Development of an expression for the MSE

All processes are zero-mean. Let the L-vector of observations be $\mathbf{y}_k = [\,y[k], y[k-1], \ldots, y[k-L+1]\,]^T$. The linear estimator is $\hat{x}[k] = \mathbf{h}^T \mathbf{y}_k$ with $\mathbf{h} \in \mathbb{R}^L$. Define

$$R_x[0] = E[x[k]^2], \qquad \mathbf{r}_{xy} = E[\mathbf{y}_k x[k]], \qquad R_y = E[\mathbf{y}_k \mathbf{y}_k^T].$$

1. Write the MSE

$$\begin{aligned} \mathrm{MSE}(\mathbf{h}) &= E\left[(x[k] - \mathbf{h}^T \mathbf{y}_k)^2\right] \\ &= E[x[k]^2] - 2\,\mathbf{h}^T E[\mathbf{y}_k x[k]] + \mathbf{h}^T E[\mathbf{y}_k \mathbf{y}_k^T]\mathbf{h} \\ &= R_x[0] - 2\,\mathbf{h}^T \mathbf{r}_{xy} + \mathbf{h}^T R_y \mathbf{h}. \end{aligned}$$

2. Minimize w.r.t. $\mathbf{h}$. Take gradient and set equal to zero:

$$\nabla_{\mathbf{h}} \mathrm{MSE} = -2\mathbf{r}_{xy} + 2R_y \mathbf{h} = 0 \quad \Longrightarrow \quad R_y \mathbf{h} = \mathbf{r}_{xy},$$

the Wiener–Hopf (normal) equations. (We assume R_y invertible so the solution is unique: $\mathbf{h} = R_y^{-1} \mathbf{r}_{xy}$.)

3. Substitute the optimal $\mathbf{h}$ into the MSE expression. Using the normal equations one has $\mathbf{h}^T R_y \mathbf{h} = \mathbf{h}^T \mathbf{r}_{xy}$. Thus

$$\mathrm{MSE}_{\min} = R_x[0] - 2\mathbf{h}^T \mathbf{r}_{xy} + \mathbf{h}^T R_y \mathbf{h} = R_x[0] - \mathbf{h}^T \mathbf{r}_{xy}.$$