III.12 (a) $J^{\alpha}(w) = \alpha \|w\|^2 + E|d - uw|^2$

$= \alpha w^H w + E[(d-uw)^H(d-uw)]$

$= \alpha w^H w + R_d + w^H R_u w - R_{du}^H w - w^H R_{ud}^H$

(Notation as used in section 8.1, Sayed)

$\Rightarrow \nabla_w (J^{\alpha}(w)) = \alpha w^H + w^H R_u - R_{du}^H$

setting the gradient to zero,

$\alpha w^H + w^H R_u - R_{du}^H = 0$

(also note that $R_d = \sigma_d^2$)

$\Rightarrow w^H (R_u + \alpha I) = R_{du}^H$

$\Rightarrow (R_u + \alpha I) w = R_{du}$

$\Rightarrow \boxed{w_{opt}^{\alpha} = (R_u + \alpha I)^{-1} R_{du}}$

- $J(w) = \alpha w^H w + R_d + w^H R_u w - R_{du}^H w - w^H R_{ud}^H$

$= \alpha w^H w + \sigma_d^2 + w^H R_u w - R_{ud} w - w^H R_{du}$

$= \alpha (w^H w) + \sigma_d^2 + w^H (R_u + \alpha I - \alpha I) w - R_{ud} w - w^H R_{du}$

$= \sigma_d^2 + w^H (R_u + \alpha I) w - R_{ud} w - w^H R_{du}$

Substitute $w = (R_u + \alpha I)^{-1} R_{du}$,

$J(w) = \sigma_d^2 + R_{ud} \underbrace{(R_u + \alpha I)^{-1} (R_u + \alpha I)}_{I} \cdot (R_u + \alpha I)^{-1} R_{du}$
$- R_{ud} (R_u + \alpha I)^{-1} R_{du}$
$- R_{ud} (R_u + \alpha I)^{-1} R_{du}$

$\Rightarrow \boxed{J(w) = \sigma_d^2 - R_{ud} (R_u + \alpha I)^{-1} R_{du}}$

- Given $\alpha > 0$. Also, $R_u > 0$ Positive definite Matrix

$\Rightarrow R_u = U D U^H$ where $U U^H = I$ and

$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ where λ_i's are eigenvalues of R_u and each $\lambda_i > 0$.

We know from properties of Matrices that the eigenvalues of R_y^{-1} are $1/\lambda_1, 1/\lambda_2 \cdots 1/\lambda_n$ and those of $(R_u+\alpha I)^{-1}$ are $(1/(\lambda_1+\alpha), 1/(\lambda_2+\alpha) \cdots 1/(\lambda_n+\alpha))$ and the eigenvectors remain the same.

$$\Rightarrow R_u^{-1} = U \begin{bmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_n \end{bmatrix} U^H, \quad (R_u+\alpha I)^{-1} = U \begin{bmatrix} \frac{1}{\lambda_1+\alpha} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n+\alpha} \end{bmatrix} U^H$$

$$\Rightarrow R_u^{-1} - (R_u+\alpha I)^{-1} = U \underbrace{\begin{bmatrix} \frac{1}{\lambda_1} - \frac{1}{\lambda_1+\alpha} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n} - \frac{1}{\lambda_n+\alpha} \end{bmatrix}}_{\text{Say } S} U^H.$$

as $\alpha > 0$, each nondiagonal element in $S = 0$ and each diagonal element in S is > 0. $\Rightarrow$ The eigenvalues of $R_u^{-1} - (R_u+\alpha I)^{-1}$ are all positive

$$\Rightarrow \quad R_u^{-1} - (R_u+\alpha I)^{-1} > 0 \rightarrow \text{Positive definite.}$$

$$\Rightarrow \quad x^H (R_u^{-1} - (R_u+\alpha I)^{-1}) x > 0 \text{ for any } x \in \mathbb{C}^{n\times 1}$$

$$\Rightarrow \quad R_{ud} (R_u^{-1} - (R_u+\alpha I)^{-1}) R_{du} > 0 \quad (R_{ud} = R_{du}^H) \rightarrow (1)$$

$$\Rightarrow J_{min} - J^{\alpha}(w^{\alpha}) = -R_{ud}(R_u^{-1} - (R_u+\alpha I)^{-1}) R_{du}$$

$$\Rightarrow J_{min} - J^{\alpha}(w^{\alpha}) < 0 \quad \Rightarrow \boxed{J^{\alpha}(w^{\alpha}) > J_{min}}$$

(b) From the previous derivation,

$$\begin{aligned} J_{min} - J^{\alpha}(w^{\alpha}) &= R_{ud}((R_u+\alpha I)^{-1} - R_u^{-1}) R_{du} \\ &= R_{ud}((R_u+\alpha I)^{-1} R_{du} - R_u^{-1} R_{du}) \\ &= R_{ud}(w^{\alpha} - w^{o}) \\ &= R_{ud}\, \delta w \end{aligned}$$

$$\Rightarrow \boxed{J_{min} - J^{\alpha}(w^{\alpha}) = R_{ud}\, \delta w}$$

(c). We know in gradient descent, we use $w_i = w_{i-1} + \mu p$. A logical choice for p is along the gradient, ie, $-[\nabla_w (J^\alpha(w^\alpha))]^H$

$$w_i^\alpha = w_{i-1}^\alpha - \mu[\nabla_w (J^\alpha(w^\alpha))]^H$$

$$\Rightarrow \quad w_i^\alpha = w_{i-1}^\alpha - \mu(\alpha w_{i-1}^\alpha + R_u w_{i-1}^\alpha - R_{du}) \quad \alpha \in \mathbb{R}^+$$

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

- So, we reach this update equation by setting P along the negative of the gradient.
- The update equation can be further simplified as

$$w_i^\alpha = [I - \mu(\alpha I + R_u)] w_{i-1}^\alpha + \mu R_{du}$$

Recall from step ⓐ that $\quad R_{du} = (R_u + \alpha I) w^\alpha_{opt}$

$$\Rightarrow \quad w_i^\alpha = [I - \mu(\alpha I + R_u)] w_{i-1}^\alpha + \mu(R_u + \alpha I) w^\alpha_{opt}$$

$$\Rightarrow \quad w_i^\alpha - w^\alpha_{opt} = [I - \mu(\alpha I + R_u)](w_{i-1}^\alpha - w^\alpha_{opt})$$

This is exactly the same equation we got for the gradient descent based update in Lecture 15 with the only difference being we have $(\alpha I + R_u)$ now instead of just R_u.
Say the eigenvalues of R_u are $\lambda_1, \lambda_2 \ldots \lambda_n$ and its corresponding eigenvectors are $u_1, u_2 \ldots u_n$. Then the eigenvalues of $(\alpha I + R_u)$ are $\lambda_1 + \alpha, \lambda_2 + \alpha \ldots \lambda_n + \alpha$ and the corresponding eigenvectors are $u_1, u_2 \ldots u_n$. So, instead of deriving everything from scratch, we can use the results from class and take care to use $\lambda_i + \alpha$ instead of λ_i whenever they occur in the results.
- We saw for the vanilla gradient descent that convergence happens if we choose $0 < \mu < 2/\lambda_{max}$. So, for our case, Convergence happens for $\boxed{0 < \mu < 2/(\lambda_{max} + \alpha)}$.

The optimal step-size for fastest convergence in the vanilla case is $\mu = \dfrac{2}{\lambda_{max} + \lambda_{min}}$.

$\Rightarrow$ The optimal step-size for fastest convergence (μ^α) in this case is

$$\frac{2}{(\lambda_{max}+\alpha) + (\lambda_{min}+\alpha)} = \boxed{\frac{2}{\lambda_{max} + \lambda_{min} + 2\alpha}}$$

(d) As seen in step (c), $\mu^0 = \dfrac{2}{\lambda_{max}+\lambda_{min}}$ and

$\mu^\alpha = \dfrac{2}{\lambda_{max} + \lambda_{min} + 2\alpha}$, i.e., $\mu^0 > \mu^\alpha$.

III. 26.

In III.12 we arrived at the update equation:

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

use instantaneous approximation and:

- set $R_{du} \approx d(i)\cdot u_i^H$
- set $R_u \approx u_i^H u_i$, we get

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

$d(i)$ is a scalar $\Rightarrow d(i)\, u_i^H = u_i^H\, d(i)$

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

(have used u^H instead of u^* to keep notation similar to that discussed in the class).

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

$$\Rightarrow d(i) - u_i w_i^\alpha = d(i) - (1-\mu\alpha)\, u_i w_{i-1}^\alpha - \mu\, \|u_i\|^2\,[d(i) - u_i w_{i-1}^\alpha]$$

(adding and subtracting $\alpha\mu d(i)$ on the RHS)

$$d(i) - u_i w^{\alpha}_{i-1} = d(i) - \alpha\mu d(i) - (1-\alpha\mu) u_i w^{\alpha}_{i-1}$$
$$- \mu \|u_i\|^2 [d(i) - u_i w^{\alpha}_{i-1}]$$
$$+ \alpha\mu d(i)$$

grouping the terms,

$$d(i) - u_i w^{\alpha}_{i-1} = d(i) [1 - \alpha\mu - \mu \|u_i\|^2]$$
$$- u_i w^{\alpha}_{i-1} [1 - \alpha\mu - \mu \|u_i\|^2]$$
$$+ \alpha\mu d(i)$$

$$\Rightarrow d(i) - u_i w^{\alpha}_{i-1} = (d(i) - u_i w^{\alpha}_{i-1}) [1 - \alpha\mu - \mu \|u_i\|^2]$$
$$+ \alpha\mu d(i)$$

$$\Rightarrow \boxed{r^{\alpha}(i) = [1 - \alpha\mu - \mu \|u_i\|^2] e^{\alpha}(i) + \alpha\mu d(i)}$$