EE 121: Digital Communications

April 22, 2008

Matched Filters

Introduction

Starting from this lecture, we focus on how to communicate over LTI channels. This lecture we focus on matched filtering. This strategy aims to mitigate ISI by *filtering* the received signal.

The Matched Filter

To motivate this linear processor, let us start with a very simple sequential communication scheme: we send independent information bits at different time instants, but *interleave* $L-1$ zeros between every information bit. In other words, we send information only once in L time samples. Of course, such a scheme entails a very low communication rate but it has the major advantage of ensuring that there is *no* ISI at the receiver. Denoting the voltages chosen by the sequential communication scheme by $d[n], n \geq 1$, the transmit voltages are

$$x[(n-1)L + \ell + 1] = \begin{cases} d[n] & \text{if } \ell = 0 \\ 0 & \text{if } \ell = 1 \dots L-1 \end{cases}, \quad n \geq 1. \tag{1}$$

Now the received voltages are

$$y[(n-1)L + \ell + 1] = h_\ell d[n] + w[(n-1)L + \ell + 1], \quad n \geq 1, \ell = 0, \dots, L-1. \tag{2}$$

So, to recover the information bits that generated the n^{th} data voltage $d[n]$ we only need to focus on L receive voltages:

$$y[(n-1)L + \ell + 1], \quad \ell = 0, \dots, L-1. \tag{3}$$

Each of these voltages is simply the (scaled) voltage of interest ($d[n]$) corrupted by additive white Gaussian noise. This situation is reminiscent of repetition coding over the AWGN channel.

We have derived the vector ML rule in an earlier lecture to be a nearest-neighbor rule where the distance is now a Euclidean one. If $d[n]$ is equal to the k^{th} possible voltage level (among the 2^B choices, if we are sending B information bits in each of the transmit voltages), say v_k, then the squared Euclidean distance between the received voltages and the vector

$$\begin{bmatrix} h_0 v_k \\ h_1 v_k \\ \vdots \\ h_{L-1} v_k \end{bmatrix} \tag{4}$$

is

$$\sum_{\ell=0}^{L-1} \left(y[(n-1)L + \ell] - h_\ell v_k \right)^2 \tag{5}$$

which can be expanded out as

$$\left(\sum_{\ell=0}^{L-1} h_\ell^2 v_k^2\right) + \left(\sum_{\ell=0}^{L-1} (y[(n-1)L+\ell])^2\right) - 2v_k \left(\sum_{\ell=0}^{L-1} h_\ell y[(n-1)L+\ell]\right). \tag{6}$$

In comparing the Euclidean distances for different choices of k (among the 2^B different choices), we only need to keep track of the *sufficient statistic*:

$$\sum_{\ell=0}^{L-1} h_\ell y[(n-1)L+\ell], \tag{7}$$

the *weighted* sum of the received voltages. Indeed, when all the filter coefficients are equal, this sufficient statistic is simply the sum of the received voltages and featured in the ML rule for repetition coding. This operation of taking a weighted sum of the received voltages, with the weights in *proportion* to the channel coefficients is known as *match filtering*; the idea is that the filter weights are *matched* to those of the channel. The output of the matched filter

$$y^{\mathrm{MF}}[n] \stackrel{\mathrm{def}}{=} \sum_{\ell=0}^{L-1} h_\ell y[(n-1)L+\ell] \tag{8}$$

$$= \left(\sum_{\ell=0}^{L-1} h_l^2\right) d[n] + \left(\sum_{\ell=0}^{L-1} h_\ell w[(n-1)L+\ell]\right). \tag{9}$$

Thus we have a plain AWGN channel between the voltage of interest $d[n]$ and the output of the matched filter $y^{\mathrm{MF}}[n]$. We could now employ the nearest-neighbor rule on the output of the matched filter (MF) to detect the information bits that generated the voltage $d[n]$. The operating SNR of this AWGN channel is

$$\mathsf{SNR}_{\mathrm{MF}} \stackrel{\mathrm{def}}{=} \left(\sum_{\ell=0}^{L-1} h_\ell^2\right) \mathsf{SNR} \tag{10}$$

where we have denoted the ratio of average transmit energy (say, E) and the noise variance σ^2) by SNR. The effect of the matched filter is to *boost* the operating SNR level by carefully harnessing the delayed copies of the transmit symbols. This SNR level directly impacts the reliability of communication; indeed, with the voltage $d[n]$ constrained to be within $\pm\sqrt{E}$ Volts and there are 2^B possible levels, the average probability of error is (from an earlier lecture)

$$\left(2 - \frac{1}{2^{B-1}}\right) Q\left(\frac{\sqrt{\mathsf{SNR}_{\mathrm{MF}}}}{2^B - 1}\right). \tag{11}$$

Note that the operation (8) is called matched filtering because it can actually be rewritten as a convolution operation followed by sampling every L time instants:

$$y^{\mathrm{MF}}[n] = \sum_{\ell=0}^{-(L-1)} h_{-\ell} y[(n-1)L-\ell] = (\tilde{h} * y)[(n-1)L]$$

where

$$\tilde{h}_\ell = h_{-\ell},$$

i.e. the channel in reverse.

Throughout this discussion, we have not had to worry about ISI: the transmission interleaved with zeros so that there is no interference between consecutive transmit voltages. Now to use sequential communication at all each time sample (and not interleave with $L-1$ zeros) while maintaining the simplicity of the matched filter receiver, the key conceptual idea is the following: even if there were interference but it was negligible compared to the additive Gaussian noise then for practical purposes it is as if it does not exist. In this situation, the MF would be quite a sensible strategy to pursue at the receiver. With sequential communication, the average energy in the interference is

$$E\left(\sum_{\ell=1}^{L-1} h_\ell^2\right) \tag{12}$$

where we have denoted the average transmit energy by E. Whenever this is small compared to the noise energy σ^2, i.e., the operating SNR

$$\mathsf{SNR} \ll \frac{1}{\sum_{\ell=1}^{L-1} h_\ell^2}, \tag{13}$$

it is reasonable to suspect that the interference can be ignored without much loss of performance. In such instances, the matched filter is quite likely to yield near-optimal performance.

Concretely, suppose the transmit voltages are generated based on sequential communication (no interleaving with zeros this time). Now the matched filter operates at every time sample and the corresponding output is

$$\begin{aligned}
y^{\mathrm{MF}}[m] &\stackrel{\mathrm{def}}{=} \sum_{\ell=0}^{L-1} h_\ell y[m+\ell] && (14)\\
&= \sum_{\ell=0}^{L-1}\sum_{\tilde{\ell}=0}^{L-1} h_\ell h_{\tilde{\ell}} x[m+\ell-\tilde{\ell}] && (15)\\
&= \left(\sum_{\ell=0}^{L-1} h_\ell^2\right) x[m] + \left(\sum_{\ell,\tilde{\ell}=0:\ell\neq\tilde{\ell}}^{L-1} h_\ell h_{\tilde{\ell}} x[m+\ell-\tilde{\ell}]\right) + \left(\sum_{\ell=0}^{L-1} h_\ell w[m+\ell]\right). && (16)
\end{aligned}$$

We observe that the channel between the transmitted voltage $x[m]$ and the received voltage $y[m]$ is not quite AWGN: the main difference is the presence of the interference

$$I[m] \stackrel{\mathrm{def}}{=} \sum_{\ell,\tilde{\ell}=0:\ell\neq\tilde{\ell}}^{L-1} h_\ell h_{\tilde{\ell}} x[m+\ell-\tilde{\ell}], \tag{17}$$

which is *discrete valued* and not Gaussian. So the ML rule for detecting $x[m]$ from $y[m]$ does not directly follow from our analysis of the AWGN channel. While in certain instances (which are explored in the homework), we see that the ML rule still comes down to the simple nearest-neighbor rule, this is not necessarily always the case.

As we have seen earlier in this course, it is useful for the communication engineer to have a simple rule of thumb to quickly measure the reliability of communication. In the AWGN channel, the operating SNR serves as a natural criterion; indeed, it is directly related to the reliability of communication. While the SNR is not so directly related to the reliability of communication in additive but non-Gaussian channels, it is very easy to measure and still serves as a surrogate to measure the reliability of communication. In the context of the additive noise plus interference channel in Equation (16), the signal energy is

$$\left(\sum_{\ell=0}^{L-1} h_\ell^2\right)^2 E, \tag{18}$$

the interference energy is

$$\sum_{\hat{\ell}\neq 0,\ \hat{\ell}=-L}^{L} \left(\sum_{\ell,\tilde{\ell}=0:\ell-\tilde{\ell}=\hat{\ell}}^{L-1} h_\ell h_{\tilde{\ell}}\right)^2 E, \tag{19}$$

and the noise energy is

$$\left(\sum_{\ell=0}^{L-1} h_\ell^2\right) \sigma^2. \tag{20}$$

So the *signal to interference plus noise ratio* (SINR) at the output of the matched filter is

$$\begin{aligned}
\mathsf{SINR}_{\mathrm{MF}} &\stackrel{\mathrm{def}}{=} \frac{\left(\sum_{\ell=0}^{L-1} h_\ell^2\right)^2 E}{\sum_{\hat{\ell}\neq 0,\hat{\ell}=-L}^{L}\left(\sum_{\ell,\tilde{\ell}=0:\ell-\tilde{\ell}=\hat{\ell}}^{L-1} h_\ell h_{\tilde{\ell}}\right)^2 E + \left(\sum_{\ell=0}^{L-1} h_\ell^2\right)\sigma^2} &(21)\\
&= \frac{c_1 \mathsf{SNR}}{c_2\mathsf{SNR}+1}, &(22)
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{SNR} &= \tfrac{E}{\sigma^2} &(23)\\
c_1 &\stackrel{\mathrm{def}}{=} \sum_{\ell=0}^{L-1} h_\ell^2 &(24)\\
c_2 &\stackrel{\mathrm{def}}{=} \frac{1}{c_1}\sum_{\hat{\ell}\neq 0,\ \hat{\ell}=-L}^{L}\left(\sum_{\ell,\tilde{\ell}=0:\ell-\tilde{\ell}=\hat{\ell}}^{L-1} h_\ell h_{\tilde{\ell}}\right)^2 &(25)
\end{aligned}$$

From Equation (22) we see clearly how the SINR at the output of the matched filter depends on the operating SNR. This function is schematically illustated in Figure 1.

Two different regimes are of keen interest:

- *Low* SNR: For small values of SNR, the SINR is close to linear in SNR. In other words, a doubling of the operating SNR also doubles the $\mathsf{SINR}_{\mathrm{MF}}$. Mathematically,

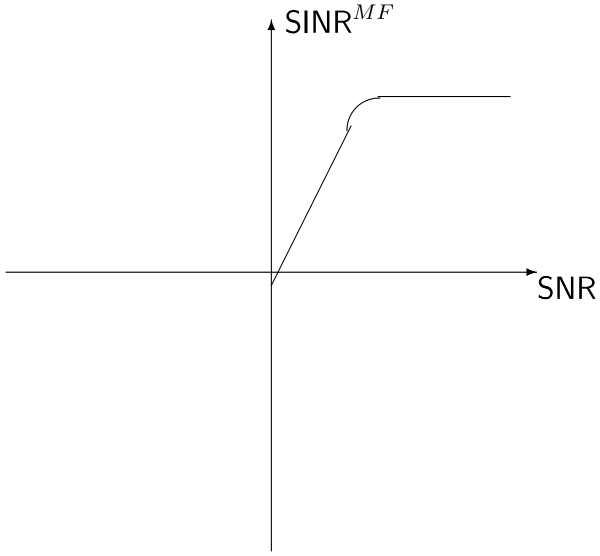

Figure 1: SINR at the output of the matched filter plotted as a function of the operating SNR.

this regime happens when the interference $c_2\mathsf{SNR} \ll 1$. In this regime, the interference is much smaller than the background noise and it makes fine sense to just ignore the interference. The channel is almost like an AWGN one and hence the linear relationship between SNR and $\mathsf{SINR}_{\mathrm{MF}}$.

- *High* SNR: For large values of SNR the $\mathsf{SINR}_{\mathrm{MF}}$ is almost constant and hardly changes. Mathematically, this regime kicks in when the interference $c_2\mathsf{SNR} \gg 1$. In this regime, the interference is much larger than noise and we pay a steep price by just ignoring its presence. The interference level is directly proportional to the transmit signal energy and hence the $\mathsf{SINR}_{\mathrm{MF}}$ is not sensitive to increases in SNR in this regime.

To ameliorate the deleterious presence of interference in the high SNR regime, one needs to explicitly deal with this by jointly detecting the sequence of symbols instead of detecting symbol-by-symbol. This will be the topic of the next lecture.