[Observe that (209) tacitly assumes that $s^{-1}(\cdot)$ exists. If it does not, then even in the absence of noise we shall be unable to determine $A$ unambiguously. If we were designing a system, we would always choose an $s(\cdot)$ that allows us to find $A$ unambiguously in the absence of noise.] If the range of $s(a)$ does not include $(1/N)\sum_{i=1}^{N} R_i$, the maximum is at an end point of the range.

We see that the maximum likelihood estimate commutes over nonlinear operations. (This is *not* true for MS or MAP estimation.) If it is unbiased, we evaluate the bound on the variance by differentiating (206):

$$\frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2} = \frac{1}{\sigma_n^2}\sum_{i=1}^{N}[R_i - s(A)]\frac{\partial^2 s(A)}{\partial A^2} - \frac{N}{\sigma_n^2}\left[\frac{\partial s(A)}{\partial A}\right]^2. \tag{210}$$

Observing that

$$E[r_i - s(A)] = E(n_i) = 0, \tag{211}$$

we obtain the following bound for any unbiased estimate,

$$\text{Var}\,[\hat{a}(\mathbf{R}) - A] \geq \frac{\sigma_n^2}{N[\partial s(A)/\partial A]^2}. \tag{212}$$

We see that the bound is exactly the same as that in Example 2 except for a factor $[\partial s(A)/\partial A]^2$. The intuitive reason for this factor and also some feeling for the conditions under which the bound will be useful may be obtained by inspecting the typical function shown in Fig. 2.22. Define

$$Y = s(A). \tag{213}$$

Then

$$r_i = Y + n_i. \tag{214}$$

The variance in estimating $Y$ is just $\sigma_n^2/N$. However, if $y_\epsilon$, the error in estimating $Y$, is small enough so that the slope is constant, then

$$A_\epsilon \simeq \frac{Y_\epsilon}{\left.\dfrac{\partial s(A)}{\partial A}\right|_{A=\hat{a}(\mathbf{R})}} \tag{215}$$

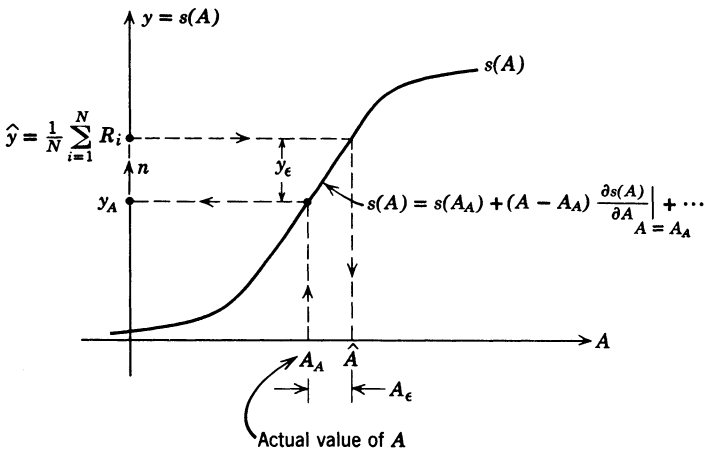


**Fig. 2.22 Behavior of error variance in the presence of small errors.**

and

$$\text{Var}\,(a_\epsilon) \cong \frac{\text{Var}\,(y_\epsilon)}{[\partial s(A)/\partial A]^2} = \frac{\sigma_n^2}{N[\partial s(A)/\partial A]^2}. \tag{216}$$

We observe that if $y_\epsilon$ is large there will no longer be a simple linear relation between $y_\epsilon$ and $a_\epsilon$. This tells us when we can expect the Cramér-Rao bound to give an accurate answer in the case in which the parameter enters the problem in a nonlinear manner. Specifically, whenever the estimation error is small, relative to $A\,\partial^2 s(A)/\partial A^2$, we should expect the actual variance to be close to the variance bound given by the Cramér-Rao inequality.

The properties of the ML estimate which are valid when the error is small are generally referred to as asymptotic. One procedure for developing them formally is to study the behavior of the estimate as the number of independent observations $N$ approaches infinity. Under reasonably general conditions the following may be proved (e.g., Cramér [9], pp. 500–504).

1. The solution of the likelihood equation (177) converges in probability to the correct value of $A$ as $N \to \infty$. *Any* estimate with this property is called consistent. Thus the ML estimate is consistent.
2. The ML estimate is asymptotically efficient; that is,

$$\lim_{N\to\infty} \frac{\text{Var}\,[\hat{a}_{\text{ml}}(\mathbf{R}) - A]}{\left(-E\left[\dfrac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2}\right]\right)^{-1}} = 1.$$

3. The ML estimate is asymptotically Gaussian, $N(A, \sigma_{a_\epsilon})$.

These properties all deal with the behavior of ML estimates for large N. They provide some motivation for using the ML estimate even when an efficient estimate does not exist.

At this point a logical question is: "Do better estimation procedures than the maximum likelihood procedure exist?" Certainly if an efficient estimate does not exist, there may be unbiased estimates with lower variances. The difficulty is that there is no general rule for finding them. In a particular situation we can try to improve on the ML estimate. In almost all cases, however, the resulting estimation rule is more complex, and therefore we emphasize the maximum likelihood technique in all of our work with real variables.

A second logical question is: "Do better lower bounds than the Cramér–Rao inequality exist?" One straightforward but computationally tedious procedure is the Bhattacharyya bound. The Cramér–Rao bound uses $\partial^2 p_{\mathbf{r}|a}(\mathbf{R}|A)/\partial A^2$. Whenever an efficient estimate does not exist, a larger bound which involves the higher partial derivatives can be obtained. Simple derivations are given in [13] and [14] and in Problems 2.4.23–24. For the cases of interest to us the computation is too involved to make the bound of much practical value. A second bound is the Barankin bound

(e.g. [15]). Its two major advantages are that it does not require the probability density to be differentiable and it gives the greatest lower bound. Its disadvantages are that it requires a maximization over a function to obtain the bound and the procedure for finding this maximum is usually not straightforward. Some simple examples are given in the problems (2.4.18–19). In most of our discussions, we emphasize the Cramér–Rao bound.

We now digress briefly to develop a similar bound on the mean-square error when the parameter is random.

***Lower Bound on the Minimum Mean-Square Error in Estimating a Random Parameter.*** In this section we prove the following theorem.

**Theorem.** Let $a$ be a random variable and $\mathbf{r}$, the observation vector. The mean-square error of any estimate $\hat{a}(\mathbf{R})$ satisfies the inequality

$$\begin{aligned} E\{[\hat{a}(\mathbf{R}) - a]^2\} &\geq \left(E\left\{\left[\frac{\partial \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A}\right]^2\right\}\right)^{-1} \\ &= \left\{-E\left[\frac{\partial^2 \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A^2}\right]\right\}^{-1}. \end{aligned} \tag{217}$$

Observe that the probability density is a joint density and that the expectation is over both $a$ and $\mathbf{r}$. The following conditions are assumed to exist:

1. $\dfrac{\partial p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A}$ is absolutely integrable with respect to $\mathbf{R}$ and $A$.

2. $\dfrac{\partial^2 p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A^2}$ is absolutely integrable with respect to $\mathbf{R}$ and $A$.

3. The conditional expectation of the error, given $A$, is

$$B(A) = \int_{-\infty}^{\infty} [\hat{a}(\mathbf{R}) - A]\, p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R}. \tag{218}$$

We assume that

$$\lim_{A \to \infty} B(A)\, p_a(A) = 0, \tag{219}$$

$$\lim_{A \to -\infty} B(A)\, p_a(A) = 0. \tag{220}$$

The proof is a simple modification of the one on p. 66. Multiply both sides of (218) by $p_a(A)$ and then differentiate with respect to $A$:

$$\begin{aligned} \frac{d}{dA}[p_a(A)\, B(A)] &= -\int_{-\infty}^{\infty} p_{\mathbf{r},a}(\mathbf{R}, A)\, d\mathbf{R} \\ &\quad + \int_{-\infty}^{\infty} \frac{\partial p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A}\, [\hat{a}(\mathbf{R}) - A]\, d\mathbf{R}. \end{aligned} \tag{221}$$

Now integrate with respect to $A$:

$$p_a(A)\, B(A)\Big|_{-\infty}^{+\infty} = -1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A}\,[\hat{a}(\mathbf{R}) - A]\, dA\, d\mathbf{R}. \quad (222)$$

The assumption in Condition 3 makes the left-hand side zero. The remaining steps are identical. The result is

$$E\{[\hat{a}(\mathbf{R}) - a]^2\} \geq \left\{E\left[\left(\frac{\partial \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A}\right)^2\right]\right\}^{-1} \quad (223)$$

or, equivalently,

$$E\{[\hat{a}(\mathbf{R}) - a]^2\} \geq \left\{-E\left[\frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2}\right] - E\left[\frac{\partial^2 \ln p_a(A)}{\partial A^2}\right]\right\}^{-1} \quad (224)$$

with equality if and only if

$$\frac{\partial \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A} = k[\hat{a}(\mathbf{R}) - A], \quad (225)$$

for all $\mathbf{R}$ and all $A$. (In the nonrandom variable case we used the Schwarz inequality on an integral over $\mathbf{R}$ so that the constant $k(A)$ could be a function of $A$. Now the integration is over both $\mathbf{R}$ and $A$ so that $k$ cannot be a function of $A$.) Differentiating again gives an equivalent condition

$$\frac{\partial^2 \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A^2} = -k. \quad (226)$$

Observe that (226) may be written in terms of the a posteriori density,

$$\frac{\partial^2 \ln p_{a|\mathbf{r}}(A|\mathbf{R})}{\partial A^2} = -k. \quad (227)$$

Integrating (227) twice and putting the result in the exponent, we have

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = \exp(-kA^2 + C_1 A + C_2) \quad (228)$$

for all $\mathbf{R}$ and $A$; but (228) is simply a statement that the a posteriori probability density of $a$ must be Gaussian for all $\mathbf{R}$ in order for an efficient estimate to exist. (Note that $C_1$ and $C_2$ are functions of $\mathbf{R}$).

Arguing as in (193)–(195), we see that if (226) is satisfied the MAP estimate will be efficient. Because the minimum MSE estimate cannot have a larger error, this tells us that $\hat{a}_{ms}(\mathbf{R}) = \hat{a}_{map}(\mathbf{R})$ whenever an efficient estimate exists. As a matter of technique, when an efficient estimate does exist, it is usually computationally easier to solve the MAP equation than it is to find the conditional mean. When an efficient estimate does not exist, we do not know how closely the mean-square error, using either $\hat{a}_{ms}(\mathbf{R})$ or $\hat{a}_{map}(\mathbf{R})$, approaches the lower bound. Asymptotic results similar to those for real variables may be derived.

### 2.4.3 Multiple Parameter Estimation

In many problems of interest we shall want to estimate more than one parameter. A familiar example is the radar problem in which we shall estimate the range and velocity of a target. Most of the ideas and techniques can be extended to this case in a straightforward manner. The model is shown in Fig. 2.23. If there are $K$ parameters, $a_1, a_2, \ldots, a_K$, we describe them by a parameter vector $\mathbf{a}$ in a $K$-dimensional space. The other elements of the model are the same as before. We shall consider both the case in which $\mathbf{a}$ is a random parameter vector and that in which $\mathbf{a}$ is a real (or nonrandom) parameter vector. Three issues are of interest. In each the result is the vector analog to a result in the scalar case.

1. Estimation procedures.
2. Measures of error.
3. Bounds on performance.

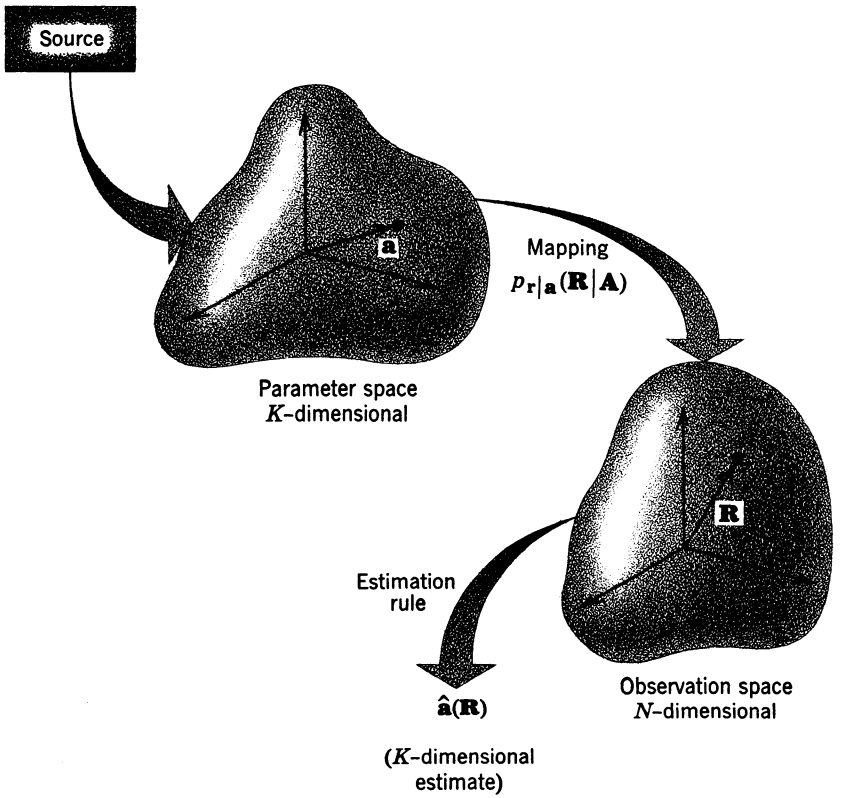


**Fig. 2.23 Multiple parameter estimation model.**

***Estimation Procedure.*** For random variables we could consider the general case of Bayes estimation in which we minimize the risk for some arbitrary scalar cost function $C(\mathbf{a}, \hat{\mathbf{a}})$, but for our purposes it is adequate to consider only cost functions that depend on the error. We define the error vector as

$$\mathbf{a}_\epsilon(\mathbf{R}) = \begin{bmatrix} \hat{a}_1(\mathbf{R}) - a_1 \\ \hat{a}_2(\mathbf{R}) - a_2 \\ \vdots \\ \hat{a}_K(\mathbf{R}) - a_K \end{bmatrix} = \hat{\mathbf{a}}(\mathbf{R}) - \mathbf{a}. \tag{229}$$

For a mean-square error criterion, the cost function is simply

$$C(\mathbf{a}_\epsilon(\mathbf{R})) \triangleq \sum_{i=1}^{K} a_{\epsilon_i}^2(\mathbf{R}) = \mathbf{a}_\epsilon^T(\mathbf{R})\, \mathbf{a}_\epsilon(\mathbf{R}). \tag{230}$$

This is just the sum of the squares of the errors. The risk is

$$\mathcal{R}_{\text{ms}} = \int_{-\infty}^{\infty}\int C(\mathbf{a}_\epsilon(\mathbf{R}))\, p_{\mathbf{r},\mathbf{a}}(\mathbf{R}, \mathbf{A})\, d\mathbf{R}\, d\mathbf{A} \tag{231}$$

or

$$\mathcal{R}_{\text{ms}} = \int_{-\infty}^{\infty} p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R} \int_{-\infty}^{\infty} \left[\sum_{i=1}^{K} (\hat{a}_i(\mathbf{R}) - A_i)^2\right] p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})\, d\mathbf{A}. \tag{232}$$

As before, we can minimize the inner integral for each $\mathbf{R}$. Because the terms in the sum are positive, we minimize them separately. This gives

$$\hat{a}_{\text{ms}_i}(\mathbf{R}) = \int_{-\infty}^{\infty} A_i p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})\, d\mathbf{A} \tag{233}$$

or

$$\hat{\mathbf{a}}_{\text{ms}}(\mathbf{R}) = \int_{-\infty}^{\infty} \mathbf{A} p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})\, d\mathbf{A}. \tag{234}$$

It is easy to show that mean-square estimation commutes over *linear transformations*. Thus, if

$$\mathbf{b} = \mathbf{D}\mathbf{a}, \tag{235}$$

where $\mathbf{D}$ is a $L \times K$ matrix, and we want to minimize

$$E[\mathbf{b}_\epsilon^T(\mathbf{R})\, \mathbf{b}_\epsilon(\mathbf{R})] = E\left[\sum_{i=1}^{L} b_{\epsilon_i}^2(\mathbf{R}),\right] \tag{236}$$

the result will be,

$$\hat{\mathbf{b}}_{\text{ms}}(\mathbf{R}) = \mathbf{D}\hat{\mathbf{a}}_{\text{ms}}(\mathbf{R}) \tag{237}$$

[see Problem 2.4.20 for the proof of (237)].

For MAP estimation we must find the value of $\mathbf{A}$ that maximizes $p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})$. If the maximum is interior and $\partial \ln p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})/\partial A_i$ exists at the maximum then a necessary condition is obtained from the MAP equations. By analogy with (137) we take the logarithm of $p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})$, differentiate with respect to each parameter $A_i$, $i = 1, 2, \ldots, K$, and set the result equal to zero. This gives a set of $K$ simultaneous equations:

$$\frac{\partial \ln p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})}{\partial A_i}\bigg|_{\mathbf{A}=\hat{\mathbf{a}}_{\mathrm{map}}(\mathbf{R})} = 0, \quad i = 1, 2, \ldots, K. \tag{238}$$

We can write (238) in a more compact manner by defining a partial derivative matrix operator

$$\nabla_{\mathbf{A}} \triangleq \begin{bmatrix} \dfrac{\partial}{\partial A_1} \\ \dfrac{\partial}{\partial A_2} \\ \vdots \\ \dfrac{\partial}{\partial A_K} \end{bmatrix}. \tag{239}$$

This operator can be applied only to $1 \times m$ matrices; for example,

$$\nabla_{\mathbf{A}}\, \mathbf{G} = \begin{bmatrix} \dfrac{\partial G_1}{\partial A_1} & \dfrac{\partial G_2}{\partial A_1} & \cdots & \dfrac{\partial G_m}{\partial A_1} \\ \vdots & & & \\ \dfrac{\partial G_1}{\partial A_K} & & & \dfrac{\partial G_m}{\partial A_K} \end{bmatrix}. \tag{240}$$

Several useful properties of $\nabla_{\mathbf{A}}$ are developed in Problems 2.4.27–28. In our case (238) becomes a single vector equation,

$$\nabla_{\mathbf{A}}[\ln p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})]|_{\mathbf{A}=\hat{\mathbf{a}}_{\mathrm{map}}(\mathbf{R})} = \mathbf{0}. \tag{241}$$

Similarly, for ML estimates we must find the value of $\mathbf{A}$ that maximizes $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$. If the maximum is interior and $\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})/\partial A_i$ exists at the maximum then a necessary condition is obtained from the likelihood equations:

$$\nabla_{\mathbf{A}}[\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})]|_{\mathbf{A}=\hat{\mathbf{a}}_{\mathrm{ml}}(\mathbf{R})} = \mathbf{0}. \tag{242}$$

In both cases we must verify that we have the absolute maximum.

***Measures of Error.*** For nonrandom variables the first measure of interest is the bias. Now the bias is a vector,

$$\mathbf{B}(\mathbf{A}) \triangleq E[\mathbf{a}_\epsilon(\mathbf{R})] = E[\hat{\mathbf{a}}(\mathbf{R})] - \mathbf{A}. \tag{243}$$

If each component of the bias vector is zero for every **A**, we say that the estimate is unbiased.

In the single parameter case a rough measure of the spread of the error was given by the variance of the estimate. In the special case in which $a_\epsilon(\mathbf{R})$ was Gaussian this provided a complete description:

$$p_{a_\epsilon}(A_\epsilon) = \frac{1}{\sqrt{2\pi}\,\sigma_{a_\epsilon}} \exp\left(-\frac{A_\epsilon^2}{2\sigma_{a_\epsilon}^2}\right). \tag{244}$$

For a vector variable the quantity analogous to the variance is the covariance matrix

$$E[(\mathbf{a}_\epsilon - \bar{\mathbf{a}}_\epsilon)(\mathbf{a}_\epsilon^T - \bar{\mathbf{a}}_\epsilon^T)] \triangleq \mathbf{\Lambda}_\epsilon, \tag{245}$$

where

$$\bar{\mathbf{a}}_\epsilon \triangleq E(\mathbf{a}_\epsilon) = \mathbf{B}(\mathbf{A}). \tag{246}$$

The best way to determine how the covariance matrix provides a measure of spread is to consider the special case in which the $a_{\epsilon_i}$ are jointly Gaussian. For algebraic simplicity we let $E(\mathbf{a}_\epsilon) = \mathbf{0}$. The joint probability density for a set of $K$ jointly Gaussian variables is

$$p_{\mathbf{a}_\epsilon}(\mathbf{A}_\epsilon) = (|2\pi|^{K/2}|\mathbf{\Lambda}_\epsilon|^{1/2})^{-1} \exp(-\tfrac{1}{2}\mathbf{A}_\epsilon^T\mathbf{\Lambda}_\epsilon^{-1}\mathbf{A}_\epsilon) \tag{247}$$

(e.g., p. 151 in Davenport and Root [1]).

The probability density for $K = 2$ is shown in Fig. 2.24*a*. In Figs. 2.24*b*,*c* we have shown the equal-probability contours of two typical densities.From (247) we observe that the equal-height contours are defined by the relation,

$$\mathbf{A}_\epsilon^T\mathbf{\Lambda}_\epsilon^{-1}\mathbf{A}_\epsilon = C^2, \tag{248}$$

which is the equation for an ellipse when $K = 2$. The ellipses move out monotonically with increasing $C$. They also have the interesting property that the probability of being inside the ellipse is only a function of $C^2$.

**Property.** For $K = 2$, the probability that the error vector lies inside an ellipse whose equation is

$$\mathbf{A}_\epsilon^T\mathbf{\Lambda}_\epsilon^{-1}\mathbf{A}_\epsilon = C^2, \tag{249}$$

is

$$P = 1 - \exp\left(-\frac{C^2}{2}\right). \tag{250}$$

*Proof.* The area inside the ellipse defined by (249) is

$$\mathcal{A} = |\mathbf{\Lambda}_\epsilon|^{1/2}\pi C^2. \tag{251}$$

The differential area between ellipses corresponding to $C$ and $C + dC$ respectively is

$$d\mathcal{A} = |\mathbf{\Lambda}_\epsilon|^{1/2}2\pi C\,dC. \tag{252}$$

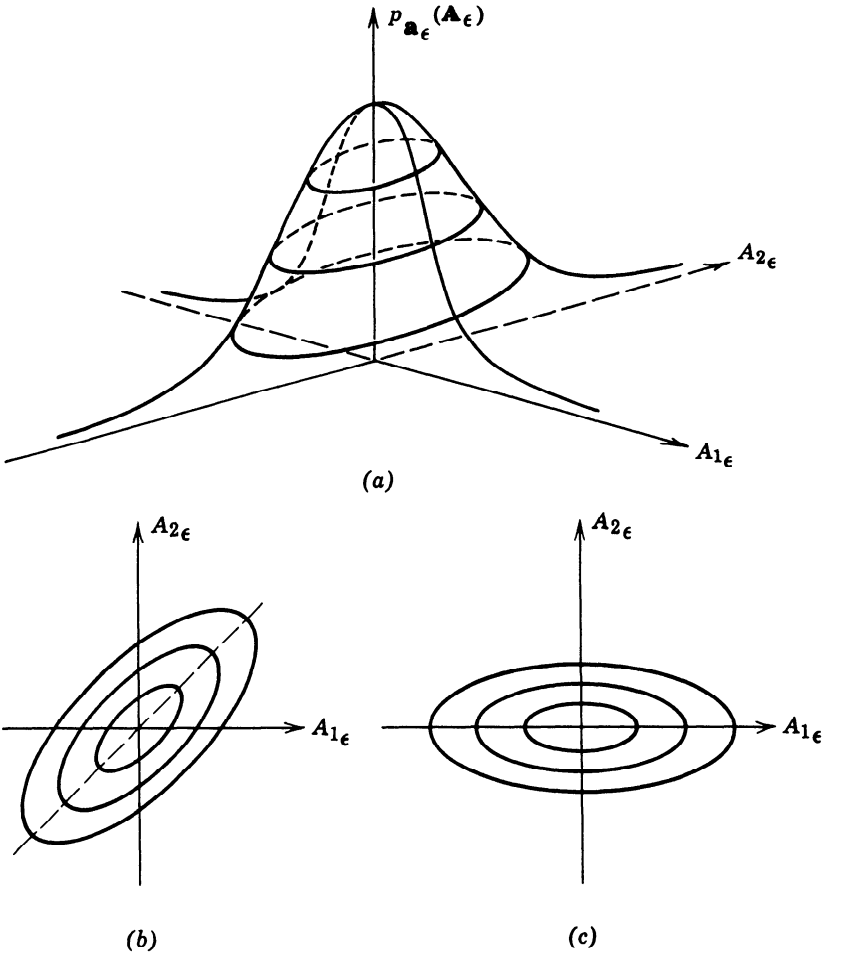


**Fig. 2.24 Gaussian densities: [*a*] two-dimensional Gaussian density; [*b*] equal-height contours, correlated variables; [*c*] equal-height contours, uncorrelated variables.**

The height of the probability density in this differential area is

$$(2\pi|\Lambda_\epsilon|^{1/2})^{-1} \exp\left(-\frac{C^2}{2}\right). \tag{253}$$

We can compute the probability of a point lying outside the ellipse by multiplying (252) by (253) and integrating from $C$ to $\infty$.

$$1 - P = \int_C^\infty X \exp\left(-\frac{X^2}{2}\right) dX = \exp\left(-\frac{C^2}{2}\right), \tag{254}$$

which is the desired result.

For this reason the ellipses described by (248) are referred to as *concentration ellipses* because they provide a measure of the concentration of the density.

A similar result holds for arbitrary $K$. Now, (248) describes an *ellipsoid.* Here the differential volume† in $K$-dimensional space is

$$dv = |\mathbf{\Lambda}_\epsilon|^{1/2} \frac{\pi^{K/2}}{\Gamma(K/2 + 1)} KC^{K-1}\, dC. \tag{255}$$

The value of the probability density on the ellipsoid is

$$[(2\pi)^{K/2}|\mathbf{\Lambda}_\epsilon|^{1/2}]^{-1} \exp\left(-\frac{C^2}{2}\right). \tag{256}$$

Therefore

$$1 - P = \frac{K}{(2)^{K/2}\Gamma(K/2 + 1)} \int_C^\infty X^{K-1} e^{-X^2/2}\, dX, \tag{257}$$

which is the desired result. We refer to these ellipsoids as *concentration ellipsoids.*

When the probability density of the error is *not* Gaussian, the concentration ellipsoid no longer specifies a unique probability. This is directly analogous to the one-dimensional case in which the variance of a non-Gaussian zero-mean random variable does not determine the probability density. We can still interpret the concentration ellipsoid as a rough measure of the spread of the errors. When the concentration ellipsoids of a given density lie wholly outside the concentration ellipsoids of a second density, we say that the second density is more concentrated than the first. With this motivation, we derive some properties and bounds pertaining to concentration ellipsoids.

***Bounds on Estimation Errors: Nonrandom Variables.*** In this section we derive two bounds. The first relates to the variance of an individual error; the second relates to the concentration ellipsoid.

**Property 1.** Consider *any* unbiased estimate of $A_i$. Then

$$\sigma_{\epsilon_i}^2 \triangleq \text{Var}\,[\hat{a}_i(\mathbf{R}) - A_i] \geq J^{ii}, \tag{258}$$

where $J^{ii}$ is the $ii$th element in the $K \times K$ square matrix $\mathbf{J}^{-1}$. The elements in $\mathbf{J}$ are

$$\begin{aligned} J_{ij} &\triangleq E\left[\frac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_i} \cdot \frac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_j}\right] \\ &= -E\left[\frac{\partial^2 \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_i\, \partial A_j}\right] \end{aligned} \tag{259}$$

† e.g., Cramér [9], p. 120, or Sommerfeld [32].

or

$$\begin{aligned}\mathbf{J} &\triangleq E(\{\nabla_{\mathbf{A}}[\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})]\}\{\nabla_{\mathbf{A}}[\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})]\}^T) \\ &= -E[\nabla_{\mathbf{A}}(\{\nabla_{\mathbf{A}}[\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})]\}^T)].\end{aligned} \tag{260}$$

The **J** matrix is commonly called *Fisher's information matrix*. The equality in (258) holds if and only if

$$\hat{a}_i(\mathbf{R}) - A_i = \sum_{j=1}^{K} k_{ij}(\mathbf{A}) \frac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_j} \tag{261}$$

for all values of $A_i$ and **R**.

In other words, the estimation error can be expressed as the weighted sum of the partial derivatives of $\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$ with respect to the various parameters.

*Proof.* Because $\hat{a}_i(\mathbf{R})$ is unbiased,

$$\int_{-\infty}^{\infty} [\hat{a}_i(\mathbf{R}) - A_i] p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})\, d\mathbf{R} = 0 \tag{262}$$

or

$$\int_{-\infty}^{\infty} \hat{a}_i(\mathbf{R}) p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})\, d\mathbf{R} = A_i. \tag{263}$$

Differentiating both sides with respect to $A_j$, we have

$$\begin{aligned}&\int_{-\infty}^{\infty} \hat{a}_i(\mathbf{R}) \frac{\partial p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_j} d\mathbf{R} \\ &\qquad = \int_{-\infty}^{\infty} \hat{a}_i(\mathbf{R}) \frac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_j} p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})\, d\mathbf{R} = \delta_{ij}.\end{aligned} \tag{264}$$

We shall prove the result for $i = 1$. We define a $K + 1$ vector

$$\mathbf{x} = \begin{bmatrix} \hat{a}_1(\mathbf{R}) - A_1 \\ \dfrac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_1} \\ \vdots \\ \dfrac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_K} \end{bmatrix}. \tag{265}$$

The covariance matrix is

$$E[\mathbf{x}\mathbf{x}^T] = \begin{bmatrix} \sigma_{\epsilon_1}^2 & 1 & 0 & 0 & 0 \\ 1 & J_{11} & J_{12} & \cdots & J_{1K} \\ 0 & \vdots & & \ddots & \vdots \\ 0 & J_{K1} & & & J_{KK} \end{bmatrix}. \tag{266}$$

[The ones and zeroes in the matrix follow from (264).] Because it is a covariance matrix, it is nonnegative definite, which implies that the determinant of the entire matrix is greater than or equal to zero. (This condition is only necessary, not sufficient, for the matrix to be nonnegative definite.)

Evaluating the determinant using a cofactor expansion, we have

$$\sigma_{\epsilon_1}{}^2|\mathbf{J}| - \text{cofactor } J_{11} \geq 0. \tag{267}$$

If we assume that $\mathbf{J}$ is nonsingular, then

$$\sigma_{\epsilon_1}{}^2 \geq \frac{\text{cofactor } J_{11}}{|\mathbf{J}|} = J^{11}, \tag{268}$$

which is the desired result. The modifications for the case when $\mathbf{J}$ is singular follow easily for any specific problem.

In order for the determinant to equal zero, the term $\hat{A}_1(\mathbf{R}) - A_1$ must be expressible as a linear combination of the other terms. This is the condition described by (261). The second line of (259) follows from the first line in a manner exactly analogous to the proof in (189)–(192). The proof for $i \neq 1$ is an obvious modification.

**Property 2.** Consider *any* unbiased estimate of $\mathbf{A}$. The concentration ellipse

$$\mathbf{A}_\epsilon{}^T \mathbf{\Lambda}_\epsilon{}^{-1} \mathbf{A}_\epsilon = C^2 \tag{269}$$

lies either outside or on the bound ellipse defined by

$$\mathbf{A}_\epsilon{}^T \mathbf{J} \mathbf{A}_\epsilon = C^2. \tag{270}$$

*Proof.* We shall go through the details for $K = 2$. By analogy with the preceding proof, we construct the covariance matrix of the vector.

$$\mathbf{x} = \begin{bmatrix} \hat{a}_1(\mathbf{R}) - A_1 \\ \hat{a}_2(\mathbf{R}) - A_2 \\ \dfrac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_1} \\ \dfrac{\partial \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_2} \end{bmatrix}. \tag{271}$$

Then

$$E[\mathbf{x}\mathbf{x}^T] = \left[\begin{array}{cc:cc} \sigma_{1_\epsilon}{}^2 & \rho\sigma_{1_\epsilon}\sigma_{2_\epsilon} & 1 & 0 \\ \rho\sigma_{1_\epsilon}\sigma_{2_\epsilon} & \sigma_{2_\epsilon}{}^2 & 0 & 1 \\ \hdashline 1 & 0 & J_{11} & J_{12} \\ 0 & 1 & J_{21} & J_{22} \end{array}\right] = \left[\begin{array}{c:c} \mathbf{\Lambda}_\epsilon & \mathbf{I} \\ \hdashline \mathbf{I} & \mathbf{J} \end{array}\right]. \tag{272}$$

The second equality defines a partition of the $4 \times 4$ matrix into four $2 \times 2$ matrices. Because it is a covariance matrix, it is nonnegative definite. Using a formula for the determinant of a partitioned matrix,† we have

$$|\mathbf{\Lambda}_\epsilon \mathbf{J} - \mathbf{I}| \geq 0 \tag{273}$$

or, assuming that $\mathbf{\Lambda}_\epsilon$ is nonsingular and applying the product rule for determinants,

$$|\mathbf{\Lambda}_\epsilon| \, |\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}| \geq 0. \tag{274}$$

This implies

$$|\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}| \geq 0. \tag{275}$$

Now consider the two ellipses. The intercept on the $A_{\epsilon_1}$ axis is

$$A_{1_\epsilon}^2 \Big|_{A_{2\epsilon}=0} = C^2 \frac{|\mathbf{\Lambda}_\epsilon|}{\sigma_2^2} \tag{276}$$

for the actual concentration ellipse and

$$A_{1_\epsilon}^2 \Big|_{A_{2\epsilon}=0} = C^2 \frac{1}{J_{11}} \tag{277}$$

for the bound ellipse.

We want to show that the actual intercept is greater than or equal to the bound intercept. This requires

$$J_{11}|\mathbf{\Lambda}_\epsilon| \geq \sigma_2^2. \tag{278}$$

This inequality follows because the determinant of the $3 \times 3$ matrix in the upper left corner of (272) is greater than or equal to zero. (Otherwise the entire matrix is not nonnegative definite, e.g. [16] or [18].) Similarly, the actual intercept on the $A_{2_\epsilon}$ axis is greater than or equal to the bound intercept. Therefore the actual ellipse is either always outside (or on) the bound ellipse *or* the two ellipses intersect.

*If* they intersect, we see from (269) and (270) that there must be a solution, $\mathbf{A}_\epsilon$, to the equation

$$\mathbf{A}_\epsilon^T \mathbf{\Lambda}_\epsilon^{-1} \mathbf{A}_\epsilon = \mathbf{A}_\epsilon^T \mathbf{J} \mathbf{A}_\epsilon \tag{279}$$

or

$$\mathbf{A}_\epsilon^T [\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}] \mathbf{A}_\epsilon \triangleq \mathbf{A}_\epsilon^T \mathbf{D} \mathbf{A}_\epsilon = 0. \tag{280}$$

In scalar notation

$$A_{1_\epsilon}^2 D_{11} + 2A_{1_\epsilon} A_{2_\epsilon} D_{12} + A_{2_\epsilon}^2 D_{22} = 0 \tag{281}$$

or, equivalently,

$$\left(\frac{A_{1_\epsilon}}{A_{2_\epsilon}}\right)^2 D_{11} + 2\left(\frac{A_{1_\epsilon}}{A_{2_\epsilon}}\right) D_{12} + D_{22} = 0. \tag{282}$$

† Bellman [16], p. 83.

Solving for $A_{1_\epsilon}/A_{2_\epsilon}$, we would obtain real roots only if the discriminant were greater than or equal to zero. This requires

$$|\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}| \leq 0. \tag{283}$$

The inequality is a contradiction of (275). One possibility is $|\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}| = 0$, but this is true only when the ellipses coincide. In this case all the estimates are efficient.

For arbitrary $K$ we can show that $\mathbf{J} - \mathbf{\Lambda}_\epsilon^{-1}$ is nonnegative definite. The implications with respect to the concentration ellipsoids are the same as for $K = 2$.

Frequently we want to estimate functions of the $K$ basic parameters rather than the parameters themselves. We denote the desired estimates as

$$\begin{aligned} d_1 &= g_{d_1}(\mathbf{A}), \\ d_2 &= g_{d_2}(\mathbf{A}), \\ &\vdots \\ d_M &= g_{d_M}(\mathbf{A}). \end{aligned} \tag{284}$$

or

$$\mathbf{d} = \mathbf{g}_d(\mathbf{A})$$

The number of estimates $M$ is not related to $K$ in general. The functions may be nonlinear. The estimation error is

$$\hat{d}_i - g_i(\mathbf{A}) \triangleq d_{\epsilon_i}. \tag{285}$$

If we assume that the estimates are unbiased and denote the error covariance matrix as $\mathbf{\Lambda}_\epsilon$, then by using methods identical to those above we can prove the following properties.

**Property 3.** The matrix

$$\mathbf{\Lambda}_\epsilon - \{\nabla_{\mathbf{A}}[\mathbf{g}_d^T(\mathbf{A})]\}^T \mathbf{J}^{-1} \{\nabla_{\mathbf{A}}[\mathbf{g}_d^T(\mathbf{A})]\} \tag{286}$$

is nonnegative definite.

This implies the following property (just multiply the second matrix out and recall that all diagonal elements of nonnegative definite matrix are nonnegative):

**Property 4.**

$$\text{Var}\,(d_{\epsilon_i}) \geq \sum_i^K \sum_j^K \frac{\partial g_{d_i}(\mathbf{A})}{\partial A_i} J^{ij} \frac{\partial g_{d_i}(\mathbf{A})}{\partial A_j}. \tag{287}$$

For the special case in which the desired functions are linear, the result in (287) can be written in a simpler form.

**Property 5.** Assume that

$$\mathbf{g}_d(\mathbf{A}) \triangleq \mathbf{G}_d\mathbf{A}, \tag{288}$$

where $\mathbf{G}_d$ is an $M \times K$ matrix. If the estimates are unbiased, then

$$\mathbf{\Lambda}_\epsilon - \mathbf{G}_d\mathbf{J}^{-1}\mathbf{G}_d^T$$

is nonnegative definite.

**Property 6.** Efficiency commutes with linear transformations but does not commute with nonlinear transformations. In other words, if $\hat{\mathbf{a}}$ is efficient, then $\hat{\mathbf{d}}$ will be efficient if and only if $\mathbf{g}_d(\mathbf{A})$ is a linear transformation.

***Bounds on Estimation Errors: Random Parameters.*** Just as in the single parameter case, the bound for random parameters is derived by a straightforward modification of the derivation for nonrandom parameters. The information matrix now consists of two parts:

$$\mathbf{J}_T \triangleq \mathbf{J}_D + \mathbf{J}_P. \tag{289}$$

The matrix $\mathbf{J}_D$ is the information matrix defined in (260) and represents information obtained from the *data*. The matrix $\mathbf{J}_P$ represents the a priori information. The elements are

$$\begin{aligned} J_{P_{ij}} &\triangleq E\left[\frac{\partial \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_i}\frac{\partial \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_j}\right] \\ &= -E\left[\frac{\partial^2 \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_i\,\partial A_j}\right]. \end{aligned} \tag{290}$$

The *correlation matrix* of the errors is

$$\mathbf{R}_\epsilon \triangleq E(\mathbf{a}_\epsilon\mathbf{a}_\epsilon^T). \tag{291}$$

The diagonal elements represent the mean-square errors and the off-diagonal elements are the cross correlations. Three properties follow easily:

**Property No. 1.**

$$E[a_{\epsilon_i}^2] \geq J_T^{ii}. \tag{292}$$

In other words, the diagonal elements in the inverse of the total information matrix are lower bounds on the corresponding mean-square errors.

**Property No. 2.** The matrix

$$\mathbf{J}_T - \mathbf{R}_\epsilon^{-1}$$

is nonnegative definite. This has the same physical interpretation as in the nonrandom parameter problem.

**Property No. 3.** If $\mathbf{J}_T = \mathbf{R}_\epsilon^{-1}$, all of the estimates are efficient. A necessary and sufficient condition for this to be true is that $p_{\mathbf{a}|\mathbf{r}}(\mathbf{A}|\mathbf{R})$ be Gaussian for all $\mathbf{R}$. This will be true if $\mathbf{J}_T$ is constant. [Modify (261), (228)].

A special case of interest occurs when the a priori density is a $K$th-order Gaussian density. Then

$$\mathbf{J}_P = \mathbf{\Lambda}_{\mathbf{a}}^{-1}, \tag{293}$$

where $\mathbf{\Lambda}_{\mathbf{a}}$ is the covariance matrix of the random parameters.

An even simpler case arises when the variables are independent Gaussian variables. Then

$$J_{P_{ij}} = \frac{1}{\sigma_{a_i}^2}\,\delta_{ij}, \tag{294}$$

Under these conditions only the diagonal terms of $\mathbf{J}_T$ are affected by the a priori information.

Results similar to Properties 3 to 6 for nonrandom parameters can be derived for the random parameter case.

### 2.4.4 Summary of Estimation Theory

In this section we developed the estimation theory results that we shall need for the problems of interest. We began our discussion with Bayes estimation of random parameters. The basic quantities needed in the model were the a priori density $p_a(A)$, the probabilistic mapping to the observation space $p_{\mathbf{r}|a}(\mathbf{R}|A)$, and a cost function $C(A_\epsilon)$. These quantities enabled us to find the risk. The estimate which minimized the risk was called a Bayes estimate and the resulting risk, the Bayes risk. Two types of Bayes estimate, the MMSE estimate (which was the mean of the a posteriori density) and the MAP estimate (the mode of the a posteriori density), were emphasized. In Properties 1 and 2 (pp. 60–61) we saw that the conditional mean was the Bayes estimate for a large class of cost functions when certain conditions on the cost function and a posteriori density were satisfied.

Turning to nonrandom parameter estimation, we introduced the idea of bias and variance as two separate error measures. The Cramér-Rao inequality provided a bound on the variance of any unbiased estimate. Whenever an efficient estimate existed, the maximum likelihood estimation procedure gave this estimate. This property of the ML estimate, coupled with its asymptotic properties, is the basis for our emphasis on ML estimates.

The extension to multiple parameter estimation involved no new concepts. Most of the properties were just multidimensional extensions of the corresponding scalar result.

It is important to emphasize the close relationship between detection and estimation theory. Both theories are based on a likelihood function or likelihood ratio, which, in turn, is derived from the probabilistic transition

mechanism. As we proceed to more difficult problems, we shall find that a large part of the work is the manipulation of this transition mechanism. In many cases the mechanism will not depend on whether the problem is one of detection or estimation. Thus the difficult part of the problem will be applicable to either problem. This close relationship will become even more obvious as we proceed. We now return to the detection theory problem and consider a more general model.

## 2.5 COMPOSITE HYPOTHESES

In Sections 2.2 and 2.3 we confined our discussion to the decision problem in which the hypotheses were simple. We now extend our discussion to the case in which the hypotheses are composite. The term composite is most easily explained by a simple example.

***Example 1.*** Under hypothesis 0 the observed variable $r$ is Gaussian with zero mean and variance $\sigma^2$. Under hypothesis 1 the observed variable $r$ is Gaussian with mean $m$ and variance $\sigma^2$. The value of $m$ can be anywhere in the interval $[M_0, M_1]$. Thus

$$\begin{aligned} H_0 &: p_{r|H_0}(R|H_0) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R^2}{2\sigma^2}\right), \\ H_1 &: p_{r|H_1,m}(R|H_1,M) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(R-M)^2}{2\sigma^2}\right), \qquad M_0 \le M \le M_1. \end{aligned} \tag{295}$$

We refer to $H_1$ as a composite hypothesis because the parameter value $M$, which characterizes the hypothesis, ranges over a set of values. A model of this decision problem is shown in Fig. 2.25*a*. The output of the source is a parameter value $M$, which we view as a point in a parameter space $\chi$. We then define the hypotheses as subspaces of $\chi$. In this case $H_0$ corresponds to the point $M = 0$ and $H_1$ corresponds to the interval $[M_0, M_1.]$ We assume that the probability density governing the mapping from the parameter space to the observation space, $p_{r|m}(R|M)$, is known for all values of $M$ in $\chi$.

The final component is a decision rule that divides the observation space into two parts which correspond to the two possible decisions. It is important to observe that we are interested *solely* in making a decision and that the actual value of $M$ is not of interest to us. For this reason the parameter $M$ is frequently referred to as an "unwanted" parameter.

The extension of these ideas to the general composite hypothesis-testing problem is straightforward. The model is shown in Fig. 2.25*b*. The output of the source is a set of parameters. We view it as a point in a parameter space $\chi$ and denote it by the vector $\boldsymbol{\theta}$. The hypotheses are subspaces of $\chi$. (In Fig. 2.25*b* we have indicated nonoverlapping spaces for convenience.) The probability density governing the mapping from the parameter space to the observation space is denoted by $p_{\mathbf{r}|\boldsymbol{\theta}}(\mathbf{R}|\boldsymbol{\theta})$ and is assumed to be known for all values of $\boldsymbol{\theta}$ in $\chi$. Once again, the final component is a decision rule.

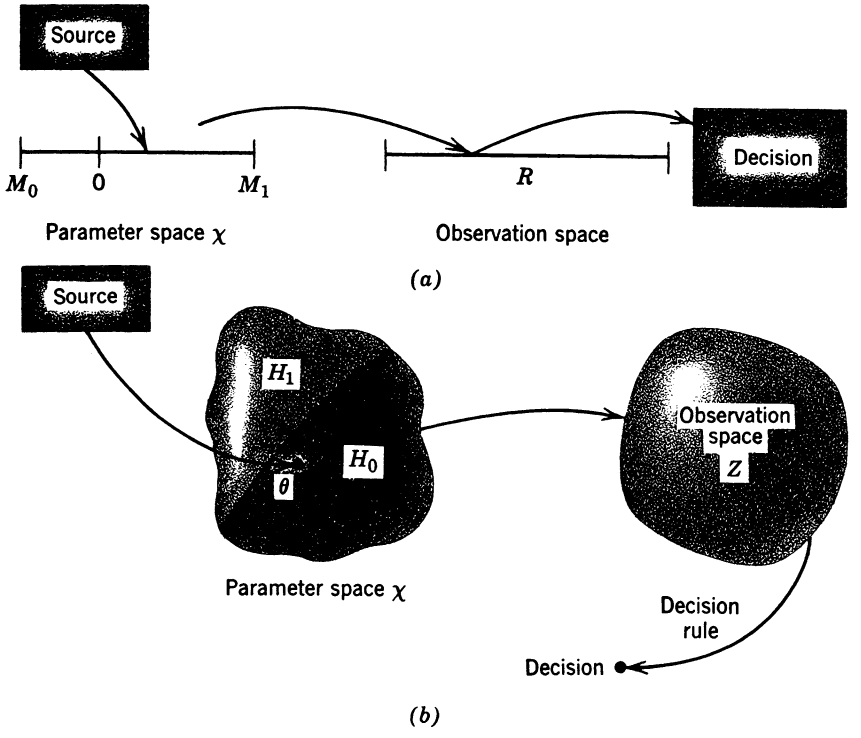


**Fig. 2.25** ***a.*** **Composite hypothesis testing problem for single-parameter example.** ***b.*** **Composite hypothesis testing problem.**

To complete the formulation, we must characterize the parameter $\boldsymbol{\theta}$. Just as in the parameter estimation case the parameter $\boldsymbol{\theta}$ may be a nonrandom or random variable. If $\boldsymbol{\theta}$ is a random variable with a known probability density, the procedure is straightforward. Denoting the probability density of $\boldsymbol{\theta}$ on the two hypotheses as $p_{\boldsymbol{\theta}|H_0}(\boldsymbol{\theta}|H_0)$ and $p_{\boldsymbol{\theta}|H_1}(\boldsymbol{\theta}|H_1)$, the likelihood ratio is

$$\Lambda(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} = \frac{\int_\chi p_{\mathbf{r}|\boldsymbol{\theta}}(\mathbf{R}|\boldsymbol{\theta},H_1)p_{\boldsymbol{\theta}|H_1}(\boldsymbol{\theta}|H_1)\,d\boldsymbol{\theta}}{\int_\chi p_{\mathbf{r}|\boldsymbol{\theta}}(\mathbf{R}|\boldsymbol{\theta},H_0)p_{\boldsymbol{\theta}|H_0}(\boldsymbol{\theta}|H_0)\,d\boldsymbol{\theta}}. \tag{296}$$

The reason for this simplicity is that the known probability density on $\boldsymbol{\theta}$ enables us to reduce the problem to a simple hypothesis-testing problem by integrating over $\boldsymbol{\theta}$. We can illustrate this procedure for the model in Example 1.

***Example 1*** *(continued.)* We assume that the probability density governing $m$ on $H_1$ is

$$p_{m|H_1}(M|H_1) = \frac{1}{\sqrt{2\pi}\,\sigma_m} \exp\left(-\frac{M^2}{2\sigma_m^2}\right), \qquad -\infty < M < \infty, \tag{297}$$

Then (296) becomes

$$\Lambda(R) = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(R-M)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi}\,\sigma_m} \exp\left(-\frac{M^2}{2\sigma_m^2}\right) dM}{\dfrac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\dfrac{R^2}{2\sigma^2}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{298}$$

Integrating and taking the logarithm of both sides, we obtain

$$R^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2\sigma^2(\sigma^2 + \sigma_m^2)}{\sigma_m^2}\left[\ln \eta + \frac{1}{2}\ln\left(1 + \frac{\sigma_m^2}{\sigma^2}\right)\right]. \tag{299}$$

This result is equivalent to Example 2 on p. 29 because the density used in (297) makes the two problems identical.

As we expected, the test uses only the magnitude of $R$ because the mean $m$ has a symmetric probability density.

For the general case given in (296) the actual calculation may be more involved, but the desired procedure is well defined.

When $\boldsymbol{\theta}$ is a random variable with an unknown density, the best test procedure is not clearly specified. One possible approach is a minimax test over the unknown density. An alternate approach is to try several densities based on any partial knowledge of $\boldsymbol{\theta}$ that is available. In many cases the test structure will be insensitive to the detailed behavior of the probability density.

The second case of interest is the case in which $\boldsymbol{\theta}$ is a nonrandom variable. Here, just as in the problem of estimating nonrandom variables, we shall try a procedure and investigate the results. A first observation is that, because $\boldsymbol{\theta}$ has no probability density over which to average, a Bayes test is not meaningful. Thus we can devote our time to Neyman-Pearson tests.

We begin our discussion by examining what we call a *perfect measurement* bound on the test performance. We illustrate this idea for the problem in Example 1.

***Example 2.*** In this case $\theta = M$.

From (295)

$$H_1: p_{r|m}(R|M) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(R-M)^2}{2\sigma^2}\right), \qquad (M_0 \le M \le M_1),$$

and (300)

$$H_0: p_{r|m}(R|M) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

where $M$ is an unknown nonrandom parameter.

It is clear that whatever test we design can never be better than a hypothetical test in which the receiver first measures $M$ perfectly (or, alternately, it is told $M$) and then designs the optimum likelihood ratio test. Thus we can bound the ROC of any test by the ROC of this fictitious perfect measurement test. For this example we could use the ROC's in Fig. 2.9*a* by letting $d^2 = M^2/\sigma^2$. Because we are interested in the behavior versus $M$, the format in Fig. 2.9*b* is more useful. This is shown in Fig. 2.26. Such a curve is called a *power function*. It is simply a plot of $P_D$ for all values of $M$ (more generally $\theta$) for various values of $P_F$. Because $H_0 = H_1$ for $M = 0$, $P_D = P_F$. The curves in Fig. 2.26 represent a bound on how well any test could do. We now want to see how close the actual test performance comes to this bound.

The best performance we could achieve would be obtained if an actual test's curves equaled the bound for all $M \in \chi$. We call such tests *uniformly most powerful* (UMP). In other words, for a given $P_F$ a UMP test has a $P_D$ greater than or equal to any other test for *all* $M \in \chi$. The conditions for a UMP test to exist can be seen in Fig. 2.27.

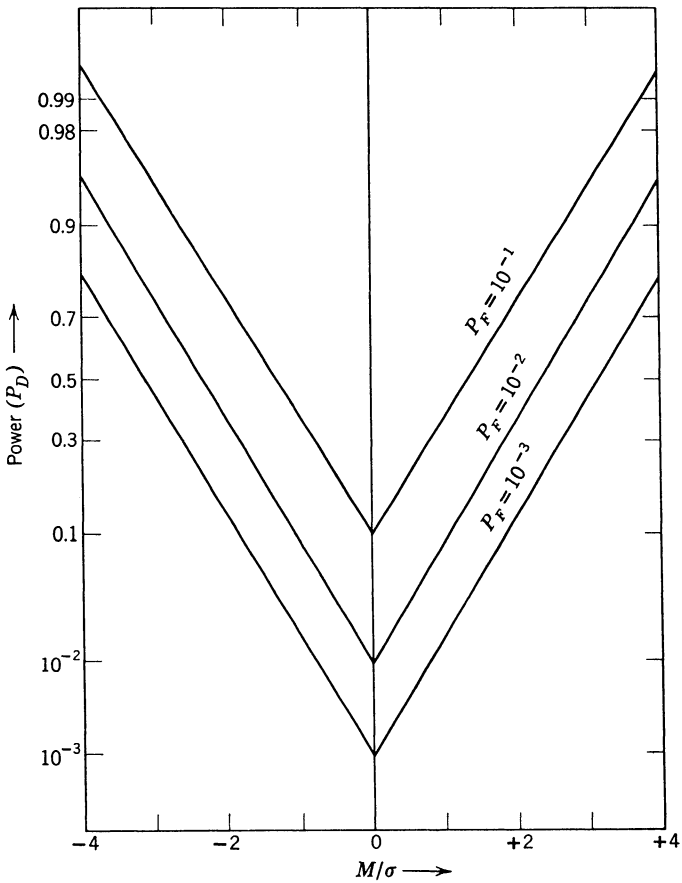


**Fig. 2.26 Power function for perfect measurement test.**

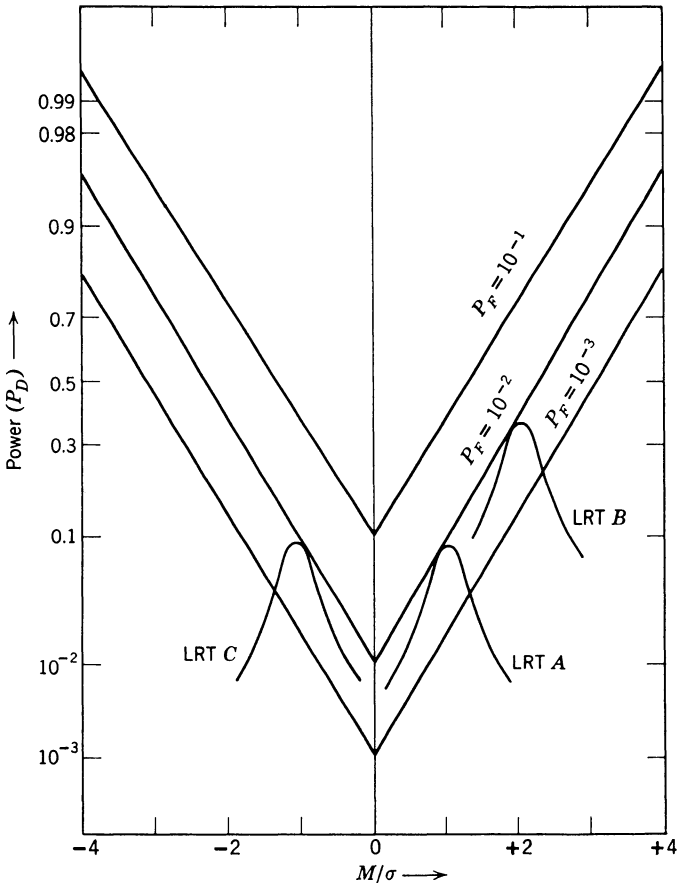


**Fig. 2.27 Power functions for various likelihood ratio tests.**

We first construct the perfect measurement bound. We next consider other possible tests and their performances. Test A is an ordinary likelihood ratio test designed under the assumption that $M = 1$. The first observation is that the power of this test equals the bound at $M = 1$, which follows from the manner in which we constructed the bound. For other values of $M$ the power of test A may or may not equal the bound. Similarly, test B is a likelihood ratio test designed under the assumption that $M = 2$, and test C is a likelihood ratio test designed under the assumption that $M = -1$. In each case their power equals the bound at their design points. (The power functions in Fig. 2.27 are drawn to emphasize this and are not quantitatively correct away from the design point. The quantitatively correct curves are shown in Fig. 2.29.) They may also equal the bound at other points. The conditions for a UMP test are now obvious. We must be able to design a complete likelihood ratio test (including the threshold) for every $M \in \chi$ without knowing $M$.

The analogous result for the general case follows easily.

It is clear that in general the bound can be reached for any particular $\boldsymbol{\theta}$ simply by designing an ordinary LRT for that particular $\boldsymbol{\theta}$. Now a UMP test must be as good as any other test for every $\boldsymbol{\theta}$. This gives us a necessary and sufficient condition for its existence.

**Property.** A UMP test exists if and only if the likelihood ratio test for every $\boldsymbol{\theta} \in \chi$ can be completely defined (including threshold) without knowledge of $\boldsymbol{\theta}$.

The "if" part of the property is obvious. The "only if" follows directly from our discussion in the preceding paragraph. If there exists some $\boldsymbol{\theta} \in \chi$ for which we cannot find the LRT without knowing $\boldsymbol{\theta}$, we should have to use some other test, because we do not know $\boldsymbol{\theta}$. This test will necessarily be inferior for that particular $\boldsymbol{\theta}$ to a LRT test designed for that particular $\theta$ and therefore is not *uniformly* most powerful.

Returning to our example and using the results in Fig. 2.8, we know that the likelihood ratio test is

$$R \underset{H_0}{\overset{H_1}{\gtrless}} \gamma^+, \tag{301}$$

and

$$P_F = \int_{\gamma^+}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R^2}{2\sigma^2}\right) dR, \qquad \text{if } M > 0. \tag{302}$$

(The superscript + emphasizes the test assumes $M > 0$. The value of $\gamma^+$ may be negative.) This is shown in Fig. 2.28*a*.

Similarly, for the case in which $M < 0$ the likelihood ratio test is

$$R \underset{H_1}{\overset{H_0}{\gtrless}} \gamma^-, \tag{303}$$

where

$$P_F = \int_{-\infty}^{\gamma^-} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R^2}{2\sigma^2}\right) dR, \qquad M < 0. \tag{304}$$

This is shown in Fig. 2.28*b*. We see that the threshold is just the negative of the threshold for $M > 0$. This reversal is done to get the largest portion of $p_{r|H_1}(R|H_1)$ inside the $H_1$ region (and therefore maximize $P_D$).

Thus, with respect to Example 1, we draw the following conclusions:

1. If $M$ can take on *only* nonnegative values (i.e., $M_0 \geq 0$), a UMP test exists [use (301)].
2. If $M$ can take on *only* nonpositive values (i.e., $M_1 \leq 0$), a UMP test exists [use (303)].
3. If $M$ can take on both negative and positive values (i.e., $M_0 < 0$ and $M_1 > 0$), then a UMP test does not exist. In Fig. 2.29 we show the power function for a likelihood ratio test designed under the assumption that $M$ was positive. For negative values of $M$, $P_D$ is less than $P_F$ because the threshold is on the wrong side.

Whenever a UMP test exists, we use it, and the test works as well as if we knew $\boldsymbol{\theta}$. A more difficult problem is presented when a UMP test does

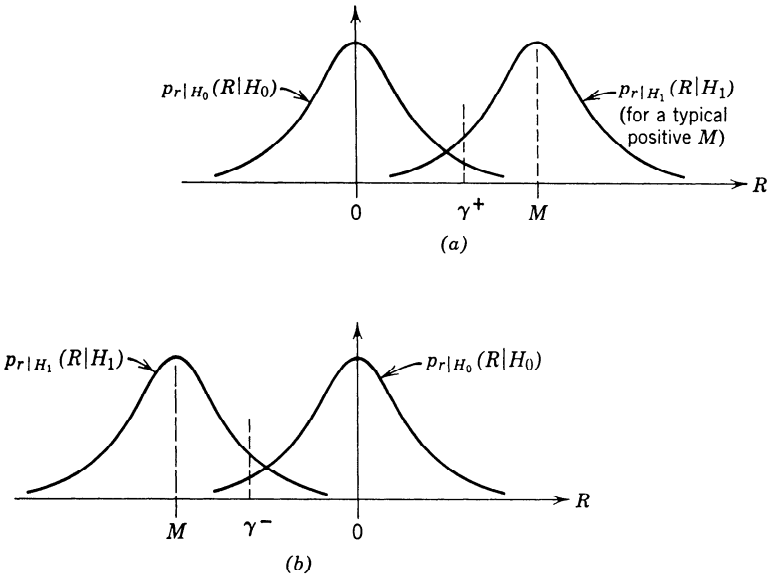


**Fig. 2.28 Effect of sign of $M$: [$a$] threshold for positive $M$; [$b$] threshold for negative $M$.**

not exist. The next step is to discuss other possible tests for the cases in which a UMP test does not exist. We confine our discussion to one possible test procedure. Others are contained in various statistics texts (e.g., Lehmann [17]) but seem to be less appropriate for the physical problems of interest in the sequel.

The perfect measurement bound suggests that a logical procedure is to estimate $\boldsymbol{\theta}$ assuming $H_1$ is true, then estimate $\boldsymbol{\theta}$ assuming $H_0$ is true, and use these estimates in a likelihood ratio test as if they were correct. If the maximum likelihood estimates discussed on p. 65 are used, the result is called a *generalized likelihood ratio test.* Specifically,

$$\Lambda_g(\mathbf{R}) = \frac{\max\limits_{\boldsymbol{\theta}_1} p_{\mathbf{r}|\boldsymbol{\theta}_1}(\mathbf{R}|\boldsymbol{\theta}_1)}{\max\limits_{\boldsymbol{\theta}_0} p_{\mathbf{r}|\boldsymbol{\theta}_0}(\mathbf{R}|\boldsymbol{\theta}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{305}$$

where $\boldsymbol{\theta}_1$ ranges over all $\boldsymbol{\theta}$ in $H_1$ and $\boldsymbol{\theta}_0$ ranges over all $\boldsymbol{\theta}$ in $H_0$. In other words, we make a ML estimate of $\boldsymbol{\theta}_1$, assuming that $H_1$ is true. We then evaluate $p_{\mathbf{r}|\boldsymbol{\theta}_1}(\mathbf{R}|\boldsymbol{\theta}_1)$ for $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1$ and use this value in the numerator. A similar procedure gives the denominator.

A simple example of a generalized LRT is obtained by using a slightly modified version of Example 1.

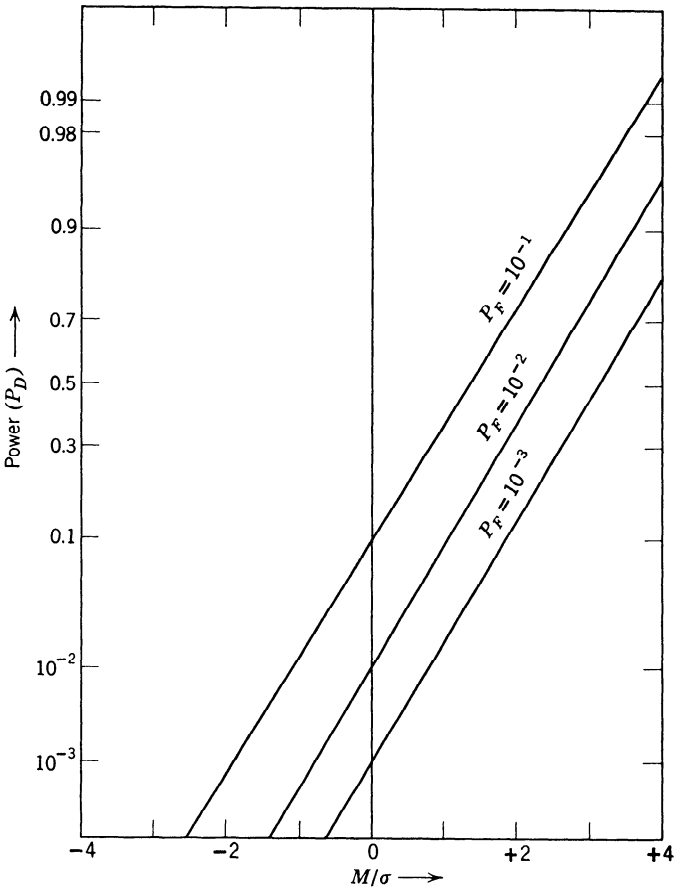


**Fig. 2.29 Performance of LRT assuming positive $M$.**

***Example 2.*** The basic probabilities are the same as in Example 1. Once again, $\theta = M$. Instead of one, we have $N$ independent observations, which we denote by the vector $\mathbf{R}$. The probability densities are,

$$p_{\mathbf{r}|m,H_1}(\mathbf{R}|M,H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(R_i - M)^2}{2\sigma^2}\right),$$
$$p_{\mathbf{r}|m,H_0}(R|M,H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right). \tag{306}$$

In this example $H_1$ is a composite hypothesis and $H_0$, a simple hypothesis. From (198)

$$\hat{M}_1 = \frac{1}{N}\sum_{i=1}^{N} R_i. \tag{307}$$

Then

$$\Lambda_g(\mathbf{R}) = \frac{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{[R_i - (1/N)\sum_{j=1}^{N} R_j]^2}{2\sigma^2}\right\}}{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-R_i^2/2\sigma^2\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{308}$$

Canceling common terms and taking the logarithm, we have

$$\ln \Lambda_g(\mathbf{R}) = \frac{1}{2\sigma^2 N}\left(\sum_{i=1}^{N} R_i\right)^2 \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma. \tag{309}$$

The left side of (309) is always greater than or equal to zero. Thus, $\gamma$ can always be chosen greater than or equal to one. Therefore, an equivalent test is

$$\left(\frac{1}{N^{1/2}} \sum_{i=1}^{N} R_i\right)^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_1^2 \tag{310}$$

where $\gamma_1 \geq 0$. Equivalently,

$$|z| \triangleq \left|\frac{1}{N^{1/2}} \sum_{i=1}^{N} R_i\right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_1. \tag{311}$$

The power function of this test follows easily. The variable $z$ has a variance equal

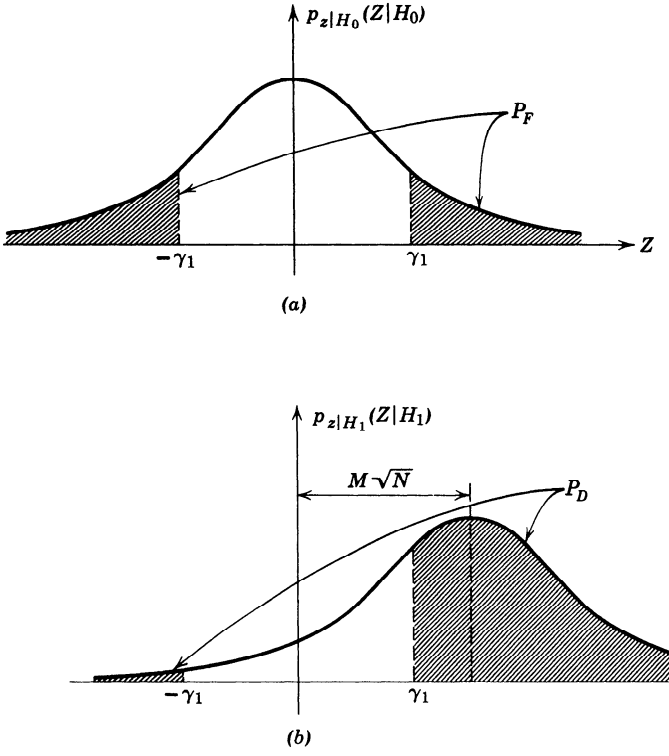


**Fig. 2.30 Errors in generalized likelihood ratio test: [a] $P_F$ calculation; [b] $P_D$ calculation.**

to $\sigma^2$. On $H_0$ its mean is zero and on $H_1$ its mean is $M\sqrt{N}$. The densities are sketched in Fig. 2.30.

$$P_F = \int_{-\infty}^{-\gamma_1} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{Z^2}{2\sigma^2}\right) dZ + \int_{\gamma_1}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{Z^2}{2\sigma^2}\right) dZ$$

$$= 2\,\mathrm{erfc}_*\left(\frac{\gamma_1}{\sigma}\right) \tag{312}$$

and

$$P_D(M) = \int_{-\infty}^{-\gamma_1} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(Z - M\sqrt{N})^2}{2\sigma^2}\right] dZ$$

$$+ \int_{\gamma_1}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(Z - M\sqrt{N})^2}{2\sigma^2}\right] dZ$$

$$= \mathrm{erfc}_*\left[\frac{\gamma_1 + M\sqrt{N}}{\sigma}\right] + \mathrm{erfc}_*\left[\frac{\gamma_1 - M\sqrt{N}}{\sigma}\right]. \tag{313}$$

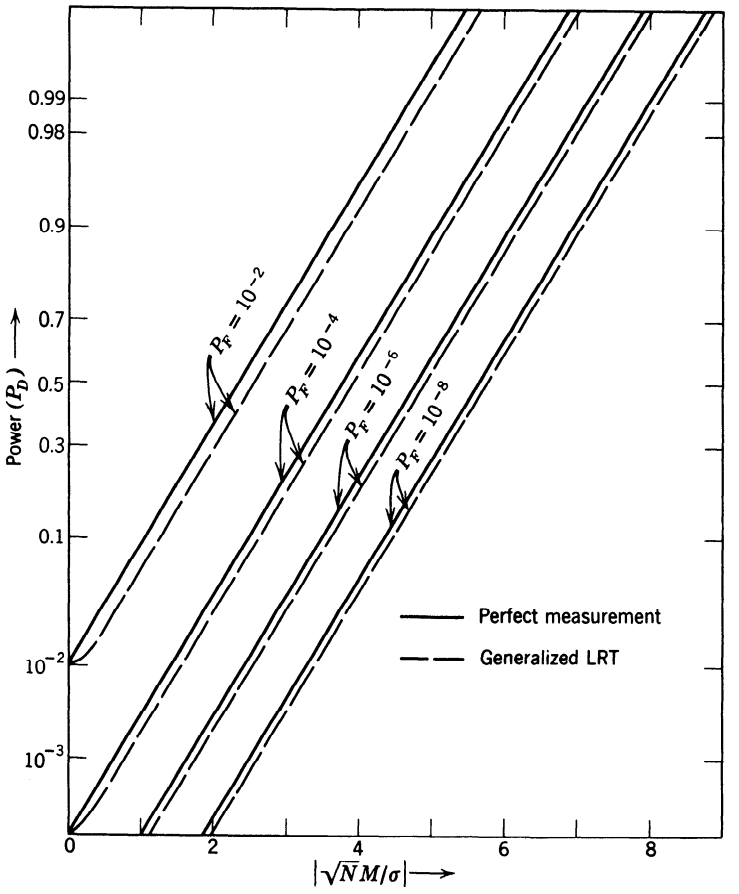


**Fig. 2.31 Power function: generalized likelihood ratio tests.**

The resulting power function is plotted in Fig. 2.31. The perfect measurement bound is shown for comparison purposes. As we would expect from our discussion of ML estimates, the difference approaches zero as $\sqrt{N}\, M/\sigma \to \infty$.

Just as there are cases in which the ML estimates give poor results, there are others in which the generalized likelihood ratio test may give bad results. In these cases we must look for other test procedures. Fortunately, in most of the physical problems of interest to us either a UMP test will exist or a generalized likelihood ratio test will give satisfactory results.

## 2.6 THE GENERAL GAUSSIAN PROBLEM

All of our discussion up to this point has dealt with arbitrary probability densities. In the binary detection case $p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)$ and $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$ were not constrained to have any particular form. Similarly, in the estimation problem $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$ was not constrained. In the classical case, constraints are not particularly necessary. When we begin our discussion of the waveform problem, we shall find that most of our discussions concentrate on problems in which the conditional density of $\mathbf{r}$ is Gaussian. We discuss this class of problem in detail in this section. The material in this section and the problems associated with it lay the groundwork for many of the results in the sequel. We begin by defining a Gaussian random vector and the general Gaussian problem.

**Definition.** A set of random variables $r_1, r_2, \ldots, r_N$ are defined as jointly Gaussian if all their linear combinations are Gaussian random variables.

**Definition.** A vector $\mathbf{r}$ is a Gaussian random vector when its components $r_1, r_2, \ldots, r_N$ are jointly Gaussian random variables.

In other words, if

$$z = \sum_{i=1}^{N} g_i r_i \triangleq \mathbf{G}^T\mathbf{r} \tag{314}$$

is a Gaussian random variable for all finite $\mathbf{G}^T$, then $\mathbf{r}$ is a Gaussian vector.

If we define

$$E(\mathbf{r}) = \mathbf{m} \tag{315}$$

and

$$\text{Cov}\,(\mathbf{r}) = E[(\mathbf{r} - \mathbf{m})(\mathbf{r}^T - \mathbf{m}^T)] \triangleq \mathbf{\Lambda}, \tag{316}$$

then (314) implies that the characteristic function of $\mathbf{r}$ is

$$M_{\mathbf{r}}(j\mathbf{v}) \triangleq E[e^{j\mathbf{v}^T\mathbf{r}}] = \exp\,(+j\mathbf{v}^T\mathbf{m} - \tfrac{1}{2}\mathbf{v}^T\mathbf{\Lambda}\mathbf{v}) \tag{317}$$

and assuming $\mathbf{\Lambda}$ is nonsingular the probability density of $\mathbf{r}$ is

$$p_{\mathbf{r}}(\mathbf{R}) = [(2\pi)^{N/2}|\mathbf{\Lambda}|^{1/2}]^{-1} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}^T)\mathbf{\Lambda}^{-1}(\mathbf{R} - \mathbf{m})\right]. \quad (318)$$

The proof is straightforward (e.g., Problem 2.6.20).

**Definition.** A hypothesis testing problem is called a general Gaussian problem if $p_{\mathbf{r}|H_i}(\mathbf{R}|H_i)$ is a Gaussian density on all hypotheses. An estimation problem is called a general Gaussian problem if $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$ has a Gaussian density for all $\mathbf{A}$.

We discuss the binary hypothesis testing version of the general Gaussian problem in detail in the text. The $M$-hypothesis and the estimation problems are developed in the problems. The basic model for the binary detection problem is straightforward. We assume that the observation space is $N$-dimensional. Points in the space are denoted by the $N$-dimensional vector (or column matrix) $\mathbf{r}$:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}. \quad (319)$$

Under the first hypothesis $H_1$ we assume that $\mathbf{r}$ is a Gaussian random vector, which is completely specified by its mean vector and covariance matrix. We denote these quantities as

$$E[\mathbf{r}|H_1] = \begin{bmatrix} E(r_1|H_1) \\ E(r_2|H_1) \\ \vdots \\ E(r_N|H_1) \end{bmatrix} \triangleq \begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{1N} \end{bmatrix} \triangleq \mathbf{m}_1. \quad (320)$$

The covariance matrix is

$$\mathbf{K}_1 \triangleq E[(\mathbf{r} - \mathbf{m}_1)(\mathbf{r}^T - \mathbf{m}_1{}^T)|H_1]$$

$$= \begin{bmatrix} {}_1K_{11} & {}_1K_{12} & {}_1K_{13} & \cdots & {}_1K_{1N} \\ {}_1K_{21} & {}_1K_{22} & \ddots & & \\ \vdots & \vdots & & & \vdots \\ {}_1K_{N1} & & & & {}_1K_{NN} \end{bmatrix}. \quad (321)$$

We define the inverse of $\mathbf{K}_1$ as $\mathbf{Q}_1$

$$\mathbf{Q}_1 \triangleq \mathbf{K}_1^{-1} \quad (322)$$

$$\mathbf{Q}_1\mathbf{K}_1 = \mathbf{K}_1\mathbf{Q}_1 = \mathbf{I}, \quad (323)$$

where **I** is the identity matrix (ones on the diagonal and zeroes elsewhere). Using (320), (321), (322), and (318), we may write the probability density of **r** on $H_1$,

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = [(2\pi)^{N/2}|\mathbf{K}_1|^{1/2}]^{-1} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_1{}^T)\mathbf{Q}_1(\mathbf{R} - \mathbf{m}_1)\right]. \quad (324)$$

Going through a similar set of definitions for $H_0$, we obtain the probability density

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = [(2\pi)^{N/2}|\mathbf{K}_0|^{1/2}]^{-1} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_0{}^T)\mathbf{Q}_0(\mathbf{R} - \mathbf{m}_0)\right]. \quad (325)$$

Using the definition in (13), the likelihood ratio test follows easily:

$$\Lambda(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} = \frac{|\mathbf{K}_0|^{1/2} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_1{}^T)\mathbf{Q}_1(\mathbf{R} - \mathbf{m}_1)\right]}{|\mathbf{K}_1|^{1/2} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_0{}^T)\mathbf{Q}_0(\mathbf{R} - \mathbf{m}_0)\right]} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (326)$$

Taking logarithms, we obtain

$$\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_0{}^T)\,\mathbf{Q}_0(\mathbf{R} - \mathbf{m}_0) - \tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_1{}^T)\,\mathbf{Q}_1(\mathbf{R} - \mathbf{m}_1) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta + \tfrac{1}{2}\ln|\mathbf{K}_1| - \tfrac{1}{2}\ln|\mathbf{K}_0| \triangleq \gamma^*. \quad (327)$$

We see that the test consists of finding the difference between two *quadratic forms*. The result in (327) is basic to many of our later discussions. For this reason we treat various cases of the general Gaussian problem in some detail. We begin with the simplest.

### 2.6.1 Equal Covariance Matrices

The first special case of interest is the one in which the covariance matrices on the two hypotheses are equal,

$$\mathbf{K}_1 = \mathbf{K}_0 \triangleq \mathbf{K}, \quad (328)$$

but the means are different.

Denote the inverse as **Q**:

$$\mathbf{Q} = \mathbf{K}^{-1}. \quad (329)$$

Substituting into (327), multiplying the matrices, canceling common terms, and using the symmetry of **Q**, we have

$$(\mathbf{m}_1{}^T - \mathbf{m}_0{}^T)\mathbf{Q}\mathbf{R} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta + \tfrac{1}{2}(\mathbf{m}_1{}^T\mathbf{Q}\mathbf{m}_1 - \mathbf{m}_0{}^T\mathbf{Q}\mathbf{m}_0) \triangleq \gamma'_*. \quad (330)$$

We can simplify this expression by defining a vector corresponding to the difference in the mean value vectors on the two hypotheses:

$$\Delta\mathbf{m} \triangleq \mathbf{m}_1 - \mathbf{m}_0. \quad (331)$$

Then (327) becomes

$$\boxed{l(\mathbf{R}) \triangleq \Delta\mathbf{m}^T\mathbf{Q}\mathbf{R} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'_*} \tag{332}$$

or, equivalently,

$$l(\mathbf{R}) \triangleq \mathbf{R}^T\mathbf{Q}\,\Delta\mathbf{m} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'_*. \tag{333}$$

The quantity on the left is a *scalar* Gaussian random variable, for it was obtained by a linear transformation of jointly Gaussian random variables. Therefore, as we discussed in Example 1 on pp. 36–38, we can completely characterize the performance of the test by the quantity $d^2$. In that example, we defined $d$ as the distance between the means on the two hypothesis when the variance was normalized to equal one. An identical definition is,

$$d^2 \triangleq \frac{[E(l|H_1) - E(l|H_0)]^2}{\text{Var}\,(l|H_0)}. \tag{334}$$

Substituting (320) into the definition of $l$, we have

$$E(l|H_1) = \Delta\mathbf{m}^T\mathbf{Q}\mathbf{m}_1 \tag{335}$$

and

$$E(l|H_0) = \Delta\mathbf{m}^T\mathbf{Q}\mathbf{m}_0. \tag{336}$$

Using (332), (333), and (336) we have

$$\text{Var}\,[l|H_0] = E\{[\Delta\mathbf{m}^T\mathbf{Q}(\mathbf{R} - \mathbf{m}_0)][(\mathbf{R}^T - \mathbf{m}_0{}^T)\mathbf{Q}\,\Delta\mathbf{m}]\}. \tag{337}$$

Using (321) to evaluate the expectation and then (323), we have

$$\text{Var}\,[l|H_0] = \Delta\mathbf{m}^T\mathbf{Q}\,\Delta\mathbf{m}. \tag{338}$$

Substituting (335), (336), and (338) into (334), we obtain

$$\boxed{d^2 = \Delta\mathbf{m}^T\mathbf{Q}\,\Delta\mathbf{m}}. \tag{339}$$

Thus the performance for the equal covariance Gaussian case is completely determined by the quadratic form in (339). We now interpret it for some cases of interest.

**Case 1. Independent Components with Equal Variance.** Each $r_i$ has the same variance $\sigma^2$ and is statistically independent. Thus

$$\mathbf{K} = \sigma^2\mathbf{I} \tag{340}$$

and

$$\mathbf{Q} = \frac{1}{\sigma^2}\mathbf{I}. \tag{341}$$

Substituting (341) into (339), we obtain

$$d^2 = \Delta\mathbf{m}^T \frac{1}{\sigma^2} \mathbf{I}\, \Delta\mathbf{m} = \frac{1}{\sigma^2} \Delta\mathbf{m}^T \Delta\mathbf{m} = \frac{1}{\sigma^2} |\Delta\mathbf{m}|^2 \tag{342}$$

or

$$\boxed{d = \frac{|\Delta\mathbf{m}|}{\sigma}}. \tag{343}$$

We see that $d$ corresponds to the *distance* between the two mean-value vectors divided by the standard deviation of $R_i$. This is shown in Fig. 2.32. In (332) we see that

$$l = \frac{1}{\sigma^2} \Delta\mathbf{m}^T\mathbf{R}. \tag{344}$$

Thus the sufficient statistic is just the dot (or scalar) product of the observed vector $\mathbf{R}$ and the mean difference vector $\Delta\mathbf{m}$.

**Case 2. Independent Components with Unequal Variances.** Here the $r_i$ are statistically independent but have unequal variances. Thus

$$\mathbf{K} = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_N^2 \end{bmatrix} \tag{345}$$

and

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & 0 \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_N^2} \end{bmatrix}. \tag{346}$$

Substituting into (339) and performing the multiplication, we have

$$d^2 = \sum_{i=1}^{N} \frac{(\Delta m_i)^2}{\sigma_i^2}. \tag{347}$$

Now the various difference components contribute to $d^2$ with weighting that is inversely proportional to the variance along that coordinate. We can also interpret the result as distance in a new coordinate system.

Let

$$\Delta\mathbf{m}' = \begin{bmatrix} \frac{1}{\sigma_1}\,\Delta m_1 \\ \frac{1}{\sigma_2}\,\Delta m_2 \\ \vdots \\ \frac{1}{\sigma_N}\,\Delta m_N \end{bmatrix} \tag{348}$$

and

$$R_i' = \frac{1}{\sigma_i} R_i. \tag{349}$$

This transformation changes the scale on each axis so that the variances are all equal to one. We see that $d$ corresponds exactly to the difference vector in this "scaled" coordinate system.

The sufficient statistic is

$$l(\mathbf{R}) = \sum_{i=1}^{N} \frac{\Delta m_i \cdot R_i}{\sigma_i^2}. \tag{350}$$

In the scaled coordinate system it is the dot product of the two vectors

$$l(\mathbf{R}') = \Delta\mathbf{m}'^T\mathbf{R}'. \tag{351}$$

**Case 3.** This is the general case. A satisfactory answer for $l$ and $d$ is already available in (332) and (339):

$$l(\mathbf{R}) = \Delta\mathbf{m}^T\mathbf{Q}\mathbf{R} \tag{352}$$

and

$$d^2 = \Delta\mathbf{m}^T\mathbf{Q}\,\Delta\mathbf{m}. \tag{353}$$

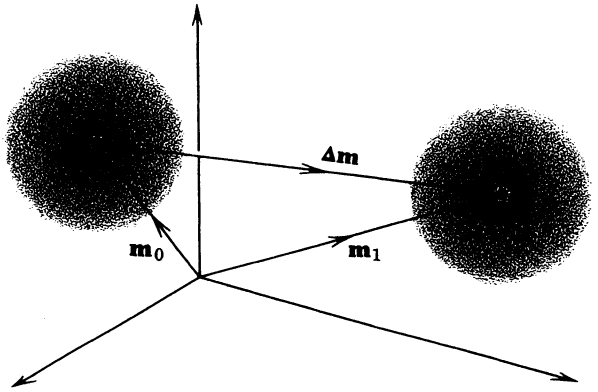


**Fig. 2.32 Mean-value vectors.**

Valuable insight into the important features of the problem can be gained by looking at it in a different manner.

The key to the simplicity in Cases 1 and 2 is the diagonal covariance matrix. This suggests that we try to represent $\mathbf{R}$ in a new coordinate system in which the components are statistically independent random variables. In Fig. 2.33*a* we show the observation in the original coordinate system. In Fig. 2.33*b* we show a new set of coordinate axes, which we denote by the orthogonal unit vectors $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N$:

$$\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j = \delta_{ij}. \tag{354}$$

We denote the observation in the new coordinate system by $\mathbf{r}'$. We want to choose the orientation of the new system so that the components $r_i'$ and $r_j'$ are uncorrelated (and therefore statistically independent, for they are Gaussian) for all $i \neq j$. In other words,

$$E[(r_i' - m_i')(r_j' - m_j')] = \lambda_i \delta_{ij}, \tag{355}$$

where

$$m_i' \triangleq E(r_i') \tag{356}$$

and

$$\text{Var}\,[r_i'] \triangleq \lambda_i. \tag{357}$$

Now the components of $\mathbf{r}'$ can be expressed simply in terms of the dot product of the original vector $\mathbf{r}$ and the various unit vectors

$$r_i' = \mathbf{r}^T \boldsymbol{\phi}_i = \boldsymbol{\phi}_i^T \mathbf{r}. \tag{358}$$

Using (358) in (355), we obtain

$$E[\boldsymbol{\phi}_i^T (\mathbf{r} - \mathbf{m})(\mathbf{r}^T - \mathbf{m}^T)\boldsymbol{\phi}_j] = \lambda_i \delta_{ij}. \tag{359}$$

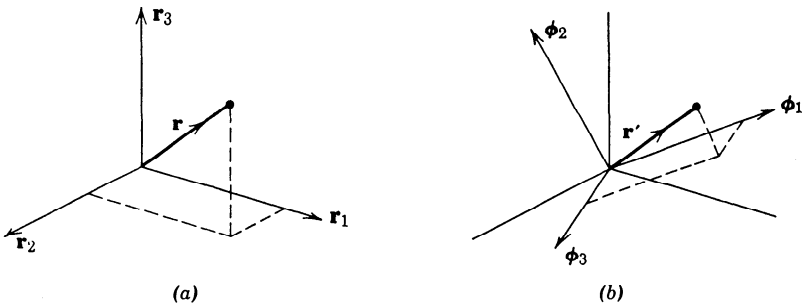


**Fig. 2.33 Coordinate systems: [*a*] original coordinate system; [*b*] new coordinate system.**

The expectation of the random part is just $\mathbf{K}$ [see (321)]. Therefore (359) becomes

$$\lambda_i \delta_{ij} = \boldsymbol{\phi}_i^T \mathbf{K} \boldsymbol{\phi}_j. \tag{360}$$

This will be satisfied if and only if

$$\lambda_j \boldsymbol{\phi}_j = \mathbf{K} \boldsymbol{\phi}_j \qquad \text{for } j = 1, 2, \ldots, N. \tag{361}$$

To check the "if" part of this result, substitute (361) into (360):

$$\lambda_i \delta_{ij} = \boldsymbol{\phi}_i^T \lambda_j \boldsymbol{\phi}_j = \lambda_j \delta_{ij}, \tag{362}$$

where the right equality follows from (354). The "only if" part follows using a simple proof by contradiction. Now (361) can be written with the $j$ subscript suppressed:

$$\boxed{\lambda \boldsymbol{\phi} = \mathbf{K} \boldsymbol{\phi}.} \tag{363}$$

We see that the question of finding the proper coordinate system reduces to the question of whether we can find $N$ solutions to (363) that satisfy (354).

It is instructive to write (363) out in detail. Each $\boldsymbol{\phi}$ is a vector with $N$ components:

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_N \end{bmatrix}. \tag{364}$$

Substituting (364) into (363), we have

$$\begin{aligned} K_{11}\phi_1 + K_{12}\phi_2 + \cdots + K_{1N}\phi_N &= \lambda\phi_1 \\ K_{21}\phi_1 + K_{22}\phi_2 + \cdots + K_{2N}\phi_N &= \lambda\phi_2 \\ \vdots \\ K_{N1}\phi_1 + K_{N2}\phi_2 + \cdots + K_{NN}\phi_N &= \lambda\phi_N \end{aligned} \tag{365}$$

We see that (365) corresponds to a set of $N$ homogeneous simultaneous equations. A nontrivial solution will exist if and only if the determinant of the coefficient matrix is zero. In other words, if and only if

$$|\mathbf{K} - \lambda \mathbf{I}| = \begin{vmatrix} K_{11} - \lambda & K_{12} & K_{13} & \cdots \\ K_{21} & K_{22} - \lambda & K_{23} & \\ K_{31} & K_{32} & \ddots & \\ \vdots & \vdots & & K_{NN} - \lambda \end{vmatrix} = 0. \tag{366}$$

We see that this is an $N$th-order polynomial in $\lambda$. The $N$ roots, denoted by $\lambda_1, \lambda_2, \ldots, \lambda_N$, are called the *eigenvalues* of the covariance matrix $\mathbf{K}$. It can be shown that the following properties are true (e.g., [16] or [18]):

1. Because $\mathbf{K}$ is symmetric, the eigenvalues are real.
2. Because $\mathbf{K}$ is a covariance matrix, the eigenvalues are nonnegative. (Otherwise we would have random variables with negative variances.)

For each $\lambda_i$ we can find a solution $\boldsymbol{\phi}_i$ to (363). Because there is an arbitrary constant associated with each solution to (363), we may choose the $\boldsymbol{\phi}_i$ to have unit length

$$\boldsymbol{\phi}_i^T \boldsymbol{\phi}_i = 1. \tag{367}$$

These solutions are called the normalized *eigenvectors* of $\mathbf{K}$. Two other properties may also be shown for symmetric matrices.

3. If the roots $\lambda_i$ are distinct, the corresponding eigenvectors are orthogonal.
4. If a particular root $\lambda_j$ is of multiplicity $M$, the $M$ associated eigenvectors are linearly independent. They can be chosen to be orthonormal.

We have now described a coordinate system in which the observations are statistically independent. The mean difference vector can be expressed as

$$\begin{aligned} \Delta m_1' &= \boldsymbol{\phi}_1^T \Delta \mathbf{m} \\ \Delta m_2' &= \boldsymbol{\phi}_2^T \Delta \mathbf{m} \\ &\vdots \\ \Delta m_N' &= \boldsymbol{\phi}_N^T \Delta \mathbf{m} \end{aligned} \tag{368}$$

or in vector notation

$$\Delta \mathbf{m}' = \begin{bmatrix} \boldsymbol{\phi}_1^T \\ \hline \boldsymbol{\phi}_2^T \\ \hline \vdots \\ \hline \boldsymbol{\phi}_N^T \end{bmatrix} \Delta \mathbf{m} \triangleq \mathbf{W}\, \Delta \mathbf{m}. \tag{369}$$

The resulting sufficient statistic in the new coordinate system is

$$l(\mathbf{R}') = \sum_{i=1}^{N} \frac{\Delta m_i' \cdot R_i'}{\lambda_i} \tag{370}$$

and $d^2$ is

$$\boxed{d^2 = \sum_{i=1}^{N} \frac{(\Delta m_i')^2}{\lambda_i}.} \tag{371}$$

The derivation leading to (371) has been somewhat involved, but the result is of fundamental importance, for it demonstrates that there always exists a coordinate system in which the random variables are uncorrelated and that the new system is related to the old system by a linear transformation. To illustrate the technique we consider a simple example.

***Example.*** For simplicity we let $N = 2$ and $\mathbf{m}_0 = \mathbf{0}$. Let

$$\mathbf{K} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{372}$$

and

$$\mathbf{m}_1 = \begin{bmatrix} m_{11} \\ m_{12} \end{bmatrix}. \tag{373}$$

To find the eigenvalues we solve

$$\begin{vmatrix} 1 - \lambda & \rho \\ \rho & 1 - \lambda \end{vmatrix} = 0 \tag{374}$$

or

$$(1 - \lambda)^2 - \rho^2 = 0. \tag{375}$$

Solving,

$$\begin{aligned} \lambda_1 &= 1 + \rho, \\ \lambda_2 &= 1 - \rho. \end{aligned} \tag{376}$$

To find $\boldsymbol{\phi}_1$ we substitute $\lambda_1$ into (365),

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} = \begin{bmatrix} (1 + \rho)\phi_{11} \\ (1 + \rho)\phi_{12} \end{bmatrix} \tag{377}$$

Solving, we obtain

$$\phi_{11} = \phi_{12}. \tag{378}$$

Normalizing gives

$$\boldsymbol{\phi}_1 = \begin{bmatrix} +\dfrac{1}{\sqrt{2}} \\ \\ +\dfrac{1}{\sqrt{2}} \end{bmatrix}. \tag{379}$$

Similarly,

$$\boldsymbol{\phi}_2 = \begin{bmatrix} +\dfrac{1}{\sqrt{2}} \\ \\ -\dfrac{1}{\sqrt{2}} \end{bmatrix}. \tag{380}$$

The old and new axes are shown in Fig. 2.34. The transformation is

$$\mathbf{W} = \left[\begin{array}{c:c} +\dfrac{1}{\sqrt{2}} & +\dfrac{1}{\sqrt{2}} \\ \hdashline +\dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{array}\right] \tag{381}$$

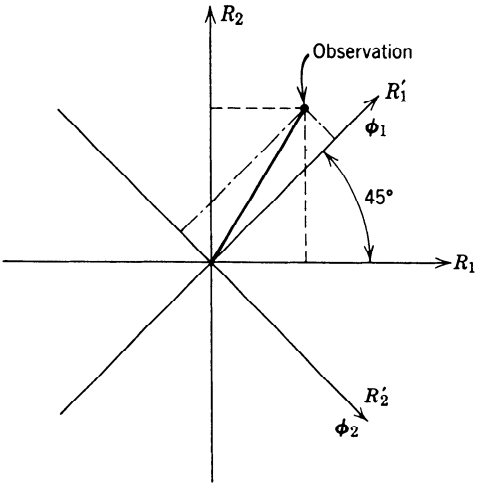


**Fig. 2.34 Rotation of axes.**

$$
\begin{aligned}
R_1' &= \frac{R_1 + R_2}{\sqrt{2}}, \\
R_2' &= \frac{R_1 - R_2}{\sqrt{2}}, \\
m_{11}' &= \frac{m_{11} + m_{12}}{\sqrt{2}}, \\
m_{12}' &= \frac{m_{11} - m_{12}}{\sqrt{2}}.
\end{aligned} \tag{382}
$$

The sufficient statistic is obtained by using (382) in (370),

$$
l(\mathbf{R}') = \frac{1}{1+\rho}\frac{(R_1 + R_2)(m_{11} + m_{12})}{2} + \frac{1}{1-\rho}\frac{(R_1 - R_2)(m_{11} - m_{12})}{2} \tag{383}
$$

and $d^2$ is

$$
d^2 = \frac{(m_{11} + m_{12})^2}{2(1+\rho)} + \frac{(m_{11} - m_{12})^2}{2(1-\rho)} = \frac{(m_{11}')^2}{(1+\rho)} + \frac{(m_{12}')^2}{(1-\rho)}. \tag{384}
$$

To illustrate a typical application in which the transformation is important we consider a simple optimization problem. The length of the mean vector is constrained,

$$
|\mathbf{m}_1|^2 = 1. \tag{385}
$$

We want to choose $m_{11}$ and $m_{12}$ to maximize $d^2$. Because our transformation is a rotation, it preserves lengths

$$
|\mathbf{m}_1'|^2 = 1. \tag{386}
$$

Looking at (384), we obtain the solution by inspection:

If $\rho > 0$, choose $m_{11}' = 0$ and $m_{12}' = 1$.

If $\rho < 0$, choose $m_{11}' = 1$ and $m_{12}' = 0$.

If $\rho = 0$, all vectors satisfying (385) give the same $d^2$.

We see that this corresponds to choosing the mean-value vector to be equal to the eigenvector with the smallest eigenvalue. This result can be easily extended to $N$ dimensions.

The result in this example is characteristic of a wide class of optimization problems in which the solution corresponds to an eigenvector (or the waveform analog to it).

In this section, we have demonstrated that when the covariance matrices on the two hypotheses are equal the sufficient statistic $l(\mathbf{R})$ is a Gaussian random variable obtained by a linear transformation of $\mathbf{R}$. The performance for any threshold setting is determined by using the value of $d^2$ given by (339) on the ROC in Fig. 2.9. Because the performance improves monotonically with increasing $d^2$, we can use any freedom in the parameters to maximize $d^2$ without considering the ROC explicitly.

### 2.6.2 Equal Mean Vectors

In the second special case of interest the mean-value vectors on the two hypotheses are equal. In other words,

$$\mathbf{m}_1 = \mathbf{m}_0 \triangleq \mathbf{m}. \tag{387}$$

Substituting (387) into (327), we have

$$\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}^T)(\mathbf{Q}_0 - \mathbf{Q}_1)(\mathbf{R} - \mathbf{m}) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta + \tfrac{1}{2} \ln \frac{|\mathbf{K}_1|}{|\mathbf{K}_0|} = \gamma^*. \tag{388}$$

Because the mean-value vectors contain no information that will tell us which hypothesis is true, the likelihood test subtracts them from the received vector. Therefore, without loss of generality, we may assume that $\mathbf{m} = \mathbf{0}$.

We denote the difference of the inverse matrices as $\Delta\mathbf{Q}$:

$$\Delta\mathbf{Q} \triangleq \mathbf{Q}_0 - \mathbf{Q}_1. \tag{389}$$

The likelihood ratio test may be written as

$$\boxed{l(\mathbf{R}) \triangleq \mathbf{R}^T \Delta\mathbf{Q}\mathbf{R} \underset{H_0}{\overset{H_1}{\gtrless}} 2\gamma^* \triangleq \gamma'.} \tag{390}$$

Note that $l(\mathbf{R})$ is the dot product of two Gaussian vectors, $\mathbf{R}^T$ and $\Delta\mathbf{Q}\mathbf{R}$. Thus, $l(\mathbf{R})$ is not a Gaussian random variable.

We now consider the behavior of this test for some interesting special cases.

**Case 1. Diagonal Covariance Matrix on $H_0$: Equal Variances.** Here the $R_i$ on $H_0$ are statistically independent variables with equal variances:

$$\mathbf{K}_0 = \sigma_n^2 \mathbf{I}. \tag{391}$$

We shall see later that (391) corresponds to the physical situation in which there is "noise" only on $H_0$. The following notation is convenient:

$$r_i = n_i, \qquad H_0. \tag{392}$$

On $H_1$ the $r_i$ contain the same variable as on $H_0$, plus additional signal components that may be correlated:

$$\begin{aligned} r_i &= s_i + n_i, \qquad H_1, \\ \mathbf{K}_1 &= \mathbf{K}_s + \sigma_n^2 \mathbf{I}, \end{aligned} \tag{393}$$

where the matrix $\mathbf{K}_s$ represents the covariance matrix of the signal components. Then

$$\mathbf{Q}_0 = \frac{1}{\sigma_n^2} \mathbf{I} \tag{394}$$

and

$$\mathbf{Q}_1 = \frac{1}{\sigma_n^2} \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{K}_s \right)^{-1}. \tag{395}$$

It is convenient to write (395) as

$$\mathbf{Q}_1 = \frac{1}{\sigma_n^2} [\mathbf{I} - \mathbf{H}], \tag{396}$$

which implies

$$\mathbf{H} = (\sigma_n^2 \mathbf{I} + \mathbf{K}_s)^{-1} \mathbf{K}_s = \mathbf{K}_s (\sigma_n^2 \mathbf{I} + \mathbf{K}_s)^{-1} = \sigma_n^2 \mathbf{Q}_0 - \mathbf{Q}_1 = \sigma_n^2 \Delta \mathbf{Q}. \tag{397}$$

The $\mathbf{H}$ matrix has an important interpretation which we shall develop later. We take the first expression in (397) as its definition. Substituting (397) into (389) and the result into (390), we have

$$\boxed{l(\mathbf{R}) = \frac{1}{\sigma_n^2} \mathbf{R}^T \mathbf{H} \mathbf{R} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'.} \tag{398}$$

Several subcases are important.

**Case 1A. Uncorrelated, Identically Distributed Signal Components.** In this case the signal components $s_i$ are independent variables with identical variances:

$$\mathbf{K}_s = \sigma_s^2 \mathbf{I}. \tag{399}$$

Then

$$\mathbf{H} = (\sigma_n^2 \mathbf{I} + \sigma_s^2 \mathbf{I})^{-1} \sigma_s^2 \mathbf{I}, \tag{400}$$

or

$$\mathbf{H} = \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2} \mathbf{I} \tag{401}$$

and

$$l(\mathbf{R}) = \frac{1}{\sigma_n^2} \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2} \mathbf{R}^T\mathbf{R} = \frac{1}{\sigma_n^2} \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2} \sum_{i=1}^{N} R_i^2. \tag{402}$$

The constant can be incorporated in the threshold to give

$$l(\mathbf{R}) \triangleq \sum_{i=1}^{N} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma''. \tag{403}$$

We now calculate the performance of the test. On both hypotheses $l(\mathbf{R})$ is the sum of the squares of $N$ Gaussian variables. The difference in the hypotheses is in the variance of the Gaussian variables. For simplicity, we shall assume that $N$ is an even integer.

To find $p_{l|H_0}(L|H_0)$ we observe that the characteristic function of each $R_i^2$ is

$$\begin{aligned} M_{R_i^2|H_0}(jv) &\triangleq \int_{-\infty}^{\infty} e^{jvR_i^2} \frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-R_i^2/2\sigma_n^2}\, dR_i \\ &= (1 - 2jv\sigma_n^2)^{-1/2}. \end{aligned} \tag{404}$$

Because of the independence of the variables, $M_{l|H_0}(jv)$ can be written as a product. Therefore

$$M_{l|H_0}(jv) = (1 - 2jv\sigma_n^2)^{-N/2}. \tag{405}$$

Taking the inverse transform, we obtain $p_{l|H_0}(L|H_0)$:

$$\begin{aligned} p_{l|H_0}(L|H_0) &= \frac{L^{N/2-1}e^{-L/2\sigma_n^2}}{2^{N/2}\sigma_n^N \Gamma\left(\frac{N}{2}\right)}, \qquad L \geq 0, \\ &= 0, \qquad L < 0, \end{aligned} \tag{406}$$

which is familiar as the $\chi^2$ (chi-square) density function with $N$ degrees of freedom. It is tabulated in several references (e.g., [19] or [3]). For $N = 2$ it is easy to check that it is the simple exponential on p. 41. Similarly,

$$\begin{aligned} p_{l|H_1}(L|H_1) &= \frac{L^{N/2-1}e^{-L/2\sigma_1^2}}{2^{N/2}\sigma_1^N \Gamma\left(\frac{N}{2}\right)}, \qquad L \geq 0, \\ &= 0, \qquad L < 0, \end{aligned} \tag{407}$$

where $\sigma_1^2 \triangleq \sigma_s^2 + \sigma_n^2$.

The expressions for $P_D$ and $P_F$ are,

$$P_D = \int_{\gamma''}^{\infty} [2^{N/2}\sigma_1^N \Gamma(N/2)]^{-1} L^{N/2-1} e^{-L/2\sigma_1^2}\, dL \tag{408}$$

and

$$P_F = \int_{\gamma''}^{\infty} [2^{N/2}\sigma_n^N \Gamma(N/2)]^{-1} L^{N/2-1} e^{-L/2\sigma_n^2}\, dL. \tag{409}$$

Construction of the ROC requires an evaluation of the two integrals. We see that for $N = 2$ we have the same problem as Example 2 on p. 41 and (408) and (409) reduce to

$$\begin{aligned} P_D &= \exp\left(-\frac{\gamma''}{2\sigma_1^2}\right), \\ P_F &= \exp\left(-\frac{\gamma''}{2\sigma_n^2}\right), \end{aligned} \tag{410}$$

and

$$P_F = (P_D)^{(1+\sigma_s^2/\sigma_n^2)}. \tag{411}$$

For the general case there are several methods of proceeding. First, let $M = N/2 - 1$ and $\gamma''' = \gamma''/2\sigma_n^2$. Then write

$$P_F = 1 - \int_0^{\gamma'''} \frac{x^M}{M!} e^{-x}\, dx. \tag{412}$$

The integral, called the incomplete Gamma function, has been tabulated by Pearson [21]:

$$I_\Gamma(u, M) \triangleq \int_0^{u\sqrt{M+1}} \frac{x^M}{M!} e^{-x}\, dx, \tag{413}$$

and

$$P_F = 1 - I_\Gamma\left(\frac{\gamma'''}{\sqrt{M+1}}, M\right). \tag{414}$$

These tables are most useful for $P_F \geq 10^{-6}$ and $M \leq 50$.

In a second approach we integrate by parts $M$ times. The result is

$$P_F = \exp(-\gamma''') \sum_{k=0}^{M} \frac{(\gamma''')^k}{k!}. \tag{415}$$

For *small* $P_F$, $\gamma'''$ is large and we can approximate the series by the last few terms,

$$P_F = \frac{(\gamma''')^M e^{-\gamma'''}}{M!}\left[1 + \frac{M}{\gamma'''} + \frac{M(M-1)}{(\gamma''')^2} + \cdots\right]. \tag{416}$$

Furthermore, we can approximate the bracket as $(1 - M/\gamma''')^{-1}$. This gives

$$P_F \cong \frac{(\gamma''')^M e^{-\gamma'''}}{M!(1 - M/\gamma''')}. \tag{417}$$

A similar expression for $P_D$ follows in which $\gamma'''$ is replaced by $\gamma^{iv} \triangleq \gamma''/2\sigma_1^2$. The approximate expression in (417) is useful for manual calculation. In actual practice, we use (415) and calculate the ROC numerically. In Fig. 2.35*a* we have plotted the receiver operating characteristic for some representative values of $N$ and $\sigma_s^2/\sigma_n^2$.

Two particularly interesting curves are those for $N = 8$, $\sigma_s^2/\sigma_n^2 = 1$ and $N = 2$, $\sigma_s^2/\sigma_n^2 = 4$. In both cases the product $N\sigma_s^2/\sigma_n^2 = 8$. We see that when the desired $P_F$ is greater than 0.3, $P_D$ is higher if the available "signal strength" is divided into more components. This suggests that for each $P_F$

and product $N\sigma_s^2/\sigma_n^2$ there should be an optimum $N$. In Chapter 4 we shall see that this problem corresponds to optimum diversity in communication systems and the optimum energy per pulse in radar. In Figs. 2.35*b* and *c* we have sketched $P_M$ as a function of $N$ for $P_F = 10^{-2}$ and $10^{-4}$, respectively, and various $N\sigma_s^2/\sigma_n^2$ products. We discuss the physical implications of these results in Chapter 4.

**Case 1B. Independent Signal Components: Unequal Variances.** In this case the signal components $s_i$ are independent variables with variances $\sigma_{s_i}^2$:

$$\mathbf{K}_s = \begin{bmatrix} \sigma_{s_1}^2 & & & 0 \\ & \sigma_{s_2}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{s_N}^2 \end{bmatrix}. \tag{418}$$

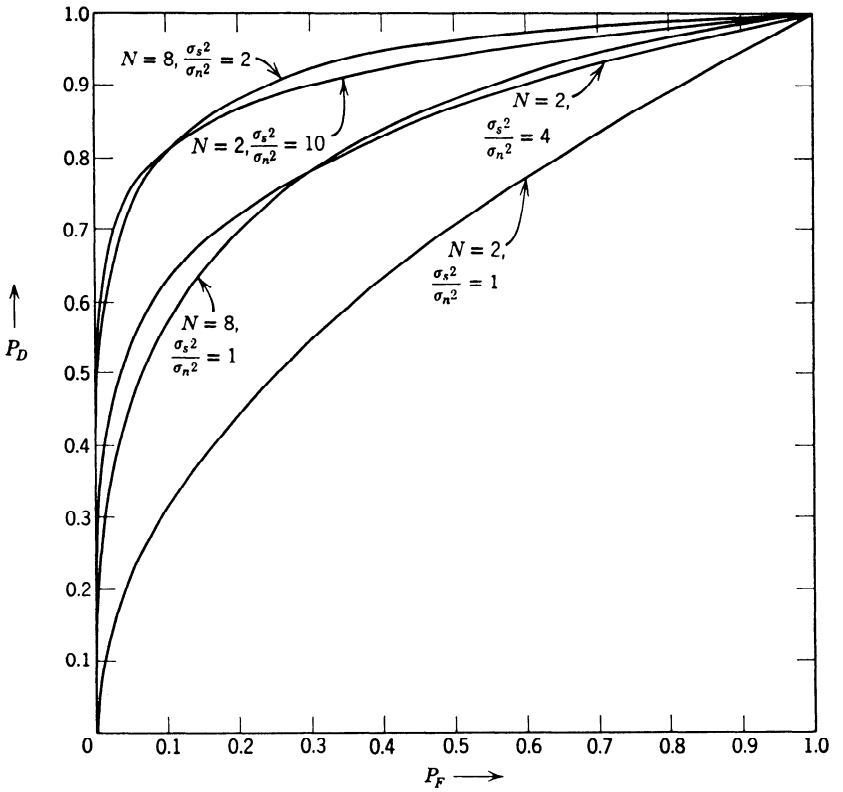


**Fig. 2.35** *a.* **Receiver operating characteristic: Gaussian variables with identical means and unequal variances on the two hypotheses.**

Then

$$\mathbf{H} = \begin{bmatrix} \dfrac{\sigma_{s_1}^2}{\sigma_n^2 + \sigma_{s_1}^2} & & & 0 \\ & \dfrac{\sigma_{s_2}^2}{\sigma_n^2 + \sigma_{s_2}^2} & & \\ & & \ddots & \\ 0 & & & \dfrac{\sigma_{s_N}^2}{\sigma_n^2 + \sigma_{s_N}^2} \end{bmatrix} \tag{419}$$

and

$$l(\mathbf{R}) = \frac{1}{\sigma_n^2} \sum_{i=1}^{N} \frac{\sigma_{s_i}^2}{\sigma_n^2 + \sigma_{s_i}^2} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'. \tag{420}$$

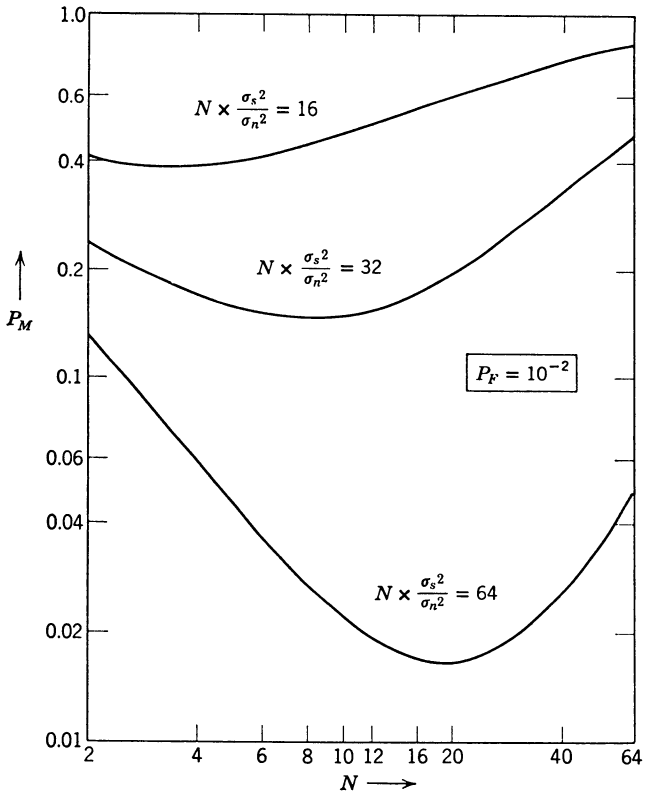


**Fig. 2.35** *b.* $P_M$ **as a function of** $N$ $[P_F = 10^{-2}]$.

The characteristic function of $l(\mathbf{R})$ follows easily, but the calculation of $P_F$ and $P_D$ is difficult. In Section 2.7 we derive approximations to the performance that lead to simpler expressions.

**Case 1C. Arbitrary Signal Components.** This is, of course, the general case. We revisit it merely to point out that it can always be reduced to Case 1B by an orthogonal transformation (see discussion on pp. 102–106).

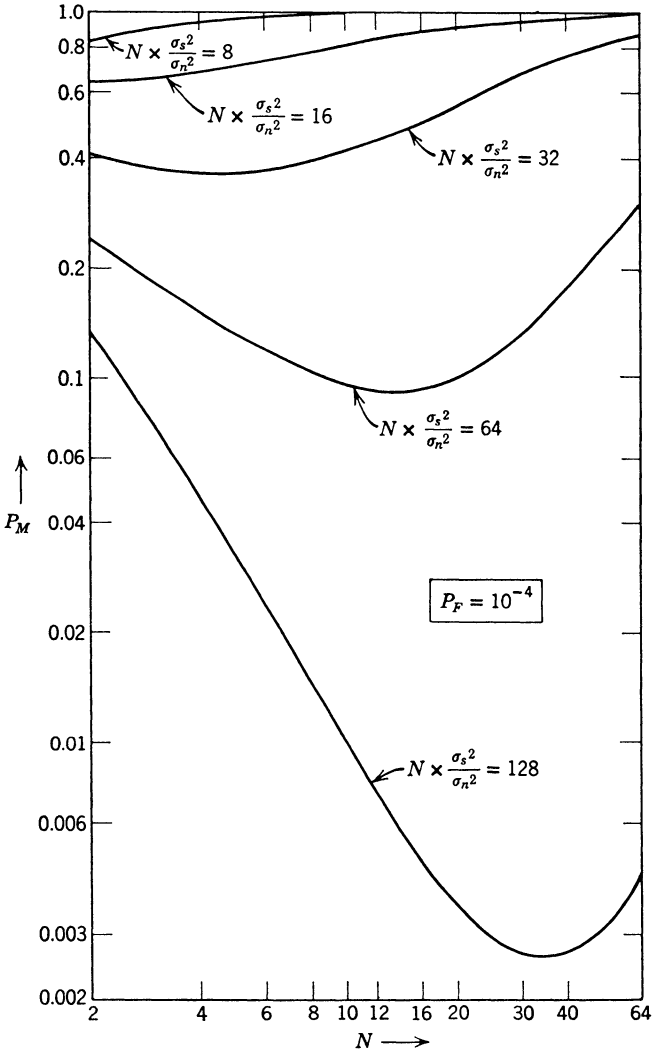


**Fig. 2.35** *c.* $P_M$ **as a function of** $N$ **[**$P_F = 10^{-4}$**].**

**Case 2. Symmetric Hypotheses, Uncorrelated Noise.** Case 1 was unsymmetric because of the noise-only hypothesis. Here we have the following hypotheses:

$$\begin{aligned} H_1: r_i &= s_i + n_i \qquad i = 1, \ldots, N \\ &\quad\ n_i \qquad i = N+1, \ldots, 2N, \\ H_0: r_i &= n_i \qquad i = 1, \ldots, N \\ &\quad\ s_i + n_i \qquad i = N+1, \ldots, 2N, \end{aligned} \tag{421}$$

where the $n_i$ are independent variables with variance $\sigma_n^2$ and the $s_i$ have a covariance matrix $\mathbf{K}_s$. Then

$$\mathbf{K}_1 = \begin{bmatrix} \sigma_n^2\mathbf{I} + \mathbf{K}_s & \mathbf{0} \\ \mathbf{0} & \sigma_n^2\mathbf{I} \end{bmatrix} \tag{422}$$

and

$$\mathbf{K}_0 = \begin{bmatrix} \sigma_n^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_n^2\mathbf{I} + \mathbf{K}_s \end{bmatrix}, \tag{423}$$

where we have partitioned the $2N \times 2N$ matrices into $N \times N$ submatrices. Then

$$\Delta\mathbf{Q} = \begin{bmatrix} \frac{1}{\sigma_n^2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & (\sigma_n^2\mathbf{I} + \mathbf{K}_s)^{-1} \end{bmatrix} - \begin{bmatrix} (\sigma_n^2\mathbf{I} + \mathbf{K}_s)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_n^2}\mathbf{I} \end{bmatrix}. \tag{424}$$

Using (397), we have

$$\Delta\mathbf{Q} = \frac{1}{\sigma_n^2}\begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & -\mathbf{H} \end{bmatrix}, \tag{425}$$

where, as previously defined in (397), $\mathbf{H}$ is

$$\mathbf{H} \triangleq (\sigma_n^2\mathbf{I} + \mathbf{K}_s)^{-1}\mathbf{K}_s. \tag{426}$$

If we partition $\mathbf{R}$ into two $N \times 1$ matrices,

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix}, \tag{427}$$

then

$$l(\mathbf{R}) = \frac{1}{\sigma_n^2}\mathbf{R}_1^T\mathbf{H}\mathbf{R}_1 - \mathbf{R}_2^T\mathbf{H}\mathbf{R}_2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'. \tag{428}$$

The special cases analogous to 1A and 1B follow easily.

**Case 2A. Uncorrelated, Identically Distributed Signal Components.** Let

$$\mathbf{K}_s = \sigma_s^2 \mathbf{I}; \tag{429}$$

then

$$l(\mathbf{R}) = \sum_{i=1}^{N} R_i^2 - \sum_{i=N+1}^{2N} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma^v. \tag{430}$$

If the hypotheses are equally likely and the criterion is minimum $\Pr(\epsilon)$, the threshold $\eta$ in the LRT is unity (see 69). From (388) and (390) we see that this will result in $\gamma^v = 0$. This case occurs frequently and leads to a simple error calculation. The test then becomes

$$l_1(\mathbf{R}) \triangleq \sum_{i=1}^{N} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \sum_{i=N+1}^{2N} R_i^2 \triangleq l_0(\mathbf{R}). \tag{431}$$

The probability of error given that $H_1$ is true is the probability that $l_0(\mathbf{R})$ is greater than $l_1(\mathbf{R})$. Because the test is symmetric with respect to the two hypotheses,

$$\Pr(\epsilon) = \tfrac{1}{2} \Pr(\epsilon|H_1) + \tfrac{1}{2} \Pr(\epsilon|H_0) = \Pr(\epsilon|H_1). \tag{432a}$$

Thus

$$\Pr(\epsilon) = \int_0^\infty dL_1 p_{l_1|H_1}(L_1|H_1) \int_{L_1}^\infty p_{l_0|H_1}(L_0|H_1)\, dL_0. \tag{432b}$$

Substituting (406) and (407) in (432*b*), recalling that $N$ is even, and evaluating the inner integral, we have

$$\Pr(\epsilon) = \int_0^\infty \frac{1}{2^{N/2}\sigma_1{}^N \Gamma(N/2)} L_1^{N/2-1} e^{-L_1/2\sigma_1{}^2} \times \left[ e^{-L_1/2\sigma_n{}^2} \sum_{k=0}^{N/2-1} \frac{(L_1/2\sigma_n{}^2)^k}{k!} \right] dL_1. \tag{432c}$$

Defining

$$\alpha = \frac{\sigma_n{}^2}{\sigma_1{}^2 + \sigma_n{}^2} = \frac{\sigma_n{}^2}{\sigma_s{}^2 + 2\sigma_n{}^2}, \tag{433}$$

and integrating, (432*c*) reduces to

$$\Pr(\epsilon) = \alpha^{N/2} \sum_{j=0}^{N/2-1} \binom{\frac{N}{2} + j - 1}{j} (1 - \alpha)^j. \tag{434}$$

This result is due to Pierce [22]. It is a closed-form expression but it is tedious to use. We delay plotting (434) until Section 2.7, in which we derive an approximate expression for comparison.

**Case 2B. Uncorrelated Signal Components: Unequal Variances.** Now,

$$\mathbf{K}_s = \begin{bmatrix} \sigma_{s_1}{}^2 & & & 0 \\ & \sigma_{s_2}{}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{s_N}{}^2 \end{bmatrix}. \tag{435}$$

It follows easily that

$$l(\mathbf{R}) = \frac{1}{\sigma_n^2}\left[\sum_{i=1}^{N} \frac{\sigma_{s_i}^2}{\sigma_n^2 + \sigma_{s_i}^2} R_i^2 - \sum_{i=N+1}^{2N} \frac{\sigma_{s_i-N}^2}{\sigma_n^2 + \sigma_{s_i-N}^2} R_i^2\right] \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'. \quad (436)$$

As in Case 1B, the performance is difficult to evaluate. The approximations developed in Section 2.7 are also useful for this case.

### 2.6.3 Summary

We have discussed in detail the general Gaussian problem and have found that the sufficient statistic was the difference between two quadratic forms:

$$l(\mathbf{R}) = \tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_0^T)\mathbf{Q}_0(\mathbf{R} - \mathbf{m}_0) - \tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_1^T)\mathbf{Q}_1(\mathbf{R} - \mathbf{m}_1). \quad (437)$$

A particularly simple special case was the one in which the covariance matrices on the two hypotheses were equal. Then

$$l(\mathbf{R}) = \tfrac{1}{2}\,\Delta\mathbf{m}^T\mathbf{Q}\mathbf{R}, \quad (438)$$

and the performance was completely characterized by the quantity $d^2$:

$$d^2 = \Delta\mathbf{m}^T\mathbf{Q}\,\Delta\mathbf{m}. \quad (439)$$

When the covariance matrices are unequal, the implementation of the likelihood ratio test is still straightforward but the performance calculations are difficult (remember that $d^2$ is no longer applicable because $l(\mathbf{R})$ is not Gaussian). In the simplest case of diagonal covariance matrices with equal elements, exact error expressions were developed. In the general case, exact expressions are possible but are too unwieldy to be useful. This inability to obtain tractable performance expressions is the motivation for discussion of performance bounds and approximations in the next section.

Before leaving the general Gaussian problem, we should point out that similar results can be obtained for the $M$-hypothesis case and for the estimation problem. Some of these results are developed in the problems.

## 2.7 PERFORMANCE BOUNDS AND APPROXIMATIONS

Up to this point we have dealt primarily with problems in which we could derive the structure of the optimum receiver and obtain relatively simple expressions for the receiver operating characteristic or the error probability.

In many cases of interest the optimum test can be derived but an exact performance calculation is impossible. For these cases we must resort to