Wiki: https://en.wikipedia.org/wiki/Minimum_mean_square_error

1 Motivation

The term MMSE more specifically refers to estimation in a Bayesian setting with quadratic cost function. The basic idea behind the Bayesian approach to estimation stems from practical situations where we often have some prior information about the parameter to be estimated. For instance, we may have prior information about the range that the parameter can assume; or we may have an old estimate of the parameter that we want to modify when a new observation is made available; or the statistics of an actual random signal such as speech. This is in contrast to the non-Bayesian approach like minimum-variance unbiased estimator (MVUE) where absolutely nothing is assumed to be known about the parameter in advance and which does not account for such situations. In the Bayesian approach, such prior information is captured by the prior probability density function of the parameters; and based directly on Bayes theorem, it allows us to make better posterior estimates as more observations become available. Thus unlike non-Bayesian approach where parameters of interest are assumed to be deterministic, but unknown constants, the Bayesian estimator seeks to estimate a parameter that is itself a random variable. Furthermore, Bayesian estimation can also deal with situations where the sequence of observations are not necessarily independent. Thus Bayesian estimation provides yet another alternative to the MVUE. This is useful when the MVUE does not exist or cannot be found.

2 Definition

Let $\boldsymbol{x}$ be a $n \times 1$ hidden random vector variable, and let $\boldsymbol{y}$ be a $m \times 1$ known random vector variable (the measurement or observation), both of them not necessarily of the same dimension. An estimator $\widehat{\boldsymbol{x}}(\boldsymbol{y})$ of $\boldsymbol{x}$ is any function of the measurement $\boldsymbol{y}$. The estimation error vector is given by $\boldsymbol{e} = \widehat{\boldsymbol{x}} - \boldsymbol{x}$ and its mean squared error (MSE) is given by the trace of error covariance matrix.

$$\begin{aligned} \boldsymbol{MSE} &= \boldsymbol{tr}\left\{\boldsymbol{E}\{(\widehat{\boldsymbol{x}} - \boldsymbol{x})(\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\boldsymbol{T}}\}\right\} \\ &= \boldsymbol{tr}\left\{\boldsymbol{E}\{(\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\boldsymbol{T}}(\widehat{\boldsymbol{x}} - \boldsymbol{x})\}\right\} \\ &= \boldsymbol{E}\left\{\boldsymbol{tr}\{(\widehat{\boldsymbol{x}} - \boldsymbol{x})(\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\boldsymbol{T}}\}\right\} \\ &= \boldsymbol{E}\{(\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\boldsymbol{T}}(\widehat{\boldsymbol{x}} - \boldsymbol{x})\} \end{aligned} \tag{2-1}$$

Where the expectationm $\boldsymbol{E}$ is taken over both $\boldsymbol{x}$ and $\boldsymbol{y}$. When $\boldsymbol{x}$ is a scalar variable, the MSE expression simplifies to $\boldsymbol{E}\{(\hat{x} - x)^2\}$. Note that MSE can equivalently be defined in other ways, since

$$\boldsymbol{tr}\left\{\boldsymbol{E}\{\boldsymbol{e}\boldsymbol{e}^{\boldsymbol{T}}\}\right\} = \boldsymbol{E}\left\{\boldsymbol{tr}\{\boldsymbol{e}\boldsymbol{e}^{\boldsymbol{T}}\}\right\} = \boldsymbol{E}\{\boldsymbol{e}^{\boldsymbol{T}}\boldsymbol{e}\} = \sum_{\boldsymbol{n}=\boldsymbol{0}}^{\boldsymbol{N}-\boldsymbol{1}} \boldsymbol{e}_{\boldsymbol{i}}^{\boldsymbol{2}} \tag{2-2}$$

The MMSE estimatoer is then defined as the estimator achieving minimal MSE:

$$\widehat{\boldsymbol{x}}_{\boldsymbol{MMSE}}(\boldsymbol{y}) = \arg\min_{\widehat{x}} \boldsymbol{MSE} \tag{2-3}$$

2.1 Properties

- Under some weak regularity assumptions, the MMSE estimator is uniquely defined, and is given by

$$\widehat{\boldsymbol{x}}_{\boldsymbol{MMSE}}(\boldsymbol{y}) = \boldsymbol{E}\{\boldsymbol{x}|\boldsymbol{y}\} \tag{2-4}$$

  In other words, the MMSE estimator is the conditional expectation of $\boldsymbol{x}$ given the known observed value of the measurements.

- The MMSE estimator is ***unbiased*** (under the regularity assumptions mentioned above):

$$\boldsymbol{E}\{\widehat{\boldsymbol{x}}_{\boldsymbol{MMSE}}(\boldsymbol{y})\} = \boldsymbol{E}\left\{\boldsymbol{E}\{\boldsymbol{x}|\boldsymbol{y}\}\right\} = \boldsymbol{E}\{\boldsymbol{x}\} \tag{2-5}$$

- The MMSE estimator is asymptotically unbiased and it converges in distribution to the normal distribution:

$$\sqrt{n}(\hat{\boldsymbol{x}}-\boldsymbol{x}) \xrightarrow{d} \boldsymbol{N}\left(0, \boldsymbol{I}^{-1}(\boldsymbol{x})\right) \tag{2-6}$$

  where $\boldsymbol{I}(\boldsymbol{x})$ is the Fisher information of $\boldsymbol{x}$. Thus, the MMSE estimator is asymptotically efficient.
- The orthogonality principle: When $\boldsymbol{x}$ is a scalar, ***an estimator constrained to be of certain form $\hat{\boldsymbol{x}} = \boldsymbol{g}(\boldsymbol{y})$ is an optimal estimator***, i.e. $\hat{\boldsymbol{x}}_{\boldsymbol{MMSE}} = \boldsymbol{g}^*(\boldsymbol{y})$ if and only if

$$\boldsymbol{E}\{(\hat{\boldsymbol{x}}_{\boldsymbol{MMSE}} - \boldsymbol{x})\boldsymbol{g}(\boldsymbol{y})\} = \boldsymbol{0} \tag{2-7}$$

  Yunshuai: This means the estimation error of x is perpendicular to the observed data, so it's optimal in linear meaning.
  For all $\boldsymbol{g}(\boldsymbol{y})$ in closed, linear subspace $V = \{\boldsymbol{g}(\boldsymbol{y})^2 | \boldsymbol{g}: \mathbb{R}^m \to \mathbb{R}, \boldsymbol{E}\{\boldsymbol{g}(\boldsymbol{y})^2\} < +\infty\}$ of the measurements. For random vectors, since the MSE for estimation of random vector decomposes into finding the MMSE estimator of the coordinates of X separately:

$$\boldsymbol{E}\left\{\left(\boldsymbol{g}_j^*(\boldsymbol{y}) - \boldsymbol{x}\right)\boldsymbol{g}_j(\boldsymbol{y})\right\} = \boldsymbol{0} \tag{2-8}$$

  For all i and j. More succinctly put, the cross-correlation between the minimum estimation errors $\hat{\boldsymbol{x}}_{\boldsymbol{MMSE}} - \boldsymbol{x}$ and the estimator $\hat{\boldsymbol{x}}$ $(= \boldsymbol{g}(\boldsymbol{y}))$ should be zero

$$\begin{aligned} &\boldsymbol{E}\{(\hat{\boldsymbol{x}}_{\boldsymbol{MMSE}} - \boldsymbol{x})\hat{\boldsymbol{x}}^{\boldsymbol{T}}\} \\ &= \boldsymbol{E}\{(\hat{\boldsymbol{x}}_{\boldsymbol{MMSE}} - \boldsymbol{x})\boldsymbol{g}^{\boldsymbol{T}}(\boldsymbol{y})\} = \boldsymbol{0} \end{aligned} \tag{2-9}$$

- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly Gaussian, then the MMSE estimator is linear, ***i.e., it has the form $\boldsymbol{W}\boldsymbol{y} + \boldsymbol{b}$ for matrix $\boldsymbol{W}$ and constant $\boldsymbol{b}$.*** This can be directly shown using the Bayes theorem. As a consequence, to find the MMSE estimator, it is sufficient to find the linear MMSE estimator.

2.2 Linear MMSE estimator

In many cases, it is not possible to determine the analytical expression of the MMSE estimator. Two basic numerical approaches to obtain the MMSE estimate depends on either finding the conditional expectation $\mathbf{E}\{\boldsymbol{x}|\boldsymbol{y}\}$ or finding the minima of MSE. Direct numerical evaluation of the conditional expectation is computationally expensive, since they often require multidimensional integration usually done via Monte Carlo methods. Another computational approach is to directly seek the minima of the MSE using techniques such as the gradient descent methods; but this method still requires the evaluation of expectation. While these numerical methods have been fruitful, a closed form expression for the MMSE estimator is nevertheless possible if we are willing to make some compromises.

One possibility is to abandon the full optimality requirements and seek a technique minimizing the MSE within a particular class of estimators, such as the class of linear estimators. Thus we postulate that the conditional expectation of $\boldsymbol{x}$ given $\boldsymbol{y}$ is a simple linear function of $\boldsymbol{y}$, $\boldsymbol{E}\{\boldsymbol{x}|\boldsymbol{y}\} = \boldsymbol{W}\boldsymbol{y} + \boldsymbol{b}$, where the measurement $\boldsymbol{y}$ is a random vector, $\boldsymbol{W}$ is a matrix and $\boldsymbol{b}$ is a vector. This can be seen as the first order Taylor approximation of $\boldsymbol{E}\{\boldsymbol{x}|\boldsymbol{y}\}$. The linear MMSE estimator is the estimator achieving minimum MSE among all estimators of such form. That is, it solves the following the optimization problem:

$$\min_{\boldsymbol{W},\boldsymbol{b}} \boldsymbol{MSE} \quad s.t. \quad \hat{\boldsymbol{x}} = \boldsymbol{W}\boldsymbol{y} + \boldsymbol{b} \tag{2-10}$$

One advantage of such linear MMSE estimator is that it is not necessary to explicitly calculate the posterior probability density function of $\boldsymbol{x}$. Such linear estimator only depends on the first two moments of $\boldsymbol{x}$ and $\boldsymbol{y}$. So although it may be convenient to assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly Gaussian, it is not necessary to make this assumption, so long as the assumed distribution has well defined first and second moments. The form of the linear estimator does not depend on the type of the assumed underlying distribution.

The expression for linear optimal $\boldsymbol{b}$ and $\boldsymbol{W}$ are given by

$$\begin{aligned} \boldsymbol{b} &= \overline{\boldsymbol{x}} - \boldsymbol{W}\overline{\boldsymbol{y}}, \\ \boldsymbol{W} &= \boldsymbol{C_{XY}}\boldsymbol{C_Y^{-1}}. \end{aligned} \tag{2-11}$$

where $\overline{\boldsymbol{x}} = \boldsymbol{E}\{\boldsymbol{x}\}, \overline{\boldsymbol{y}} = \boldsymbol{E}\{\boldsymbol{y}\}$, the $\boldsymbol{C_{XY}}$ is cross-covariance matrix between $\boldsymbol{x}$ and $\boldsymbol{y}$, the $\boldsymbol{C_Y}$ is auto-covariance matrix of $\boldsymbol{y}$.

==

2.2.1 Proof of (2-11) (Yunshuai):

$$\begin{aligned} \hat{x}_i &= \boldsymbol{W_{i,:}}\,\boldsymbol{y} + b_i \\ e_i &= \hat{x}_i - x_i \\ S &= \|e_i\|^2 = e_i^* e_i \\ &= (\hat{x}_i - x_i)^H(\hat{x}_i - x_i) \\ &= \left(\boldsymbol{W_{i,:}}\,\boldsymbol{y} + b_i - x_i\right)^H\left(\boldsymbol{W_{i,:}}\,\boldsymbol{y} + b_i - x_i\right) \\ &\overset{\overline{scalar}}{=} \left(\boldsymbol{W_{i,:}}\,\boldsymbol{y} + b_i - x_i\right)\left(\boldsymbol{W_{i,:}}\,\boldsymbol{y} + b_i - x_i\right)^H \\ &= \boldsymbol{W_{i,:}}\,\boldsymbol{y}\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} + \boldsymbol{W_{i,:}}\,\boldsymbol{y}(b_i - x_i)^H + (b_i - x_i)\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} + (b_i - x_i)(b_i - x_i)^H \end{aligned} \tag{2-12}$$

Then the expectation of the square of the error is

$$\begin{aligned} \boldsymbol{E}\{S\} &= \boldsymbol{E}\left\{\boldsymbol{W_{i,:}}\,\boldsymbol{y}\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} + \boldsymbol{W_{i,:}}\,\boldsymbol{y}(b_i - x_i)^H + (b_i - x_i)\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} + (b_i - x_i)(b_i - x_i)^H\right\} \\ &= E\{\boldsymbol{W_{i,:}}\boldsymbol{y}\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\boldsymbol{y}x_i^H\} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}b_i^H + b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} - \boldsymbol{E}\{x_i\boldsymbol{y^H}\}\boldsymbol{W_{i,:}^H} + b_ib_i^H - b_i\bar{x}_i^H - \bar{x}_ib_i^H + C_{X_iX_i} \\ &= \boldsymbol{W_{i,:}}\boldsymbol{C_Y}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\boldsymbol{C_{X_iY}} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}b_i^H + b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} - \boldsymbol{C_{X_iY}^H}\boldsymbol{W_{i,:}^H} + b_ib_i^H - b_i\bar{x}_i^H - \bar{x}_ib_i^H + C_{X_iX_i} \end{aligned} \tag{2-13}$$

Calculate the derivation to b_i and force it to zero to obtain the MMSE b_i

$$\begin{aligned} \frac{\partial \boldsymbol{E}\{S\}}{\partial b_i} &= 2\left(\boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\right)^{\boldsymbol{H}} + 2b_i^H - 2\bar{x}_i^H = 0 \\ &\Rightarrow \hat{b}_i = \bar{x}_i - \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}} \\ &\Rightarrow \widehat{\boldsymbol{b}} = \overline{\boldsymbol{x}} - \boldsymbol{W}\overline{\boldsymbol{y}} \end{aligned} \tag{2-14}$$

Change (2-13) *to*

$$\begin{aligned} \boldsymbol{E}\{S\} &= E\left\{\boldsymbol{W_{i,:}}\boldsymbol{y}\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\boldsymbol{y}x_i^H\right\} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}b_i^H + b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} - \boldsymbol{E}\{x_i\boldsymbol{y^H}\}\boldsymbol{W_{i,:}^H} + K \\ &= E\left\{\boldsymbol{W_{i,:}}(\boldsymbol{y} - \overline{\boldsymbol{y}})(\boldsymbol{y} - \overline{\boldsymbol{y}})^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}(\boldsymbol{y} - \overline{\boldsymbol{y}})(x_i - \bar{x}_i)^H - (x_i - \bar{x}_i)(\boldsymbol{y} - \overline{\boldsymbol{y}})^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H}\right\} \\ &\qquad + \boldsymbol{E}\left\{\boldsymbol{W_{i,:}}\boldsymbol{y}\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\boldsymbol{y^H}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H}\right\} \\ &\qquad - \boldsymbol{E}\left\{\boldsymbol{W_{i,:}}\boldsymbol{y}\bar{x}_i^H + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}x_i^H - \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\bar{x}_i^H\right\} - \boldsymbol{E}\{x_i\overline{\boldsymbol{y}}^{\boldsymbol{H}} + \bar{x}_i\boldsymbol{y^H} - \bar{x}_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\}\boldsymbol{W_{i,:}^H} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}b_i^H \\ &\qquad + b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} + K \\ &= \left(\boldsymbol{W_{i,:}}\boldsymbol{C_Y}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\boldsymbol{C_{X_iY}} - \boldsymbol{C_{X_iY}^H}\boldsymbol{W_{i,:}^H}\right) + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} - \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}\bar{x}_i^H - \bar{x}_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} + \boldsymbol{W_{i,:}}\overline{\boldsymbol{y}}b_i^H \\ &\qquad + b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}}\boldsymbol{W_{i,:}^H} + K \end{aligned} \tag{2-15}$$

Where $K = b_ib_i^H - b_i\bar{x}_i^H - \bar{x}_ib_i^H + C_{X_iX_i}$

Calculate the derivation to $\boldsymbol{W_{i,:}^H}$ and force it to zero to obtain the MMSE $\boldsymbol{W_{i,:}^H}$

$$
\begin{aligned}
\frac{\partial \boldsymbol{E}\{S\}}{\partial \boldsymbol{W}_{i,:}^{\boldsymbol{H}}} &= 2\boldsymbol{W}_{i,:}\boldsymbol{C}_Y - 2\boldsymbol{C}_{X_iY}^{\boldsymbol{H}} + 2\boldsymbol{W}_{i,:}\overline{\boldsymbol{y}}\overline{\boldsymbol{y}}^{\boldsymbol{H}} - 2\bar{x}_i\overline{\boldsymbol{y}}^{\boldsymbol{H}} + 2b_i\overline{\boldsymbol{y}}^{\boldsymbol{H}} = 0 \\
&\Longleftrightarrow 2\boldsymbol{W}_{i,:}\boldsymbol{C}_Y - 2\boldsymbol{C}_{X_iY}^{\boldsymbol{H}} + 2\left(\boldsymbol{W}_{i,:}\overline{\boldsymbol{y}} - \bar{x}_i + b_i\right) = 0 \\
&\Longleftrightarrow 2\boldsymbol{W}_{i,:}\boldsymbol{C}_Y - 2\boldsymbol{C}_{X_iY}^{\boldsymbol{H}} + 0 = 0 \\
&\Rightarrow \boldsymbol{W}_{i,:}\boldsymbol{C}_Y = \boldsymbol{C}_{X_iY}^{\boldsymbol{H}} \\
&\Rightarrow \boldsymbol{W}\boldsymbol{C}_Y = \boldsymbol{C}_{XY} \\
&\Rightarrow \boldsymbol{W} = \boldsymbol{C}_{XY}\boldsymbol{C}_Y^{-1}
\end{aligned} \tag{2-16}
$$

Substitue (2-14) into (2-16) ($\boldsymbol{W}_{i,:}\overline{\boldsymbol{y}} - \bar{x}_i + b_i = 0$)

$$
\begin{aligned}
\frac{\partial \boldsymbol{E}\{S\}}{\partial \boldsymbol{W}_{i,:}^{\boldsymbol{H}}} &= \boldsymbol{W}_{i,:}\boldsymbol{C}_Y - 2\boldsymbol{C}_{X_iY}^{\boldsymbol{H}} = 0 \\
&\Rightarrow \boldsymbol{W}_{i,:}\boldsymbol{C}_Y = \boldsymbol{C}_{X_iY}^{\boldsymbol{H}} \\
&\Rightarrow \boldsymbol{W}\boldsymbol{C}_Y = \boldsymbol{C}_{XY} \\
&\Rightarrow \boldsymbol{W} = \boldsymbol{C}_{XY}\boldsymbol{C}_Y^{-1}
\end{aligned} \tag{2-17}
$$

So the linear MMSE solution is

$$
\begin{aligned}
\widehat{\boldsymbol{b}}_{mmse} &= \overline{\boldsymbol{x}} - \boldsymbol{W}_{mmse}\overline{\boldsymbol{y}} \\
\boldsymbol{W}_{mmse} &= \boldsymbol{C}_{XY}\boldsymbol{C}_Y^{-1}
\end{aligned} \tag{2-18}
$$

And then the MMSE estimation result is

$$
\begin{aligned}
\widehat{\boldsymbol{x}} &= \boldsymbol{W}_{mmse}\boldsymbol{y} + \overline{\boldsymbol{x}} - \boldsymbol{W}_{mmse}\overline{\boldsymbol{y}} \\
&= \boldsymbol{W}_{mmse}(\boldsymbol{y} - \overline{\boldsymbol{y}}) + \overline{\boldsymbol{x}}
\end{aligned} \tag{2-19}
$$

2.2.2 Common form of the MMSE estimator in wireless communication system

In wireless communication system, signal model in below is widely used

$$
\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{z} \tag{2-20}
$$

Where $\boldsymbol{s}$ is the transmitted signal, $\boldsymbol{H}$ is the channel response written in matrix form, $\boldsymbol{z}$ is the additive noise or noise with interference (color noise), then

$$
\begin{aligned}
\widehat{\boldsymbol{b}}_{mmse} &= \overline{\boldsymbol{s}} - \boldsymbol{W}_{mmse}\overline{\boldsymbol{y}} \\
\boldsymbol{W}_{mmse} &= \boldsymbol{C}_{SY}\boldsymbol{C}_Y^{-1}
\end{aligned} \tag{2-21}
$$

Usually it assumes $\boldsymbol{s}$ and $\boldsymbol{w}$ to Independent Identical Distribution (i.i.d.), and especially obey zero mean normal distribution $CN(0, \sigma^2)$, then the linear MMSE can be simplified

$$
\begin{aligned}
\overline{\boldsymbol{s}} &= \boldsymbol{0} \\
\overline{\boldsymbol{y}} &= \boldsymbol{H}\boldsymbol{0} + \boldsymbol{0} = \boldsymbol{0} \\
\widehat{\boldsymbol{b}}_{mmse} &= \boldsymbol{0}
\end{aligned} \tag{2-22}
$$

And

$$\begin{aligned}
C_{SY} &= E\{s(Hs+z)^H\} = C_S H^H \\
&= \sigma_s^2 I_{n\times n} H^H = \sigma_s^2 H^H \\
C_Y &= E\{(Hs+z)(Hs+z)^H\} \\
&= HC_S H^H + C_z \\
&\xRightarrow{only\ awgn\ noise} H\sigma_s^2 I_{n\times n} H^H + \sigma_n^2 I_{n\times n} \\
&= H\sigma_s^2 H^H + \sigma_n^2
\end{aligned} \tag{2-23}$$

Then the MMSE estimator/filters becomes to

$$\begin{aligned}
W_{mmse} &= C_S H^H (HC_S H^H + C_z)^{-1} \\
&= \sigma_s^2 H^H (H\sigma_s^2 H^H + \sigma_n^2 I)^{-1} \\
&= H^H \left(HH^H + \frac{\sigma_n^2}{\sigma_s^2} I\right)^{-1} \\
&= H^H \left(HH^H + \frac{1}{snr} I\right)^{-1}
\end{aligned} \tag{2-24}$$

2.2.3 Alternative form for LMMSE

$$\begin{aligned}
W_{mmse} &= (H^H C_z^{-1} H + C_s^{-1})^{-1} H^H C_z^{-1} \\
&\xRightarrow{only\ awgn\ noise} \left(H^H H + \frac{\sigma_n^2}{\sigma_s^2} I\right)^{-1} H^H
\end{aligned} \tag{2-25}$$

We can use the matrix identity given by Luenberger [1, P90] to derive (2-25)

$$RW^H (WRW^H + Q)^{-1} = (W^H Q^{-1} W + R^{-1})^{-1} W^H Q^{-1} \tag{2-26}$$

Proving:

$$\begin{aligned}
& W^H Q^{-1} WRW^H = W^H Q^{-1} WRW^H \\
\Longleftrightarrow\ & W^H Q^{-1} WRW^H + R^{-1} RW^H = W^H Q^{-1} WRW^H + W^H Q^{-1} Q \\
\Longleftrightarrow\ & (W^H Q^{-1} W + R^{-1}) RW^H = W^H Q^{-1} (WRW^H + Q) \\
\Longleftrightarrow\ & (W^H Q^{-1} W + R^{-1}) RW^H (WRW^H + Q)^{-1} = W^H Q^{-1} \\
\Longleftrightarrow\ & RW^H (WRW^H + Q)^{-1} = (W^H Q^{-1} W + R^{-1})^{-1} W^H Q^{-1}
\end{aligned} \tag{2-27}$$

For (2-24), $W = H, R = C_s, Q = C_z$, by using the identity, we can get

$$\begin{aligned}
W_{mmse} &= (H^H C_z^{-1} H + C_s^{-1})^{-1} H^H C_z^{-1} \\
&\xRightarrow{only\ awgn\ noise} H^H (H(\sigma_n^2 I)^{-1} H^H + (\sigma_s^2 I)^{-1})^{-1} (\sigma_n^2 I)^{-1}
\end{aligned} \tag{2-28}$$

$$= \left(H^H H + \frac{\sigma_n^2}{\sigma_s^2} I\right)^{-1} H^H$$
$$= \left(H^H H + \frac{1}{snr} I\right)^{-1} H^H$$

Wonderful result.

Actually

$$\begin{aligned} H^H \left(HH^H + \frac{1}{snr} I\right)^{-1} &= \left(H^H H + \frac{1}{snr} I\right)^{-1} H^H \\ \Longleftrightarrow \left(H^H H + \frac{1}{snr} I\right) H^H &= H^H \left(HH^H + \frac{1}{snr} I\right) \\ \Longleftrightarrow H^H H H^H + \frac{1}{snr} H^H &= H^H H H^H + \frac{1}{snr} H^H \end{aligned} \tag{2-29}$$

In many applications the number of observations is signiflcantly larger than the number of variables to estimate. As an example, if 10 observations are collected to estimate 2 variables; the flrst form requires 10×10 matrix inversion, the second form requires 2×2 matrix inversion which makes a lot of difierence in the implementation.

[1] D. G. Luenberger, *Optimization by vector space methods*. Wiley, 1990

2.2.4 Whitening if interference exist

If there isn't addictive white noise but with some colorful interference noise, some solution uses whitening method

$$\begin{aligned} y &= Hs + z \\ \underline{y} = Wy &= WHs + Wz \\ &= \underline{H}s + \underline{z} \end{aligned} \tag{2-30}$$

Where

$$\begin{aligned} W^H W &= C_z^{-1} \\ C_{\underline{z}} &= W C_z W^H \\ &= I \end{aligned} \tag{2-31}$$

Proof (2-13)

$$\begin{aligned} W^H W &= C_z^{-1} \\ \Longleftrightarrow W^H W C_z &= I \\ \Longleftrightarrow W^H W C_z W^H &= I W^H = W^H I \\ \Longleftrightarrow W C_z W^H &= I \end{aligned} \tag{2-32}$$

If we apply MMSE estimation for $\underline{\boldsymbol{y}}$, then the solution is

$$\begin{aligned}\underline{\boldsymbol{W}}_{mmse} &= \left(\underline{\boldsymbol{H}}^H \boldsymbol{C}_{\underline{z}}^{-1}\underline{\boldsymbol{H}} + \boldsymbol{C}_s^{-1}\right)^{-1}\underline{\boldsymbol{H}}^H \boldsymbol{C}_{\underline{z}}^{-1} \\ &= (\boldsymbol{H}^H\boldsymbol{W}^H\boldsymbol{I}\boldsymbol{W}\boldsymbol{H} + \boldsymbol{C}_s^{-1})^{-1}\boldsymbol{H}^H\boldsymbol{W}^H \\ &= (\boldsymbol{H}^H\boldsymbol{C}_z^{-1}\boldsymbol{H} + \boldsymbol{C}_s^{-1})^{-1}\boldsymbol{H}^H\boldsymbol{W}^H\end{aligned} \tag{2-33}$$

Then the MMSE estimation of $\boldsymbol{s}$ is

$$\begin{aligned}\boldsymbol{s}_{mmse} &= \underline{\boldsymbol{W}}_{mmse}\underline{\boldsymbol{y}} \\ &= (\boldsymbol{H}^H\boldsymbol{C}_z^{-1}\boldsymbol{H} + \boldsymbol{C}_s^{-1})^{-1}\boldsymbol{H}^H\boldsymbol{W}^H\boldsymbol{W}\boldsymbol{y} \\ &= (\boldsymbol{H}^H\boldsymbol{C}_z^{-1}\boldsymbol{H} + \boldsymbol{C}_s^{-1})^{-1}\boldsymbol{H}^H\boldsymbol{C}_z^{-1}\boldsymbol{y} \\ &= \boldsymbol{W}_{mmse}\boldsymbol{y}\end{aligned} \tag{2-34}$$

2.2.5 One alternative/simple whitening form

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{H}\boldsymbol{s} + \boldsymbol{z} \\ \boldsymbol{W}^H\boldsymbol{W} &= \boldsymbol{C}_z^{-1}\end{aligned} \tag{2-35}$$

Let $\boldsymbol{W}' = \boldsymbol{Q}^H\boldsymbol{W}$, where $\boldsymbol{W}\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{R}$, then is one whitening filtering matrix which whitens $\boldsymbol{z}$ and triangularize the channel $\boldsymbol{H}$ simultaneously.

$$\begin{aligned}(\boldsymbol{W}')^H\boldsymbol{W}' &= \boldsymbol{W}^H\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{W} = \boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{C}_z^{-1} \\ \underline{\boldsymbol{y}} = \boldsymbol{W}'\boldsymbol{y} &= \boldsymbol{Q}^H\boldsymbol{W}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{W}'\boldsymbol{z} \\ &= \boldsymbol{Q}^H\boldsymbol{Q}\boldsymbol{R}\boldsymbol{s} + \underline{\boldsymbol{z}} \\ &= \boldsymbol{R}\boldsymbol{s} + \underline{\boldsymbol{z}}\end{aligned} \tag{2-36}$$

Where $\boldsymbol{R}$ is one up-triangular matrix, which can simply the equation.

2.2.6 Computation of the MMSE estimator

The key operation is the matrix inversion, both (2-24), (2-25) can be solved by Cholesky decomposition, SVD decomposition …

For the form shown in (2-24), we can use LR decomposition to solve the MMSE equation

$$\begin{aligned}\underline{\boldsymbol{s}}_{mmse} = \begin{bmatrix}\boldsymbol{s}_{mmse} \\ \boldsymbol{J}_{m\times 1}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{H}^H + \dfrac{\boldsymbol{1}}{\boldsymbol{snr}}\right)^{-1}\boldsymbol{y} \\ \boldsymbol{J}_{m\times 1}\end{bmatrix} = \underline{\boldsymbol{W}}_{mmse}\boldsymbol{y} \\ &= \left[\boldsymbol{H}, \frac{\boldsymbol{1}}{\boldsymbol{snr}}\boldsymbol{I}_m\right]^H\left(\left[\boldsymbol{H}, \frac{\boldsymbol{1}}{\boldsymbol{snr}}\boldsymbol{I}_m\right]\left[\boldsymbol{H}, \frac{\boldsymbol{1}}{\boldsymbol{snr}}\boldsymbol{I}_m\right]^H\right)^{-1}\boldsymbol{y} \\ &= (\boldsymbol{L}\boldsymbol{R})^H(\boldsymbol{L}\boldsymbol{R}(\boldsymbol{L}\boldsymbol{R})^H)^{-1}\boldsymbol{y} \\ &= \boldsymbol{R}^H\boldsymbol{L}^H(\boldsymbol{L}\boldsymbol{L}^H)^{-1}\boldsymbol{y} \\ &\stackrel{?}{=} \boldsymbol{R}^H\boldsymbol{L}^{-1}\boldsymbol{y}\end{aligned} \tag{2-37}$$

$$\Rightarrow \boldsymbol{LR\underline{s}_{mmse}} = \boldsymbol{y}$$

One can first use forward-substitution to solve the $\boldsymbol{R\underline{s}_{mmse}} = \boldsymbol{t}$, then slove the $\boldsymbol{\underline{s}_{mmse}} = \boldsymbol{R^H t}$

For the form shown in (2-25), we can use QR decomposition to solve the MMSE equation

$$\begin{aligned} \boldsymbol{s_{mmse}} &= \boldsymbol{W_{mmse} y} \\ &= \left(\boldsymbol{H^H H} + \frac{\mathbf{1}}{\boldsymbol{snr}}\right)^{-1} \boldsymbol{H^H y} \\ &= \left(\begin{bmatrix} \boldsymbol{H} \\ \frac{\mathbf{1}}{\boldsymbol{snr}} \boldsymbol{I_m} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{H} \\ \frac{\mathbf{1}}{\boldsymbol{snr}} \boldsymbol{I_m} \end{bmatrix}\right)^{-1} \begin{bmatrix} \boldsymbol{H} \\ \frac{\mathbf{1}}{\boldsymbol{snr}} \boldsymbol{I_m} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0}_{\boldsymbol{m}} \end{bmatrix} \\ &= (\boldsymbol{R^H Q^H Q R})^{-1} \boldsymbol{R^H Q^H} \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0}_{\boldsymbol{m}} \end{bmatrix} \\ &= \boldsymbol{R^{-1} Q^H} \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0}_{\boldsymbol{m}} \end{bmatrix} \end{aligned} \tag{2-38}$$

One can first calculate $\boldsymbol{t} = \boldsymbol{Q^H} \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0}_{\boldsymbol{m}} \end{bmatrix}$, then uses the backward-substitution to solve the $\boldsymbol{Rs_{mmse}} = \boldsymbol{t}$

==

Thus the expression for linear MMSE estimator, its mean, and its auto-covariance is given by

$$\begin{aligned} \widehat{\boldsymbol{x}} &= \boldsymbol{W}(\boldsymbol{y} - \overline{\boldsymbol{y}}) + \overline{\boldsymbol{x}}, \\ \boldsymbol{E}\{\widehat{\boldsymbol{x}}\} &= \overline{\boldsymbol{x}}, \\ \boldsymbol{C}_{\widehat{X}} &= \boldsymbol{C_{XY} C_Y^{-1} C_{YX}} \end{aligned} \tag{2-39}$$

Where the $\boldsymbol{C_{YX}}$ is cross-covariance matrix between $\boldsymbol{y}$ and $\boldsymbol{x}$.

Lastly, the error covariance and minimum mean square error achievable by such estimator is

$$\begin{aligned} \boldsymbol{C_e} = \boldsymbol{C_X} - \boldsymbol{C}_{\widehat{X}} &= \boldsymbol{C_X} - \boldsymbol{C_{XY} C_Y^{-1} C_{YX}}, \\ \boldsymbol{LMMSE} &= \boldsymbol{tr}\{\boldsymbol{C_e}\} \end{aligned} \tag{2-40}$$

For the special case when both x and y are scalars, the above relations simplify to

$$\begin{aligned} \hat{x} &= \frac{\sigma_{XY}}{\sigma_Y^2}(y - \bar{y}) + \bar{x}, \\ \sigma_e^2 &= \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} \end{aligned} \tag{2-41}$$

2.3 Computation

Standard method like Gauss elimination can be used to solve the matrix equation for $\mathbf{W}$. A more numerically stable method is provided by QR decomposition method. Since the matrix $\mathbf{C}_{\boldsymbol{Y}}$ is a symmetric positive definite matrix, $\boldsymbol{W}$ can be solved twice as fast with theCholesky decomposition, while for large sparse systems conjugate gradient method is more effective. Levinson recursion is a fast method when $\boldsymbol{C_Y}$ is also a Toeplitz matrix. This can happen when $\boldsymbol{y}$ is a wide sense stationary process. In such stationary cases, these estimators are also referred to as Wiener-Kolmogorov filters.

2.4 Linear MMSE estimator for linear observation process (known linear equations)

Let us further model the underlying process of observation as a linear process: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{z}$ (or wrote as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}$ in some places), where $\boldsymbol{A}$ is a known matrix and $\boldsymbol{z}$ is random noise vector with the mean $\boldsymbol{E}\{\boldsymbol{z}\} = \boldsymbol{0}$ and cross-covariance $\boldsymbol{C}_{\boldsymbol{XZ}} = \boldsymbol{0}$. Here the required mean and the covariance matrices will be

$$
\begin{aligned}
\boldsymbol{E}\{\boldsymbol{y}\} &= \boldsymbol{A}\boldsymbol{E}\{\boldsymbol{x}\} = \boldsymbol{A}\overline{\boldsymbol{x}} \\
\boldsymbol{C}_{\boldsymbol{Y}} &= \boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{Z}} \\
\boldsymbol{C}_{\boldsymbol{XY}} &= \boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}}
\end{aligned} \tag{2-42}
$$

Thus the expression for the linear MMSE estimator matrix $\boldsymbol{W}$ shown in (2-11) further modifies to

$$
\boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}}(\boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{Z}})^{-1} \tag{2-43}
$$

Putting evering into the expression for $\widehat{\boldsymbol{x}}$ in (2-10), we get

$$
\widehat{\boldsymbol{x}} = \boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}}(\boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{Z}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\overline{\boldsymbol{x}}) + \overline{\boldsymbol{x}} \tag{2-44}
$$

(Yunshuai) Uses the linear equation denotion of $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}$ whcih is common used in communication system, the equation (2-44) becomes to

$$
\begin{aligned}
\boldsymbol{W} &= \boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{H}^{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{H}^{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{w}})^{-1} \\
&= \boldsymbol{\sigma}_s^2\boldsymbol{I}_{n\times n}\boldsymbol{H}^{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{\sigma}_s^2\boldsymbol{I}_{n\times n}\boldsymbol{H}^{\boldsymbol{H}} + \boldsymbol{\sigma}_n^2)^{-1} \\
&= \boldsymbol{\sigma}_s^2\boldsymbol{H}^{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{\sigma}_s^2\boldsymbol{H}^{\boldsymbol{H}} + \boldsymbol{\sigma}_n^2)^{-1} \\
&= \boldsymbol{H}^{\boldsymbol{H}}\left(\boldsymbol{H}\boldsymbol{H}^{\boldsymbol{H}} + \frac{\boldsymbol{\sigma}_n^2}{\boldsymbol{\sigma}_s^2}\right)^{-1} \\
&= \boldsymbol{H}^{\boldsymbol{H}}\left(\boldsymbol{H}\boldsymbol{H}^{\boldsymbol{H}} + \frac{\boldsymbol{1}}{\boldsymbol{snr}}\right)^{-1}
\end{aligned} \tag{2-45}
$$

With the assumption that $\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{\sigma}_s^2\boldsymbol{I}_{n\times n}$ and $\boldsymbol{C}_{\boldsymbol{w}} = \boldsymbol{\sigma}_n^2\boldsymbol{I}_{n\times n}$, where $\boldsymbol{\sigma}_s^2$ and $\boldsymbol{\sigma}_n^2$ are the signal and noise variance/power.

Lastly, the error covariance is

$$
\begin{aligned}
\boldsymbol{C}_{\boldsymbol{e}} &= \boldsymbol{C}_{\boldsymbol{X}} - \boldsymbol{C}_{\widehat{\boldsymbol{X}}} \\
&= \boldsymbol{C}_{\boldsymbol{X}} - \boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}}(\boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^{\boldsymbol{H}} + \boldsymbol{C}_{\boldsymbol{Z}})^{-1}\boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}
\end{aligned} \tag{2-46}
$$

The significant difference between the estimation problem treated above and those of least squares and Gauss-Markov estimate is that the number of observations *m*, (i.e. the dimension of) need not be at least as large as the number of unknowns, *n*, (i.e. the dimension of). The estimate for the linear observation process exists so long as the *m*-by-*m* matrix exists; this is the case for any *m* if, for instance, is positive definite. Physically the reason for this property is that since is now a random variable, it is possible to form a meaningful estimate (namely its mean) even with no measurements. Every new measurement simply provides additional information which may modify our original estimate. Another feature of this estimate is that for *m* < *n*, there need be no measurement error. Thus, we may have, because as long as is positive definite, the estimate still exists. Lastly, this technique can handle cases where the noise is correlated.

2.5 Alternative form

An alternative form of expression can be obtained by using the matrix identity

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{C_X A^H}(\boldsymbol{A C_X A^H} + \boldsymbol{C_Z})^{-1} \\ &= \left(\boldsymbol{A^H C_Z^{-1} A} + \boldsymbol{C_X^{-1}}\right)^{-1} \boldsymbol{A^H C_Z^{-1}} \end{aligned} \tag{2-47}$$

which can be established by post-multiplying by $(\boldsymbol{A C_X A^H} + \boldsymbol{C_Z})$ and pre-multiplying by $\left(\boldsymbol{A^H C_Z^{-1} A} + \boldsymbol{C_X^{-1}}\right)$, to obtain

$$\boldsymbol{W} = \left(\boldsymbol{A^H C_Z^{-1} A} + \boldsymbol{C_X^{-1}}\right)^{-1} \boldsymbol{A^H C_Z^{-1}} \tag{2-48}$$

And

$$\boldsymbol{C_e} = \left(\boldsymbol{A^H C_Z^{-1} A} + \boldsymbol{C_X^{-1}}\right)^{-1} \mathbf{x} \tag{2-49}$$

In this form the above expression can be easily compared with weighed least square and Gauss-Markov estimate.

In particular, when , corresponding to infinite variance of the apriori information concerning , the result is identical to the weighed linear least square estimate with as the weight matrix. Moreover, if the components of are uncorrelated and have equal variance such that where is an identity matrix, then is identical to the ordinary least square estimate.

3 Wiener filtering

3.1 Signal model

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{z} \tag{3-1}$$

Where $\boldsymbol{s}, \boldsymbol{z}$ are independent and identically distributed (i.i.d).

3.2 Solutions

The Wiener filter problem has solutions for three possible cases: one where a noncausal filter is acceptable (requiring an infinite amount of both past and future data), the case where a causal filter is desired (using an infinite amount of past data), and the finite impulse response (FIR) case where only input data is used (ie. the result or output is not fed back into the filter as in the IIR case). The first case is simple to solve but is not suited for real-time applications. Wiener's main accomplishment was solving the case where the causality requirement is in effect, and in an appendix of Wiener's book Levinson gave the FIR solution.

3.3 FIR wiener filtering

$$\hat{\boldsymbol{s}} = \boldsymbol{W}\boldsymbol{y} \tag{3-2}$$

Then error will be $\boldsymbol{e} = \hat{\boldsymbol{s}} - \boldsymbol{s}$

Cost function of mean square error

$$\boldsymbol{J} = \min_{\boldsymbol{w},\boldsymbol{b}} \boldsymbol{E}[\boldsymbol{e}\boldsymbol{e}^{\boldsymbol{H}}] \tag{3-3}$$

$$= \boldsymbol{E}[(\boldsymbol{Wy}-\boldsymbol{s})(\boldsymbol{Wy}-\boldsymbol{s})^{\boldsymbol{H}}]$$
$$= \boldsymbol{W}\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{W}^{\boldsymbol{H}} - \boldsymbol{W}\boldsymbol{C}_{\boldsymbol{YS}} - \boldsymbol{C}_{\boldsymbol{SY}}\boldsymbol{W}^{\boldsymbol{H}}$$

Calculate the derivation to $\boldsymbol{W}$ and force it to zero to obtain the MMSE $\boldsymbol{W}$

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{W}} = \boldsymbol{2}\boldsymbol{W}\boldsymbol{C}_{\boldsymbol{Y}} - \boldsymbol{2}\boldsymbol{C}_{\boldsymbol{YS}} = \boldsymbol{0} \tag{3-4}$$

Then

$$\boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{YS}}\boldsymbol{C}_{\boldsymbol{Y}}^{-\boldsymbol{1}} \tag{3-5}$$

Another form if assume

$$\hat{\boldsymbol{s}} = \boldsymbol{W}^{\boldsymbol{H}}\boldsymbol{y} \tag{3-6}$$

Then the FIR wiener filter is

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{C}_{\boldsymbol{Y}}^{-\boldsymbol{1}}\boldsymbol{C}_{\boldsymbol{SY}} \\ &= \boldsymbol{C}_{\boldsymbol{Y}}^{-\boldsymbol{1}}(\boldsymbol{C}_{\boldsymbol{Y}} - \boldsymbol{C}_{\boldsymbol{Z}}) \\ &= \boldsymbol{I} - \boldsymbol{C}_{\boldsymbol{Y}}^{-\boldsymbol{1}}\boldsymbol{C}_{\boldsymbol{Z}} \end{aligned} \tag{3-7}$$