EE 631: Estimation and Detection
Part 2

Dr. Mehrdad Soumekh

Bayes decision rule (continued)

In the last section we noted that the Bayes decision rule can be written as:

$$I_k(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} I_m(\vec{R})$$

which means that our task is to select the hypothesis with the minimum $I_i(\vec{R})$.

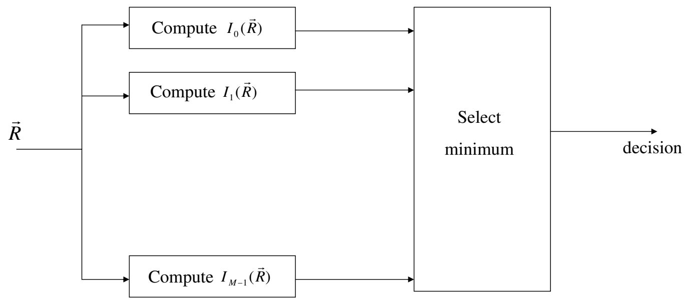

Figure 1. Block diagram of the decision process

The above decision rule is re-written as:

$$\sum_j P_j(C_{kj} - C_{jj})p(\vec{R}\big|H_j) \gtrless^{\text{not } H_k}_{\text{not } H_m} \sum_j P_j(C_{mj} - C_{jj})p(\vec{R}\big|H_j)$$

After cancelling the common terms, we obtain

$$(C_{km} - C_{mm})P_m p(\vec{R}\big|H_m) \gtrless^{\text{not } H_k}_{\text{not } H_m} (C_{mk} - C_{kk})P_k p(\vec{R}\big|H_k) + \sum_{j\neq m, j\neq k} (C_{mj} - C_{kj})P_j p(\vec{R}\big|H_j)$$

Define the ith likelihood function via:

$$\Lambda_i(\vec{R}) \triangleq \frac{p(\vec{R}\big|H_i)}{p(\vec{R}\big|H_0)}$$

Therefore, the test becomes:

$$(C_{km} - C_{mm})P_m \Lambda_m(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} (C_{mk} - C_{kk})P_k \Lambda_k(\vec{R}) + \sum_{j\neq m, j\neq k} (C_{mj} - C_{kj})P_j \Lambda_j(\vec{R})$$

- Note that there are $\frac{M(M-1)}{2}$ such inequalities.
- Note that the decision space is the $(M-1)$ dimensional space of the likelihood function.

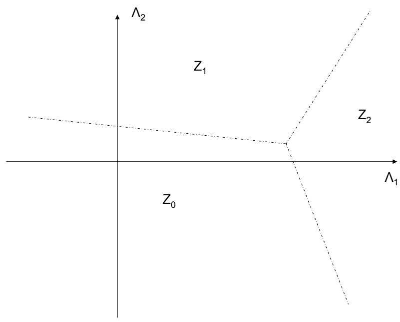

Figure 2. Decision regions

Special cases - Minimum probability of error criterion:-

Cost functions are defined via:

$$C_{ij} = \begin{cases} 1 & ; i \neq j \\ 0 & ; i = j \end{cases}$$

i.e. unit cost function for all wrong decisions.
In this case, the risk function becomes:

$$\mathbb{R} = \sum_j P_j \sum_i C_{ij} \int_{\mathbb{Z}_i} p(\vec{R} \big| H_j) d\vec{R}$$

$$= \sum_j P_j \underbrace{\sum_{i \neq j} \underbrace{\int_{\mathbb{Z}_i} p(\vec{R} \big| H_j) d\vec{R}}_{\text{Prob[choose } H_i \big| H_j \text{ is true]}}}_{\text{Prob[error } \big| H_j \text{ is true]}}$$

$$= \sum_j \text{Prob[error } \big| H_j \text{ is true]} = P(error)$$

- Thus, minimizing the risk function with $C_{ij} = \delta_{ij}$ is equivalent to minimizing the overall probability (expected value) of error percentage.
After substituting $C_{ij} = \delta_{ij}$ in

$$I_k(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} I_m(\vec{R})$$

we obtain:

$$P_m \Lambda_m(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} P_k \Lambda_k(\vec{R})$$

$$P_m p(\vec{R} \big| H_m) \gtrless^{\text{not } H_k}_{\text{not } H_m} P_k p(\vec{R} \big| H_k) \tag{1}$$

Thus, for a given measurement $\vec{R}$, select the hypothesis that maximizes

$$P_i \Lambda_i(\vec{R}), \forall i = 0, 1, \cdots, M-1$$

Rewrite equation (1) using the log-likelihood function:

$$\ln P_m + \ln \Lambda_m(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} \ln P_k + \ln \Lambda_k(\vec{R})$$

Going back to the original channel pdf's, we may also write the decision rule via:

$$P_m p(\vec{R}|H_m) \gtrless^{\text{not } H_k}_{\text{not } H_m} P_k p(\vec{R}|H_k)$$

Dividing both sides by $p(\vec{R})$, where $p(\vec{R})=\sum_{j=0}^{M-1} p(\vec{R}|H_j)$, we get:

$$\frac{P_m p(\vec{R}|H_m)}{p(\vec{R})} \gtrless^{\text{not } H_k}_{\text{not } H_m} \frac{P_k p(\vec{R}|H_k)}{p(\vec{R})}$$

Recall: $p(a|b) = \frac{p(a)p(b|a)}{p(b)}$
Associate $a \to H_m$, $b \to \vec{R}$

$$\therefore p(H_m|\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} p(H_k|\vec{R})$$

- Therefore, the decision rule is to select the hypothesis that yields the maximum a posteriori pdf (i.e. the probability of a hypothesis given an observation).

- This is called the *Maximum A Posteriori Probability* (MAP) decision rule.

Summary
Minimum probability of error decision rule and MAP decision rule are the same; they are special cases of the Bayes decision rule with uniform cost for all incorrect decisions.

Example
Binary decision, i.e $M = 2 \Rightarrow$ only one likelihood function.

$$\Lambda_1(\vec{R}) = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \triangleq \Lambda(\vec{R})$$

$\therefore$ The Bayes decision becomes

$$\Lambda(\vec{R}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \triangleq \eta$$

η, the threshold is set by the user based on the a priori probabilities (i.e P_0 and $P_1 = 1 - P_0$) and the assigned cost functions.

Receiver
The receiver structure is as shown in figure 3.

Special case

MAP decision rule: substitute $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$.

$$\Rightarrow \eta = \frac{P_0(1-0)}{P_1(1-0)} = \frac{P_0}{P_1} = \frac{P_0}{(1-P_0)}$$

Log likelihood form:

$$L(\vec{R}) \triangleq \ln \Lambda(\vec{R}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \ln \eta \triangleq \xi$$

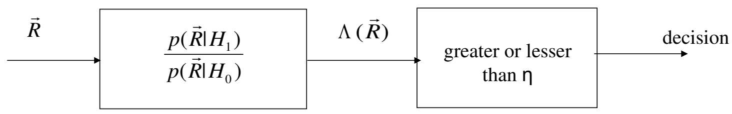

Figure 3. Receiver

where

$$\xi = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

e.g.:

$$H_0 : \vec{R} = \vec{S_0} + \vec{N}$$
$$H_1 : \vec{R} = \vec{S_1} + \vec{N}$$

where

$$\vec{S_0} = \begin{bmatrix} S_{01} \\ S_{02} \\ \vdots \\ S_{0N} \end{bmatrix}, \vec{S_1} = \begin{bmatrix} S_{11} \\ S_{12} \\ \vdots \\ S_{1N} \end{bmatrix}, \vec{N} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix}$$

<u>Source and channel model</u>

n_i's are independent identically distributed (iid) normal or Gaussian r.v.'s with zero mean and variance σ_N^2; $n_i \sim N(0, \sigma_N^2)$, $\forall\ i$.

$$E(n_i n_j) = \sigma_N^2 \delta_{ij}$$

$$E(n_i) = 0$$

$$p(\vec{N}) = \frac{1}{(\sqrt{2\pi}\sigma_N)^N} exp\left[\frac{-\sum_{i=1}^N n_i^2}{2\sigma_n^2}\right]$$

S_0 and S_1 are deterministic (known) vectors, e.g. samples of two known signals.

The observation pdf:

$$H_k : \vec{R} = \vec{S_k} + \vec{N}; k = 0, 1$$

$\Rightarrow \vec{R}|H_k$ is also normal multivariate with mean $E(\vec{R}|H_k) = \vec{S_k}$ and covariance matrix:

$$Cov = \begin{bmatrix} \sigma_n^2 & 0 & \cdots & 0 \\ 0 & \sigma_n^2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

In this case its pdf is:

$$p(\vec{R}|\vec{H_k}) = \frac{1}{(\sqrt{2\pi}\sigma_N)^N} exp\left[\frac{-\sum_{i=1}^{N}(r_i - S_{ki})^2}{2\sigma_n^2}\right]$$

where

$$\vec{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$

Likelihood function:

$$\Lambda(\vec{R}) = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)}$$

$$= \frac{exp\left[-\sum_{i=1}^{N}(r_i - S_{1i})^2/2\sigma_n^2\right]}{exp\left[-\sum_{i=1}^{N}(r_i - S_{0i})^2/2\sigma_n^2\right]}$$

Log-likelihood function:

$$L(\vec{R}) = \ln(\Lambda\vec{R})$$

$$\Rightarrow \frac{\sum_{i=1}^{N}(r_i - S_{0i})^2}{2\sigma_n^2} - \frac{\sum_{i=1}^{N}(r_i - S_{1i})^2}{2\sigma_n^2} \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi$$

$$\Rightarrow \frac{\sum_{i=1}^{N} 2r_i(S_{1i} - S_{0i}) - (S_{1i}^2 - S_{0i}^2)}{2\sigma_n^2} \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi$$

Rewrite the above via:

$$\sum_{i=1}^{N} 2r_i(S_{1i} - S_{0i}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi\sigma_n^2 + \frac{1}{2}\sum_{i=1}^{N} S_{1i}^2 - \frac{1}{2}\sum_{i=1}^{N} S_{0i}^2$$

Define

$$E_0 \triangleq \sum_{i=1}^{N} S_{0i}^2 : \text{energy of } \vec{S_0}$$

$$E_1 \triangleq \sum_{i=1}^{N} S_{1i}^2 : \text{energy of } \vec{S_1}$$

and note that $\sum_{i=1}^{N} r_i S_{ki} = \vec{R}^T\vec{S_k} \triangleq <\vec{R}, \vec{S_k}>$, also known as the projection of S_k onto $\vec{R}$. The decision equation becomes:

$$(S_1 - S_0)^T\vec{R} \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi\sigma_n^2 + \frac{1}{2}(E_1 - E_0)$$

Note that $E_k = <\vec{S_k}, \vec{S_k}> = \vec{S_k}^T\vec{S_k}$.
Therefore the projection of $\vec{R}$ into $\vec{S_1} - \vec{S_0}$ is the only information required (sufficient statistic) for decision making.

Unipolar ASK:

$$\vec{S_0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \vec{S_1} = \begin{bmatrix} A \\ A \\ \vdots \\ A \end{bmatrix}$$

To transmit 0 or 1, an ON/OFF signaling scheme is required.
Therefore, the decision rule becomes:

$$\sum_{i=1}^{N} r_i(A - 0) \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi\sigma_n^2 + \frac{1}{2}(NA^2 - 0)$$

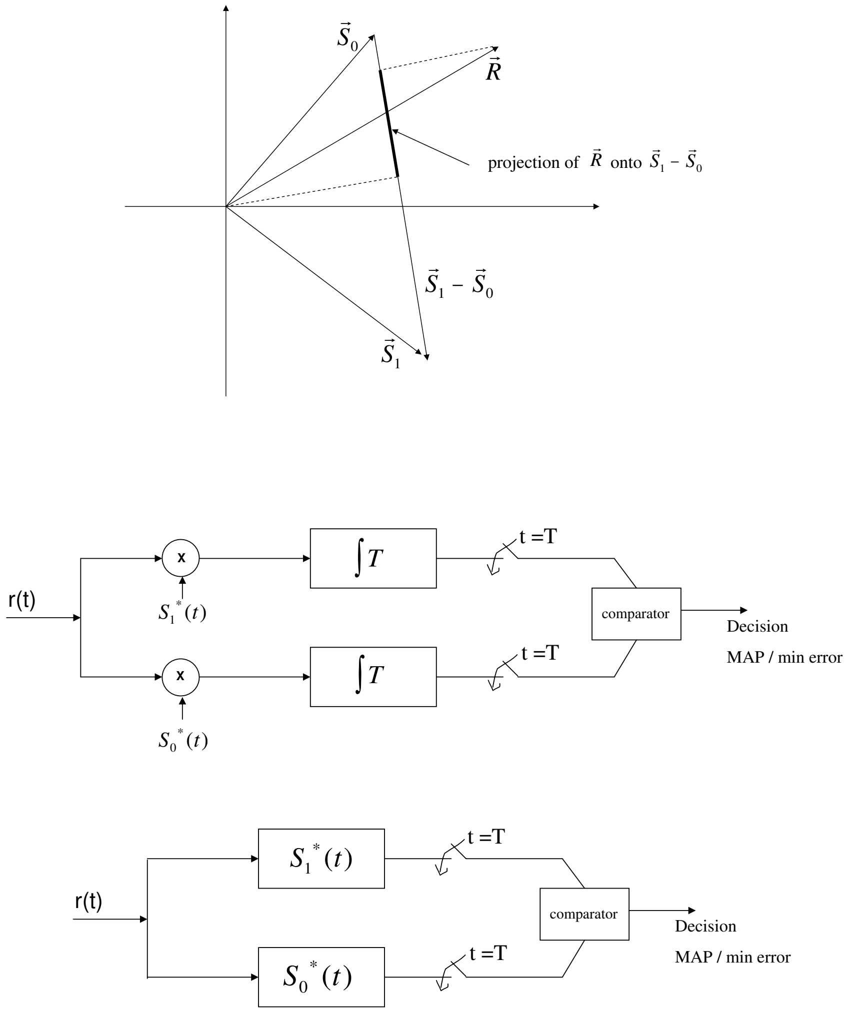

Figure 5. Continuous domain inversion: Matched filtering

Dividing both sides by $N.A$

$$\underbrace{\frac{1}{N}\sum_{i=1}^{N} r_i}_{\bar{R}\text{: avg value of } \vec{R}} \quad \gtrless^{\text{select } H_1}_{\text{select } H_0} \quad \underbrace{\frac{\xi\sigma_n^2}{NA}}_{\text{bias term}} + \frac{A}{2} \triangleq \gamma$$

$$\bar{R} \gtrless^{\text{select } H_1}_{\text{select } H_0} \gamma$$

Note that for the decision, it is sufficient to reduce the processing by using the scalar $\bar{R}$ instead of the N-dimensional $\vec{R}$.
Pdf of $\bar{R}$ is given by:

$$H_0 : \bar{R} \sim N(0, \frac{\sigma_n^2}{N}), r_i = 0 + n_i$$

$$H_1 : \bar{R} \sim N(A, \frac{\sigma_n^2}{N}), r_i = A + n_i$$

$$\Rightarrow p(\bar{R}\big|H_0) = \frac{1}{\sqrt{2\pi}\frac{\sigma_n}{\sqrt{N}}} exp\Big[- \frac{\bar{R}^2}{2\frac{\sigma_n^2}{N}} \Big]$$

$$p(\bar{R}\big|H_1) = \frac{1}{\sqrt{2\pi}\frac{\sigma_n}{\sqrt{N}}} exp\Big[-\frac{(\bar{R}-A)^2}{2\frac{\sigma_n^2}{N}}\Big]$$

<u>Special case</u>: MAP or min. probability of error criterion with $P_0 = P_1 = \frac{1}{2}$ yields:

$$\eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = 1 \Rightarrow \xi = 0$$

$$\gamma = \frac{A}{2}$$

Probability of false alarm:

$$P_F = P(error|H_0) = \int_{\gamma}^{\infty} p(\bar{R}\big|H_0)d\bar{R}$$
$$= erfc^*(\frac{\gamma}{\frac{\sigma_n}{\sqrt{N}}})$$

Probability of miss:

$$P_M = P(error|H_1) = \int_{-\infty}^{\gamma} p(\bar{R}\big|H_1)d\bar{R}$$
$$= erfc^*(\frac{(A-r)}{\frac{\sigma_n}{\sqrt{N}}})$$

Probability of detection:

$$P_D = 1 - P_M = \int_{\gamma}^{\infty} p(\bar{R}\big|H_1)d\bar{R}$$

Probability of error:

$$P_E = P_0.P(error\big|H_0) + P_1.P(error\big|H_1)$$

<u>Bipolar ASK:</u>

$$\vec{S_0} = \begin{bmatrix} -A \\ -A \\ \vdots \\ -A \end{bmatrix}, \vec{S_1} = \begin{bmatrix} A \\ A \\ \vdots \\ A \end{bmatrix}$$

$\Rightarrow E_0 = NA^2$, $E_1 = NA^2$. Therefore, the decision equation becomes:

$$\sum_{i=1}^{N} r_i[A - (-A)] \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi\sigma_n^2 + \frac{1}{2}(NA^2 - NA^2)$$

or

$$2A\sum_{i=1}^{N} r_i \gtrless^{\text{select } H_1}_{\text{select } H_0} \xi\sigma_n^2$$

or

$$\bar{R} \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{\xi\sigma_n^2}{2AN} \triangleq \gamma$$

Pdf of $\bar{R}$ is given by:

$$H_0 : \bar{R} \sim N(-A, \frac{\sigma_n^2}{N}), r_i = 0 + n_i$$

$$H_1 : \bar{R} \sim N(A, \frac{\sigma_n^2}{N}), r_i = A + n_i$$

$$\Rightarrow p(\bar{R}\big|H_k) = \frac{1}{\sqrt{2\pi}\frac{\sigma_n}{\sqrt{N}}} exp\left[-\frac{(\bar{R}-S)^2}{2\frac{\sigma_n^2}{N}}\right]$$

where

$$S = \begin{cases} -A & ; \text{under } H_0 \\ A & ; \text{under } H_1 \end{cases}$$

For $\gamma = 0$ or MAP or min probability of error with $C_{ij} = \delta_{ij}$,
$P(error|H_0) = P(error|H_1) : Bipolar < P(error|H_0) = P(error|H_1) : Unipolar$ Average transmitted energy:

$$E_b \triangleq P_0 E_0 + P_1 E_1$$

For $P_0 = P_1 = \frac{1}{2}$

$$E_b = \begin{cases} \frac{1}{2}.0 + \frac{1}{2}NA^2 = \frac{NA^2}{2} & : \text{unipolar} \\ \frac{1}{2}NA^2 + \frac{1}{2}NA^2 = NA^2 & : \text{bipolar} \end{cases}$$

Bipolar is better in probability of error performance since we use more energy in the transmission. Even if we adjust the A value to use the same average energy in both cases, still the bipolar performs better. The only scenario where the unipolar scheme is preferable are the asynchronous (non-coherent) systems.

Sufficient statistic

- In the previous section, we formulated testing of hypotheses based on an N-dimensional observed vector $\vec{R}$.
- In an example, we demonstrated how the average value of the elements of $\vec{R}$ is sufficient for the receiver to make a decision.
- We now present the general concept for what is referred to as sufficient statistic in decision theory.

Consider the transformation of $\vec{R}$ denoted by:

$$\vec{W}_{L\times 1} = T[\vec{R}_{N\times 1}]$$

where $L \leq N$; thus the transformation is not necessarily reversible. For the time being assume W to be of dimension $N \times 1$ and that the inverse of $T[.]$ exists.
We partition $\vec{W}$ into two parts:

$$\vec{W}_{N\times 1} = [\vec{W}_{1L\times 1}, \vec{W}_{2(N-L)\times 1}]$$

We can write the likelihood function via:

$$\Lambda\vec{R} \triangleq \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)}$$

We also know that if $y = g(x)$, then

$$p_Y(y) = \frac{p_X(x)}{\frac{dg(x)}{dx}}$$

If $X \rightarrow \vec{W}$ and $Y \rightarrow \vec{R}$, then

$$p(\vec{R}|H_i) = \frac{p(\vec{W}|H_i)}{J}$$

where J is the Jacobian of the transformation from $\vec{W}$ to $\vec{R}$. The Jacobian function is invariant of H_i. Substitute in the likelihood function in terms of functions of $\vec{W}$.

$$\Lambda\vec{R} = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)}$$

$$= \frac{\frac{p(\vec{W}|H_1)}{J}}{\frac{p(\vec{W}|H_0)}{J}}$$

$$= \frac{p(\vec{W}|H_1)}{p(\vec{W}|H_0)}$$

Thus the test can be performed via processing the likelihood function for $\vec{W}$. Using the Bayes theorem, we have:

$$p(\vec{W}|H_i) = p([\vec{W}_1, \vec{W}_2]|H_i)$$

$$= p(\vec{W}_1|H_i).p(\vec{W}_2|\vec{W}_1, H_i)$$

Suppose there exists a partitioning of $\vec{W}$ such that

$$p(\vec{W}_2|\vec{W}_2, H_i) = p(\vec{W}_2|\vec{W}_1)$$

i.e. it is invariant in the hypotheses. Using this in the likelihood function

$$\Lambda_{\vec{R}}(\vec{R}) = \Lambda_{\vec{W}}(\vec{W})$$

$$= \frac{p(\vec{W}_1|H_1).p(\vec{W}_2|\vec{W}_1, H_1)}{p(\vec{W}_1|H_0).p(\vec{W}_2|\vec{W}_2, H_0)}$$

$$= \frac{p(\vec{W}_1|H_1)}{p(\vec{W}_1|H_0)} \triangleq \Lambda_{\vec{W}_1}(\vec{W}_1) \gtrless^{\text{select } H_1}_{\text{select } H_0} \eta$$

i.e the test is invariant in $\vec{W}_2$. In this case $\vec{W}_1$ is called sufficient statistic to construct the test.
e.g.: Binary hypothesis testing

$$H_0 : r_i = n_i$$

$$H_1 : r_i = n_i + A$$

where $i = 1, 2, \cdots, N$ and $n_i \sim N(0, \sigma_n^2)$ and i.i.d $\forall\ i$. A is a constant. The sufficient statistic is $w_1 = \sum_{i=1}^{N} r_i$. It does not matter what $\vec{W}_2$ is.

$$\vec{W} = \begin{bmatrix} \vec{W}_1 \\ \vec{W}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{T[.]} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_n \end{bmatrix}$$

e.g.

$$H_0 : \vec{R} = \vec{S_0} + \vec{N}$$

$$H_1 : \vec{R} = \vec{S_1} + \vec{N}$$

where $\vec{N} \sim N(0, \sigma_n^2)$. S_0 and S_1 are constants. In that case, we showed that the decision is based on

$$l(R) \triangleq (S_1 - S_0)^T \vec{R} \gtrless^{\text{select } H_1}_{\text{select } H_0} \text{ threshold}$$

The threshold is a scalar that is sufficient statistic for this detection problem.