Department of Electrical Communication Engineering
Indian Institute of Science

E9 211 Adaptive Signal Processing

3 Dec 2020, Take-home mid-term assessment solutions

This exam has two questions (20 points).

This is an open book assessment with a turn in time of 24 hrs. Make sure the report is turned in before 9 am, 4 December 2020. This is a hard deadline. Your report should be a single PDF file, legible, scanned/pictured under good lighting conditions with your full name. The report should be submitted via MS Teams. Late or plagiarized submissions will not be graded.

Question 1 (10 points)

Suppose we receive signals from two interfering sources s_1 and s_2 in noise as

$$\mathbf{x} = \mathbf{a}_1 s_1 + \mathbf{a}_2 s_2 + \mathbf{n},$$

where the column vector $\mathbf{x}$ collects observations from an array of length M. The array steering (column) vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are assumed to be known. The receiver noise $\mathbf{n}$ is assumed to be zero mean with covariance matrix $\sigma^2\mathbf{I}$. Here, $\mathbf{I}$ is the identity matrix of size M. We assume that s_1, s_2, and noise are mutually uncorrelated. The sources have unit power.

We are interested in recovering the source symbols s_1 using a linear receiver as

$$\hat{s}_1 = \mathbf{w}^H \mathbf{x}$$

using the following constrained beamformers.

(4pts) (a) Find a receiver $\mathbf{w}$ that has a unity gain (distortionless response) towards the direction of source s_1 and minimizes the mean squared error. Show that this receiver minimizes the interference plus noise power at the output of the beamformer, where we treat s_2 as interference.

(4pts) (b) Find a unit norm beamformer, i.e., $\|\mathbf{w}\|_2 = 1$, that maximizes the signal to interference plus noise ratio, where the signal and interference plus noise components in $\mathbf{w}^H\mathbf{x}$ are $\mathbf{w}^H\mathbf{a}_1\mathbf{a}_1^H\mathbf{w}$ and $\mathbf{w}^H[\mathbf{a}_2\mathbf{a}_2^H + \sigma^2\mathbf{I}]\mathbf{w}$, respectively.

(2pts) (c) Compare the beamformers computed in part (a) and (b) of this question in terms of the signal to interference plus noise ratio.

Solutions

1. The desired receiver is expected to have a distortionless response towards source s_1. In other words, we have $\mathbf{w}^{\mathrm{H}}\mathbf{a}_1 = 1$. The desired beamformer can be obtained by solving the following optimization problem:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \mathbb{E}[|\mathbf{w}^{\mathrm{H}}\mathbf{x} - s_1|^2] \quad \text{such that} \quad \mathbf{w}^{\mathrm{H}}\mathbf{a}_1 = 1 \tag{1}$$

We note that

$$\begin{aligned} \mathbb{E}[|\mathbf{w}^{\mathrm{H}}\mathbf{x} - s_1|^2] &= \mathbb{E}[(\mathbf{w}^{\mathrm{H}}\mathbf{x} - s_1)(\mathbf{w}^{\mathrm{H}}\mathbf{x} - s_1)^{\mathrm{H}}] \\ &= \mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{r}_{xs} - \mathbf{r}_{xs}^{\mathrm{H}}\mathbf{w} + \mathbb{E}[|s_1|^2] \\ &= \mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} - 1 - 1 + 1, \end{aligned}$$

where $\mathbf{R}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]$, $\mathbf{r}_{xs} = \mathbb{E}[\mathbf{x}s_1^*] = \mathbb{E}[(\mathbf{a}_1 s_1 + \mathbf{a}_2 s_2 + \mathbf{n})s_1^*] = \mathbf{a}_1$, and $\mathbb{E}[|s_1|^2] = 1$. Using the method of Lagrange multiplier, we can transform the constrained optimization problem in (1) to the following form

$$\underset{\mathbf{w},\lambda}{\text{minimize}} \quad J(\mathbf{w}, \lambda) = \mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} - 1 - 1 + 1 + \lambda(1 - \mathbf{w}^{\mathrm{H}}\mathbf{a}_1)$$

Computing the gradient of $J(\mathbf{w}, \lambda)$ w.r.t $\mathbf{w}^{\mathrm{H}}$ and setting to zero yields

$$\mathbf{w}_{\text{opt}} = (\lambda)\mathbf{R}_x^{-1}\mathbf{a}_1$$

Using the constraint $\mathbf{a}_1^{\mathrm{H}}\mathbf{w}_{\text{opt}} = 1$, we note that

$$\lambda = \frac{1}{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a}_1}.$$

Thus, the desired beamformer is given by

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_x^{-1}\mathbf{a}_1}{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a}_1} \tag{2}$$

We now show that this receiver minimizes the interference plus noise power at the output of the beamformer. To begin with, let us write the received signal as

$$\mathbf{x} = \mathbf{d} + \mathbf{z},$$

where $\mathbf{d} = \mathbf{a}_1 s_1$ is the desired signal and $\mathbf{z}$ is the sum of interference and noise. Then we have $\mathbf{R}_x = \mathbf{R}_d + \mathbf{R}_z$ where $\mathbf{R}_d = \mathbb{E}[\mathbf{d}\mathbf{d}^{\mathrm{H}}] = \mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}$ and $\mathbf{R}_z = \mathbb{E}[\mathbf{z}\mathbf{z}^{\mathrm{H}}]$. We note that

$$\mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} = \mathbf{w}^{\mathrm{H}}\mathbf{R}_d\mathbf{w} + \mathbf{w}^{\mathrm{H}}\mathbf{R}_z\mathbf{w} = 1 + \mathbf{w}^{\mathrm{H}}\mathbf{R}_z\mathbf{w}.$$

when the constraint $\mathbf{a}_1^{\mathrm{H}}\mathbf{w} = 1$ is satisfied. The term $\mathbf{w}^{\mathrm{H}}\mathbf{R}_z\mathbf{w}$ corresponds to the interference plus noise power at the output of the beamformer. Hence we can conclude that the receiver given in (2) minimizes the sum of interference and noise power at the output of the beamformer.

2. The problem of obtaining the beamformer to maximize the SINR may be written as

$$\text{maximize} \quad \frac{\mathbf{w}^{\mathrm{H}}\mathbf{R}_d\mathbf{w}}{\mathbf{w}^{\mathrm{H}}\mathbf{R}_z\mathbf{w}}. \tag{3}$$

We can observe that the solution to the above optimization problem is independent of scale (i.e., if $\mathbf{w}_{\text{opt}}$ is a solution, $\beta\mathbf{w}_{\text{opt}}$ is a solution as well for any $\beta \in \mathbb{C}$.) So instead of explicitly solving for a constrained optimization problem with $\|\mathbf{w}\| = 1$, we can instead solve for the above un-constrained problem and then normalize the solution to make it unit norm.

There are multiple ways to solve the above optimization problem.

Method 1

Let the projection of $\mathbf{w}_{\text{opt}}$ onto $\mathbf{a}_1$ be α (i.e., $\mathbf{w}_{\text{opt}}^{\mathrm{H}}\mathbf{a}_1 = \alpha$) where $\mathbf{w}_{\text{opt}}$ is the solution and $\alpha \in \mathbb{R},\ \alpha > 0$. Then the numerator of (3) will be α^2 at the solution point (i.e., when $\mathbf{w} = \mathbf{w}_{\text{opt}}$). In that case, we can modify the optimization problem in (3) as follows

$$\text{minimize} \quad \mathbf{w}^{\mathrm{H}}\mathbf{R}_z\mathbf{w} \qquad \text{subject to} \quad \mathbf{w}^{\mathrm{H}}\mathbf{a}_1 = \alpha \tag{4}$$

Using the method of Lagrange multiplier, we can obtain the solution to (4) as

$$\mathbf{w}_{\text{opt}} = \alpha\frac{\mathbf{R}_z^{-1}\mathbf{a}_1}{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1} \tag{5}$$

We note that

$$\|\mathbf{w}_{\text{opt}}\| = \|\mathbf{R}_z^{-1}\mathbf{a}_1\|\frac{\alpha}{|\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1|}.$$

Since we are interested in a unit norm solution, we can normalize the above solution (i.e., $\mathbf{w}_{\text{opt}} \leftarrow \mathbf{w}_{\text{opt}}/\|\mathbf{w}_{\text{opt}}\|$) to obtain

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_z^{-1}\mathbf{a}_1}{\|\mathbf{R}_z^{-1}\mathbf{a}_1\|} \tag{6}$$

Method 2

The optimization problem in (3) is the well known *generalized Rayleigh quotient*. The solution is given by the generalized eigen vector corresponding to the largest eigen value of the eigen system $\{\mathbf{R}_d, \mathbf{R}_z\}$. Since $\mathbf{R}_z$ is invertible, this means that the solution $\mathbf{w}_{\text{opt}}$ will be the eigen vector of $\mathbf{R}_z^{-1}\mathbf{R}_d$ corresponding to the largest eigen value. We note that $\mathbf{R}_z^{-1}\mathbf{R}_d = \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}$ is a rank-one matrix. Since the eigen vectors of a matrix lies in the column span of that matrix, it is straightforward to observe that the required eigen vector is given by

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_z^{-1}\mathbf{a}_1}{\|\mathbf{R}_z^{-1}\mathbf{a}_1\|} \tag{7}$$

3. SINR of the receiver in part a (2) may be computed as follows.

$$\mathrm{SINR}_1 = \frac{\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_d\mathbf{w}_{\mathrm{opt}}}{\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_z\mathbf{w}_{\mathrm{opt}}}.$$

Due to the unity gain constraint, we have $\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_d\mathbf{w}_{\mathrm{opt}} = 1$. Denominator of SINR_1 can be computed as

$$\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_z\mathbf{w}_{\mathrm{opt}} = \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{R}_z\mathbf{R}_x^{-1}\mathbf{a}_1}{|\mathbf{a}_1\mathbf{R}_x^{-1}\mathbf{a}_1|^2}.$$

Hence, we get

$$\mathrm{SINR}_1 = \frac{|\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a}_1|^2}{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{R}_z\mathbf{R}_x^{-1}\mathbf{a}_1} \tag{8}$$

Similarly, we can compute the SINR_2 as

$$\mathrm{SINR}_2 = \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{R}_d\mathbf{R}_z^{-1}\mathbf{a}_1}{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{R}_z\mathbf{R}_z^{-1}\mathbf{a}_1} = \mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 \tag{9}$$

It is possible to show that $\mathrm{SINR}_1 = \mathrm{SINR}_2$. To begin with, we can make use of the matrix inversion lemma to express $\mathbf{R}_x^{-1}$ in terms of $\mathbf{R}_z^{-1}$. We have

$$\begin{aligned}\mathbf{R}_x^{-1} &= (\mathbf{R}_z + \mathbf{a}_1\mathbf{a}_1^{\mathrm{H}})^{-1} \\ &= \mathbf{R}_z^{-1} - \frac{\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1} \\ &= \frac{\mathbf{R}_z^{-1} + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1} \end{aligned} \tag{10}$$

We can now write

$$\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a}_1 = \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 + \mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1 - \mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}$$

We note that $\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1$ is a scalar. Hence with a simple rearrangement and cancellation, we obtain

$$\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a} = \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1} \tag{11}$$

We also have

$$\begin{aligned}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{R}_z\mathbf{R}_x^{-1}\mathbf{a}_1 &= \mathbf{a}_1^{\mathrm{H}}\Big(\frac{\mathbf{R}_z^{-1} + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big)\mathbf{R}_z\Big(\frac{\mathbf{R}_z^{-1} + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big)\mathbf{a}_1 \\ &= \mathbf{a}_1^{\mathrm{H}}\Big(\frac{\mathbf{I} + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{R}_z - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big)\Big(\frac{\mathbf{R}_z^{-1} + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big)\mathbf{a}_1 \\ &= \Big(\frac{\mathbf{a}_1^{\mathrm{H}} + \mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{R}_z - \mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big)\Big(\frac{\mathbf{R}_z^{-1}\mathbf{a}_1 + \mathbf{R}_z^{-1}\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1 - \mathbf{R}_z^{-1}\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}{1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}\Big) \\ &= \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}{(1+\mathbf{a}_1^{\mathrm{H}}\mathbf{R}_z^{-1}\mathbf{a}_1)^2} \end{aligned} \tag{12}$$

Substituting (11) and (12) in (8) yields

$$\text{SINR}_1 = \frac{(\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1)^2}{(1+\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1)^2}\frac{(1+\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1)^2}{\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1} = \mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 = \text{SINR}_2$$

Hence we conclude that the SINR of both beamformers that we have computed in this question are the same.

Alternative method

In question 1, we have also mentioned that the receiver that minimize the MSE with the distortion less constraint will also minimize the interference plus noise power at the output of the beamformer. We have also mentioned that the optimal beamformer can be obtained by solving an equivalent optimization problem

$$\text{minimize} \quad \mathbf{w}^{\text{H}}\mathbf{R}_z\mathbf{w} + \tilde{\lambda}(1 - \mathbf{w}^{\text{H}}\mathbf{a}_1).$$

Solution to the above optimization problem is

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_z^{-1}\mathbf{a}_1}{\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1}.$$

Thus the resulting SINR may be computed as

$$\begin{aligned}\text{SINR}_1 &= \frac{\mathbf{w}_{\text{opt}}^{\text{H}}\mathbf{R}_d\mathbf{w}_{\text{opt}}}{\mathbf{w}_{\text{opt}}^{\text{H}}\mathbf{R}_z\mathbf{w}_{\text{opt}}} \\ &= \frac{1}{\frac{(\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1)^2}{\mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{R}_z\mathbf{R}_z^{-1}\mathbf{a}_1}} \\ &= \mathbf{a}_1^{\text{H}}\mathbf{R}_z^{-1}\mathbf{a}_1 = \text{SINR}_2\end{aligned}$$

Question 2 (10 points)

Consider a first-order autoregressive, i.e., AR(1), process $x(n)$ that has an autocorrelation sequence

$$r_x(k) = \alpha^{|k|}.$$

We make noisy measurements of $x(n)$ as

$$y(n) = x(n) + v(n)$$

where $v(n)$ is zero mean white noise with a variance of σ^2 and $v(n)$ is uncorrelated with $x(n)$. We find the optimum first-order linear predictor of the form

$$\hat{x}(n+1) = w(0)y(n) + w(1)y(n-1) = \mathbf{w}^H\mathbf{y}$$

where $\mathbf{w} = [w(0)\ w(1)]^T$ and $\mathbf{y} = [y(n)\ y(n-1)]^T$.

(3pts) (a) Derive the optimum $\mathbf{w}$ by minimizing the mean-squared error

$$J = E\{|\hat{x}(n+1) - x(n+1)|^2\}.$$

What happens to $\mathbf{w}$ in the noise-free case $\sigma^2 \to 0$. Why?

(3pts) (b) Develop a steepest gradient descent algorithm that determines $\mathbf{w}$ iteratively. Provide a condition on the step-size μ in terms of α in order to guarantee convergence. Provide the value of the step-size that yields fastest convergence, and also the resulting time-constant.

(4pts) (c) Consider a modified cost function

$$J = E\{|\hat{x}(n+1) - x(n+1)|^2\} + \beta\|\mathbf{w}\|^2.$$

Design a steepest gradient descent algorithm that determines $\mathbf{w}$ iteratively. Find the value of the step-size that yields fastest convergence and compare with the optimum μ from the previous question. Comment on the convergence rate (i.e., the time constant) for $\beta > 0$ and argue when is the modified cost function useful.

Solutions

(a) Let us define $s = x(n+1)$. Then the cost function can be written as

$$J(\mathbf{w}) = \mathbb{E}[|\mathbf{w}^{\mathrm{H}}\mathbf{y} - s|^2]$$

Solution to the above optimization problem, $\mathbf{w}_{\text{opt}}$, is given by

$$\mathbf{w}_{\text{opt}}^{\mathrm{H}} = \mathbf{r}_{sy}\mathbf{R}_y^{-1},$$

where $\mathbf{R}_y = \mathbb{E}[\mathbf{y}\mathbf{y}^{\mathrm{H}}]$ and $\mathbf{r}_{sy} = \mathbb{E}[s\mathbf{y}^{\mathrm{H}}]$. Let us define the autocorrelation of the process $y(n)$ as $r_y(k) = \mathbb{E}[y(n)y^*(n-k)]$. Then it is straightforward to observe that

$$\mathbf{R}_y = \begin{bmatrix} r_y(0) & r_y(1) \\ r_y^*(1) & r_y(0) \end{bmatrix} \quad \text{and} \quad \mathbf{r}_{sy} = \begin{bmatrix} r_y(1) & r_y(2) \end{bmatrix}$$

Since the process $v(n)$ is white and uncorrelated with $x(n)$, we can compute

$$\begin{aligned} r_y(0) &= \mathbb{E}[(x(n)+v(n))(x(n)+v(n))^*] = r_x(0) + \sigma^2 = 1 + \sigma^2 \\ r_y(1) &= \mathbb{E}[(x(n)+v(n))(x(n-1)+v(n-1))^*] = r_x(1) = \alpha \\ r_y(2) &= \mathbb{E}[(x(n)+v(n))(x(n-2)+v(n-2))^*] = r_x(2) = \alpha^2 \end{aligned}$$

Optimal weights are computed as

$$\begin{aligned} \mathbf{w}_{\text{opt}}^{\mathrm{H}} &= \begin{bmatrix} \alpha & \alpha^2 \end{bmatrix} \begin{bmatrix} 1+\sigma^2 & \alpha \\ \alpha & 1+\sigma^2 \end{bmatrix}^{-1} \\ &= \frac{1}{(1+\sigma^2)^2 - \alpha^2} \begin{bmatrix} \alpha(1+\sigma^2) - \alpha^3 & \alpha^2\sigma^2 \end{bmatrix} \end{aligned}$$

When $\sigma^2 \to 0$, we have $\mathbf{w}_{\text{opt}}^{\mathrm{H}} \to \begin{bmatrix} \alpha & 0 \end{bmatrix}$. This is happening because $x(n)$ is an AR(1) process. For an AR(1) process, the current sample depends only on the previous sample. Hence one filter tap is sufficient.

(b) Update equations of the steepest descent algorithm is given by

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \mu(\mathbf{R}_y\mathbf{w}^{(k)} - \mathbf{r}_{ys})$$

The eigenvalues of $\mathbf{R}_y$ are $\lambda_{min} = 1 + \sigma^2 - |\alpha|$ and $\lambda_{max} = 1 + \sigma^2 + |\alpha|$. Hence the iterations will converge if

$$0 < \mu < \frac{2}{1 + \sigma^2 + |\alpha|}.$$

For the fastest convergence, we need to select the step size as

$$\begin{aligned}\mu_{opt} &= \frac{2}{\lambda_{min} + \lambda_{max}} \\ &= \frac{1}{1 + \sigma^2},\end{aligned}$$

and the corresponding time constant (say τ_1) is

$$\tau_1 = \frac{-1}{\ln(|1 - \mu_{opt}\lambda_{min}|)} = \frac{-1}{\ln(|(1 - \mu_{opt}\lambda_{max}|)} = \frac{-1}{\ln(||\alpha|/(1 + \sigma^2)|)}$$

(c) We can re-write the modified cost function (mentioned in the question) as

$$J(\mathbf{w}) = \mathbf{w}^{\mathrm{H}}\mathbf{R}_y\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{r}_{ys} - \mathbf{r}_{ys}^{\mathrm{H}}\mathbf{w} + r_x(0) + \beta\mathbf{w}^{\mathrm{H}}\mathbf{w}$$

Computing the gradient of $J(\mathbf{w})$ w.r.t. $\mathbf{w}^{\mathrm{H}}$ yields

$$\nabla J(\mathbf{w}) = \mathbf{R}_y\mathbf{w} - \mathbf{r}_{ys} + \beta\mathbf{w}$$

The update equations of the steepest descent algorithm to iteratively determine $\mathbf{w}$ is then given by

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \mu((\beta\mathbf{I} + \mathbf{R}_y)\mathbf{w}^{(k)} - \mathbf{r}_{ys})$$

The eigenvalues of $(\beta\mathbf{I}+\mathbf{R}_y)$ are $\lambda_{min} = 1+\sigma^2-|\alpha|+\beta$ and $\lambda_{max} = 1+\sigma^2+|\alpha|+\beta$. For the fastest convergence, we choose the step size as

$$\mu_{opt} = \frac{1}{1 + \sigma^2 + \beta}.$$

Since $\beta > 0$, this step size is smaller that the one we used in the previous question. Time constant (say τ_2) for this iteration can be found out as

$$\tau_2 = \frac{-1}{\ln(|1 - \mu_{opt}\lambda_{min}|)} = \frac{-1}{\ln(||\alpha|/(1 + \sigma^2 + \beta)|)}$$

Since $\beta > 0$, we can observe that $\tau_1 > \tau_2$. In other words, the steepest descent method for the modified cost function converges faster. This is also expected since the eigen spread of $\mathbf{R}_y + \beta\mathbf{I}$ is less than that of $\mathbf{R}_y$. Modified cost function is useful in scenarios where the matrix $\mathbf{R}_y$ is ill conditioned. We can improve the condition number (and thus the convergence speed) by diagonally loading $\mathbf{R}_y$ with $\beta > 0$.