ADAPTIVE FILTERS

Ali H. Sayed

Electrical Engineering Department
University of California, Los Angeles

A. H. Sayed, *Adaptive Filters*, Wiley, NJ, 2008.

ADDITIONAL PROBLEMS
version: March 2015

This file contains additional problems that instructors may find useful. The problems originate from various examination questions on adaptation and learning in the graduate level course *EE210A: Adaptation and Learning* taught by the author at UCLA Electrical Engineering. The problems are generally designed to test deeper understanding of core concepts in adaptation and estimation theories. This file is updated on a regular basis.

The statement of each problem is preceded by the chapter number(s) whose material is tested by the problem.

Solutions are available to instructors only. Please contact the author at sayed@ee.ucla.edu.

ADDITIONAL PROBLEMS

Last updated March 2015

1. (**Chapter 29**) Consider a regularized least-squares problem of the form:

$$\min_w \left[\rho\lambda^N\|w\|^2 \;+\; \sum_{m=0}^{N-1}\lambda^{N-1-m}\left(d(m)-u_m w\right)^2\right]$$

Show that it is equivalent to the following constrained optimization problem:

$$\min_w \left[\sum_{m=0}^{N-1}\lambda^{N-1-m}\left(d(m)-u_m w\right)^2\right], \qquad \text{subject to } \|w\|^2\le\alpha$$

where $\alpha > 0$ is a given constant. What is the relation between the regularization factor ρ and α? Show that there is a one-to-one correspondence between the values of ρ and α. What is the significance of this result?

2. (**Chapter 10**) Let W denote an $M \times M$ square invertible matrix with individual entries W_{mk}.

   (a) Show that

$$\frac{\partial \det(W)}{\partial W_{mk}} \;=\; \det(W)\cdot\left[W^{-1}\right]_{km}$$

   in terms of the $(k,m)-$th entry of W^{-1}. What would the result be if W is an orthogonal matrix?

   (b) Consider a random variable $\boldsymbol{y}$ with probability density function, $f_{\boldsymbol{y}}(y)$, and cumulative density function, $F_{\boldsymbol{y}}(y)$. Assume we select the cumulative density function to be in the following logistic form:

$$F(y) \;\triangleq\; \int_{-\infty}^{y} f_{\boldsymbol{y}}(t)dt \;=\; \frac{1}{1+e^{-y}}$$

   If we express the pdf in the form $f_{\boldsymbol{y}}(y) = e^{-q(y)}$, for some function $q(y)$, what are the expressions for $q(y)$ and its derivative, $g(y) = q'(y)$?

   (c) Let $\{w_m^{\mathsf{T}}\}$ denote the rows of W, for $m = 1, 2, \ldots, M$. Consider the following cost function with a matrix argument:

$$J(W) \;\triangleq\; -\ln|\det(W)| \;-\; \sum_{m=1}^{M} \mathsf{E}\,\ln f_{\boldsymbol{y}}\left(w_m^{\mathsf{T}}\boldsymbol{h}\right)$$

   where $\boldsymbol{h} \in \mathbb{R}^M$ represents the feature data, assumed to be independently and identically distributed. Derive a stochastic-gradient recursion for updating W from training data $\{h_n\}$, $n = 0, 1, 2, \ldots, N-1$.

3. (**Chapters 16, 23**) Consider the LMS recursion

$$\boldsymbol{w_i} = \boldsymbol{w_{i-1}} + \boldsymbol{\mu(i)}\boldsymbol{u_i^{\mathsf{T}}}\left(\boldsymbol{d(i)} - \boldsymbol{u_i}\boldsymbol{w_{i-1}}\right)$$

where all variables are real-valued. The step-size $\mu(i)$ is taken to be an independent and identically distributed (i.i.d.) random process with mean $\bar{\mu}$ and variance σ_μ^2.

The regression process $\boldsymbol{u}_i$ is i.i.d. Gaussian with zero mean and covariance matrix $R_u = \sigma_u^2 I > 0$. Moreover, the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ is assumed to satisfy the stationary data model $\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i)$, where $\boldsymbol{u}_i$ and $\boldsymbol{v}(j)$ are independent of each other for all i and j. The power of the zero-mean process $\boldsymbol{v}(i)$ is denoted by σ_v^2. In addition, the step-size variable $\mu(i)$ is assumed to be independent of all random variables in the learning algorithm for any time instant.

(a) Determine exact conditions to ensure mean convergence of $\boldsymbol{w}_i$ towards w^o.

(b) Determine a recursion for $\mathsf{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$.

(c) Determine conditions to ensure the convergence of $\mathsf{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$ to a steady-state value.

(d) Use the recursion of part (b) to determine an exact closed-form expression for the MSD metric of the algorithm, which is defined as the limiting value of $\mathsf{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$ as $i \to \infty$.

(e) Determine an approximation for the MSD metric to first-order in $\bar{\mu}$.

(f) Determine an approximation for the convergence rate to first-order in $\bar{\mu}$.

(g) Assume $\mu(i)$ is Bernoulli and assumes the values μ and 0 with probabilities p and $1-p$, respectively. What are the values of $\bar{\mu}$ and σ_μ^2 in this case? Consider further the traditional LMS algorithm with $\mu(i)$ replaced by the constant value μ. How do the MSD values of these two implementations, with $\mu(i)$ and μ, compare to each other?

4. (**Chapter 1**) All variables in this problem, whether random or deterministic, are scalar and real-valued.

(a) Consider two zero-mean random variables $\{\boldsymbol{x}, \boldsymbol{y}\}$ and let $f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)$ denote the conditional pdf of $\boldsymbol{x}$ given $\boldsymbol{y}$. We already know that the estimator for $\boldsymbol{x}$ that minimizes the mean-square-error, $\mathsf{E}\,(\boldsymbol{x} - h(\boldsymbol{y}))^2$, over $h(\cdot)$ is given by the mean of the conditional distribution $f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)$. We denote this solution by $\widehat{\boldsymbol{x}}_{\text{mean}} = \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})$. Show that the solution to the following alternative problem, where the quadratic measure is replaced by the absolute measure:

$$\widehat{\boldsymbol{x}}_{\text{median}} \;=\; \arg\min_{h(\cdot)} \; \mathsf{E}\,|\boldsymbol{x} - h(\boldsymbol{y})|$$

is given by the median of the same conditional distribution, $f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)$. We recall that, for continuous distributions, the median is the point a such that

$$\int_{-\infty}^{a} f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)dx \;=\; \int_{a}^{\infty} f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)dx \;=\; 1/2$$

(b) Consider now N noisy measurements of an unknown scalar variable x, say, $y(n) = x + v(n)$, and formulate the following two optimization problems:

$$\begin{aligned} \widehat{x}_{\text{mean}} \;&\stackrel{\Delta}{=}\; \arg\min_{x} \; \frac{1}{N}\sum_{n=1}^{N}(y(n) - x)^2 \\ \widehat{x}_{\text{median}} \;&\stackrel{\Delta}{=}\; \arg\min_{x} \; \frac{1}{N}\sum_{n=1}^{N}|y(n) - x| \end{aligned}$$

(b.1) Show that $\widehat{x}_{\text{mean}}$ is equal to the mean of the noisy observations, i.e., $\widehat{x}_{\text{mean}} = \frac{1}{N}\sum_{n=1}^{N} y(n)$.

(b.2) Show that $\widehat{x}_{\text{median}}$ is equal to the median of the observations, i.e.,

$$\widehat{x}_{\text{median}} = \text{median}\{y(1), y(2), \ldots, y(N)\}$$

where the median is such that an equal number of observations exists to its left and to its right.

5. (**Chapter 30**) All data are real-valued. Consider a regularized least-squares problem of the form:

$$\min_{w,\theta} \left[\rho\lambda^N \|w\|^2 \; + \; \sum_{m=0}^{N-1} \lambda^{N-1-m} \left(d(m) - \theta - u_m w\right)^2 \right]$$

where $\rho > 0$, θ is a scalar, and regularization is applied to w only. Both w and θ are design parameters. Let $\{\widehat{\theta}(N-1), \widehat{w}_{N-1}\}$ denote the estimates for $\{\theta, w\}$ that are based on data up to time $N-1$. Determine recursions that update $\{\widehat{\theta}(N-1), \widehat{w}_{N-1}\}$ to $\{\widehat{\theta}(N), \widehat{w}_N\}$.

6. (**Chapter 29**) All variables are real-valued. Assume H is a square orthogonal matrix, i.e., $H^{\mathsf{T}}H = I = HH^{\mathsf{T}}$. Let $\overline{y}$ denote the transformed vector $\overline{y} = H^{\mathsf{T}}y$, and consider the regularized least-squares problem:

$$\min_{w} \; \left[\rho\|w\|_1 \; + \; \|y - Hw\|^2 \right]$$

where $\rho > 0$, and $\|w\|_1$ denotes the ℓ_1-norm of its vector argument, i.e., it is equal to the sum of the magnitudes of the entries of w. Show that the entries of the solution $\widehat{w} \in \mathbb{R}^M$ are given by:

$$\widehat{w}_m \; = \; \text{sign}(\overline{y}_m) \cdot \left\{ |\overline{y}_m| - \frac{\rho}{2} \right\}_+$$

where the notation $\{x\}_+ = x$ when $x \geq 0$ and is zero otherwise.

7. (**Chapters 4, 7**) All variables are zero mean. Show that for any three random variables $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ it holds that

$$\widehat{\boldsymbol{x}}_{y,z} \; = \; \widehat{\boldsymbol{x}}_y \; + \; \widehat{(\widetilde{\boldsymbol{x}}_y)}_{\widetilde{z}_y}$$

where

$$\begin{cases} \widehat{\boldsymbol{x}}_{y,z} & = & \text{linear least-mean-squares estimator (l.l.m.s.e) of } \boldsymbol{x} \text{ given } \{\boldsymbol{z}, \boldsymbol{y}\}. \\ \widehat{\boldsymbol{x}}_y & = & \text{l.l.m.s.e of } \boldsymbol{x} \text{ given } \boldsymbol{y}. \\ \widehat{\boldsymbol{z}}_y & = & \text{l.l.m.s.e of } \boldsymbol{z} \text{ given } \boldsymbol{y}. \\ \widetilde{\boldsymbol{x}}_y & = & \boldsymbol{x} - \widehat{\boldsymbol{x}}_y \\ \widetilde{\boldsymbol{z}}_y & = & \boldsymbol{z} - \widehat{\boldsymbol{z}}_y \\ \widehat{(\widetilde{\boldsymbol{x}}_y)}_{\widetilde{z}_y} & = & \text{l.l.m.s.e of } \widetilde{\boldsymbol{x}}_y \text{ given } \widetilde{\boldsymbol{z}}_y. \end{cases}$$

What is the geometric interpretation of this result?

8. (**Chapter 5**) All variables are real-valued. Let $\boldsymbol{y}(n) = x + \boldsymbol{v}(n)$, where x is an unknown scalar constant and $\boldsymbol{v}(n)$ is zero-mean white noise with power σ_v^2. An estimator for x is constructed recursively in the following manner:

$$\widehat{\boldsymbol{x}}(n) \; = \; (1-\alpha)\widehat{\boldsymbol{x}}(n-1) \; + \; \alpha\, \boldsymbol{y}(n), \quad n \geq 0$$

starting from $\widehat{\boldsymbol{x}}(-1) = 0$ and where $0 < \alpha < 1$. Determine the steady-state mean and variance of $\widehat{\boldsymbol{x}}(n)$ as $n \to \infty$. Any optimal choice for α?

9. **(Chapter 23)** Consider an LMS update of the form

$$\boldsymbol{w_i} \;=\; \boldsymbol{w_{i-1}} \;+\; \boldsymbol{\mu \cdot \alpha \cdot u_i^*(d(i) - u_i w_{i-1})}, \qquad \boldsymbol{i \geq 0}$$

where the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the stationary data model from Sec. 15.2 in the text, and α is a binary random variable assuming the values 1 and 0 with probabilities p and $1-p$, respectively. In other words, the filter performs updates p fraction of the time. Assume further that the regressors $\{\boldsymbol{u}_i\}$ are independent and identically distributed and arise from a circular Gaussian distribution with covariance matrix R_u.

(a) Determine the condition on the step-size μ to ensure stability in the mean.

(b) Determine the condition on the step-size μ to ensure stability in the mean-square-error sense.

(c) What is the convergence rate of the algorithm?

(d) Determine the learning curve of the filter, i.e., a recursion for $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$ over time.

(e) Determine expressions for the EMSE and MSD performance of the filter for sufficiently small step-sizes.

10. **(Chapter 23)** Consider a network of two LMS-type nodes running the following interlaced recursions:

$$\begin{aligned}
\boldsymbol{w}_{1,i} &= a\boldsymbol{w}_{1,i-1} + (1-a)\boldsymbol{w}_{2,i-1} \;+\; \mu \boldsymbol{u}_{1,i}^*(\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i}\boldsymbol{w}_{1,i-1}) \\
\boldsymbol{w}_{2,i} &= (1-a)\boldsymbol{w}_{1,i-1} + a\boldsymbol{w}_{2,i-1} \;+\; \mu \boldsymbol{u}_{2,i}^*(\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i}\boldsymbol{w}_{2,i-1})
\end{aligned}$$

where the data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ at node k satisfy the stationary data model from Sec. 15.2 in the text with the same vector w^o and with noise variances $\sigma_{v,k}^2$ and regression covariance matrices $R_{u,k}$ for $k = 1, 2$. Moreover, the combination coefficient a satisfies $0 \leq a \leq 1$ and $\boldsymbol{w}_{k,i}$ denotes the weight estimator for w^o computed by node k at time i. Assume the $\{\boldsymbol{u}_{k,i}\}$ and the $\{\boldsymbol{v}_k(i)\}$ random processes are each independent over both time and space, and that $\{\boldsymbol{u}_{k,i}, \boldsymbol{v}_m(j)\}$ are independent of each other for all i, j, k, m.

(a) Determine the condition on the step-size μ to ensure stability in the mean.

(b) Determine the condition on the step-size μ to ensure stability in the mean-square-error sense.

(c) What is the convergence rate of the algorithm?

(d) Determine the learning curve of the filter, i.e., a recursion for the average excess-mean-square-error of both nodes over time.

(e) Determine an expression for the MSD performance of the filter for sufficiently small step-sizes, which is defined as

$$\text{MSD} \;\stackrel{\Delta}{=}\; \lim_{i\to\infty} \frac{1}{2}\sum_{k=1}^{2} \mathsf{E}\,\|\boldsymbol{w}^o - \boldsymbol{w}_{k,i}\|^2$$

(f) Optimize the MSD over a.

(g) Assume $a = 1$ so that both nodes end up running separate LMS recursions. Assume further that $R_{u,1} = R_{u,2} \equiv R_u$. Determine the condition on μ to ensure stability in the mean of both independent filters.

(h) Is the condition of part (g) sufficient to ensure stability of the interlaced filters when $a \neq 1$? Prove or give a counter-example.

(i) Assume $a = 1/2$ and $R_{u,1} = R_{u,2} \equiv R_u$. Compare the MSD performance of the individual LMS filters (corresponding to $a = 1$) with the performance of the cooperative filter using $a = 1/2$.

11. **(Chapter 30)** Two least-squares estimators are out of sync. At any time i, estimator $\#1$ computes the estimate $w_{1,0:i-1}$ that corresponds to the solution of

$$\min_w \left[\delta\lambda^i \|w\|^2 + \sum_{j=0}^{i-1} \lambda^{i-1-j} |d(j) - u_j w|^2 \right] \implies w_{1,0:i-1}$$

where $\delta > 0$ and λ is the forgetting factor. Note that $w_{1,0:i-1}$ is an estimate that is based on measurements between times $j = 0$ and $j = i - 1$. On the other hand, estimator $\#2$ computes the estimate $w_{2,1:i}$ that corresponds to the solution of

$$\min_w \left[\delta\lambda^i \|w\|^2 + \sum_{j=1}^{i} \lambda^{i-j} |d(j) - u_j w|^2 \right] \implies w_{2,1:i}$$

Here, $w_{2,1:i}$ is an estimate that is based on measurements between times $j = 1$ and $j = i$. Can you use the available estimates $\{w_{1,0:i-1}, w_{2,1:i}, i \geq 0\}$ to construct the recursive solution of

$$\min_w \left[\delta\lambda^{i+1} \|w\|^2 + \sum_{j=0}^{i} \lambda^{i-j} |d(j) - u_j w|^2 \right] \implies w_i$$

where w_i is an estimate that is based on all data up to time i? If so, explain the construction. If not, explain why not.

12. **(Appendix B and Chapter 23)** Let H denote a positive-definite Hermitian matrix and let G denote a Hermitian square matrix of compatible dimensions. Show that

$$HG \geq 0 \text{ if, and only if, } G \geq 0$$

where the notation $A \geq 0$ means that all eigenvalues of matrix A are nonnegative.

13. **(Chapters 1 and 3)** A random variable $\boldsymbol{x}$ assumes the value $+1$ with probability p and the value -1 with probability $1-p$. The distribution of a second random variable $\boldsymbol{v}$ depends on the value assumed by $\boldsymbol{x}$. If $\boldsymbol{x} = +1$, then $\boldsymbol{v}$ is normal with zero mean and variance σ_a^2, i.e., $\mathcal{N}(0, \sigma_a^2)$, with probability a and uniformly distributed within the interval $[-\sigma_a, \sigma_a]$ with probability $1-a$. On the other hand, if $\boldsymbol{x} = -1$, then $\boldsymbol{v}$ is $\mathcal{N}(0, \sigma_b^2)$ with probability b and uniformly distributed within the interval $[-\sigma_b, \sigma_b]$ with probability $1 - b$. Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$.

(a) Determine $\mathsf{E}(\boldsymbol{v}|x = +1)$, $\mathsf{E}(\boldsymbol{v}|x = -1)$, and $\mathsf{E}(\boldsymbol{v})$.

(b) Determine $\mathsf{E}(\boldsymbol{y}|x = +1)$, $\mathsf{E}(\boldsymbol{y}|x = -1)$, and $\mathsf{E}(\boldsymbol{y})$.

(c) Determine the optimal mean-square-error estimator, $\hat{\boldsymbol{x}}_{\text{opt}} = \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})$.

(d) Compute σ_x^2, σ_y^2, and $\sigma_{xy} = \mathsf{E}\,(\boldsymbol{x}\boldsymbol{y})$.

(e) Determine the linear least-mean-squares-error (l.l.m.s.e.) estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$.

(f) Compute the minimum mean-square-error (m.m.s.e.) of part (e).

14. **(Chapters 3 and 7)** The output of an FIR filter is described by

$$\boldsymbol{y}(i) = \boldsymbol{x}(i) + \frac{1}{2}\boldsymbol{x}(i-1) + \boldsymbol{v}(i), \quad i > -\infty$$

where $\boldsymbol{v}(i)$ is a zero-mean white random process with variance σ_v^2 and $\boldsymbol{x}(i)$ is also a zero-mean white random process with variance σ_x^2. The variables $\boldsymbol{v}(i)$ and $\boldsymbol{x}(j)$ are independent for all i and j. The filter is assumed to be operating since the remote past so that all processes can be assumed to wide sense stationary. Let

$$\boldsymbol{s}_i \triangleq \begin{bmatrix} \boldsymbol{x}(i) \\ \boldsymbol{x}(i-1) \end{bmatrix}$$

(a) Find the l.l.m.s.e. estimator of $\boldsymbol{s}_i$ given the observations $\{\boldsymbol{y}(i), \boldsymbol{y}(i-1)\}$. What is the resulting m.m.s.e.?

(b) Find the l.l.m.s.e. estimator of $\boldsymbol{s}_i$ given $\{\boldsymbol{y}(i), \boldsymbol{y}(i-1), \boldsymbol{y}(i-2)\}$. What is the resulting m.m.s.e.?

(c) Find the l.l.m.s.e. estimator of $\boldsymbol{s}_i$ given $\{\boldsymbol{y}(i), \boldsymbol{y}(i-1), \ldots, \boldsymbol{y}(i-m)\}$. What is the resulting m.m.s.e.?

(d) Optimize the m.m.s.e. of part (c) over m.

(e) Find the innovations process corresponding to the random process $\boldsymbol{y}(i)$.

15. **(Chapter 7)** Consider the following state-space model:

$$\begin{aligned} \boldsymbol{x}_{i+1} &= \alpha \boldsymbol{x}_i + \boldsymbol{u}_i + d, \quad |\alpha| < 1 \\ \boldsymbol{y}(i) &= \mathbb{1}^{\mathsf{T}}(\boldsymbol{x}_i + \boldsymbol{x}_{i-1}) + \boldsymbol{v}(i), \quad i \geq 0 \end{aligned}$$

with a constant driving factor d of size $n \times 1$, and where $\mathbb{1}$ is the column vector with unit entries and

$$\mathsf{E}\begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}(i) \\ \boldsymbol{x}_o \\ 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{v}(j) \\ \boldsymbol{x}_o \end{bmatrix}^* = \begin{bmatrix} qI_n\delta_{ij} & 0 & 0 \\ 0 & r\delta_{ij} & 0 \\ 0 & 0 & \pi_o I_n \\ 0 & 0 & 0 \end{bmatrix}$$

where $\{q, r, \pi_o\}$ are positive scalars. Determine the innovations process of $\{\boldsymbol{y}(i)\}$.

16. **(Chapters 3, 10, and 16)** All variables are zero-mean. Consider a complex-valued scalar random variable $\boldsymbol{d}$ and a complex-valued $1 \times M$ regression vector $\boldsymbol{u}$. Let

$$\widehat{\boldsymbol{d}} = \boldsymbol{u}w^o$$

denote the linear least-mean-squares error (l.l.m.s.e.) estimator of $\boldsymbol{d}$ given $\boldsymbol{u}$ for some $M \times 1$ vector w^o. Consider additionally the problem of estimating separately the real and imaginary parts of $\boldsymbol{d}$ using knowledge of the real and imaginary parts of $\boldsymbol{u}$, also in the linear least-mean-squares error sense, namely,

$$\widehat{\boldsymbol{d}}_{\text{real}} = \begin{bmatrix} \mathsf{Re}(\boldsymbol{u}) & \mathsf{Im}(\boldsymbol{u}) \end{bmatrix} w^o_{\text{real}}, \quad \widehat{\boldsymbol{d}}_{\text{imag}} = \begin{bmatrix} \mathsf{Re}(\boldsymbol{u}) & \mathsf{Im}(\boldsymbol{u}) \end{bmatrix} w^o_{\text{imag}}$$

for some $2M \times 1$ vectors w^o_{real} and w^o_{imag}.

(a) Argue that estimating the real and imaginary parts of $\boldsymbol{d}$ from the real and imaginary parts of $\boldsymbol{u}$ is equivalent to estimating the real and imaginary parts of $\boldsymbol{d}$ from $\{\boldsymbol{u}, \boldsymbol{u}^*\}$.

(b) What are the optimal choices for w^o, w^o_{real} and w^o_{imag}?

(c) Let $\widehat{\boldsymbol{d}}_2 = \widehat{\boldsymbol{d}}_{\text{real}} + j\widehat{\boldsymbol{d}}_{\text{imag}}$ denote the estimator that is obtained for $\boldsymbol{d}$ from this second construction. What is the corresponding m.m.s.e.? How does it compare to the m.m.s.e. obtained for $\widehat{\boldsymbol{d}} = \boldsymbol{u}w^o$? Under what conditions will both constructions lead to the same m.m.s.e.?

(d) Derive an LMS-type filter for updating separately $M \times 1$ complex weight vectors a^o and b^o when the data $\boldsymbol{d}(i)$, at successive time instants i, are related to the regressors $\{\boldsymbol{u}_i, \boldsymbol{u}_i^*\}$ through the linear model:

$$\boldsymbol{d}(i) = \boldsymbol{u}_i a^o \; + \; (\boldsymbol{u}_i^*)^{\mathsf{T}} b^o \; + \; \boldsymbol{v}(i)$$

Assume the noise is white with variance σ_v^2 (its real and imaginary parts are independent of each other). Assume further that the regressor sequence is temporally white and independent of the noise process.

(e) Assume sufficiently small step-sizes. Derive an expression for the EMSE of this so-called widely-linear LMS filter. How does the performance compare to that delivered by the regular LMS filter, when we assume that $\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i)$ and estimate w^o via the LMS iteration? Under what conditions will both filters have similar EMSE performance?

17. **(Chapter 30)** Node #1 observes even-indexed data $\{d(2n), u_{2n}\}$ for $n \geq 0$ and computes the recursive least-squares solution of

$$\min_w \left[\delta \cdot \lambda^{2n+1} \cdot \|w\|^2 \; + \; \sum_{j=0}^{n} \lambda^{2n-2j} |d(2j) - u_{2j} w|^2 \right] \implies w_{2n}$$

where λ is the forgetting factor. Note that w_{2n} is an estimate that is based solely on the even-indexed data. Likewise, node #2 observes odd-indexed data $\{d(2n+1), u_{2n+1}\}$ for $n \geq 0$ and computes the recursive least-squares solution of

$$\min_w \left[\delta \cdot \lambda^{2n+2} \cdot \|w\|^2 \; + \; \sum_{j=0}^{n} \lambda^{2n-2j} |d(2j+1) - u_{2j+1} w|^2 \right] \implies w_{2n+1}$$

Here, w_{2n+1} is an estimate that is based solely on the odd-indexed data. Can you use the available estimates $\{w_{2n}, w_{2n+1}, n \geq 0\}$ to construct the recursive solution of

$$\min_w \left[\delta \cdot \lambda^{i+1} \cdot \|w\|^2 \; + \; \sum_{j=0}^{i} \lambda^{i-j} |d(j) - u_j w|^2 \right] \implies w_i$$

where w_i is an estimate that is based on all data (both even and odd-indexed) up to time i? If so, explain the construction. If not, explain why not.

18. **(Chapter 23)** Consider a collection of N agents, indexed by $k = 1, 2, \ldots, N$, with each agent k observing zero-mean random processes $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ over time i . The regression data $\{\boldsymbol{u}_{k,i}\}$ are assumed to be temporally white and spatially independent, i.e.,

$$\mathsf{E}\,\boldsymbol{u}_{k,i}\boldsymbol{u}^*_{m,j} = R_{u,k}\delta_{km}\delta_{ij}$$

It is assumed that the data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ satisfy the linear model

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i}w^o + \boldsymbol{v}_k(i)$$

where the noise $\boldsymbol{v}_k(i)$ has zero mean, and is also temporally white and spatially independent:

$$\mathsf{E}\,\boldsymbol{v}_k(i)\boldsymbol{v}^*_m(j) = \sigma^2_{v,k}\delta_{km}\delta_{ij}$$

The noise process is further assumed to be independent of all other random processes. The agents are interested in determining w^o by seeking the $M \times 1$ column vector w that minimizes $\mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2$. Two modes of operation are considered.

In mode **A**, each agent k operates individually and uses the standard LMS iteration:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k\boldsymbol{u}^*_{k,i}[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}], \quad i \geq 0$$

where μ_k is the positive step-size used by node k and $\boldsymbol{w}_{k,i}$ is the estimator of w^o at time i.

In mode **B**, the agents cooperate with each other as follows. The nodes are assumed to be connected by some topology and each node is allowed to share information with its neighbors. The set of neighbors of any node k is denoted by $\mathcal{N}_k$ and it consists of all nodes that are connected to k by edges. Each agent then runs instead the following diffusion LMS iteration:

$$\begin{aligned}
\psi_{k,i} &= \boldsymbol{w}_{k,i-1} + \mu_k\boldsymbol{u}^*_{k,i}[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}], \;\; i \geq 0 \\
\boldsymbol{w}_{k,i} &= \sum_{\ell=1}^{N} a_{\ell k}\psi_{\ell,i}
\end{aligned}$$

where the coefficients $\{a_{\ell k}\}$ satisfy

$$a_{\ell k} \geq 0, \quad \sum_{\ell\in\mathcal{N}_k} a_{\ell k} = 1, \quad a_{\ell k} = 0 \text{ if } \ell \notin \mathcal{N}_k$$

The corresponding matrix $A = [a_{\ell k}]$ is said to be left-stochastic (the entries of each of its columns add up to one). In this algorithm, the first step is similar to the LMS iteration except that it generates an intermediate vector estimate $\psi_{k,i}$. The second step combines all intermediate estimates from the neighbors of node k to obtain $\boldsymbol{w}_{k,i}$.

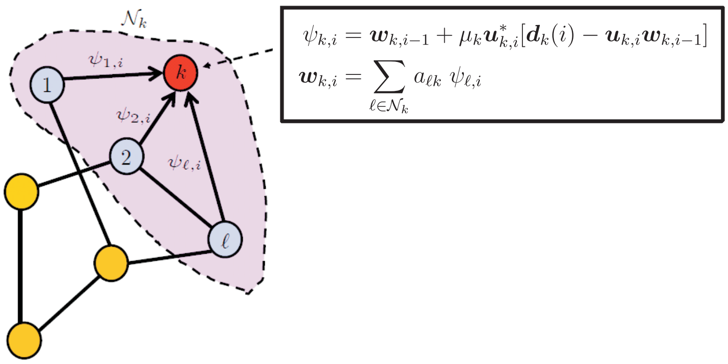


FIGURE 0.1 A network of cooperating nodes running the diffusion LMS algorithm.

(a) Introduce the weight error vector

$$\widetilde{\boldsymbol{w}}_i \triangleq \begin{bmatrix} \widetilde{\boldsymbol{w}}_{1,i} \\ \widetilde{\boldsymbol{w}}_{2,i} \\ \vdots \\ \widetilde{\boldsymbol{w}}_{N,i} \end{bmatrix}$$

and define the extended matrix $\mathcal{A} = A \otimes I$ in terms of the Kronecker product operation $\otimes$. Find recursions for $\mathsf{E}\,\widetilde{\boldsymbol{w}}_i$ under both modes of operation.

(b) Provide conditions on the step-size parameters $\{\mu_k\}$ to ensure mean stability in each mode of operation.

(c) Which mode of operation exhibits faster convergence of $\mathsf{E}\,\widetilde{\boldsymbol{w}}_i$ towards zero?

(d) Prove or give a counter-example: the mean weight-error vector $\mathsf{E}\,\widetilde{\boldsymbol{w}}_{k,i}$ of each node k will converge faster towards zero in the diffusion LMS case than in the non-cooperative LMS case.

(e) Assume each node employs a more sophisticated diffusion mechanism of the following alternative form:

$$\begin{aligned} \psi_{k,i} &= \boldsymbol{w}_{k,i-1} + \mu_k \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \boldsymbol{u}_{\ell,i}^*[\boldsymbol{d}_\ell(i) - \boldsymbol{u}_{\ell,i}\boldsymbol{w}_{k,i-1}], \quad i \geq 0 \\ \boldsymbol{w}_{k,i} &= \sum_{\ell \in \mathcal{N}_k} a_{\ell k}\psi_{\ell,i} \end{aligned}$$

where the coefficients $\{c_{\ell k}\}$ satisfy

$$c_{\ell k} \geq 0, \quad \sum_{k=1}^{N} c_{\ell k} = 1, \quad \sum_{\ell=1}^{N} c_{\ell k} = 1, \quad c_{\ell k} = 0 \text{ if } \ell \notin \mathcal{N}_k$$

The matrix C is said to be doubly-stochastic (each of its rows and columns add up to one). Does using $C \neq I$ improve the rate of convergence of $\mathsf{E}\,\widetilde{\boldsymbol{w}}_i$ towards zero over the case $C = I$ in the diffusion case?

(f) Prove or give a counter-example: Using $C \neq I$ in the diffusion case always results in faster convergence of $\mathsf{E}\,\widetilde{\boldsymbol{w}}_i$ towards zero relative to the non-cooperative case.

19. **(Appendix B)** Conditions of the following form are generally useful in the study of stable adaptive schemes over networks. Let A_1 and A_2 be $N \times N$ left-stochastic matrices (i.e., the entries on each of the columns of A_1 and A_2 add up to one). Define $\mathcal{A}_1 = A_1 \otimes I$ and $\mathcal{A}_2 = A_2 \otimes I$. Prove or disprove: For matrices of compatible dimensions, if $\mathcal{D}$ is any stable matrix, then $\mathcal{A}_1^T \mathcal{D} \mathcal{A}_2^T$ is also a stable matrix.

20. **(Chapters 1 and 3)** A random variable $\boldsymbol{x}$ assumes the value $+1$ with probability p and the value -1 with probability $1-p$. The variable $\boldsymbol{x}$ is observed under additive noise, say, as $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{v}$ has mean $\bar{v}$ and variance σ_v^2. Both $\boldsymbol{x}$ and $\boldsymbol{v}$ are independent of each other. In this problem, we consider different probability density functions (pdf) for $\boldsymbol{v}$.

    (a) Assume initially that $\boldsymbol{v}$ is Gaussian. What is the optimal mean-square-error estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$? What is the corresponding minimum mean-square-error (m.m.s.e.)?

    (b) Assume instead that $\boldsymbol{v}$ is uniformly distributed. What is the optimal mean-square-error estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$? What is the corresponding m.m.s.e.?

    (c) Assume now that $\boldsymbol{v}$ is exponentially distributed. What is the optimal mean-square-error estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$? What is the corresponding m.m.s.e.?

    (d) Which noise distribution results in the smallest m.m.s.e.? How would your conclusion differ if we instead compute the m.m.s.e. values that result from the optimal *linear* mean-square-error estimators in the three cases (a)–(c)?

    (e) Over all possible pdfs, determine the pdf of noise that results in the smallest m.m.s.e. for optimal estimation.

21. **(Chapter 3)** Consider two nodes $\{A_1, A_2\}$ and assume each node has an unbiased estimator, $\{\boldsymbol{w}_k, k = 1, 2\}$ for some $M \times 1$ column vector w^o. Let $\{P_k, k = 1, 2\}$ denote the error covariance matrix, $P_k = \mathsf{E}\,(w^o - \boldsymbol{w}_k)(w^o - \boldsymbol{w}_k)^*$. Assume the errors of the two estimators are uncorrelated, i.e., $\boldsymbol{E}(w^o - \boldsymbol{w}_1)(w^o - \boldsymbol{w}_2)^* = 0$. Consider an aggregate estimator of the form

$$\widehat{\boldsymbol{w}} = \alpha \boldsymbol{w}_1 \; + \; (1-\alpha)\boldsymbol{w}_2$$

    (a) If α is nonnegative, determine the optimal scalar α that minimizes the mean-square-error, i.e.,

$$\min_{\alpha \geq 0} \; \mathsf{E}\,\|w^o - \widehat{\boldsymbol{w}}\|^2$$

    (b) Repeat part (a) when α is not restricted to being nonnegative. When would a negative α be advantageous?

    (c) Let $P = \mathsf{E}\,(w^o - \widehat{\boldsymbol{w}})(w^o - \widehat{\boldsymbol{w}})^*$. How does P compare to P_1 and P_2 in both cases (a) and (b)?

    (d) Now assume the errors of the two estimators are *correlated* instead, i.e., $\mathsf{E}\,(w^o - \boldsymbol{w}_1)(w^o - \boldsymbol{w}_2)^* = C$. Repeat parts (a)–(c).

22. **(Chapter 5)** All variables are zero-mean. Let

$$\begin{bmatrix} \boldsymbol{y}_a \\ \boldsymbol{y} \\ \boldsymbol{y}_b \end{bmatrix} = \begin{bmatrix} H_a \\ H \\ H_b \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \boldsymbol{v}_a \\ \boldsymbol{v} \\ \boldsymbol{v}_b \end{bmatrix}$$

where $\{\boldsymbol{v}_a, \boldsymbol{v}, \boldsymbol{v}_b\}$, are uncorrelated with $\boldsymbol{x}$ and have zero mean and covariance matrices:

$$\mathsf{E}\begin{bmatrix}\boldsymbol{v}_a\\ \boldsymbol{v}\\ \boldsymbol{v}_b\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_a\\ \boldsymbol{v}\\ \boldsymbol{v}_b\end{bmatrix}^* = \begin{bmatrix}R_a & S_a & 0\\ S_a^* & R & S_b\\ 0 & S_b^* & R_b\end{bmatrix}$$

Let $\widehat{\boldsymbol{x}}_{\boldsymbol{y}_a,\boldsymbol{y}}$ denote the linear estimator of $\boldsymbol{x}$ given $\{\boldsymbol{y}_a, \boldsymbol{y}\}$. Let $\widehat{\boldsymbol{x}}_{\boldsymbol{y}_b,\boldsymbol{y}}$ denote the linear estimator of $\boldsymbol{x}$ given $\{\boldsymbol{y}_b, \boldsymbol{y}\}$. Can you relate these estimators, and their minimum-mean-square-error, to each other?

23. **(Chapters 17 and 24)** Consider an LMS filter with a real-valued regression vector $\boldsymbol{u}_i = [\boldsymbol{u}(i)\ \boldsymbol{u}(i-1)]$, where the entries of $\boldsymbol{u}_i$ arise from a wide-sense stationary stochastic process described by

$$\boldsymbol{u}(i) = \boldsymbol{A}\cos(\omega i + \boldsymbol{\theta})$$

The random variable $\boldsymbol{A}$ has mean A and variance σ_a^2. The random variable $\boldsymbol{\theta}$ is uniformly distributed over $[-\pi, \pi]$. The random variables $\boldsymbol{A}$ and $\boldsymbol{\theta}$ are independent of each other and of all other random variables appearing in this problem.

(a) Use the separation principle to find an expression for the EMSE of the filter. How does it depend on ω?

(b) Use the small-step size approximation to determine an expression for the learning curve of the filter.

(c) How does the convergence rate of the filter depend on ω?

24. **(Chapter 30)** Consider the optimization problem

$$\min_{w}\ \lambda^{N+1}w^*\Pi w + \mathsf{E}\left(\sum_{i=0}^{N}\lambda^{N-i}|d(i) - \boldsymbol{\alpha}u_i w|^2\right) \Longrightarrow w_N$$

where the data $\{d(i), u_i\}$ are deterministic (not random) measurements with $d(i)$ a scalar and u_i a row vector of size $1 \times M$. The random variable $\boldsymbol{\alpha}$ is Bernoulli and assumes the value 1 with probability p and the value 0 with probability $1-p$; it is used to model a faulty sensor – when the sensor fails, no regression data is measured. Let w_N denote the solution. Can you determine a recursion to go from w_N to w_{N+1}?

25. **(Chapters 15, 16, 22, 23)** Consider three nodes, $k = 1$, 2, and 3. Each node k collects data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ that satisfy the linear model $\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i}w^o + \boldsymbol{v}_k(i)$, where w^o is $M \times 1$ and $\boldsymbol{u}_{k,i}$ is $1 \times M$. The processes $\boldsymbol{v}_k(i)$ are zero mean white noises that are independent of each other and have variances $\sigma_{v,k}^2$. At every iteration i, the nodes exchange the output estimation errors $\{\boldsymbol{e}_k(i)\}$, defined as

$$\boldsymbol{e}_k(i) = \boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}$$

and each node updates its weight estimate according to the following rule:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu\boldsymbol{u}_{k,i}^*\boldsymbol{e}(i)$$

where

$$\boldsymbol{e}(i) = \alpha_1\boldsymbol{e}_1(i) + \alpha_2\boldsymbol{e}_2(i) + \alpha_3\boldsymbol{e}_3(i), \qquad \alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \text{and} \quad \alpha_k \geq 0$$

Under reasonable assumptions:

(a) Determine expressions for the individual EMSEs of the nodes. Determine the average EMSE.

(b) Determine a condition on the step-size μ to ensure mean-square convergence of all nodes.

(c) Can you pick values for $\{\alpha_1, \alpha_2, \alpha_3\}$ to minimize the average EMSE?

26. **(Chapter 3)** Consider a collection of N independent and identically-distributed real-valued random variables, $\{\boldsymbol{y}(n),\ n = 0, 1, \ldots, N-1\}$. Each $\boldsymbol{y}(n)$ has a Gaussian distribution with zero mean and variance σ^2. We want to use the observations $\{\boldsymbol{y}(n)\}$ to estimate the variance σ^2 in the following manner:

$$\widehat{\boldsymbol{\sigma}}^2 = \alpha \cdot \left(\sum_{n=0}^{N-1} \boldsymbol{y}^2(n) \right)$$

for some scalar parameter α to be determined.

(a) What is the mean of the estimator $\widehat{\boldsymbol{\sigma}}^2$ in terms of α and σ^2?

(b) Evaluate the mean-square-error MSE below in terms of α and σ^2:

$$\mathsf{MSE} = \mathsf{E}|\widehat{\boldsymbol{\sigma}}^2 - \sigma^2|^2$$

(c) Determine the optimal scalar α that minimizes the MSE. Is the corresponding estimator biased or unbiased?

(d) For what value of α would the estimator be unbiased? What is the MSE of this estimator and how does it compare to the MSE of the estimator from part (c)?

(e) What do you learn from this problem?

27. **(Chapter 5)** Let $\boldsymbol{y}_1 = H_1\boldsymbol{x} + \boldsymbol{v}_1$ and $\boldsymbol{y}_2 = H_2\boldsymbol{x} + \boldsymbol{v}_2$ denote two linear observation models with the same unknown random vector $\boldsymbol{x}$. All random variables have zero-mean. The covariance and cross-covariance matrices of $\{\boldsymbol{x}, \boldsymbol{v}_1, \boldsymbol{v}_2\}$ are denoted by

$$\mathsf{E}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{v}_1\\ \boldsymbol{v}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{v}_1\\ \boldsymbol{v}_2\end{bmatrix}^* = \begin{bmatrix}R_x & 0 & 0\\ 0 & R_1 & C\\ 0 & C^* & R_2\end{bmatrix}$$

In particular, observe that we are assuming the noises to be correlated with $C = \mathsf{E}\boldsymbol{v}_1\boldsymbol{v}_2^*$. All covariance matrices are assumed to be invertible whenever necessary.

(a) Show how you would replace the observation vectors $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$ by two other observation vectors $\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$ of similar dimensions such that they satisfy linear models of the form

$$\boldsymbol{z}_1 = G_1\boldsymbol{x} + \boldsymbol{w}_1, \quad \boldsymbol{z}_2 = H_2\boldsymbol{x} + \boldsymbol{w}_2$$

for a matrix G_1 to be specified, and where the noises $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ are now uncorrelated. What are the covariance matrices of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in terms of R_1 and R_2?

(b) Let $\hat{\boldsymbol{x}}_1$ be the linear least-mean-squares estimator (l.l.m.s.e.) of $\boldsymbol{x}$ given $\boldsymbol{z}_1$ with error covariance matrix P_1. Similarly, let $\hat{\boldsymbol{x}}_2$ be the l.l.m.s.e. of $\boldsymbol{x}$ given

$\boldsymbol{z}_2$ with error covariance matrix P_2. Let further $\hat{\boldsymbol{x}}$ denote the l.l.m.s.e. of $\boldsymbol{x}$ given $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$ with error covariance matrix P. Determine expressions for $\hat{\boldsymbol{x}}$ and P in terms of $\{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, P_1, P_2, C, R_x, R_1, R_2\}$.

28. **(Chapter 5)** Let $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{v}$. All random variables have zero-mean. The covariance and cross-covariance matrices of $\{\boldsymbol{x}, \boldsymbol{v}\}$ are denoted by

$$\mathsf{E}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{v}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{v}\end{bmatrix}^* = \begin{bmatrix}R_x & C\\ C^* & R_v\end{bmatrix}$$

with positive-definite R_x and R_v.

(a) What is the l.l.m.s.e. of $\boldsymbol{x}$ given $\boldsymbol{y}$? What is the corresponding m.m.s.e.?

(b) A new scalar observation is added to $\boldsymbol{y}$ and a new row vector is added to H such that

$$\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{\alpha}\end{bmatrix} = \begin{bmatrix}H\\ u\end{bmatrix}\boldsymbol{x} + \begin{bmatrix}\boldsymbol{v}\\ \gamma\end{bmatrix}$$

where γ is uncorrelated with all other variables and has variance σ_γ^2. Let $\hat{\boldsymbol{x}}_{\text{new}}$ denote the new estimator of $\boldsymbol{x}$ given $\{\boldsymbol{y}, \boldsymbol{\alpha}\}$. Relate $\hat{\boldsymbol{x}}_{\text{new}}$ to $\hat{\boldsymbol{x}}$ from part (a). Relate also their m.m.s.e.

29. **(Chapters 8, 10)** The initial condition is $w_{-1} = 0$. Consider the following two algorithms:

$$\text{Alg. I}: \quad \left\{\begin{array}{l}\text{For every time } i \geq 0:\\ \quad \text{Start with } w_{0,i} = w_{i-1} \text{ and iterate over } k = 1, 2, \ldots, N:\\ \quad w_{k,i} = w_{k-1,i} + \mu u_{k,i}^*[d_k(i) - u_{k,i}w_{i-1}]\\ \quad \text{set } w_i \leftarrow w_{N,i}\\ \text{repeat}\end{array}\right.$$

and

$$\text{Alg. II}: \quad \left\{\begin{array}{l}\text{For every time } i \geq 0:\\ \quad \text{Start with } \phi_{0,i} = w_{i-1} \text{ and iterate over } k = 1, 2, \ldots, N:\\ \quad \phi_{k,i} = \phi_{k-1,i} + \mu u_{k,i}^*[d_k(i) - u_{k,i}\phi_{k-1,i}]\\ \quad \text{set } w_i \leftarrow \phi_{N,i}\\ \text{repeat}\end{array}\right.$$

Which of the above algorithms would result directly from a stochastic gradient approximation to the problem:

$$\min_w \sum_{k=1}^{N} \mathsf{E}|\boldsymbol{d}_k - \boldsymbol{u}_k w|^2$$

where all variables are zero-mean and $\{d_k(i), u_{k,i}\}$ are realizations of $\{\boldsymbol{d}_k, \boldsymbol{u}_k\}$. What is the difference between both algorithms?

30. **(Chapters 22, 24)** Consider an LMS filter with a real-valued regression vector $\boldsymbol{u}_i = [\boldsymbol{u}_1(i) \quad \boldsymbol{u}_2(i)]$. The first entry $\boldsymbol{u}_1(i)$ is an i.i.d. random variable distributed according to

$$\boldsymbol{u}_1(i) = \begin{cases} a & \text{with probability } 1/3 \\ -\dfrac{a}{2} & \text{with probability } 2/3 \end{cases}$$

The second entry $\boldsymbol{u}_2(i)$ is also i.i.d., independent of $\boldsymbol{u}_1(i)$ and has the same distribution as $\boldsymbol{u}_1(i)$.

(a) Find the exact condition that the step-size μ should satisfy to ensure mean-square stability.

(b) Find also exact expressions for the filter's EMSE and MSD.

(c) Find the number of iterations that the filter needs for its mean-square error to be within 10% of its final value.

31. **(Chapters 30, 35)** Consider an unknown $M \times 1$ vector $w = \text{col}\{w_1, w_2\}$, where w_1 is $L \times 1$. Introduce the least-squares problem:

$$\min_{w} \; w_1^* \Pi w_1 + \|y_N - H_N w\|^2 + \|d_N - G_N w_1\|^2$$

where

$$y_N = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad d_N = \begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(N) \end{bmatrix}, \quad H_N = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{bmatrix}, \quad G_N = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_N \end{bmatrix}$$

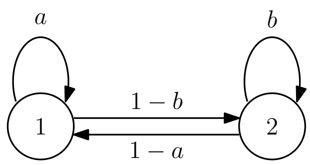
Let w_N denote the solution and let $\zeta(N)$ be the resulting minimum cost.

(a) Relate w_N to w_{N-1}.

(b) Relate $\zeta(N)$ to $\zeta(N-1)$.

(c) Motivate and derive an array algorithm for solving the least-squares problem recursively.

32. **(Chapters 16, 23)** Consider the network shown in the figure with two adaptive nodes. The network functions as follows.



Consider node 1. It starts with a weight estimator $\boldsymbol{w}_{i-1}^{(1)}$ and updates it to $\boldsymbol{w}_i^{(1)}$ in two steps. First, node 1 combines its estimator with the estimator of node 2 and generates an intermediate estimator $\phi_{i-1}^{(1)}$, say,

$$\boldsymbol{\phi}_{i-1}^{(1)} = a\boldsymbol{w}_{i-1}^{(1)} + (1-a)\boldsymbol{w}_{i-1}^{(2)}$$

where a is some real scalar. Node 2 performs a similar step to obtain its intermediate estimator:

$$\boldsymbol{\phi}_{i-1}^{(2)} = (1-b)\boldsymbol{w}_{i-1}^{(1)} + b\boldsymbol{w}_{i-1}^{(2)}$$

Subsequently, each node performs an LMS-type update:

$$\begin{aligned} \boldsymbol{w}_i^{(1)} &= \phi_{i-1}^{(1)} + \mu_1 \boldsymbol{u}_{1,i}^* [\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i}\phi_{i-1}^{(1)}] \\ \boldsymbol{w}_i^{(2)} &= \phi_{i-1}^{(2)} + \mu_2 \boldsymbol{u}_{2,i}^* [\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i}\phi_{i-1}^{(2)}] \end{aligned}$$

where the $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ are the data collected at node k; they are assumed to satisfy the linear model

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_k(i)$$

where the noise components are temporally and spatially white with variances $\sigma_{v,k}^2$. The noise components are further assumed to be independent of all other data. All data are zero-mean and circular. Assume all regressors are Gaussian i.i.d. with a diagonal covariance matrix $\sigma_u^2 \cdot \mathrm{I}$. For all parts below, except for the bonus part, assume $a = b$.

(a) Find conditions on the step-sizes $\{\mu_1, \mu_2\}$ and the combination scalar a that will ensure that the estimators $\boldsymbol{w}_i^{(1)}$ and $\boldsymbol{w}_i^{(2)}$ are asymptotically unbiased for any initial condition.

(b) Assume from now on that $\mu_1 = \mu_2 = \mu$. Find conditions on the step-size μ and the combination scalar a that will ensure that the estimators $\boldsymbol{w}_i^{(1)}$ and $\boldsymbol{w}_i^{(2)}$ converge in the mean-square sense.

(c) Find expressions for the steady-state EMSE and MSD for each node in the network.

(d) Find the optimal value of a that minimizes the individual EMSEs.

(e) How do the EMSE values from part (c) depend on the relation between $\sigma_{v,1}^2$ and $\sigma_{v,2}^2$? In particular, assume $\sigma_{v,1}^2 = \alpha \cdot \sigma_{v,2}^2$ and compare the EMSE of node 1 with the EMSE of node 2 as a function of the positive scalar α for all three cases: $\alpha > 1$, $\alpha = 1$, and $\alpha < 1$.

(f) Repeat part (e) for the optimal EMSE values from part (d).

(g) Assume $\sigma_{v,1}^2 = \sigma_{v,2}^2 = \sigma_v^2$. How does the EMSE obtained in part (c) compare to the value that would result if the nodes were operating independently of each other and without combining their estimators (i.e., using $a = 1$)?

(h) Assume $a \neq b$. Repeat parts (c) and (e).

33. **(Chapters 1, 2)** The state of some system of interest is described by a binary variable θ, which can be either 0 or 1 with equal probability. Let $\boldsymbol{y}$ be a random variable that is observed according to the following probability distribution:

	$\boldsymbol{y} = 0$	$\boldsymbol{y} = 1$
$\theta = 0$	q	$1 - q$
$\theta = 1$	$1 - q$	q

We collect N independent observations $\{\boldsymbol{y}(1), \ldots, \boldsymbol{y}(N)\}$. We wish to employ these observations in order to learn about the state of the system. Assume the true state is $\theta = 0$.

(a) Find the optimal mean-square-error estimator of θ given these observations, namely,

$$\widehat{\theta}_N = \mathsf{E}\,[\theta|\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(N)]$$

(b) Assume $q \neq 0.5$. Show that $\widehat{\theta}_N$ decays to zero exponentially in the mean-square-error sense as $N \to \infty$. That is, verify that the mean-square error converges to zero at an exponential rate. What happens when $q = 0.5$?

(c) Find an expression for the variance of $\widehat{\theta}_N$. Find its limit as $N \to \infty$.

(d) Why are these results useful?

34. **(Chapters 3-5)** The state of some system of interest is described by a Gaussian random variable θ with mean $\bar{\theta}$ and variance σ_θ^2. An observation of θ is collected under additive white Gaussian noise, namely,

$$\boldsymbol{y} = \boldsymbol{\theta} + \boldsymbol{v}$$

where $\boldsymbol{v}$ has zero-mean and variance σ_v^2 and is independent of θ. All variables are real-valued.

(a) Show that the optimal mean-square-error estimator, $\widehat{\boldsymbol{\theta}} = \mathsf{E}\,(\boldsymbol{\theta}|\boldsymbol{y})$, has a Gaussian distribution. Find the mean and variance of this distribution in terms of $\mathsf{SNR} = \sigma_\theta^2/\sigma_v^2$.

(b) Find the optimal mean-square-error estimator of θ given N independent observations $\{\boldsymbol{y}(1), \ldots, \boldsymbol{y}(N)\}$.

(c) Determine the variance of the estimator in part (b) and evaluate its limit as $N \to \infty$.

35. **(Chapters 3-6)** Consider N nodes $\{A_1, A_2, \ldots, A_N\}$. Each node has an unbiased estimate of some unknown column vector w^o. We denote the individual estimator at node A_k by $\boldsymbol{w}_k$. We also denote the covariance matrix of $\boldsymbol{w}_k$ by P_k and the cross-covariance matrix of $\boldsymbol{w}_k$ and $\boldsymbol{w}_\ell$ by $P_{k\ell}$. A node S wishes to combine the estimators $\{\boldsymbol{w}_k, k = 1, \ldots, N\}$ as follows:

$$\widehat{\boldsymbol{w}}_S = \sum_{k=1}^{N} a_k \boldsymbol{w}_k$$

in order to optimize the cost function

$$\min_{\{a_k\}} \mathsf{E}\left\| w^o - \sum_{k=1}^{N} a_k \boldsymbol{w}_k \right\|^2$$

where the $\{a_k\}$ are real-valued scalars.

(a) Find a condition on the coefficients $\{a_k\}$ to ensure that the resulting $\widehat{\boldsymbol{w}}_S$ is an unbiased estimator for w^o.

(b) Under condition (a), find the optimal coefficients $\{a_k\}$. Your solution should not depend on w^o.

(c) Assume the reliability of each estimator $\boldsymbol{w}_k$ is measured by the scalar $\sigma_k^2 = \text{Tr}(P_k)$. The smaller the σ_k^2, the more reliable the estimator is. What is the relation between the optimal coefficients $\{a_k\}$ and the reliability factors $\{\sigma_k^2\}$?

(d) Evaluate the reliability of the estimator $\widehat{\boldsymbol{w}}_S$.

(e) Motivate and derive an adaptive filter for updating the coefficients $\{a_k\}$ in part (b).

(f) How is the estimator of part (b) different from the unbiased linear least-mean-squares estimator of w^o based on the $\{\boldsymbol{w}_k\}$? Find the latter estimator.

(g) Find the minimum MSEs of the estimators in parts (b) and (f) for the case where $P_{k\ell} = 0$ when $\ell \neq k$. Specialize your result to the case $P_k = P$ for all k and compare the resulting MSEs.

36. **(Chapters 1, 3, 5)** Consider two real-valued scalar random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The random variable $\boldsymbol{x}_1$ assumes the value $+1$ with probability p and the value -1 with probability $1-p$. The random variable $\boldsymbol{x}_2$ is distributed as follows:

$$\text{if } \boldsymbol{x}_1 = +1 \text{ then} \quad \boldsymbol{x}_2 = \begin{cases} +2 \text{ with probability } q \\ -2 \text{ with probability } 1-q \end{cases}$$

$$\text{if } \boldsymbol{x}_1 = -1 \text{ then} \quad \boldsymbol{x}_2 = \begin{cases} +3 \text{ with probability } r \\ -3 \text{ with probability } 1-r \end{cases}$$

Consider further the variables

$$\boldsymbol{y}_1 = \boldsymbol{x}_1 + \boldsymbol{v}_1 \quad \text{and} \quad \boldsymbol{y}_2 = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{v}_2$$

where $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ are independent zero-mean Gaussian random variables with unit variance. The variables $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ are independent of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

(a) Express the pdfs of the individual random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in terms of delta functions.

(b) Find the joint pdf of $(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

(c) Find the joint pdf of $(\boldsymbol{y}_1, \boldsymbol{y}_2)$.

(d) Find the joint pdf of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2)$.

(e) Find the conditional pdf of $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ given $(\boldsymbol{y}_1, \boldsymbol{y}_2)$.

(f) Find the minimum mean-square error estimator of $\boldsymbol{x}_2$ given $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$.

(g) Find the minimum mean-square error estimator of $\boldsymbol{x}_2$ given $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_1\}$.

(h) Find the linear least-mean-squares error estimator of $\boldsymbol{x}_2$ given $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_1\}$.

37. **(Chapters 15, 16)** Consider a NLMS recursion of the form

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \frac{\mu \boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2}[\boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}]$$

with $1 \times M$ regression vectors $\boldsymbol{u}_i$ and step-size μ. Each entry of $\boldsymbol{u}_i$ has the form $re^{j\theta}$, where θ is uniformly distributed over $[0, 2\pi]$ and $r > 0$. In other words, the entries of $\boldsymbol{u}_i$ lie on a circle of radius r. Assume the data $\boldsymbol{d}(i)$ satisfy the stationary data model of Section 6.2.

(a) Find an *exact* expression for the EMSE of NLMS under such conditions.

(b) Does the value of r have an influence on the EMSE? Is there an optimal choice for r?

(c) The entries of the regression vectors are further assumed to be independent of each other. Find an *exact* condition on the step-size μ to ensure mean-square convergence.

(d) Which algorithm will have the lower MSE in steady-state for the same step-size: LMS or NLMS?

(e) Evaluate the number of iterations that are needed for LMS to be within 5% of its EMSE? What about NLMS? Which algorithm converges faster? Assume the same step-size for both algorithms.

38. **(Chapters 35, 40, 41)** Consider a least-squares problem of the form

$$\min_{w_M}\left[\tau\lambda^{i+1}\|w_M\|^2 \;+\; \sum_{j=0}^{i}\lambda^{i-j}|d(j) - u_{M,j}w_M|^2\right]$$

where $\tau > 0$ is a regularization parameter, $0 \ll \lambda \leq 1$ is a forgetting factor, and $u_{M,j}$ is a $1 \times M$ regression vector. We denote the solution vector by $w_{M,i}$. Introduce the data matrix

$$H_{M,i} = \begin{bmatrix} u_{M,0} \\ u_{M,1} \\ \vdots \\ u_{M,i} \end{bmatrix} \qquad ((i+1) \times M)$$

and partition $H_{M+1,i}$ as follows:

$$H_{M+1,i} \;=\; \begin{bmatrix} x_{0,i} & \bar{H}_{M,i} \end{bmatrix} \;=\; \begin{bmatrix} H_{M,i} & x_{M,i} \end{bmatrix}$$

where $\{x_{0,i}, x_{M,i}\}$ denote the leading and trailing columns of $H_{M+1,i}$. It is assumed that the regression data satisfy the following structural relation:

$$\bar{H}_{M,i} = \begin{bmatrix} 0 \\ H_{M,i-1}\Phi_M^{-1} \end{bmatrix}$$

where Φ_M is an $M \times M$ invertible matrix of the form

$$\Phi_M = \begin{bmatrix} F_{M-1} & \\ & \mu \end{bmatrix}$$

where F_{M-1} is a unitary matrix and μ is a scalar.

(a) Derive an *a-posteriori-based* lattice filter for order-updating the least-squares solution.

(b) Derive an array-based lattice filter for order-updating the same least-squares solution.

(c) Can you think of a situation where such structural relations arise?

39. **(Chapters 29, 30)** Consider a least-squares problem of the form

$$\min_{w} \left[\delta\|w\|^2 \; + \; \sum_{j=0}^{N} \lambda^{N-j} |d(j) - u_j w|^2 \right]$$

where $\delta > 0$ is a regularization parameter, u_j is a $1 \times M$ regression vector, and $0 \ll \lambda \leq 1$ is a forgetting factor defined as follows:

$$\lambda = \begin{cases} \lambda_e & \text{for } j \text{ even} \\ \lambda_o & \text{for } j \text{ odd} \end{cases}$$

Let w_N denote the solution to the above least-squares problem. Can you derive a recursive least-squares solution that updates w_N to w_{N+1}?

40. **(Chapters 7 and 26)** Consider a standard state-space model of the form:

$$\begin{aligned} \boldsymbol{x}_{i+1} &= F_i \boldsymbol{x}_i + G \boldsymbol{n}_i \\ \boldsymbol{y}_i &= H \boldsymbol{x}_i + \boldsymbol{v}_i \end{aligned}$$

where

$$\mathsf{E} \begin{bmatrix} \boldsymbol{n}_i \\ \boldsymbol{v}_i \\ \boldsymbol{x}_o \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{n}_j \\ \boldsymbol{v}_j \\ \boldsymbol{x}_o \end{bmatrix}^* = \begin{bmatrix} Q\delta_{ij} & S\delta_{ij} & 0 \\ S^*\delta_{ij} & R\delta_{ij} & 0 \\ 0 & 0 & \Pi_o \\ 0 & 0 & 0 \end{bmatrix}$$

where all parameters are time-invariant, with the exception of F_i, which is defined below.

(a) Assume first that $F_i = F_1$ with probability p and $F_i = F_2$ with probability $1 - p$, where F_1 and F_2 are some constant matrices. Determine a recursive procedure for constructing the innovations process.

(b) Assume instead that $F_{2i} = F_1$ and $F_{2i+1} = F_2$. Determine a recursive procedure for constructing the innovations process.

(c) Assume the filters converge to steady-state in both cases so that the corresponding output processes $\{\boldsymbol{y}_i\}$ are stationary. Determine their auto-correlation sequences $\{R_y(k)\}$, and the $z-$transforms of these auto-correlation sequences (also called z-spectra).

(d) Under part (c), determine the corresponding spectral factors and the pre-whitening filters.

41. **(Chapter 33)** Consider two $n \times M$ matrices A and B with $n \leq M$. Lemma 33.1 in the textbook establishes that the equality $AA^* = BB^*$ holds if, and only if, there exists an $M \times M$ unitary matrix Θ such that $A = B\Theta$. Is Θ unique?

42. **(Chapters 15, 16, 23)** Consider the following constrained LMS recursion

$$w_i = w_{i-1} + \mu \left[I - \frac{cc^*}{\|c\|^2} \right] u_i^* (d(i) - u_i w_{i-1}), \qquad c^* w_{-1} = 1$$

which results from considering an instantaneous approximation for the following linear least-mean-squares estimation problem:

$$\min_{w} \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2 \quad \text{subject to} \quad \sum_{k=1}^{M} c(k)w(k) = 1$$

where the $\{w(k)\}$ denote the individual entries of w and the $\{c(k)\}$ are the scalar entries of the column vector c. Moreover, $\boldsymbol{d}$ denotes a scalar zero-mean random variable with variance σ_d^2, and $\boldsymbol{u}$ denotes a $1 \times M$ zero-mean random vector with covariance matrix $R_u = \mathsf{E}\,\boldsymbol{u}^*\boldsymbol{u} > 0$. Assume all data are circular Gaussian.

(a) Perform a transient mean-square-error analysis of the adaptive filter and provide conditions on the step-size μ in order to ensure that the filter is mean-square stable. Specify clearly the conditions on the data that you are assuming for your analysis.

(b) Derive expressions for the EMSE and the MSD of the filter.

(c) Derive an expression for the learning curve of the filter.

43. **(Chapters 29, 32)** Let $\widehat{w}$ denote the solution to the following regularized least-squares problem

$$\min_{w} \left[w^*\Pi w \;+\; (y - Hw)^*W(y - Hw)\right]$$

where $W > 0$ and $\Pi > 0$. Let $\widehat{y} = H\widehat{w}$ denote the resulting estimate of y and let ξ denote the corresponding minimum cost. Now consider the extended problem

$$\min_{w_z} \left\{ w_z^*\Pi_z w_z \;+\; \left(\begin{bmatrix} y \\ d \end{bmatrix} - \begin{bmatrix} h_a & H & h_b \\ \alpha_a & u & \alpha_b \end{bmatrix} w_z\right)^* W_z \left(\begin{bmatrix} y \\ d \end{bmatrix} - \begin{bmatrix} h_a & H & h_b \\ \alpha_a & u & \alpha_b \end{bmatrix} w_z\right)\right\}$$

where a and b are positive scalars, h_a and h_b are column vectors, α_a and α_b are scalars, d is a scalar, u is a row vector, and

$$\Pi_z = \begin{bmatrix} a & & \\ & \Pi & \\ & & b \end{bmatrix}, \qquad W_z = \begin{bmatrix} W & \\ & 1 \end{bmatrix}$$

Let

$$\widehat{y}_z = \begin{bmatrix} h_a & H & h_b \\ \alpha_a & u & \alpha_b \end{bmatrix} \widehat{w}_z$$

and let ξ_z denote the corresponding minimum cost of the extended problem.

(a) Relate $\{\widehat{w}_z, \widehat{y}_z, \xi_z\}$ to $\{\widehat{w}, \widehat{y}, \xi\}$.

(b) Can you motivate and derive an array algorithm to update the solution from $\widehat{w}$ to $\widehat{w}_z$?

44. **(Chapter 30)** Let w_i denote the solution to the following regularized least-squares problem:

$$\min_{w} \left[\lambda^{i+1}w^*\Pi w + \|y_i - H_i w\|^2\right]$$

where λ is a real scalar such that $0 < \lambda \leq 1$, Π is a positive-definite diagonal matrix, and where y_i and H_i are related to y_{i-1} and H_{i-1} as follows:

$$y_i = \begin{bmatrix} y_{i-1} \\ d(i) \end{bmatrix} \qquad H_i = \begin{bmatrix} H_{i-1} \\ u_i \end{bmatrix}$$

with $d(i)$ denoting a scalar and u_i denoting a row vector. We also define the matrix

$$P_i = \left[\lambda^{i+1}\Pi + H_i^* H_i\right]^{-1}$$

a) Assume that at time i, it holds that $H_i^* H_i > 0$. Let $w_{u,i}$ denote the solution to the following un-regularized least-squares problem:

$$\min_w \|y_i - H_i w\|^2$$

and let $P_{u,i} = [H_i^* H_i]^{-1}$. Provide a recursive algorithm to compute the un-regularized quantities $\{w_{u,i}, P_{u,i}\}$ from the regularized quantities $\{w_i, P_i\}$.

b) Derive a recursive algorithm to update the regularized solution, i.e., to compute $\{w_i, P_i\}$ from $\{w_{i-1}, P_{i-1}\}$.

In both parts (a) and (b), your algorithm should use a series of rank-1 updates, and direct matrix inversions are not allowed.

45. **(Chapters 2, 4)** Consider noisy observations $\boldsymbol{y}(i) = \boldsymbol{x} + \boldsymbol{v}(i)$, where $\boldsymbol{x}$ and $\boldsymbol{v}(i)$ are independent random variables, $\boldsymbol{v}(i)$ is a white random process with zero mean and distributed as follows:

   $\boldsymbol{v}(i)$ is Gaussian with variance σ_v^2 with probability q
   $\boldsymbol{v}(i)$ is uniformly distributed over $[-a, a]$ with probability $1 - q$

   Moreover, $\boldsymbol{x}$ assumes the values $\{1+j, 1-j, -1+j, -1-j\}$ with equal probability. The value of $\boldsymbol{x}$ is the same for all measurements $\{\boldsymbol{y}(i)\}$.

   (a) Find an expression for the optimal least-mean-squares estimate of $\boldsymbol{x}$ given a collection of N observations $\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(N-1)\}$.

   (b) Find the linear least-mean-squares estimate of $\boldsymbol{x}$ given the combined observations $\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(N-1)\}$ and $\{\boldsymbol{y}^2(0), \boldsymbol{y}^2(1), \ldots, \boldsymbol{y}^2(N-1)\}$.

46. **(Chapters 3, 4)** Let $\boldsymbol{x}$ be a zero-mean random variable with an $M \times M$ positive-definite covariance matrix R_x. Let $\hat{\boldsymbol{x}}_{\boldsymbol{y}_1}$ denote the linear least-mean-squares estimator of $\boldsymbol{x}$ given a zero-mean observation $\boldsymbol{y}_1$ with covariance matrix R_{y_1}. Likewise, let $\hat{\boldsymbol{x}}_{\boldsymbol{y}_2}$ denote the linear least-mean-squares estimator of $\boldsymbol{x}$ given another zero-mean observation $\boldsymbol{y}_2$ with covariance matrix R_{y_2}. Let $R_{y_1,y_2} = \mathsf{E}\,\boldsymbol{y}_1\boldsymbol{y}_2^*$. We want to determine another estimator for $\boldsymbol{x}$ by combining $\hat{\boldsymbol{x}}_{y_1}$ and $\hat{\boldsymbol{x}}_{y_2}$ in a convex manner as follows:

   $$\hat{\boldsymbol{x}} = \lambda\hat{\boldsymbol{x}}_{y_1} + (1-\lambda)\hat{\boldsymbol{x}}_{y_2}$$

   where λ is a real scalar lying inside the interval $0 \leq \lambda \leq 1$.

   (a) Determine the value of λ that results in an estimator $\hat{\boldsymbol{x}}$ with the smallest mean-square error.

(b) If λ is allowed to be any arbitrary real scalar (not necessarily limited to the range $[0, 1]$), how much smaller can the mean-square-error be?

47. **(Chapters 8, 10)** Let $\boldsymbol{d}$ denote a scalar zero-mean random variable with variance σ_d^2, and let $\boldsymbol{u}$ denote a $1 \times M$ zero-mean random vector with covariance matrix $R_u = \mathsf{E}\,\boldsymbol{u}^*\boldsymbol{u} > 0$. Consider the optimization problem

$$\min_w \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2 \quad \text{subject to} \quad \sum_{k=1}^{M} c(k)w(k) = 1$$

where the $\{w(k)\}$ denote the individual entries of w and the $\{c(k)\}$ are scaling coefficients.

(a) Derive a stochastic-gradient algorithm for approximating the optimal solution w^o in terms of realizations $\{d(i), u_i\}$ for $\{\boldsymbol{d}, \boldsymbol{u}\}$, and starting from an initial condition w_{-1} that satisfies the constraint.

(b) Derive an approximate expression for the EMSE of the filter for sufficiently small-step-sizes.

(c) Derive an optimal choice for the coefficients $\{c(k)\}$ in order to result in the smallest EMSE.

(d) Can you repeat parts (a)-(c) when the $\{c(k)\}$ are required to be nonnegative scalars?