Minimum Symbol-Error-Rate Equalisation

Sheng Chen

Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

The equalisation topic is well researched and a variety of solutions are available. The MAP sequence detector provides the lowest symbol error rate (SER) attainable, and the MLSE offers a near optimal solution. However, these optimal techniques are not yet practical for high-level modulation schemes, due to their computational complexity. Linear equaliser or linear-combiner DFE are practical schemes for high-level modulation systems, as they readily meet real-time computation requirements. This research re-visits the linear equaliser and conventional DFE. Classically, the minimum mean square error (MMSE) solution is regarded as the "optimal" solution for the linear equaliser or DFE. For the MMSE to be optimal, the probability distribution of the equaliser soft output must be Gaussian. As the PDF of the equaliser output is clear non-Gaussian (a sum of mixed Gaussians), the MMSE solution can be far away from the optimal solution. Adopting the non-Gaussian approach naturally leads to the optimal minimum SER (MSER) equaliser.

We will first consider the real-valued channel which generates the received signal sample of:

$$r(k) = \sum_{i=0}^{n_h-1} h_i s(k-i) + n(k) \quad (1)$$

where h_i are the CIR taps, n_h is the CIR length, the AWGN $n(k)$ has variance σ_n^2, and the L-PAM symbol $s(k) \in \mathcal{S} \triangleq \{s_l = 2l - L - 1, 1 \le l \le L\}$. The linear equaliser with an order m has the form:

$$y(k) = \mathbf{w}^T \mathbf{r}(k) \quad (2)$$

where $\mathbf{r}(k) = [r(k)\ r(k-1) \cdots r(k-m+1)]^T$, and $\mathbf{w} = [w_0\ w_1 \cdots w_{m-1}]^T$ is the weight vector. The equaliser output $y(k)$ is passed to a threshold detector which provides an estimate $\hat{s}(k-d)$ of the transmitted symbol $s(k-d)$, where $0 \le d \le m + n_h - 2$ is the decision delay.

The received signal vector can be expressed as:

$$\mathbf{r}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k) \quad (3)$$

where $\mathbf{n}(k) = [n(k)\ n(k-1) \cdots n(k-m+1)]^T$, the $m \times (m + n_h - 1)$ CIR matrix $\mathbf{H} = [H_{i,q}]$ is a Toeplitz matrix satisfying $H_{i,q} = h_{q-i}$ if $0 \le q - i \le n_h - 1$ and $H_{i,q} = 0$ otherwise. As $\mathbf{s}(k)$ has $N_s = L^{m+n_h-1}$ combinations, denoted as $\mathbf{s}_q$, $1 \le q \le N_s$, $\bar{\mathbf{r}}(k)$ takes values from the channel state set

$$\mathcal{R} \triangleq \{\bar{\mathbf{r}}_q = \mathbf{H}\mathbf{s}_q,\ 1 \le q \le N_s\}\,. \quad (4)$$

Similarly, express $y(k)$ as

$$y(k) = \mathbf{w}^T(\bar{\mathbf{r}}(k) + \mathbf{n}(k)) = \bar{y}(k) + e(k) \quad (5)$$

where $e(k)$ is Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$, and $\bar{y}(k)$ takes values from the set $\mathcal{Y} \triangleq \{\bar{y}_q = \mathbf{w}^T\bar{\mathbf{r}}_q, 1 \le q \le N_s\}$, which can be divided into M subsets

$$\mathcal{Y}_l \triangleq \{\bar{y}_q \in \mathcal{Y} | s(k-d) = s_l\},\ 1 \le l \le L\,. \quad (6)$$

Let the combined impulse response of the equaliser and channel be $\mathbf{c}$, which is given by

$$\mathbf{c}^T = \mathbf{w}^T\mathbf{H} = [c_0\ c_1 \cdots c_{m+n_h-2}] \quad (7)$$

Then $y(k)$ can be expressed as

$$y(k) = c_d s(k-d) + \sum_{i \ne d} c_i s(k-i) + e(k)\,. \quad (8)$$

The first term in (8) is the desired signal, and the second term is the residual ISI. Thus the optimal decision making is

$$\hat{s}(k-d) = \begin{cases} s_1, & \text{if } y(k) \le (s_1+1)c_d\,, \\ s_l, & \text{if } (s_l-1)c_d < y(k) \le (s_l+1)c_d \\ & \text{for } l = 2, \cdots L-1\,, \\ s_L, & \text{if } y(k) > (s_L-1)c_d\,. \end{cases} \quad (9)$$

Observe that the PDF of $y(k)$ is given by

$$p_y(x) = \frac{1}{\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \frac{1}{N_s} \sum_{l=1}^{L} \sum_{i=1}^{N_{sb}} \exp\left(-\frac{\left(x - \bar{y}_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \quad (10)$$

where $N_{sb} = N_s/L$ and $\bar{y}_i^{(l)} \in \mathcal{Y}_l$. Using the property $\mathcal{Y}_{l+1} = \mathcal{Y}_l + 2c_d$ and noting the symmetric distribution of $\mathcal{Y}_l$ around the symbol point $c_d s_l$, it can be shown that the SER is:

$$P_E(\mathbf{w}) = \frac{\gamma}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) \quad (11)$$

where $\gamma = 2(L-1)/L$, and

$$g_{l,i}(\mathbf{w}) = \frac{\bar{y}_i^{(l)} - c_d(s_l - 1)}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\,. \quad (12)$$

The MSER solution $\mathbf{w}_{\text{MSER}}$ that minimizes the SER (11) can readily be obtained using a gradient-based numerical optimization algorithm, such as the simplified conjugated gradient algorithm with reseting the search direction to the negative gradient $-\nabla P_E(\mathbf{w})$ every I iterations. As the SER is invariant to a positive scaling of $\mathbf{w}$, it is computationally advantageous to normalize the weight vector to $\mathbf{w}^T\mathbf{w} = 1$.

For block-data adaptation, a channel estimate can be identified and the MSER solution is obtained by optimization. Alternatively, a kernel density or Parzen window estimate approach can be adopted. An estimated PDF of $p_y(x)$ is

$$\hat{p}_y(x) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{K}\sum_{k=1}^{K}\exp\left(-\frac{(x-y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right) \quad (13)$$

where K is the length of training samples, and the radius parameter ρ_n is related to σ_n. From this estimated PDF, the estimated SER expression is given by

$$\hat{P}_E(\mathbf{w}) = \frac{\gamma}{K}\sum_{k=1}^{K} Q(\hat{g}_k(\mathbf{w})) \quad (14)$$

where

$$\hat{g}_k(\mathbf{w}) = \frac{y(k) - \hat{c}_d(s(k-d)-1)}{\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\,, \quad (15)$$

$\hat{c}_d = \mathbf{w}^T\hat{\mathbf{h}}_d$, and $\hat{\mathbf{h}}_d$ an estimate for the d-th column $\mathbf{h}_d$ of $\mathbf{H}$. Given the gradient $\nabla\hat{P}_E(\mathbf{w})$, the estimated MSER solution can be obtained. To derive a sample-by-sample adaptive algorithm, consider a single-sample estimate of $p_y(x)$

$$\hat{p}_y(x,k) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\exp\left(-\frac{(x-y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right)\,. \quad (16)$$

With a re-scaling after each update to ensure $\mathbf{w}^T\mathbf{w} = 1$, the instantaneous stochastic gradient is given by

$$\nabla\hat{P}_E(\mathbf{w},k) = \frac{\gamma}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{(y(k)-\hat{c}_d(s(k-d)-1))^2}{2\rho_n^2}\right)\times \left((y(k)-\hat{c}_d(s(k-d)-1))\mathbf{w} - \mathbf{r}(k) + (s(k-d)-1)\hat{\mathbf{h}}_d\right)\,. \quad (17)$$

This leads to the least SER (LSER) algorithm

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\nabla\hat{P}_E(\mathbf{w}(k),k)\,. \quad (18)$$

The adaptive gain μ and width ρ_n are the two algorithm parameters that need to be set appropriately.

We next address the DFE defined by:

$$y(k) = \mathbf{w}^T\mathbf{r}(k) + \mathbf{b}^T\hat{\mathbf{s}}_b(k) \quad (19)$$

where $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1)\cdots\hat{s}(k-d-n_b)]^T$ is the past detected symbol vector with n_b being the feedback order, and $\mathbf{b} = [b_1\cdots b_{n_b}]^T$ the feedback filter coefficient vector. We will choose $d = n_h - 1$, $m = n_h$ and $n_b = n_h - 1$, as this choice of the DFF structure parameters is sufficient to guarantee linear separability. Define $\mathbf{s}_f(k) = [s(k)\cdots s(k-d)]^T$ and partition the CIR matrix $\mathbf{H} = [\mathbf{H}_1 \mid \mathbf{H}_2]$, where $\mathbf{H}_1$ has a dimension of $m\times(d+1)$ and $\mathbf{H}_2$ a dimension of $m\times n_b$. Under the assumption that the past decisions are correct, that is, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k) = [s(k-d-1)\cdots s(k-d-n_b)]^T$, the received signal vector can be expressed as

$$\mathbf{r}(k) = \mathbf{H}_1\mathbf{s}_f(k) + \mathbf{H}_2\hat{\mathbf{s}}_b(k) + \mathbf{n}(k)\,. \quad (20)$$

Thus, the decision feedback can be viewed to translate the original observation space $\mathbf{r}(k)$ into a new space $\mathbf{r}^{'}(k)$:

$$\mathbf{r}^{'}(k) \triangleq \mathbf{r}(k) - \mathbf{H}_2\hat{\mathbf{s}}_b(k)\,. \quad (21)$$

In this translated observation space, the DFE (19) becomes a "linear equaliser":

$$y(k) = \mathbf{w}^T\mathbf{r}^{'}(k) = \tilde{y}(k) + e(k)\,. \quad (22)$$

Notice that the feedback filter coefficients do not disappear. They in fact have been set to their optimal values, which are the related channel taps. All the results for the linear equaliser are readily applicable. The SER expression (11) gives the lower-bound of the SER for the DFE with the weight vector $\mathbf{w}$, under the assumption of correct symbols being fed back.

We finally turn to the complex-valued channel. For $M = L^2$, the M-QAM symbol set is defined by

$$\mathcal{S} \triangleq \{s_{l,q} = u_l + ju_q,\ \ 1\le l,q\le L\} \quad (23)$$

with $u_l = 2l-L-1$ and $u_q = 2q-L-1$. As $c_d = c_{R_d} + jc_{I_d}$ generally involves a rotation of the symbol set, it is desired to perform a de-rotation of $\mathbf{w}$

$$\mathbf{w} = \frac{|c_d|\,\mathbf{w}}{c_d} \quad (24)$$

so that $c_{I_d} = 0$. With this measure, the optimal decision rule for $\hat{s}_R(k-d)$ is according to

$$\hat{s}_R(k-d) = \begin{cases} u_1, & \text{if } y_R(k) \le (u_1+1)c_{R_d}\,, \\ u_l, & \text{if } (u_l-1)c_{R_d} < y_R(k) \le (u_l+1)c_{R_d} \\ & \text{for } l = 2,\cdots L-1\,, \\ u_L, & \text{if } y_R(k) > (u_L-1)c_{R_d} \end{cases} \quad (25)$$

and a similar rule is used for $\hat{s}_I(k-d)$. Furthermore, the SER is given by

$$P_E(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}) - P_{E_R}(\mathbf{w})P_{E_I}(\mathbf{w}) \quad (26)$$

where $P_{E_R}(\mathbf{w}) = \text{Prob}\{\hat{s}_R(k-d) \neq s_R(k-d)\}$ and $P_{E_I}(\mathbf{w}) = \text{Prob}\{\hat{s}_I(k-d) \neq s_I(k-d)\}$ can similarly be derived based on the PAM result. The MSER is defined as the solution that minimizes the upper-bound of the SER

$$P_{E_B}(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}) \quad (27)$$

and the adaptive MSER algorithm can similarly be derived.

Example 1: The channel is $H(z) = 1.0 + 0.5z^{-1}$ with 4-PAM, the linear equaliser has $m = 2$ and $d = 0$.

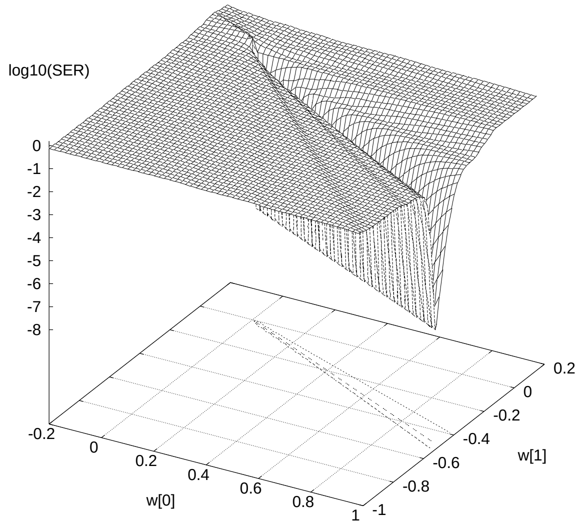

Fig. 1. The SER surface for Example 1 with SNR$= 35$ dB. The MMSE solution $\mathbf{w}_{\text{MMSE}} = [0.929 \ -0.371]^T$, which has been normalized, with $\log_{10} P_E(\mathbf{w}_{\text{MMSE}}) = -2.7593$. The MSER solution $\mathbf{w}_{\text{MSER}} = \alpha[0.896 \ -0.445]^T$, $\alpha > 0$, with $\log_{10} P_E(\mathbf{w}_{\text{MSER}}) = -7.1566$.

Example 2: The channel is $H(z) = 0.3 + 1.0z^{-1} - 0.3z^{-2}$ with 8-PAM, the DFE has $m = 3$, $d = 2$ and $n_b = 2$.

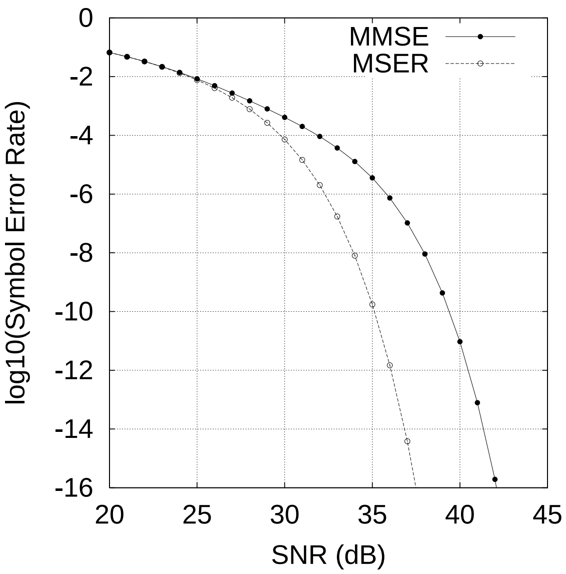

Fig. 2. Lower-bound symbol error rate comparison for Example 2, assuming correct symbols being fad back.

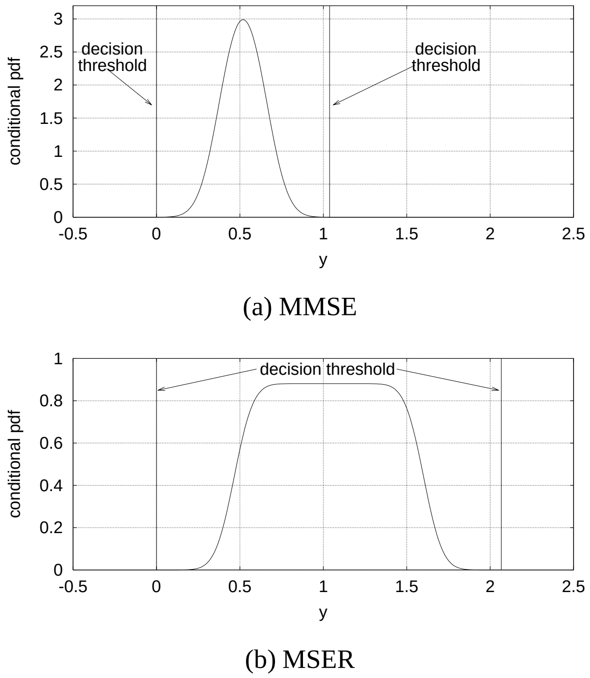

Fig. 3. Conditional PDF of the equaliser output given $s(k-d) = 1$ for Example 2 with SNR$= 34$ dB. The weight vector is normalized to a unit length.

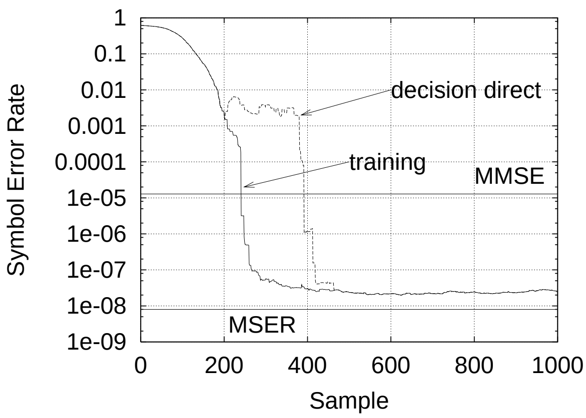

Fig. 4. Learning curves of the LSER for Example 2 with SNR$= 34$ dB, averaged over 100 runs. Initial $\mathbf{w} = [-0.01 \ 0.01 \ 0.01]^T$. Dashed curve: after 200-sample training, switched to decision-directed adaptation with $\hat{s}(k-d)$ substituting $s(k-d)$.