EE224A Fall 2009: Homework #3

Assigned: 09/22/2009 --- Due: 09/29/2009 in class *before* lecture

Problem 1

Let $\mathbf{u} = [u_1 u_2 u_3]^T$ and $\mathbf{v} = A[\mathbf{u}]$ where A is some linear transformation. Let the auto-correlation matrix of $\mathbf{u}$ be:

$$R = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

where $\rho = 0.9$.

a. Find the KLT of $\mathbf{u}$ (i.e. A is the KLT).

b. Given an average of R bits/sample, find the distortion attained by a Shannon Rate-Distortion quantization of $\mathbf{v} = KLT[\mathbf{u}]$. Assume $\mathbf{u}$ to be jointly Gaussian.

c. Compare the performance with that of $A = I$.

d. Evaluate (b) and (c) for $R = 1$ bits/sample.

Problem 2

Suppose that we want to send at the same time two independent binary symbols $X_1 \in \{-1, 1\}$ and $X_2 \in \{-1, 1\}$ over an additive Gaussian channel, where -1 and $+1$ are equally likely. A clever communication engineer proposes a transmission scheme that uses twice the channel to send at each time a linear combination of X_1 and X_2. The channel input output relation is given by:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ -h_2 & h_1 \end{pmatrix} * \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where Y_1 and Y_2 are the channel outputs, the $h_i, i = 1, 2$ are known real constant, and $Z_i, i = 1, 2$ are iid standard normal random variables.

We are interested in detecting both X_1 and X_2 (i.e. the vector $(X_1, X_2)^T$).

(a) Show that the vector detection problem can be decomposed into two scalar detection problems.

Hint: observe that the rows of the matrix H are orthogonal.

(b) Give the decision rule for detecting X_1 and write the error probability using the Q-function.

Problem 3

Consider a modified version of Example 3.2 of the course note (Gallager) where the source output is no longer 0 or 1 but takes values in the set $S = \{1, 2, \ldots, I\}$. If the source output is $i \in S$, the modulator produces the real vector $\vec{a}_i = (a_{i1}, \ldots, a_{in})$. Given that the source output is i the observation is $\vec{Y} = \vec{a}_i + \vec{Z}$ where the noise is assumed to be $\mathcal{N}(0, I_n)$.

(a) Show that the ML decision rule is given by

$$\hat{i} = argmin_i(\| \vec{Y} - \vec{a}_i \|^2)$$

where $\| \cdot \|$ is the Euclidean distance in R^n.
(b) For $n = 2$, use a one dimensional sufficient statistics to compute the decision rule.
(c) Compare your decision rule to the one in (a). Explain.

Problem 4

Consider sending a single bit using repetition coding over an AGN channel over two time instants. There is an energy constraint of E Joules in each time instant. The Gaussian noise is independent from time to time but the variance is different: say σ_1^2 at time 1 and σ_2^2 at time 2.

(a) Derive the ML rule based on the two received voltages y_1 and y_2. Does the rule make intuitive sense?

(b) Compute the average error probability of the ML rule.

Problem 5

Consider detecting whether voltages A_1 or A_2 was sent over a channel with independent Gaussian noise. The Gaussian noise has mean A_3 volts and variance σ^2. The voltages A_1 and A_2 are equally likely (i.e., they occur with probability 0.5 each).

(a) What decision rule would you use to minimize the average error probability? You can use any result derived in class or in the textbook but it needs to be stated precisely and connected to your derivation.

(b) What is the average error probability for the decision rule you derived in the previous part? Use the Q function notation.

(c) If the transmit voltage is A_1 then the average transmit energy is proportional to A_1^2. If the voltage is held constant for 1 second, then the energy is exactly A_1^2. Calculate the *average* transmit energy (the transmit energy is random, because the transmit voltage can be either A_1 or A_2) in this set up.

(d) Now suppose you have a constraint that the average transmit energy cannot be more than E. How would you pick the voltage levels A_1 and A_2 to meet the average energy constraint and still arrive at the smallest error probability using the decision rule derived earlier?

Appendix (Gallager course notes Example 3.2

Example 3.2 Now we look at the vector version of Figure 3.1. If the source output is 0, (H=0), the modulator produces the real vector $\vec{a} = (a_1, \ldots, a_n)^T$. If the source output is 1, the real vector $b = (b_1, ..., b_n)^T$ is produced. $Z = (Z_1, ..., Z_n)^T$ is a noise r$\vec{v}$ assumed to be $\mathcal{N}(0, \sigma^2 I)$. That is, $Z_1, \ldots Z_n$ are IID Gaussian rv's, also independent of H. The observation r$\vec{v}$ $\vec{Y}$ is $\vec{b} + \vec{Z}$ or $\vec{a} + \vec{Z}$, depending on whether $H = 1$ or 0. Thus, given H=1, $\vec{Y} \sim \mathcal{N}(\vec{b}, \sigma^2 I_n)$, so that

$$p_{\vec{Y}|H}(\vec{y} \mid 1) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \sum_{k=1}^{n} \frac{-(y_k - b_k)^2}{2\sigma^2} \tag{3.22}$$

)

Similarly, given $H = 0$, $\vec{Y} \sim \mathcal{N}(\vec{a}, \sigma^2 I)$, so that

$$p_{\vec{Y}|H}(\vec{y} \mid 0) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \sum_{k=1}^{n} \frac{-(y_k - a_k)^2}{2\sigma^2} \tag{3.23}$$

The likelihood ratio is then given by

$$\begin{aligned} \Lambda(\vec{y}) &= \exp \sum_{k=1}^{n} \frac{(y_k - a_k)^2 - (y_k - b_k)^2}{2\sigma^2} && (3.24) \\ &= \exp \sum_{k=1}^{n} \frac{2(b_k - a_k)y_k + (a_k^2 - b_k^2)}{2\sigma^2} \\ &= \exp \left[\frac{(\vec{b} - \vec{a})^T \vec{y}}{\sigma^2} + \frac{\vec{a}^T\vec{a} - \vec{b}^T\vec{b}}{2\sigma^2} \right] && (3.25) \end{aligned}$$

Substituting this into (3.6) and taking the logarithm of both sides,

$$\mathrm{LLR}(\vec{y}) = \frac{(\vec{b} - \vec{a})^T \vec{y}}{\sigma^2} + \frac{\vec{a}^T\vec{a} - \vec{b}^T\vec{b}}{2\sigma^2} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\gtrless}} \ln \frac{P_0}{P_1} = \ln(\eta) \tag{3.26}$$

$$\frac{(b-a)^T}{\sigma^2}\left(y - \frac{(b+a)}{2}\right) \gtrless \ln \eta$$

It can be seen that the test involves the observation $\vec{y}$ only in terms of the inner product $(\vec{b} - \vec{a})^T \vec{y}$, so we can rewrite (3.26) in the form

$$(\vec{b} - \vec{a})^T \vec{y} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\gtrless}} \sigma^2 \ln(\eta) + \frac{(\vec{b}^T\vec{b} - \vec{a}^T\vec{a})}{2} = \phi \tag{3.27}$$

These equations are interpreted in Figure 3.3. Contours of equal probability density for $p_{\vec{Y}|H}(\vec{y} \mid 0)$ are concentric spherical shells centered at $\vec{a}$, whereas contours of equal probability density for $p_{\vec{Y}|H}(\vec{y} \mid 1)$ are concentric spherical shells centered at $\vec{b}$. As can be seen from (3.24), the locus of points of constant likelihood ratio are points $\vec{y}$ for which the squared distance to $\vec{a}$, less the squared distance to $\vec{b}$, is a constant. This set of points forms a straight line for the two dimensional case shown in Figure 3.3. In general, as seen analytically by (3.25), points of constant likelihood ratio are points for which $(\vec{b} - \vec{a})^T \vec{y}$ is constant, and this is the equation of an affine space.[1]

We have seen from (3.27) that comparing $\Lambda(\vec{y})$ to the threshold η is equivalent to comparing $(\vec{b} - \vec{a})^T \vec{y}$ to the threshold ϕ. Thus the affine space $(\vec{b} - \vec{a})^T \vec{y} = \phi$ separates the observation space into two regions, where $\hat{H} = 1$ for $(\vec{b} - \vec{a})^T \vec{y} \geq \phi$ and $\hat{H} = 0$ otherwise.

[1] In linear algebra, an $n-1$ dimensional hyperplane in n dimensional space is by definition a linear space in its own right; such a space must include the origin, and be spanned by $n-1$ vectors. The translation of a hyperplane away from the origin is called an *affine space*. For the case here, points $\vec{y}$ for which $(\vec{b} - \vec{a})^T \vec{y} = 0$ form a hyperplane of points perpendicular to $(\vec{b} - \vec{a})$. The set of points for which $(\vec{b} - \vec{a})^T \vec{y} = c$, for some constant c, is thus an affine space.

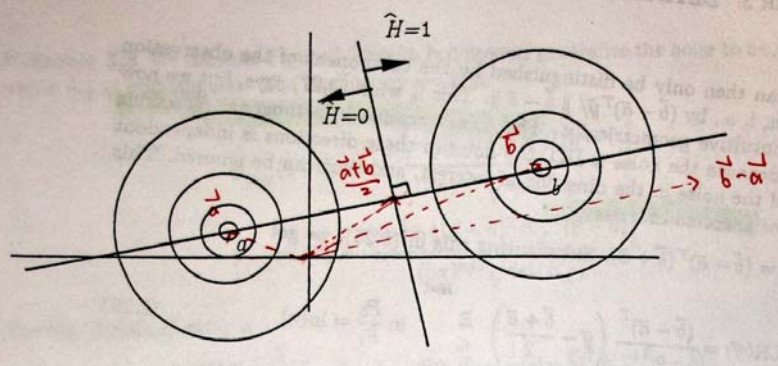

Figure 3.3: Decision regions for binary signals in additive IID Gaussian noise.

We also see from (3.26) that $\Lambda(\vec{y})$ can be calculated from $(\vec{b}-\vec{a})^T\vec{y}$, so that $(\vec{b}-\vec{a})^T\vec{y}$ is a sufficient statistic. This says that the threshold test for this problem is based on the value of a single number which is simply a linear combination of the observed variables. Note that each observation y_k is weighted by $(b_k - a_k)$ in forming the sufficient statistic. This makes sense intuitively, since if $b_k - a_k$ is very small, the observation y_k is mostly noise, whereas if $b_k - a_k$ is large, the observation gives a much better indication of which hypothesis is correct.

We can view $(\vec{b}-\vec{a})^T\vec{y}$ as the correlation[2] between $\vec{b}-\vec{a}$ and the observation $\vec{y}$. Thus a threshold detector, for this additive Gaussian noise case, is often called a *correlation detector* in communication theory. Often $\vec{b}$ and $\vec{a}$ are separately correlated with $\vec{y}$ and the results compared; this is also called a correlation detector.

If we view $\vec{y}$ as a discrete time sequence $y_1, \ldots, y_n$, then we can also visualize performing this correlation function by convolving $y_1, \ldots, y_n$ with $(b_n - a_n), (b_{n-1} - a_{n-1}), \ldots, (b_1 - a_1)$. This is the output, at the appropriate sampling time, of a digital filter with the impulse response $(b_n - a_n), \ldots, (b_1 - a_1)$. A filter with this impulse response is said to be a *matched filter* to $(b_1 - a_1), \ldots, (b_n - a_n)$. We will look at correlation detectors and matched filters again later when we consider detection of waveforms. The important point to note here, however, is that both the correlation detector and the matched filter simply compute the inner product $(\vec{b}-\vec{a})^T\vec{y}$.

Another way of viewing (3.27), and the most fundamental, is to view it in a different co-ordinate basis. That is, view the observation $\vec{y}$ as a point in n dimensional space represented in a particular co-ordinate system. Consider a different orthonormal basis where one of the basis elements is $(\vec{b}-\vec{a})/\parallel \vec{b}-\vec{a} \parallel$, where $\parallel \vec{b}-\vec{a} \parallel$ is the length of $\vec{b}-\vec{a}$,

$$\parallel \vec{b}-\vec{a} \parallel = \sqrt{(\vec{b}-\vec{a})^T(\vec{b}-\vec{a})}, \tag{3.28}$$

Thus $(\vec{b}-a)/\parallel \vec{b}-\vec{a} \parallel$ is the vector $\vec{b}-\vec{a}$ normalized to unit length.

[2]For the moment, we ignore any similarity between this use of the word correlation as an inner product and the use of correlation as an expectation between random variables; we discuss this similarity later.

The two hypotheses can then only be distinguished by the component of the observation vector in this direction, i. e., by $(\vec{b}-\vec{a})^T\vec{y}/\parallel \vec{b}-\vec{a}\parallel$. This is what (3.27) says, but we now see that this is very intuitive geometrically. The measurements in orthogonal directions only measure noise. Because the noise is IID, the noise in these directions is independent of both the signal and the noise in the direction of interest, and thus can be ignored. This is sometimes called the theorem of irrelevance.

Note that $\vec{b}^T\vec{b}-\vec{a}^T\vec{a}=(\vec{b}-\vec{a})^T(\vec{b}+\vec{a})$. Substituting this in (3.26), we get

$$\mathrm{LLR}(\vec{y})=\frac{(\vec{b}-\vec{a})^T}{\sigma^2}\left(\vec{y}-\frac{\vec{b}+\vec{a}}{2}\right)\underset{\hat{H}=0}{\overset{\hat{H}=1}{\gtrless}}\ln\frac{P_0}{P_1}=\ln(\eta) \tag{3.29}$$

This says that for ML detection, where $\ln\eta=0$, the decision regions are separated by the affine space that forms the perpendicular bisector between $\vec{a}$ and $\vec{b}$.

Finally, we use (3.29) to evaluate $\Pr(e\mid H{=}0)$. $\mathrm{E}[\vec{Y}-(\vec{b}+\vec{a})/2\mid H{=}0)]=(\vec{a}-\vec{b})/2$, so

$$\mathrm{E}[\mathrm{LLR}(\vec{Y})\mid H{=}0]=\frac{-(\vec{b}-\vec{a})^T(\vec{b}-\vec{a})}{2\sigma^2}$$

Defining γ as

$$\gamma=\frac{||\vec{b}-\vec{a}||}{\sigma} \tag{3.30}$$

this simplifies to $\mathrm{E}\left[\mathrm{LLR}(\vec{Y})\,|\,H{=}0\right]=-\gamma^2/2$. Similarly, we see that $\mathrm{VAR}\left[\mathrm{LLR}(\vec{Y})\,|\,H{=}0\right]=\gamma^2$. Thus, conditional on $H=0$, $\mathrm{LLR}(\vec{Y})\sim\mathcal{N}(-\gamma^2/2,\gamma^2)$. The probability of error can then be found (see Exercise 3.1 as

$$\Pr(e\mid H{=}0)=\Pr\left[\mathrm{LLR}(\vec{Y})\geq\ln(\eta)\ \mid H{=}0\right]=Q\left(\frac{\ln(\eta)}{\gamma}+\frac{\gamma}{2}\right) \tag{3.31}$$

Analyzing $\mathrm{LLR}(\vec{Y})$ conditional on $H=1$ in the same way, we find that, conditional on $H=1$, $\mathrm{LLR}(\vec{Y})\sim\mathcal{N}(\gamma^2/2,\gamma^2)$, and it follows that

$$\Pr(e\mid H=1)=Q\left(\frac{-\ln(\eta)}{\gamma}+\frac{\gamma}{2}\right) \tag{3.32}$$

Note that both error probabilities are functions only of $\gamma=\parallel\vec{b}-\vec{a}\parallel/\sigma$. This is not surprising in terms of our geometric interpretation. $\parallel\vec{b}-\vec{a}\parallel$ is the distance from $\vec{a}$ to $\vec{b}$, and this is normalized by the standard deviation of the noise. That is, if we measure both $\parallel\vec{b}-\vec{a}\parallel$ and σ in some other units, the error probability cannot change. We can interpret $\parallel\vec{b}-\vec{a}\parallel^2$ as the energy in the difference between the signals. We can also interpret σ^2 as the energy per measurement of the noise. This says that what is relevant is not the number of different measurement values (i. e., n), but rather the total signal difference energy used over the set of measurements. With IID Gaussian noise, this signal difference energy can be split up in any way without affecting the error probability. This is why the signal to noise ratio is such an important parameter in digital communication.