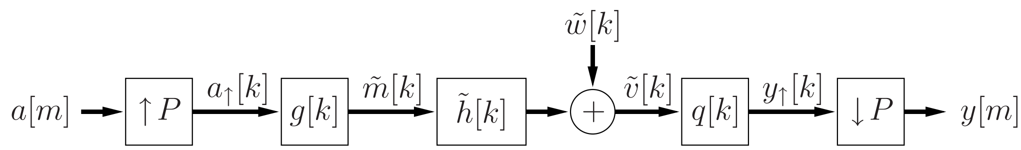
MMSE Equalizer Design

Phil Schniter

March 6, 2008



For a trivial channel (i.e., $\tilde{h}[k] = \delta[k]$), we know that the use of square-root raised-cosine (SRRC) pulses at transmitter and receiver suppresses inter-symbol interference (ISI) and maximizes the received signal-to-noise ratio (SNR) in the presence of white noise $\{\tilde{w}[k]\}$. With a non-trivial channel, however, we need to re-visit the design of the receiver pulse $\{q[k]\}$, which is called an "equalizer" when it tries to compensate for the channel.

Here we design the minimum mean-squared error (MMSE) equalizer coefficients $\{q[k]\}$ assuming that the input symbols $\{a[n]\}$ and the noise $\{\tilde{w}[k]\}$ are white random sequences that are uncorrelated with each other. This means that

$$\mathrm{E}\{a[m]a^*[n]\} = \sigma_a^2 \delta[m-n] \quad (1)$$
$$\mathrm{E}\{\tilde{w}[k]\tilde{w}^*[l]\} = \sigma_w^2 \delta[k-l] \quad (2)$$
$$\mathrm{E}\{a[m]\tilde{w}^*[l]\} = 0 \quad (3)$$

for some positive variances σ_a^2 and σ_w^2. For practical implementation, we will consider a causal equalizer with length N_q, so that $q[k] = 0$ for $k < 0$ and $k \geq N_q$. To simplify the derivation, we combine the transmitted pulse $g[k]$ and the complex-baseband channel $\tilde{h}[k]$ into the "effective channel"

$$h[k] := g[k] * \tilde{h}[k]$$

and assume that this effective channel is causal with finite length N_h. Throughout, we assume that the effective channel coefficients $\{h[k]\}$, as well as the variances σ_a^2 and σ_w^2, are known. Learning these quantities is a separate (and often challenging) problem.

Notice that, because the effective channel is causal and length N_h, it can delay the upsampled input signal $a_\uparrow[k]$ by between 0 and $N_h - 1$ samples. Since it is difficult to compensate for this delay with a causal equalizer, we will allow for the presence of end-to-end system delay. Thus, our goal is to make $y[m] \approx a[m-\Delta]$ for some integer $\Delta \geq 0$. Throughout the design, we assume that Δ has been chosen for us, although eventually we shall see how to optimize Δ.

Recall that if $y[m] = a[m-\Delta]$, then we will be able to make perfect decisions on the symbols $a[m]$ from the output sequence $y[m]$. However, we would never expect a perfect output in the presence of noise. Thus, we take as our objective the minimization of the error signal

$$e[m] := y[m] - a[m-\Delta].$$

In particular, we minimize the mean squared error (MSE)

$$\mathcal{E} := \mathrm{E}\{|e[m]|^2\}.$$

We saw earlier that, if $e[m]$ can be modelled as a zero-mean Gaussian random variable (with variance $\sigma_e^2 = \mathcal{E}$), then the symbol error rate (SER) decreases as $\mathcal{E}/\sigma_a^2$ decreases. Thus, there is good reason to minimize $\mathcal{E}$.

Our eventual goal is to derive an expression for the MSE $\mathcal{E}$ from which the equalizer coefficients can be optimized. But first we notice that, due to the stationary of $\{a[m]\}$ and $\{\tilde{w}[k]\}$ (i.e., the time-invariance of their statistics) and the LTI nature of our filters, the statistics of $\{e[m]\}$ will also be time invariant, allowing us to write $\mathcal{E} = \mathrm{E}\{|e[0]|^2\}$. This allows us to focus on $e[0]$ instead of $e[m]$, which simplifies the development.

The next step is then to find an expression for $e[0]$. From the block diagram,

$$\begin{aligned} e[0] &= y[0] - a[-\Delta] && (4) \\ &= \sum_{l=0}^{N_q-1} q[l]\tilde{v}[-l] - a[-\Delta] && (5) \\ \tilde{v}[k] &= \tilde{w}[k] + \sum_{l=-\infty}^{\infty} a_{\uparrow}[l]h[k-l] && (6) \\ &= \tilde{w}[k] + \sum_{n=-\infty}^{\infty} a[n]h[k-nP], && (7) \end{aligned}$$

where in (7) we used the fact that $a_{\uparrow}[l] = 0$ when l is not a multiple of P and that $a_{\uparrow}[nP] = a[n]$. Though (7) is written with an infinite summation, it turns out that most values of n will not contribute. Due to the causality and length-N_h of $h[n]$, the values of n which lead to contributions to $\tilde{v}[k]$ ensure that

$$\begin{aligned} & 0 \le k - nP \le N_h - 1 \quad \text{for at least one } k \in \{0,-1,\dots,-N_q+1\} && (8) \\ \Leftrightarrow\ & \frac{k}{P} \ge n \ge -\frac{N_h - 1 - k}{P} \quad \text{for at least one } k \in \{0,-1,\dots,-N_q-1\} && (9) \\ \Leftrightarrow\ & 0 \ge n \ge -\underbrace{\left\lfloor \frac{N_h + N_q - 2}{P} \right\rfloor}_{:= N_a - 1}, && (10) \end{aligned}$$

where N_a denotes the number of contributing symbols. In other words,

$$\tilde{v}[k] = \tilde{w}[k] + \sum_{n=1-N_a}^{0} a[n]h[k-nP]. \quad (11)$$

Next we use a vector formulation to simplify the development. We start by rewriting (5) as

$$e[0] = \underbrace{\left[q[0] \quad q[1] \quad \cdots \quad q[N_q-1]\right]}_{:= \underline{q}^T} \begin{bmatrix} \tilde{v}[0] \\ \tilde{v}[-1] \\ \vdots \\ \tilde{v}[1-N_q] \end{bmatrix} - a[-\Delta], \tag{12}$$

where, for $l \in \{0, \ldots, N_q - 1\}$,

$$\tilde{v}[-l] = \tilde{w}[-l] + \underbrace{\left[h[-l] \quad h[-l+P] \quad \cdots \quad h[-l+(N_a-1)P]\right]}_{:= \underline{h}_{-l}^T} \underbrace{\begin{bmatrix} a[0] \\ a[-1] \\ \vdots \\ a[1-N_a] \end{bmatrix}}_{:= \underline{a}} \tag{13}$$

so that

$$\begin{bmatrix} \tilde{v}[0] \\ \tilde{v}[-1] \\ \vdots \\ \tilde{v}[1-N_q] \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{w}[0] \\ \tilde{w}[-1] \\ \vdots \\ \tilde{w}[1-N_q] \end{bmatrix}}_{:= \underline{w}} + \underbrace{\begin{bmatrix} \underline{h}_0^T \\ \underline{h}_{-1}^T \\ \vdots \\ \underline{h}_{1-N_q}^T \end{bmatrix}}_{:= \underline{H}} \underline{a}. \tag{14}$$

In (14), the row vectors $\{\underline{h}_{-l}^T\}_{l=0}^{N_q-1}$ were combined to form the $N_q \times N_a$ matrix $\underline{H}$. Throughout, we use underlined lower-case letters to represent column vectors, and underlined upper-case letters to represent matrices. Plugging (14) into (12), we get

$$e[0] = \underline{q}^T(\underline{w} + \underline{H}\,\underline{a}) - a[-\Delta]. \tag{15}$$

Defining $\underline{\delta}_\Delta$ as the column vector with a 1 in the Δ^{th} place[1] and 0's elsewhere, we can write $\underline{\delta}_\Delta^T \underline{a} = a[-\Delta]$, which yields the final expression for the time-0 error:

$$\begin{aligned} e[0] &= \underline{q}^T(\underline{w} + \underline{H}\,\underline{a}) - \underline{\delta}_\Delta^T \underline{a} && (16) \\ &= \underline{q}^T \underline{w} + (\underline{q}^T \underline{H} - \underline{\delta}_\Delta^T)\underline{a}. && (17) \end{aligned}$$

Next, we derive an expression for the MSE $\mathcal{E}$. Notice that

$$\begin{aligned} \mathcal{E} &= \mathrm{E}\{|e[0]|^2\} = \mathrm{E}\{e[0]e^*[0]\} && (18) \\ &= \mathrm{E}\left\{ \left[\underline{q}^T\underline{w} + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\right] \left[\underline{q}^T\underline{w} + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\right]^* \right\} && (19) \\ &= \mathrm{E}\left\{ \left[\underline{q}^T\underline{w} + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\right] \left[\underline{w}^T\underline{q} + \underline{a}^T(\underline{H}^T\underline{q} - \underline{\delta}_\Delta)\right]^* \right\} && (20) \\ &= \mathrm{E}\left\{ \left[\underline{q}^T\underline{w} + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\right] \left[\underline{w}^H\underline{q}^* + \underline{a}^H(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta)\right] \right\}, && (21) \end{aligned}$$

[1] Note that $\Delta = 0$ indicates the first place, so that $\underline{\delta}_0 = [1, 0, \ldots, 0]^T$.

where in (20) we transposed the scalar quantities on the right (e.g., $\underline{q}^T\underline{w} = (\underline{q}^T\underline{w})^T = \underline{w}^T\underline{q}$) and in (21) we distributed the complex conjugate, using the "Hermitian transpose" notation $(\cdot)^H := (\cdot)^{T*}$. Expanding (21) gives

$$\begin{aligned}
\mathcal{E} &= \mathrm{E}\left\{\underline{q}^T\underline{w}\underline{w}^H\underline{q}^*\right\} + \mathrm{E}\left\{\underline{q}^T\underline{w}\underline{a}^H(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta)\right\} \\
&\quad + \mathrm{E}\left\{(\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\underline{w}^H\underline{q}^*\right\} + \mathrm{E}\left\{(\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\underline{a}\underline{a}^H(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta)\right\} \qquad (22) \\
&= \underline{q}^T\,\mathrm{E}\left\{\underline{w}\underline{w}^H\right\}\underline{q}^* + \underline{q}^T\,\mathrm{E}\left\{\underline{w}\underline{a}^H\right\}(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta) \\
&\quad + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\,\mathrm{E}\left\{\underline{a}\underline{w}^H\right\}\underline{q}^* + (\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)\,\mathrm{E}\left\{\underline{a}\underline{a}^H\right\}(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta), \qquad (23)
\end{aligned}$$

where in (23) we moved the non-random quantities through the expectations.

Equation (23) can be simplified using the statistical properties of the random vectors $\underline{a}$ and $\underline{w}$, which were each constructed from white sequences that are uncorrelated with each other. In particular, notice that

$$\begin{aligned}
&\mathrm{E}\left\{\underline{w}\underline{w}^H\right\} \\
&= \mathrm{E}\left\{\begin{bmatrix}\tilde{w}[0] \\ \tilde{w}[-1] \\ \vdots \\ \tilde{w}[1-N_q]\end{bmatrix}\begin{bmatrix}\tilde{w}^*[0] & \tilde{w}^*[-1] & \cdots & \tilde{w}^*[1-N_q]\end{bmatrix}\right\} \qquad (24) \\
&= \mathrm{E}\left\{\begin{bmatrix}\tilde{w}[0]\tilde{w}^*[0] & \tilde{w}[0]\tilde{w}^*[-1] & \cdots & \tilde{w}[0]\tilde{w}^*[1-N_q] \\ \tilde{w}[-1]\tilde{w}^*[0] & \tilde{w}[-1]\tilde{w}^*[-1] & \cdots & \tilde{w}[-1]\tilde{w}^*[1-N_q] \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{w}[1-N_q]\tilde{w}^*[0] & \tilde{w}[1-N_q]\tilde{w}^*[-1] & \cdots & \tilde{w}[1-N_q]\tilde{w}^*[1-N_q]\end{bmatrix}\right\} \qquad (25) \\
&= \begin{bmatrix}\mathrm{E}\{\tilde{w}[0]\tilde{w}^*[0]\} & \mathrm{E}\{\tilde{w}[0]\tilde{w}^*[-1]\} & \cdots & \mathrm{E}\{\tilde{w}[0]\tilde{w}^*[1-N_q]\} \\ \mathrm{E}\{\tilde{w}[-1]\tilde{w}^*[0]\} & \mathrm{E}\{\tilde{w}[-1]\tilde{w}^*[-1]\} & \cdots & \mathrm{E}\{\tilde{w}[-1]\tilde{w}^*[1-N_q]\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\{\tilde{w}[1-N_q]\tilde{w}^*[0]\} & \mathrm{E}\{\tilde{w}[1-N_q]\tilde{w}^*[-1]\} & \cdots & \mathrm{E}\{\tilde{w}[1-N_q]\tilde{w}^*[1-N_q]\}\end{bmatrix} \qquad (26) \\
&= \text{a matrix whose } (m,n)^{th} \text{ entry equals } \mathrm{E}\{\tilde{w}[-m]\tilde{w}^*[-n]\} = \sigma_w^2\delta[n-m] \qquad (27) \\
&= \sigma_w^2\underline{I}, \qquad (28)
\end{aligned}$$

where $\underline{I}$ denotes the identity matrix. The same reasoning can be used to show

$$\mathrm{E}\{\underline{a}\underline{a}^H\} = \sigma_a^2\underline{I}.$$

Similarly,

$$\begin{aligned}
\mathrm{E}\{\underline{a}\underline{w}^H\} &= \text{a matrix whose } (m,n)^{th} \text{ entry equals } \mathrm{E}\{a[-m]\tilde{w}^*[-n]\} = 0 \qquad (29) \\
&= \underline{0}, \qquad (30)
\end{aligned}$$

and $\mathrm{E}\{\underline{w}\underline{a}^H\} = (\mathrm{E}\{\underline{a}\underline{w}^H\})^H = \underline{0}^H = \underline{0}$. Applying these relationships to (23), we get

$$\begin{aligned}
\mathcal{E} &= \sigma_w^2\underline{q}^T\underline{q}^* + \sigma_a^2(\underline{q}^T\underline{H} - \underline{\delta}_\Delta^T)(\underline{H}^H\underline{q}^* - \underline{\delta}_\Delta) \qquad (31) \\
&= \sigma_w^2\|\underline{q}\|^2 + \sigma_a^2\|\underline{H}^T\underline{q} - \underline{\delta}_\Delta\|^2, \qquad (32)
\end{aligned}$$

which shows that the MSE consists of $\sigma_w^2\|\underline{q}\|^2$ (due to noise) plus $\sigma_a^2\|\underline{H}^T\underline{q}-\underline{\delta}_\Delta\|^2$ (due to ISI and error in the end-to-end gain). In general, minimizing the sum of these two components requires making a tradeoff between them. For example, setting $\underline{q}=\underline{0}$ would cancel the noise component but result in an end-to-end gain of zero. Similarly, choosing $\underline{q}$ such that $\underline{H}^T\underline{q}=\underline{\delta}_D$ (if this is even possible) can amplify the noise.

To proceed further, we rewrite the MSE as follows.

$$
\begin{aligned}
\mathcal{E} &= \underline{q}^T\underbrace{(\sigma_w^2\underline{I}+\sigma_a^2\underline{H}\,\underline{H}^H)}_{:=\underline{A}}\underline{q}^* - \underline{q}^T\underbrace{\underline{H}\underline{\delta}_\Delta\sigma_a^2}_{:=\underline{b}} - \sigma_a^2\underline{\delta}_\Delta^T\underline{H}^H\underline{q}^* + \sigma_a^2\underbrace{\underline{\delta}_\Delta^T\underline{\delta}_\Delta}_{=1} &(33)\\
&= \underline{q}^T\underline{A}\underline{q}^* - \underline{q}^T\underline{b} - \underline{b}^H\underline{q}^* + \sigma_a^2. &(34)
\end{aligned}
$$

By completing-the-square,[2] the MSE expression (34) can be put into a convenient form, from which the MSE-minimizing $\underline{q}^*$ will become readily apparent. To do this, it is essential that the matrix $\underline{A}$ can[3] be decomposed as $\underline{A}=\underline{B}\,\underline{B}^H$, so that

$$
\begin{aligned}
\mathcal{E} &= \underline{q}^T\underline{B}\,\underline{B}^H\underline{q}^* - \underline{q}^T\underline{b} - \underline{b}^H\underline{q}^* + \sigma_a^2 &(35)\\
&= \underline{q}^T\underline{B}\,\underline{B}^H\underline{q}^* - \underline{q}^T\underline{B}\,\underline{B}^{-1}\underline{b} - \underline{b}^H\underline{B}^{-H}\underline{B}^H\underline{q}^* + \sigma_a^2 &(36)\\
&= (\underline{q}^T\underline{B} - \underline{b}^H\underline{B}^{-H})(\underline{B}^H\underline{q}^* - \underline{B}^{-1}\underline{b}) - \underline{b}^H\underbrace{\underline{B}^{-H}\underline{B}^{-1}}_{=\underline{A}^{-1}}\underline{b} + \sigma_a^2 &(37)\\
&= \underbrace{(\underline{B}^H\underline{q}^* - \underline{B}^{-1}\underline{b})^H(\underline{B}^H\underline{q}^* - \underline{B}^{-1}\underline{b})}_{\geq 0} + \underbrace{\sigma_a^2 - \underline{b}^H\underline{A}^{-1}\underline{b}}_{\mathcal{E}_{\min}}. &(38)
\end{aligned}
$$

Note that the equalizer parameters only affect the first term in (38), which is non-negative. So, to minimize $\mathcal{E}$ via choice of $\underline{q}$, the best we can do is to set the first term in (38) to zero, at which point the second term specifies the minimum possible $\mathcal{E}$. Thus, the MSE-minimizing equalizer parameters are those which give $\underline{B}^H\underline{q}_{\min}^* = \underline{B}^{-1}\underline{b}$, i.e.,

$$
\begin{aligned}
\underline{q}_{\min}^* &= \underline{B}^{-H}\underline{B}^{-1}\underline{b} \;=\; \underline{A}^{-1}\underline{b} \;=\; (\sigma_w^2\underline{I}+\sigma_a^2\underline{H}\,\underline{H}^H)^{-1}\underline{H}\underline{\delta}_\Delta\sigma_a^2 &(39)\\
&= \left(\frac{\sigma_w^2}{\sigma_a^2}\underline{I}+\underline{H}\,\underline{H}^H\right)^{-1}\underline{H}\underline{\delta}_\Delta &(40)
\end{aligned}
$$

and the minimum MSE is

$$
\begin{aligned}
\mathcal{E}_{\min} &= \sigma_a^2 - \underline{b}^H\underline{A}^{-1}\underline{b} &(41)\\
&= \sigma_a^2 - \sigma_a^2\underline{\delta}_\Delta^T\underline{H}^H(\sigma_w^2\underline{I}+\sigma_a^2\underline{H}\,\underline{H}^H)^{-1}\underline{H}\underline{\delta}_\Delta\sigma_a^2 &(42)\\
&= \sigma_a^2\left(1-\underline{\delta}_\Delta^T\underline{H}^H\left(\frac{\sigma_w^2}{\sigma_a^2}\underline{I}+\underline{H}\,\underline{H}^H\right)^{-1}\underline{H}\underline{\delta}_\Delta\right). &(43)
\end{aligned}
$$

[2]For scalar quantities, this means writing $x^2-2xy+z=(x-y)^2-y^2+z$.

[3]To see this, we can use the singular value decomposition (SVD) $\underline{H}=\underline{U}\,\underline{S}\,\underline{V}^H$, where $\underline{U}$ and $\underline{V}$ are unitary and $\underline{S}$ is non-negative diagonal, to write $\underline{A}=(\sigma_w^2\underline{U}\,\underline{U}^H+\sigma_a^2\underline{U}\,\underline{S}\,\underline{V}^H\underline{V}\,\underline{S}\,\underline{U}^H)=\underline{U}(\sigma_w^2\underline{I}+\sigma_a^2\underline{S}^2)\underline{U}^H=\underline{U}\Sigma^2\underline{U}^H$. Thus, if we define $\underline{B}:=\underline{U}\Sigma$, then $\underline{A}=\underline{B}\,\underline{B}^H$. Note that $\underline{B}$ is guaranteed invertible when $\sigma_w^2>0$.

Finally, we can see how the delay Δ can be optimized. Notice from (43) that the term

$$\underline{\delta}_{\Delta}^{T} \underline{H}^{H} \left(\frac{\sigma_w^2}{\sigma_a^2} \underline{I} + \underline{H}\underline{H}^{H} \right)^{-1} \underline{H} \underline{\delta}_{\Delta}$$

is simply the Δ^{th} diagonal element of the matrix $\underline{H}^{H} \left(\frac{\sigma_w^2}{\sigma_a^2} \underline{I} + \underline{H}\underline{H}^{H} \right)^{-1} \underline{H}$, and that $\mathcal{E}_{\min}$ decreases as this term gets bigger. Thus, the MSE-minimizing Δ is simply the index of the maximum diagonal element of the matrix $\underline{H}^{H} \left(\frac{\sigma_w^2}{\sigma_a^2} \underline{I} + \underline{H}\underline{H}^{H} \right)^{-1} \underline{H}$.