Problems and Computer Projects

PROBLEMS

Problem III.1 (Multiple step-sizes) Refer to expression (8.18) and choose the matrix B as $B = \text{diag}\{b(1), b(2), \ldots, b(M)\}$ with $b(i) > 0$. In this case, the recursion (8.20) is replaced by $w_i = w_{i-1} + \mu B[R_{du} - R_u w_{i-1}]$. This scheme associates one step-size with each entry of the weight vector w_i. Follow the discussion in Sec. 8.2 and derive a necessary and sufficient condition on μ in order to guarantee convergence of w_i to $w^o = R_u^{-1} R_{du}$.

Problem III.2 (Product of infinitely many numbers) Consider a scalar recursion of the form $x(i) = a(i)x(i-1)$ for $i \geq 0$, and assume $a(i) = e^{-1/(i+1)^2}$.

(a) Verify that $0 < a(i) < 1$ for all finite i.

(b) Let $p(i) = \prod_{j=0}^{i} a(j)$. Show that $p(i)$ converges to $e^{-\pi^2/6}$, which is a finite positive number. *Hint:* The series $\sum_{j=1}^{\infty}(1/j^2)$ converges to $\pi^2/6$.

Problem III.3 (Optimal step-size) Refer to expression (9.16) for the optimal step-size. Verify that it is equivalent to the following:

$$\mu^o(i) = \frac{\|\nabla_w J(w_{i-1})\|^2}{[\nabla_w J(w_{i-1})]\; R_u \;[\nabla_w J(w_{i-1})]^*}$$

in terms of the squared Euclidean norm of the gradient vector in the numerator, and the weighted squared Euclidean norm of the same vector in the denominator.

Problem III.4 (Convergent step-size sequence) Consider the steepest-descent algorithm (9.13) with a time-variant step-size. Assume that $\mu(i)$ converges to a positive value, say, $\mu(i) \longrightarrow \alpha > 0$ as $i \longrightarrow \infty$. Show that if α satisfies $\alpha < 2/\lambda_{\max}$, then w_i converges to w^o.

Problem III.5 (Optimal step-size) Consider the optimal step-size (9.16) in the iteration-dependent case of the steepest-descent algorithm. Show that $1/\lambda_{\max} \leq \mu^o(i) \leq 1/\lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and smallest eigenvalues of R_u. Conclude that $\sum_{i=0}^{\infty} \mu^o(i) = \infty$.

Problem III.6 (Contour curves) Consider the steepest-descent algorithm (9.13) with the optimal time-variant step-size (9.16).

(a) Show that $J(w_i) = J(w_{i-1}) - \mu^o(i)\tilde{w}_{i-1}^* R_u^2 \tilde{w}_{i-1}$.

(b) Assume $M = 2$ (i.e., the size of w^o is 2×1). Sketch the elliptic contours of constant mean-square error and explain how the optimized algorithm moves from one elliptic curve to another.

Problem III.7 (Interfering signals) Refer to Prob. I.14. Suggest a steepest-descent algorithm for estimating $\boldsymbol{x}_1$ from $\boldsymbol{y}$.

Problem III.8 (Prediction problem) A zero-mean stationary random process $\{\boldsymbol{u}(\cdot)\}$ is generated by passing a zero-mean white sequence $\{\boldsymbol{v}(\cdot)\}$ with variance σ_v^2 through a second-order auto-regressive model, namely, $\boldsymbol{u}(i) + \alpha \boldsymbol{u}(i-1) + \beta \boldsymbol{u}(i-2) = \boldsymbol{v}(i),\ i > -\infty$, where α and β are real numbers such that the roots of the characteristic equation $1 + \alpha z^{-1} + \beta z^{-2} = 0$ are strictly inside the unit circle. We wish to design a second-order predictor for the process $u(\cdot)$ of the form $\hat{u}(i) = \begin{bmatrix} u(i-1) & u(i-2) \end{bmatrix} w^o$, for some 2×1 vector w^o.

(a) Verify that $\{\alpha, \beta\}$ must satisfy $|\beta| < 1$ and $|\alpha| < 1 + \beta$.

(b) Define the data vector $u = \begin{bmatrix} u(i-1) & u(i-2) \end{bmatrix}$ and the desired signal $d = u(i)$. Let $R_u = \mathsf{E}\, u^* u$ and $R_{du} = \mathsf{E}\, du^*$. Show that

$$\left[\begin{array}{c|c} R_u & R_{du} \end{array}\right] = \frac{\sigma_v^2}{(1-\beta)[(1+\beta)^2 - \alpha^2]} \left[\begin{array}{cc|c} 1+\beta & -\alpha & -\alpha \\ -\alpha & 1+\beta & \alpha^2 - \beta^2 - \beta \end{array}\right]$$

Establish that $(1-\beta)[(1+\beta)^2 - \alpha^2] > 0$.

(c) Show that the optimal weight vector w^o is given by $w^o = \text{col}\{-\alpha, -\beta\}$. Could you have guessed this answer more directly without evaluating the product $R_u^{-1} R_{du}$?

(d) Verify that the eigenvalue spread of R_u is $\rho = (\beta + 1 + |\alpha|)/(\beta + 1 - |\alpha|)$. Design a steepest-descent algorithm that determines w^o iteratively. Provide a condition on the step-size μ in terms of $\{\alpha, \beta\}$ in order to guarantee convergence.

(e) Show that the value of the step-size that yields fastest convergence, and the resulting time-constant, are

$$\mu^o = \frac{1-\beta}{1+\beta} \frac{(1+\beta)^2 - \alpha^2}{\sigma_v^2}, \qquad \bar{\tau}^o = \frac{-1}{2\ln\left(|\alpha|/(\beta+1)\right)}$$

Problem III.9 (Logarithm function) Establish that the logarithm function satisfies the following properties:

(a) For any $y \geq 0$, it holds that $\ln(1-y) \leq -y$.

(b) There exist $0 < b < 1$ and $a > 0$ such that $-ay \leq \ln(1-y)$ for any $y \in [0, b]$.

Problem III.10 (Convergence proof) The purpose of this problem is to establish the *statement* of Thm. 9.1. Let $\tilde{w}_i = w^o - w_i$ and introduce the eigen-decomposition $R_u = U \Lambda U^*$. Define the transformed vector $x_i = U^* \tilde{w}_i$ and let $x_k(i)$ denote its k−th entry.

(a) Assume first that the step-size sequence is divergent and let us establish that w_i converges to w^o. Thus since $\lambda_k > 0$, and using the fact that $\mu(i) \to 0$, conclude that there exists a large enough i_o such that $0 < 1 - \mu(i)\lambda_k \leq 1$ for all k and for all $i > i_o$. Use the result of part (a) of Prob. III.9 to conclude that $\sum_{j=i_o+1}^{\infty} \ln(1 - \mu(j)\lambda_k) = -\infty$.

(b) Verify that $x_k(i)$ satisfies the recursion $x_k(i) = (1 - \mu(i)\lambda_k) x_k(i-1)$ and conclude that

$$\lim_{i \to \infty} \ln \left[\frac{|x_k(i)|}{|x_k(i_o)|} \right] = -\infty$$

and, consequently, $x_k(i) \to 0$. *Remark.* The argument in parts (a) and (b) shows that if the step-size sequence is divergent then $\tilde{w}_i \to 0$. We now examine the converse statement.

(c) Let us now assume that w_i converges to w^o and let us establish that the step-size sequence is divergent. Thus assume $x_k(i) \to 0$ and show that it implies $\sum_{j=i_o+1}^{\infty} \ln(1 - \mu(j)\lambda_k) = -\infty$.

(d) Assume that i_o is large enough so that not only $0 < 1 - \mu(j)\lambda_k < 1$ for all k and $j > i_o$, but also $\mu(j)\lambda_k < b$. Use the result of part (b) of Prob. III.9 to conclude that

$$-a\lambda_k \sum_{j=i_o+1}^{\infty} \mu(j) \leq \sum_{j=i_o+1}^{\infty} \ln(1 - \mu(j)\lambda_k) = -\infty$$

That is, conclude that the sequence $\{\mu(i)\}$ is divergent.

Remark. The above arguments are patterned along those given in Macchi (1995, pp. 65–67).

Problem III.11 (Regularized Newton's method) Consider the steepest-descent recursion of Remark 2,

$$w_i = w_{i-1} - \mu[\epsilon \mathrm{I} + \nabla_w^2 J(w_{i-1})]^{-1} \left[\nabla J(w_{i-1})\right]^*, \quad w_{-1} = \text{initial guess}$$

for some $\epsilon > 0$. For the quadratic cost function $J(w)$ of (8.8) we have $\nabla_w^2 J(w_{i-1}) = R_u > 0$ and $\nabla_w J(w_{i-1}) = (R_{du} - R_u w_{i-1})^*$.

(a) Show that a necessary and sufficient condition on μ that guarantees convergence of w_i to the minimizing argument of $J(w)$ is $0 < \mu < 2 + \epsilon/\lambda_{\max}$.

(b) Find the optimum step-size μ^o at which the convergence rate is maximized.

Problem III.12 (Leaky variant of steepest-descent) Consider the modified optimization problem

$$\min_w \left[J^\alpha(w) \triangleq \alpha\|w\|^2 + \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2 \right]$$

where α is a positive real number and $J^\alpha(w)$ is the new cost function (it is dependent on α). In the text we studied the case $\alpha = 0$ (see expression (8.7) for $J(w)$). The above modified cost function penalizes the energy (or squared norm) of the vector w, and is therefore useful in situations where we want to avoid solutions with potentially large norms.

(a) Show that the optimal solution is given by $w^\alpha = [R_u + \alpha \mathrm{I}]^{-1} R_{du}$. Compute the resulting minimum cost, $J^\alpha(w^\alpha)$, and show that $J^\alpha(w^\alpha) > \sigma_d^2 - R_{ud}R_u^{-1}R_{du} = J_{\min}$, where $J_{\min}$ is the minimum cost associated with $J(w)$ (cf. (8.11)).

(b) Let $\delta w = w^\alpha - w^o$ denote the difference between the new solution w^α and the linear least-mean squares solution w^o of (8.4). Show that $R_{ud}\delta w = J_{\min} - J^\alpha(w^\alpha)$.

(c) Justify the validity of the following steepest-descent method for determining w^α:

$$\boxed{w_i^\alpha = (1 - \mu\alpha)w_{i-1}^\alpha + \mu[R_{du} - R_u w_{i-1}^\alpha], \quad w_{-1}^\alpha = \text{initial guess}}$$

Show that w_i^α converges to w^α if, and only if, $0 < \mu < 2/(\lambda_{\max} + \alpha)$. Show also that the optimal step-size for fastest convergence is $\mu^\alpha = 2/(\lambda_{\max} + \lambda_{\min} + 2\alpha)$.

(d) Let μ^o denote the optimal step-size choice for the standard steepest-descent method (with $\alpha = 0$, i.e., cf. (9.5)). Compare μ^α and μ^o.

Remark. We see from the result in part (a) that the inclusion of the term $\alpha\|w\|^2$ in the cost function $J^\alpha(w)$ has the effect of modifying the input covariance matrix from R_u to $R_u + \alpha\mathrm{I}$. This can be interpreted as adding a noise vector $\boldsymbol{v}$ to $\boldsymbol{u}$, with the individual entries of $\boldsymbol{v}$ arising from a zero-mean white-noise process with variance α. This process of disturbing the input $\boldsymbol{u}$ with entries from a white-noise sequence is known as *dithering*. Its effect is to provide a mechanism for controlling the size of the solution vector w^α; it also results in a covariance matrix with smaller eigenvalue spread (see Prob. III.13). Its disadvantage is that the optimal solution w^α will be distinct from the desired solution w^o.

Problem III.13 (One effect of dithering) Consider the same setting as in Prob. III.12. Show that the eigenvalue spread of $R_u + \alpha\mathrm{I}$ is smaller than the eigenvalue spread of R_u.

Problem III.14 (l_1-norm of complex variables) Let x be a nonzero real-valued variable. Verify that, for $x \neq 0$,

$$\frac{\mathsf{d}|x|}{\mathsf{d}x} = \mathsf{sign}(x) \quad \text{where} \quad \mathsf{sign}(x) \triangleq \begin{cases} +1 & x > 0 \\ -1 & x < 0 \end{cases}$$

[At $x = 0$, we shall define from now on $\mathsf{sign}(0) = 0$.] Now assume x is complex-valued with real and imaginary parts denoted by x_r and x_i, respectively. Define its l_1-norm as follows $|x| = |x_\mathsf{r}| + |x_\mathsf{i}|$. That is, we add the absolute values of its real and imaginary parts. [The result can be interpreted as the l_1-norm of the vector $\text{col}\{x_\mathsf{r}, x_\mathsf{i}\}$.] Show that the complex gradient (cf. Chapter C) of $|x|$ with respect to $x \neq 0$ is given by

$$\frac{\mathsf{d}|x|}{\mathsf{d}x} = \frac{1}{2}\left[\mathsf{sign}(x_\mathsf{r}) - j\,\mathsf{sign}(x_\mathsf{i})\right]$$

Problem III.15 (Sign-error algorithm) Consider two zero-mean random variables $\boldsymbol{d}$ and $\boldsymbol{u}$ where $\boldsymbol{d}$ is scalar-valued and $\boldsymbol{u}$ is a row vector. We are interested in minimizing the expected value of the l_1-norm of the error $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{u}w$ (cf. Prob. III.14), i.e., $\min_w \mathsf{E}\,|\boldsymbol{e}|$.

(a) Let $\boldsymbol{e} = \boldsymbol{e}_r + j\boldsymbol{e}_i$. Show that the complex-gradient of the cost function $J(w) = \mathsf{E}\,|\boldsymbol{e}|$ with respect to w is given by $\nabla_w J(w) = -\mathsf{E}\left(\boldsymbol{u}\;[\mathsf{sign}(\boldsymbol{e}_r) - j\mathsf{sign}(\boldsymbol{e}_i)]\right)/2$.

(b) Conclude that a steepest-descent method can be obtained via the recursion

$$w_i = w_{i-1} + \frac{\mu}{2}\mathsf{E}\,\boldsymbol{u}^*\left\{\mathsf{sign}[\boldsymbol{e}_r(i)] + j\mathsf{sign}[\boldsymbol{e}_i(i)]\right\}$$

for some positive step-size μ and where $\boldsymbol{e}(i) = \boldsymbol{d} - \boldsymbol{u}w_{i-1} = \boldsymbol{e}_r(i) + j\boldsymbol{e}_i(i)$.

Remark. A more compact representation can be obtained if we define the sign of a complex number as follows:

$$\mathsf{csgn}(x) \triangleq \mathsf{sign}(x_r) + j\mathsf{sign}(x_i)$$

Observe that we are writing csgn instead of sign; we reserve the notation sign for the sign of a real number. With this definition, we can rewrite the above steepest-descent recursion as

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^*\mathsf{csgn}[\boldsymbol{e}(i)]}$$

in terms of $\boldsymbol{e}(i)$ and for a scaled step-size, which we still denote by μ.

Problem III.16 (Least-mean-fourth (LMF) criterion) Consider zero-mean random variables $\boldsymbol{d}$ and $\boldsymbol{u}$, with $\boldsymbol{u}$ a row vector. An optimal weight vector w^o is to be chosen by minimizing the cost function

$$\min_w \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^{2L}$$

for some positive integer L. Observe that we are now minimizing the moment of order $2L$ of the error signal rather than its variance, as was done in Sec. 8.1.

(a) Argue that the cost function $\mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^{2L}$ is convex in w and that, therefore, it cannot have local minima.

(b) Assume in this part that $\{\boldsymbol{d}, \boldsymbol{u}, w\}$ are real and scalar-valued. Assume further that $\boldsymbol{d}$ and $\boldsymbol{u}$ are jointly Gaussian and that $L = 2$. Verify that the cost function in this case reduces to

$$J(w) = 3\sigma_d^4 - 12\sigma_d^2\sigma_{du}w + 6(\sigma_u^2\sigma_d^2 + 2\sigma_{du}^2)w^2 - 12\sigma_u^2\sigma_{du}w^3 + 3\sigma_u^4 w^4$$

where $\sigma_u^2 = \mathsf{E}\,\boldsymbol{u}^2$, $\sigma_d = \mathsf{E}\,\boldsymbol{d}^2$ and $\sigma_{du} = \mathsf{E}\,\boldsymbol{du}$. Show that $w^o = \sigma_{du}/\sigma_u^2$ is a minimum of $J(w)$. Is it a global minimum? Assume $\sigma_d^2 = 1$, $\sigma_u^2 = 1$ and $\sigma_{du} = 0.7$. Plot $J(w)$.

(c) Let z be a complex-valued variable and consider the function $f(z) = |z|^{2L}$. Verify that $\mathsf{d}f/\mathsf{d}z = Lz^*|z|^{2(L-1)}$.

(d) Let $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{u}w$ and denote the cost function by $J(w) = \mathsf{E}\,|\boldsymbol{e}|^{2L}$. Use the composition rule of differentiation to verify that $\nabla_w J(w) = -L\;\mathsf{E}\,\boldsymbol{ue}^*|\boldsymbol{e}|^{2(L-1)}$. Conclude that a steepest-descent implementation for finding a minimizing solution of $J(w)$ is given by (note that we blended the constant L into μ):

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^*\boldsymbol{e}(i)\;|\boldsymbol{e}(i)|^{2(L-1)}}$$

for some step-size $\mu > 0$ and where $\boldsymbol{e}(i) = \boldsymbol{d} - \boldsymbol{u}w_{i-1}$.

Remark. When $L = 2$, the criterion corresponds to solving a least-mean-fourth (LMF) estimation problem. The recursion in this case would take the form

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^*\boldsymbol{e}(i)\;|\boldsymbol{e}(i)|^2}$$

Note also that when all variables are real-valued, the recursion of part (d) reduces to

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^\mathsf{T}\boldsymbol{e}^{2L-1}(i)}$$

and in the LMF case it becomes

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^\mathsf{T}\boldsymbol{e}^3(i)}$$

Problem III.17 (Least-mean-mixed-norm (LMMN) criterion) Consider zero-mean random variables $\boldsymbol{d}$ and $\boldsymbol{u}$, where $\boldsymbol{u}$ is a row vector. An optimal weight vector w^o is to be chosen so as to minimize the cost function

$$\min_{w} \mathsf{E}\left[\delta|\boldsymbol{e}|^2 + \frac{1}{2}(1-\delta)|\boldsymbol{e}|^4\right]$$

where $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{u}w$ and $0 \leq \delta \leq 1$. Observe that this cost function reduces to the least-mean-squares and least-mean-fourth criteria at the extreme points $\delta = 1$ and $\delta = 0$, respectively. Other values of δ allow for a tradeoff between both criteria. Verify that

$$\nabla_w J(w) = -\mathsf{E}\boldsymbol{u}\left[\delta\boldsymbol{e}^* + (1-\delta)\boldsymbol{e}^*|\boldsymbol{e}|^2\right]$$

Conclude that a steepest-descent implementation for finding a minimizing solution of $J(w)$ is given by

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^*\boldsymbol{e}(i)\left[\delta + (1-\delta)|\boldsymbol{e}(i)|^2\right]}$$

for some step-size $\mu > 0$ and where $\boldsymbol{e}(i) = \boldsymbol{d} - \boldsymbol{u}w_{i-1}$. *Remark.* For real-data, the above recursion reduces to

$$\boxed{w_i = w_{i-1} + \mu\mathsf{E}\,\boldsymbol{u}^{\mathsf{T}}\left[\delta\boldsymbol{e}(i) + (1-\delta)\boldsymbol{e}^3(i)\right]}$$

Problem III.18 (Constant-modulus criterion) Let $\boldsymbol{u}$ be a zero-mean row vector and w an unknown column vector that we wish to determine so as to minimize the cost function $J(w) = \mathsf{E}\left(\gamma - |\boldsymbol{u}w|^2\right)^2$, for a given constant number γ.

(a) Let $R_u = \mathsf{E}\,\boldsymbol{u}^*\boldsymbol{u}$. Show that

$$J(w) = \gamma^2 + w^*[-2\gamma R_u + \mathsf{E}\,(\boldsymbol{u}^*\boldsymbol{u}\,|\boldsymbol{u}w|^2)]w$$

and

$$\nabla_w J(w) = -2w^*\left(\gamma R_u - \mathsf{E}\left(\boldsymbol{u}^*\boldsymbol{u}|\boldsymbol{u}w|^2\right)\right)$$

Conclude that a steepest-descent algorithm for the minimization of $J(w)$ is given by

$$\boxed{w_i = w_{i-1} + \mu\left(\gamma R_u w_{i-1} - \mathsf{E}\left[\boldsymbol{u}^*\boldsymbol{u}w_{i-1}|\boldsymbol{u}w_{i-1}|^2\right]\right)}$$

for some step-size $\mu > 0$.

(b) Assume that w is two-dimensional, say with entries $w = \text{col}\{\alpha, \beta\}$, and that $\boldsymbol{u}$ is a circular Gaussian random vector (cf. Sec. A.5) with $R_u = \text{diag}\{1, 2\}$.

(b.1) Verify that $\mathsf{E}\,|\boldsymbol{u}w|^2 = |\alpha|^2 + 2|\beta|^2$ and $\mathsf{E}\,|\boldsymbol{u}w|^4 = 2\left(|\alpha|^2 + 2|\beta|^2\right)^2$. Conclude that $J(w) = \gamma^2 + 2\left(|\alpha|^2 + 2|\beta|^2\right)^2 - 2\gamma\left(|\alpha|^2 + 2|\beta|^2\right)$. *Remark.* If $\boldsymbol{z}$ is a zero-mean scalar complex-valued Gaussian random variable with variance $\mathsf{E}\,|\boldsymbol{z}|^2 = \sigma_z^2$, then its fourth-moment is $\mathsf{E}\,|\boldsymbol{z}|^4 = 2\sigma_z^4$. If $\boldsymbol{z}$ were real-valued instead, its fourth moment would be $\mathsf{E}\,|\boldsymbol{z}|^4 = 3\sigma_z^4$. Verify these claims.

(b.2) Conclude also that $\nabla_w J(w) = 2(-\gamma + 2|\alpha|^2 + 4|\beta|^2)\begin{bmatrix}\alpha^* & 2\beta^*\end{bmatrix}$. Argue that $J(w)$ has a local maximum at the point $\alpha = \beta = 0$ and global minima at all points lying on the ellipse $|\alpha|^2 + 2|\beta|^2 = \gamma/2$. Argue further that the minimum value of $J(w)$ is equal to $\gamma^2/2$.

(b.3) Conclude that the steepest-descent algorithm of part (a) reduces in this case to the form

$$\begin{bmatrix}\alpha(i)\\ \beta(i)\end{bmatrix} = \begin{bmatrix}\alpha(i-1)\\ \beta(i-1)\end{bmatrix} + \mu\left(\gamma - 2|\alpha(i-1)|^2 - 4|\beta(i-1)|^2\right)\begin{bmatrix}\alpha(i-1)\\ 2\beta(i-1)\end{bmatrix}$$

and that the weight estimates are always real-valued if the initial condition is real-valued. This problem is pursued further ahead in a computer project.

Remark. By minimizing the cost function $J(w)$ we are forcing the magnitude of uw to be close to the constant $\sqrt{\gamma}$; hence, the name constant-modulus criterion. Such cost functions are used in blind channel equalization applications — see Computer Project III.4.

Problem III.19 (Reduced-constellation criterion) Let u be a zero-mean row vector and w an unknown column vector that we wish to determine so as to minimize the cost function

$$J(w) = \mathsf{E}\left[\frac{1}{2}|uw|^2 - \gamma|uw|_1\right]$$

for a given constant number γ and in terms of the l_1-norm of uw, as defined in Prob. III.14. Let $R_u = \mathsf{E}\,u^*u$.

(a) Show that a steepest-descent algorithm for the minimization of $J(w)$ is given by

$$\boxed{w_i = w_{i-1} - \mu\left[\,R_u w_{i-1} - \gamma\mathsf{E}\ \left(\mathsf{csgn}(uw_{i-1})u^*\right)\,\right]}$$

for some step-size $\mu > 0$.

(b) Show that minimizing $J'(w) = \mathsf{E}\,|uw - \gamma \cdot \mathsf{csgn}(uw)|^2$ is equivalent to minimizing the cost function above. Conclude that $J(w)$ is attempting to minimize the distance between uw and the four points $\{\pm\gamma \pm j\gamma\}$.

Remark. Compare the above conclusion with the constant-modulus criterion of Prob. III.18, which attempts to minimize the distance between uw and all points on the circle of radius $\sqrt{\gamma}$. Both criteria are useful for blind channel equalization.

Problem III.20 (Multi-modulus criterion) Let u be a zero-mean row vector and w an unknown column vector that we wish to determine so as to minimize the cost function

$$J(w) = \mathsf{E}\left\{\ [(\mathsf{Re}(uw))^2 - \gamma]^2 \ + \ [(\mathsf{Im}(uw))^2 - \gamma]^2\ \right\}$$

for a given constant number γ, and where $\mathsf{Re}(\cdot)$ and $\mathsf{Im}(\cdot)$ denote the real and imaginary parts of their arguments. Let $R_u = \mathsf{E}\,u^*u$. Show that a steepest-descent algorithm for the minimization of $J(w)$ is given by

$$a = \mathsf{Re}(uw_{i-1}), \quad b = \mathsf{Im}(uw_{i-1}), \quad e(i) = a[\gamma - a^2] + jb[\gamma - b^2], \quad w_i = w_{i-1} + \mu\mathsf{E}\,u^*e(i)$$

for some step-size $\mu > 0$. *Remark.* By minimizing the cost function $J(w)$ we are forcing the real and imaginary parts of uw to be close to the values $\pm\sqrt{\gamma}$. In other words, $J(w)$ attempts to minimize the distance between uw and the horizontal and vertical lines located at $\pm\sqrt{\gamma}$.

Problem III.21 (Another constant-modulus criterion) Let u be a zero-mean row vector and w an unknown column vector that we wish to determine so as to minimize the cost function

$$J(w) = \mathsf{E}\left[\frac{1}{2}|uw|^2 - \gamma|uw|\right]$$

for a given constant number γ and in terms of the magnitude of uw. Let $R_u = \mathsf{E}\,u^*u$.

(a) Show that a steepest-descent algorithm for the minimization of $J(w)$ is given by

$$\boxed{w_i = w_{i-1} - \mu\left[\,R_u w_{i-1} - \gamma\mathsf{E}\ \left(\frac{uw}{|uw|}u^*\right)\right]}$$

for some step-size $\mu > 0$.

(b) Show that minimizing $J'(w) = \mathsf{E}\,(\gamma - |uw|)^2$ is equivalent to minimizing the cost function above. Conclude that $J(w)$ is attempting to minimize the distance between $|uw|$ and the circle of radius γ.

Problem III.22 (Constrained mean-square error) Consider the constrained optimization problem described in Prob. II.36, namely,

$$\boxed{\min_{w} \mathsf{E}\,|d - uw|^2 \quad \text{subject to } c^*w = \alpha}$$

where c is a known $M \times 1$ vector and α is a known real scalar. Verify that a steepest-descent recursion for estimating the optimal solution w^o is given by

$$\boxed{w_i = w_{i-1} + \mu\left[\mathrm{I} - \frac{cc^*}{\|c\|^2}\right](R_{du} - R_u w_{i-1})}$$

where $R_u = \mathsf{E}\,u^*u > 0$ and $R_{du} = \mathsf{E}\,du^*$. *Hint:* Use the extended cost function $J_e(w, \lambda)$ of Prob. II.36 and enforce the condition that the successive weight vectors $\{w_i\}$, including the initial condition, must satisfy the constraint $c^*w_i = \alpha$.

Problem III.23 (Homogeneous difference equation) Consider the homogeneous difference equation $x_i = Ax_{i-1}$ with arbitrary initial vector x_{-1}. Show that x_i tends to the zero vector if, and only if, all the eigenvalues of A have strictly less than unit magnitude. *Hint:* One possibility is to use the Jordan decomposition of the matrix A — see Prob. II.2.

Problem III.24 (Power estimate) Refer to (11.9) and assume that the entries $\{u(j)\}$ arise from observations of a white random sequence $\{\boldsymbol{u}(j)\}$ with variance σ_u^2. In this way, the quantity $p(i)$ can be interpreted as a realization of a random variable $\boldsymbol{p}(i)$ satisfying $\boldsymbol{p}(i) = \beta\boldsymbol{p}(i-1) + (1-\beta)|\boldsymbol{u}(i)|^2$, $\boldsymbol{p}(-1) = 0$. Show that $\mathsf{E}\,\boldsymbol{p}(i) \longrightarrow \sigma_u^2$ as $i \longrightarrow \infty$.

Problem III.25 (Perturbation property of ϵ-NLMS) Follow the derivation in Sec. 10.4 and show that the solution to the optimization problem

$$\min_{w_i} \|w_i - w_{i-1}\|^2 \quad \text{subject to } r(i) = \left(1 - \frac{\mu\|u_i\|^2}{\epsilon + \|u_i\|^2}\right)e(i)$$

is given by the ϵ−NLMS recursion (11.4).

Problem III.26 (Leaky-LMS) The leaky-LMS algorithm is the stochastic gradient version of the steepest-descent method of Prob. III.12. Replace R_{du} by $d(i)u_i^*$, R_u by $u_i^*u_i$, and verify that this leads to the following recursion:

$$\boxed{w_i^\alpha = (1 - \mu\alpha)w_{i-1}^\alpha + \mu u_i^*[d(i) - u_i w_{i-1}^\alpha], \quad i \geq 0}$$

Define the *a posteriori* and *a priori* output errors $r^\alpha(i) = d(i) - u_i w_i^\alpha$ and $e^\alpha(i) = d(i) - u_i w_{i-1}^\alpha$. Verify that $r^\alpha(i) = \left[1 - \alpha\mu - \mu\|u_i\|^2\right]e^\alpha(i) + \alpha\mu d(i)$.

Problem III.27 (Drift problem) Assume the regressors u_i are scalars, say $u(i)$, and given by $u(i) = 1/\sqrt{i+1}$. Let $\mu = 1$ and $d(i) = u(i)w^o + v(i)$ with $w^o = 0$ and $v(i) = c$ for all i.

(a) Verify that the LMS update for the weight estimate $w(i)$ gives

$$w(i) = \frac{c}{i+1}\sum_{k=1}^{i+1}\sqrt{k} \;\geq\; \frac{2c}{3}\sqrt{i+1}$$

and, therefore, $w(i) \to \infty$ as $i \to \infty$ (no matter how small c is).

(b) Consider instead the leaky-LMS update of Prob. III.26 for a generic step-size μ,

$$w^\alpha(i) = \left(1 - \mu\alpha - \frac{\mu}{i+1}\right)w^\alpha(i-1) + \frac{\mu c}{\sqrt{i+1}}$$

Show that this recursion results in a bounded sequence $\{w^\alpha(i)\}$ if $0 < \mu < 2/(\alpha+1)$.

Remark. This example shows that the weight estimates computed by LMS can grow slowly to large values, even when the noise is small — see Prob. IV.39 for a more detailed example.

Problem III.28 (Constrained LMS) Refer to the steepest-descent algorithm of Prob. III.22, which pertains to the constrained optimization problem

$$\boxed{\min_{w} \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2 \quad \text{subject to } c^* w = \alpha}$$

where c is a known $M \times 1$ vector and α is a known real scalar. Verify that a stochastic-gradient algorithm for approximating the optimal solution w^o is given by

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

in terms of realizations $\{d(i), u_i\}$ for $\{\boldsymbol{d}, \boldsymbol{u}\}$, and starting from an initial condition that satisfies $c^* w_{-1} = \alpha$.

Problem III.29 (Sign-error LMS) Refer to the statement of Prob. III.15. Show that the corresponding stochastic-gradient method is given by the following so-called sign-error LMS algorithm,

$$\boxed{w_i = w_{i-1} + \mu u_i^* \mathsf{csgn}[d(i) - u_i w_{i-1}], \quad i \geq 0}$$

where the complex-sign function is as defined in (12.1).

Problem III.30 (LMF algorithm) Refer to Prob. III.16 and assume $L = 2$. Argue that a stochastic gradient implementation is given by

$$\boxed{w_i = w_{i-1} + \mu u_i^* e(i) |e(i)|^2, \qquad e(i) = d(i) - u_i w_{i-1}, \quad i \geq 0}$$

Problem III.31 (LMMN algorithm) Refer to Prob. III.17. Argue that a stochastic-gradient implementation is given by

$$\boxed{w_i = w_{i-1} + \mu u_i^* e(i) [\delta + (1-\delta)|e(i)|^2], \qquad e(i) = d(i) - u_i w_{i-1}, \quad i \geq 0}$$

Problem III.32 (Least mean-phase (LMP) algorithm) Let $\boldsymbol{d}$ and $\boldsymbol{u}$ be zero-mean random variables, with $\boldsymbol{d}$ being a scalar and $\boldsymbol{u}$ a $1 \times M$ vector. Introduce the phase-error cost function $J_{\text{pe}}(w) = \mathsf{E}\,|\text{phase}(\boldsymbol{d}) - \text{phase}(\boldsymbol{u}w)|^m = \mathsf{E}\,|\angle \boldsymbol{d} - \angle \boldsymbol{u}w|^m$, where $m = 1, 2$ and w is an unknown weight vector to be estimated. Consider further the squared-error cost function $J_{\text{se}}(w) = \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2$ and let $J(w) = k_1 J_{\text{se}}(w) + k_2 J_{\text{pe}}(w)$, where k_1 and k_2 define the contribution of each term to the overall cost function.

(a) Verify that for $m = 2$, a stochastic gradient implementation is given by

$$w_i = w_{i-1} + \mu_1 u_i^* (d(i) - u_i w_{i-1}) + \mu_2 (\angle d(i) - \angle u_i w_{i-1}) \frac{j u_i^*}{(u_i w_{i-1})^*}$$

where μ_1 and μ_2 are step-size parameters.

(b) Likewise, verify that $m = 1$, we obtain instead

$$w_i = w_{i-1} + \mu_1 u_i^* (d(i) - u_i w_{i-1}) + \mu_2 \mathsf{sign}(\angle d(i) - \angle u_i w_{i-1}) \frac{j u_i^*}{(u_i w_{i-1})^*}$$

Remark. In some applications, the squared-error is not the primary parameter affecting the performance of the system. The information bits may be carried over the phase of the transmitted signal, in which case it is useful to consider cost functions that relate to both the error magnitude and the phase error — see Tarighat and Sayed (2004).

Problem III.33 (Constant-modulus algorithm) The constant-modulus algorithm CMA2-2 is a stochastic gradient version of the steepest-descent algorithm developed in Prob. III.18.

(a) Replace the term $\gamma R_u w_{i-1} - \mathsf{E}[\, \boldsymbol{u}^* \boldsymbol{u} w_{i-1} |\boldsymbol{u} w_{i-1}|^2]$ by the instantaneous approximation $\gamma u_i^* u_i w_{i-1} - u_i^* u_i w_{i-1} |u_i w_{i-1}|^2$, and define $z(i) = u_i w_{i-1}$. Verify that this leads to the recursion

$$\boxed{w_i = w_{i-1} + \mu u_i^* z(i)\left[\gamma - |z(i)|^2\right], \quad z(i) = u_i w_{i-1}, \quad i \geq 0}$$

Remark. This recursion is known as CMA2-2. The numbers 2-2 refer to the fact that the cost function in this case (cf. Prob. III.18) is of the form $\mathsf{E}\left(\gamma - |\boldsymbol{u}w|^2\right)^2$, which is a special case of the more general cost function $\mathsf{E}\,(\gamma - |\boldsymbol{u}w|^p)^q$ for the values $p = 2$ and $q = 2$.

(b) Can you guarantee, as in the steepest-descent method of Prob. III.18, that the estimates w_i in CMA2-2 will always be real-valued for any real-valued initial condition w_{-1}? Justify your answer.

Problem III.34 (Reduced-constellation algorithm) The reduced-constellation algorithm (RCA) is a stochastic gradient version of the steepest-descent algorithm developed in Prob. III.19. Replace the expectations in the recursion of part (a) of Prob. III.19 by instantaneous approximations and verify that this substitution leads to the following recursion. Let $z(i) = u_i w_{i-1}$. Then

$$\boxed{w_i = w_{i-1} + \mu u_i^* \left(\gamma \mathsf{csgn}(z(i)) \; - \; z(i)\right)}$$

for some step-size $\mu > 0$.

Remark. Recall from the discussion in Prob. III.19 that RCA attempts to minimize the distance between the output of the filter and four points in the complex plane, namely, $\gamma(\pm 1 \pm j)$. In other words, for multi-level constellations, RCA attempts to minimize the mean-square error between the output of the filter and a reduced number of symbols, which may not belong to the original signal constellation. Compared with CMA, the RCA method is simple to implement but tends to face convergence difficulties (see, e.g., Werner et al. (1999)).

Problem III.35 (Multi-modulus algorithm) The multi-modulus algorithm (MMA) is a stochastic-gradient version of the steepest-descent algorithm developed in Prob. III.20. Replace the expectations in the recursion of Prob. III.20 by instantaneous approximations and verify that this substitution leads to the following algorithm. Let $z(i) = u_i w_{i-1}$. Then

$$\left\{\begin{array}{l} z(i) = u_i w_{i-1} \\ a(i) = \mathsf{Re}[z(i)] \\ b(i) = \mathsf{Im}[z(i)] \\ e(i) = a(i)[\gamma - a^2(i)] \; + \; jb(i)[\gamma - b^2(i)] \\ w_i = w_{i-1} + \mu u_i^* e(i) \end{array}\right.$$

for some step-size $\mu > 0$. How is this recursion different from that of CMA2-2 from Prob. III.33?

Remark. The MMA scheme was proposed by Yang, Werner, and Dumont (1997). Recall from Prob. III.20 that MMA attempts to minimize the distance between the output of the filter and the horizontal and vertical lines located at $\pm\sqrt{\gamma}$. Specifically, RCA attempts to minimize the dispersion between the real and imaginary parts of the filter output around the value of γ.

Problem III.36 (Another constant-modulus algorithm) The constant modulus algorithm CMA1-2 is a stochastic-gradient version of the steepest-descent algorithm developed in Prob. III.21.

(a) Replace the expectations in the recursion of part (a) of Prob. III.21 by instantaneous approximations and verify that this substitution leads to the following recursion. Let $z(i) = u_i w_{i-1}$.

Then

$$\boxed{w_i = w_{i-1} + \mu u_i^* \left(\gamma \frac{z(i)}{|z(i)|} - z(i)\right)}$$

for some step-size $\mu > 0$ (when $z(i) = 0$ we set $w_i = w_{i-1}$).

Remark. This recursion is known as CMA1-2 because the cost function in this case (cf. Prob. III.21) is of the form $\mathsf{E}\,(\gamma - |\boldsymbol{u}w|)^2$, which is a special case of the more general cost function $\mathsf{E}\,(\gamma - |\boldsymbol{u}w|^p)^q$ for the values $p = 1$ and $q = 2$.

(b) Let $w_i = w_{i-1} + \delta w$. Given w_{i-1}, show that the weight vector w_i with the smallest perturbation δw that solves the optimization problem

$$\min_{w_i} |u_i w_i - u_i w_{i-1}|^2 \qquad \text{subject to } |u_i w_i| = \gamma$$

is given by

$$w_i = w_{i-1} + \frac{u_i^*}{\|u_i\|^2}\left(\gamma \frac{z(i)}{|z(i)|} - z(i)\right)$$

Remark. Inserting a step-size μ into the above recursion leads to the normalized CMA algorithm,

$$\boxed{w_i = w_{i-1} + \mu \frac{u_i^*}{\|u_i\|^2}\left(\gamma \frac{z(i)}{|z(i)|} - z(i)\right)}$$

Problem III.37 (Gauss-Newton algorithm) A stochastic-gradient method that is similar in nature to the RLS algorithm of Sec. 14.1 can be obtained in the following manner.

(a) Let $\epsilon(i) = \lambda^{i+1}\epsilon/(i+1)$ and $\mu(i) = \mu$. Use the same approximations as in the RLS case for R_u and $(R_{du} - R_u w_{i-1})$ to verify that the regularized Newton's recursion (14.1) reduces to

$$w_i = w_{i-1} + \mu \Phi_i^{-1}\, u_i^*[d(i) - u_i w_{i-1}]$$

where

$$\Phi_i \triangleq \left(\frac{\lambda^{i+1}\epsilon}{(i+1)}\mathrm{I} + \frac{1}{i+1}\sum_{j=0}^{i} \lambda^{i-j} u_j^* u_j\right), \qquad i \geq 0$$

(b) Show that Φ_i satisfies the recursion $\Phi_i = \lambda[1 - \alpha(i)]\Phi_{i-1} + \alpha(i) u_i^* u_i$ for $i \geq 1$, with initial condition $\Phi_0 = \lambda\epsilon\mathrm{I} + u_0^* u_0$, and where $\alpha(i) = 1/(i+1)$.

(c) Define $P_i = \Phi_i^{-1}$. Show that

$$\begin{aligned} P_i &= \frac{\lambda^{-1}}{[1-\alpha(i)]}\left[P_{i-1} - \frac{\lambda^{-1} P_{i-1} u_i^* u_i P_{i-1}}{\frac{[1-\alpha(i)]}{\alpha(i)} + \lambda^{-1} u_i P_{i-1} u_i^*}\right], \qquad i \geq 1 \\ w_i &= w_{i-1} + \mu P_i\, u_i^*[d(i) - u_i w_{i-1}], \qquad i \geq 0 \end{aligned}$$

with initial condition

$$P_0 = \frac{1}{\lambda\epsilon}\left[\mathrm{I} - \frac{u_0^* u_0}{\lambda\epsilon + \|u_0\|^2}\right]$$

(d) Repeat the calculations in Tables 5.7 and 5.8 and estimate the amount of computations that are required per iteration for both cases of real and complex-valued data.

Problem III.38 (Simplified GN algorithm) Consider the same setting as in Prob. III.37. An alternative form of the GN algorithm can be obtained by replacing $\alpha(i)$ by a constant positive number α (usually small, say $0 < \alpha < 0.1$).

(a) Verify that the recursions of part (c) of Prob. III.37 reduce to

$$\begin{aligned} P_i &= \frac{\lambda^{-1}}{1-\alpha}\left[P_{i-1} - \frac{\lambda^{-1} P_{i-1} u_i^* u_i P_{i-1}}{\frac{(1-\alpha)}{\alpha} + \lambda^{-1} u_i P_{i-1} u_i^*}\right], \qquad P_{-1} = \epsilon^{-1}\mathrm{I} \\ w_i &= w_{i-1} + \mu P_i\, u_i^*[d(i) - u_i w_{i-1}], \qquad i \geq 0 \end{aligned}$$

(b) Show further that the above recursions could have been derived directly from Newton's recursion (14.1) by using the following sample average approximation for R_u, $\widehat{R}_u = \alpha\sum_{j=0}^{i}[\lambda(1-\alpha)]^{i-j}u_j^*u_j$, along with the choice $\epsilon(i) = \lambda^{(i+1)}(1-\alpha)^{(i+1)}\epsilon$.

Problem III.39 (Sample covariance matrices) Consider the sample covariance matrix $\widehat{R}_u = \frac{1}{i+1}\sum_{j=0}^{i}u_j^*u_j$ used in Sec. 14.1 (with exponential weighting). Let us denote $\widehat{R}_u$ by $\widehat{R}_{u,i}$ to indicate that it is based on data up to time i.

(a) Verify that $\widehat{R}_{u,i}$ satisfies the recursion $\widehat{R}_{u,i} = \frac{i}{i+1}\widehat{R}_{u,i-1} + \frac{1}{i+1}u_i^*u_i$.

(b) Consider instead the sample covariance matrix $\widehat{R}_{u,i} = \alpha\sum_{j=0}^{i}(1-\alpha)^{i-j}u_j^*u_j$ used in Prob. III.38. Verify now that $\widehat{R}_{u,i} = (1-\alpha)\widehat{R}_{u,i-1} + \alpha u_i^*u_i$.

Problem III.40 (Conversion factor) Show that the RLS algorithm can be written in the equivalent form

$$w_i = w_{i-1} + \frac{\lambda^{-1}P_{i-1}\,u_i^*}{1+\lambda^{-1}u_iP_{i-1}u_i^*}\,[d(i) - u_iw_{i-1}]\,, \quad i \geq 0.$$

Show further that $r(i) = \gamma(i)e(i)$ where $\gamma(i) = 1/(1+\lambda^{-1}u_iP_{i-1}u_i^*)$ and $r(i)$ and $e(i)$ denote the *a posteriori* and *a priori* output errors, $r(i) = d(i) - u_iw_i$ and $e(i) = d(i) - u_iw_{i-1}$. Conclude that, for all i, $|r(i)| \leq |e(i)|$. *Remark.* The coefficient $\gamma(i) = 1/(1+\lambda^{-1}u_iP_{i-1}u_i^*)$ is called the *conversion factor* since it converts the *a priori* error to the *a posteriori error.* We shall have more to say about the RLS algorithm, its properties, and its variants in Parts VII (*Least-Squares Methods)* through X (*Lattice Filters*).

Problem III.41 (Bound on the step-size for ϵ−APA) Refer to the discussion in Sec. 13.3 and to the definitions of the error vectors $\{e_i, r_i\}$.

(a) Use the constraint in (13.8) to show that $\|r_i\|^2 < \|e_i\|^2$ if, and only if, the matrix $\mathrm{I} - A$ is positive-definite, where

$$A \triangleq \left(\mathrm{I} - \mu U_iU_i^*(\epsilon\mathrm{I} + U_iU_i^*)^{-1}\right)^*\left(\mathrm{I} - \mu U_iU_i^*(\epsilon\mathrm{I} + U_iU_i^*)^{-1}\right)$$

Show further that $\|r_i\|^2 = \|e_i\|^2$ if, and only if, $e_i = 0$.

(b) Let $U_iU_i^* = V_i\Gamma_iV_i^*$ denote the eigen-decomposition of the $K\times K$ matrix $U_iU_i^*$, where V_i is unitary and $\Gamma_i = \text{diag}\{\gamma_0(i), \gamma_1(i), \ldots, \gamma_{K-1}(i)\}$ contains the corresponding eigenvalues. Show that $\mathrm{I} - A = \mu V_i\Gamma_i'(2\mathrm{I} - \mu\Gamma_i')V_i^*$, where $\Gamma_i' = \Gamma_i(\epsilon\mathrm{I} + \Gamma_i)^{-1}$.

(c) Conclude that $\mathrm{I} - A > 0$ if, and only if,

$$0 < \mu < \min_{0\leq k\leq K-1}\frac{2(\epsilon + \gamma_k(i))}{\gamma_k(i)}$$

Conclude that $0 < \mu < 2$ guarantees $\|r_i\|^2 \leq \|e_i\|^2$.

Problem III.42 (Least-perturbation property of ϵ-APA) Refer to the optimization problem (13.8).

(a) Introduce the difference $\delta w = w_i - w_{i-1}$. Show that the constraint amounts to the requirement $U_i\delta w = \mu U_iU_i^*(\epsilon\mathrm{I} + U_iU_i^*)^{-1}e_i$.

(b) Verify that the choice $\delta w^o = \mu U_i^*(\epsilon\mathrm{I} + U_iU_i^*)^{-1}e_i$ satisfies the constraint.

(c) Now complete the argument, as in Sec. 10.4, to show that the solution of (13.8) leads to the ϵ-APA recursion (13.5).

Problem III.43 (Gram-Schmidt orthogonalization) Consider three row vectors $\{u_1, u_2, u_3\}$ and define the transformed vectors (also called residuals):

$$\tilde{u}_1 \;=\; u_1, \quad \tilde{u}_2 = u_2 - u_2\frac{\tilde{u}_1^*\tilde{u}_1}{\|\tilde{u}_1\|^2}, \quad \tilde{u}_3 = u_3 - u_3\frac{\tilde{u}_1^*\tilde{u}_1}{\|\tilde{u}_1\|^2} - u_3\frac{\tilde{u}_2^*\tilde{u}_2}{\|\tilde{u}_2\|^2}$$

(a) Verify that the residual vectors so obtained are orthogonal to each other, i.e., show that $\tilde{u}_i\tilde{u}_j^* = 0$ for $i \neq j$.

(b) Verify further that $\{u_1, u_2, u_3\}$ and $\{\tilde{u}_1, \tilde{u}_2, \tilde{u}_3\}$ are related via an invertible lower triangular matrix as follows:

$$\begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -\frac{u_2\tilde{u}_1^*}{\|\tilde{u}_1\|^2} & 1 & \\ -\frac{u_3}{\|\tilde{u}_1\|^2}\left(\mathrm{I} - \frac{\tilde{u}_2^* u_2}{\|\tilde{u}_2\|^2}\right)\tilde{u}_1^* & -\frac{u_3\tilde{u}_2^*}{\|\tilde{u}_2\|^2} & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Problem III.44 (APA with orthogonal correction factors) Consider an APA update of order $K = 3$, i.e., $e_i = d_i - U_i w_{i-1}$ and $w_i = w_{i-1} + \mu U_i^* \left(U_i U_i^*\right)^{-1} e_i$, where

$$U_i = \begin{bmatrix} u_i \\ u_{i-1} \\ u_{i-2} \end{bmatrix}, \qquad d_i = \begin{bmatrix} d(i) \\ d(i-1) \\ d(i-2) \end{bmatrix}$$

In this problem we want to show that APA can be implemented in an equivalent form that involves only orthogonal regression vectors. So assume that, at each iteration i, the regressors $\{u_i, u_{i-2}, u_{i-3}\}$ are first orthogonalized, as described in Prob. III.43, and let $\{\tilde{u}_i, \tilde{u}_{i-1}, \tilde{u}_{i-2}\}$ denote the corresponding residual vectors (with $\tilde{u}_i = u_i$). Let L_i denote the lower-triangular matrix relating *the residuals* to the regressors, i.e.,

$$\tilde{U}_i \triangleq \begin{bmatrix} \tilde{u}_i \\ \tilde{u}_{i-1} \\ \tilde{u}_{i-2} \end{bmatrix} = L_i U_i$$

(a) Verify that the APA update can be written as $w_i = w_{i-1} + \mu \tilde{U}_i^* \left(\tilde{U}_i \tilde{U}_i^*\right)^{-1} L_i e_i$, where $\tilde{U}_i \tilde{U}_i^*$ is diagonal and given by $\tilde{U}_i \tilde{U}_i^* = \text{diag}\left\{ \|\tilde{u}_i\|^2, \|\tilde{u}_{i-1}\|^2, \ldots, \|\tilde{u}_{i-2}\|^2 \right\}$.

(b) Show that the entries of $L_i e_i$ are given by $L_i e_i = \text{col}\{e(i), e^{(1)}(i-1), e^{(2)}(i-2)\}$, where

$$e(i) = d(i) - u_i w_{i-1}, \quad e^{(1)}(i-1) = d(i-1) - u_{i-1} w_{i-1}^{(1)}, \quad e^{(2)}(i-2) = d(i-2) - u_{i-2} w_{i-1}^{(2)}$$

and $\{w_{i-1}^{(1)}, w_{i-1}^{(2)}\}$ are intermediate corrections to w_{i-1} defined by

$$w_{i-1}^{(1)} \triangleq w_{i-1} + \frac{\tilde{u}_i^*}{\|\tilde{u}_i\|^2} e(i), \qquad w_{i-1}^{(2)} \triangleq w_{i-1}^{(1)} + \frac{\tilde{u}_{i-1}^*}{\|\tilde{u}_{i-1}\|^2} e^{(1)}(i-1)$$

(c) Conclude that a general $K-$th order APA update can be equivalently implemented as follows. For each time instant i, start with $w_{i-1}^{(0)} = w_{i-1}$ and repeat for $k = 0, 1, \ldots, K-1$:

$$\boxed{\begin{aligned} e^{(k)}(i-k) &= d(i-k) - u_{i-k} w_{i-1}^{(k)} \\ w_{i-1}^{(k+1)} &= w_{i-1}^{(k)} + \frac{\tilde{u}_{i-k}^*}{\|\tilde{u}_{i-k}\|^2} e^{(k)}(i-k) \end{aligned}}$$

Then set

$$\boxed{w_i = w_{i-1} + \mu \sum_{k=0}^{K-1} \frac{\tilde{u}_{i-k}^*}{\|\tilde{u}_{i-k}\|^2} e^{(k)}(i-k)}$$

Remark. This form of the algorithm is sometimes referred to as APA or NLMS with orthogonal correction factors (APA-OCF or NLMS-OCF) since the regressors $\{\tilde{u}_i\}$ that are used in the update of the weight vector are orthogonal to each other.

(d) Verify further from part (c) that w_i is given by the convex combination

$$\boxed{w_i = (1-\mu) w_{i-1} \; + \; \mu w_{i-1}^{(K)}}$$

Problem III.45 (LMS as a notch filter) Consider an LMS adaptive filter with a regression vector u_i with shift-structure, i.e., its entries are delayed replicas of an input sequence $\{u(i)\}$, as in an FIR implementation,

$$u_i = \begin{bmatrix} u(i) & u(i-1) & \dots & u(i-M+1) \end{bmatrix}$$

Assume $u(i)$ is sinusoidal, say $u(i) = e^{j\omega_o i}$ for some ω_o. Assume also that the filter is initially at rest. Let w_i denote the coefficients of the LMS filter, which are adapted according to the rule $w_i = w_{i-1} + \mu u_i^*[d(i) - u_i w_{i-1}]$. Although the coefficients of the filter vary with time, due to the adaptation process, it turns out that the input-output mapping is actually time-invariant in this case.

(a) Show that the transfer function from the desired signal $d(i)$ to the error signal $e(i) = d(i) - u_i w_{i-1}$ is given by

$$\frac{E(z)}{D(z)} = \frac{z - e^{j\omega_o}}{z + (\mu M - 1)e^{j\omega_o}}$$

(b) Assume the adaptive filter is trained instead by $\epsilon-$NLMS. Show that the same transfer function becomes

$$\frac{E(z)}{D(z)} = \frac{z - e^{j\omega_o}}{z + \left(\frac{\mu M}{\epsilon + M} - 1\right) e^{j\omega_o}}$$

What does this result reduce to when $M \to \infty$?

(c) Assume the filter is trained using the power normalized $\epsilon-$NLMS algorithm (cf. Alg. 11.2). Show that in the limit, as $i \to \infty$, the transfer function from $d(i)$ to $e(i)$ is again given by

$$\frac{E(z)}{D(z)} = \frac{z - e^{j\omega_o}}{z + \left(\frac{\mu M}{\epsilon + M} - 1\right) e^{j\omega_o}}$$

Remark. The results indicate that, with a sinusoidal excitation, the LMS filter behaves like a notch filter with a notch frequency at ω_o. Applications to adaptive noise cancelling of sinusoidal interferences can be found in Widrow et al. (1976) and Glover (1977).

Problem III.46 (Adaptive line enhancement) Let $d(i) = s(i) + v(i)$ denote a zero-mean random sequence that consists of two components: a signal component, $s(i)$, and a noise component, $v(i)$. Let $r_s(k)$ and $r_v(k)$ denote the auto-correlation sequences of $\{s(i), v(i)\}$, assumed stationary, i.e., $r_s(k) = \mathsf{E}\, s(i)s^*(i-k)$ and $r_v(k) = \mathsf{E}\, v(i)v^*(i-k)$.

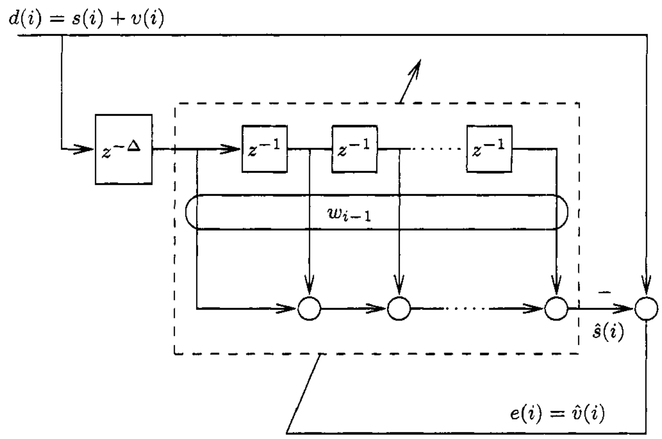

FIGURE III.1 An adaptive structure for line signal enhancement.

Assume that $r_s(k)$ is negligible for $k > \delta_s$, while $r_v(k)$ is negligible for $k > \delta_n$ with $\delta_s \gg \delta_n$. We say that $\boldsymbol{s}(i)$ corresponds to the narrowband component of $\boldsymbol{d}(i)$ while $\boldsymbol{v}(i)$ corresponds to the wideband component of $\boldsymbol{d}(i)$. The adaptive structure of Fig. III.1 is suggested for use in separating the signal $\boldsymbol{s}(i)$ from the noise $\boldsymbol{v}(i)$. As the figure indicates, realizations $d(i)$ are used as the reference sequence, and a delayed replica of these same realizations is used as input to the tapped delay line. The value of the delay Δ is chosen to satisfy $\delta_n < \Delta < \delta_s$, and the filter taps are trained using LMS, for example. It is claimed that the output of the filter, $u_i w_{i-1}$, provides estimates of the signal component, while the error signal, $e(i)$, provides estimates of the noise component. That is, $\hat{v}(i) = e(i)$ and $\hat{s}(i) = u_i w_{i-1}$. Justify the validity of this claim. *Hint:* Refer to Prob. II.13.

Remark. The adaptive structure described in this problem is known as an adaptive line enhancer (ALE); it permits separating a narrowband signal (e.g., a sinusoid) from a wideband noise signal. The ALE was originally developed by Widrow et al. (1975). Its performance was later studied in some detail by Zeidler et al. (1978), Rickard and Zeidler (1979), Treichler (1979), and Zeidler (1990).

COMPUTER PROJECTS

Project III.1 (Constant-modulus criterion) Refer to Prob. III.18, where we introduced the cost function $J(w) = \mathsf{E}\left(\gamma - |\boldsymbol{u}w|^2\right)^2$, for a given positive constant γ. This cost arises in the context of blind equalization where it is used to derive blind adaptive filters — see, e.g., Prob. III.33. In this project, we use $J(w)$ to highlight some of the issues that arise in the design of steepest-descent methods.

(a) Assume first that w is one-dimensional and $\boldsymbol{u}$ is scalar-valued with variance $\sigma_u^2 = \mathsf{E}|\boldsymbol{u}|^2$. Assume further that $\boldsymbol{u}$ and w are real-valued and that $\boldsymbol{u}$ is Gaussian so that its fourth moment is given by $\mathsf{E}\,\boldsymbol{u}^4 = 3\sigma_u^4$. Verify that, under these conditions, the cost function $J(w)$ evaluates to $J(w) = \gamma^2 - 2\gamma\sigma_u^2 w^2 + 3\sigma_u^4 w^4$. Conclude that $J(w)$ has a local maximum at $w = 0$ and two global minima at $w^o = \pm\sqrt{\gamma/3\sigma_u^2}$. Plot $J(w)$ and determine the values of w^o using $\gamma = 1$ and $\sigma_u^2 = 0.5$. Find also the corresponding minimum cost.

(b) Argue that a steepest-descent method for minimizing $J(w)$ can be taken as

$$w(i) = w(i-1) + 4\mu\sigma_u^2 w(i-1)[\gamma - 3\sigma_u^2 w^2(i-1)], \quad w(-1) = \text{initial guess}$$

with scalar estimates $\{w(i)\}$. Does this method converge to a global minimum if the initial weight guess is chosen as $w(-1) = 0$? Why?

(c) Simulate the steepest-descent recursion of part (b) in the following cases and comment on its behavior in each case:

1. $\mu = 0.2$ and $w(-1) = 0.3$ or $w(-1) = -0.3$.
2. $\mu = 0.6$ and $w(-1) = 0.3$ or $w(-1) = -0.3$.
3. $\mu = 1$ and $w(-1) = -0.2$.

In each case, plot the evolution of $w(i)$ as a function of time. Plot also the graph $J[w(i)] \times w(i)$.

(d) Now consider the setting of part (b) in Prob. III.18 where w is two-dimensional.

(d.1) Generate a plot of the contour curves of the cost function for $\gamma = 1$, i.e.,

$$J(w) = \gamma^2 + 2\left(|\alpha|^2 + 2|\beta|^2\right)^2 - 2\gamma\left(|\alpha|^2 + 2|\beta|^2\right)$$

(d.2) Choose $\mu = 0.02$ and apply the resulting algorithm,

$$\begin{bmatrix} \alpha(i) \\ \beta(i) \end{bmatrix} = \begin{bmatrix} \alpha(i-1) \\ \beta(i-1) \end{bmatrix} + \mu\left(\gamma - 2|\alpha(i-1)|^2 - 4|\beta(i-1)|^2\right)\begin{bmatrix} \alpha(i-1) \\ 2\beta(i-1) \end{bmatrix}$$

starting from the initial conditions $w_{-1} = \text{col}\{1, -0.25\}$ and $w_{-1} = \text{col}\{-1, 0.25\}$. Iterate over a period of length $N = 1000$. Plot the trajectory of the weight estimates superimposed on the contour curves of $J(w)$.

Adaptive filters are used in many applications and we cannot attempt to cover all of them in a textbook. Instead, we shall illustrate the use of adaptive filters in selected applications of heightened interest, including channel equalization, channel estimation, and echo cancellation. The computer projects in this section will focus on channel equalization. In later parts of the book, the computer projects will consider applications involving line echo cancellation, channel tracking, channel estimation, acoustic echo cancellation, and active noise control. In addition, in some of the problems, other applications are considered, such as adaptive line enhancement in Prob. III.46.

Project III.2 (Constant-modulus algorithm) In Prob. III.18 we introduced the constant-modulus criterion

$$\min_{w} \mathsf{E}\left(\gamma - |\boldsymbol{u}w|^2\right)^2$$

and developed a steepest-descent method for it, namely

$$w_i = w_{i-1} + \mu\left(\gamma R_u w_{i-1} - \mathsf{E}\left[\boldsymbol{u}^*\boldsymbol{u}w_{i-1}|\boldsymbol{u}w_{i-1}|^2\right]\right)$$

We reconsidered this method in Prob. III.33 and derived the corresponding stochastic gradient approximation, known as CMA2-2, namely,

$$w_i = w_{i-1} + \mu u_i^* z(i)\left[\gamma - |z(i)|^2\right], \quad z(i) = u_i w_{i-1}$$

In this project we compare the performance of the steepest-descent method and the CMA2-2 recursion. For this purpose, we set $\gamma = 1$ and let $\boldsymbol{u}$ be a 2-dimensional circular Gaussian random vector with covariance matrix $R_u = \text{diag}\{1, 0\}$. We showed in Prob. III.18 that, under these conditions, the steepest-descent method collapses to the form

$$\begin{bmatrix} \alpha(i) \\ \beta(i) \end{bmatrix} = \begin{bmatrix} \alpha(i-1) \\ \beta(i-1) \end{bmatrix} + \mu\left(\gamma - 2|\alpha(i-1)|^2 - 4|\beta(i-1)|^2\right)\begin{bmatrix} \alpha(i-1) \\ 2\beta(i-1) \end{bmatrix}$$

where $w_i = \text{col}\{\alpha(i), \beta(i)\}$. In addition, we showed in Prob. III.18 that the corresponding cost function evaluates to

$$J(w_i) = \gamma^2 + 2\left(|\alpha(i)|^2 + 2|\beta(i)|^2\right)^2 - 2\gamma\left(|\alpha(i)|^2 + 2|\beta(i)|^2\right)$$

(a) Plot the learning curve $J(i) = J(w_{i-1})$ for the steepest-descent method; here w_{i-1} is the weight estimate that results from the steepest-descent recursion. Plot also an ensemble-average learning curve $\widehat{J}(i)$ for the CMA2-2 algorithm that is generated as follows:

$$\widehat{J}(i) = \frac{1}{L}\sum_{j=1}^{L}\left(\gamma - \left|z^{(j)}(i)\right|^2\right)^2, \quad i \geq 0$$

with the data generated from L experiments, say of duration N each. Choose $L = 200$ and $N = 500$. Use $\mu = 0.001$ and $w_{-1} = \text{col}\{-0.8, 0.8\}$. *Remark.* We are assuming complex-valued regressors $\boldsymbol{u}$, with the two entries of $\boldsymbol{u}$ having variances $\{1, 2\}$. In order to generate such regressors, create four separate *real-valued* zero-mean and independent Gaussian numbers $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{q}\}$ with variances $\{1/2, 1/2, 1, 1\}$, respectively, and then set $\boldsymbol{u} = [\boldsymbol{a} + j\boldsymbol{b} \quad \boldsymbol{p} + j\boldsymbol{q}]$.

(b) Plot the contour curves of the cost function $J(w) = \mathsf{E}\left(\gamma - |\boldsymbol{u}w|^2\right)^2$ superimposed on the four weight trajectories that are generated by CMA2-2 for the following four choices of initial conditions:

$$w_{-1} \in \left\{\begin{bmatrix} 0.8 & 0.8 \end{bmatrix}, \begin{bmatrix} -0.8 & 0.8 \end{bmatrix}, \begin{bmatrix} 0.8 & -0.8 \end{bmatrix}, \begin{bmatrix} -0.8 & -0.8 \end{bmatrix}\right\}$$

Use $\mu = 0.001$ and $N = 1000$ iterations in each case. Print the final value of the weight vector estimate in each case. Show also the trajectories of the weight estimates that are generated by the steepest-descent method. *Remark.* Although the weight estimates in the steepest-descent method are always real-valued for a real-valued initial condition w_{-1}, the same is not true for CMA2-2. However, the imaginary parts of the successive weight estimates will be small compared to the real parts. For this reason, when plotting the weight trajectories, we shall ignore the imaginary parts.

Project III.3 (Adaptive channel equalization) In Computer Projects II.1 and II.3 we dealt with the design of minimum mean-square error equalizers. In this project we examine the design of *adaptive* equalizers. We consider the same channel of Computer Project II.3, namely,

$$C(z) = 0.5 + 1.2z^{-1} + 1.5z^{-2} - z^{-3}$$

and proceed to design an adaptive linear equalizer for it. The equalizer structure is shown in Fig. III.2. Symbols $\{s(i)\}$ are transmitted through the channel and corrupted by additive complex-valued white noise $\{v(i)\}$. The received signal $\{u(i)\}$ is processed by the FIR equalizer to generate estimates $\{\hat{s}(i-\Delta)\}$, which are fed into a decision device. The equalizer possesses two modes of operation: a training mode during which a delayed replica of the input sequence is used as a reference sequence, and a decision-directed mode during which the output of the decision-device replaces the reference sequence. The input sequence $\{s(i)\}$ is chosen from a quadrature-amplitude modulation (QAM) constellation (e.g., 4-QAM, 16-QAM, 64-QAM, or 256-QAM).

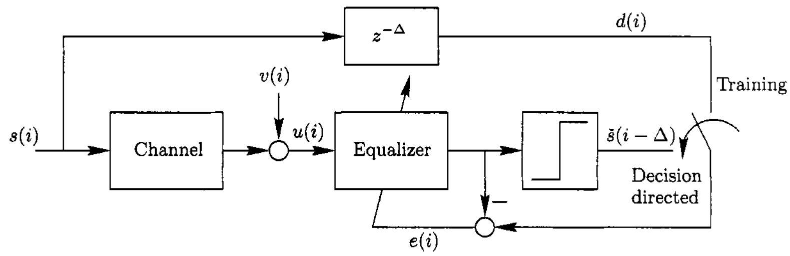

FIGURE III.2 An adaptive linear equalizer operating in two modes: training mode and decision-direction mode.

(a) Write a program that trains the adaptive filter with 500 symbols from a QPSK constellation, followed by decision-directed operation during 5000 symbols from a 16-QAM constellation. Choose the noise variance σ_v^2 in order to enforce an SNR level of 30 dB at the input of the equalizer. Note that symbols chosen from QAM constellations do not have unit variance. For this reason, the noise variance needs to be adjusted properly for different QAM orders in order to enforce the desired SNR level — see Prob. II.16. Choose $\Delta = 15$ and equalizer length $L = 35$. Use ϵ−NLMS to train the equalizer with step-size $\mu = 0.4$ and $\epsilon = 10^{-6}$. Plot the scatter diagrams of $\{s(i), u(i), \hat{s}(i-\Delta)\}$.

(b) For the same setting as part (a), plot and compare the scatter diagrams that would result at the output of the equalizer if training is performed only for 150, 300, and 500 iterations. Repeat the simulations using LMS with $\mu = 0.001$.

(c) Now assume the transmitted data are generated from a 256-QAM constellation rather than a 16-QAM constellation. Plot the scatter diagrams of the output of the equalizer, when trained with ϵ-NLMS using 500 training symbols.

(d) Generate symbol-error-rate (SER) curves versus signal-to-noise ratio (SNR) at the input of the equalizer for (4, 16, 64, 256)-QAM data. Let the SNR vary between 5 dB and 30 dB in increments of 1 dB.

(e) Continue with SNR at 30 dB. Design a decision-feedback equalizer with $L = 10$ feedforward taps and $Q = 2$ feedback taps. Use $\Delta = 7$ and plot the resulting scatter diagram of the output of the equalizer. Repeat for $L = 20$, $Q = 2$ and $\Delta = 10$. In both cases, choose the transmitted data from a 64-QAM constellation.

(f) Generate SER curves versus SNR at the input of the DFE for $(4, 16, 64, 256)$-QAM data. Let the SNR vary between 5 dB and 30 dB. Compare the performance of the DFE with that of the linear equalizer of part (d).

(g) Load the file channel, which contains the impulse response sequence of a more challenging channel with spectral nulls. Set the SNR level at the input of the equalizer to 40 dB and select a linear equalizer structure with 55 taps. Set also the delay at $\Delta = 30$. Train the equalizer using $\epsilon-$NLMS for 2000 iterations before switching to decision-directed operation. Plot the resulting scatter diagram of the output of the equalizer. Now train it again using RLS for 100 iterations before switching to decision-directed operation, and plot the resulting scatter diagram. Compare both diagrams.

Project III.4 (Blind adaptive equalization) We consider the same channel used in Computer Project III.3,

$$C(z) = 0.5 + 1.2z^{-1} + 1.5z^{-2} - z^{-3}$$

and proceed to design blind adaptive equalizers for it. The equalizer structure is shown in Fig. III.3. Symbols $\{s(i)\}$ are transmitted through the channel and corrupted by additive complex-valued white noise $\{v(i)\}$. The received signal $\{u(i)\}$ is processed by a linear equalizer, whose outputs $\{z(i)\}$ are fed into a decision device to generate $\{\check{s}(i-\Delta)\}$. These signals are delayed decisions and the value of Δ is determined by the delay that the signals undergo when travelling through the channel and the equalizer. In this project, the equalizer is supposed to operate blindly, i.e., without a reference sequence and therefore without a training mode. Most blind algorithms use the output of the equalizer, $z(i)$, to generate an error signal $e(i)$, which is used to adapt the equalizer coefficients according to the rule

$$w_i = w_{i-1} + \mu u_i^* e(i)$$

where u_i is the regressor at time i. Some blind algorithms use the output of the decision device, $\check{s}(i-\Delta)$, to evaluate $e(i)$ (e.g., the "stop-and-go" variant described in part (d) below).

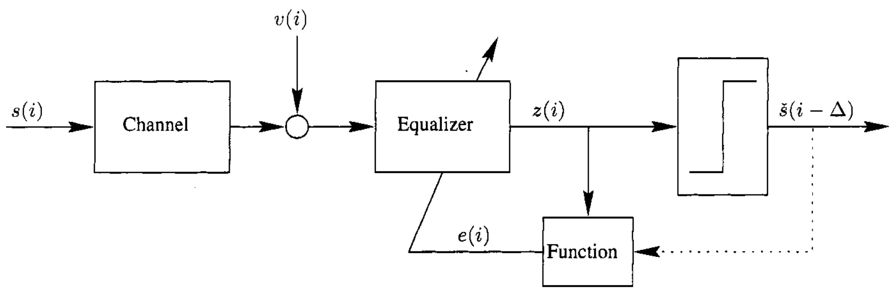

FIGURE III.3 A representation of a general structure for blind adaptive equalization.

(a) Write a program that transmits 10000 QPSK symbols through the channel and trains a 35-tap equalizer using CMA2-2,

$$w_i = w_{i-1} + \mu u_i^* z(i)\left[\gamma - |z(i)|^2\right], \quad z(i) = u_i w_{i-1}$$

Choose the value of γ as $\mathsf{E}\,|s|^4/\mathsf{E}\,|s|^2$, which is in terms of the second and fourth-order moments of the symbol constellation. For QPSK data, $\gamma = 1$. Set the SNR level at the input of the equalizer to 30 dB and use $\mu = 0.001$. Plot the impulse responses of the channel, the equalizer, and the combination channel-equalizer at the end of the adaptation process. How much delay does the signal undergo in travelling through the channel and the equalizer? Plot the scatter diagrams of the transmitted sequence, the received sequence, and the sequence at the output of the equalizer. Ignore the first 2000 transmissions and count the number of erroneous decisions in the remaining decisions (you should take into account the delay introduced by the channel-equalizer system).

(b) Repeat the simulations of part (a) using 16-QAM data, for which $\gamma = 13.2$ (verify this value). Use $\mu = 0.000001$. Run the simulation for 30000 symbols and ignore the first 15000 for error calculation. These numbers are larger than in the QPSK case of part (a), and the step-size is also significantly smaller, since the equalizer converges at a slower pace now.

(c) Repeat the simulations of part (b) using the multi-modulus algorithm (MMA) of Prob. III.35.

(d) Repeat the simulations of part (b) using the following three additional blind adaptive algorithms:

(d.1) CMA1-2 from Prob. III.36, where γ is now chosen as $\gamma = \mathsf{E}\,|\boldsymbol{s}|^2/\mathsf{E}\,|\boldsymbol{s}|$ in terms of the second moment of the symbol constellation divided by the mean of the magnitude of the symbols. For 16-QAM data we find $\gamma = 3.3385$ (verify this value). Use $\mu = 0.0001$ and increase the SNR level at the input of the equalizer to 60 dB. Simulate for 30000 iterations and plot the scatter diagram of the output of the equalizer after ignoring the first 15000 samples.

(d.2) The reduced constellation algorithm (RCA) of Prob. III.34, where γ is now chosen as $\gamma = \mathsf{E}\,|\boldsymbol{s}|^2/\mathsf{E}\,|\boldsymbol{s}|_1$, in terms of the second moment of the symbol constellation divided by the mean of the l_1-norm of the symbols (remember that the l_1 norm of a complex number amounts to adding the absolute values of its real and imaginary parts, as in Prob. III.14). For 16-QAM data we find $\gamma = 2.5$ (verify this value). Use the same step-size and same simulation duration as part (d.1).

(d.3) The "stop-and-go" algorithm is a blind adaptation scheme that employs the decision-directed error,

$$e_d(i) = \check{s}(i-\Delta) - z(i)$$

It also employs a flag to indicate how reliable $e_d(i)$ is. The flag is set by comparing $e_d(i)$ to another error signal, say the one used by the RCA recursion,

$$e_s(i) = \gamma \mathsf{csgn}(z(i)) - z(i), \qquad \gamma = \mathsf{E}\,|\boldsymbol{s}|^2/\mathsf{E}\,|\boldsymbol{s}|_1$$

If the complex signs of $\{e_d(i), e_s(i)\}$ differ, then $e_d(i)$ is assumed unreliable and the flag is set to zero (see Picchi and Prati (1987)). More explicitly, the stop-and-go recursion takes the form:

$$\left\{\begin{array}{l} z(i) = u_i w_{i-1} \\ e_d(i) = \check{s}(i-\Delta) - z(i) \qquad \text{(decision-directed error)} \\ e_s(i) = \gamma \mathsf{csgn}(z(i)) - z(i) \qquad \text{(RCA error)} \\ \\ f(i) = \left\{\begin{array}{ll} 1 & \text{if } \mathsf{csgn}(e_d(i)) = \mathsf{csgn}(e_s(i)) \\ \\ 0 & \text{if } \mathsf{csgn}(e_d(i)) \neq \mathsf{csgn}(e_s(i)) \end{array}\right. \qquad \text{(a flag)} \\ \\ e(i) = f(i) e_d(i) \\ w_i = w_{i-1} + \mu u_i^* e(i) \end{array}\right.$$

Use the same step-size and simulation duration as in part (d.1).