Recap of the Development from Wiener Filters to Steepest Descent to LMS

Figure 5.1. Statistical Filtering Setup

Wiener Filter

Given **R**, **p** we design our filter to be optimal in the least-mean-square error sense, i.e. the Wiener filter

$$\begin{aligned}\mathbf{w} &= \mathbf{P}^{-1}\boldsymbol{\pi} \\ &= \boldsymbol{\omega}_o\end{aligned} .$$

This choice of **w** guarantees (by virtue of the quadratic error surface) that our MSE is minimized

$$\begin{aligned}J &= E[\varepsilon(\nu)\varepsilon^*(\nu)] \\ &= \vartheta_{\mu\iota\nu} \\ &= \sigma_\delta^2 - \boldsymbol{\pi}^H \mathbf{P}^{-1}\boldsymbol{\pi}\end{aligned} .$$

Figure. Sample quadratic error surface, MSE

Steepest Descent

Given **R**, **p** we may *search* for the optimal filter (Wiener filter) with the following update

$$\mathbf{w}(n+1) = \boldsymbol{\omega}(\nu) + \mu[\boldsymbol{\pi} - \mathbf{P}\boldsymbol{\omega}(\nu)]$$

with

$$0 < \mu < \frac{2}{\lambda_{\mu\alpha\xi}} .$$

Using this update with the proper choice of μ, we guarantee that

$$\lim_{n\to\infty} \vartheta(\nu) = \vartheta_{\mu\iota\nu} .$$

We note that $J(n)$ consists of a sum of dying exponentials (modes or transients). For fixed μ, the fastest decaying mode corresponds to the λ_{max} of **R** whereas the slowest decaying mode corresponds to the λ_{min} . For fixed μ the convergence rate depends on χ. For fixed χ the larger μ the faster the we converge.

Figure. Sample quadratic error surface with trajectory, MSE

Least-Mean-Square Algorithm

If we are not given **R**, **p** we can adjust our filter $\hat{\mathbf{w}}(n)$ so as to *attempt* to minimize *J*(*n*). In this case we rely on instantaneous estimates of **R** and **p** to arrive at the update equation

$$\hat{\mathbf{w}}(n+1) = \hat{\boldsymbol{\omega}}(\nu) + \mu\boldsymbol{\upsilon}(\nu)\varepsilon^{*}(\nu).$$

1) If we choose

$$0 < \mu < \frac{2}{\lambda_{\mu\alpha\xi}}$$

we have convergence of the mean (weak)

$$\lim_{n\to\infty} E[\hat{\boldsymbol{\omega}}(\nu)] = \boldsymbol{\omega}_{o}.$$

2) If we choose

$$0 < \mu < \frac{2}{3\tau\rho(\mathbf{P})}$$

we have convergence in the mean square (strong). Mathematically

$$J(n) = \vartheta_{\mu\iota\nu} + \vartheta_{\varepsilon\xi}(\nu)$$

and

$$\lim_{n\to\infty} \vartheta_{\varepsilon\xi}(\nu) = \chi o\nu\sigma\tau\alpha\nu$$

Again, we note that $J_{ex}(n)$ and hence *J*(*n*) consists of a sum of dying exponentials (modes or transients). For fixed μ the convergence rate depends on χ.

For fixed χ the larger μ the faster the we converge in the sense that $\lim_{n\to\infty} \vartheta_{\varepsilon\xi}(\nu) = \chi o\nu\sigma\tau\alpha\nu$. However, larger μ implies larger $\lim_{n\to\infty} \vartheta_{\varepsilon\xi}(\nu)$. These are competing requirements since we wish to operate as close to J_{min} as possible but get to optimal operating conditions as quick as possible.

Figure. Sample quadratic error surface with trajectory, MSE

LMS Working Rules

In practice we'll select

$$0 < \mu < \frac{2}{3\tau\rho(\mathbf{P})}$$

so that

$$M = \frac{\vartheta_{\varepsilon\xi}(\infty)}{\vartheta_{\mu\iota\nu}} \approx 10\%.$$

When we compare other adjustment algorithms we must keep $J_{ex}(n)$ in mind. Remember as suggested many years ago in the *IEEE Trans. ASSP*, "...comparison of convergence rates is meaningless without specifying the level of misadjustment!"

As mentioned earlier LMS is the benchmark for which all other adaptive algorithms are judged against. For the remainder of the class (adaptive filtering portion) we will examine algorithms that attempt to converge faster (for fixed misadjustment) or are more robust.

LMS Experiment (from Diniz--similar examples in Haykin)

Consider the following channel equalization application

Figure.

where we know (but LMS doesn't) that

$$H_2(z) = \frac{\zeta^{-1}}{1+0.8\zeta^{-1}} .$$

Also assume an adaptive filter of length 2 and an additive Gaussian noise in the channel with $\sigma_{v_2}^2 = 1$.

We assume input to the channel can be modeled as an ARMA process (driven by Gaussian white noise with $\sigma_{v_1}^2 = 1$)

Figure. Input signal model.

where

$$H_1(z) = \frac{\zeta^{-1}}{1-0.5\zeta^{-1}} .$$

For our analysis we can show

$$\mathbf{R} = \begin{bmatrix} 1.6873 & -0.7937 \\ -0.7937 & 1.6873 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} 0.9524 \\ 0.4762 \end{bmatrix},$$

and

$$\mathbf{w}_o = \begin{bmatrix} 0.8953 \\ 0.7034 \end{bmatrix}.$$

Choose $\mu = 1 / (40\text{tr}(\mathbf{R})) = 0.0074$.

Figures. Trajectory and MSE for single run, MSE for ensemble-averaged (25 runs).