Institut EURECOM 11/06/2007
SSP Statistical Signal Processing

TD2: Optimal and Adaptive Filtering, Equalization

1 Wiener Filtering

Problem 1. A Wiener filtering problem

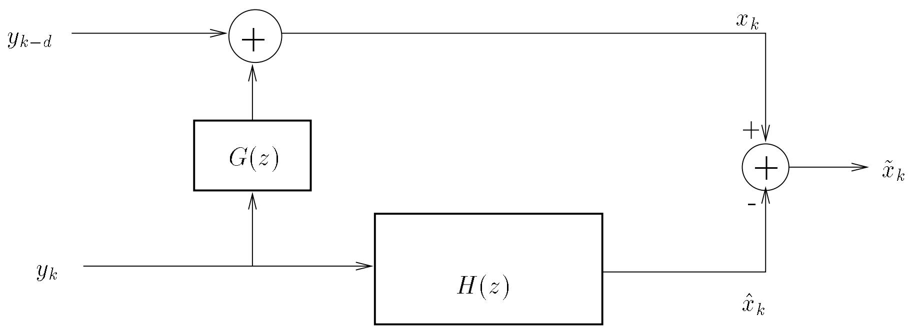


Figure 1: Wiener filtering problem.

In Fig. 1, a Wiener filtering problem is sketched. We can measure a signal y_k but we are interested in a related signal x_k that is indicated in the figure. $G(z)$ is the transfer function of some linear time-invariant filter and d is some delay. What is the Wiener filter $H(z)$ (in terms of the quantities indicated in the figure) for optimally estimating x_k from y_k and what is the associated MMSE?

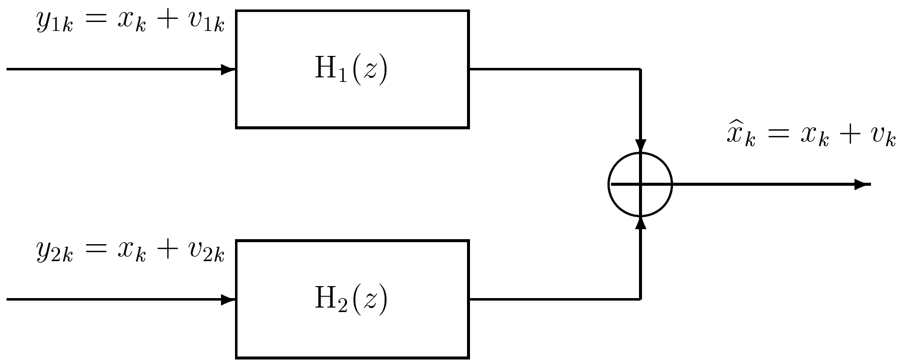
Problem 2. Constrained and Unconstrained Two-Channel Wiener Filtering

Consider the problem of combining two independent noisy measurements of the same signal as depicted in the figure below.

(a) Often a spectral description of x_k is not available, or it may not be properly modeled as a random process. Furthermore, one may require that the signal be filtered without distortion. We consider x_k to be deterministic in this case, and constrain the two filters to satisfy

$$\mathrm{H}_1(z) + \mathrm{H}_2(z) = 1, \quad \forall z \quad , \tag{1}$$

so that x_k passes undistortedly (that is, $\widehat{x}_k = x_k +$ noise). Due to the presence of more than one filter, we can take here a point of view that is intermediate between the classical filtering point of view of (no) distortion and the statistical MSE point of view. We can



constrain the filters in order to have no signal distortion. But due to the presence of more than one filter, this no distortion requirement does not fix all the degrees of freedom. We can use the remaining degrees of freedom to minimize the MSE.
Show how to choose $\mathrm{H}_1(z)$ so as to minimize the MSE, the variance of the output noise $v_k = -\widetilde{x}_k$. Calculate the associated MMSE.

(b) Consider now on the contrary that x_k is a random process and suppose that we have its spectral description $S_{xx}(f)$ available after all. We can now formulate a Wiener filtering problem for the generalized case where we have a vector process available, namely in this case

$$y_k = \begin{bmatrix} y_{1k} \\ y_{2k} \end{bmatrix} . \tag{2}$$

Using appropriately a vector valued transfer function

$$\mathrm{H}(z) = [\mathrm{H}_1(z) \;\; \mathrm{H}_2(z)] \, , \tag{3}$$

we can form a linear estimator

$$\widehat{x}_k = \mathrm{H}(q)\, y_k = \mathrm{H}_1(q)\, y_{1k} + \mathrm{H}_2(q)\, y_{2k} \, . \tag{4}$$

Using the short-cut frequency domain derivation for the optimal Wiener filter, in which at every frequency f the filter $H(f)$ is just a vector of (complex) coefficients to form the LMMSE estimate of the scalar $X(f)$ given the vector $Y(f)$, we find straightforwardly the following generalization

$$H(f) = S_{xy}(f)\, S_{yy}^{-1}(f) = R_{X(f)Y(f)}\, R_{Y(f)Y(f)}^{-1} \tag{5}$$

where in this example

$$S_{xy}(f) = [S_{xy_1}(f) \;\; S_{xy_2}(f)] \;\; , \quad S_{yy}(f) = \begin{bmatrix} S_{y_1y_1}(f) & S_{y_1y_2}(f) \\ S_{y_2y_1}(f) & S_{y_2y_2}(f) \end{bmatrix} . \tag{6}$$

Show that for the signal in noise problem at hand (with n_{1k} and n_{2k} being uncorrelated), we get

$$S_{xy}(f) = S_{xx}(f)\,[1 \;\; 1] \;\; , \quad S_{yy}(f) = S_{xx}(f) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} S_{n_1n_1}(f) & 0 \\ 0 & S_{n_2n_2}(f) \end{bmatrix} . \tag{7}$$

Find the Wiener filter $H(f)$ and compare to the solution of (a). Find the associated MMSE, the variance of $\widetilde{x}_k = x_k - \widehat{x}_k$ when the Wiener filter is used for $\widehat{x}_k$, and compare to the solution of (a).

2 Equalization and Wiener Filtering

Problem 3. Equalization of a First-Order FIR Channel

Consider the following simple channel $\mathrm{C}(z) = 1 - az^{-1}$. Compute the linear equalizer transfer function $\mathrm{H}(z)$ and the associated MSE for the ZF, MMSE and UMMSE criteria. Comment especially on what happens when $|a| \to 1$.

Problem 4. Wiener Filtering and Zero-Forcing Linear Equalization of a Second-Order FIR Channel

(a) Consider the signal in noise case ($y_k = x_k + v_k$). Show that when the noise v_k is white with variance σ_v^2, the MMSE of the Wiener filter turns out to be

$$\mathrm{MMSE} \ = E\, \tilde{x}_k^2 = \sigma_v^2\, h_0$$

where h_k is the impulse response of the non-causal Wiener filter.

(b) Consider the following discrete-time channel

$$\mathrm{C}(z) = 1 - \frac{5}{2} z^{-1} + z^{-2}\ .$$

Compute the impulse response h_k of the (non-causal) zero-forcing linear equalizer.

(c) For the same channel, in which we assume the variance of the additive white noise to be σ_v^2, compute the MSE of the zero-forcing (ZF) linear equalizer (LE).

(d) Compute the SNR for the ZF-LE, and the Matched Filter Bound (MFB) and compare.

3 Steepest-Descent and Adaptive Filtering Algorithms

Problem 5. Steepest-descent algorithm

Consider the steepest-descent algorithm for iterating towards the Wiener FIR filter for the case of two filter coefficients and the input having the covariance sequence $r_{yy}(k) = \rho^{|k|}$. Determine both the maximum stepsize μ for convergence and the stepsize that gives the fastest convergence as a function of ρ. With this stepsize, what is the slowest mode ?

Problem 6. An application of the LMS algorithm

Assume we are applying the LMS algorithm to a problem in which the input signal is of the form $y_k = a + w_k$ where $a \sim \mathcal{N}(0, \sigma_a^2)$ is a random dc component, the $w_k \sim \mathcal{N}(0, \sigma_w^2)$ are white noise samples and a is independent of the w_k.

(a) Find the mean E y_k.

(b) Find the covariance matrix R_{YY}.

(c) Find the eigenvalues of R_{YY} (order them so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$) and associated eigenvectors $V_1, \ldots, V_N$.

(d) Consider the analysis obtained by using the averaging theorem. When running the LMS algorithm in this particular case, what are the bounds on the stepsize μ so that convergence occurs (in other words, $? < \mu < ?$).

(e) What is the value for μ for fastest convergence?

Problem 7. Stochastic Newton Algorithm

Consider the filter adaptation approach in which the filter coefficients H_k are obtained by minimizing the criterion

$$\xi_k^{SN}(H) \;=\; \left(x_k - Y_k^T H\right)^2 + \left(H - H_{k-1}\right)^T A_k \left(H - H_{k-1}\right) \tag{8}$$

where $A_k = A_k^T > 0$ is a symmetric positive definite matrix.

(a) Find $H_k = \arg\min_H \xi_k^{SN}(H)$ and express it in terms of H_{k-1}.

(b) Show how the LMS algorithm and the RLS algorithm fall out as special cases of the approach in (8). Specifically, show for which value of A_k the update of H_k you found in (a) coincides with

(i) the LMS algorithm,

(ii) the RLS algorithm.

(c) Take $A_k \;=\; \left(\frac{1}{\mu} - Y_k^T R_{YY}^{-1} Y_k\right) R_{YY}$ in (8).

(i) Show that

$$\frac{1}{1 + Y_k^T A_k^{-1} Y_k} A_k^{-1} \;=\; \mu\, R_{YY}^{-1}\,. \tag{9}$$

(ii) Use this result in the update equation for H_k you found in (a). Investigate the convergence dynamics (different modes) of $E\,H_k$, using the independence assumption.

(iii) What are the conditions on μ for convergence of $E\,H_k$?

(iv) Assuming convergence, what is the steady-state value $E\,H_\infty$?

(v) Which value of μ leads to the fastest convergence of $E\,H_k$?

(vi) With this optimal μ, the convergence of $E\,H_k$ occurs in how many updates?

Problem 8. Sign LMS Algorithm

Apply the derivation technique for the LMS algorithm to the following criterion: $\min E\left|x_k - H^T Y_k\right|$. Comment on how this leads to an adaptive algorithm that can be considered as a computational simplification of the LMS algorithm.

Knowing that the MSE criterion is well adjusted to the case where the measurement noise $\widetilde{x}_k$ is Gaussian, to what kind of distribution of $\widetilde{x}_k$ would the criterion considered here be adapted?

Problem 9. Leaky LMS Algorithm

Apply the derivation technique of the LMS algorithm to the following criterion:

$$(x_k - H^T Y_k)^2 + \rho \, \|H\|^2 \ .$$

In the usual system identification set-up with a fixed optimal filter H^o, does the mean of the resulting parameter estimation error $\widetilde{H}_k$ converge to zero?
Explain why this algorithm may have beneficial properties when the input covariance matrix R_{YY} is singular.

Problem 10. Exponentially Weighted RLS Algorithm

Derive the exponentially weighted Recursive Least-Squares algorithm, which minimizes the following criterion recursively:

$$\xi_k \left(H\right) = \sum_{i=1}^{k} \lambda^{k-i} \, \left(x_i - H^T Y_i\right)^2 + \lambda^k \, \left(H - H_0\right)^T R_0 \left(H - H_0\right) \ .$$

Problem 11. LMS as an Optimization Problem

Show that the LMS algorithm can be obtained exactly from the following optimization problem

$$H_k = \arg\min_H \left\{ \left(x_k - Y_k^T H\right)^2 + \left(\frac{1}{\mu} - Y_k^T Y_k\right) \left(H - H_{k-1}\right)^T \left(H - H_{k-1}\right) \right\} \ . \tag{10}$$

This shows the LMS update as the result of a compromise between taking into account the new data at time k and adhering to the previous estimate H_{k-1}. However, the weighting factor of the second term is not guaranteed to be positive and if it is not, the Hessian of the optimization problem in (10) will not be positive definite, meaning that the extremum at H_k will not be a minimum. This illustrates again the tricky stability issue of the LMS algorithm (unless μ is very small). Note that $\mu(\frac{1}{\mu} - Y_k^T Y_k) = 1 - \mu Y_k^T Y_k$ is the one eigenvalue different from one of the transition matrix $I - \mu Y_k Y_k^T$.

Problem 12. Approximate Analysis of LMS for Large Stepsize

Consider the LMS algorithm in the system identification set-up with fixed optimal parameters. Using the independence assumption, but not the averaging theorem, the correlation matrix of the parameter estimation error vector $\widetilde{H}_k$ is found to satisfy the recursion

$$\begin{aligned} C_k &= E\left(\left[I - \mu Y_k Y_k^T\right] \, \widetilde{H}_{k-1} \widetilde{H}_{k-1}^T \, \left[I - \mu Y_k Y_k^T\right]\right) + \mu^2 \xi^o R_{YY} \\ &= E_{Y_k}\left(\left[I - \mu Y_k Y_k^T\right] C_{k-1} \left[I - \mu Y_k Y_k^T\right]\right) + \mu^2 \xi^o R_{YY} \end{aligned} \ . \tag{11}$$

The remaining expectation operator E_{Y_k} is over the elements of the input vector $Y_k = [y_k \; y_{k-1} \cdots y_{k-N+1}]^T$.
Working out the different terms, we get

$$C_k \ = \ C_{k-1} - \mu R_{YY} C_{k-1} - \mu C_{k-1} R_{YY} + \mu^2 \, E_{Y_k} \left\{ Y_k Y_k^T \, Y_k^T C_{k-1} Y_k \right\} + \mu^2 \xi^o R_{YY} \ . \tag{12}$$

Using a Law of Large Numbers, we have the convergence

$$\frac{1}{N} \, Y_k^T C_{k-1} Y_k \ \overset{N\to\infty}{\longrightarrow} \ E \, \frac{1}{N} \, Y_k^T C_{k-1} Y_k \ = \ \frac{1}{N} \, \mathrm{tr}\left\{C_{k-1} R_{YY}\right\} \tag{13}$$

where the convergence is according to one of the types of convergence of sequences of random variables. Using this result, we can approximate

$$Y_k^T C_{k-1} Y_k \;\approx\; \mathrm{tr}\,\{C_{k-1} R_{YY}\} \tag{14}$$

and hence (12) becomes approximately

$$C_k \;=\; C_{k-1} - \mu R_{YY} C_{k-1} - \mu C_{k-1} R_{YY} + \mu^2\, R_{YY}\,\mathrm{tr}\,\{C_{k-1} R_{YY}\} + \mu^2 \xi^o R_{YY}\;. \tag{15}$$

(a) Using the eigen decomposition $R_{YY} = V \Lambda V^T$, show that by multiplying (15) to the left with V^T and to the right with V we get

$$V^T C_k V \;=\; V^T C_{k-1} V - \mu \Lambda\, V^T C_{k-1} V - \mu V^T C_{k-1} V\,\Lambda + \mu^2\,\Lambda\,\mathrm{tr}\left\{V^T C_{k-1} V\,\Lambda\right\} + \mu^2 \xi^o \Lambda\;. \tag{16}$$

(b) Now introduce the notation

$$D_k = \mathrm{diag}\left(V^T C_k V\right) = \begin{bmatrix} E\,v_1^2(k) \\ \vdots \\ E\,v_N^2(k) \end{bmatrix} \;,\quad \lambda = \mathrm{diag}\,(\Lambda) = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} \tag{17}$$

where we have used the Matlab operator diag(.) which, when operating on a square matrix, selects its diagonal elements and puts them in a column vector. Show that from (16), we obtain

$$D_k \;=\; \left[I - 2\mu\Lambda + \mu^2 \lambda\lambda^T\right] D_{k-1} + \mu^2 \xi^o \lambda\;. \tag{18}$$

Show also that the expression for the MSE (learning curve) becomes

$$\xi_k \;=\; \xi^o + \mathrm{tr}(\Lambda\, V^T C_k V) \;=\; \xi^o + \lambda^T D_k\;. \tag{19}$$

(c) Assuming convergence, show that we get for the steady-state values

$$D_\infty \;=\; \frac{\mu}{2 - \mu\lambda^T \mathbf{1}}\,\xi^o \mathbf{1}\;,\;\; \xi_\infty \;=\; \xi^o + \lambda^T D_\infty \;=\; \xi^o[1 + \frac{\mu\,\lambda^T \mathbf{1}}{2 - \mu\,\lambda^T \mathbf{1}}]\;. \tag{20}$$

(d) The dynamics of the convergence are governed by the eigenvalues z_i of the system matrix $\Phi \;=\; I - 2\mu\Lambda + \mu^2 \lambda\lambda^T$. We have $z_i \equiv 1$ (stability boundary) for $\mu = 0$ and $z_i \lessapprox 1$ (exponentially stable) for $\mu \gtrapprox 0$. The maximal value μ_{max} of μ for which we reach the boundary of stability again can be found as the smallest positive value of μ for which D_∞ diverges. So what is μ_{max}? Note that in particular for a white input ($\lambda_N = \lambda_1$) this value for μ_{max} is N times smaller than the value obtained using the averaging theory (the μ_{max} of steepest-descent). For a white input, the value for μ_{max} obtained here is quite close to the true value. As the input signal gets more and more colored (λ_1/λ_N increases), the ratio of μ_{max} obtained for steepest-descent and the μ_{max} obtained here decreases. However, as the input gets more colored, the true μ_{max} gets smaller than the value obtained here (and is not known in general).

(e) Using a perturbation theorem due to Wilkinson for the eigenvalues of a matrix perturbed by a rank one matrix, one can state

$$z_i \;=\; 1 - 2\mu\lambda_i + m_i\,\mu^2 \lambda^T \lambda \;\;,\quad m_i(\mu) \geq 0\;,\; \sum_{i=1}^{N} m_i = 1\;. \tag{21}$$

When for the maximal stepsize we reach the stability border, do we have $z_1(\mu_{max}) = -1$ or $z_N(\mu_{max}) = 1$? Analyze in particular the white input case ($\lambda_N = \lambda_1$).