2.6 The optimum filtering solution is defined by the Wiener-Hopf equation

$$\mathbf{R}\mathbf{w}_o = \mathbf{p} \qquad (1)$$

for which the minimum mean-square error equals

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H \mathbf{w}_o \qquad (2)$$

Combine Eqs. (1) and (2) into a single relation:

$$\begin{bmatrix} \sigma_d^2 & \mathbf{p}^H \\ \mathbf{p} & \mathbf{R} \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{w}_o \end{bmatrix} = \begin{bmatrix} J_{\min} \\ \mathbf{0} \end{bmatrix}$$

Define

$$\mathbf{A} = \begin{bmatrix} \sigma_d^2 & \mathbf{p}^H \\ \mathbf{p} & \mathbf{R} \end{bmatrix} \qquad (3)$$

Since

$$\sigma_d^2 = E[d(n)d^*(n)],$$

$$\mathbf{p} = E[\mathbf{u}(n)d^*(n)] \text{ , and}$$

$$\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^*(n)],$$

we may rewrite Eq. (3) as

$$\mathbf{A} = \begin{bmatrix} E[d(n)d^*(n)] & E[d(n)]\mathbf{u}^H(n) \\ E[\mathbf{u}(n)d^*(n)] & E[\mathbf{u}(n)\mathbf{u}^H(n)] \end{bmatrix}$$

$$= E\left\{\begin{bmatrix} d(n) \\ \mathbf{u}(n) \end{bmatrix}\begin{bmatrix} d^*(n), \ \mathbf{u}^H(n) \end{bmatrix}\right\}$$

The minimum mean-square error equals

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H\mathbf{w}_o \qquad (4)$$

Eliminate σ_d^2 between Eqs. (1) and (4):

$$J(\mathbf{w}) = J_{\min} + \mathbf{p}^H\mathbf{w}_o - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w} \qquad (5)$$

Eliminate $\mathbf{p}$ between (2) and (5):

$$J(\mathbf{w}) = J_{\min} + \mathbf{w}_o^H\mathbf{R}\mathbf{w}_o - \mathbf{w}_o^H\mathbf{R}\mathbf{w} - \mathbf{w}^H\mathbf{R}\mathbf{w}_o + \mathbf{w}^H\mathbf{R}\mathbf{w} \qquad (6)$$

where we have used the property $\mathbf{R}^H = \mathbf{R}$. We may rewrite Eq. (6) simply as

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H\mathbf{R}(\mathbf{w} - \mathbf{w}_o)$$

which clearly show that $J(\mathbf{w}_o) = J_{\min}$.

2.7 The minimum mean-square error equals

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H\mathbf{R}^{-1}\mathbf{p} \qquad (1)$$

Using the spectral theorem, we may express the correlation matrix $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{Q}\Lambda\mathbf{Q}^H$$

$$= \sum_{k=1}^{M} \lambda_k \mathbf{q}_k \mathbf{q}_k^H$$

Hence, the inverse of $\mathbf{R}$ equals

$$\mathbf{R}^{-1} = \sum_{k=1}^{M} \frac{1}{\lambda_k}\mathbf{q}_k\mathbf{q}_k^H \qquad (2)$$

$$\frac{\partial J}{\partial w} = 2r(0)(w - w_o)$$

4.3 (a) There is a single mode with eigenvalue $\lambda_1 = r(0)$, and $q_1 = 1$, Hence,

$$J(n) = J_{\min} + \lambda_1 |v_1(n)|^2$$

where $v_1(n) = q_1(w_o - w(n)) = (w_o - w(n))$

(b) $$\frac{\partial J(n)}{\partial \lambda_1} = |v_1(n)|^2 = (w_o - w(n))^2$$

4.4 The estimation error $e(n)$ equals

$$e(n) = d(n) - \mathbf{w}^H(n)\mathbf{u}(n)$$

where $d(n)$ is the desired response, $\mathbf{w}(n)$ is the tap-weight vector, and $\mathbf{u}(n)$ is the tap-input vector. Hence, the gradient of the instantaneous squared error equals

$$\hat{\nabla} J(n) = \frac{\partial}{\partial \mathbf{w}}[|e(n)|^2] = \frac{\partial}{\partial \mathbf{w}}[e(n)e^*(n)]$$

$$= e(n)\frac{\partial e^*(n)}{\partial \mathbf{w}} + e^*(n)\frac{\partial e(n)}{\partial \mathbf{w}}$$

$$= -2e^*(n)\mathbf{u}(n) = 2\mathbf{u}(n)d^*(n) + 2\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)$$

4.5 Consider the approximation to the inverse of the correlation matrix:

$$\mathbf{R}^{-1}(n+1) \approx \mu \sum_{k=0}^{n} (\mathbf{I} - \mu\mathbf{R})^k$$

where μ is a positive constant bounded in value as

$$0 < \mu < \frac{2}{\lambda_{\max}}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$. Note that according to this approximation, we have $\mathbf{R}^{-1}(1) = \mu\mathbf{I}$. Correspondingly, we may approximate the optimum Wiener solution as

$$\mathbf{w}(n+1) = \mathbf{R}^{-1}(n+1)\mathbf{p}$$

$$\approx \mu \sum_{k=0}^{n} (\mathbf{I} - \mu\mathbf{R})^k \mathbf{p}$$

$$= \mu\mathbf{p} + \mu \sum_{k=1}^{n} (\mathbf{I} - \mu\mathbf{R})^k \mathbf{p}$$

In the second term, put $k = i+1$ or $i = k-1$:

$$\mathbf{w}(n+1) = \mu\mathbf{p} + \mu(\mathbf{I} - \mu\mathbf{R}) \sum_{k=1}^{n} (\mathbf{I} - \mu\mathbf{R})^i \mathbf{p}$$

$$= \mu\mathbf{p} + (\mathbf{I} - \mu\mathbf{R})\mathbf{w}(n) \qquad (1)$$

where, in the second line, we have used the fact that

$$\mu \sum_{i=0}^{n} (\mathbf{I} - \mu\mathbf{R})^i \mathbf{p} = \mathbf{w}(n)$$

Hence, rearranging Eq. (1):

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu[\mathbf{p} - \mathbf{R}\mathbf{w}(n)]$$

which is the standard formula for the steepest descent algorithm.

4.6 $$J(\mathbf{w}(n+1)) = J(\mathbf{w}(n)) - \frac{1}{2}\mu\|\mathbf{g}(n)\|^2$$

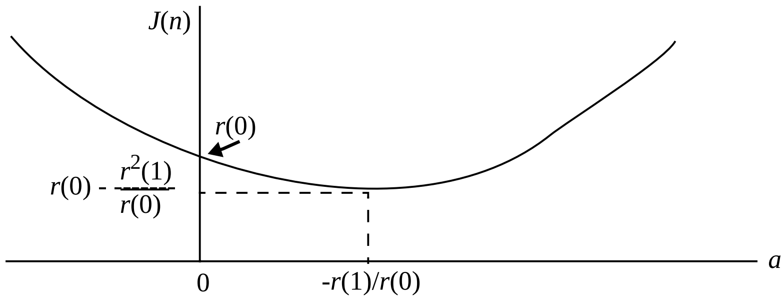
For stability of the steepest-descent algorithm, we therefore require

$$J(\mathbf{w}(n+1)) < J(\mathbf{w}(n))$$

To satisfy this requirement, the step-size parameter μ should be positive, since $\mu\|\mathbf{g}(n)\|^2 > 0$.

Hence, the steepest-descent algorithm becomes unstable when the step-size parameter is negative.

The corresponding plot of the error performance surface is therefore



(c) The condition on the step-size parameter is

$$0 < \mu < \frac{2}{r(0)}$$

4.10 The second-order AR process $u(n)$ is described by the difference equation

$$u(n) = -0.5u(n-1) + u(n-2) + v(n) \tag{1}$$

Hence,

$$w_1 = -0.5, \quad w_2 = 1$$

and the AR parameters equal

$$a_1 = 0.5, \quad a_2 = -1$$

Accordingly, we write the Yule-Walker equations as

$$\begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \end{bmatrix} \tag{2}$$

$$\sigma_v^2 = \sum_{k=0}^{2} a_k r(k)$$

$$1 = a_0 r(0) + a_1 r(1) + a_2 r(2)$$

$$= r(0) + 0.5r(1) - r(2) \tag{3}$$

Eqs. (1), (2) and (3) yield

$r(0) = 0$
$r(1) = 1$
$r(2) = -1/2$

Hence,

$$\mathbf{R} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The eigenvalues of **R** are -1, +1. For convergence of the steepest descent algorithm:

$$0 < \mu < \frac{2}{\lambda_{max}}$$

where λ_{max} is the largest eigenvalue of the correlation matrix **R**. Hence, with $\lambda_{max} = 1$.

$0 < \mu < 2$

4.11 $$u(n\text{-}2) = 0.5u(n\text{-}1) + u(n) - v(n) \qquad (1)$$

Hence,

$$w_1 = 1 \quad w_2 = 0.5$$

Accordingly, we may write $\mathbf{R}\mathbf{w}_b = \mathbf{r}^{B*}$ as

$$\begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} r(2) \\ r(1) \end{bmatrix}$$

$$\sigma_v^2 = \sum_{k=0}^{2} a_k r(k)$$

$$1 = r(0) - r(1) - 0.5r(2) \qquad (2)$$

$r(0) = 0$
$r(1) = -2/3$

Therefore, $\mathbf{R} = \begin{bmatrix} 0 & -2/3 \\ -2/3 & 0 \end{bmatrix}$

components subtract coherently, thereby yielding the average power $(A^2/2)(1-a)^2$. Hence, we may express the cost function J as

$$J \approx \sigma_\nu^2 + \left(\frac{2a^2}{M}\right)\sigma_\nu^2 + \frac{A^2}{2}(1-a)^2$$

Differentiating J with respect to a and setting the result equal to zero yields the optimum scale factor

$$a_{\text{opt}} = \frac{A^2}{A^2 + 4(\sigma_\nu^2/M)}$$

$$= \frac{(A^2/2\sigma_\nu^2)(M/2)}{1 + (A^2/2\sigma_\nu^2)(M/2)}$$

$$= \frac{(M/2)\text{SNR}}{1 + (M/2)\text{SNR}}$$

where $\text{SNR} = A^2/2\sigma_\nu^2$

5.5 The index of performance equals

$$J(\mathbf{w}, K) = E[e^{2K}(n)], \qquad K = 1, 2, 3, \cdots$$

The estimation error $e(n)$ equals

$$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{u}(n) \tag{1}$$

where $d(n)$ is the desired response, $\mathbf{w}(n)$ is the tap-weight vector of the transversal filter, and $\mathbf{u}(n)$ is the tap-input vector. In accordance with the multiple linear regression model for $d(n)$, we have

$$d(n) = \mathbf{w}_o^T(n)\mathbf{u}(n) \tag{2}$$

where $\mathbf{w}_o$ is the parameter vector, and $v(n)$ is a white-noise process of zero mean and variance σ_v^2.

(a) The instantaneous gradient vector equals

$$\hat{\nabla}(n, K) = \frac{\partial}{\partial \mathbf{w}} J(\mathbf{w}, K)$$

$$= \frac{\partial}{\partial \mathbf{w}}[e^{2K}(n)]$$

$$= 2Ke^{2K\text{-}1}(n)\frac{\partial e(n)}{\partial \mathbf{w}}$$

$$= -2Ke^{2K\text{-}1}(n)\mathbf{u}(n)$$

Hence, we may express the new adaptation rule for the estimate of the tap-weight vector as

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) - \frac{1}{2}\mu(-\hat{\nabla}(n, K))$$

$$= \hat{\mathbf{w}}(n) + \mu K\mathbf{u}(n)e^{2K\text{-}1}(n) \tag{3}$$

(b) Eliminate $d(n)$ between Eqs. (1) and (2), with the estimate $\hat{\mathbf{w}}(n)$ used in place of $\mathbf{w}$(n):

$$e(n) = (\mathbf{w}_o - \hat{\mathbf{w}}(n))^T\mathbf{u}(n) + v(n)$$

$$= \in^T(n)\mathbf{u}(n) + v(n)$$

$$= \mathbf{u}^T(n) \in (n) + v(n) \tag{4}$$

Subtract $\mathbf{w}_o$ from both sides of Eq. (3):

$$\in (n+1) = \in (n)\text{-}\mu K\mathbf{u}(n)e^{2K\text{-}1}(n) \tag{5}$$

For the case when $\in(n)$ is close to zero (i.e., $\hat{\mathbf{w}}(n)$ is close to $\mathbf{w}_o$), we may use Eq. (4) to write

$$e^{2K\text{-}1}(n) = [\mathbf{u}^T(n) \in (n) + v(n)]^{2K\text{-}1}$$

$$= v^{2K\text{-}1}(n)\left[1 + \frac{\mathbf{u}^T(n) \in (n)}{v(n)}\right]^{2K\text{-}1}$$

$$\approx v^{2K\text{-}1}(n)\left[1 + (2K\text{-}1)\frac{\mathbf{u}^T(n)\in(n)}{v(n)}\right]$$

$$= v^{2K\text{-}1}(n) + (2K\text{-}1)\mathbf{u}^T(n)\in(n)v^{2(K\text{-}1)}(n) \qquad (6)$$

Substitute Eq. (6) into (5):

$$\in(n+1) \approx [\mathbf{I} - \mu K(2K\text{-}1)v^{2(K\text{-}1)}(n)\mathbf{u}(n)\mathbf{u}^T(n)]\in(n)\text{-}\mu K v^{2K\text{-}1}\mathbf{u}(n)$$

Taking the expectation of both sides of this relation and recognizing that (1) $\in(n)$ is independent of $\mathbf{u}(n)$ by low-pass filtering action of the filter, (2) $\mathbf{u}(n)$ is independent of $v(n)$ by assumption, and (3) $\mathbf{u}(n)$ has zero mean, we get

$$E[\in(n+1)] = \{\mathbf{I} - \mu K(2K\text{-}1)E[v^{2(K\text{-}1)}(n)]\mathbf{R}\}E[\in(n)] \qquad (7)$$

where

$$\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^T(n)]$$

(c) Let

$$\mathbf{R} = \mathbf{Q}\Lambda\mathbf{Q}^T \qquad (8)$$

where Λ is the diagonal matrix of eigenvalues of $\mathbf{R}$, and $\mathbf{Q}$ is a matrix whose column vectors equal the associated eigenvectors. Hence, substituting Eq. (8) in (7) and using

$$\upsilon(n) = \mathbf{Q}^T E[\in(n)]$$

we get

$$\upsilon(n+1) = \{\mathbf{I} - \mu K(2K\text{-}1)E[v^{2(K\text{-}1)}(n)]\Lambda\}\upsilon(n)$$

That is, the *i*th element of this equation is

$$\upsilon_i(n+1) = \left\{1 - \mu K(2K\text{-}1)E[v^{2(K\text{-}1)}(n)]\lambda_i\right\}\upsilon_i(n) \qquad (9)$$

where $\upsilon_i(n)$ is the *i*th element of $\upsilon(n)$, and $i = 1, 2, \cdots, M$. Solving the first-order difference equation (9):

$$\upsilon_i(n) = \left\{1 - \mu K(2K\text{-}1)E[v^{2(K\text{-}1)}(n)]\lambda_i\right\}^{n\text{-}1} \upsilon_i(0)$$

where $\upsilon_i(0)$ is the initial value of $\upsilon_i(n)$. Hence, for $\upsilon_i(n)$ to converge, we require that

$$\left|1 - \mu K(2K\text{-}1)E[v^{2(K\text{-}1)}(n)]\right|\lambda_{\max} < 1$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$. This condition on μ may be rewritten as

$$0 < \mu < \frac{2}{K(2K\text{-}1)\lambda_{\max}E[v^{2(K\text{-}1)}(n)]} \tag{10}$$

When this condition is satisfied, we find that

$$\upsilon_i(\infty) \to 0 \qquad \text{for all } i$$

That is, $\in(\infty) \to 0$ and, correspondingly, $\hat{\mathbf{w}}(\infty) \to \mathbf{w}_o$.

(d) For $K = 1$, the results described in Eqs. (3), (7) and (10) reduce as follows, respectively,

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu\mathbf{u}(n)e(n)$$

$$E[\in(n+1)] = (\mathbf{I} - \mu\mathbf{R})E[\in(n)]$$

$$0 < \mu < \frac{2}{\lambda_{\max}}$$

These results are recognized to be the same as those for the conventional LMS algorithm for real-valued data.

5.6 (a) We start with the equation

$$E[\in(n+1)] = (\mathbf{I} - \mu\mathbf{R})E[\in(n)] \tag{1}$$

where

$$\varepsilon(n) = \mathbf{w}_o - \hat{\mathbf{w}}(n)$$

We note that

$$\approx \frac{\mu J_{\min}}{2} \quad \text{for small } \mu.$$

(b) From the Lyapunov equation derived in Problem 5.10, we have

$$\mathbf{R}\mathbf{K}_0(n) + \mathbf{K}_0(n)\mathbf{R} \approx \mu J_{\min}\mathbf{R}, \qquad \mu \ \text{small}$$

where only the first term of the summation in the right-hand side of Eq. 8 in the solution to Problem 5.10 is retained. Taking the trace of both sides of this equation, and recognizing that

$$\mathrm{tr}[\mathbf{R}\mathbf{K}_0(n)] = \mathrm{tr}[\mathbf{K}_0(n)\mathbf{R}]$$

we get for $n = \infty$:

$$2\mathrm{tr}[\mathbf{R}\mathbf{K}_0(\infty)] \approx \mu J_{\min}\mathrm{tr}[\mathbf{R}]$$

From Eq. (5.90) of the text,

$$J_{ex}(\infty) = \mathrm{tr}[\mathbf{R}\mathbf{K}_0(n)] = \frac{\mu}{2}J_{\min}\mathrm{tr}[\mathbf{R}]$$

Hence, the misadjustment is

$$M \approx \frac{J_{ex}(\infty)}{J_{\min}}$$

$$= \frac{\mu}{2}\mathrm{tr}[\mathbf{R}]$$

5.13 The error correlation matrix **K**(n) equals

$$\mathbf{K}(n) = E[\in(n)\in^H(n)]$$

The trace of **K**(n) equals

$$\mathrm{tr}[\mathbf{K}(n)] = \mathrm{tr}\{E[\in(n)\in^H(n)]\}$$

$$= E\{\mathrm{tr}[\in(n)\in^H(n)]\}$$

Since

$$\text{tr}[\epsilon(n)\,\epsilon^H(n)] = \text{tr}[\epsilon^H(n)\,\epsilon(n)]$$

we may express the trace of $\mathbf{K}(n)$ as

$$\text{tr}[\mathbf{K}(n)] = E\{\text{tr}[\epsilon^H(n)\,\epsilon(n)]\}$$

$$= \text{tr}\{E[\epsilon^H(n)\epsilon(n)]\}$$

The inner product $\epsilon^H(n)\epsilon(n)$ equals the squared norm of $\epsilon(n)$ which is a scalar. Hence

$$\text{tr}[\mathbf{K}(n)] = E[\|\epsilon(n)\|^2] \tag{1}$$

From convergence analysis of the LMS algorithm, we have

$$\mathbf{K}(n+1) = (\mathbf{I}-\mu\mathbf{R})\mathbf{K}(n)(\mathbf{I}-\mu\mathbf{R}) + \mu^2 J_{\min}\mathbf{R} \tag{2}$$

Initially, $\|\epsilon(n)\|$ is so large that we may justifiably ignore the term $\mu^2 J_{\min}\mathbf{R}$, in which case Eq. (2) may be approximated as

$$\mathbf{K}(n+1) = (\mathbf{I}-\mu\mathbf{R})\mathbf{K}(n)(\mathbf{I}-\mu\mathbf{R}), \qquad n \text{ small} \tag{3}$$

Assuming that

$$\mathbf{R} = \sigma_u^2\mathbf{I}$$

we may further reduce Eq. (3) to

$$\mathbf{K}(n+1) \approx (1-\mu\sigma_u^2)^2\mathbf{K}(n)$$

Thus, in light of Eq. (1) we may write

$$E[\|\epsilon(n+1)\|^2] \approx (1-\mu\sigma_u^2)^2 E[\|\epsilon(n)\|^2], \qquad n \text{ small}$$

The convergence ratio is therefore approximately

$$c(n) = \frac{E[\| \in (n+1) \|^2]}{E[\| \in (n) \|^2]}$$

$$\approx (1 - \mu\sigma_u^2)^2$$

n

w