ECE 361: Digital Communication

Lecture 14: Interference Management at all SNRs: Minimum Mean Square Error (MMSE) Filter

Introduction

In the previous lectures, we have seen two linear processing techniques that deal with ISI: the matched filter (which worked well at low SNRs) and the zero forcing equalizer (which worked well at high SNRs). The matched filter harnessed the multiple delayed copies of each transmit symbol. The zero forcing equalizer worked to entirely null all the multiple delayed copies (and hence zero out the interference). In this lecture we study the *optimal* linear filter that balances both these effects: the need to harness the multiple delayed copies while still mitigating the effect of interference. As usual, our performance metric will be the SINR of the transmit symbol at the output of the filter. This filter, known as the minimum mean square error (MMSE) equalizer, in conjunction with the successive interference cancelation (SIC) technique, is routinely used in practical communication schemes over wireline channels (an example: the voiceband telephone line modem).

We have seen that the matched filter ignores the interference and combines the received voltage points to collect the signal energy. This strategy works well at low SNRs. In particular, the SINR at the output of matched filter is of the form

$$\mathsf{SINR}_{MF} = \frac{a\mathsf{SNR}}{1 + b\mathsf{SNR}} \tag{1}$$

where a and b are the appropriate constants (that are independent of SNR).

The zero forcing equalizer works well in high SNR regime. It essentially ignores noise and removes ISI by solving an overspecified set of linear equations. The SINR at the output of ZFE is of the form

$$\mathsf{SINR}_{ZFE} = c\mathsf{SNR} \tag{2}$$

where c is the appropriate constant.

Both these equalizers are simple linear processing that perform well in different SNR regimes. In this lecture, we study a linear processing strategy that combines the best of both the equalizers and strictly outperform them in both regimes. We will motivate it by studying an example.

Getting Started

To motivate the design of this filter, consider the simple 2-tap ISI channel:

$$y[m] = h_0 x[m] + h_1 x[m-1] + w[m], \quad m \geq 1. \tag{3}$$

Suppose we communicate information sequentially, say one bit at a time: so $x[m]$ is $\pm\sqrt{E}$ for each m and independent over different time samples. We want to filter only the first two received voltages $y[1], y[2]$ to estimate the first transmit voltage $x[1]$:

$$\begin{aligned} y[1] &= h_0 x[1] + w[1] && (4)\\ y[2] &= h_1 x[1] + (h_0 x[2] + w[2]). && (5) \end{aligned}$$

The matched filter receiver would be

$$y^{\mathrm{MF}}[1] = h_0 y[1] + h_1 y[2], \tag{6}$$

while the ZFE would be

$$y^{\mathrm{ZFE}}[1] = y[1]. \tag{7}$$

We are interested in choosing a_1, a_2 such that using the filter

$$\begin{aligned} y^{\{a_1,a_2\}}[1] &\stackrel{\mathrm{def}}{=} a_1 y[1] + a_2 y[2] && (8)\\ &= (a_1 h_0 + a_2 h_1)\, x[1] + (a_1 w[1] + a_2 h_0 x[2] + a_2 w[2]) && (9) \end{aligned}$$

has the largest SINR at its output in terms of the transmit voltage $x[1]$:

$$\mathsf{SINR}_{\{a_1,a_2\}} \stackrel{\mathrm{def}}{=} \frac{(a_1 h_0 + a_2 h_1)^2\, \mathsf{SNR}}{a_1^2 + a_2^2 + a_2^2 h_0^2 \mathsf{SNR}}. \tag{10}$$

Here we have denoted the ratio of E to σ^2 by SNR, as usual. There is an alternative nomenclature for the filter that has the largest SINR: the MMSE equalizer. Why the filter that has the largest SINR also has the "minimum mean squared error" is explored in a homework exercise.

To get a feel for the appropriate choice of the filter coefficients a_1 and a_2, consider the following somewhat abstract representation of Equations (4) and (5):

$$\begin{aligned} y[1] &= h_0 x[1] + z[1] && (11)\\ y[2] &= h_1 x[1] + z[2]. && (12) \end{aligned}$$

Specifically, in terms of Equations (4) and (5), we have

$$\begin{aligned} z[1] &= w[1] && (13)\\ z[2] &= h_0 x[2] + w[2]. && (14) \end{aligned}$$

Observe that the "interference plus noise" terms $z[1], z[2]$ are statistically uncorrelated (even statistically independent in this case). Let us denote the variance of $z[i]$ by $\sigma^2_{z[i]}$ for $i = 1, 2$. In the context of Equations (4) and (5) we have

$$\begin{aligned} \sigma^2_{z[1]} &= \sigma^2 && (15)\\ \sigma^2_{z[2]} &= h_0^2 E + \sigma^2. && (16) \end{aligned}$$

Using the linear filter in Equation (8):

$$y^{\{a_1,a_2\}}[1] = (a_1 h_0 + a_2 h_1)\, x[1] + (a_1 z[1] + a_2 z[2])\,, \tag{17}$$

we see that the signal energy is

$$(a_1 h_0 + a_2 h_1)^2\, E. \tag{18}$$

On the other hand, interference plus noise energy is

$$a_1^2 \sigma^2_{z[1]} + a_2^2 \sigma^2_{z[2]}, \tag{19}$$

where we used the property that $z[1], z[2]$ are uncorrelated. Thus the output SINR is

$$\mathsf{SINR}_{\{a_1,a_2\}} = \frac{(a_1 h_0 + a_2 h_1)^2 E}{\left(a_1^2 \sigma_{z[1]}^2 + a_2^2 \sigma_{z[2]}^2\right)}. \tag{20}$$

The MMSE equalizer chooses a_1 and a_2 that will maximize SINR. Note that only the *ratio* of the linear coefficients a_1, a_2 decides the SINR. But how do the matched filter (MF) and ZFE choose these coefficients?

We know that the matched filter ignores the interference. Thus, it assumes $z[1]$ and $z[2]$ to have same variance σ^2. We have seen from Cauchy-Schwartz inequality that the coefficients that maximize the SNR are $a_1 = h_0$ and $a_2 = h_1$. We conclude:

> The matched filter would be indeed the MMSE equalizer, i.e., possess the largest output SINR among all linear filters, if the interference plus noise terms are uncorrelated and have the same variance.

On the other hand, ZFE ignores the noise and treats Equations (4) and (5) as "deterministic". Clearly, it chooses only $y[1]$ to decode $x[1]$ and hence $a_1 = 1$ and $a_2 = 0$.

Unlike the matched filter, MMSE equalizer maximizes SINR even when $\sigma_{z[1]}^2$ and $\sigma_{z[2]}^2$ are not necessarily the same. The key idea in deriving this filter is to *scale* the received voltages such that the scaled interference plus noise terms have the same variance. Consider

$$\tilde{y}[1] \stackrel{\text{def}}{=} y[1] = h_0 x[1] + z[1] \tag{21}$$

$$\tilde{y}[2] \stackrel{\text{def}}{=} c y[2] = c h_1 x[1] + c z[2]. \tag{22}$$

we choose c such that $z[1]$ and $cz[2]$ have the same variance, i.e.,

$$c^2 (h_0^2 E + \sigma^2) = \sigma^2 \tag{23}$$

$$c^2 = \frac{1}{1 + h_0^2 \mathsf{SNR}}. \tag{24}$$

Now both the $z[1]$ and $cz[2]$ have variance σ^2. We know from Lecture 10 that the matched filter of the scaled received voltages will maximize the SINR. Hence,

$$y_{\text{MMSE}}[1] = h_0 \tilde{y}[1] + c h_1 \tilde{y}[2] \tag{25}$$

$$= h_0 y[1] + c^2 h_1 y[2]. \tag{26}$$

Thus, the optimal choice of the coefficients a_1 and a_2 are

$$a_1^* = h_0, \quad \text{and} \quad a_2^* = c^2 h_1 = \frac{h_1}{1 + h_0^2 \mathsf{SNR}}. \tag{27}$$

It is instructive to check what happens to the MMSE filter coefficients at high and low SNRs:

- At low SNRs, when $\mathsf{SNR} \ll 1$, we see from Equation (27) that

$$a_1 = h_0 \tag{28}$$

$$a_2 \approx h_1. \tag{29}$$

In other words, the MMSE filter is just the regular *matched filter* at low SNRs.

- At high SNRs, when $\mathsf{SNR} \gg 1$, we see from Equation (27) that

$$a_1 = h_0 \quad (30)$$
$$a_2 \approx 0. \quad (31)$$

In other words, the MMSE filter is just the regular *zero forcing equalizer* at high SNRs.

This formally justifies our intuition (from Lectures 10 and 11) that the matched filter and zero forcing equalizers were performing quite well at low and high SNRs respectively.

From Equation (20), the SINR achieved by the MMSE equalizer is

$$\mathsf{SINR}_{\text{MMSE}} = \frac{\left(a_1^* h_0 + a_2^* h_1\right)^2 E}{\left((a_1^*)^2\sigma^2 + (a_2^*)^2(h_0^2 E + \sigma^2)\right)} \quad (32)$$

$$\mathsf{SINR}_{\text{MMSE}} = \frac{(h_0^2 + c^2 h_1^2)^2 \mathsf{SNR}}{(h_0^2 + c^2 h_1^2)} \quad (33)$$

$$= (h_0^2 + c^2 h_1^2)\mathsf{SNR} \quad (34)$$

$$= h_0^2 \mathsf{SNR} + \frac{h_1^2 \mathsf{SNR}}{1 + h_0^2 \mathsf{SNR}}. \quad (35)$$

Note that the first term corresponds to the SINR that ZFE provides, whereas the second term can be seen as a contribution of the matched filter. Thus, MMSE combines both the equalizers and outperforms them at all SNRs.

A General Derivation

While our derivation of the MMSE filter was very concrete, it was for a rather special scenario where the interference plus noise terms were uncorrelated and all we had to do was just normalize the variances of the noise plus interference terms. More generally, they would be correlated – for example, consider the 3-tap ISI channel:

$$y[m] = h_0 x[m] + h_1 x[m-1] + h_2 x[m-2] + w[m] \quad (36)$$

where we want to estimate $x[1]$ from $y[1], y[2], y[3]$:

$$y[1] = h_0 x[1] + w[1] \quad (37)$$
$$y[2] = h_1 x[1] + (h_0 x[2] + w[2]) \quad (38)$$
$$y[3] = h_2 x[1] + (h_0 x[3] + h_1 x[2] + w[3]). \quad (39)$$

It helps to write the received voltages in a *vector* form:

$$\begin{bmatrix} y[1] \\ y[2] \\ y[3] \end{bmatrix} = \begin{bmatrix} h_0 \\ h_1 \\ h_2 \end{bmatrix} x[1] + \begin{bmatrix} z[1] \\ z[2] \\ z[3] \end{bmatrix}, \quad (40)$$

where $z[1], z[2], z[3]$ are the interference plus noise terms. If $z[1], z[2], z[3]$ are all uncorrelated, then we could have derived the MMSE filter by:

- scaling the received voltages so that the variances of $z[1], z[2], z[3]$ are all the same;
- then matched filtering the scaled received voltages.

But in this case, we see that though $z[1]$ is independent of $z[2]$ and $z[3]$, $z[2]$ and $z[3]$ are correlated since the transmit voltage $x[2]$ appears in both of them. How does one derive the MMSE filter for such a situation?

It would help if we can "whiten" $y[2]$ and $y[3]$, i.e., ensure that the interference plus noise terms in them are uncorrelated. We should exploit the correlation between $z[2]$ and $z[3]$. Ideally we would like to get, by appropriate linear processing of y_2 and y_3,

$$\hat{y}[3] = \alpha x[1] + \hat{z}[3] \tag{41}$$

such that $z[1], z[2]$ and $\hat{z}[3]$ are uncorrelated. Here α is the coefficient that naturally arises when this process is completed (and, as we will see below, is uniquely defined).

The correlation between $z[2]$ and $z[3]$ is

$$\begin{aligned} \mathbb{E}\left[z[2]z[3]\right] &= \mathbb{E}\left[h_0 h_1 (x[2])^2\right] && (42)\\ &= h_0 h_1 E. && (43) \end{aligned}$$

Let's obtain $\hat{y}[3]$ by linear processing between $y[2]$ and $y[3]$.

$$\begin{aligned} \hat{y}[3] &= y[3] - a y[2] && (44)\\ &= (h_2 - a h_1)x[1] + (z[3] - a z[2]). && (45) \end{aligned}$$

Let $\hat{z}[3] \stackrel{\text{def}}{=} (z[3] - a z[2])$. We want to choose a such that $z[2]$ and $\hat{z}[3]$ are uncorrelated, i.e., we need

$$\begin{aligned} \mathbb{E}\left[z[2]\hat{z}[3]\right] &= 0 && (46)\\ \mathbb{E}\left[z[2]z[3]\right] &= a\mathbb{E}\left[(z[2])^2\right] && (47)\\ h_0 h_1 E &= a(h_0^2 E + \sigma^2) && (48)\\ a &= \frac{h_0 h_1 E}{h_0^2 E + \sigma^2} = \frac{h_0 h_1 \mathsf{SNR}}{h_0^2 \mathsf{SNR} + 1}. && (49) \end{aligned}$$

With this choice of a, we have that

$$\mathbb{E}\left[z[2]\hat{z}[3]\right] = 0. \tag{50}$$

Further, the variance of $\hat{z}[3]$ is

$$\mathbb{E}\left[(\hat{z}[3])^2\right] = \mathbb{E}\left[(z[3] - a z[2])^2\right] \tag{51}$$

Substituting for a and $\mathbb{E}\left[z[2]z[3]\right]$, we get

$$\begin{aligned} \mathbb{E}\left[(\hat{z}[3])^2\right] &= (h_0^2 + h_1^2)E + \sigma^2 - \frac{h_0^2 h_1^2 E^2}{(h_0^2 E + \sigma^2)} && (52)\\ &= (h_0^2 E + \sigma^2) + \frac{h_1^2 E \sigma^2}{(h_0^2 E + \sigma^2)}. && (53) \end{aligned}$$

We now have three noisy observations of $x[1]$, namely, $y[1], y[2]$ and $\hat{y}[3]$. These are corrupted by the noises $z[1], z[2]$ and $\hat{z}[3]$ that are uncorrelated. We now just need to normalize their variances and then choose the coefficients of $x[1]$ to be the optimal choice of a_1, a_2 and a_3 (matched filtering).

Thus, we have

$$\tilde{y}[1] = \quad y[1] = \quad h_0 x[1] + z[1] \tag{54}$$

$$\tilde{y}[2] = \quad c_2 y[2] = \quad c_2 h_1 x[1] + c_2 z[2] \tag{55}$$

$$\tilde{y}[3] = \quad c_3 \hat{y}[3] = \quad c_3 (h_2 - a h_1) x[1] + c_3 \hat{z}[3] \tag{56}$$

where c_2 and c_3 are chosen to normalize the variances. As before

$$c_2^2 (h_0^2 E + \sigma^2) \;=\; \sigma^2 \tag{57}$$

$$c_2^2 \;=\; \frac{1}{1 + h_0^2 \mathsf{SNR}}. \tag{58}$$

Here c_3 is such that

$$c_3^2 \mathbb{E}\left[(\hat{z}[3])^2 \right] \;=\; \sigma^2 \tag{59}$$

$$c_3^2 \;=\; \frac{1 + h_0^2 \mathsf{SNR}}{(1 + h_0^2 \mathsf{SNR})^2 + h_1^2 \mathsf{SNR}}. \tag{60}$$

MMSE equalizer is then

$$y_{\text{MMSE}}[1] \;=\; h_0 \tilde{y}[1] + c_2 h_1 \tilde{y}[2] + c_3 (h_2 - a h_1) \tilde{y}[3]. \tag{61}$$

More generally, when we have an arbitrary number of taps L (not simply less than 4) and when we are interested in a general transmit voltage $x[m]$ (not just for $m = 1$), there is a straightforward procedure to derive the MMSE filter without resorting to numerical optimization. This procedure is best explained using tools from linear algebra and is explored in some of your reference books.

Looking Ahead

So far, we have let the receiver handle much of the burden of mitigating ISI. In the next lecture, we will see how the transmitter can also work to address the ISI issue. Specifically, the interference is from symbols sent in the past and, surely, the transmitter is already aware of what interference it has caused. If the transmitter is savvy about what it transmits, then the ISI can be controlled to be within acceptable limits. Such transmitter strategies (that go beyond the sequential and block communication schemes we have seen in the context of the AWGN channel) are known as *precoding*. We will see this next.