Helsinki University of Technology
Signal Processing Laboratory

S-38.411 Signal Processing in Telecommunications I

Spring 2000

Lecture 7: Adaptive equalizers 2

Prof. Timo I. Laakso
timo.laakso@hut.fi, Tel. 451 2473
http://wooster.hut.fi/studies.html

Timetable

L1 Introduction; models for channels and communication systems
L2 Channel capacity
L3 Transmit and receive filters for bandlimited AWGN channels
L4 Optimal linear equalizers for linear channels 1
L5 Optimal linear equalizers for linear channels 2
L6 Adaptive equalizers 1
L7 Adaptive equalizers 2
L8 Nonlinear receivers 1: DFE equalizers
L9 Nonlinear receivers 2: Viterbi algorithm
L10 GL1: DSP for Fixed Networks / *Matti Lehtimäki, Nokia Networks*
L11 GL2: DSP for Digital Subscriber Lines / *Janne Väänänen, Tellabs*
L12 GL3: DSP for CDMA Mobile Systems / *Kari Kalliojärvi, NRC*
L13 Course review, questions, feedback
E 24.5. (Wed) 9-12 S4 **Exam**

Contents of Lecture 7

Helsinki University of Technology
Signal Processing Laboratory

I. Estimation of autocorrelation matrix

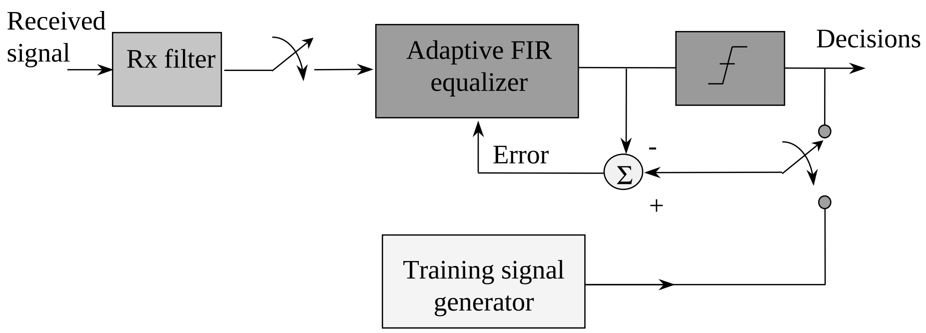
Estimation of autocorrelation matrix

- In the previous lecture, we derived the MSE gradient algorithm for adaptation of the FIR equalizer
- MSEG requires knowledge of channel in form of Rx signal autocorrelation matrix and crosscorrelation vector
- Separate estimation cumbersome and adds delay
- Goal of this lecture:

  Improve the MSEG algorithm to eliminate separate estimation of autocorrelation matrix and crosscorrelation vector

Estimation of autocorrelation matrix...

- System model:



Estimation of autocorrelation matrix...

- Consider estimation of autocorrelation matrix
- Diagonal term:

$$R_0 = \mathrm{E}\left[r^2(k)\right] \approx \frac{1}{L}\sum_{l=0}^{L-1} r^2(k-l) = \hat{R}_0$$

- Whole matrix:

$$\hat{\mathbf{R}} = \frac{1}{L}\sum_{l=0}^{L-1} \mathbf{r}(k-l)\mathbf{r}^{\mathrm{T}}(k-l)$$

- Crosscorrelation vector:

$$\hat{\mathbf{p}} = \frac{1}{L}\sum_{l=0}^{L-1} \mathbf{r}(k-l)a_{k-l}$$

Estimation of autocorrelation matrix...

Problems with averaging estimation:

- Choice of averaging window *L*
  - long averaging window adds delay and degrades tracking capability
  - short window gives noisy estimates
- Solution: *integrate* estimation and adaptation into one algorithm

II. Stochastic Gradient Algorithm

Stochastic Gradient Algorithm

- General MSE expression:

$$
\begin{aligned}
\mathrm{E}\left[e^2(k)\right] &= \mathrm{E}\left[\left(\mathbf{h}_\mathrm{R}^\mathrm{T}\mathbf{r}(k) - a_k\right)^2\right] \\
&= \mathrm{E}\left[a_k^2\right] - 2\mathbf{h}_\mathrm{R}^\mathrm{T}\mathrm{E}\left[\mathbf{r}(k)a_k\right] + \mathbf{h}_\mathrm{R}^\mathrm{T}\mathrm{E}\left[\mathbf{r}(k)\mathbf{r}^\mathrm{T}(k)\right]\mathbf{h}_\mathrm{R} \\
&= \mathrm{E}\left[a_k^2\right] - 2\mathbf{h}_\mathrm{R}^\mathrm{T}\mathbf{p} + \mathbf{h}_\mathrm{R}^\mathrm{T}\mathbf{R}\mathbf{h}_\mathrm{R}
\end{aligned}
$$

- Problem: evaluating the expectations
- Solution: forget the expectation and use *instantaneous* error instead

Stochastic Gradient Algorithm...

- Instantaneous squared error (ISE):

$$e^2(k) = \left(\mathbf{h}_R^T \mathbf{r}(k) - a_k\right)^2$$
$$= a_k^2 - 2\mathbf{h}_R^T \mathbf{r}(k) a_k + \mathbf{h}_R^T \mathbf{r}(k) \mathbf{r}^T(k) \mathbf{h}_R$$

- Gradient:

$$\nabla_{\mathbf{h}_R}\left[e^2(k)\right] = 2\mathbf{r}(k)\mathbf{r}^T(k)\mathbf{h}_R - 2\mathbf{r}(k)a_k$$

- Gradient of ISE can be viewed as *stochastic estimate* for the exact MSE gradient

Stochastic Gradient Algorithm...

- From ISE, the stochastic estimates for the autocorrelation matrix and crosscorrelation vector are:

$$\hat{\mathbf{R}} = \mathbf{r}(k)\mathbf{r}^T(k)$$
$$\hat{\mathbf{p}} = \mathbf{r}(k)a_k$$

- Simplified form for stochastic gradient:

$$\nabla_{\mathbf{h}_R}\left[e^2(k)\right] = -2\mathbf{r}(k)(a_k - \mathbf{r}^T(k)\mathbf{h}_R) = -2\mathbf{r}(k)e(k)$$

- Use gradient of ISE to construct a stochastic gradient (SG) algorithm

Stochastic Gradient Algorithm...

- Stochastic gradient (SG) algorithm:

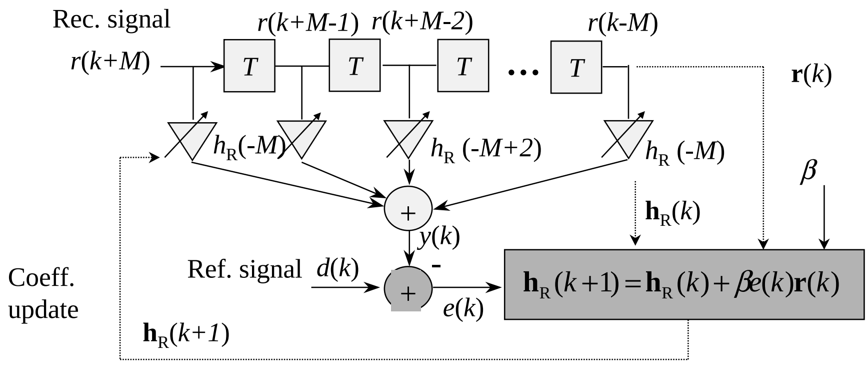
$$\mathbf{h}_R(k+1) = \mathbf{h}_R(k) - \frac{\beta}{2}\nabla_{\mathbf{h}_R} e^2(k)$$
$$= \mathbf{h}_R(k) + \beta e(k)\mathbf{r}(k)$$

- Also known as Least Mean Squares (LMS) algorithm
- Steps of filtering and adaptation:

$$y(k) = \mathbf{h}_R^T(k)\mathbf{r}(k)$$
$$e(k) = d(k) - y(k) = a_k - y(k)$$
$$\mathbf{h}_R(k+1) = \mathbf{h}_R(k) + \beta e(k)\mathbf{r}(k)$$

Stochastic Gradient Algorithm...

- Computations for SG algorithm:



Stochastic Gradient Algorithm...

Properties of SG algorithm:

- No separate estimation needed: all signals of coefficient update equation are known
- Continuous averaging of channel data integrated in the algorithm
- Iteration index and time index *k* are the same
- Noisy gradient estimate $\rightarrow$ convergence is *stochastic* (the result can sometimes get worse!)

III. Analysis of SG algorithm

Analysis of SG algorithm

- Let us study *coefficient error vector*:

$$\Delta \mathbf{h}_{\mathrm{R}}(k) = \mathbf{h}_{\mathrm{R}}(k) - \mathbf{h}_{\mathrm{R,\,opt}}$$

- SG update equation:

$$\mathbf{h}_{\mathrm{R}}(k+1) = \mathbf{h}_{\mathrm{R}}(k) + \beta e(k)\mathbf{r}(k)$$

- Elaborate *coefficient error update* equation as:

$$\Delta\mathbf{h}_{\mathrm{R}}(k+1) = \left(I - \beta\mathbf{r}(k)\mathbf{r}(k)^{\mathrm{T}}\right)\Delta\mathbf{h}_{\mathrm{R}}(k) + \beta e_{\mathrm{opt}}(k)\mathbf{r}(k)$$

where the error of the optimal solution is $e_{\mathrm{opt}}(k) = a_k - \mathbf{r}^{\mathrm{T}}(k)\mathbf{h}_{\mathrm{R,opt}}$

Analysis of SG algorithm...

- Define squared norm for coefficient error vector:

$$\left\|\Delta \mathbf{h}_{\mathrm{R}}(k)\right\|^2 = \Delta \mathbf{h}_{\mathrm{R}}^{\mathrm{T}}(k)\Delta \mathbf{h}_{\mathrm{R}}(k)$$

- When small, the solution is close to optimum!
- Relation between MSE and coefficient error:

$$\begin{aligned}\mathrm{E}\left[e^2(k)\right] &= E_{\mathrm{MIN}} + \left(\mathbf{h}_{\mathrm{R}} - \mathbf{h}_{\mathrm{R,opt}}\right)^{\mathrm{T}}\mathbf{R}\left(\mathbf{h}_{\mathrm{R}} - \mathbf{h}_{\mathrm{R,opt}}\right) \\ &= E_{\mathrm{MIN}} + \Delta\mathbf{h}_{\mathrm{R}}^{\mathrm{T}}\mathbf{R}\Delta\mathbf{h}_{\mathrm{R}}\end{aligned}$$

Analysis of SG algorithm...

- Take expectation w.r.t. the coefficient process:

$$\mathrm{E}\left[e^2(k)\right] = E_{\mathrm{MIN}} + \mathrm{E}\left[\Delta\mathbf{h}_{\mathrm{R}}^{\mathrm{T}}\mathbf{R}\Delta\mathbf{h}_{\mathrm{R}}\right]$$
$$= E_{\mathrm{MIN}} + E_{\mathrm{EX}}$$

- The latter term is called *Excess MSE*:
- Excess MSE tells how much error is caused by the adaptation process

Analysis of SG algorithm...

- In general, the Excess MSE is hard to analyze exactly
- Study white noise input: $\mathbf{R} = R_0\mathbf{I}, \quad R_0 = \mathrm{E}\left[r^2(k)\right]$
- Now the Excess MSE is

$$\mathrm{E}\left[e^2(k)\right] = E_{\mathrm{MIN}} + R_0\mathrm{E}\left[\left\|\Delta\mathbf{h}_{\mathrm{R}}\right\|^2\right]$$

- Following result can be obtained (see Lee-Messerschmitt, Appendix 11-A):

$$\mathrm{E}\left[\left\|\Delta\mathbf{h}_{\mathrm{R}}(k+1)\right\|^2\right] = \gamma\mathrm{E}\left[\left\|\Delta\mathbf{h}_{\mathrm{R}}(k)\right\|^2\right] + \beta^2 N R_0 E_{\mathrm{min}}$$

where $\gamma = 1 - 2\beta R_0 + \beta^2 N R_0^2$

Analysis of SG algorithm...

- Condition for convergence:

$$|\gamma| < 1$$

$$\Leftrightarrow \; -1 < 1 - 2\beta R_0 + \beta^2 N R_0^2 < 1$$

- This is equivalent to requiring

$$0 < \beta < \frac{2}{N R_0}$$

- Compare with MSEG criterion:

$$0 < \beta < \frac{2}{\lambda_{\mathrm{MAX}}} = \frac{2}{R_0}$$

Analysis of SG algorithm...

- The max step size of SG algorithm is proportional to $1/N$ (N = no. of filter coefficients)
- Extra tolerance is needed because of the stochastic nature of the algorithm

Summary

Today we discussed:

Adaptive equalizers 2

I. Estimation of autocorrelation matrix

II. Stochastic Gradient algorithm

III. Analysis of SG algorithm

Next week: Decision-feedback equalizers