Stochastic Least-Symbol-Error-Rate Adaptive Equalization for Pulse-Amplitude Modulation

Sheng Chen[†], Bernard Mulgrew[‡] and Ljaos Hanzo[†]

[†] Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.
sqc@ecs.soton.ac.uk lh@ecs.soton.ac.uk

[‡] Department of Electronics and Electrical Engineering
University of Edinburgh, Edinburgh EH9 3JL, U.K.
Bernie.Mulgrew@ee.ed.ac.uk

Presented at ICASSP2002, May 13-17, 2002, Orlando, USA

Support of U.K. Royal Academy of Engineering under an international travel grant (IJB/AH/ITG 02-011) is gratefully acknowledged

Motivations

Equalization topic is well researched, and a variety of solutions exists. BUT

- For high-level modulation, MAP/MLSE sequence detector too complex

  Even MAP or Bayesian symbol-detector too complex

- Affordable: linear equalizer and decision feedback equalizer

  Classically, MMSE solution is regarded as optimum

  MMSE would be optimum only if equalizer soft output were Gaussian

$\star$ Adopting to non-Gaussian nature leads to optimal MSER solution for linear equalizer and DFE

Some Previous Works

- Yao, *IEEE Trans. Information Theory* 1972
- Shamash & Yao, *ICC'74*
- Chen *et al.*, *ICC'96*, *IEE Proc. Communications* 1998
- Yeh & Barry, *ICC'97*, *ICC'98*⋆, *IEEE Trans. Communications* 2000
- Chen & Mulgrew, *IEE Proc. Communications* 1999⋆
- Mulgrew & Chen, *IEEE Symp. ASSPCC* 2000, *Signal Processing* 2001

⋆: for multi-level PAM schemes

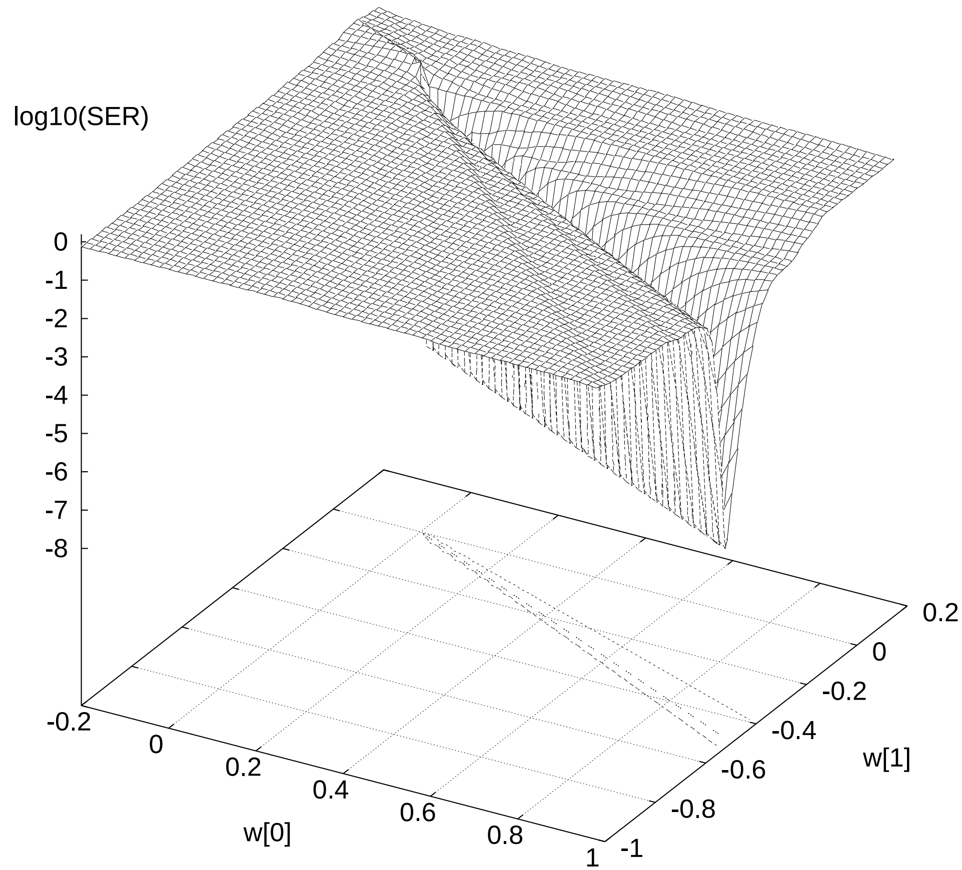
A Toy Example

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

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



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

PAM Channel Model

- Channel of length n_h

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

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

- Linear equaliser of order m

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

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

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

As $\mathbf{s}(k) \in \{\mathbf{s}_q, 1 \leq q \leq N_s\}$ where $N_s = L^{m+n_h-1}$,

$$\bar{\mathbf{r}}(k) \in \mathcal{R} \stackrel{\triangle}{=} \{\bar{\mathbf{r}}_q = \mathbf{H}\mathbf{s}_q, 1 \leq q \leq N_s\}$$

- Express equaliser output

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

  ⋆ $e(k)$: Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$

  ⋆ $\bar{y}(k) \in \mathcal{Y} \triangleq \{\bar{y}_q = \mathbf{w}^T\bar{\mathbf{r}}_q, 1 \leq q \leq N_s\}$, which can be divided into M subsets

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

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

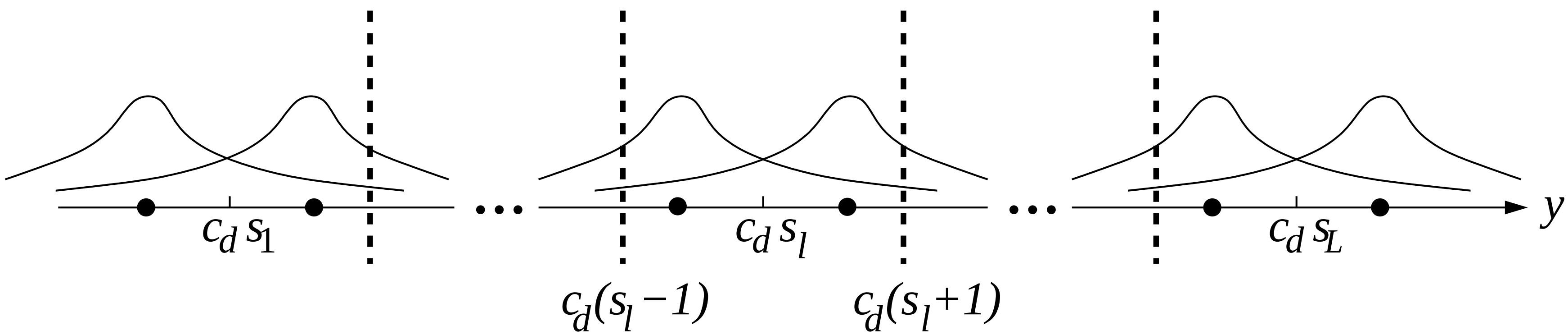
$$y(k) = c_d s(k-d) + \sum_{i \neq d} c_i s(k-i) + e(k)$$

- Optimal decision making

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

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:

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

Gradient of $P_E(\mathbf{w})$

$$\nabla P_E(\mathbf{w}) = \frac{\gamma}{\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{N_{sb}}\sum_{i=1}^{N_{sb}} \exp\left(-\frac{(\bar{y}_i^{(l)} - c_d(s_l-1))^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \times$$

$$\left(\frac{(\bar{y}_i^{(l)} - c_d(s_l-1))}{\mathbf{w}^T\mathbf{w}}\mathbf{w} - \bar{\mathbf{r}}_i^{(l)} + (s_l-1)\mathbf{h}_d\right)$$

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

Block Adaptation

- Identify channel $\longrightarrow P_E(\mathbf{w}) \longrightarrow$ optimisation
- Alternatively, kernel density or Parzen window estimate approach

An estimated PDF of $p_y(x)$

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

K: sample length, and ρ_n: radius parameter. From $\hat{p}_y(x)$, estimated SER

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

where

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

$\hat{c}_d = \mathbf{w}^T\hat{\mathbf{h}}_d$, and $\hat{\mathbf{h}}_d$ an estimate for the d-th column $\mathbf{h}_d$ of $\mathbf{H}$

Sample-by-Sample Adaptation: LSER

Single-sample estimate of $p_y(x)$

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

With a re-scaling after each update to ensure $\mathbf{w}^T\mathbf{w}=1$, and using instantaneous stochastic gradient, $\longrightarrow$ LSER:

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

$$\left(\mathbf{r}(k) - (y(k)-\hat{c}_d(k)(s(k-d)-1))\mathbf{w}(k) - (s(k-d)-1)\hat{\mathbf{h}}_d(k)\right)$$

$$\mathbf{w}(k+1) = \frac{\mathbf{w}(k+1)}{\sqrt{\mathbf{w}^T(k+1)\mathbf{w}(k+1)}}$$

Adaptive gain μ and width ρ_n need to be set appropriately

Sample-by-Sample Adaptation: ALSER

Single-sample estimate of $p_y(x)$

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

Using instantaneous stochastic gradient, $\longrightarrow$ ALSER:

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

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

⋆ No need for normalization to simplify complexity

⋆ Although using ρ_n rather than $\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$ appears to involve more approximation, ALSER seems to work well — not restrict to unit length makes it easier to converge to a MSER

Extension to DFE

"Linear-combiner" DFE:

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

where $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n_b)]^T$ and $\mathbf{b} = [b_1 \cdots b_{n_b}]^T$

- Choose $d = n_h - 1$, $m = n_h$ and $n_b = n_h - 1$
- Define $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$ and partition $\mathbf{H} = [\mathbf{H}_1 \mid \mathbf{H}_2]$

Under assumption $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n_b)]^T$,

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

Define translated observation space

$$\mathbf{r}'(k) \stackrel{\triangle}{=} \mathbf{r}(k) - \mathbf{H}_2 \hat{\mathbf{s}}_b(k) = \tilde{\mathbf{r}}(k) + \mathbf{n}(k)$$

DFE becomes a "linear equaliser":

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

⋆ Feedback filter coefficients do not disappear. They have been set to their optimal values.

As $\mathbf{s}_f(k) \in \{\mathbf{s}_{f,q}, 1 \le q \le N_f\}$ with $N_f = L^{d+1}$

$$\tilde{\mathbf{r}}(k) \in \tilde{\mathcal{R}} \stackrel{\triangle}{=} \{\tilde{\mathbf{r}}_q = \mathbf{H}_1 \mathbf{s}_{f,q}, 1 \le q \le N_f\}$$

$\tilde{y}(k) \in \tilde{\mathcal{Y}} \stackrel{\triangle}{=} \{\tilde{y}_q = \mathbf{w}^T \tilde{\mathbf{r}}_q, 1 \le q \le N_f\}$ which can be partitioned into L subsets

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

⋆ $\tilde{\mathcal{Y}}_l$ are always linearly separable. All results of linear equaliser are applicable.

Lower bound SER for DFE $\mathbf{w}$ under assumption of correct symbol feedback

$$P_E(\mathbf{w}) = \frac{\gamma}{N_{fsb}} \sum_{i=1}^{N_{fsb}} Q(\tilde{g}_{l,i}(\mathbf{w}))$$

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

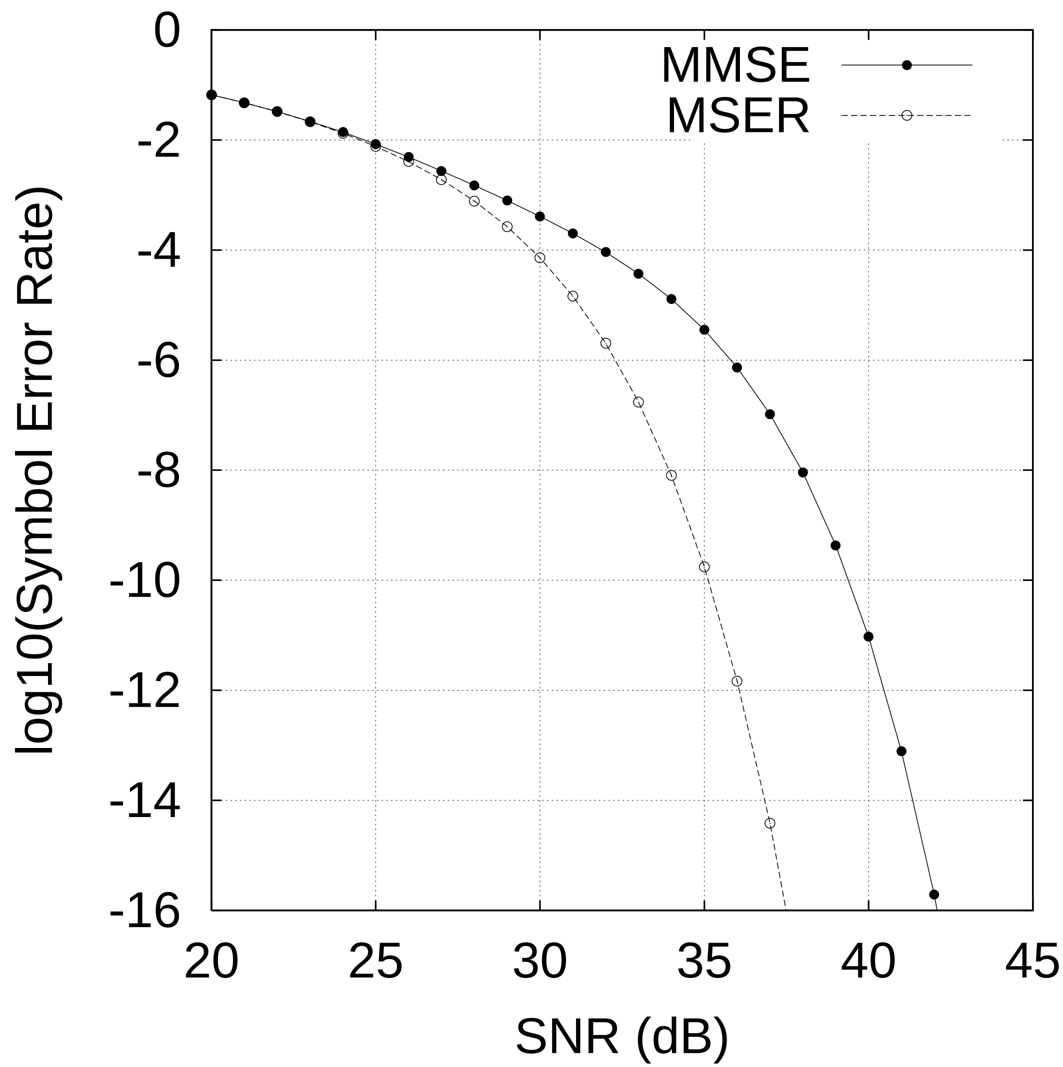
$\tilde{y}_i^{(l)} \in \tilde{\mathcal{Y}}_l$, and $N_{fsb} = N_f / L = L^d$ is number of points in $\tilde{\mathcal{Y}}_l$

An 8-PAM DFE Example

- Lower-Bound SER Comparison

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

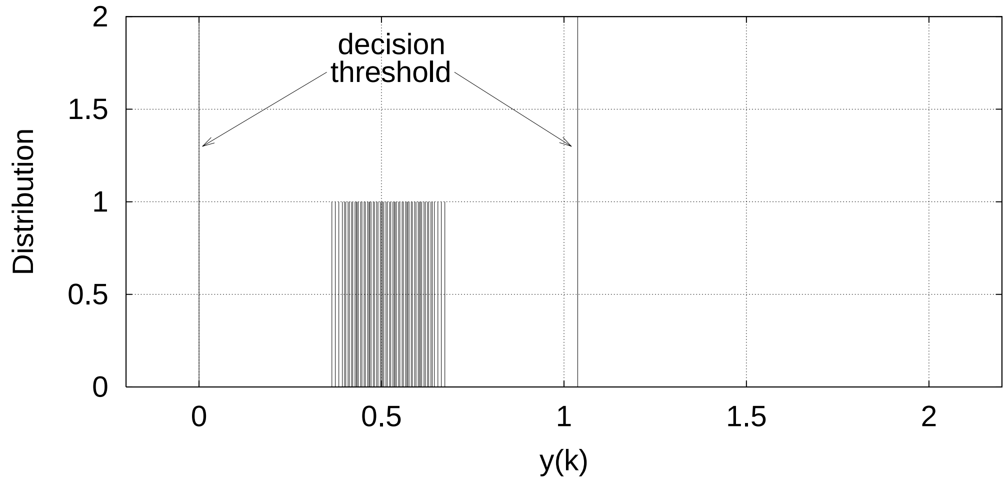
DFE:
$m = 3,\ d = 2,\ n_b = 2$



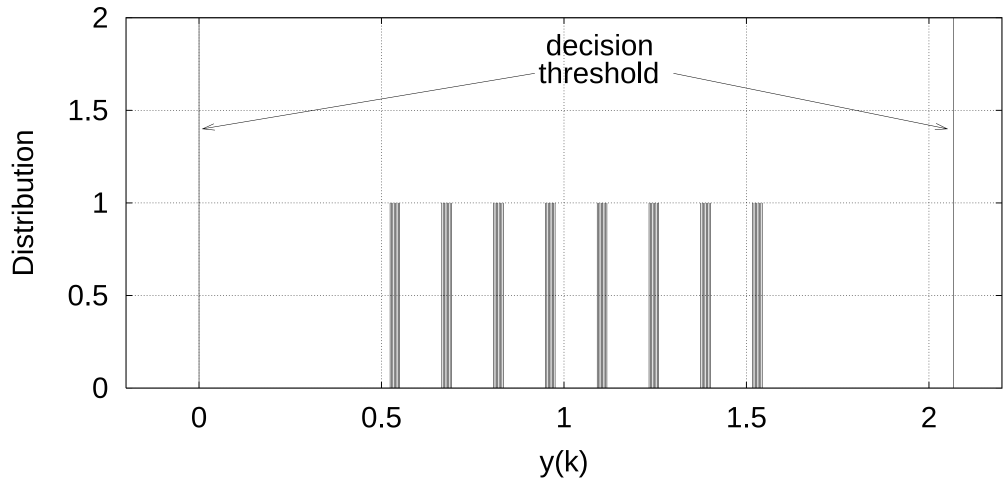
- Distribution of Subset $\tilde{\mathcal{Y}}_5$ $(s_5 = 1)$, 64 points, SNR$=34$ dB

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

(a) MMSE



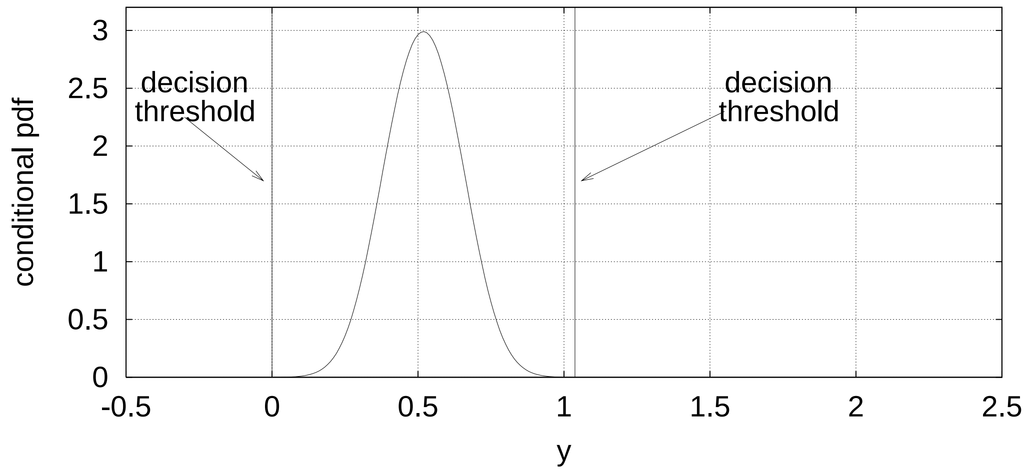
(b) MSER



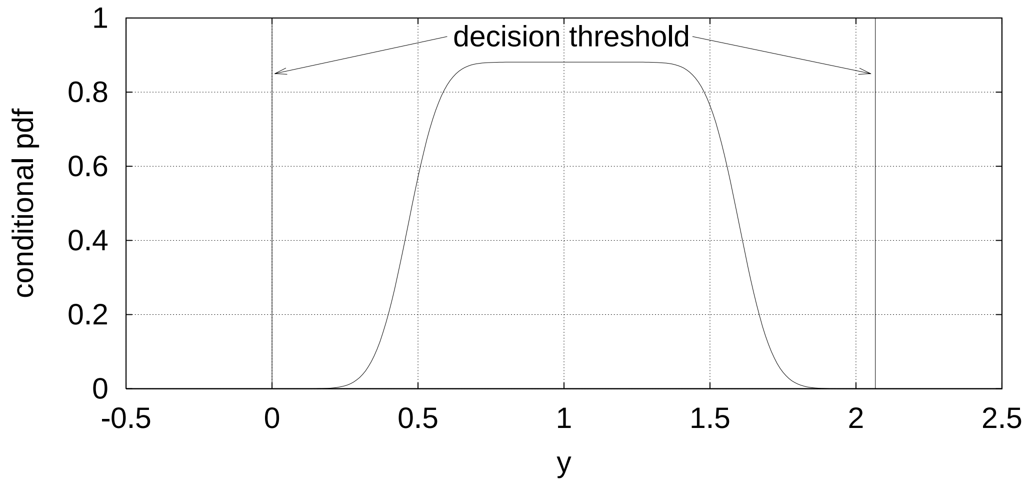
- Conditional PDF given $s(k-d) = 1$, SNR=34 dB

normalized $\mathbf{w}_{\mathrm{MMSE}}^T = [-0.0578\ 0.2085\ 0.9763]$, $\mathbf{w}_{\mathrm{MSER}}^T = [-0.2365\ 0.7946\ 0.5592]$

(a) MMSE

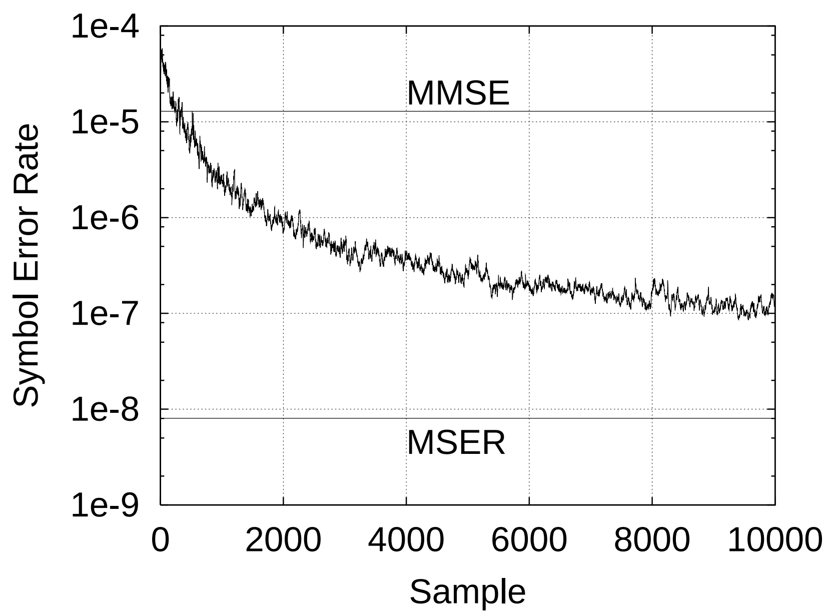


(b) MSER

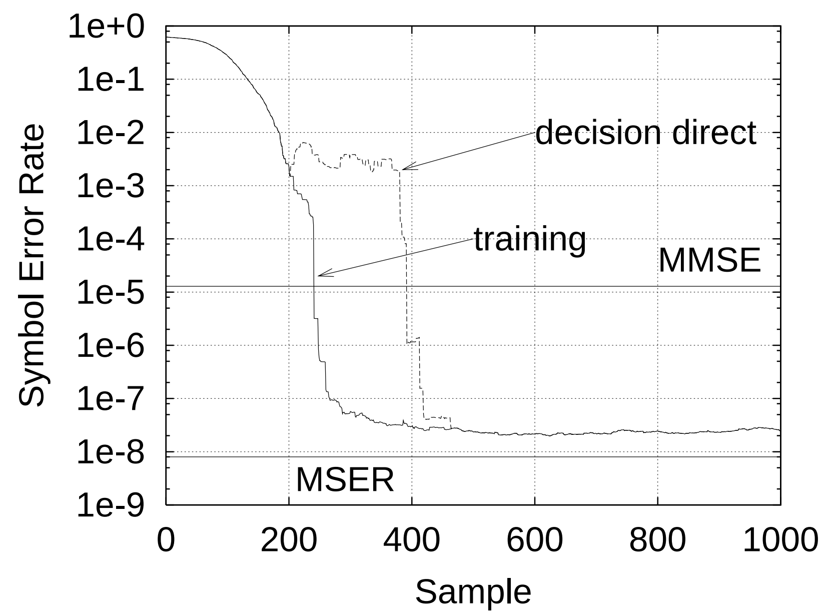


• Learning Curves of **LSER** Averaged Over 100 Runs, SNR$=34$ dB

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



(a)



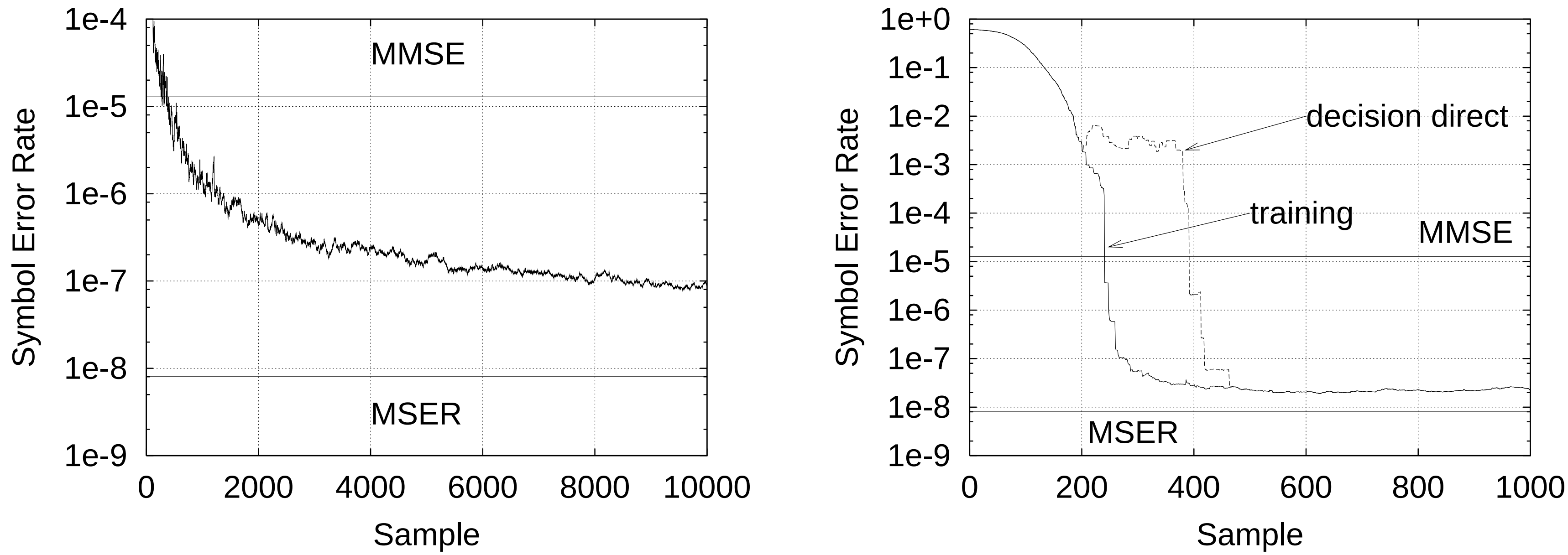
(b)

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

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

Initial weight: (a) $\mathbf{w}_{\text{MMSE}}$, (b) $[-0.01\ 0.01\ 0.01]^T$ Weight normalization not applied



(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 $s(k-d)$

Compared with LSER, no performance degradation, much simpler

Conclusions

- MMSE views equaliser output as Gaussian and tries to fit parameters to true non-Gaussian PDF in a way so that it looks as closely as possible to a Gaussian one (without making noise sky high)
- Non-Gaussian approach leads naturally to MSER
- For high-level PAM modulation schemes, MSER equalisation solution has being derived

  Effective sample-by-sample adaptation has been developed

  Unlike MSE surface which is quadratic, SER surface is highly complex

  Initial equaliser weight values can critically influence convergence speed

  ALSER is particular promising: simpler computation