Introduction to Adaptive Equalization

Arnaldo Spalvieri

Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, ITALY

MSE Linear Equalizer in FIR Form

Let the received sequence be

$$y_k = \sum_{i=-\infty}^{\infty} g_i a_{k-i} + w_k,$$

where $\{a_i\}$ is i.i.d. with zero mean and variance σ_a^2 and $\{w_k\}$ is AWGN with variance σ_w^2. At the receive side, we aim at recoverying the sequence

$$a_{k-d},$$

where d is the decision delay. We therefore look for the 1-causal FIR filter $\mathbf{c} = (c_1, c_2, \cdots c_n)^T$ that minimizes the mean square error:

$$J(\mathbf{c}) = E\{(x_k - a_{k-d})^2\}, \qquad (1)$$

where

$$x_k = \sum_{i=1}^{n} = c_i y_{k-i}.$$

MSE Linear Equalizer in FIR Form

The partial derivative of the cost function w.r.t. c_j is

$$
\begin{aligned}
\frac{\partial J(\mathbf{c})}{\partial c_j} &= E\left\{\frac{\partial (x_k - a_{k-d})^2}{\partial c_j}\right\} = E\left\{(x_k - a_{k-d}) y_{k-j}\right\} \\
&= E\left\{\left(\sum_{i=1}^{n} c_i y_{k-i} - a_{k-d}\right) y_{k-j}\right\} \\
&= \sum_{i=1}^{n} c_i E\left\{y_{k-i} y_{k-j}\right\} - E\left\{a_{k-d} y_{k-j}\right\} \\
&= \sum_{i=1}^{n} c_i \psi_{i-j} - \sigma_a^2 g_{d-j},
\end{aligned}
$$

where

$$
\psi_{i-j} = E\{y_k y_{k+i-j}\} = \sigma_a^2 \sum_{k=-\infty}^{\infty} g_k g_{k+i-j} + \sigma_w^2 \delta_{i,j}.
$$

MSE Linear Equalizer in FIR Form

It is easily seen that the unique extremant of the quadratic form $J(\mathbf{c})$ is the global minimum, which is found by setting to zero the gradient:

$$\nabla_{\mathbf{c}} J(\mathbf{c}_{opt}) = 2\mathbf{\Psi}\mathbf{c}_{opt} - 2\sigma_a^2 \mathbf{g}^- = \mathbf{0}, \quad \mathbf{c}_{opt} = \sigma_a^2 \mathbf{\Psi}^{-1}\mathbf{g}^-,$$

where $\mathbf{\Psi}$ is the autocorrelation matrix of y,

$$\mathbf{\Psi} = \begin{pmatrix} \psi_0 & \psi_{-1} & \cdots & \psi_{1-n} \\ \psi_1 & \psi_0 & \cdots & \psi_{2-n} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{n-1} & \psi_{n-2} & \cdots & \psi_0 \end{pmatrix}$$

and

$$\sigma_a^2 \mathbf{g}^- = \sigma_a^2 (g_{d-1}, g_{d-2}, \cdots, g_{d-n}).$$

Properties of the Autocorrelation Matrix

The autocorrelation matrix has the following properties.

- $\boldsymbol{\Psi}$ is Toeplitz, meaning that all the elements along the same diagonal have the same value. A very special feature of an $n \cdot n$ Toeplitz matrix is that its inverse can be computed with complexity $\mathcal{O}(n^2)$ (Rybicki, see Numerical Recipes in C, chapter 2).
- Being the autocorrelation matrix Toeplitz and being the autocorrelation even ($\psi_k = \psi_{-k}$), $\boldsymbol{\Psi}^T = \boldsymbol{\Psi}$.
- $\boldsymbol{\Psi}$ is positive semidefinite: $\mathbf{x}^T \boldsymbol{\Psi} \mathbf{x} \geq 0, \ \forall\, \mathbf{x}$.
- When $\boldsymbol{\Psi}$ is positive definite, its inverse $\boldsymbol{\Psi}^{-1}$ is positive definite.
- When $\boldsymbol{\Psi}$ is positive definite, $\boldsymbol{\Psi}^{-1} = (\boldsymbol{\Psi}^{-1})^T$, but $\boldsymbol{\Psi}^{-1}$ is not Toeplitz.

Minimum MSE

By computing the expectation (1), one finds that the mean square error is

$$J(\mathbf{c}) = \mathbf{c}^T \boldsymbol{\Psi} \mathbf{c} + \sigma_a^2 - 2\sigma_a^2 \mathbf{c}^T \mathbf{g}^-$$

therefore, substituting the optimal coefficient's vector, one finds that the minimum mean square error is

$$\begin{aligned} J(\mathbf{c}_{opt}) &= \sigma_a^2 (\mathbf{g}^-)^T \boldsymbol{\Psi}^{-1} \mathbf{g}^- \sigma_a^2 + \sigma_a^2 - 2\sigma_a^4 (\mathbf{g}^-)^T \boldsymbol{\Psi}^{-1} \mathbf{g}^- \\ &= \sigma_a^2 - \sigma_a^4 (\mathbf{g}^-)^T \boldsymbol{\Psi}^{-1} \mathbf{g}^- \leq \sigma_a^2. \end{aligned}$$

Equality holds when $\mathbf{c}_{opt} = \mathbf{0}$, that is when the AWGN is so strong that the optimum is achieved by putting to zero the output.

Gradient Algorithm

The optimal set of coefficients can be found by iteratively moving a small step in the direction of the gradient descent

$$\mathbf{c}_{j+1} = \mathbf{c}_j - \frac{\gamma}{2}\nabla_{\mathbf{c}} J(\mathbf{c}_j) = \mathbf{c}_j - \gamma(\boldsymbol{\Psi}\mathbf{c}_j - \sigma_a^2 \mathbf{g}^-), \;\; j = 0, 1, \cdots ,$$

where $\gamma > 0$ is a small real number called *step size*, and $\mathbf{c}_j$ is the coefficient's vector at the j-th iteration. With this algorithm, inversion of the autocorrelation matrix is not necessary. Caution should be taken in the choice of γ to guarantee convergence of the algorithm.

Orthogonality Principle

The gradient can be written as

$$0.5 \cdot \nabla_{\mathbf{c}} J(\mathbf{c}) = \mathbf{\Psi}_y \mathbf{c} - \sigma_a^2 \mathbf{g}^- = E\{\mathbf{y}_k(\mathbf{y}_k^T \mathbf{c} - a_{k-d})\}$$
$$= E\{\mathbf{y}_k(x_k - a_{k-d})\},$$

where

$$\mathbf{y}_k = (y_{k-1}, y_{k-2}, \cdots, y_{k-n})^T.$$

For optimal $\mathbf{c}$ one has

$$\nabla_{\mathbf{c}} J(\mathbf{c}_{opt}) = 2 \cdot E\{\mathbf{y}_k(x_k - a_{k-d})\} = \mathbf{0},$$

which shows that the error $e_k = x_k - a_{k-d}$ is *orthogonal* to the received signal.

Stochastic Gradient Algorithm

An unbiased estimate of the expected value is obtained by a time average over N samples:

$$E\{\mathbf{y}_k(x_k - a_{k-d})\} = \frac{1}{N}\sum_{i=k_0}^{N+k_0-1} \mathbf{y}_i(x_i - a_{i-d}) + \boldsymbol{\eta}_{k_0},$$

where $\boldsymbol{\eta}_{k_0}$ is a random vector with zero mean. The most common implementation is with $N = 1$:

$$E\{\mathbf{y}_k(x_k - a_{k-d})\} = \mathbf{y}_k(x_k - a_{k-d}) + \boldsymbol{\eta}_k,$$

leading to the stochastic gradient algorithm

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \gamma\mathbf{y}_k(x_k - a_{k-d}).$$

Stochastic Gradient Algorithm

Note that the explicit knowledge of the channel's coefficients and of the autocorrelation matrix in not necessary. It has been substituted by the knowledge of the transmitted data sequence. One of the consequences of the random nature of the algorithm is that there is an excess MSE even when the algorithm has reached the steady state. Specifically, due to the presence of $\boldsymbol{\eta}_k$, at the steady-state, that is when the expectation inside the brackets in the last equation is zero, the coefficients wander around $\mathbf{c}_{opt}$. The step size γ regulates the compromise between the detrimental effect of wandering and the capability of the equalizer of tracking a time-varying channel.

Popular Variants: Polarity-type Algorithm

It is common to take $E\{|\Re\{x_k - a_{k-d}\}| + |\Im\{x_k - a_{k-d}\}|\}$ as a cost function, leading to the iterative algorithm

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \gamma \mathbf{y}_k \cdot csign(x_k - a_{k-d}),$$

where

$$csign(x) = sign(\Re\{x\}) + j \cdot sign(\Im\{x\}).$$

This simplifies the hardware, because the complex multiplication becomes a change of sign followed by a sum. An extreme simplification is obtained by taking only the complex sign of the complex product, leading to a so-called bang-bang algorithm.

Popular Variants: Blind Equalization

Commonly, the error is computed using decisions in place of data. However, during acquisition decisions may be not reliable. Moreover, also carrier recovery can be not at the steady state. One can minimize the phase independent cost function

$$\text{CMA}_{p,q}(\mathbf{c}) = E\{||x_k|^p - C|^q\},$$

where C is a constant that regulates the output amplitude. The most common implementation is with $p = 2$, $q = 1$, leading to

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \gamma \mathbf{y}_k x_k sign(|x_k|^2 - C).$$

It can be shown that, as $n \to \infty$, $N_0 \to 0$, the algorithm converges to the zero-forcing equalizer, with a decision delay that depends on the initialization.