Solutions to the Problems in Adaptive Signal Processing by Bernard Widrow and Samuel D. Stearns

John Weatherwax

To my family.

Introduction

This is a solution manual to some of the problems in the excellent textbook:

Adaptive Signal Processing
by Bernard Widrow and Samuel D. Stearns

I'm currently working aggressively on finishing more of the problems in this book. In the meantime I'm publishing my partial results for any student who does not want to wait for the full book to be finished.

One of the benefits of this manual is that I heavily use the `R` statistical language to perform any of the needed numerical computations (rather than do them "by-hand"). Thus if you work though this manual you will be learning the `R` language at the same time as you learn statistics. The `R` programming language is one of the most desired skills for anyone who hopes to use data/statistics in their future career. The `R` code can be found at the following location:

`https://waxworksmath.com/Authors/N_Z/Widrow/widrow.html`

As a final comment, I've worked hard to make these notes as good as I can, but I have no illusions that they are perfect. If you feel that that there is a better way to accomplish or explain an exercise or derivation presented in these notes; or that one or more of the explanations is unclear, incomplete, or misleading, please tell me. If you find an error of any kind – technical, grammatical, typographical, whatever – please tell me that, too. I'll gladly add to the acknowledgments in later printings the name of the first person to bring each problem to my attention.

Chapter 2: The Adaptive Linear Combiner

Exercise Solutions

Exercise 1

Part (a): Selecting almost any two matrices will show that $AB \neq BA$ in general. For example if we take

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$$
$$B = \begin{bmatrix} 1 & 0 \\ -1 & 3 \end{bmatrix},$$

then we see that

$$AB = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 3 \end{bmatrix}$$
$$BA = \begin{bmatrix} 1 & 0 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -2 & 2 \end{bmatrix}.$$

Note that $AB \neq BA$.

Part (b): This follows from the definition of the matrix product. The ijth element of the product $A(B+C)$ can be written as

$$\begin{aligned} (A(B+C))_{ij} &= \sum_k A_{ik}(B+C)_{kj} \\ &= \sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\ &= (AB)_{ij} + (AC)_{ij}\,. \end{aligned}$$

The last line is the sum of the ijth elements of the product AB and AC.

Part (c): We first show that $(AB)^T = B^T A^T$. Recall that the ijth element of the product AB is

$$(AB)_{ij} = \sum_k A_{ik}B_{kj}\,.$$

This means that the ijth element of $(AB)^T$ is then

$$(AB)_{ji} = \sum_k A_{jk}B_{ki}\,.$$

Next, the product of $B^T A^T$ has an ijth element given by

$$(B^T A^T)_{ij} = \sum_k (B^T)_{ik}(A^T)_{kj} = \sum_k B_{ki}A_{jk} = \sum_k A_{jk}B_{kj}\,.$$

Notice that this is equal to the ijth element of $(AB)^T$.

Next we show that $(AB)^{-1} = B^{-1}A^{-1}$. To do that we will simply verify that the matrix $B^{-1}A^{-1}$ has the required properties to be the inverse of the matrix AB. We have

$$B^{-1}A^{-1} \cdot AB = I \,,$$

and that

$$AB \cdot B^{-1}A^{-1} = I \,,$$

showing that the matrix product $B^{-1}A^{-1}$ is the matrix inverse of AB.

Part (d): If A has an inverse then

$$A \cdot A^{-1} = I \,.$$

If we take the transpose of this and use the above condition on the transpose of a product we have

$$(A^{-1})^T A^T = I \,.$$

As A is symmetric this means that

$$(A^{-1})^T A = I \,.$$

If we multiply "on-the-left" by A^{-1} we get

$$(A^{-1})^T = A^{-1} \,.$$

This states that A^{-1} is symmetric.

Exercise 2

The book's equation 2.13 is given by

$$\text{MSE} = E[d_k^2] + W^T R W - 2P^T W \,. \tag{1}$$

If we write this in terms of the components of the matrices and the vectors above we have

$$\text{MSE} = E[d_k^2] + \sum_{t,s} w_t R_{ts} w_s - 2 \sum_t P_t w_t \,.$$

This means that taking the derivative of this with respect to w_l gives

$$\begin{aligned}
\frac{\partial \text{MSE}}{\partial w_l} &= \sum_{t,s} \delta_{tl} R_{ts} w_s + \sum_{t,s} w_t R_{ts} \delta_{sl} - 2 \sum_t P_t \delta_{tl} \\
&= \sum_s R_{ls} w_s + \sum_t w_t R_{tl} - 2P_l \,.
\end{aligned}$$

Now the matrix R is symmetric so $R_{tl} = R_{lt}$ and the above is

$$\begin{aligned} \frac{\partial \text{MSE}}{\partial w_l} &= \sum_s R_{ls} w_s + \sum_t R_{lt} w_t - 2P_l \\ &= 2 \sum_s R_{ls} w_s - 2P_l \,. \end{aligned}$$

In vector form this is

$$\frac{\partial \text{MSE}}{\partial W} = 2RW - 2P \,,$$

as we were to show.

Exercise 3

For this example we have expressions for R, P, and W^* given in the book. Using the calculated weight vector W^* we would have that

$$\begin{aligned} y_k &= w_0 x_k + w_1 x_{k-1} \\ &= 2.752764 x_k - 3.402603 x_{k-1} \\ &= 2.752764 \sin\left(\frac{\pi k}{5}\right) - 3.402603 \sin\left(\frac{\pi(k-1)}{5}\right) \,. \end{aligned}$$

In the `R` code for this problem we plot this as a function of k along with d_k and we see that the two curves overlap (are identical) as they should be.

Exercise 4

From Figure 2.6 (and the discussion in the text) the square of the root-mean-square or MSE is denoted by ξ and for this example we have

$$\xi = 0.5(w_0^2 + w_1^2) + w_0 w_1 \cos\left(\frac{2\pi}{N}\right) + 2w_1 \sin\left(\frac{2\pi}{N}\right) + 2 \,.$$

Then to satisfy the conditions requested we should have

$$\sqrt{\xi} = 2 \quad \text{so} \quad \xi = 4 \,.$$

Then any weights (w_0, w_1) that satisfies

$$0.5(w_0^2 + w_1^2) + w_0 w_1 \cos\left(\frac{2\pi}{N}\right) + 2w_1 \sin\left(\frac{2\pi}{N}\right) = 2 \,,$$

would work. There will be a family of numbers that will make this true.

Exercise 5

We have $N = 8$ and from Eq. 2.25 when $w_1 = 0$ we get

$$\nabla = \begin{bmatrix} w_0 \\ w_0 \cos\left(\frac{2\pi}{N}\right) + 2\sin\left(\frac{2\pi}{N}\right) \end{bmatrix} .$$

From the books expression for the MSE ξ when $w_1 = 0$ we have

$$\xi = 2 + 0.5 w_0^2 .$$

If $\xi = 2$ then $w_0 = 0$. If $\xi = 4$ then $w_0 = 2$. Using these we can evaluate ∇. We find

```
[1] 0.000000 1.414214
[1] 2.000000 2.828427
```

The gradient is steeper in the second case because when $\xi = 2$ we are at a higher location on the parabolic bowl (further from the minimum).

Exercise 6

When the switch is open there is no signal from x_k (only a delayed signal). This means that the output from the system at time k is

$$y_k = w_1 x_{k-1} .$$

The vector X_k is then $X_k = [x_{k-1}]$ (a scalar). This means that

$$R = r = E[x_{k-1}^2] = 1$$
$$P = p = E[d_k x_{k-1}] = 1 .$$

The equation for the MSE (that we want to minimize) is then

$$\text{MSE} = \xi = E[\varepsilon_k^2] = E[d_k^2] + W^T R W - 2P^T W , \tag{2}$$

or with $W = w_1$ this becomes

$$\xi = 4 + r w_1^2 - 2p w_1 = 4 + w_1^2 - 2w_1 .$$

The value of w_1 that minimizes this is $w_1^* = 1$.

Exercise 7

When the switch is closed there is now an input from x_k. This means that the output from the system at time k is

$$y_k = x_k + w_1 x_{k-1} .$$

The vector X_k is then $X_k^T = [x_k\,, x_{k-1}]$ and the weight vector is $W^T = [1\,, w_1]$. This means that

$$\begin{aligned} R &= E[X_k X_k^T] \\ &= \begin{bmatrix} E[x_k^2] & E[x_k x_{k-1}] \\ E[x_k x_{k-1}] & E[x_{k-1}^2] \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \end{aligned}$$

and

$$P = E[d_k X_k] = \begin{bmatrix} E[d_k x_k] \\ E[d_k x_{k-1}] \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The equation for the MSE (that we want to minimize) is then

$$\text{MSE} = \xi = E[\varepsilon_k^2] = E[d_k^2] + W^T R W - 2P^T W\,,$$

or with $W = \begin{bmatrix} 1 \\ w_1 \end{bmatrix}$ this becomes

$$\begin{aligned} \xi &= 4 + \begin{bmatrix} 1 & w_1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ w_1 \end{bmatrix} - 2 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ w_1 \end{bmatrix} \\ &= 4 + \begin{bmatrix} 1 & w_1 \end{bmatrix} \begin{bmatrix} 1 + 0.5 w_1 \\ 0.5 + w_1 \end{bmatrix} - 2(-1 + w_1) \\ &= 4 + 1 + 0.5 w_1 + 0.5 w_1 + w_1^2 + 2 - 2 w_1 \\ &= 7 - w_1 + w_1^2\,. \end{aligned}$$

The value of w_1 that minimizes this is $w_1^* = \frac{1}{2}$.

Exercise 8

From the previous exercise we see that $w^* = 1$ and using that we find $\xi_{\text{min}} = 3$.

Exercise 9

We are told to take $N = 5$ with

$$d_k = 2 \cos\left(\frac{2\pi k}{N}\right).$$

For this function, the book argues that $E[d_k^2] = 2$. Now in Figure 2.6 the book compute the needed statistics for this exercise. From that figure we have

$$\begin{aligned} R &= r = E[x_{k-1}^2] = 0.5 \\ P &= p = E[d_k x_{k-1}] = -\sin\left(\frac{2\pi}{N}\right). \end{aligned}$$

The equation for the MSE (that we want to minimize) is then

$$\text{MSE} = \xi = E[\varepsilon_k^2] = E[d_k^2] + W^T R W - 2P^T W \,, \tag{3}$$

or with $W = w_1$ this becomes

$$\xi = 2 + r w_1^2 - 2 p w_1 = 2 + 0.5 w_1^2 + 2 \sin\left(\frac{2\pi}{N}\right) w_1 \,.$$

The value of w_1 that minimizes this is

$$w_1^* = -2 \sin\left(\frac{2\pi}{N}\right) \,.$$

The smallest MSE would be the above expression when $w_1 = w_1^*$.

Exercise 10

We are again told to take $N = 5$ with

$$d_k = 2 \cos\left(\frac{2\pi k}{N}\right) \,.$$

For this function, the book argues that $E[d_k^2] = 2$. Now in Figure 2.6 the book compute the needed statistics for this exercise. From that figure we have

$$R = \begin{bmatrix} E[x_k^2] & E[x_k x_{k-1}] \\ E[x_k x_{k-1}] & E[x_{k-1}^2] \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \cos\left(\frac{2\pi}{N}\right) \\ 0.5 \cos\left(\frac{2\pi}{N}\right) & 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1545085 \\ 0.1545085 & 0.5 \end{bmatrix}$$

$$P = \begin{bmatrix} E[d_k x_k] \\ E[d_k x_{k-1}] \end{bmatrix} = \begin{bmatrix} 0 \\ -\sin\left(\frac{2\pi}{N}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ -0.9510565 \end{bmatrix} .$$

The equation for the MSE (that we want to minimize) is then

$$\text{MSE} = \xi = E[\varepsilon_k^2] = E[d_k^2] + W^T R W - 2P^T W \,,$$

or with $W = \begin{bmatrix} 1 \\ w_1 \end{bmatrix}$ (as specified by the architecture given in Exercise 6 with the switch closed i.e. Exercise 7) this becomes

$$\begin{aligned} \xi &= 2 + \begin{bmatrix} 1 & w_1 \end{bmatrix} \begin{bmatrix} 0.5 & 0.1545085 \\ 0.1545085 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ w_1 \end{bmatrix} - 2 \begin{bmatrix} 0 & -0.9510565 \end{bmatrix} \begin{bmatrix} 1 \\ w_1 \end{bmatrix} \\ &= 2.5 + 2.21113 w_1 + 0.5 w_1^2 \,. \end{aligned}$$

The value of w_1 that minimizes this is $w_1^* = -2.21113$ and gives $\xi_{\min} = 0.055452$.

Exercise 13

The output of the linear filter is

$$y_k = w_0 x_{0k} + w_1 x_{1k} \,,$$

so that the error is

$$\begin{aligned}\varepsilon_k &= d_k - y_k = d_k - \left[\begin{array}{cc} w_0 & w_1 \end{array} \right] \left[\begin{array}{c} x_{0k} \\ x_{1k} \end{array} \right] \\ &= d_k - w_0 x_{0k} - w_1 x_{1k} \,.\end{aligned}$$

Part (a): If we start to compute ε_k^4 we get

$$\begin{aligned}\varepsilon_k^2 &= (d_k - w_0 x_{0k} - w_1 x_{1k})(d_k - w_0 x_{0k} - w_1 x_{1k}) \\ &= d_k^2 + w_0^2 x_{0k}^2 + w_1^2 x_{1k}^2 - 2 w_0 d_k x_{0k} - 2 w_1 d_k x_{1k} + 2 w_0 w_1 x_{0k} x_{1k} \,.\end{aligned}$$

Next we would need to compute have

$$\varepsilon_k^3 = (d_k^2 + w_0^2 x_{0k}^2 + w_1^2 x_{1k}^2 - 2 w_0 d_k x_{0k} - 2 w_1 d_k x_{1k} + 2 w_0 w_1 x_{0k} x_{1k})(d_k - w_0 x_{0k} - w_1 x_{1k}) \,.$$

We would need to multiply by another $d_k - w_0 x_{0k} - w_1 x_{1k}$ to get ε_k^4 and then take the expectation of the expression that results.

Part (b): From the above notice that there will be *fourth* powers of w_0 and w_1 and so the expression $E[\varepsilon_k^4]$ is *not* quadratic.

Part (c): Heuristically, to find the minimum we would need to satisfy the first order optimization conditions (that the first derivative is equal to zero). The first derivative (set equal to zero) of a fourth order function will be a third order equation which will have *three* roots in general. Thus the optimum will not be unique and is not unimodal.

Chapter 3: Properties of the Quadratic Performance Surface

Exercise Solutions

Exercise 1

Part (a): Eq. 3.30 is

$$V^T R V = \text{constant}\,. \tag{4}$$

When there are two weights this would be

$$\begin{bmatrix} v_0 & v_1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \end{bmatrix} = F\,,$$

where F is our constant. This is equal to

$$\begin{bmatrix} v_0 & v_1 \end{bmatrix} \begin{bmatrix} r_{11}v_0 + r_{12}v_1 \\ r_{12}v_0 + r_{22}v_1 \end{bmatrix} = F\,,$$

or

$$r_{11}v_0^2 + r_{12}v_0v_1 + r_{12}v_0v_1 + r_{22}v_1^2 = F\,,$$

or

$$r_{11}v_0^2 + 2r_{12}v_0v_1 + r_{22}v_1^2 = F\,.$$

Now following the hint, this will be an ellipse if using the $A \equiv r_{11}$, $B \equiv 2r_{12}$, and $C \equiv r_{22}$ that this expression indicates we have

$$(2r_{12})^2 - 4r_{11}r_{22} < 0\,.$$

This can be written as

$$r_{11}r_{22} - r_{12}^2 > 0\,.$$

This is *equivalent* to the condition $|R| > 0$ that the matrix R must satisfy to be positive definite. This means that the expression above is an ellipse.

Part (b): When there is one weight Equation 4 would be

$$r_{11}v_0^2 = F\,,$$

where F is a constant. The solutions v_0 to the above are the points $\pm\sqrt{\frac{F}{r_{11}}}$.

Exercise 2

Lets define the function F of the weights V as

$$F(V) = V^T R V\,,$$

then taking the derivative of F with respect to the i-th component of V i.e. v_i we have

$$\frac{\partial F}{\partial v_i} = e_i^T R V + V^T R e_i = 2 e_i^T R V \,.$$

Here e_i is the vector of all zeros but with a one in the i-th location. Forming the vector

$$\begin{bmatrix} \frac{\partial F}{\partial v_0} \\ \frac{\partial F}{\partial v_1} \\ \vdots \\ \frac{\partial F}{\partial v_L} \end{bmatrix} ,$$

we find it is equal to

$$\begin{bmatrix} 2e_0^T R V \\ 2e_1^T R V \\ \vdots \\ 2e_L^T R V \end{bmatrix} = 2 \begin{bmatrix} e_0^T \\ e_1^T \\ \vdots \\ e_L^T \end{bmatrix} R V = 2 R V \,,$$

which is the book's Eq. 3.31.

Exercise 3

Part (a): The characteristic equation for this R is given by

$$|R - \lambda I| = \begin{vmatrix} a - \lambda & b \\ b & a - \lambda \end{vmatrix} = 0 \,.$$

Expanding this we get

$$(a - \lambda)^2 - b^2 = 0 \,,$$

or

$$\lambda^2 - 2a\lambda + a^2 - b^2 = 0 \,.$$

Part (b): The characteristic equation for this R is given by

$$|R - \lambda I| = \begin{vmatrix} a - \lambda & b & c \\ b & a - \lambda & b \\ c & b & a - \lambda \end{vmatrix} = 0 \,.$$

Expanding the determinant along the first column this is

$$(a - \lambda) \begin{vmatrix} a - \lambda & b \\ b & a - \lambda \end{vmatrix} - b \begin{vmatrix} b & c \\ b & a - \lambda \end{vmatrix} + c \begin{vmatrix} b & c \\ a - \lambda & b \end{vmatrix} = 0 \,,$$

or expanding the 2×2 determinant we get

$$(a - \lambda)((a - \lambda)^2 - b^2) - b(b(a - \lambda) - bc) + c(b^2 - c(a - \lambda)) = 0 \,,$$

or in great detail

$$(a-\lambda)(a^2-2a\lambda+\lambda^2-b^2)-b(ab-b\lambda-bc)+cb^2-c^2(a-\lambda)=0\,,$$

or

$$a^3-2a^2\lambda+a\lambda^2-ab^2-a^2\lambda+2a\lambda^2-\lambda^3+b^2\lambda-ab^2+b^2\lambda+b^2c+cb^2-ac^2+c^2\lambda=0\,.$$

or

$$-\lambda^3+(a+2a)\lambda^2+(-2a^2-a^2+b^2+b^2+c^2)\lambda+a^3-ab^2-ab^2+b^2c+b^2c-ac^2=0\,,$$

or

$$-\lambda^3+3a\lambda^2+(-3a^2+2b^2+c^2)\lambda+a^3-2ab^2+2b^2c-ac^2=0\,,$$

or

$$\lambda^3-3a\lambda^2+(3a^2-2b^2-c^2)\lambda+(-a^3+2ab^2-2b^2c+ac^2)=0\,.$$

Exercise 4

For this R we have

$$|R-\lambda I|=0\Rightarrow\left|\begin{array}{cc}3-\lambda & 2\\ 2 & 3-\lambda\end{array}\right|=0\,.$$

Expanding and simplifying we get

$$(3-\lambda)^2=4\quad\text{so}\quad\lambda=3\pm2\,.$$

This means that $\lambda\in\{1,5\}$.

Exercise 5

For this R we have

$$|R-\lambda I|=0\Rightarrow\left|\begin{array}{cc}3-\lambda & 1\\ 1 & 3-\lambda\end{array}\right|=0\,.$$

Expanding and simplifying we get

$$(3-\lambda)^2=1\quad\text{so}\quad\lambda=3\pm1\,.$$

This means that $\lambda\in\{2,4\}$.

Exercise 6

Part (a): We have

$$|R-\lambda I|=\left|\begin{array}{cc}a-\lambda & b\\ b & c-\lambda\end{array}\right|=(a-\lambda)(c-\lambda)-b^2\,.$$

Setting this equal to zero and expanding gives

$$ac - (a+c)\lambda + \lambda^2 - b^2 = 0\,,$$

or

$$\lambda^2 - (a+c)\lambda + ac - b^2 = 0\,.$$

Part (b): We have

$$\begin{aligned}
|R - \lambda I| &= \left|\begin{array}{ccc} a-\lambda & b & c \\ b & d-\lambda & e \\ c & e & f-\lambda \end{array}\right| \\
&= (a-\lambda)\left|\begin{array}{cc} d-\lambda & e \\ e & f-\lambda \end{array}\right| - b\left|\begin{array}{cc} b & e \\ c & f-\lambda \end{array}\right| + c\left|\begin{array}{cc} b & d-\lambda \\ c & e \end{array}\right| \\
&= (a-\lambda)[(d-\lambda)(f-\lambda) - e^2] - b[b(f-\lambda) - ec] + c[be - c(d-\lambda)] \\
&= (a-\lambda)[df - (d+f)\lambda + \lambda^2 - e^2] - b[bf - b\lambda - ec] + c[be - cd + c\lambda] \\
&= adf - a(d+f)\lambda + a\lambda^2 - ae^2 - df\lambda + (d+f)\lambda^2 - \lambda^3 + e^2\lambda - b^2 f + b^2\lambda + bce + bce - c^2 d + c^2\lambda \\
&= -\lambda^3 + (a+d+f)\lambda^2 + (b^2 + c^2 + e^2 - ad - af - df)\lambda + adf + 2bce - ae^2 - b^2 f - c^2 d\,.
\end{aligned}$$

This would be set equal to zero to form the cubic equation we would solve for λ.

Exercise 7

Recall that for a linear combiner with $L = 1$ the matrix R is given by

$$R = E\left[\begin{array}{cc} x_{0k}^2 & x_{0k}x_{1k} \\ x_{1k}x_{0k} & x_{1k}^2 \end{array}\right]\,. \tag{5}$$

For a *single-input* linear combiner by stationarity we expect $E[x_{0k}^2] = E[x_{1k}^2]$ and thus this matrix takes the form

$$R = \left[\begin{array}{cc} a & b \\ b & a \end{array}\right]\,, \tag{6}$$

for some constants a and b.

For a linear combiner with $L = 2$ the matrix R is given by

$$R = E\left[\begin{array}{ccc} x_{0k}^2 & x_{0k}x_{1k} & x_{0k}x_{2k} \\ x_{1k}x_{0k} & x_{1k}^2 & x_{1k}x_{2k} \\ x_{2k}x_{0k} & x_{2k}x_{1k} & x_{2k}^2 \end{array}\right]\,. \tag{7}$$

For a *single-input* linear combiner by stationarity we expect $E[x_{0k}^2] = E[x_{1k}^2] = E[x_{2k}^2]$ and

$$E[x_{0k}x_{1k}] = E[x_{1k}x_{2k}]\,,$$

and thus this matrix takes the form

$$R = \left[\begin{array}{ccc} a & b & c \\ b & a & b \\ c & b & a \end{array}\right]\,,$$

for some constants a, b, and c. Thus the matrices in Exercise 3 are direct analogs of these and the matrices in Exercise 6 are under certain constraints on their values.

For a *multiple-input* linear combiner with $L = 1$ we don't have $E[x_{0k}^2] = E[x_{1k}^2]$ since x_{0k} is from a different signal than x_{1k} and not just a lagged value of x_k thus R in this case looks like

$$R = \begin{bmatrix} a & b \\ b & c \end{bmatrix} .$$

By the same logic for a multiple-input linear combiner with $L = 2$ many of the expectation are now different and we have

$$R = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} .$$

These matrices better matched the ones given in Exercise 6.

Exercises 8-13

Its helpful to be able to perform these calculations using a programming language as its unlikely that for larger systems a person would be performing these operations "by hand". Thus for these problems I've performed them using the `python` programming language. Please see the code `exercises_8_13.py` where we perform all requested manipulations.

Exercise 14

From the problem statement we are told that the input correlation matrix R is given by

$$R = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} ,$$

with the vector of cross correlations given by

$$P = E[d_k X_k] = \begin{bmatrix} E[d_k x_{0k}] & E[d_k x_{1k}] \end{bmatrix}^T = \begin{bmatrix} 6 & 4 \end{bmatrix}^T .$$

Part (a): The mean-square error ξ we seek to minimize is given by

$$\begin{aligned} \text{MSE} = \xi = E[\varepsilon_k] &= E[d_k^2] + W^T R W - 2P^T W \\ &= 36 + \begin{bmatrix} w_0 & w_1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} - 2 \begin{bmatrix} 6 & 4 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \\ &= 2w_0^2 + 2w_0 w_1 + 3w_1^2 - 12w_0 - 8w_1 + 36 , \end{aligned}$$

when we simplify.

Part (b): The optimum weight vector w^* is given by

$$w^* = R^{-1}P = \begin{bmatrix} 2.8 \\ 0.4 \end{bmatrix},$$

when we invert and multiply.

Part (c): Putting those values into the above expression for ξ gives $\xi = 17.6$.

Part (d): The eigenvalues and eigenvectors of R are given by

```
> ev
eigen() decomposition
$values
[1] 3.618034 1.381966

$vectors
          [,1]       [,2]
[1,] 0.5257311 -0.8506508
[2,] 0.8506508  0.5257311
```

Exercise 15

A single input linear combiner with two weights will have a correlation matrix R of the form given by Equation 6 or

$$R = \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

To compute the eigenvalues we need to solve $|R - \lambda I| =$ which is

$$\begin{vmatrix} a-\lambda & b \\ b & a-\lambda \end{vmatrix} = (a-\lambda)^2 - b^2 = 0\,.$$

Solving this gives

$$\lambda = a \pm b\,.$$

For the first eigenvector let $\lambda_1 = a - b$ so that

$$R - \lambda_1 I = \begin{bmatrix} a & b \\ b & a \end{bmatrix} - \begin{bmatrix} a-b & 0 \\ 0 & a-b \end{bmatrix} = \begin{bmatrix} b & b \\ b & b \end{bmatrix} = b \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

This has a null vector of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ which normalized is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

For the second eigenvector we have $\lambda_2 = a + b$ so that

$$R - \lambda_2 I = \begin{bmatrix} a & b \\ b & a \end{bmatrix} - \begin{bmatrix} a+b & 0 \\ 0 & a+b \end{bmatrix} = \begin{bmatrix} -b & b \\ b & -b \end{bmatrix} = b \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

This has a null vector of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ which normalized is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Chapter 4: Searching the Performance Surface

Notes on the Text

Notes on Gradient Search by Newton's Method

Using the drawing in the text note that the tangent line through the point $(w_0, f(w_0))$ will have a slope given by $f'(w_0)$. When we extend the tangent line until it intersects the x-axis (at a point $(w_1, 0)$) we can write this slope as $f'(w_0) = \tan(\theta)$ where θ is the angle from the x-axis to the tangent line. Now in the right triangle $\triangle ABC$ with $A = (w_1, 0)$, $B = (w_0, 0)$, and $C = (w_0, f(w_0))$ we can use the definition of the tangent to write

$$\tan(\theta) = f'(w_0) = \frac{f(w_0)}{w_0 - w_1}\,.$$

Solving this for w_1 gives

$$w_1 = w_0 - \frac{f(w_0)}{f'(w_0)}\,, \tag{8}$$

which is the expression in the book for one step of Newton's method.

Newton's Method for Optimization when the Error Surface is a Quadratic

When our error surface is quadratic i.e. $\xi(w) = \lambda(w - w^*)^2$ then the derivatives of $\xi(w)$ with respect to w are

$$\frac{d\xi}{dw} = 2\lambda(w - w^*)\,,$$

and

$$\frac{d^2\xi}{dw^2} = 2\lambda\,.$$

Now using Equation 8 one step of Newton's method with $f(w) = \xi'(w)$ on this quadratic surface is then

$$w_1 = w_0 - \frac{\xi'(w_0)}{\xi''(w_0)} = w_0 - \frac{2\lambda(w_0 - w^*)}{2\lambda} = w^*\,,$$

showing that we step immediately to the optimal solution from any initial guess.

Newton's Method in Multidimensional Space

Recall from Eq. 2.17 that we have the optimal multidimensional weights given by

$$W^* = R^{-1}P\,, \tag{9}$$

and the gradient of the error surface is given by Equation 1 or

$$\nabla = 2RW - 2P\,,$$

then multiplying this on the left by $\frac{1}{2}R^{-1}$ we get

$$\frac{1}{2}R^{-1}\nabla = W - R^{-1}P = W - W^* .$$

This means that

$$W^* = W - \frac{1}{2}R^{-1}\nabla , \tag{10}$$

which is the book's Eq. 4.30.

Gradient Search by the Method of Steepest Descent

The method of steepest decent is to update the weights W_k using

$$W_{k+1} = W_k - \mu\nabla_k , \tag{11}$$

with μ the stepsize constant and ∇_k given by

$$\nabla_k = 2RW_k - 2P .$$

Recalling Eq. 9 as $P = RR^{-1}P = RW^*$ we can write the weight update equation as

$$\begin{aligned} W_{k+1} &= W_k - \mu(2RW_k - 2P) = W_k - \mu(2RW_k - 2RW^*) \\ &= W_k + 2\mu R(W^* - W_k) = W_k - 2\mu RV_k , \end{aligned} \tag{12}$$

which is the book's Eq. 4.37. Grouping all terms involving W_k together we get

$$W_{k+1} = (I - 2\mu R)W_k + 2\mu RW^* , \tag{13}$$

which is the book's Eq. 4.38.

Recall that $V_k = W_k - W^*$ so $W_k = V_k + W^*$ and Equation 13 can be written as

$$\begin{aligned} V_{k+1} + W^* &= (I - 2\mu R)(V_k + W^*) + 2\mu RW^* \\ &= V_k + W^* - 2\mu RV_t - 2\mu RW^* + 2\mu RW^* , \end{aligned}$$

which simplifies to

$$V_{k+1} = (I - 2\mu R)V_k , \tag{14}$$

which is the book's Eq. 4.39.

As argued in the text for convergence we need to have

$$|1 - 2\mu\lambda_i| < 1 ,$$

for all i. We can write these inequalities as

$$-1 < 1 - 2\mu\lambda_i < +1 ,$$

or

$$-2 < -2\mu\lambda_i < 0 ,$$

or

$$0 < \mu\lambda_i < 1 ,$$

or

$$0 < \mu < \frac{1}{\lambda_i} ,$$

for all i. This means that

$$0 < \mu < \frac{1}{\lambda_{\max}} , \tag{15}$$

which is the book's Eq. 4.45. This is a condition on the stepsize parameter μ for convergence.

Starting with the vector "solution" to Equation 14 which is

$$V'_k = (I - 2\mu\Lambda)^k V'_0 , \tag{16}$$

if we multiply this on the left by Q we get

$$QV'_k = Q(I - 2\mu\Lambda)^k V'_0 . \tag{17}$$

Then recalling that $\Lambda = Q^{-1}RQ$ with $V' = Q^{-1}(W - W^*)$ so that the above becomes

$$W_k - W^* = Q(I - 2\mu\Lambda)^k Q^{-1}(W_0 - W^*) , \tag{18}$$

which is Eq. 4.50. Now using the relationship that $(QAQ^{-1})^k = QA^kQ^{-1}$ for the matrix A where $A = I - 2\mu\Lambda$ Equation 18 becomes

$$\begin{aligned} W_k &= W^* + (Q(I - 2\mu\Lambda)Q^{-1})^k (W_0 - W^*) \\ &= W^* + (I - 2\mu R)^k (W_0 - W^*) , \end{aligned} \tag{19}$$

which is the book's Eq. 4.52.

Comparison of Learning Curves

Now recall that the quadratic mean-square-error function is $\xi = \xi_{\min} + V^T RV$ and $V = W - W^*$ when we use the Newton update equation

$$W_k = W^* + (1 - 2\mu)^k (W_0 - W^*) , \tag{20}$$

we find that ξ_k is given by

$$\begin{aligned} \xi_k &= \xi_{\min} + (1 - 2\mu)^k (W_0 - W^*)^T R (1 - 2\mu)^k (W_0 - W^*) \\ &= \xi_{\min} + (1 - 2\mu)^{2k} V_0^T R V_0 , \end{aligned} \tag{21}$$

which is the book's Eq. 4.54 and represents the learning curve for Newton's method on a quadratic surface.

To derive the learning curve for steepest decent recall that we can write the error surface as $\xi = \xi_{\min} + V'^T \Lambda V'$ and under steepest decent we have Equation 16 or

$$V'_k = (I - 2\mu\Lambda)^k V'_0 ,$$

so that the error surface at each iteration ξ_k becomes

$$\xi_k = \xi_{\min} + V_0'^T (I - 2\mu\Lambda)^k \Lambda (I - 2\mu\Lambda)^k V_0' \tag{22}$$

$$= \xi_{\min} + V_0'^T (I - 2\mu\Lambda)^{2k} \Lambda V_0' \,, \tag{23}$$

using the fact that the product of diagonal matrices are commutative.

Exercise Solutions

Exercise 1

The example weight performance surface given in the book takes the form

$$\xi = \xi_{\min} + \lambda (w - w^*)^2 \,.$$

For the numbers given here this takes the form

$$\xi = 0 + 0.1(w - 2)^2 = 0.1(w - 2)^2 \,.$$

Exercise 2

Gradient search will update the weights according to

$$w_{k+1} = w_k - \mu \nabla_k \,.$$

With this performance surface we have

$$\nabla_k = \left. \frac{d\xi}{dw} \right|_{w_k} = 0.2(w_k - 2) \,.$$

Thus the updates take the form

$$w_{k+1} = w_k - 0.2\mu (w_k - 2) \,. \tag{24}$$

If $\mu = 4$ then we get

$$w_{k+1} = w_k - 0.8(w_k - 2) = 0.2 w_k + 1.6 \,.$$

Iterating this with $w_0 = 0$ we get

$$\begin{aligned} w_1 &= 0 + 1.6 = 1.6 \\ w_2 &= 0.2(1.6) + 1.6 = 1.92 \\ w_3 &= 0.2(1.92) + 1.6 = 1.984 \\ w_4 &= 0.2(1.984) + 1.6 = 1.9968 \\ w_5 &= 0.2(1.9968) + 1.6 = 1.99936 \,. \end{aligned}$$

Exercise 3

With $\mu = 8$ Equation 24 becomes

$$w_{k+1} = w_k - 0.2(8)(w_k - 2) = w_k = 1.6(w_k - 2) = -0.6w_k + 3.2\,.$$

Iterating this with $w_0 = 0$ gives

$$\begin{aligned} w_1 &= 3.2 \\ w_2 &= -0.6(3.2) + 3.2 = 1.28 \\ w_3 &= -0.6(1.28) + 3.2 = 2.432 \\ w_4 &= -0.6(2.432) + 3.2 = 1.7408 \\ w_5 &= -0.6(1.7408) + 3.2 = 2.15552\,. \end{aligned}$$

Exercise 4

Lets write this performance surface as

$$\begin{aligned} \xi &= 0.4w^2 + 4w + 11 = 0.4(w^2 + 10w) + 11 = 0.4[(w+5)^2 - 25] + 11 \\ &= 0.4(w+5)^2 + 1\,. \end{aligned}$$

Comparing this to

$$\xi = \xi_{\text{min}} + \lambda(w - w^*)^2\,, \tag{25}$$

we see that $\lambda = 0.4$. From the discussion in the book we recall that the overdamped region is given by

$$0 < \mu < \frac{1}{2\lambda}\,.$$

For this value of λ we find $\frac{1}{2\lambda} = 1.25$.

Exercise 5

Note that with a convergence parameter of $\mu = 1.5 > \frac{1}{2\lambda} = 1.25$ these iterations will be be "underdamped" and we expect the convergence of w_k to oscillate "around" w^*. From the book the error surface at each iteration for simple gradient search is given by

$$\xi_k = \xi_{\text{min}} + \lambda(w_0 - w^*)^2(1 - 2\mu\lambda)^{2k}\,. \tag{26}$$

For Exercise 4 we have $\xi_{\text{min}} = +1$, $\lambda = 0.4$, $w_0 = 0$, and $w^* = -5$ so the above becomes

$$\xi_k = 1 + 10(-0.2)^{2k} = 1 + 10(0.04^k)\,.$$

We can plot this with the following `R` code

```
ks = 1:10
xi_k = 1 + 10 * (0.04^ks)
plot(ks, xi_k, type='b')
grid()
```

This plot looks much like similar ones given presented in this section.

Exercise 6

Newton's formula for "finding" an optimum of the function $\xi(w)$ is given by

$$w_{k+1} = w_k - \frac{\xi'(w_k)}{\xi''(w_k)} . \tag{27}$$

To create a "discrete version" of this we need to approximate the derivatives above with differences. We will approximate the first derivative using

$$\xi'(w_k) = \frac{\xi(w_k) - \xi(w_{k-1})}{w_k - w_{k-1}} . \tag{28}$$

Using this we can approximate the second derivative using

$$\begin{aligned}
\xi''(w_k) &= \frac{\xi'(w_k) - \xi'(w_{k-1})}{w_k - w_{k-1}} \\
&= \frac{1}{w_k - w_{k-1}} \left[\left(\frac{\xi(w_k) - \xi(w_{k-1})}{w_k - w_{k-1}} \right) - \left(\frac{\xi(w_{k-1}) - \xi(w_{k-2})}{w_{k-1} - w_{k-2}} \right) \right] \\
&= \frac{(\xi(w_k) - \xi(w_{k-1}))(w_{k-1} - w_{k-2}) - (\xi(w_{k-1}) - \xi(w_{k-2}))(w_k - w_{k-1})}{(w_k - w_{k-1})^2 (w_{k-1} - w_{k-2})} .
\end{aligned}$$

We can then place these two approximations into Equation 27 to derive the "discrete" Newton's algorithm.

Exercise 7

The surface in Figure 4.5 is given by the formula

$$\xi = 1 - \frac{1}{26} [(1 - w^2)(4 + 3w)^2 + 1] .$$

Newton's method for finding an optima of $\xi(w)$ is given in Equation 27 or

$$w_{k+1} = w_k - \frac{\xi'(w_k)}{\xi''(w_k)} . \tag{29}$$

To evaluate the above derivatives we will write $\xi(w)$ above as

$$\begin{aligned}\xi(w) &= 1 - \frac{1}{26}[(1-w^2)(16+24w+9w^2)+1] \\ &= 1 - \frac{1}{26}[16+24w+9w^2-16w^2-24w^3-9w^4+1] \\ &= 1 - \frac{1}{26}[17+24w-7w^2-24w^3-9w^4]\,.\end{aligned}$$

Using this we have that

$$\begin{aligned}\xi'(w) &= -\frac{1}{26}[24-14w-72w^2-36w^3] \\ &= -\frac{1}{13}[12-7w-36w^2-18w^3]\,,\end{aligned}$$

and

$$\xi''(w) = -\frac{1}{13}(-7-72w-54w^2)\,.$$

With these derivatives Newton's method is $\xi(w)$ is given by

$$\begin{aligned}w_{k+1} &= w_k - \frac{12-7w_k-36w_k^2-18w_k^3}{-7-72w_k-54w_k^2} \\ &= w_k - \frac{18w_k^3+36w_k^2+7w_k-12}{54w_k^2+72w_k+7}\,.\end{aligned}$$

Notice that a root of the numerator is $w = -\frac{4}{3}$. This means that $w + \frac{4}{3}$ is a factor of the numerator. If we factor this out we can write Newton's method above as

$$w_{k+1} = w_k - \frac{(3w_k+4)(6w_k^2+4w_k-3)}{54w_k^2+72w_k+7}\,.$$

Exercise 8

This problem is worked in the `R` code `chap_4_exercise_8_N_9.R`.

Starting with $w_0 = 0$ (and using the above formula) the first seven weight values are given by

```
[1] "w_1= 1.714286"
[1] "w_2= 1.034719"
[1] "w_3= 0.649067"
[1] "w_4= 0.483907"
[1] "w_5= 0.449825"
[1] "w_6= 0.448405"
[1] "w_7= 0.448403"
```

Exercise 9

This problem is worked in the R code chap_4_exercise_8_N_9.R.

Here we print the second approximate optimum w_1, and the "final" approximate optimum when we start with various initial values w_0. Running the above R code gives

```
[1] "w0= +0.0000; w1= +1.7143; w_inf= +0.4484"
[1] "w0= -0.0800; w1= +7.7018; w_inf= +0.4484"
[1] "w0= -0.1400; w1= -6.2361; w_inf= -1.3335"
[1] "w0= -1.2000; w1= -0.9951; w_inf= -1.1151"
[1] "w0= -1.3000; w1= -1.3416; w_inf= -1.3333"
```

This shows the classic result that Newton's algorithm will converge to a given root if you start the iteration process "close enough". Starting farther away from the root will find a "different" one.

Exercise 10

Figure 3.2 is a plot of the quadratic MSE surface ξ given by

$$\xi = 42 + \begin{bmatrix} w_0 & w_1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} - 2 \begin{bmatrix} 7 & 8 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} . \tag{30}$$

Comparing this to Equation 9 we see that $R = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $P = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$ and $E[d_k^2] = 42$.

Now the book's Eq. 4.31 is given by

$$W_{k+1} = W_k - \frac{1}{2} R^{-1} \nabla_k \, , \tag{31}$$

and for a quadratic error surface we have

$$\nabla_k = 2RW_k - 2P \, . \tag{32}$$

Now for the numbers given in this exercise the inverse of R is given by

$$R^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} ,$$

and the gradient is given by

$$\nabla_k = 2 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - 2 \begin{bmatrix} 7 \\ 8 \end{bmatrix} .$$

Thus $\frac{1}{2}R^{-1}\nabla_k$ is given by

$$\frac{1}{2}R^{-1}\nabla_k = \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - R^{-1}\begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Thus Equation 31 becomes

$$W_{k+1} = \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \left(\begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

which is $\begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix}$. This means that from any location W_k in one iteration we find the optimum.

Exercise 11

From the book Eq. 4.32 (with a quadratic MSE surface) is given by

$$\begin{aligned} W_{k+1} &= W_k - \mu R^{-1}\nabla_k \quad \text{and} \\ \nabla_k &= 2RW_k - 2P\,. \end{aligned} \tag{33}$$

This exercise seeks to iterate this expression starting with $W_0 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ to find the value of W_{20}.

This problem is worked in the R code `chap_4_exercise_11.R`. When that problem is run we get the following

```
     0   1    2     3      4       5        20
[1,] 5 4.4 3.92 3.536 3.2288 2.98304 2.034588
[2,] 2 2.2 2.36 2.488 2.5904 2.67232 2.988471
```

Here each column is the value of W_k and the "index" on the column tells the value of k. I print the first five values of W_k and then W_{20}.

Exercise 12

Newton's method with a convergence parameter μ is given by Equation 33. For the variables given here we find

$$\begin{aligned} W_{k+1} &= W_k - \mu R^{-1}\nabla_k \\ &= \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \mu\begin{bmatrix} \rho_0 & \rho_1 \\ \rho_1 & \rho_0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} w_{0,k} \\ w_{1,k} \end{bmatrix} - \mu\begin{bmatrix} \rho_0\alpha + \rho_1\beta \\ \rho_1\alpha + \rho_0\beta \end{bmatrix} \\ &= \begin{bmatrix} w_{0,k} - \mu(\rho_0\alpha + \rho_1\beta) \\ w_{1,k} - \mu(\rho_1\alpha + \rho_0\beta) \end{bmatrix}. \end{aligned}$$

Next steepest decent with a convergence parameter μ in this case is given by

$$
\begin{aligned}
W_{k+1} &= W_k - \mu \nabla_k \\
&= \left[\begin{array}{c} w_0 \\ w_1 \end{array}\right] - \mu \left[\begin{array}{c} \alpha \\ \beta \end{array}\right] = \left[\begin{array}{c} w_0 - \mu\alpha \\ w_1 - \mu\beta \end{array}\right].
\end{aligned}
$$

Exercise 13

When applied to a quadratic MSE surface (where the derivative can be evaluated explicitly) the book's Eq. 4.34 (Newton's method) is given by

$$
W_{k+1} = (1 - 2\mu) W_k + 2\mu W^* , \quad (34)
$$

while the book's Eq. 4.38 (the steepest-descent method) is given by Equation 13 or

$$
W_{k+1} = (I - 2\mu R) W_k + 2\mu R W^* .
$$

For this problem we have $W^* = R^{-1}P = \left[\begin{array}{c} 2 \\ 3 \end{array}\right]$ when we compute. Then in terms of the components of W_k Newton's method Equation 34 is

$$
\left[\begin{array}{c} w_{0,k+1} \\ w_{1,k+1} \end{array}\right] = (1 - 2\mu) \left[\begin{array}{c} w_{0,k} \\ w_{1,k} \end{array}\right] + 2\mu \left[\begin{array}{c} 2 \\ 3 \end{array}\right].
$$

Notice that there is no "cross coupling" in these equations that is $w_{i,k}$ and $w_{j,k}$ are not dependent on each other if $i \neq j$.

In terms of the components of W_k the steepest-decent algorithm Equation 13 is

$$
\begin{aligned}
\left[\begin{array}{c} w_{0,k+1} \\ w_{1,k+1} \end{array}\right] &= \left(\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] - 2\mu \left[\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}\right] \right) \left[\begin{array}{c} w_{0,k} \\ w_{1,k} \end{array}\right] + 2\mu \left[\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}\right] \left[\begin{array}{c} 2 \\ 3 \end{array}\right] \\
&= \left[\begin{array}{cc} 1 - 4\mu & -2\mu \\ -2\mu & 1 - 4\mu \end{array}\right] \left[\begin{array}{c} w_{0,k} \\ w_{1,k} \end{array}\right] + 2\mu \left[\begin{array}{c} 7 \\ 7 \end{array}\right].
\end{aligned}
$$

This are the two equations

$$
\begin{aligned}
w_{0,k+1} &= (1 - 4\mu) w_{0,k} - 2\mu w_{1,k} + 14\mu \\
w_{1,k+1} &= -2\mu w_{0,k} + (1 - 4\mu) w_{1,k} + 14\mu .
\end{aligned}
$$

These equations have "cross coupling" in that the equation for $w_{i,k}$ depends on $w_{j,k}$ for $j \neq i$.

Exercise 14

The steepest-decent algorithm is given by Equation 11 or

$$
W_{k+1} = W_k - \mu \nabla_k ,
$$

and when our MSE surface is quadratic we have the gradient ∇_k given by

$$\nabla_k = 2RW_k - 2P\,. \tag{35}$$

We seek to iterate the above starting with $W_0 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$.

This problem is worked in the `R` code `chap_4_exercise_14.R`. When that problem is run we get the following

```
     0   1    2     3      4       5        20
[1,] 5 4.0 3.44 3.088 2.8448 2.66560 2.023058
[2,] 2 1.8 1.88 2.040 2.2064 2.35488 2.976942
```

Here each column is the value of W_k and the "index" on the column tells the value of k. I print the first five values of W_k and then W_{20}.

Exercise 15

Numerically this problem is worked in the `R` code `chap_4_exercise_15.R`. Running that code we plot the learning curve for this problem.

Here we derive the analytic representation of the learning curve for Newton's method. Recall that the book's Eq. 3.44 is given by Equation 30 which we can write as

$$\xi = \xi_{\text{min}} + V^T RV\,,$$

were $V = W - W^* = W - R^{-1}P$ and

$$\xi_{\text{min}} = E[d_k^2] - P^T W^* = E[d_k^2] - P^T R^{-1} P\,.$$

From the text the learning curve at each iteration W_k takes the form given in Equation 21. For this problem we can evaluate the given expressions above. We find

$$\begin{aligned}
W^* = R^{-1}P &= \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\
\xi_{\text{min}} = 42 - \begin{bmatrix} 7 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} &= 4 \\
V_0 = W_0 - W^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} &= \begin{bmatrix} -2 \\ -3 \end{bmatrix} \\
V_0^T RV_0 = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} &= 38\,.
\end{aligned}$$

Thus Equation 21 becomes

$$\xi_k = 4 + (1 - 0.1)^{2k}(38) = 4 + 38(0.9)^{2k}\,.$$

Exercise 16

Newton's method is the derivation of the book's Equation 4.54 which is presented in the above notes ending at Equation 21. The steepest-descent method is the derivation of the book's Equation 4.58 which is presented in the above notes ending at Equation 23.

Exercise 17

Here we derive the analytic representation of the learning curve for the steepest-decent method. Recall that the steepest-decent method has a learning-curve that looks like Equation 23. To "use" this expression we first note that we have calculated $\xi_{\rm min} = 4$ (in a previous problem), that $V_0' = Q^{-1}(W_0 - W^*)$, and $R = Q\Lambda Q^{-1}$. For this matrix R (and given in the book) we have

$$\Lambda = \left[\begin{array}{cc} 1 & 0 \\ 0 & 3 \end{array}\right] \quad \text{with} \quad Q = \frac{1}{\sqrt{2}}\left[\begin{array}{cc} 1 & 1 \\ -1 & 1 \end{array}\right] \quad \text{and} \quad Q^{-1} = Q^T = \frac{1}{\sqrt{2}}\left[\begin{array}{cc} 1 & -1 \\ 1 & 1 \end{array}\right].$$

We are told to take $W_0 = \left[\begin{array}{c} 0 \\ 0 \end{array}\right]$ and $W^* = \left[\begin{array}{c} 2 \\ 3 \end{array}\right]$. Using these we find

$$V_0' = \frac{1}{\sqrt{2}}\left[\begin{array}{cc} 1 & -1 \\ 1 & 1 \end{array}\right]\left[\begin{array}{c} -2 \\ -3 \end{array}\right] = \frac{1}{\sqrt{2}}\left[\begin{array}{c} 1 \\ -5 \end{array}\right],$$

and

$$I - 2\mu\Lambda = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] - 0.1\left[\begin{array}{cc} 1 & 0 \\ 0 & 3 \end{array}\right] = \left[\begin{array}{cc} 0.9 & 0 \\ 0 & 0.7 \end{array}\right].$$

This means that

$$(I - 2\mu\Lambda)^{2k} = \left[\begin{array}{cc} 0.9^{2k} & 0 \\ 0 & 0.7^{2k} \end{array}\right],$$

so that

$$(I - 2\mu\Lambda)^{2k}\Lambda = \left[\begin{array}{cc} 0.9^{2k} & 0 \\ 0 & 0.7^{2k} \end{array}\right]\left[\begin{array}{cc} 1 & 0 \\ 0 & 3 \end{array}\right] = \left[\begin{array}{cc} 0.9^{2k} & 0 \\ 0 & 3(0.7)^{2k} \end{array}\right],$$

and

$$\begin{aligned} V_0'^T(I - 2\mu\Lambda)^{2k}\Lambda V_0' &= \frac{1}{\sqrt{2}}\left[\begin{array}{cc} 1 & -5 \end{array}\right]\left[\begin{array}{cc} 0.9^{2k} & 0 \\ 0 & 3(0.7)^{2k} \end{array}\right]\left[\begin{array}{c} 1 \\ -5 \end{array}\right]\frac{1}{\sqrt{2}} \\ &= \frac{1}{2}(0.9^{2k} + 75(0.7^{2k}))\,. \end{aligned}$$

Adding $\xi_{\rm min}$ to this gives Equation 23 or the desired expression for ξ_k.

Exercise 18

Now $(I - 2\mu\Lambda)^{2k}\Lambda$ is a diagonal matrix with diagonal elements given by

$$(1 - 2\mu\lambda_n)^{2k}\lambda_n \quad \text{for} \quad 0 \le n \le L\,.$$

The scalar $V_0'^T(I - 2\mu\Lambda)^{2k}\Lambda V_0'$ is then the sum of each of the diagonal elements above multiplied by ${v_{0n}'}^2$ which is the book's Eq. 4.59.

Exercise 19

Now W_k is dimensionless while the dimensions of MSE is signal power. This means that ∇_k also has dimensions of signal power. Then in gradient-decent given by Equation 11 we see that μ must have units of reciprocal signal power.

As $R = E[X_k X_k^T]$ we see that it has dimensions of signal power, so R^{-1} has dimensions of reciprocal signal power. This means that $R^{-1}\nabla_k$ is dimensionless so in Newton's method given by Equation 33 we see that μ must be dimensionless.

Chapter 5: Gradient Estimation and Its Effects on Adaptation

Notes on the Text

Notes on the Variance of the Gradient Estimate

In computing α_r if r is even then we have

$$\begin{aligned}
\alpha_r &= \frac{1}{2\sigma\sqrt{3}}\left(\frac{\epsilon^{r+1}}{r+1}\right|_{-\sigma\sqrt{3}}^{\sigma\sqrt{3}} = \frac{1}{2\sigma\sqrt{3}(r+1)}\left(\sigma^{r+1}+\sigma^{r+1}\right)(\sqrt{3})^{r+1} \\
&= \frac{(\sqrt{3})^r\sigma^r}{r+1} = \frac{\sigma^r 3^{r/2}}{r+1}\,.
\end{aligned}$$

Then $\text{Var}\left(\hat{\xi}\right)$ is given by

$$\text{Var}\left(\hat{\xi}\right) = \frac{1}{N}\left[\frac{\sigma^4 3^2}{5} - \left(\frac{\sigma^2 3}{3}\right)^2\right] = \frac{\sigma^4}{N}\left[\frac{9}{5}-\frac{9}{3}\right] = \frac{4\sigma^4}{5N}\,,$$

when we simplify.

Notes on the Derivation of $\text{cov}\left[V_k'\right]$

Here I work thought the algebra needed to derive the results in this section. We start with

$$\text{cov}\left[V_k'\right] = E[V_k' V_k'^T]\,,$$

then using the books Eq. 5.39 or

$$V_k' = (1-2\mu)V_{k-1}' - \mu\Lambda^{-1}N_{k-1}'\,, \tag{36}$$

the above is

$$\begin{aligned}
V_k' V_k'^T &= [(1-2\mu)V_{k-1}' - \mu\Lambda^{-1}N_{k-1}'][(1-2\mu)V_{k-1}'^T - \mu N_{k-1}'^T(\Lambda^{-1})^T] \\
&= (1-2\mu)^2 V_{k-1}'V_{k-1}'^T + \mu^2\Lambda^{-1}N_{k-1}'N_{k-1}'^T(\Lambda^{-1})^T \\
&- \mu(1-2\mu)\Lambda^{-1}N_{k-1}'V_{k-1}'^T - \mu(1-2\mu)\Lambda^{-1}V_{k-1}'N_{k-1}'^T(\Lambda^{-1})^T\,,
\end{aligned} \tag{37}$$

which is the books Eq. 5.49. Taking the expectation of this we get

$$\begin{aligned}
\text{cov}\left[V_k'\right] &= (1-2\mu)^2\text{cov}\left[V_k'\right] + \mu^2 E[(\Lambda^{-1}N_{k-1}')(\Lambda^{-1}N_{k-1}')^T] \\
&= (1-2\mu)^2\text{cov}\left[V_k'\right] + \mu^2\Lambda^{-1}\text{cov}\left[N_{k-1}'\right]\Lambda^{-1} \\
&= (1-2\mu)^2\text{cov}\left[V_k'\right] + \mu^2(\Lambda^{-1})^2\text{cov}\left[N_{k-1}'\right]\,.
\end{aligned}$$

Here we have used the fact that both Λ^{-1} and $\text{cov}\left[N'_{k-1}\right]$ are diagonal matrices and thus commute. Expanding the argument in front of $\text{cov}\left[V'_k\right]$ we get

$$\text{cov}\left[V'_k\right] = (1 - 4\mu + 4\mu^2)\text{cov}\left[V'_{k-1}\right] + \mu^2(\Lambda^{-1})^2\text{cov}\left[N'_{k-1}\right] .$$

On canceling $\text{cov}\left[V'_k\right]$ from both sides this can be written as

$$(4\mu - 4\mu^2)\text{cov}\left[V'_{k-1}\right] = \mu^2(\Lambda^{-1})^2\text{cov}\left[N'_{k-1}\right] .$$

Solving this for $\text{cov}\left[V'_{k-1}\right]$ we get

$$\text{cov}\left[V'_{k-1}\right] = \frac{\mu(\Lambda^{-1})^2}{4(1-\mu)}\text{cov}\left[N'_{k-1}\right] , \tag{38}$$

which is the books Eq. 5.50.

To derive the similar result in the steepest decent case we recall the books Eq. 5.45 or

$$V'_k = (I - 2\mu\Lambda)V'_{k-1} - \mu N'_{k-1} , \tag{39}$$

and put this in the above expression for $V'_k V'^T_k$ to get

$$\begin{aligned} V'_k V'^T_k &= (I - 2\mu\Lambda)^2 V'_{k-1} V'^T_{k-1} (I - 2\mu\Lambda)^T - \mu(I - 2\mu\Lambda)V'_{k-1}N'^T_{k-1} \\ &- \mu N'_{k-1}(I - 2\mu\Lambda)^T + \mu^2 N'_{k-1}N'^T_{k-1} , \end{aligned} \tag{40}$$

which is the books Eq. 5.51. Taking the expectation of this we get

$$\text{cov}\left[V'_k\right] = (I - 2\mu\Lambda)^2\text{cov}\left[V'_k\right] + \mu^2\text{cov}\left[N'_k\right] .$$

If we expand the first term on the right-hand-side of the above we get

$$\text{cov}\left[V'_k\right] = \text{cov}\left[V'_k\right] - 4\mu\Lambda\text{cov}\left[V'_k\right] + 4\mu^2\Lambda^2\text{cov}\left[V'_k\right] + \mu^2\text{cov}\left[N'_k\right] .$$

Solving for $\text{cov}\left[V'_k\right]$ we get

$$\text{cov}\left[V'_k\right] = \frac{\mu}{4}(\Lambda - \mu\Lambda^2)^{-1}\text{cov}\left[N'_k\right] , \tag{41}$$

which is the books Eq. 5.52.

Now recalling that $N' = Q^{-1}N$ (so that $N = QN'$) and that $\hat{\nabla}_k = \nabla_k + N_k$ we have

$$N'_k = Q^{-1}N_k = Q^{-1}(\hat{\nabla}_k - \nabla_k) ,$$

so that

$$\begin{aligned} \text{cov}\left[N'_k\right] &= Q^{-1}E[(\hat{\nabla}_k - \nabla_k)(\hat{\nabla}_k - \nabla_k)^T]Q \\ &= Q^{-1}\text{cov}\left[\hat{\nabla}_k\right]Q = \frac{\xi^2_{\min}}{N\delta^2}I , \end{aligned} \tag{42}$$

which is the books Eq. 5.53.

Using Equation 42 in Equation 38 (Newton's method) we have that

$$\text{cov}\,[V_k'] = \frac{\mu(\Lambda^{-1})^2}{4(1-\mu)} \frac{\xi_{\text{min}}^2}{N\delta^2}\,, \tag{43}$$

which is the books Eq. 5.54.

Using Equation 42 in Equation 41 (the steepest decent method) we have that

$$\text{cov}\,[V_k'] = \frac{\mu}{4}(\Lambda - \mu\Lambda^2)^{-1} \frac{\xi_{\text{min}}^2}{N\delta^2}\,, \tag{44}$$

which is the books Eq. 5.55.

To express these in terms of the unprimed coordinate system we recall that

$$\text{cov}\,[V_k] = E[V_k V_k^T] = QE[V_k' V_k'^T]Q^{-1} = Q\text{cov}\,[V_k']\,Q^{-1}\,, \tag{45}$$

which is the books Eq. 5.56. This equation allows one to derive the books Eq. 5.57 and Eq. 5.58.

Lets now evaluate the excess MSE from $E[V_k'\Lambda V_k']$ for Newton's method where

$$V_k' = -\mu\Lambda^{-1}\sum_{n=0}^{\infty}(1-2\mu)^n N_{k-n-1}'\,, \tag{46}$$

to get

$$\begin{aligned}
\text{excess MSE} &= E[V_k'\Lambda V_k] \\
&= \mu^2 E\left[\left(\sum_{n=0}^{\infty} r^n N_{k-n-1}'^T \Lambda^{-1}\right)\Lambda\left(\sum_{m=0}^{\infty} r^m N_{k-m-1}'^T \Lambda^{-1}\right)\right] \\
&= \mu^2 \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} r^{n+m} E[N_{k-n-1}'^T \Lambda^{-1} N_{k-m-1}']\,.
\end{aligned}$$

As only the terms with $n = m$ don't evaluate to zero the above becomes

$$\text{excess MSE} = \mu^2 \sum_{n=0}^{\infty} r^{2n} E[N_{k-n-1}'^T \Lambda^{-1} N_{k-n-1}']\,. \tag{47}$$

The rest of the steps are explained in the book.

To evaluate the excess MSE from $E[V_k'\Lambda V_k']$ for the method of steepest decent where

$$V_k' = -\mu\sum_{n=0}^{\infty}(I-2\mu\Lambda)^n N_{k-n-1}'\,, \tag{48}$$

to get

$$\begin{aligned}
\text{excess MSE} &= E[V_k'\Lambda V_k] \\
&= \mu^2 E\left[\sum_{n=0}^{\infty} N_{k-n-1}'^T[(I-2\mu\Lambda)^T]^n \times \Lambda \times \sum_{m=0}^{\infty}(I-2\mu\Lambda)^m N_{k-m-1}'\right] \\
&= \mu^2 \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} E[N_{k-n-1}'^T \Lambda (I-2\mu\Lambda)^{n+m} N_{k-m-1}']\,.
\end{aligned} \tag{49}$$

Again as only the terms with $n = m$ don't evaluate to zero the above becomes

$$\begin{aligned} \text{excess MSE} &= \mu^2 \sum_{n=0}^{\infty} E[N_k'^T \Lambda (I - 2\mu\Lambda)^{2n} N_k'] \\ &= \mu^2 E\left[N_k'^T \Lambda \left(\sum_{n=0}^{\infty} (I - 2\mu\Lambda)^{2n} \right) N_k' \right] . \end{aligned} \quad (50)$$

Now using $\sum_{n=0}^{\infty} D^n = (I - D)^{-1}$ we have

$$\begin{aligned} \sum_{n=0}^{\infty} (I - 2\mu\Lambda)^{2n} &= \sum_{n=0}^{\infty} [(I - 2\mu\Lambda)^2]^n = [I - (I - 2\mu\Lambda)^2]^{-1} \\ &= [I - (I - 4\mu\Lambda + 4\mu^2\Lambda^2)]^{-1} \\ &= (4\mu\Lambda - 4\mu^2\Lambda^2)^{-1} \\ &= \frac{1}{4\mu}(\Lambda - \mu\Lambda^2)^{-1} , \end{aligned} \quad (51)$$

which is the books Eq. 5.78. The rest of the steps are explained in the book.

Exercise Solutions

Exercise 1

When the performance surface ξ is a quadratic i.e. when it takes the form $\xi(v) = \xi_{\min} + \lambda v^2$ then the first and second derivatives are

$$\begin{aligned} \xi'(v) &= 2\lambda v \\ \xi''(v) &= 2\lambda . \end{aligned}$$

Notice that when using the central difference formulas from the book on a quadratic surface like $\xi(v)$ we get the *exact* derivatives above

$$\begin{aligned} \frac{d\xi}{dv} &\approx \frac{\xi(v+\delta) - \xi(v-\delta)}{2\delta} = 2\lambda v \\ \frac{d^2\xi}{dv^2} &\approx \frac{\xi(v+\delta) + 2\xi(v) - \xi(v-\delta)}{\delta^2} = 2\lambda . \end{aligned}$$

These calculations are done in this chapter.

Exercise 2

We can write this $\xi(w)$ as

$$\begin{aligned} \xi(v) &= 5(w^2 - 4w) + 23 \\ &= 5(w^2 - 4w + 4) - 20 + 23 = 5(w-2)^2 + 3 . \end{aligned}$$

Thus we see that $\xi_{\min} = 3$, $w^* = 2$, and $\lambda = 5$.

If we now compute the performance penalty γ using

$$\gamma = \lambda \delta^2 \,, \tag{52}$$

we find $\gamma = 5$.

Exercise 3

As the performance penalty can be written as $\gamma = \lambda \delta^2$ and $\lambda > 0$ this function is convex up. If $\gamma < 0$ then $\lambda < 0$ and $\xi(v) = \xi_{\min} + \lambda v^2$ points "downwards" which is not possible for the type of error surfaces we are discussing in this book.

Exercise 4

From the text the perturbation "P" is defined as

$$P = \frac{\gamma}{\xi_{\min}} = \frac{\lambda \delta^2}{\xi_{\min}} \,. \tag{53}$$

For the numbers in Exercise 2 this means that $P = \frac{5(1^2)}{3} = \frac{5}{3}$.

Exercise 5

We can write this in the standard form as

$$\begin{aligned} \xi &= 2w_0^2 + 2w_1^2 + 2w_0 w_1 - 14w_0 - 16w_1 + 42 \\ &= \left[\begin{array}{cc} w_0 & w_1 \end{array}\right] \left[\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}\right] \left[\begin{array}{c} w_0 \\ w_1 \end{array}\right] - 2 \left[\begin{array}{cc} 7 & 8 \end{array}\right] \left[\begin{array}{c} w_0 \\ w_1 \end{array}\right] + 42 \,, \end{aligned}$$

so that $L = 1$, $R = \left[\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}\right]$ and $P = \left[\begin{array}{c} 7 \\ 8 \end{array}\right]$. From other problems with this covariance matrix R we can show that R has eigenvalues of 1 and 3. For these two values for the eigenvalues we compute

$$\begin{aligned} \lambda_{\text{av}} &= \frac{1}{2}(1+3) = 2 \\ \left(\frac{1}{\lambda}\right)_{\text{av}} &= \frac{1}{2}\left(\frac{1}{1} + \frac{1}{3}\right) = \frac{2}{3} \,. \end{aligned}$$

From the section entitled "A Second Example" in Chapter 3 we can show that for this optimization problem that $\xi_{\min} = 4$.

Then the perturbation P is given by

$$P = \frac{\delta^2 \lambda_{\text{av}}}{\xi_{\min}} . \tag{54}$$

Using the numbers above we find $P = \frac{2\delta^2}{4} = \frac{\delta^2}{2}$.

Exercise 6

Adding another weight will increase the dimension of R and that might change the value of the average eigenvalue or λ_{av} of R. Equation 54 then shows how that would change P. If the trace of R does not change with this additional weight then the average eigenvalue will get smaller so P will get smaller.

Exercise 7

If we have that $\varepsilon_k \sim U(1,3)$ then

$$\alpha_4 = \int {\varepsilon_k}^4 p(\varepsilon_k) d\varepsilon_k = \int_1^3 {\varepsilon_k}^4 \left(\frac{1}{3-1}\right) d\varepsilon_k = \frac{1}{2}\left(\frac{{\varepsilon_k}^5}{5}\right|_1^3 = \frac{242}{10} = 24.2 .$$

Exercise 8

The mean and variance of $\varepsilon_k \sim U(1,3)$ are given by

$$\begin{aligned} \alpha_1 &= E[\varepsilon_k] = 2 \\ \sigma^2 &= \frac{(3-1)^2}{12} = \frac{4}{12} = \frac{1}{3} . \end{aligned}$$

Based on these numbers for this exercise we are to assume that $\varepsilon_k \sim \mathcal{N}\left(2, \frac{1}{3}\right)$.

Recall that the fourth moment of a normal random variable $\mathcal{N}(\mu, \sigma^2)$ is

$$E[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 .$$

Thus for the case considered here we would find

$$\alpha_4 = E[\varepsilon_k^4] = 2^4 + 6(2^2)\left(\frac{1}{3}\right) + 3\left(\frac{1}{9}\right) = \frac{73}{3} = 24.3333 .$$

Exercise 9

Here we are told that $\varepsilon \sim \mathcal{N}(0, 3)$. From the text we have

$$\text{Var}\left(\hat{\xi}\right) = \frac{\alpha_4 - \alpha_2^2}{N} = \frac{3\sigma_\varepsilon^4 - \sigma_\varepsilon^4}{N} = \frac{2\sigma_\varepsilon^4}{N}\,. \quad (55)$$

For the numbers given in this problem we have

$$\text{Var}\left(\hat{\xi}\right) = \frac{2\sigma_\varepsilon^4}{10} = \frac{\sigma_\varepsilon^4}{5} = \frac{3^2}{5} = \frac{9}{5}\,.$$

Exercise 10

Recall that the **non-central** moments of a random variable ε are defined as

$$\begin{aligned}
\alpha_1 &= E[\varepsilon] \\
\alpha_2 &= E[\varepsilon^2] \\
\alpha_3 &= E[\varepsilon^3] \\
&\vdots \\
\alpha_p &= E[\varepsilon^p]\,.
\end{aligned}$$

Using these we can write the variance σ^2 as

$$\sigma^2 = E[(\varepsilon - \alpha_1)^2] = E[\varepsilon^2] - \alpha_1^2 = \alpha_2 - \alpha_1^2\,.$$

Next recall that Eq. 5.25 gives

$$\text{Var}\left(\hat{\xi}\right) = \text{Var}\left(\hat{\alpha}_2\right) = \frac{\alpha_4 - \alpha_2^2}{N}\,. \quad (56)$$

If we have that $\varepsilon_k \sim \mathcal{N}(\alpha_1, \sigma^2)$ then we can expand the higher moment expressions above in terms of lower moments using (from a text on statistics)

$$\begin{aligned}
\alpha_4 &= \alpha_1^4 + 6\alpha_1^2\sigma^2 + 3\sigma^4 \\
\alpha_2 &= \alpha_1^2 + \sigma^2\,,
\end{aligned}$$

Then using these in Equation 56 I compute

$$\begin{aligned}
\text{Var}\left(\hat{\xi}\right) &= \frac{\alpha_1^4 + 6\alpha_1^2\sigma^2 + 3\sigma^4 - (\alpha_1^2 + \sigma^2)^2}{N} = \frac{4\alpha_1^2\sigma^2 + 2\sigma^4}{N} = \frac{2(2\alpha_1^2\sigma^2 + \sigma^4)}{N} \\
&= \left(\frac{2\alpha_1^2\sigma^2 + \sigma^4}{\xi^2}\right)\left(\frac{2\xi^2}{N}\right)\,.
\end{aligned}$$

Now when ε_k has $\alpha_1 \neq 0$ then

$$\xi = E[\varepsilon_k^2] = \alpha_1^2 + \sigma^2 \quad \text{so} \quad \xi^2 = (\alpha_1^2 + \sigma^2)^2\,, \quad (57)$$

Thus we have

$$\text{Var}\left(\hat{\xi}\right) = \left(\frac{2\alpha_1^2\sigma^2 + \sigma^4}{(\alpha_1^2+\sigma^2)^2}\right)\left(\frac{2\xi^2}{N}\right) = \frac{1+2\left(\frac{\alpha_1^2}{\sigma^2}\right)}{\left(1+\frac{\alpha_1^2}{\sigma^2}\right)^2}\left(\frac{2\xi^2}{N}\right).$$

Setting this equal to $\frac{K\xi^2}{N}$ we see that

$$\frac{K}{2} = \frac{1+2\left(\frac{\alpha_1^2}{\sigma^2}\right)}{\left(1+\frac{\alpha_1^2}{\sigma^2}\right)^2} < \frac{1+2\left(\frac{\alpha_1^2}{\sigma^2}\right)+\left(\frac{\alpha_1^2}{\sigma^2}\right)^2}{\left(1+\frac{\alpha_1^2}{\sigma^2}\right)^2} = \frac{\left(1+\frac{\alpha_1^2}{\sigma^2}\right)^2}{\left(1+\frac{\alpha_1^2}{\sigma^2}\right)^2} = 1\,.$$

Thus we have that $K < 2$.

Exercise 11

Part (a): This derived from Part (d) of this exercise when we take $\alpha_1 = 0$.

Part (b): This derived from Part (f) of this exercise when we take $\alpha_1 = 0$.

Part (c): This is done in the previous exercise above.

Part (d): From the diagram the density $p(\varepsilon)$ takes the form

$$p(\varepsilon) = \begin{cases} \frac{1}{6\sigma^2}(x-\alpha_1+\sigma\sqrt{6}) & \alpha_1-\sigma\sqrt{6} < \varepsilon < \alpha_1 \\ \frac{1}{\sigma\sqrt{6}} - \frac{1}{6\sigma^2}(x-\alpha_1) & \alpha_1 < \varepsilon < \alpha_1+\sigma\sqrt{6} \end{cases}$$

This means that

$$E[\varepsilon^2] = \int_{\alpha_1-\sigma\sqrt{6}}^{\alpha_1} \varepsilon^2 p(\varepsilon)d\varepsilon + \int_{\alpha_1}^{\alpha_1+\sigma\sqrt{6}} \varepsilon^2 p(\varepsilon)d\varepsilon$$

The expression for $E[\varepsilon^4]$ is computed in the same way. These integrals are tedious to evaluate. Using Mathematica (after factoring out $6\sigma^2$) I find

$$\begin{aligned} E[\varepsilon^2] &= \frac{6\alpha_1^2\sigma^2 + 6\sigma^4}{6\sigma^2} = \alpha_1^2 + \sigma^2 \\ E[\varepsilon^4] &= \frac{1}{6\sigma^2}\left(\frac{6}{5}\sigma^2(5\alpha_1^4 + 30\alpha_1^2\sigma^2 + 12\sigma^4\right) \\ &= \alpha_1^4 + 6\alpha_1^2\sigma^2 + \frac{12}{5}\sigma^4\,. \end{aligned}$$

This means that

$$E[\varepsilon^4] - E[\varepsilon^2]^2 = 4\alpha_1^2\sigma^2 + \frac{7}{5}\sigma^4\,,$$

when we simplify. Next using Equation 56 and 57 we compute

$$\text{Var}\left(\hat{\xi}\right) = \left(\frac{\xi^2}{N}\right)\left(\frac{4\alpha_1^2\sigma^2 + \frac{7}{5}\sigma^4}{(\alpha_1^2+\sigma^2)^2}\right) = \frac{\xi^2}{5N}\left(\frac{7+20\alpha_1^2/\sigma^2}{(1+\alpha_1^2/\sigma^2)^2}\right)\,. \quad (58)$$

Part (e): For this density we find

$$E[\varepsilon^2] = \frac{1}{2\sigma\sqrt{3}} \int_{\alpha_1-\sigma\sqrt{3}}^{\alpha_1+\sigma\sqrt{3}} \varepsilon^2 d\varepsilon = \alpha_1^2 + \sigma^2$$

$$E[\varepsilon^4] = \frac{1}{2\sigma\sqrt{3}} \int_{\alpha_1-\sigma\sqrt{3}}^{\alpha_1+\sigma\sqrt{3}} \varepsilon^4 d\varepsilon = \frac{1}{5}(5\alpha_1^4 + 30\alpha_1^2\sigma^2 + 9\sigma^4)\,.$$

This means that

$$E[\varepsilon^4] - E[\varepsilon^2]^2 = 4\alpha_1^2\sigma^2 + \frac{4}{5}\sigma^4\,,$$

when we simplify. Next using Equation 56 and 57 we compute

$$\mathrm{Var}\left(\hat{\xi}\right) = \left(\frac{4\xi^2}{N}\right)\left(\frac{\alpha_1^2\sigma^2 + \frac{1}{5}\sigma^4}{(\alpha_1^2+\sigma^2)^2}\right) = \frac{4\xi^2}{5N}\left(\frac{1+5\alpha_1^2/\sigma^2}{(1+\alpha_1^2/\sigma^2)^2}\right)\,. \tag{59}$$

Part (f): For this density we find that

$$E[\varepsilon^2] = \frac{1}{2}(\alpha_1-\sigma)^2 + \frac{1}{2}(\alpha_1+\sigma)^2$$

$$E[\varepsilon^4] = \frac{1}{2}(\alpha_1-\sigma)^4 + \frac{1}{2}(\alpha_1+\sigma)^4\,.$$

Then using these in Equation 56 we get

$$\begin{aligned}\mathrm{Var}\left(\hat{\xi}\right) &= \frac{1}{4N}\left[2(\alpha_1-\sigma)^4 + 2(\alpha_1+\sigma)^4 - \left((\alpha_1-\sigma)^2 + (\alpha_1+\sigma)^2\right)^2\right] \\ &= \frac{1}{4N}(16\alpha_1^2\sigma^2) = \frac{4}{N}\alpha_1^2\sigma^2\,.\end{aligned}$$

Using Equation 57 we can write this as

$$\begin{aligned}\mathrm{Var}\left(\hat{\xi}\right) &= \frac{\xi^2}{N}\left(\frac{4\alpha_1^2\sigma^2}{\xi^2}\right) = \frac{\xi^2}{N}\left(\frac{4\alpha_1^2\sigma^2}{(\alpha_1^2+\sigma^2)^2}\right) \\ &= \frac{\xi^2}{N}\left(\frac{4\alpha_1^2/\sigma^2}{\left(1+\alpha_1^2/\sigma^2\right)^2}\right)\,,\end{aligned} \tag{60}$$

which is the expression desired.

Part (g): In Equation 58 if we define $r \equiv \alpha_1^2/\sigma^2$ then we see that $r > 0$ and in this case when $\mathrm{Var}\left(\hat{\xi}\right) = \frac{K\xi^2}{N}$ that K is given by

$$K(r) = \frac{7+20r}{5(1+r)^2}\,.$$

One optimal value for $K(r)$ could be when $r = 0$ where $K(0) = \frac{7}{5} = 1.4$. Using calculus to look for other extreme values for $K(r)$ we need to solve

$$K'(r) = \frac{20}{5(1+r)^2} - \frac{2(7+20r)}{5(1+r)^3} = 0\,,$$

which has $r = \frac{3}{10}$. The value of K at this r is

$$K\left(\frac{3}{10}\right) = \frac{20}{13} = 1.53846\,.$$

For this value of r we find

$$\frac{\alpha_1}{\sigma} = \sqrt{\frac{3}{10}} = 0.547723\,.$$

Part (h): In Equation 59 if we define $r \equiv \alpha_1^2/\sigma^2$ then we see that $r > 0$ and in this case when $\text{Var}\left(\hat{\xi}\right) = \frac{K\xi^2}{N}$ that K is given by

$$K(r) = \frac{4 + 20r}{5(1 + r)^2}\,.$$

One optimal value for $K(r)$ could be when $r = 0$ where $K(0) = \frac{4}{5} = 0.8$. Using calculus to look for other extreme values for $K(r)$ we need to solve

$$K'(r) = \frac{20}{5(1 + r)^2} - \frac{2(4 + 20r)}{5(1 + r)^3} = 0\,,$$

which has $r = \frac{3}{5}$. The value of K at this r is

$$K\left(\frac{3}{5}\right) = \frac{5}{4} = 1.25\,.$$

For this value of r we find

$$\frac{\alpha_1}{\sigma} = \sqrt{\frac{5}{4}} = 1.11803\,.$$

Exercise 12

If $K = 0$ then we must have

$$\frac{\alpha_1^2}{\sigma^2} \to \infty \quad \text{so} \quad \left|\frac{\alpha_1}{\sigma}\right| \to \infty\,.$$

If $K = 2$ then defining $r \equiv \frac{\alpha_1^2}{\sigma^2}$ we must have

$$\frac{2 + 4r}{(1 + r)^2} = 2\,.$$

This has the solution $r = 0$ or equivalently $\left|\frac{\alpha_1}{\sigma}\right| \to 0$.

Exercise 13

Part (a): This is worked in Part (g) in Exercise 11 above.

Part (b): This is worked in Part (h) in Exercise 11 above.

Part (c): In Equation 60 if we define $r \equiv \alpha_1^2/\sigma^2$ then we see that $r > 0$ and in this case when $\text{Var}\left(\hat{\xi}\right) = \frac{K\xi^2}{N}$ that K is given by

$$K(r) = \frac{4r}{(1+r)^2}\,.$$

One optimal value for $K(r)$ could be when $r = 0$ where $K(0) = 0$. Using calculus to look for other extreme values for $K(r)$ we need to solve

$$K'(r) = \frac{4}{(1+r)^2} - \frac{8r}{(1+r)^3} = 0\,,$$

which has $r = 1$. The value of K at this r is

$$K(1) = 1\,.$$

For this value of r we find

$$\frac{\alpha_1}{\sigma} = \sqrt{1} = 1\,.$$

Exercise 14

From the book, the variance of the derivative estimate is given by

$$\text{Var}\left(\frac{\partial\hat{\xi}}{\partial v}\right) = \frac{\xi_{\min}^2}{N\delta^2}\,. \tag{61}$$

On Page 35 we have specified the parameters needed to evaluate the above and with $N = 5$ we find

$$\text{Var}\left(\frac{\partial\hat{\xi}}{\partial v}\right) = \frac{3^2}{5(1^2)} = 1.8\,.$$

Exercise 15

From the book, the covariance of the derivative estimate is given by

$$\text{Var}\left(\hat{\nabla}_k\right) = \frac{\xi_{\min}^2}{N\delta^2}\mathbf{I}\,. \tag{62}$$

On Page 34 we have specified the parameters needed to evaluate the above and with $N = 50$ we find

$$\text{Var}\left(\hat{\nabla}_k\right) = \frac{4^2}{50\delta^2}\mathbf{I} = \frac{8}{25\delta^2}\mathbf{I}\,.$$

Exercise 16

For $k = 0$ this is

$$x_0 = by_0 .$$

For $k = 1$ this is

$$x_1 = ax_0 + by_1 = aby_0 + by_1 = b(y_1 + ay_0) .$$

For $k = 2$ this is

$$x_2 = ab(y_1 + ay_0) + by_2 = b(y_2 + ay_1 + a^2 y_0) .$$

Based on these examples it looks like the general solution is

$$x_k = b \sum_{i=0}^{k} y_{k-i} a^i . \tag{63}$$

We can check this is a solution to the given first order recurrence relation by computing $ax_{k-1} + by_k$. We find

$$\begin{aligned} ax_{k-1} + by_k &= a \left(b \sum_{i=0}^{k-1} y_{k-1-i} a^i \right) + by_k \\ &= b \left(y_k + \sum_{i=0}^{k-1} y_{k-1-i} a^{i+1} \right) \\ &= b \left(y_k + \sum_{i=1}^{k} y_{k-i} a^i \right) = b \sum_{i=0}^{k} y_{k-i} a^i , \end{aligned}$$

which is the same as Equation 63 showing that we have indeed found a solution.

Exercise 17

Recall that in the one-dimensional case that

$$\xi(v) = \xi_{\min} + \lambda (w - w^*)^2 ,$$

and thus $\mathbf{R} = \lambda$ and so $R^{-1} = \lambda^{-1}$.

From the book for Newton's method the weight-vector covariance is given by

$$\operatorname{cov}[\mathbf{V}_k] = \frac{\mu \xi_{\min}^2 (R^{-1})^2}{4N\delta^2 (1 - \mu)} . \tag{64}$$

For the steepest-descent method the weight-vector covariance is given by

$$\operatorname{cov}[\mathbf{V}_k] = \frac{\mu \xi_{\min}^2 (R - \mu R^2)^{-1}}{4N\delta^2} . \tag{65}$$

Exercise 18

The book's Equation 5.54 is

$$\text{cov}\,[V_k'] = \frac{\mu(\Lambda^{-1})^2\xi^2_{\text{min}}}{4N\delta^2(1-\mu)}\,, \tag{66}$$

Now recall that $V_k = QV_k'$ so that $V_k' = Q^T V_k$. Because of this we have

$$\text{cov}\,[V_k'] = E[(x-\mu)(x-\mu)^T] = \text{cov}\left[Q^T V_k\right] = Q^T \text{cov}\,[V_k]\,Q\,.$$

Thus Equation 66 in terms of V_k becomes

$$Q^T \text{cov}\,[V_k]\,Q = \frac{\mu(\Lambda^{-1})^2\xi^2_{\text{min}}}{4N\delta^2(1-\mu)}\,,$$

or solving for $\text{cov}\,[V_k]$ and using $Q^T Q = I$ we have

$$Q(\Lambda^{-1})^2 Q^T = Q\Lambda^{-1}Q^T Q\Lambda^{-1}Q^T = (R^{-1})^2\,.$$

In the above we have noted that from $R = Q\Lambda Q^T$ we have $\Lambda = Q^T RQ$ so $\Lambda^{-1} = Q^T R^{-1} Q$. Using this we have

$$\text{cov}\,[V_k] = \frac{\mu(R^{-1})^2\xi^2_{\text{min}}}{4N\delta^2(1-\mu)}\,,$$

as we were to show.

Deriving Equation 5.58 is done in the same way.

Exercise 19

On Page 35 we have determined the parameters for Exercise 5.

Recall that for the steepest-decent algorithm the stability range for μ is

$$0 < \mu < \frac{1}{\lambda_{\text{max}}}\,,$$

For Exercise 5 we have that $\lambda_{\text{max}} = 3$ so from the problem statement we have $\mu = \frac{1}{2}\left(\frac{1}{3}\right) = \frac{1}{6}$. To use Equation 65 we need to compute

$$(R - \mu R^2)^{-1} = \frac{1}{15}\begin{bmatrix} 14 & -4 \\ -4 & 14 \end{bmatrix}.$$

Using this with the rest of Equation 65 and $N = 10$ we compute

$$\text{cov}\,[V] = \frac{1}{225\delta^2}\begin{bmatrix} 14 & -4 \\ -4 & 14 \end{bmatrix}.$$

Exercise 20

The (i,i)th element of the matrix $\sum_{n=0}^{\infty} D^n$ would be

$$\sum_{n=0}^{\infty} d_{ii}^n = \frac{1}{1-d_{ii}}\,,$$

with all other elements (i,j) zero. This means that

$$\begin{aligned}\sum_{n=0}^{\infty} D^n &= \mathrm{Diag}\left(\frac{1}{1-d_{11}}\,,\frac{1}{1-d_{22}}\,,\frac{1}{1-d_{33}}\,,\cdots,\frac{1}{1-d_{nn}}\right)\\ &= \left(\mathrm{Diag}\left(1-d_{11}\,,1-d_{22}\,,1-d_{33}\,,\ldots,1-d_{nn}\right)\right)^{-1}\\ &= (I-D)^{-1}\,,\end{aligned}$$

as we were to show.

For this to be true we must have

$$|d_{ii}| < 1\,,$$

for all i.

Exercise 21

Part (a): Recall that for Newton's method we have

$$\text{excess MSE} = \frac{(L+1)\xi_{\min}\lambda_{\text{av}}\left(\frac{1}{\lambda}\right)_{\text{av}}}{8NP\tau}\,. \tag{67}$$

with P given by Equation 54 or

$$P = \frac{\delta^2}{\xi_{\min}}\frac{\sum_{n=0}^{L}\lambda_n}{L+1} = \frac{\delta^2}{\xi_{\min}}\lambda_{\text{av}}\,.$$

The stability of the convergence parameters μ in Newton's method is $0<\mu<1$ so for this problem we are told to take $\mu = \frac{1}{2}$.

For Newton's method, the time constant of weight convergence is given by

$$\mu = \frac{1}{2\tau} \quad \text{or} \quad \tau = \frac{1}{2\mu}\,.$$

Using the numbers needed to compute the above we find

$$\begin{aligned}P &= \frac{\delta^2(2)}{2} = \delta^2\\ \tau &= \frac{1}{2(1/2)} = 1\\ \text{excess MSE} &= \frac{(2)(2)(2)\left(\frac{2}{3}\right)}{8(10)\delta^2(1)} = \frac{1}{15\delta^2}\,.\end{aligned}$$

Part (b): Recall that for the steepest decent method we have

$$\text{excess MSE} = \frac{(L+1)\xi_{\text{min}}}{8P}\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}} . \tag{68}$$

with P again given by Equation 54. We can compute $\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}}$ using

$$\mu\lambda_{\text{av}} \approx \frac{N(L+1)}{2}\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}} . \tag{69}$$

The stability of the convergence parameters μ under the steepest decent method is

$$0 < \mu < \frac{1}{\lambda_{\text{max}}} = \frac{1}{3},$$

so for this problem we are told to take $\mu = \frac{1}{2}\left(\frac{1}{3}\right) = \frac{1}{6}$. From the numbers for Exercise 5 we compute

$$\begin{aligned}\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}} &= \frac{2(1/6)(2)}{10(2)} = \frac{1}{30} \quad \text{so} \\ \text{excess MSE} &= \frac{2^2(2)}{8\delta^2}\left(\frac{1}{6}\right) = \frac{1}{6\delta^2} .\end{aligned}$$

Exercise 22

The time constant for weight convergence is denoted τ and the time constant for the learning curve is denoted τ_{mse}. In this problem we have $L = 0$ (one weight) and $\mu = 0.01$.

Part (a): For Newton's method we have $\tau = \frac{1}{2\mu} = 50$ and $\tau_{\text{mse}} = \frac{\tau}{2} = 25$.

Part (b): For the method of steepest decent we have $\tau_0 = \frac{1}{2\mu\lambda_0} = \frac{50}{\lambda_0}$ (since there is only one eigenvalue) and $(\tau_{\text{mse}})_0 = \frac{\tau_0}{2} = \frac{25}{\lambda_0}$.

Exercise 23

Part (a): For Newton's method we have

$$T_{\text{mse}} = 2(L+1)N\tau_{\text{mse}} , \tag{70}$$

which can be computed given the numbers in this problem and the answers in the previous problem.

Part (b): For the method of steepest decent we have

$$(T_{\text{mse}})_n = 2N(L+1)(\tau_{\text{mse}})_n , \tag{71}$$

which can be computed given the numbers in this problem and the answers in the previous problem.

Exercise 24

We would evaluate the formulas for the excess mean-square error in Exercise 21 with $\delta = 0.05$ and $N = 5$ (rather than $N = 10$).

Exercise 25

The correlation matrix R is given by

$$R = E\begin{bmatrix} x_0^2 & x_0x_1 \\ x_1x_0 & x_1^2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}.$$

This has eigenvalues $\lambda \in \{1, 5\}$ so

$$\lambda_{\text{av}} = \frac{1}{2}(1+5) = 3$$
$$\left(\frac{1}{\lambda}\right)_{\text{av}} = \frac{1}{2}\left(\frac{1}{1} + \frac{1}{5}\right) = \frac{3}{5}.$$

Part (a): The misadjustment for Newton's method is given by

$$M \approx \frac{(L+1)\lambda_{\text{av}}\left(\frac{1}{\lambda}\right)_{\text{av}}}{8NP\tau}. \tag{72}$$

and

$$\tau = \frac{1}{2\mu}, \tag{73}$$

For the numbers given in this Exercise we compute

$$M = \frac{2(3)\left(\frac{3}{5}\right)}{8(80)0.05\left(\frac{1}{2(0.01)}\right)},$$

which could be evaluated.

Part (b): The misadjustment for the steepest decent method method is given by

$$M \approx \frac{(L+1)^2}{8P}\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}}. \tag{74}$$

Here we use Equation 71 to compute $(T_{\text{mse}})_n$ for $n = 0$ and $n = 1$ as

$$(T_{\text{mse}})_n = 2N(L+1)(\tau_{\text{mse}})_n = N(L+1)\tau_n = N(L+1)\left(\frac{1}{2\mu\lambda_n}\right).$$

We can then compute $\frac{1}{(T_{\text{mse}})_n}$ for each n and average these numbers as needed to compute $\left(\frac{1}{T_{\text{mse}}}\right)_{\text{av}}$. Evaluating the above for M is simple to do.

Chapter 6: The LMS Algorithm

Exercise Solutions

Exercise 1

WWX: Working here.

Chapter 7: The z-Transform in Adaptive Signal Processing

Exercise Solutions

Exercise 1

WWX: Working here.

References

[1] G. Corliss. Which root does the bisection algorithm find? *SIAM Review*, 19(2):325–327, 1977.

[2] W. Ferrar. *A text-book of convergence*. The Clarendon Press, 1938.