Minimum Symbol-Error-Rate Equalisation

Sheng Chen

Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, U.K.
E-mail: sqc@ecs.soton.ac.uk

Presented at Non-linear and Non-Gaussian Signal Processing Workshop
Peebles Hydro, Scotland, July 8-9 2002

Motivations

Linear filtering at receiver (equalisation, multiuser detection, beam-forming)

$$y(k) = \bar{y}(k) + e(k)$$

Filter output generally non-Gaussian (sum of Gaussian distributions)

- Zero forcing: Gaussian but noise enhancement too serious
- MMSE: classically regarded as optimal $\rightarrow$ non-optimal in terms of symbol error rate!
- Adopt to non-Gaussian view naturally leads to MSER approach

Some Previous Works

- "MBER" almost as old as "adaptive equalisation"

  Yao, *IEEE Trans. Information Theory* 1972, Shamash & Yao, *ICC'74*

- More recently

  Chen *et al.*, *ICC'96*, *IEE Proc. Communications* 1998;

  Yeh & Barry, *ICC'97*, *IEEE Trans. Communications* 2000;

  Mulgrew & Chen, *IEEE Symp. ASSPCC* 2000, *Signal Processing* 2001

Mostly on binary modulation (BPSK)

⋆ This work: multilevel modulation schemes (PAM and QAM)

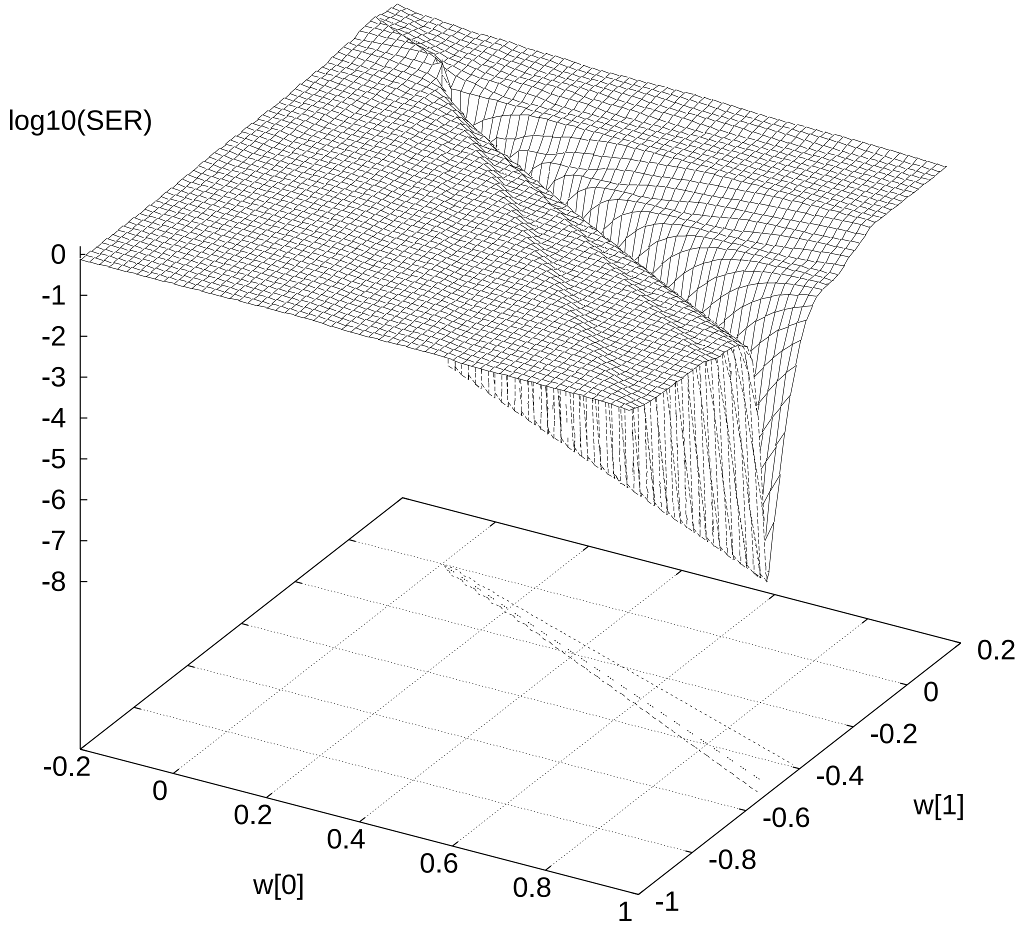
A Toy Example

Two-tap channel $1.0 + 0.5z^{-1}$
with 4-PAM and SNR$= 35$ dB

Two-tap $m = 2$ linear equaliser
with decision delay $d = 0$

Normalized MMSE:
$\mathbf{w}_{\mathrm{MMSE}}^T = [0.9285 \quad - 0.3713]$
with $\log_{10}(\mathrm{SER}) = -2.7593$

MSER ($\alpha > 0$):
$\mathbf{w}_{\mathrm{MSER}}^T = \alpha[0.8957 \quad - 0.4447]$
with $\log_{10}(\mathrm{SER}) = -7.1566$



- MSER solutions form a half line, origin is singular point

Real-Valued L-PAM Channel

- Channel of length n_h

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

$$s(k) \in \mathcal{S} \stackrel{\triangle}{=} \{s_l = 2l - L - 1, 1 \le l \le L\}$$

- Linear equaliser of order m

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

$\mathbf{r}(k) = [r(k) \cdots r(k-m+1)]^T$, $\mathbf{w} = [w_0 \cdots w_{m-1}]^T$, and decision delay d

$\star$ $e(k)$: Gaussian with zero mean and variance $\sigma_n^2 \mathbf{w}^T \mathbf{w}$, σ_n^2 being variance of $n(k)$

$\star$ $\bar{y}(k) \in \mathcal{Y} \stackrel{\triangle}{=} \{\bar{y}_q, 1 \le q \le N_s = L^{m+n_h-1}\}$, which can be divided into L subsets

$$\mathcal{Y}_l \stackrel{\triangle}{=} \{\bar{y}_q^{(l)} \in \mathcal{Y} | s(k-d) = s_l\}, 1 \le l \le L$$

Observation vector

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

- Combined impulse response $\mathbf{c}^T = \mathbf{w}^T\mathbf{H} = [c_0 \; c_1 \cdots c_{m+n_h-2}]$. Then

$$y(k) = c_d s(k-d) + \sum_{i \neq d} c_i s(k-i) + e(k)$$

- Optimal decision making

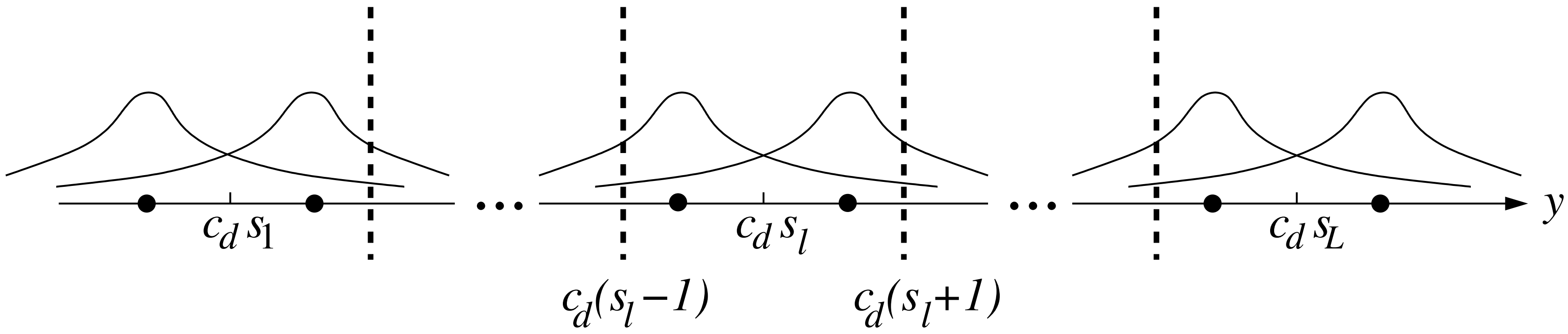
$$\hat{s}(k-d) = \begin{cases} s_1, & \text{if } \; y(k) \leq (s_1+1)c_d \, , \\ s_l, & \text{if } \; (s_l-1)c_d < y(k) \leq (s_l+1)c_d \\ & \quad \text{for } l = 2, \cdots L-1 \, , \\ s_L, & \text{if } \; y(k) > (s_L-1)c_d \, . \end{cases}$$

Unlike binary case, main tap of combined impulse response c_d needed!

Two Useful Properties

- Shifting: $\mathcal{Y}_{l+1} = \mathcal{Y}_l + 2c_d$

- Symmetry: distribution of $\mathcal{Y}_l$ is symmetric around $c_d s_l$.



⋆ For linear equaliser to work, $\mathcal{Y}_l$, $1 \leq l \leq L$, must be *linearly separable*

This is not guaranteed

⋆ In DFE, linear separability is guaranteed

SER Expression

PDF of $y(k)$

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

where $N_{sb} = N_s/L$ is number of points in $\mathcal{Y}_l$ and $\bar{y}_i^{(l)} \in \mathcal{Y}_l$.

Utilizing shifting and symmetric properties, SER of equaliser $\mathbf{w}$ is:

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

where Q is usual Q-function, $\gamma = 2(L-1)/L$, and

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

MSER Solution

◯ MSER solution is defined as:

$$\mathbf{w}_{\text{MSER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w})$$

- Use simplified conjugated gradient algorithm with periodically reseting search direction to negative gradient
- Computation is on single subset $\mathcal{Y}_l$, and further simplification by using $\mathcal{Y}_l$ with $s_l = 1$
- SER is invariant to a positive scaling of $\mathbf{w}$, computationally advantageous to normalize weight vector to $\mathbf{w}^T\mathbf{w} = 1$.

◯ Readily extend to DFE (with lower bound SER)

◯ Block adaptation: identify channel or Parzen window estimate of PDF $p_y(x) \longrightarrow P_E(\mathbf{w}) \longrightarrow$ optimisation

Stochastic Adaptation

Single-sample estimate of $p_y(x)$

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

Using instantaneous stochastic gradient $\longrightarrow$ LSER:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\gamma}{\sqrt{2\pi}\rho_n} \exp\left(-\frac{\left(y(k) - \hat{c}_d(k)(s(k-d)-1)\right)^2}{2\rho_n^2}\right) \times$$

$$\Big(\mathbf{r}(k) - (s(k-d)-1)\hat{\mathbf{h}}_d(k)\Big)$$

$\hat{c}_d(k)$: estimate of c_d

$\hat{\mathbf{h}}_d(k)$: estimate of d-th column of $\mathbf{H}$

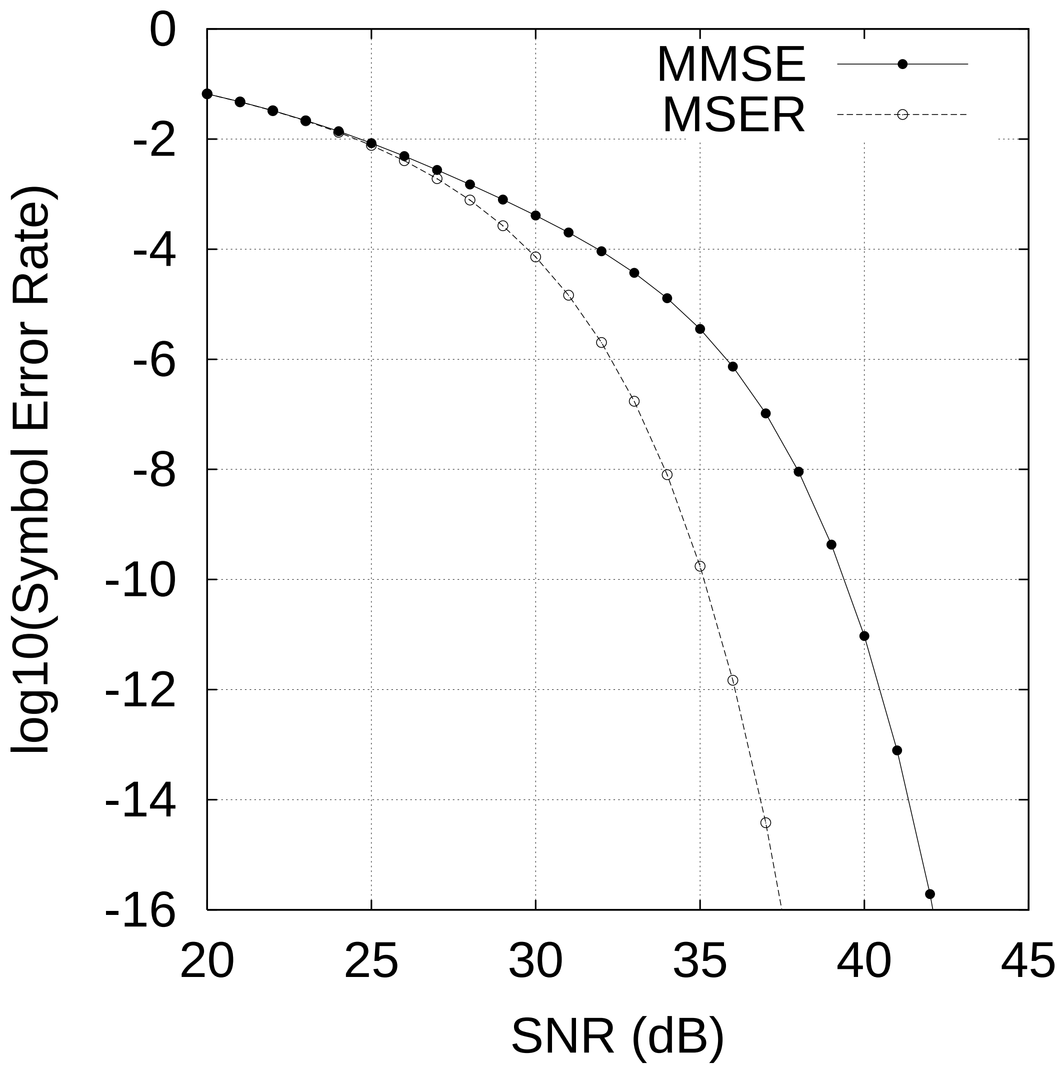
Adaptive gain μ and kernel width ρ_n need to be set appropriately

An 8-PAM DFE Example

- Lower-Bound SER Comparison

Channel:
$0.3 + 1.0z^{-1} - 0.3z^{-2}$
with 8-PAM

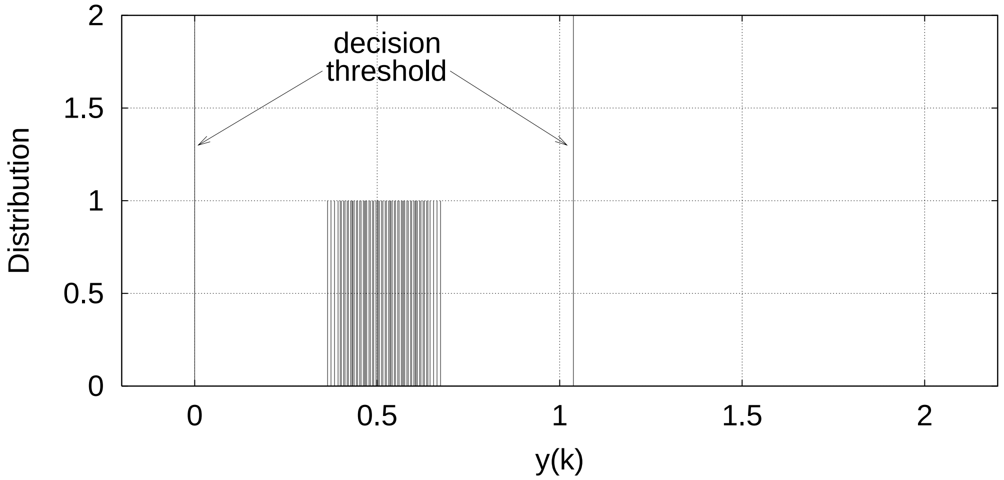
DFE:
$m = 3,\ d = 2,\ n_b = 2$



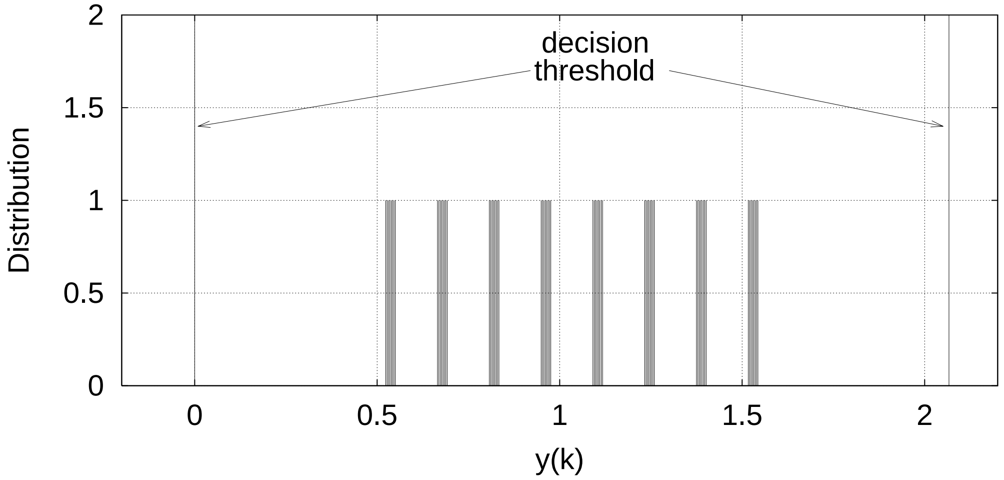
- Distribution of Subset $\mathcal{Y}_5$ $(s_5 = 1)$, 64 points, SNR$=34$ dB

Weight vector has been normalized to a unit length, a point plotted as a unit impulse.

(a) MMSE

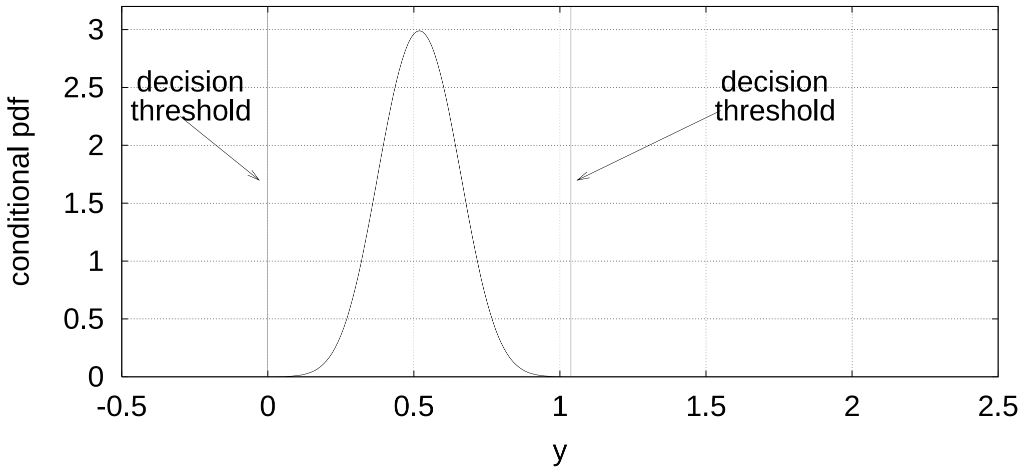


(b) MSER

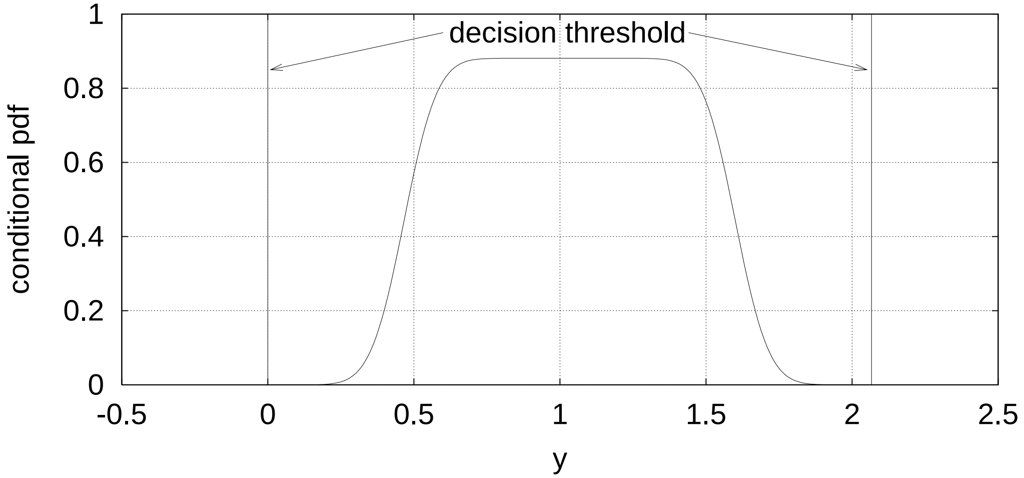


- Conditional PDF given $s(k-d)=1$, SNR=34 dB. Weight vector normalized, $\mathbf{w}_{\text{MMSE}}^T = [-0.0578\ 0.2085\ 0.9763]$, $\mathbf{w}_{\text{MSER}}^T = [-0.2365\ 0.7946\ 0.5592]$

(a) MMSE

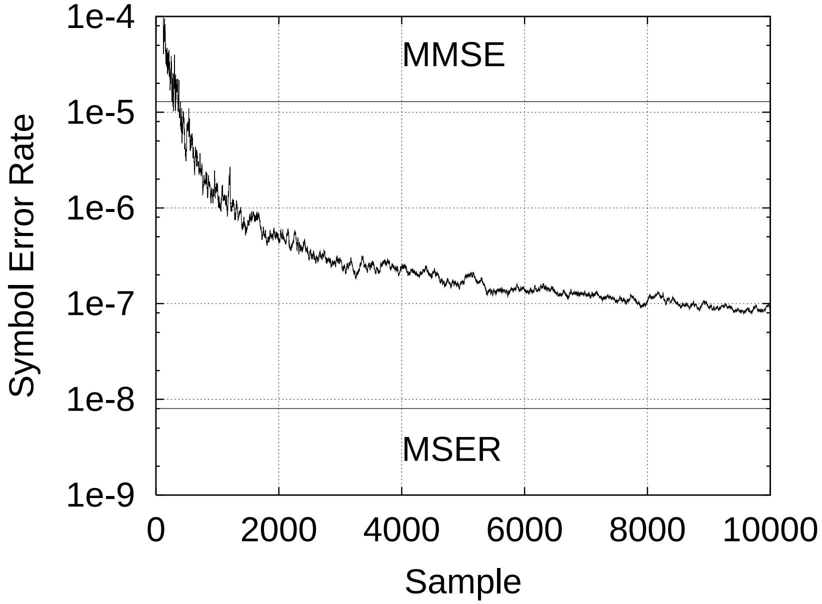


(b) MSER

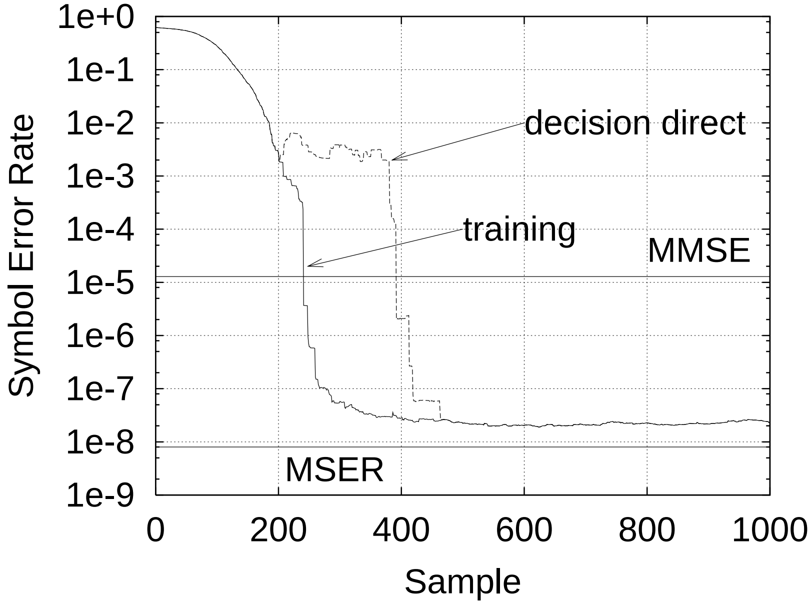


- Learning Curves of LSER Averaged Over 100 Runs, SNR$=34$ dB

Initial weight: (a) $\mathbf{w}_{\text{MMSE}}$, (b) $[-0.01\ 0.01\ 0.01]^T$



(a)



(b)

In (a) training and decision directed indistinguishable, in (b) dashed curve: after 200-sample training, switched to decision-directed with $\hat{s}(k-d)$ substituting $s(k-d)$

Initial value is critical for convergence, MMSE not necessarily good initial choice

Complex-Valued Channel and QAM

Recall $\mathbf{c} = \mathbf{w}^T \mathbf{H}$,

$$y(k) = c_d s(k-d) + \sum_{i \neq d} c_i s(k-i) + e(k)$$

Generally $c_d = c_{R_d} + j c_{I_d}$, advisable always perform a rotation to be sure $c_{I_d} = 0$:

$$\mathbf{w} = \frac{|c_d| \mathbf{w}}{c_d}$$

- Then I and Q decisions decoupled, real and imaginary parts are PAMs and

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

- MSER defined as solution that minimises upper-bound SER:

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

Conclusions

- MSER equalisation solution for high-level modulation schemes
  - Effective sample-by-sample adaptation has been developed
  - Unlike MSE surface which is quadratic, SER surface is highly complex
  - Initial equaliser weight values can influence convergence speed

- Generalization to adaptive filtering at various communication receivers
  - Traditional MMSE filtering sub-optimal because conditional filter output is non-Gaussian
  - Proposed approach is based on non-Gaussian nature and can be referred to as adaptive minimum error rate filtering