where μ_0 is a fixed number. In this case $\Lambda_0 = \{\mu_0\}$ is simple and $\Lambda_1 = (\mu_0, \infty)$ is composite.

From the example in the last Section, it is known that for each $\theta \in \Lambda_1$, the most power α-level test of H_0 vs. $Y \sim \mathcal{N}(\theta, \sigma^2)$ has critical region $\Gamma_\theta = \{y \in \Gamma \mid y > \sigma \Phi^{-1}(1-\alpha) + \mu_0\}$

(II.E.22)

One can see that this region Γ_θ does not depend on $\theta \in \Lambda_1$. This gives a UMP test that is denoted by $\tilde{\delta}_1$ and

$$P_D(\tilde{\delta}_1; \theta) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\theta - \mu_0}{\sigma}\right).$$

* However, for the same family of distributions as above, but suppose that we consider the hypothesis pair

$$H_0: \ \theta = \mu_0$$

vs.

$$H_1: \ \theta \neq \mu_0$$

In this case $\Lambda_1 = (-\infty, \mu_0) \cup (\mu_0, \infty)$.

For $\theta > \mu_0$, the most power critical region is given above in (II.E.22)

But for $\theta < \mu_0$, it is not hard to have that the most powerful α-level test has critical region

$$\Gamma_\theta = \{ y \in \Gamma \mid y < \sigma \, \Phi^{-1}(\alpha) + \mu_0 \}, \quad (\text{II}.E.24)$$

which is, although, als independent of θ, but much different from the previous one for $\theta > \mu_0$ in (II.E.22).

$\Rightarrow$ The UMP test doesnot exist.

If we denote $\tilde{\delta}_2$ as the test with critical region in (II.E.24), then we have

$$P_D(\tilde{\delta}_2;\theta) = \Phi\left(\Phi^{-1}(\alpha) - \frac{\theta-\mu_0}{\sigma}\right)$$

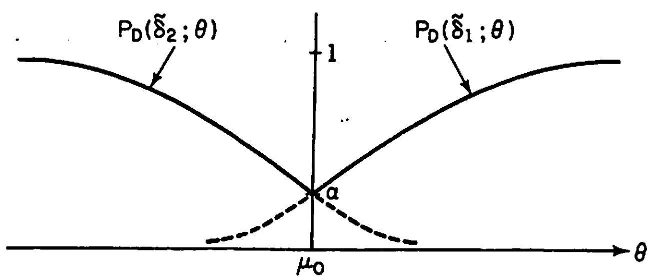


FIGURE II.E.2. Power curves for test of $\theta = \mu_0$ versus $\theta > \mu_0$ and $\theta = \mu_0$ versus $\theta < \mu_0$, for location testing with Gaussian error.

Remark: A more reasonable test for the above hypothesis pair than $\tilde{\delta}_1$ or $\tilde{\delta}_2$ is one that compares $|y - \mu_0|$ to a threshold.

* The above example shows that UMP criterion is too strong. One can relax it by using other criteria. One such a condition is unbiasedness.
$P_D(\tilde{\delta};\theta) \geq \alpha \quad \forall \theta \in \Lambda_1$ in addition to the constraint $P_F(\tilde{\delta};\theta) \leq \alpha \quad \forall \theta \in \Lambda_0$.

* Locally Most Powerful (LMP) Test

In the example of parameter set $\Lambda = [\theta_0, \infty)$ with $\Lambda_0 = \{\theta_0\}$ and $\Lambda_1 = (\theta_0, \infty)$ for the test

$$H_0 : \theta = \theta_0$$
vs.
$$H_1 : \theta > \theta_0 ,$$

One may be more interested in the region when θ is near θ_0, where θ is, for example, a signal amplitude.

In this case, assume $P_D(\tilde{\delta};\theta)$ is smooth enough so that it has Taylor series:

$$P_D(\tilde{\delta};\theta) = P_D(\tilde{\delta};\theta_0) + (\theta-\theta_0) P_D'(\tilde{\delta};\theta_0) + O((\theta-\theta_0)^2)$$

where $P_D'(\tilde{\delta},\theta) = \dfrac{\partial P_D(\tilde{\delta};\theta)}{\partial\theta}$ (note $P_D(\tilde{\delta};\theta_0) = P_F(\tilde{\delta};\theta_0)$)

$\Rightarrow$ For all size α-level tests, when θ is near θ_0, $P_D(\tilde{\delta};\theta)$ is

$$P_D(\tilde{\delta};\theta) \cong \alpha + (\theta-\theta_0)\, P_D'(\tilde{\delta};\theta_0).$$

$\Rightarrow$ For θ near θ_0 we can achieve approximate maximum power with size α by choosing $\tilde{\delta}$ to maximize $P_D'(\tilde{\delta};\theta_0)$

A test that maximizes $P_D'(\tilde{\delta};\theta_0)$ subject to false-alarm constraint $P_F(\tilde{\delta}) \le \alpha$ is called an α-level locally most powerful (LMP) test or a locally optimum test.

Note that $P_D'(\tilde{\delta};\theta_0)$ has only one paramet[er] θ_0.

Let us see the general structure of LMP tests:

Assume P_θ has density p_θ for each $\theta \in \Lambda_1$.
Then

$$P_D(\tilde{\delta};\theta) = E_\theta\{\tilde{\delta}(Y)\} = \int_\Gamma \tilde{\delta}(y)\, p_\theta(y)\, \mu(dy).$$

Assume $p_\theta(y)$ has sufficient regularity (smoothness) in terms of θ such that we may interchange the order of integration and differentiation:

$$P_D'(\tilde{\delta};\theta) = \int_\Gamma \tilde{\delta}(y) \frac{\partial}{\partial\theta} p_\theta(y)\Big|_{\theta=\theta_0} \mu(dy).$$

Compared this with the Neyman-Pearson test, we find that the α-level LMP test is similar to the Neyman-Pearson test if we replace $p_1(y)$ with

$$\left.\frac{\partial p_\theta(y)}{\partial\theta}\right|_{\theta=\theta_0}$$

$\Rightarrow$ An α-level LMP test is given

$$\tilde{\delta}_{lo}(y) = \begin{cases} 1 & \text{if } \left.\frac{\partial}{\partial\theta} p_\theta(y)\right|_{\theta=\theta_0} > \eta\, p_{\theta_0}(y), \\ \gamma & \quad \cdot\cdot \quad = \quad \cdot\cdot \\ 0 & \quad \cdot\cdot \quad < \quad \cdot\cdot \end{cases}$$

where η and γ are chosen so that $P_F(\tilde{\delta}_{lo}) = \alpha$.

* Generalized Likelihood-Ratio Test (GLRT)

When there is no any optimality, a test is often used is to compare

$$\frac{\max_{\theta\in\Lambda_1} p_\theta(y)}{\max_{\theta\in\Lambda_0} p_\theta(y)}$$

to a threshold. This test is called generalized likelihood-ratio test (GLRT), or a maximum-likelihood test. It is quite useful in signal processing.

Chapter 3 Signal Detection in Discrete Time

In the last chapter, what we talked about was when we have only one observation y at one time. In this chapter, we will talk about the case when we have a sequence of observations at a sequence of time or space.

§1 Models and Detector Structures

* Models: an observed continuous-time waveform that consists of one or two possible signals corrupted by additive noise.

Objective: to decide which of the possible signals is present by processing a finite number, say n, of samples taken from the observed waveform.

* Hypothesis pair for the observation space $(\Gamma, \mathcal{G}) = (\mathbb{R}^n, \mathcal{B}^n)$:

$$H_0 : Y_k = N_k + S_{0k}, \quad k = 1, 2, \cdots, n$$

vs. $$H_1 : Y_k = N_k + S_{1k}, \quad k = 1, 2, \cdots, n,$$

where $\underline{Y} = (Y_1, \cdots, Y_n)^T$ is an observation vector consisting of the samples from the observed waveform, $\underline{N} = (N_1, \cdots, N_n)^T$ is a vector of noise samples, and $\underline{S}_0 = (S_{01}, \cdots, S_{0n})^T$ and $\underline{S}_1 = (S_{11}, \cdots, S_{1n})^T$

are vectors of signals from the two possible signals

* Three basic types for signals $\underline{s}_0$ and $\underline{s}_1$:
  1) They are completely known (deterministic);
  2) They can be known except for a set of unknown (possibly random) parameters;
  3) They can be completely random and thus specified only by their probability distributions

* Sometimes (e.g. radar/sonar applications) one of the signals, usually $\underline{s}_0$, is identically zero, and in this case, we try to detect a signal embedded in noise.

* Assume noise is independent of signals and its distribution does not depend on which hypothesis is true. This assumption holds for most applications but a few in radar/sonar applications where the noise may be partially composed of spurious signal reflections from the ground or from objects other than the intended targets. Such noises are, for example, speckles in radar.

* Assume the noise distribution is determined by a density P_N in $\mathbb{R}^n$.

* Given $\underline{S}_j = s_j \in \mathbb{R}^n$, the observation $\underline{Y}$ has conditional density (under H_j):

$$p_{\underline{N}}(\underline{y} - \underline{s}_j), \quad \underline{y} \in \mathbb{R}^n.$$

$\Rightarrow$ The density of $\underline{Y}$ under H_j is

$$p_j(\underline{y}) = E\{p_{\underline{N}}(\underline{y} - \underline{S}_j)\}, \quad \underline{y} \in \mathbb{R}^n,$$

where the expectation is with respect to the signal $\underline{S}_j$.

The likelihood ratio is

$$L(\underline{y}) = \frac{E\{p_{\underline{N}}(\underline{y} - \underline{S}_1)\}}{E\{p_{\underline{N}}(\underline{y} - \underline{S}_0)\}}, \quad \underline{y} \in \mathbb{R}^n.$$

Case 1: Detection of Deterministic Signals in Independent Noise

* $\underline{S}_0$ and $\underline{S}_1$ are completely deterministic : $\underline{S}_j = s_j \in \mathbb{R}^n$ known. This is sometimes known as the <u>coherent detection problem</u>.

* $$L(\underline{y}) = \frac{p_{\underline{N}}(\underline{y} - \underline{s}_1)}{p_{\underline{N}}(\underline{y} - \underline{s}_0)}$$

$$= \frac{p_{\underline{N}}(y_1 - s_{11}, \ldots, y_n - s_{1n})}{p_{\underline{N}}(y_1 - s_{01}, \ldots, y_n - s_{0n})} \qquad (\text{III}.B.5)$$

In this case, the optimization detector can be determined if $p_{\underline{N}}$ is known.

* When n, the number of observation, is large, the detection problem (II. B.5) could be complicated where n-fold integrals $\int_{\{r(\underline{y})>\tau\}} p_j(\underline{y})\,\mu(d\underline{y})$ may be needed, for example.

* A simplified detection for (II. B.5) is when the noise samples $N_1, \cdots, N_n$ are statistically independent. In this case,

$$p_{\underline{N}}(\underline{y}) = \prod_{k=1}^{n} p_{N_k}(y_k),$$

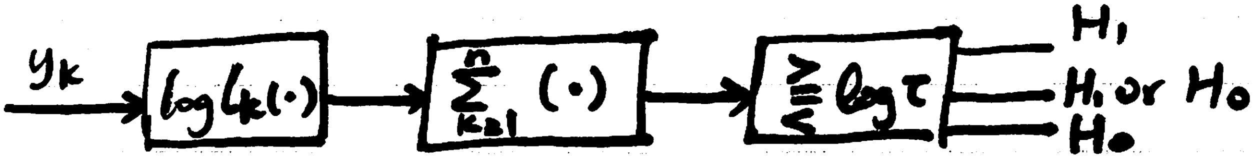
where p_{N_k} is the marginal density of N_k.

$$\Rightarrow L(\underline{y}) = \prod_{k=1}^{n} L_k(y_k)$$

with $$L_k(y_k) = \frac{p_{N_k}(y_k - s_{1k})}{p_{N_k}(y_k - s_{0k})}$$

* Since $\log(x)$ is a strictly increasing function of x, comparison of $L(\underline{y})$ to a threshold τ is equivalent to comparison of $\log L(\underline{y})$ to the threshold $\log(\tau)$. Thus the optimum tests are

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^{n} \log L_k(y_k) > \log\tau \\ \gamma & \quad\cdots\quad = \cdots \\ 0 & \quad\cdots\quad < \cdots \end{cases}$$



Detector Structure for Coherent Signals in Independent Noise

* Example 1: Coherent Detection in i.i.d. Gaussian Noise

* $N_1, \cdots, N_n$ are i.i.d. with marginal distribution $\mathcal{N}(0, \sigma^2)$. This noise model is common in some applications, such as in communications systems for thermal noise generated by the motion of the electrons in the receiver electronics

* Without loss of generality, let $\underline{s}_0 = \underline{0}$ and $\underline{s}_1 = \underline{s}$ (Otherwise we may let $\underline{y} = \underline{y} - \underline{s}_0$, then $\underline{s} = \underline{s}_1 - \underline{s}_0$)

$\Rightarrow$ From the previous chapter study (II.B.27),

$$L_k(y_k) = s_k(y_k - s_k/2)/\sigma^2,$$

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1, & \text{if } \sum_{k=1}^{n} s_k\left(y_k - \frac{s_k}{2}\right) > \tau' \\ \gamma, & \qquad \cdots \qquad = \cdots \\ 0 & \qquad \cdots \qquad < \cdots \end{cases}$$

where $\tau' = \sigma^2 \log \tau$.

* Also $-\frac{1}{2}\sum_{k=1}^{n} s_k^2$ can be absorbed into the threshold :

$$\tau'' = \tau' + \frac{1}{2}\sum_{k=1}^{n} s_k^2$$

$$\Rightarrow \tilde{\delta}_0(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^{n} s_k y_k > \tau'' \\ \gamma & \text{if } \sum_{k=1}^{n} s_k y_k = \tau'' \\ 0 & \text{if } \sum_{k=1}^{n} s_k y_k < \tau'' \end{cases}$$

This structure is called correlation detector or simply correlator.

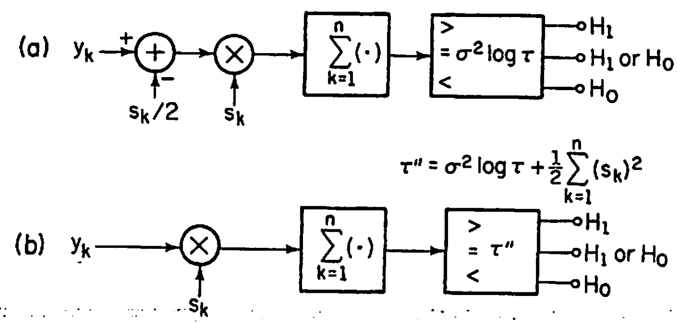


FIGURE III.B.2. Optimum detector for coherent signals i.i.d. Gaussian noise.

* An important feature of the above optimum for Gaussian noise is that it is a linear system output compared to a threshold : write

$$\sum_{k=1}^{n} s_k y_k = \sum_{k=-\infty}^{n} h_{n-k}\, y_k ,$$

where $h_k = \begin{cases} s_{n-k}, & \text{for } 0 \le k \le n-1, \\ 0, & \text{otherwise} \end{cases}$

$\Rightarrow$ The detector can be viewed as a system that inputs the observation sequence $y_1, \ldots, y_n$ to a digital linear filter and then samples the output at time n for comparison to a threshold.

This structure is known as a matched filter.

* *Example 2*: Coherent Detection in i.i.d. Laplacian Noise

$N_1, \ldots, N_n$ are i.i.d. but with the Laplacian marginal probability density:

$$p_{N_k}(y_k) = \frac{\alpha}{2} e^{-\alpha |y_k|}, \quad y_k \in \mathbb{R},$$

where $\alpha > 0$ is a scale parameter of the density.

This model is sometimes used to represent the behavior of <u>impulse noises</u> in communications receivers.

In comparison with Gaussian model, it has longer "tails" representing higher probabilities of large observations.

$$\log L_k(y_k) = \alpha \left(|y_k| - |y_k - s_k| \right)$$

$$= \begin{cases} -\alpha |s_k| & \text{if } \operatorname{sgn}(s_k)\, y_k \le 0 \\ \alpha |2y_k - s_k| & \text{if } 0 < \operatorname{sgn}(s_k)\, y_k < |s_k| \\ +\alpha |s_k| & \text{if } \operatorname{sgn}(s_k)\, y_k \ge |s_k| \end{cases}$$

where sgn denotes the signum function

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

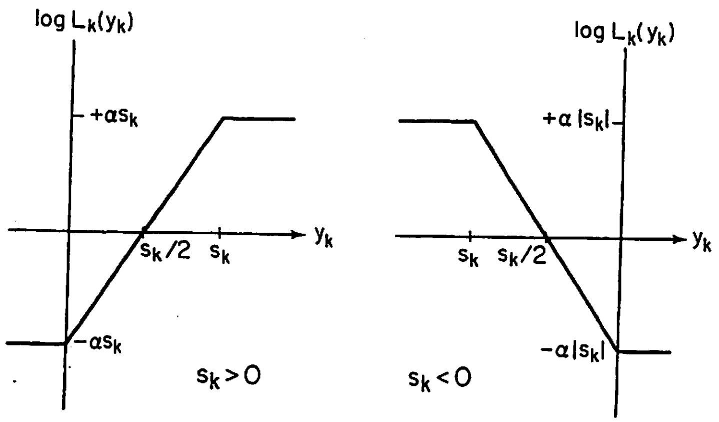


FIGURE III.B.3. Per-sample log-likelihood ratio for coherent detection in Laplacian noise.

$$\Rightarrow \tilde{\delta}_0(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^{n} \mathrm{sgn}(s_k)\, l_k\left(y_k - \frac{s_k}{2}\right) > \tau', \\ \gamma & \quad\quad \vdots \quad\quad\quad\quad\quad\quad\quad\quad = \ \vdots \\ 0 & \quad\quad \vdots \quad\quad\quad\quad\quad\quad\quad\quad < \ \vdots \end{cases}$$

where $\tau' = \log \tau / (2\alpha)$

$$l_k(x) = \begin{cases} -\frac{|s_k|}{2} & \text{if } x < -\frac{|s_k|}{2} \\ x & \text{if } -\frac{|s_k|}{2} < x < \frac{|s_k|}{2} \\ \frac{|s_k|}{2} & \text{if } x > \frac{|s_k|}{2} \end{cases}$$

Note that $l_k(x)$ is linear in x if $x \in \left(\frac{-|s_k|}{2}, \frac{|s_k|}{2}\right)$ and otherwise $l_k(x) = \frac{\mathrm{sgn}(x)}{2} |s_k|$.

It is known as a <u>soft limiter</u> or <u>amplifier limiter</u>.

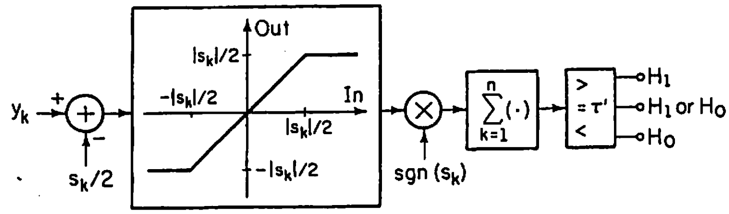


FIGURE III.B.4. Optimum detector for coherent signals in Laplacian noise.

* The above two detectors for Gaussian and Laplacian noises center the observations by subtracting $s_k/2$ from each y_k. In Gaussian case, it then correlates the centered data with the known signal and then compare the output of this correlation with a threshold. In Laplacian case, it soft-limits the centered data and then correlates these soft-limited observations with the sequence of signal signs. The effect of this soft limiting is to reduce the effect of large observations on the sum, thus making this system more tolerant to large values.

* Example 3 : Locally Optimum Detection of Coherent Signals in i.i.d. Noise

In many detection problems, the form of the received signal is known but not its amplitude.

Consider the following composite hypothesis testing problem:

$$H_0 : Y_k = N_k, \quad k = 1, 2, \cdots, n$$

vs. $$H_1 : Y_k = N_k + \theta S_k, \quad k = 1, 2, \cdots, n, \theta > 0,$$

where $\underline{S} = (S_1, \cdots, S_n)^T$ is a known signal, $\underline{N} = (N_1, \cdots, N_n)^T$ is a continuous random noise vector with i.i.d. components and marginal probability density function P_{N_k}, θ is a signal strength parameter.

Given θ, the likelihood ratio between H_0 and H_1 is

$$L_\theta(\underline{y}) = \prod_{k=1}^{n} \frac{P_{N_k}(y_k - \theta S_k)}{P_{N_k}(y_k)}.$$

The critical region $\Gamma_\theta = \{ \underline{y} \in \mathbb{R}^n \mid L_\theta(\underline{y}) > \tau \}$ generally depends on θ except for some special cases (the Gaussian noise case in Example 2). UMP tests exist only for particular noise models. But, LMP tests have simple structure as follows.

Assume the sufficient regularity on P_{N_k}. The locally optimum test for H_0 vs. H_1 is

$$\tilde{\delta}_{lo}(\underline{y}) = \begin{cases} 1 & \text{if } \frac{\partial}{\partial\theta} L_\theta(y)\Big|_{\theta=0} > \tau \\ \gamma & \quad \cdot\cdot \quad = \quad \cdot\cdot \\ 0 & \quad \cdot\cdot \quad < \quad \cdot\cdot \end{cases}$$

$$\frac{\partial}{\partial\theta} L_\theta(\underline{y})\Big|_{\theta=0} = \sum_{k=1}^{n} s_k \, g_{lo}(y_k)$$

where $g_{lo}(x) \triangleq - p_{N_1}'(x)/p_{N_1}(x)$

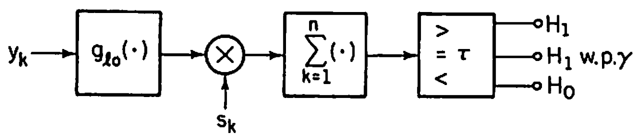


FIGURE III.B.5. Locally optimum detector structure for coherent signals in i.i.d. noise.

The detector consists of the memoryless nonlinearity g_{lo} followed by a correlator, a combination known as a nonlinear correlator.

* Like the likelihood ratio, the locally optimum nonlinearity g_{lo} shapes the observations to reduce the detrimental effects of the noise as much as possible.

* In the case of Gaussian noise $\mathcal{N}(0, \sigma^2)$,

$$g_{lo} = x/\sigma^2$$

and the detector is the correlation detector. This must be the case since the UMP exists in this case and it is the same as the LMP.

(*) In the case of Laplacian noise,

$$g_{lo}(x) = \alpha \, \mathrm{sgn}(x),$$

then the locally optimum detector correlates the signal with the sequence of signs of the observations. In this case, the function $g_{lo}(x)$ is known as a hard limiter

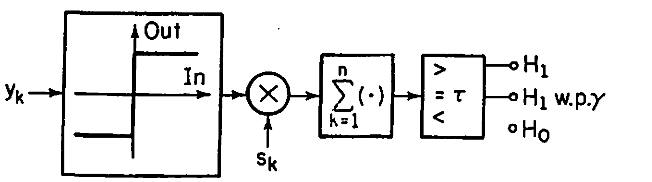


FIGURE III.B.6. Locally optimum detector for Laplacian noise.

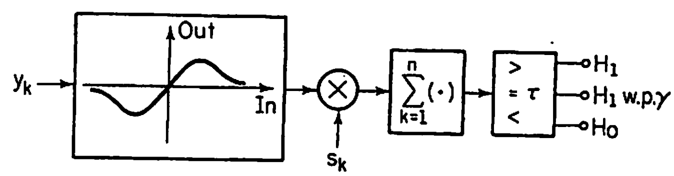


FIGURE III.B.7. Locally optimum detector for Cauchy noise.

(*) An even heavier-tailed noise model than the Laplacian is Cauchy noise model with Cauchy distribution:

$$p_{N_1}(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

$$\Rightarrow g_{lo}(x) = \frac{2x}{1+x^2}$$

is approximately linear near $x=0$ but decays to zero when x is large.

$\Rightarrow$ This detector ignores observations with large magnitudes.

An approximation to this nonlinearity is the so-called noise blanker

$$g(x) = \begin{cases} x & \text{if } |x| \le K \\ 0 & \text{if } |x| > K \end{cases}$$

for a constant $K > 0$.

This is used to combat extremely impulsive noise.

* Case 2: Detection of Deterministic Signals in Gaussian Noise

The noise samples $N_1, \dots, N_n$ are not i.i.d. but multivariate Gaussian with mean $\underline{0}$ and covariance matrix Σ_N

$Y \sim N(\underline{s}_j, \Sigma_N)$ under H_j for $j = 0$ and 1.

The likelihood ratio is

$$L(\underline{y}) = \frac{p_1(\underline{y})}{p_0(\underline{y})}$$

$$= \frac{\frac{1}{(2\pi)^{n/2}|\Sigma_N|^{1/2}} \exp\{-\frac{1}{2}(\underline{y}-\underline{s}_1)^T \Sigma_N^{-1}(\underline{y}-\underline{s}_1)\}}{\frac{1}{(2\pi)^{n/2}|\Sigma_N|^{1/2}} \exp\{-\frac{1}{2}(\underline{y}-\underline{s}_0)^T \Sigma_N^{-1}(\underline{y}-\underline{s}_0)\}}$$

$$= \exp\{\underline{s}_1^T \Sigma_N^{-1}\underline{y} - \underline{s}_0^T \Sigma_N^{-1}\underline{y} - \tfrac{1}{2}\underline{s}_1^T \Sigma_N^{-1}\underline{s}_1 + \tfrac{1}{2}\underline{s}_0^T \Sigma_N^{-1}\underline{s}_0\}$$

$$= \exp\{(\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1}(\underline{y} - \tfrac{\underline{s}_0 + \underline{s}_1}{2})\}, \quad \underline{y} \in \mathbb{R}^n,$$

where we use $\underline{s}_j^T \Sigma_N^{-1} \underline{y} = \underline{y}^T \Sigma_N^{-1} \underline{s}_j$ since Σ_N is a symmetric matrix.

Comparing to the scalar counterpart, the locations μ_0 and μ_1 are replaced by location vectors $\underline{s}_0$ and $\underline{s}_1$ and the noise variance σ^2 is replaced by the noise covariance matrix Σ_N.

The optimum tests

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1 & \text{if } (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{y} > \tau' \\ \gamma & \quad \cdot\cdot \qquad\qquad = \cdot\cdot \\ 0 & \quad \cdot\cdot \qquad\qquad < \cdot\cdot \end{cases}$$

where $\tau' = \log \tau + \frac{1}{2} (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_0 + \underline{s}_1)$.

* $(\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{y} = \tilde{\underline{s}}^T \underline{y} = \sum_{k=1}^{n} \tilde{s}_k y_k$,

where $\tilde{\underline{s}} = \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$ is the "pseudo signal". This is also the correlation detector but the signal is replaced by the pseudo signal.

* For the performance analysis, let us consider

$$T(\underline{Y}) \triangleq (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{Y}$$

Which is a linear combination of the Gaussian random vector $\underline{Y}$, is also a Gaussian.

$$E\{T(\underline{Y})|H_j\} = E\{\underline{\tilde{S}}^T \underline{Y} | H_j\}$$
$$= \underline{\tilde{S}}^T E\{\underline{Y} | H_j\}$$
$$= \underline{\tilde{S}}^T E\{\underline{N}|H_j\} + \underline{\tilde{S}}^T \underline{S}_j$$
$$= \underline{\tilde{S}}^T \underline{S}_j \triangleq \tilde{\mu}_j$$

$$Var(T(\underline{Y})|H_j) = E\{(\underline{\tilde{S}}^T\underline{Y} - \underline{\tilde{S}}^T\underline{S}_j)^2 | H_j\}$$
$$= E\{(\underline{\tilde{S}}^T\underline{N})^2\} = E\{\underline{\tilde{S}}^T \underline{N}\,\underline{N}^T \underline{\tilde{S}}\}$$
$$= \underline{\tilde{S}}^T E\{\underline{N}\,\underline{N}^T\} \underline{\tilde{S}} = \underline{\tilde{S}}^T \Sigma_N \underline{\tilde{S}}$$
$$= (\underline{S}_1 - \underline{S}_0)^T \Sigma_N^{-1} (\underline{S}_1 - \underline{S}_0) \triangleq d^2$$

This variance of $T(\underline{Y})$ is independent of the hypothesis.

$d > 0$ unless $\underline{S}_1 = \underline{S}_0$

$T(\underline{Y}) \sim N(\tilde{\mu}_j, d^2)$ under H_j, $j=0,1$.

$$* \; P_j(\Gamma_1) = \frac{1}{\sqrt{2\pi}\, d} \int_{\tau'}^{\infty} e^{-(x-\tilde{\mu}_j)^2/(2d^2)}\, dx$$
$$= 1 - \Phi\left(\frac{\tau' - \tilde{\mu}_j}{d}\right)$$

$$= \begin{cases} 1-\Phi\left(\frac{\log\tau}{d}+\frac{d}{2}\right) & \text{for } j=0 \\ 1-\Phi\left(\frac{\log\tau}{d}-\frac{d}{2}\right) & \text{for } j=1 \end{cases}$$

Which is the same as the simple scalar case with only the difference in d.

* For the Neyman–Pearson testing, it is also similar.

* Remark 1 : Interpretation of d^2

For simplicity, let us assume $\underline{s}_0=0$, $\underline{s}=\underline{s}$ and the noise is i.i.d. $N(0,\sigma^2)$. Then, $\Sigma_N=\sigma^2 I$ where I denotes the $n\times n$ identity matrix.

$$d^2=(\underline{s}_1-\underline{s}_0)^T\Sigma_N^{-1}(\underline{s}_1-\underline{s}_0)$$

$$=\frac{\underline{s}^T\underline{s}}{\sigma^2}=\frac{1}{\sigma^2}\sum_{k=1}^{n}s_k^2=n\,\frac{\overline{s^2}}{\sigma^2}$$

where $\overline{s^2}=\frac{1}{n}\sum_{k=1}^{n}s_k^2$ is the average signal power. $\sigma^2=\frac{1}{n}\sum_{k=1}^{n}E(N_k^2)$ is the average noise power.

$\Rightarrow$ d^2 is the signal-to-noise average power ratio times the number of samples.

The non-i.i.d. noise case is similar.

* d^2 has another interpretation for the i.i.d. case for two general signals:

$$d^2 = \frac{1}{\sigma^2} \|\underline{s}_1 - \underline{s}_0\|^2$$

where $\|\underline{s}_1 - \underline{s}_0\|$ denotes the Euclidean distance between the signal vectors $\underline{s}_0$ and $\underline{s}_1$ given by

$$\|\underline{s}_1 - \underline{s}_0\| = \left[\sum_{k=1}^{n} (s_{1k} - s_{0k})^2 \right]^{\frac{1}{2}}$$

The father apart the signal vectors are, the better performance can be achieved.

* Remark 2: Reduction to the i.i.d. Noise Case

Σ_N is an $n \times n$ symmetric positive-definite matrix. Let $\lambda_1, \cdots, \lambda_n$ be its eigenvalues and $\underline{v}_1, \cdots, \underline{v}_n$ be their corresponding eigenvectors. Since Σ_N is symmetric positive definite, $\lambda_1 \cdots \lambda_n$ are positive, $\underline{v}_1, \cdots, \underline{v}_n$ can be chosen orthonormal, i.e., $\underline{v}_k^T \underline{v}_l = 0$ if $k \neq l$ and $\underline{v}_k^T \underline{v}_k = 1$ for $1 \le k, l \le n$. Furthermore, $\Sigma_N = \sum_{k=1}^{n} \lambda_k \underline{v}_k \underline{v}_k^T$,

which is called the spectrum decomposition of Σ_N.

$\Rightarrow \quad \Sigma_N^{-1} = \sum_{k=1}^{n} \lambda_k^{-1} \underline{v}_k \underline{v}_k^T$

$\Rightarrow$ The optimum detection statistic $T(\underline{Y})$ can be written:

$$T(\underline{Y}) = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{y} = \sum_{k=1}^{n} (\hat{s}_{1k} - \hat{s}_{0k}) \hat{y}_k$$

where $\hat{y}_k = \underline{v}_k^T \underline{y} / \sqrt{\lambda_k}, \quad k = 1, \cdots, n$

$\hat{s}_{jk} = \underline{v}_k^T \underline{s}_j / \sqrt{\lambda_k}, \quad k = 1, \cdots, n, \quad j = 0, 1.$

* Note: $\underline{y} = \Sigma_{k=1}^{n} \sqrt{\lambda_k}\ \hat{y}_k\ \underline{v}_k$

$\Rightarrow$ $\hat{\underline{y}}$ is an equivalent observation to $\underline{y}$

Let $\hat{Y}$ be the random vector corresponding to $\hat{\underline{y}}$

$\Rightarrow$ The hypothesis pair for Y becomes

$$H_0 : \hat{\underline{Y}} = \hat{\underline{N}} + \hat{\underline{S}}_0$$

$$\text{vs} \quad H_1 : \hat{\underline{Y}} = \hat{\underline{N}} + \hat{\underline{S}}_1$$

where $\hat{N}_k = \underline{v}_k^T \underline{N} / \sqrt{\lambda_k}$ is Gaussian random variable since it is a linear combination of $\underline{N}$.

$$E\{\hat{N}_k \hat{N}_l\} = E\{\underline{v}_k^T \underline{N}\, \underline{v}_l^T \underline{N}\} / \sqrt{\lambda_k \lambda_l}$$

$$= E\{\underline{v}_k^T \underline{N}\, \underline{N}^T \underline{v}_l\} / \sqrt{\lambda_k \lambda_l}$$

$$= \underline{v}_k^T\ E\{\underline{N}\,\underline{N}^T\}\ \underline{v}_l / \sqrt{\lambda_k \lambda_l}$$

$$= \underline{v}_k^T\ \Sigma_N\ \underline{v}_l / \sqrt{\lambda_k \lambda_l}$$

$$= \underline{v}_k^T \left(\sum_{i=1}^{n} \lambda_i \underline{v}_i \underline{v}_i^T \right) \underline{v}_\ell \Big/ \sqrt{\lambda_k \lambda_\ell}$$

$$= \underline{v}_k^T \underline{v}_\ell \sqrt{\lambda_\ell / \lambda_k}$$

$$= \begin{cases} 1, & \text{if } k = \ell \\ 0, & \text{if } k \neq \ell \end{cases}$$

$\Rightarrow \hat{N}_1, \cdots, \hat{N}_n$ are i.i.d. $\mathcal{N}(0,1)$ random variables

$\Rightarrow$ By the appropriate linear tranformation of $\underline{Y}$,

dependent Gaussian noise $\longrightarrow$ independent Gaussian noise

* The original standard coordinate system in $\mathbb{R}^n$ is changed to a different coordinate system in which the usual axes are aligned with the vectors $\underline{v}_1, \cdots, \underline{v}_n$ so that the noise coordinates $\hat{N}_1, \cdots, \hat{N}_n$ are i.i.d.

* Another way to look at this change of coordinates is to write $\Sigma_N = B^2$, where $B = \sum_k^n \lambda_k^{1/2} \underline{v}_k \underline{v}_k^T$ that is called the square root of Σ_N.

$$B^{-1} = \sum_{k=1}^{n} \lambda_k^{-1/2} \underline{v}_k \underline{v}_k^T, \quad \Sigma_N^{-1} = (B^{-1})^2$$

If we define : $\underline{s}_j^* = B^{-1} \underline{s}_j$, $\underline{y}^* = B^{-1} \underline{y}$, then, $T(\underline{y}) = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{y} = (\underline{s}_1^* - \underline{s}_0^*)^T \underline{y}^*$

and under H_j, we have

$$\underline{Y}^* = \underline{N}^* + \underline{s}_j^* \quad \text{with } \underline{N}^* = B^{-1}\underline{N}$$

and
$$E\{\underline{N}^*(\underline{N}^*)^T\} = E\{B^{-1}\underline{N}\,\underline{N}^T B^{-1}\}$$
$$= B^{-1}E\{\underline{N}\,\underline{N}^T\}B^{-1} = B^{-1}\Sigma_N B^{-1} = B^{-1}BBB^{-1} = I.$$

$\Rightarrow N_1^*, \ldots, N_n^*$ are i.i.d. $\mathcal{N}(0,1)$.

$\Rightarrow$ This is similar to the observation $\hat{\underline{Y}}$. In fact, $\underline{Y}^*$ and $\hat{\underline{Y}}$ are the same random vectors in two different coordinate systems since $\underline{Y}^* = \sum_{k=1}^{n} \hat{Y}_k \underline{v}_k$ and $\hat{\underline{Y}} = \sum_{k=1}^{n} \hat{Y}_k \underline{e}_k$ where $\{\underline{e}_k\}$ are standard basis vectors for $\mathbb{R}^n$, i.e., $\underline{e}_k$ is all 0's except for a 1 in its kth component.

* Cholesky decomposition and Whitening Filter

Since Σ_N is symmetric positive definite, it can be written as $\Sigma_N = CC^T$

where C is an $n \times n$ invertible lower triangular matrix (i.e., all above-diagonal elements of C are zero). This is called the Cholesky decomposition of Σ_N. There are standard ways to find Cholesky decomposition. $\Sigma_N^{-1} = (C^T)^{-1} C^{-1} = (C^{-1})^T C^{-1}$.

Define $\bar{\underline{Y}} = C^{-1} \underline{Y} = C^{-1} \underline{N} + C^{-1} \underline{s}_j \triangleq \bar{\underline{N}} + \bar{\underline{s}}_j$

$\Rightarrow$ $\bar{\underline{N}} \sim N(0, I)$

$\Rightarrow$ i.i.d. noise and $T(\underline{Y}) = (\bar{\underline{s}}_1 - \bar{\underline{s}}_0)^T \bar{\underline{Y}}$.

C is lower triangular $\Rightarrow$ C^{-1} is also lower triangular.

$$\Rightarrow \quad \bar{y}_k = \sum_{l=1}^{k} h_{k,l} \, y_l$$

where $h_{k,l}$ is the k-lth element of C^{-1}. The above is a causal operation and it shows that $\bar{y}_1, \cdots, \bar{y}_n$ can be produced by a causal, but possibly time-variant, linear filteration of $y_1, \cdots, y_n$. Since the output noise of this filter is white, it is also known as <u>whitening filter</u>.

$\Rightarrow$ The optimum detector structure can be represented as the causal linear filter with impulse response $\{h_{k,l}\}$ driven by $y_1, \cdots, y_n$ and followed by a correlator in which the filter output is correlated with $(\bar{s}_{11} - \bar{s}_{01}), \cdots, (\bar{s}_{1n} - \bar{s}_{0n})$, the output of the same filter driven by the difference signal $(s_{11} - s_{01}), \cdots, (s_{1n} - s_{0n})$.

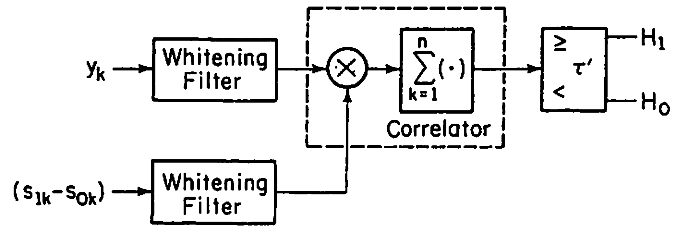


FIGURE III.B.8. Optimum detector for coherent signals in dependent Gaussian noise.

* SNR : $d^2 = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$

$$= \|\hat{\underline{s}}_1 - \hat{\underline{s}}_0\|^2 = \|\underline{s}_1^* - \underline{s}_0^*\|^2 = \|\bar{\underline{s}}_1 - \bar{\underline{s}}_0\|^2$$

The performance of coherent detection in dependent noise depends on how far apart the signals are when transformed to a coordinate system in which the noise components are i.i.d.

* Remark 3 : Signal Selection

The performance of the optimum coherent detection in Gaussian noise can be improved by increasing

$$d^2 = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$$

where $\underline{s}_1$ and $\underline{s}_0$ are two signals. In many applications, one may be able to have the flexibility to choose signals $\underline{s}_1$ and $\underline{s}_0$. In this case, given the noise covariance matrix, Σ_N, these two signals can be chosen optimally to optimize the performance by maximizing d^2.

To maximize d^2 for a given Σ_N (a given noise statistics), for any vector $\underline{x} \in \mathbb{R}^n$, we have

$$\underline{x}^T \Sigma_N^{-1} \underline{x} = \sum_{k=1}^{n} \lambda_k^{-1} \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x} \le \lambda_{min}^{-1} \sum_{k=1}^{n} \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x}$$

where $\lambda_{min} = \min\{\lambda_1, \cdots, \lambda_n\}$. Since

$$\sum_{k=1}^{n} \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x} = \underline{x}^T \underbrace{\sum_{k=1}^{n} \underline{v}_k \underline{v}_k^T}_{= I} \underline{x} = \underline{x}^T \underline{x} = \|x\|^2,$$

we have $\underline{x}^T \Sigma_N^{-1} \underline{x} \le \lambda_{min}^{-1} \|x\|^2$

where the equality holds if and only if $\underline{x}$ is proportional to an eigenvector corresponding to the eigenvalue λ_{min} [If there are more than one eigenvector corresponding to λ_{min}, $\underline{x}$ can be any linear combination of them].

$\Rightarrow$ To maximize d^2, $\underline{s}_1 - \underline{s}_0$ should be an eigenvector corresponding to the minimum eigenvalue of Σ_N.

The eigenvalues of Σ_N are measures of the noise power in the direction of their corresponding eigenvectors.

$\Rightarrow$ Putting the signal difference along the minimum eigenvalue eigenvector is equivalent to signaling the least noisy direction.

If we do that, we have

$$d^2 = \frac{1}{\lambda_{min}} \|\underline{s}_1 - \underline{s}_0\|^2.$$

Once we have chosen the direction of the signal difference $\underline{s}_1 - \underline{s}_0$, we can further optimize performance by maximizing $\|\underline{s}_1 - \underline{s}_0\|^2$.

Obviously, this quantity can be arbitrarily large if there is no contraints on the signals. However, signals are usually constrained by their total power.

Suppose we constrain $\|\underline{s}_1\|^2 \le P$ and $\|\underline{s}_0\|^2 \le P$ with $0 < P < \infty$. Then,

$$d^2 = \frac{1}{\lambda_{min}} (\underline{s}_1 - \underline{s}_0)^T (\underline{s}_1 - \underline{s}_0)$$

$$= \frac{1}{\lambda_{min}} \left(\|\underline{s}_1\|^2 + \|\underline{s}_0\|^2 - 2\, \underline{s}_1^T \underline{s}_0 \right)$$

Since $\underline{s}_1^T \underline{s}_0$ is the inner product of $\underline{s}_1$ and $\underline{s}_0$, with fixed $\|\underline{s}_1\|$ and $\|\underline{s}_0\|$, $\underline{s}_1^T \underline{s}_0$ is minimized if $\underline{s}_1$ and $\underline{s}_0$ are in opposite directions, i.e., if $\underline{s}_0 = \alpha \underline{s}_1$ with $\alpha < 0$. In this case,

$$d^2 = \frac{1}{\lambda_{min}} \left(\|\underline{s}_1\|^2 - 2\alpha \|\underline{s}_1\|^2 + \alpha^2 \|\underline{s}_1\|^2 \right)$$

$$= \frac{\|\underline{s}_1\|^2 (1-\alpha)^2}{\lambda_{min}}$$

$$\alpha = -\frac{\|\underline{s}_0\|}{\|\underline{s}_1\|} \Rightarrow d^2 = \frac{\left(\|\underline{s}_1\| + \|\underline{s}_0\|\right)^2}{\lambda_{min}}$$

$\Rightarrow$ d^2 is further maximized by choosing
$\|\underline{s}_0\|^2 = \|\underline{s}_1\|^2 = P$

$\Rightarrow$ $\alpha = -1$, i.e., $\underline{s}_0 = -\underline{s}_1$, and

$$\max_{\|\underline{s}_j\|^2 \le P} d^2 = \frac{4P}{\lambda_{min}}$$

$\Rightarrow$ $\underline{s}_1 = c\,\underline{v}_{min}$, $\underline{s}_0 = -\underline{s}_1$
where c is chosen so that $\|\underline{s}_1\|^2 = \|\underline{s}_0\|^2 = P$, i.e.
$c = \sqrt{P} / \|\underline{v}_{min}\|$

$\Rightarrow$ $\underline{s}_1 = \sqrt{P}\,\dfrac{\underline{v}_{min}}{\|\underline{v}_{min}\|}$ and $\underline{s}_0 = -\underline{s}_1$.

BPSK

* Example : Optimum Signals for Two-Sample Detection

Consider the case $n=2$ with $\Sigma_N = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $|\rho| < 1$. It is not hard to calculate

$\lambda_1 = \sigma^2(1-\rho)$, $\lambda_2 = \sigma^2(1+\rho)$

$\underline{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\underline{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\Rightarrow$ If $\rho > 0$, $\lambda_{min} = \lambda_1$ and optimum signals are
$\underline{s}_1 = \sqrt{\frac{P}{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\underline{s}_0 = \sqrt{\frac{P}{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

If $\rho < 0$, $\lambda_{min} = \lambda_2$ and optimum signals are

$$\underline{s}_1 = \sqrt{\tfrac{P}{2}}\begin{pmatrix}1\\1\end{pmatrix} \text{ and } \underline{s}_0 = \sqrt{\tfrac{P}{2}}\begin{pmatrix}-1\\-1\end{pmatrix}$$

$$\Rightarrow \quad d^2 = \frac{4P}{\sigma^2(1-|\rho|)}$$

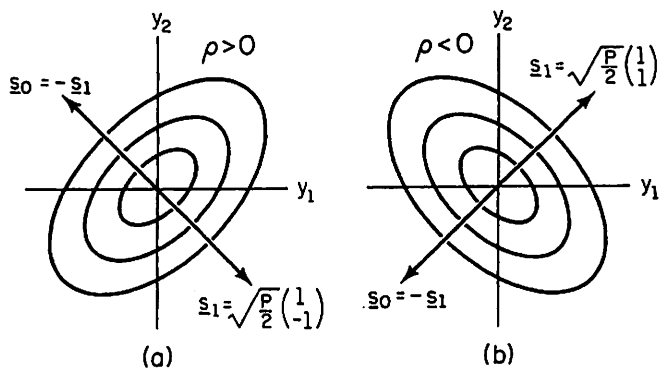


FIGURE III.B.9. Illustration of optimum signals for Gaussian noise with $\Sigma_N = \sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$.

In either case, the signal vectors are in the directions in which the noise density falls off the fastest.

Note: In this case, one only needs to know the sign of ρ, not its actual value, to choose the optimal signals.

HWK:

* Case 3: Detection of Signals with Random Parameters

* In many applications, signals to detect may not be completely known but there are a set of unknown parameters. Examples of such signals include signals in communication systems where one of two waveforms (representing "0" and "1", respectively) is modulated onto a sinusoidal carrier at the transmitter and the receiver needs to decide which was sent. Although the two signaling waveforms and the carrier frequency are known at the receiver the amplitude and phase of the carrier may not be. These unknown quantities represent unknown parameters.

* The hypothesis problem can be written as

$$H_0: \quad Y_k = N_k + s_{0k}(\theta), \quad k = 1, 2, \cdots, n,$$

vs.

$$H_1: \quad Y_k = N_k + s_{1k}(\theta), \quad k = 1, 2, \cdots, n.$$

where $\underline{s}_0(\theta)$ and $\underline{s}_1(\theta)$ are known vector-valued function of θ, which is an unknown parameter taking values in a parameter set Λ. Assume θ is random and we use Θ to represent the random variable, with density W_j under hypothesis H_j.

The likelihood ratio is

$$L(\underline{y}) = \frac{E_1\{p_N(\underline{y}-\underline{s}_1(\theta))\}}{E_0\{p_N(\underline{y}-\underline{s}_0(\theta))\}}$$

$$= \frac{\int_\Lambda p_N(\underline{y}-\underline{s}_1(\theta))\, w_1(\theta)\,\mu(d\theta)}{\int_\Lambda p_N(\underline{y}-\underline{s}_0(\theta))\, W_0(\theta)\,\mu(d\theta)}$$

* For convenience, assume $\underline{s}_0(\theta)=\underline{0}$ and $\underline{s}_1(\theta) \triangleq \underline{s}(\theta)$ and other cases can be handled similarly. Then

$$L(\underline{y}) = \int_\Lambda \frac{p_N(\underline{y}-\underline{s}(\theta))}{p_N(\underline{y})}\, w(\theta)\,\mu(d\theta)$$

$$= \int_\Lambda L_\theta(\underline{y})\, w(\theta)\,\mu(d\theta) \qquad (\text{III}.B.62)$$

Where $L_\theta(\underline{y})$ is the likelihood ratio conditional on $\Theta=\theta$ and the subscript 1 in $W_1(\theta)$ is dropped.

The above likelihood ratio $L(\underline{y})$ in (III.B.62) is simply the averaged (over θ) likelihood ratio for known θ.

With θ known, the hypothesis testing is a deterministic signal problem and $L_\theta(\underline{y})$ can be found by the previously studied cases. For example, with i.i.d. $N(0,\sigma^2)$ noise,

$$L(\underline{y}) = \int_\Lambda \exp\{[\underline{s}^T(\theta)\underline{y} - \tfrac{1}{2}\|\underline{s}(\theta)\|^2]/\sigma^2\}\, w(\theta)\,\mu(d\theta).$$

* Example: Noncoherent Detection a Modulated Sinusoidal Carrier

Consider the signal pair $\underline{s}_0(\theta)=\underline{0}$ and $\underline{s}_1(\theta)=\underline{s}(\theta)$ with

$$s_k(\theta)=a_k \sin[(k-1)\omega_c T_s+\theta],\ k=1,2,\cdots,n,$$

where $a_1, a_2, \cdots, a_n$ is a known amplitude sequence, Θ is a random phase angle independent of the noise and uniformly distributed on $[0, 2\pi]$, ω_c and T_s are a known carrier frequency and sampling interval length with the relationship

$$n\omega_c T_s = m2\pi$$

for some integer m (i.e., there are an integral number of periods of sinusoid in the time interval $[0, nT_s]$). Also assume that the number of samples taken per cycle of the sinusoid, i.e., $\frac{n}{m}$, is an integer than 1.

These signals provide a model, for example, for a digital signaling scheme in which a "0" is transmitted by sending nothing during the interval $[0, nT_s]$ and a "1" is transmitted by sending a signal $a(t)$ modulated onto a sinusoidal carrier of frequency ω_c. This signaling scheme is known as on-off keying (OOK). In this case, the sequence $a_1, \cdots, a_n$ is the sampled waveform $a(t)$, i.e., $a_k = a((k-1)T_s)$ and Θ represents the

the phase angle of the carrier, which is assumed unknown here.

Detection of a modulated carrier in which the carrier phase is unknown at the receiver is called <u>noncoherent detection</u>. The assumption that the phase angle is uniform on $[0, 2\pi)$ represents a belief that all phases are equally likely to occur, which is a reasonable assumption due to the folding process of a phase.

* Assume i.i.d. $N(0, \sigma^2)$ noise, the likelihood ratio for the above problem is

$$L(\underline{y}) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{ \frac{1}{\sigma^2} \left(\sum_{k=1}^{n} y_k s_k(\theta) - \frac{1}{2} \sum_{k=1}^{n} s_k^2(\theta) \right) \right\} d\theta$$

$$\sum_{k=1}^{n} y_k s_k(\theta) = y_c \sin\theta + y_s \cos\theta$$

where

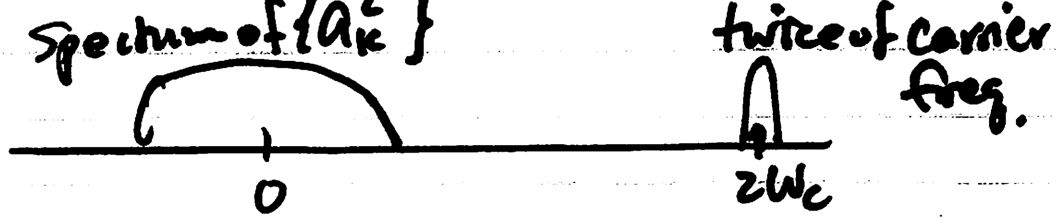
$$y_c \triangleq \sum_{k=1}^{n} a_k y_k \cos[(k-1)\omega_c T_s]$$

$$y_s \triangleq \sum_{k=1}^{n} a_k y_k \sin[(k-1)\omega_c T_s]$$

and

$$-\frac{1}{2} \sum_{k=1}^{n} s_k^2(\theta) = -\frac{1}{4} \sum_{k=1}^{n} a_k^2 + \underbrace{\frac{1}{4} \sum_{k=1}^{n} a_k^2 \cos(2(k-1)\omega_c T_s + 2\theta)}_{\to 0}$$

sequence $a_1^2, \dots, a_n^2$ has low frequency components compared to the twice of the carrier frequency ω_c. The summation is the 0th term of the convolution of these two signals.



$\Rightarrow$ The spectrum of the convolution of $\{a_k^2\}$ and $\cos(2(k-1)W_c+\theta)$ is the product of that of a_k^2 and the twice of the carrier freq. (these two do not overlap)

$\Rightarrow$ This product is 0

$\Rightarrow$ The convolution is 0

$$\Rightarrow L(\underline{y}) = e^{-n\overline{a^2}/(4\sigma^2)} \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{\frac{1}{\sigma^2}(y_c \sin\theta + y_s\cos\theta)\right\} d\theta$$

where $\overline{a^2} = \frac{1}{n}\sum_{k=1}^{n} a_k^2$

This is similar to the radius test example before.

$$L(\underline{y}) = e^{-n\overline{a^2}/(4\sigma^2)}\, I_0(r/\sigma^2)$$

Where $r = [y_c^2 + y_s^2]^{\frac{1}{2}}$ and I_0 is the zeroth-order modified Bessel function of the first kind. Since $I_0(\cdot)$ is monotonic, the optimum test is

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1, & \text{if } r > r' \triangleq I_0^{-1}\left(\tau e^{n\overline{a^2}/(4\sigma^2)}\right) \\ \gamma, & \quad\; = \\ 0, & \quad\; < \end{cases}$$

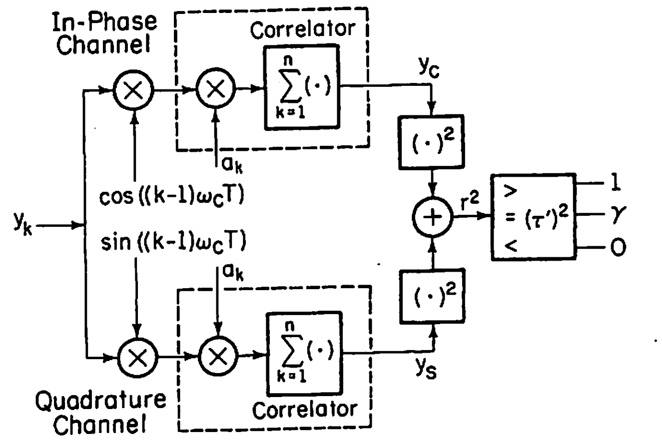


FIGURE III.B.10. Optimum system for noncoherent detection of a modulated sinusoid in i.i.d. Gaussian noise.

Note: the observed signal $y_1, \dots, y_n$ is split into two channels, one of which multiplies each y_k by $\cos((k-1)\omega_c T_s)$ and the other multiplies each y_k by $\sin((k-1)\omega_c T_s)$, called in-phase and quadrature channels, respectively. Each channel then correlates with a_k. The channel outputs are then combined to give r, which is compare to a threshold. This is known as an <u>envelope detector</u>.

Intuitively, when a signal is present, each channel picks up an amount of the signal energy depending on the actual phase angle of the carrier. Regardless of the carrier phase, the combination of the two channels pick up all the signal energy.

* Performance Analysis

We need to calculate $P_j(R > \tau') = P_j(R^2 > \tau'^2)$, $j = 0, 1$.

$R^2 = Y_c^2 + Y_s^2$

$$Y_c = \sum_{k=1}^{n} a_k Y_k \cos((k-1) w_c T_s)$$

$$Y_s = \sum_{k=1}^{n} a_k Y_k \sin((k-1) W_c T_s)$$

Under H_0, $\underline{Y} \sim N(0, \sigma^2 I)$, $\underline{Y} = (Y_1, \ldots, Y_n)^T$.

$\Rightarrow Y_c, Y_s$ are both Gaussian under H_0.

$$E\{Y_c | H_0\} = \sum_{k=1}^{n} a_k E\{N_k\} \cos((k-1) W_c T_s) = 0$$

$$Var\{Y_c | H_0\} = E\{Y_c^2 | H_0\} = \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_k a_\ell E\{N_k N_\ell\}$$

$$\times \cos[(k-1) w_c T_s] \cos[(\ell-1) w_c T_s]$$

$$= \sigma^2 \sum_{k=1}^{n} a_k^2 \cos^2[(k-1) w_c T_s] = \frac{n \sigma^2 \overline{a^2}}{2}$$

$$\frac{1}{2}\{1 + \cos[2(k-1) w_c T_s]\}$$

$$\sum_{k=1}^{n} a_k^2 \cos[2(k-1) w_c T_s] \to 0$$

$E\{Y_s | H_0\} = 0$

$Var\{Y_s | H_0\} = Var\{Y_c | H_0\}$

$Cov\{Y_c, Y_s | H_0\} = E\{Y_c Y_s | H_0\}$

$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_k a_\ell E[N_k N_\ell] \cos((k-1) w_c T_s) \sin((\ell-1) W_c T_s)$$

$$= \sigma^2 \sum_{k=1}^{n} a_k^2 \cos((k-1)w_c T_s)\sin((k-1)w_c T_s)$$

$$= \frac{\sigma^2}{2} \sum_{k=1}^{n} a_k^2 \sin[2(k-1)w_c T_s] \longrightarrow 0$$

$$\Rightarrow P_F(\Gamma_1) = \iint_{y_s^2+y_s^2 \geq (\tau')^2} \frac{1}{\pi n \sigma^2 \overline{a^2}} e^{-(y_c^2+y_s^2)/(n\sigma^2\overline{a^2})} dy_c dy_s$$

$$= \frac{1}{\pi n \sigma^2 \overline{a^2}} \int_0^{2\pi}\int_{\tau'}^{\infty} r e^{-r^2/(n\sigma^2\overline{a^2})} dr\, d\psi$$

$$= e^{-(\tau')^2/(n\sigma^2\overline{a^2})}$$

false alarm probability

Under H_0, Y_c and Y_s are independent $N(0, n\sigma^2\overline{a^2}/2)$ random variables.

To determine the detection probability, we need to find the joint pdf of Y_c and Y_s under H_1.

Given $\Theta = \theta$, $\underline{Y}$ has a conditional $N(\underline{s}(\theta), \sigma^2 I)$

$$E\{Y_c | H_1, \Theta=\theta\} = \sum_{k=1}^{n} a_k E\{Y_k | H_1, \Theta=\theta\} \cos((k-1)w_c T_s)$$

$$= \sum_{k=1}^{n} a_k^2 \sin((k-1)w_c T_s + \theta)\cos((k-1)w_c T_s)$$

$$= \frac{n\overline{a^2}}{2} \sin\theta$$

$$E\{Y_s/H_1, \Theta=\theta\} = \frac{n\overline{a^2}}{2}\cos\theta$$

With θ fixed, the covariances under H_1 of
variances
Y_c and Y_s are unchanged since Θ is only a shift in mean.

$$p_{Y_c,Y_s}(y_c, y_s|H_1)$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\pi n\sigma^2\overline{a^2}}\, e^{-\frac{1}{n\sigma^2\overline{a^2}}\, q(y_c, y_s;\, \frac{n\overline{a^2}}{2}, \theta)}\, d\theta$$

$$= p_{Y_c,Y_s}(y_c, y_s|H_0)\, e^{-n\overline{a^2}/(4\sigma^2)}\, I_0\left(\frac{[y_c^2+y_s^2]^{\frac{1}{2}}}{\sigma^2}\right)$$

where $q(y_c, y_s;\, \frac{n\overline{a^2}}{2}, \theta)$

$$= \left[\left(y_c - \frac{n\overline{a^2}}{2}\cos\theta\right)^2 + \left(y_s - \frac{n\overline{a^2}}{2}\sin\theta\right)^2\right]$$

$$\Rightarrow P_D(\tilde{\delta}_0) = P_1(\Gamma_1) = \iint_{y_c^2+y_s^2>(\tau')^2} p_{Y_c,Y_s}(y_c, y_s|H_1)\,dy_c\,dy_s$$

$$= \frac{e^{-n\overline{a^2}/(4\sigma^2)}}{\pi n\sigma^2\overline{a^2}}\int_0^{2\pi}\int_{\tau'}^{\infty} r e^{-r^2/(n\sigma^2\overline{a^2})}\, I_0\left(\frac{r}{\sigma^2}\right) dr\,d\psi$$

$$= \int_{\tau_0}^{\infty} x e^{-(x^2+b^2)/2}\, I_0(bx)\,dx \triangleq Q(b, \tau_0)$$

where $b^2 = n\overline{a^2}/(2\sigma^2)$, $\tau_0 = \tau'/(\sigma^2 b)$ and $x = \frac{r}{\sigma^2 b}$

The above function Q is known as Marcum's Q-function. Note $Q(0, \tau_0) = P_F(\tilde{\delta}_0)$

For α-level Neyman–Pearson detection,

$$\tau' = [n \sigma^2 \overline{a^2} \log(1/\alpha)]^{\frac{1}{2}}$$

$$P_D(\tilde{\delta}_0) = Q\left(b, \left[2 \log \frac{1}{\alpha}\right]^{\frac{1}{2}}\right)$$

Fig is similar to Fig. II.D.4.

Note : the average signal energy is

$$E\left\{\frac{1}{n}\sum_{k=1}^{n} s_k^2(\Theta)\right\} = \frac{1}{n}\frac{1}{2\pi}\int_0^{2\pi}\sum_{k=1}^{n} a_k^2 \sin^2((k-1)\omega_c T_s + \theta)\, d\theta$$

$$= \frac{\overline{a^2}}{2}$$

$\Rightarrow b^2 = \dfrac{n\overline{a^2}}{2\sigma^2}$ is SNR similar to d^2 in the coherent detection problem

Note when θ is known,

$$d^2 = \frac{1}{\sigma^2}\sum_{k=1}^{n} s_k^2(\theta) = \frac{n\overline{a^2}}{2\sigma^2} = b^2$$

$\boxed{\text{SNRs are the same.}}$

$$\sum_{k=1}^{n} a_k^2 \underbrace{\sin^2((k-1)\omega_c T_s + \theta)}_{= \frac{1}{2}(1 - \cos 2((k-1)\omega_c T_s + \theta))}$$

If we set $Q[b, [2\log\frac{1}{\alpha}]^{\frac{1}{2}}]$ ← noncoherent

$$= 1 - \Phi[\Phi^{-1}(1-\alpha) - d]$$ ← coherent

$\Rightarrow$ $b \approx d + 0.4$.

$\Rightarrow$ If we wish to use a noncoherent technique slightly higher SNR is required to achieve the same performance as the corresponding coherent technique.

The disadvantage of a coherent system is that some means for deriving the carrier phase must be provided.

* When two signals are of the same form but different amplitudes:

$$S_{jk}(\theta) = a_{jk} \sin[(k-1)\omega_c T_s + \theta], \quad k=1,2,\dots,n, \; j=0,1$$

For i.i.d. noise $N(0,\sigma^2)$, the likelihood ratio

$$L(\underline{y}) = \frac{e^{-n\overline{a_1^2}/(4\sigma^2)}\, I_0(r_1/\sigma^2)}{e^{-n\overline{a_0^2}/(4\sigma^2)}\, I_0(r_0/\sigma^2)}$$

where $\overline{a_j^2} = \frac{1}{n}\sum_{k=1}^{n} a_{jk}^2$, $r_j = [\, y_{c_j}^2 + y_{s_j}^2 \,]^{\frac{1}{2}}$

$$y_{c_j} = \sum_{k=1}^{n} a_{jk}\, y_k \cos((k-1)\omega_c T_s)$$

$$y_{s_j} = \sum_{k=1}^{n} a_{jk}\, y_k \sin((k-1)\omega_c T_s)$$

The optimum detection involves combining the output of two systems like the fig previously, one matched to each of the amplitude sequences.

When the two sequences have balanced energy, i.e., $\overline{a_0^2} = \overline{a_1^2}$, assume uniform cost and equal priors ($\tau = 1$), then the Bayes test is

$$\delta_B(\underline{y}) = \begin{cases} 1, & \text{if } \tau_1 > \tau_0 \\ 0 \text{ or } 1, & \text{if } \tau_1 = \tau_0 \\ 0, & \text{if } \tau_1 < \tau_0 \end{cases}.$$

When two amplitude sequences orthogonal, i.e.,

$$\sum_{k=1}^{n} a_{jk} a_{\ell k} = 0 \quad \text{if } j \neq \ell$$

and $\sum_{k=1}^{n} a_{jk} a_{\ell k} \sin((k-1)\omega_c T_s + \theta) = 0$ for all θ,

then $P_e = \frac{1}{2} e^{-b^2/4}$

where $b^2 = \dfrac{n\overline{a_1^2}}{2\sigma^2} \quad \left(\equiv \dfrac{n\overline{a_0^2}}{2\sigma^2}\right)$

* Case 4: Detection of Stochastic Signals

Both signals and noise are Gaussian random vectors and the hypothesis test problem:

$$H_0 : \underline{Y} \sim N(\underline{\mu}_0, \Sigma_0)$$

$$\text{vs.} \quad H_1 : \underline{Y} \sim N(\underline{\mu}_1, \Sigma_1)$$

The logarithm of the likelihood ratio is given by

$$\log L(\underline{y}) = \frac{1}{2}\log\frac{|\Sigma_0|}{|\Sigma_1|} + \frac{1}{2}(\underline{y}-\underline{\mu}_0)^T\Sigma_0^{-1}(\underline{y}-\underline{\mu}_0) - \frac{1}{2}(\underline{y}-\underline{\mu}_1)^T\Sigma_1^{-1}(\underline{y}-\underline{\mu}_1)$$

$$= \frac{1}{2}\underline{y}^T[\Sigma_0^{-1}-\Sigma_1^{-1}]\underline{y} + [\underline{\mu}_1^T\Sigma_1^{-1}-\underline{\mu}_0^T\Sigma_0^{-1}]\underline{y} + C$$

where $C = \frac{1}{2}\log\frac{|\Sigma_0|}{|\Sigma_1|} + \underline{\mu}_0^T\Sigma_0^{-1}\underline{\mu}_0 - \underline{\mu}_1^T\Sigma_1^{-1}\underline{\mu}_1$

quadratic term of $\underline{y}$ linear term of $\underline{y}$

If the two covariance matrices are the same, $\Sigma_0 = \Sigma_1 = \Sigma$, then the quadratic term disappears, and we have a linear test statistic

$$(\underline{\mu}_1 - \underline{\mu}_0)^T\Sigma^{-1}\underline{y}$$

where C can be absorbed into the threshold. This is the case of coherent detection in Gaussian noise we studied before.

If the mean vectors are the same under both hypotheses, $\underline{\mu}_0 = \underline{\mu}_1$, we can take them to be $\underline{0}$ and then $L(\underline{y})$ is only quadratic.

Detection of zero-mean stochastic signals in Gaussian noise:

$$H_0 : \underline{Y} = \underline{N}$$

vs.

$$H_1 : \underline{Y} = \underline{N} + \underline{S}$$

where $\underline{N} \sim N(\underline{0}, \sigma^2 I)$ and $\underline{S} \sim N(\underline{0}, \Sigma_S)$ and $\underline{N}$ and $\underline{S}$ are independent. When $\underline{N}$ is not i.i.d., it can be pre-whitened to fit the above model.

With the above model, $\Sigma_0 = \sigma^2 I$, $\Sigma_1 = \sigma^2 I + \Sigma_S$.
The optimum tests are of the form:

$$\tilde{\delta}_0(\underline{y}) = \begin{cases} 1, & \text{if } \underline{y}^T Q \underline{y} > \tau' \\ \gamma, & \quad \text{"} \quad = \text{"} \\ 0, & \quad \text{"} \quad < \text{"} \end{cases}$$

quadratic detector.

where $\tau' \triangleq 2(\log \tau - C)$

$$Q \triangleq \sigma^{-2} I - (\sigma^2 I + \Sigma_S)^{-1} = \sigma^{-2} \Sigma_S (\sigma^2 I + \Sigma_S)^{-1}$$

If the signal samples are i.i.d. $N(0, \sigma_S^2)$ random variables, then $\Sigma_S = \sigma_S^2 I$ and

$$\underline{y}^T Q \, \underline{y} = \frac{\sigma_s^2}{\sigma^2(\sigma^2+\sigma_s^2)} \sum_{k=1}^{n} y_k^2 .$$

In this case, the optimum detector compares the quantity $\sum_{k=1}^{n} y_k^2$ to a threshold.

This detector is known as an energy detector (also known as a radiometer).

For the performance analysis, we need to calculate

$$P_j(\underline{Y}^T Q \, \underline{Y} > \tau'), \quad j = 0, 1.$$

Let $\lambda_1, \cdots, \lambda_n$ and $\underline{v}_1, \cdots, \underline{v}_n$ be the eigenvalues and corresponding orthonormal eigenvectors of the signal covariance matrix Σ_s. Then

$$\Sigma_s = \sum_{k=1}^{n} \lambda_k \underline{v}_k \underline{v}_k^T$$

$$I = \sum_{k=1}^{n} \underline{v}_k \underline{v}_k^T$$

$$\Rightarrow (\sigma^2 I + \Sigma_s)^{-1} = \sum_{k=1}^{n} (\sigma^2 + \lambda_k)^{-1} \underline{v}_k \underline{v}_k^T$$

and

$$Q = \sum_{k=1}^{n} \frac{1}{\sigma^2} \underline{v}_k \underline{v}_k^T - \sum_{k=1}^{n} \frac{1}{\sigma^2+\lambda_k} \underline{v}_k \underline{v}_k^T$$

$$= \sum_{k=1}^{n} \frac{\lambda_k}{\sigma^2(\sigma^2+\lambda_k)} \underline{v}_k \underline{v}_k^T$$

Then, $\underline{y}^T Q \underline{y} = \sum_{k=1}^{n} (\bar{y}_k)^2$

where $\bar{y}_k = \left[\dfrac{\lambda_k}{\sigma^2(\sigma^2+\lambda_k)}\right]^{\frac{1}{2}} \underline{v}_k^T \underline{y}$

Their corresponding random variables $\bar{Y}_1, \cdots, \bar{Y}_n$ are independent Gaussian with variance

$$\sigma_{jk}^2 \triangleq \mathrm{Var}(\bar{Y}_k \mid H_j) = \begin{cases} \dfrac{\lambda_k}{\sigma^2+\lambda_k} & \text{if } j=0 \\ \dfrac{\lambda_k}{\sigma^2} & \text{if } j=1 \end{cases}$$

$\Rightarrow$ under H_j, $\underline{Y}^T Q \underline{Y}$ is the sum of independent random variables in which the kth term's square root has $N(0, \sigma_{jk}^2)$ distribution.

Let $T_k \triangleq \bar{Y}_k^2$ and $T \triangleq \sum_{k=1}^{n} T_k$

$$p_{T_k}(t \mid H_j) = \begin{cases} \dfrac{1}{\sqrt{2\pi t}\,\sigma_{jk}} e^{-t/(2\sigma_{jk}^2)}, & t>0 \\ 0, & t \le 0 \end{cases}$$

which is a gamma $(\frac{1}{2}, \frac{1}{2\sigma_{jk}^2})$ density.

$$p_T = p_{T_1} * p_{T_2} * \cdots * p_{T_n}$$

$$= \mathcal{F}^{-1}\left\{\prod_{k=1}^{n} \phi_{T_k}\right\}$$

where $\phi_{T_k}(u) = \mathcal{F}\{p_{T_k}\}(u) = E\{e^{iuT_k}\}$ is the characteristic function of T_k.