Lecture 13: Detection of (Purely) Stochastic Signals in Noise

23 Feb 2016

The problem of "detection of purely stochastic signals in noise" is frequently seen in applications like Radio Astronomy, Sonar etc., where we deal with highly variable and turbulent signals.

More generally, we have a hypothesis testing problem (in Gaussian setting) which looks like:

$$\mathcal{H}_0 : Y \sim \mathcal{N}(\mu_0, \Sigma_0), \tag{1}$$
$$\mathcal{H}_1 : Y \sim \mathcal{N}(\mu_1, \Sigma_1). \tag{2}$$

The simplest form of detection in a purely stochastic signal case is given by:

$$\mathcal{H}_0 : Y = N \in \mathbb{R}^n, \tag{3}$$
$$versus \tag{4}$$
$$\mathcal{H}_1 : Y = S + N \in \mathbb{R}^n, \tag{5}$$

where,

$$S \sim \mathcal{N}(0, \Sigma_s), \tag{6}$$
$$N \sim \mathcal{N}(0, \sigma^2 I). \tag{7}$$

Above mentioned case is a special case of the general setting and also, $S \amalg N$ (S is independent of N). Going forward, we would like to do the likelihood ratio test, i.e., finding the log likelihood ratio (for the general case), given by:

$$\log L(y) = \frac{1}{2} \log \frac{|\Sigma_0|}{|\Sigma_1|} + \frac{1}{2}(y - \mu_0)^T \Sigma_0^{-1}(y - \mu_0) - \frac{1}{2}(y - \mu_1)^T \Sigma_1^{-1}(y - \mu_1). \tag{8}$$

This equation is quadratic in y. Rewriting the above equation, we get

$$\log L(y) = \frac{1}{2} y^T [\Sigma_0^{-1} - \Sigma_1^{-1}] y + [\mu_1^T \Sigma_1^{-1} - \mu_0^T \Sigma_0^{-1}] y + c, \tag{9}$$

where,

$$c := \frac{1}{2}\log\frac{|\Sigma_0|}{|\Sigma_1|} + \mu_0^T\Sigma_0^{-1}\mu_0 - \mu_1^T\Sigma_1^{-1}\mu_1.$$

In the section below, we consider two special cases.

Special Cases:

1. Same Covariances: $\Sigma_0 = \Sigma_1 = \Sigma$.

   In such a case, we observe that the quadratic terms vanish and only the linear terms remain. It is like passing the output through a linear filter and then thresholding it.

   The Test Statistic in this case is given by $T(y) = (\mu_1 - \mu_0)^T\Sigma^{-1}y$.

2. Same Means: $\mu_0 = \mu_1 = 0$.

   In this case, we are only left with the quadratic terms. The Test Statistic in this case is given by $T(y) = y^T(\Sigma_0^{-1} - \Sigma_1^{-1})y$. Going back to our initial problem of detection in the setting of eqn. (6), assuming Σ_s is invertible, we can map Σ_0 and Σ_1 as follows:

$$\begin{aligned}\Sigma_0 &= \sigma^2 I,\\ \Sigma_1 &= \sigma^2 I + \Sigma_s.\end{aligned}$$

   Working further, we end up getting a detector (in the likelihood ratio form) given by,

$$\tilde{\delta}_0(y) = \begin{cases} 1, & y^TQy > \tau' \\ \gamma, & y^TQy = \tau' \\ 0, & y^TQy < \tau' \end{cases} \qquad \text{(where } \tau' = 2(\log\tau - C))$$

   Where,

$$\begin{aligned} Q &= \Sigma_0^{-1} - \Sigma_1^{-1},\\ &= \sigma^{-2}I(\sigma^2 I + \Sigma_s)(\sigma^2 I + \Sigma_s)^{-1} - (\sigma^2 I + \Sigma_s)^{-1},\\ &= (\sigma^{-2}\Sigma_s)(\sigma^2 I + \Sigma_s)^{-1}. \end{aligned} \tag{10}$$

   Remark 1. This Q is called as Quadratic detector.

   Remark 2. This is also called as an energy detector, because it looks like we are getting a term, which looks like the norm squared of y, due to the presence of the quadratic terms.

   Remark 3. This finds application in the field of radio astronomy too. When a radio system is connected with such a detector, it is called a "Radio Meter".

1 Performance analysis of the detector

To analyze the performance of the detector, let us consider the Neyman-Pearson hypothesis testing of the Quadratic detector. We need to calculate $P_j(y^TQY > \tau'),\ \forall j \in 0, 1$.

Assuming Σ_s is a Hermitian Matrix, we can decompose Σ_s into orthonormal eigen vectors using the Singular Value Decomposition (SVD),

$$\Sigma_s = \sum_{k=1}^{n} \lambda_k V_k V_k^T, \tag{11}$$

where $V_1, V_2, ..., V_n$ are orthonormal. Since they are orthonormal, we can write,

$$I = \sum_{k=1}^{n} V_k V_k^T.$$

Now, we can write,

$$(\sigma^2 I + \Sigma_s)^{-1} = \sum_{K=1}^{n} \frac{1}{\sigma^2 + \lambda_k} V_k V_k^T.$$

Therefore, eqn. (10) becomes,

$$Q = \sum_{k=1}^{n} \frac{\lambda_k}{\sigma^2(\sigma^2 + \lambda_k)} V_k V_k^T. \tag{12}$$

As mentioned earlier, this is called an energy detector because,

$$y^T Q y = \sum_{k=1}^{n} y_k^2,$$

where, $y_k = \sqrt{\frac{\lambda_k}{\sigma^2(\sigma^2+\lambda_k)}} V_k^T y$, i.e., find energy of y along all directions $\{V_k\}$.

Note 1. When $\lambda_k \gg \sigma^2$ i.e., when $\sqrt{\frac{\lambda_k}{\sigma^2(\sigma^2+\lambda_k)}} V_k^T y$ increases, it implies that the contribution from the particular direction V_k increases, which in turn indicates that more weightage is given to the signal component along V_k. Note that $\lambda_k \gg \sigma^2$ means that signal power along V_k is more compared to the noise power.

Note 2. $y_k,\ k = 1, 2, \ldots, n$ are independent (because they are projected in orthonormal directions), zero mean Gaussian random variables under both the hypotheses H_0 and H_1.

$$\sigma_{jk}^2 = Var\left(y_k|H_j\right) = \begin{cases} \frac{\lambda_k}{\lambda_k+\sigma^2}, & j = 0, \\ \frac{\lambda_k}{\sigma^2}, & j = 1. \end{cases}$$

Now, the test statistic is a random variable which is sum of squared Gaussians, pdf of $T_k(y) = y_k^2$ is given by,

$$P_{T_k}\left(t|H_j\right) = \begin{cases} \frac{1}{\sigma_{jk}\sqrt{2\pi t}} e^{-\frac{t}{2\sigma_{jk}^2}}, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{13}$$

This is a Gamma distribution with parameters $\left(\frac{1}{2}, \frac{1}{2\sum_{jk}^2}\right)$.

Definition 1.1. We define the characteristic function and the Fourier Transform of the Characteristic function as follows:

$$\Phi_x(t) = E[e^{jtx}],$$

$$f_X(x) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-jut}\Phi_x(t)dt.$$

Using the above definitions, we can write the pdf of $T = \Sigma_{k=1}^n T_k$ as,

$$P_T = P_{T_1} * P_{T_2} * P_{T_3} * .... * P_{T_k},$$

Using Fourier transform relations,

$$P_T(t|H_j) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jut}\prod_{k=1}^{n}\left(1 - 2ju\sigma_{jk}^2\right)^{\frac{1}{2}}\ du \tag{14}$$

This is a method of finding $P_T(t|H_j)$ in the Fourier Domain. We may not be able to solve this every time, but for a special case, where

$$\sigma_{j1} = \sigma_{j2} = \sigma_{j3} = ... = \sigma_{jn} = \sigma_j,\ \ j \in \{0, 1\}, \tag{15}$$

we can find this quantity,

$$P_T(t|H_j) = \begin{cases} \frac{1}{(2\sigma_j^2)^{\frac{n}{2}}}\frac{1}{\Gamma(\frac{n}{2})} t^{(\frac{n}{2}-1)} e^{\frac{-t}{2\sigma_j^2}}, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{16}$$

Therefore, $P_j(Y^TQY > \tau')$= 1 - $\Gamma\left(\frac{n}{2}, \frac{\tau'}{2\sigma_j^2}\right)$, where, $\Gamma\left(\frac{n}{2}, \frac{\tau'}{2\sigma_j^2}\right)$ is the CDF of Gamma distribution with parameters $(\frac{n}{2}, \frac{1}{2\sigma_j^2})$ evaluated at τ'. This is also called the incomplete Gamma function at τ'.

Example 1.2. For Neyman-Pearson case with $P_{F.A} = \alpha$, we choose $\tau' = 2\sigma_0^2 \Gamma^{-1}\left(\frac{n}{2}, 1-\alpha\right)$. Correspondingly,

$$P_D(\tilde{\delta}) = 1 - \Gamma\left(\frac{n}{2}, \frac{\sigma_0^2}{\sigma_1^2}\Gamma^{-1}\left(\frac{n}{2}, 1-\alpha\right)\right).$$

Note 3. Detection Performance for fixed α is governed by, (i) the number of samples n, the more the number of samples we can generate, the more energy we have and better performance, and (ii) we have the ratio $\frac{\sigma_0^2}{\sigma_1^2}$ which governs the performance of the detector.

Summarizing what we discussed in the lecture, we have the hypothesis structure given by,

$$\begin{aligned} \mathcal{H}_0 &: Y = N \in \mathbb{R}^n, \\ versus& \\ \mathcal{H}_1 &: Y = S + N \in \mathbb{R}^n, \end{aligned}$$

where,

$$\begin{aligned} S &\sim \mathcal{N}(0, \Sigma_s), \\ N &\sim \mathcal{N}(0, \sigma^2 I), \end{aligned}$$

and $S \amalg N$. The optimum detector is given by,

$$y^T Q Y > \tau',$$

where, $Q = (\sigma^{-2}\Sigma_s)(\sigma^2 I + \Sigma_s)^{-1}$. So looking back, there are a couple of ways we can interpret the Detector performance.

1. Interpretation 1: (Weighted Energy Detector): The detector structure is shown in fig. 1.

2. Interpretation 2: "Estimator-Correlator": The quantity $\tilde{s} = \Sigma_s(\sigma^2 I + \Sigma_s)^{-1} y$ is called a "Wiener filter". This is a very good estimator of the actual signal.

   Note 4. This quantity is also the minimum mean squared error estimator of S given $Y = S + N$.

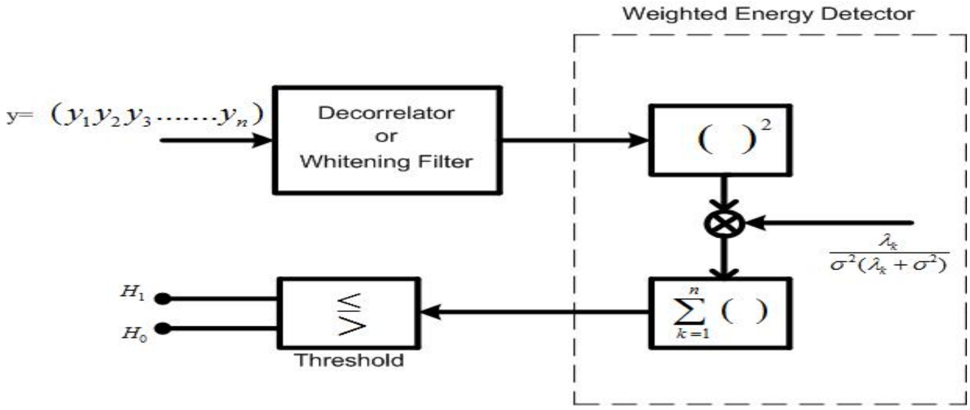

Figure 1: Optimum "weighted energy detector"

Figure 1: Weighted energy detector

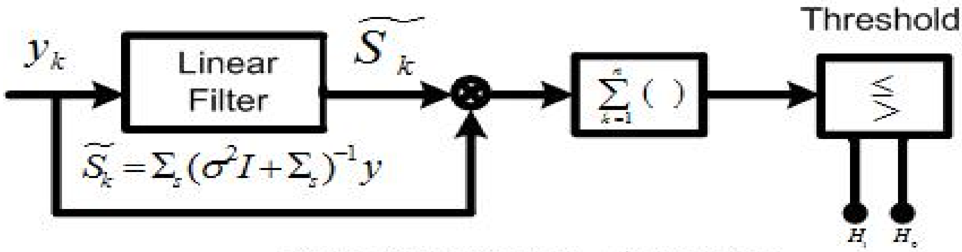

Figure 2: Estimator - Correlator.

2 Locally optimum detection of Stochastic Signals:

Consider the case, similar to a Composite Hypothesis Test:

$$\mathcal{H}_0 : Y = N \in \mathbb{R}^n,$$

$$versus$$

$$\mathcal{H}_1 : Y = N + \sqrt{Q}S \in \mathbb{R}^n, \; Q > 0,$$

where,

$$N \sim \mathcal{N}(0, \sigma^2 I),$$

$$S \sim \mathcal{N}(0, \Sigma_s).$$

The above situation can be looked as a case, where a signal is coming through a channel, which is blind to us, in terms of attenuation etc.

For a fixed $Q > 0$ (i.e., $\mathcal{H}_1$) *versus* $Q = 0$ (i.e., $\mathcal{H}_0$), the detection statistic is, $T(y) = y^T Q\Sigma_s(I + Q\Sigma_s)^{-1}y$. As we see the statistic is dependent on Q. We see that $P(y : y^T\Sigma_s(I + Q\Sigma_s)^{-1}y > \tau') = \alpha$ is dependent on Q and hence no Universally Most Powerful(UMP) test exists (i.e., one single test does not exist) $\forall Q > 0$.

For finding the Locally most powerful test(LMP), let us consider the case of $Q = 0$ *versus* $Q \neq 0$ (i.e., approximately $Q = 0$). For finding the LMP, we differentiate $y^T\Sigma_s(I + Q\Sigma_s)^{-1}y$ and set $Q = 0$. By doing so, we get the LMP test statistic as $T(y) = 2y^T\Sigma_s y$ and we use this quantity to compare with the threshold.

Consider a scaled version of the statistic i.e., $T(y) = \frac{1}{n}y^T\Sigma_s y$. Suppose Σ_s is a matrix such that its $(k, l)th$ element $\rho_{k,l}$ depends only on $k - l$. If such a condition occurs, this signal is said to be taken from a wide sense stationary (WSS) process i.e.,

$$\rho_{k,l} \triangleq \rho_{k-l}.$$

Using the above WSS structure to the white noise that we earlier considered, we can write

$$\begin{aligned} T(y) &= \frac{1}{n}y^T\Sigma_s y \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} y_k y_l \rho_{k-l} \\ &= \rho_0\hat{\rho}_0 + 2\sum_{k=1}^{n-1}\rho_k\hat{\rho}_k. \end{aligned}$$

where, $\hat{\rho}_k :\, = \frac{1}{n-k}\sum_{l=1}^{n-k} y_l y_{l+k},\ \forall k \in [n]$, which the sample estimate of correlation at lag k. For $n \gg k$, $\hat{\rho}_k$ turns out to be a good estimate of $E[Y_l Y_{l+k}]$.

From the above equations, we can see that $T(Y)$ estimates the covariance structure of the signal y and correlates it with signal covariance $\{\rho_k\}_{k=1}^n$. Roughly, if $\hat{\rho}_k$ is "accurate", then,

$$T(y) = \begin{cases} 1, \text{ under } H_0 \\ \rho_0 + Q.\left(\rho_0^2 + 2\sum\limits_{k=1}^{n-1}\rho_k^2\right), \text{ under } H_1, \end{cases}$$

where $\hat{\rho}_k = E[Y_l Y_{l+k}]$. Since we are dealing with autocorrelation functions and could end up getting equations which involve Convolutions, we choose an intelligent method of going to the frequency domain, by taking the Fourier transform of the autocorrelation function which gives us the power spectral density(PSD).

$$T(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(\omega)\hat{\phi}(\omega)d\omega,$$

where $\phi(\omega)$ is the PSD of S (signal) at ω, and,

$$\hat{\phi}(\omega) = \frac{1}{n}\left|\sum_{k=1}^{n} y_k e^{j\omega k}\right|^2.$$

This computes the spectrum of our signal and correlates with the spectrum of the desired signal.