Notes on Linear Minimum Mean Square Error Estimators

Çağatay Candan

January 2, 2011

Abstract

Some connections between linear minimum mean square error estimators, maximum output SNR filters and the least square solutions are presented. The notes have been prepared to be distributed with EE 503 (METU, Electrical Engin.) lecture notes.

1 Linear Minimum Mean Square Error Estimators

The following signal model is assumed:

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{v} \tag{1}$$

Here $\mathbf{r}$ is a N $\times$ 1 column vector denoting the observations. In this model, $\mathbf{s}$ is the desired signal vector, $\mathbf{v}$ is the interference vector on the observations. We assume that $\mathbf{s}$ and $\mathbf{v}$ are uncorrelated as in all stochastic filtering applications.

It can be noted that the observations are linear combinations of desired quantities and interference. Here the matrix $\mathbf{H}$ is an N $\times$ M matrix (N can be less than or greater than M, hence the system can be under or over determined). $\mathbf{H}$ can be considered as the channel or the observation system or a linear combiner for the modes where the modes are the columns of $\mathbf{H}$. The column vector $\mathbf{s}$ contains M entries each of which is to be estimated.

In class, we have derived the optimal linear minimum mean square error estimator as the solution of the following system of equations:

$$\mathbf{R_r}\mathbf{W} = \mathbf{R_{rs}} \tag{2}$$

Here $\mathbf{R_r} = E\{\mathbf{r}\mathbf{r^H}\}$ is the auto-correlation matrix of the observations and $\mathbf{R_{rs}} = E\{\mathbf{r}\mathbf{s^H}\}$ is the cross-correlation matrix of observations and the desired variables. The minimum mean square error estimate for $\mathbf{s}$ is given as follows:

$$\widehat{\mathbf{s}} = \mathbf{W^H}\mathbf{r} \tag{3}$$

The error covariance matrix for the minimum error estimator is as shown below:

$$\begin{aligned}
\mathbf{R_e} = E\{\underbrace{(\mathbf{s}-\widehat{\mathbf{s}})}_{\mathbf{e}}\underbrace{(\mathbf{s}-\widehat{\mathbf{s}})^{\mathbf{H}}}_{\mathbf{e^H}}\} = E\{\mathbf{e}(\mathbf{s^H}-\mathbf{r^H}\mathbf{W})\} &\overset{(a)}{=} E\{\mathbf{e}\mathbf{s^H}\} - \underbrace{E\{\mathbf{e}\mathbf{r^H}\}}_{\mathbf{0}}\mathbf{W} \\
&= E\{(\mathbf{s}-\mathbf{W^H}\mathbf{r})\mathbf{s^H}\} \\
&= \mathbf{R_s} - \mathbf{W^H}\mathbf{R_{rs}}
\end{aligned} \tag{4}$$

The zero matrix on the right hand side of equation shown with (a) is due to the orthogonality condition of the optimal estimator.

Next, we explicitly calculate the estimator in terms of $\mathbf{H}$, $\mathbf{R_s}$ and $\mathbf{R_v}$ matrices. The following can be easily verified:

$$\begin{aligned}
\mathbf{R_r} &= \mathbf{H}\mathbf{R_s}\mathbf{H^H} + \mathbf{R_v} \\
\mathbf{R_{rs}} &= \mathbf{H}\mathbf{R_s}
\end{aligned} \tag{5}$$

Then the optimal filter is given as follows:

$$\mathbf{W} = (\mathbf{H}\mathbf{R_s}\mathbf{H^H} + \mathbf{R_v})^{-1}\mathbf{H}\mathbf{R_s} \tag{6}$$

The estimator and its error covariance matrix is then

$$\begin{array}{rcl} & \text{Linear Min. MMSE Estimator (Form 1)} & \\ \widehat{\mathbf{s}} & = & \mathbf{W^H r} = \mathbf{R_s H^H}(\mathbf{H R_s H^H} + \mathbf{R_v})^{-1}\mathbf{r} \\ \mathbf{R_e} & = & \mathbf{R_s} - \mathbf{R_s H^H}(\mathbf{H R_s H^H} + \mathbf{R_v})^{-1}\mathbf{H R_s} \end{array} \tag{7}$$

This concludes the derivation of the estimator. In the following section, we present an alternative form which is mathematically identical but has some implementation advantages if $\mathbf{H}$ matrix is tall and long ($N > M$ or the number of observations > the number of unknowns) which is the case of interest for most of the time.

1.1 An Alternative Form for Linear MMSE Estimator

We follow the derivation given by Luenberger in [1, p.90].

The matrix identity

$$\mathbf{R W^H}(\mathbf{W R W^H} + \mathbf{Q})^{-1} = (\mathbf{W^H Q^{-1} W} + \mathbf{R^{-1}})^{-1}\mathbf{W^H Q^{-1}} \tag{8}$$

can be proven easily by pre and post multiplying both sides of the relation by $\mathbf{W^H Q^{-1} W} + \mathbf{R^{-1}}$ and $\mathbf{W R W^H} + \mathbf{Q}$ respectively. Using this identity we can write the following:

$$\mathbf{R_s H^H}(\mathbf{H R_s H^H} + \mathbf{R_v})^{-1} = (\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})^{-1}\mathbf{H^H R_v^{-1}} \tag{9}$$

Replacing the related terms in (7), we get the second form of the estimator:

$$\begin{array}{rcl} & \text{Linear Min. MMSE Estimator (Form 2)} & \\ \widehat{\mathbf{s}} & = & \mathbf{W^H r} = (\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})^{-1}\mathbf{H^H R_v^{-1} r} \\ \mathbf{R_e} & = & \mathbf{R_s} - (\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})^{-1}\mathbf{H^H R_v^{-1} H R_s} \\ & = & (\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})^{-1}\left((\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})\mathbf{R_s} - \mathbf{H^H R_v^{-1} H R_s}\right) \\ & = & (\mathbf{H^H R_v^{-1} H} + \mathbf{R_s^{-1}})^{-1} \end{array} \tag{10}$$

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

2 Maximum SNR Filters

In many applications, SNR determines the performance of a system. Therefore its improvement is a major goal in signal processing.

We assume that N observations of a WSS random process $s[n]$ are made under additive noise process $v[n]$ which is also WSS and uncorrelated with $s[n]$.

$$r[n] = s[n] + v[n] \tag{11}$$

The SNR of the observations is defined as the input SNR.

$$(\text{SNR})_{\text{in}} = \frac{E\{s^2[n]\}}{E\{v^2[n]\}} = \frac{\sigma_s^2}{\sigma_v^2} \tag{12}$$

(It should be noted that both $s[n]$ and $v[n]$ are assumed to be zero-mean processes as we have assumed during the lectures.)

As we know, the output of a LTI filter with the input $r[n]$ is known to be WSS stationary and its variance can be calculated as follows.

$$y[n] = \sum_{k=0}^{N-1} h_k^* r[n-k] = \mathbf{h^H r[n]} \tag{13}$$

Here we assume that the LTI filter is FIR and with the impulse response of h_n. The vectors $\mathbf{h}$ and $\mathbf{r[n]}$ appearing on the right most side of (13) are column vectors which are defined as $\mathbf{h^T} = [h_0 \dots h_{N-1}]^T$ and $\mathbf{r[n]^T} = [r[n]\, r[n-1] \dots r[n-(N-1)]\,]^T$.

The signal and noise term at the filter output can then be expressed as

$$y[n] = \underbrace{\mathbf{h^H s[n]}}_{\text{signal}} + \underbrace{\mathbf{h^H v[n]}}_{\text{noise}} \tag{14}$$

The output SNR becomes as follows:

$$(\text{SNR})_{\text{out}} = \frac{E\{|\mathbf{h^H s[n]}|^2\}}{E\{|\mathbf{h^H v[n]}|^2\}} = \frac{\mathbf{h^H R_s h}}{\mathbf{h^H R_v h}} \tag{15}$$

The filter maximizing the output SNR can be found as the solution of the following optimization problem:

$$\mathbf{h}_* \quad = \quad \arg\max_h \frac{\mathbf{h^H R_s h}}{\mathbf{h^H R_v h}} \tag{16}$$

The filter maximizing this ratio is called as the maximum SNR filter $\mathbf{h}_*$ and it can be shown that it is the eigenvector of $\mathbf{R_v^{-1} R_s}$ with the maximum eigenvalue.

$$\mathbf{h}_* \propto \text{eigenvector of } (\mathbf{R_v^{-1} R_s}) \text{ with the maximum eigenvalue} \tag{17}$$

It should be noted that $\mathbf{h}_*$ is uniquely determined apart from a constant scaling. It is clear that by scaling $\mathbf{h}_*$ with an arbitrary complex constant, we can achieve the same output SNR.

More information on the discussed optimization problem can be found in the linear algebra books under the topic of Rayleigh quotient. It should also be noted that $\mathbf{h}_*$ is also called the generalized eigenvector of $\mathbf{R_v}$ and $\mathbf{R_s}$ with the maximum eigenvalue. (Matlab's "eig.m" function has an option of calculating the generalized eigenvectors.)

This concludes the discussion of maximum-SNR filters. Below we examine a special case which naturally emerges in many applications.

In many applications, the auto-correlation matrix of $s[n]$ is rank-1, that is

$$\mathbf{R_s} = \sigma_s^2 \mathbf{p p^H} \tag{18}$$

For this special case, the maximum SNR filter and the maximum SNR value can be written as follows:

$$\begin{aligned} \mathbf{h}_* &\propto \mathbf{R_v^{-1} p} \\ \max(\text{SNR})_{\text{out}} &= \mathbf{p^H R_v^{-1} p} \sigma_s^2 \end{aligned} \tag{19}$$

It should be noted that the maximum-SNR filter is well known whitened matched filter for this case. It is the matched filter for the white noise case.

3 Connection Between Max-SNR and Min-MMSE Filters for Processes with Rank-1 Spectral Expansion

It is intuitively clear that filtering the observations to remove the noise is helpful to improve the SNR. In this section, we show that the min-MMSE filters and Max-SNR filters are (apart from a constant *complex* scaling) identical for rank-1 $\mathbf{R_s}$ matrices. Hence min-MMSE filter is equivalent to Max-SNR filter in the sense of optimizing the SNR for this type of processes.

Let s[n] be a rank-1 process observed under additive white noise,

$$\mathbf{r} = \mathbf{p}s + \mathbf{v} \tag{20}$$

Here $\mathbf{r}, \mathbf{p}$ and $\mathbf{v}$ are length N column vectors; s is a zero-mean scalar with variance σ_s^2 and $\mathbf{v}$ is a complex valued process with zero-mean and covariance matrix $\mathbf{R_v}$.

By substituting $\mathbf{H} \rightarrow \mathbf{p}$, $\mathbf{R_s} \rightarrow \sigma_s^2$ in (10), we get

$$\widehat{s} \;=\; \mathbf{w}_{\text{MMSE}}^{\mathbf{H}}\mathbf{r} = \frac{1}{\mathbf{p^H R_v^{-1} p} + \frac{1}{\sigma_s^2}} \underbrace{\mathbf{p^H R_v^{-1}}}_{\mathbf{w}_{\mathbf{max-SNR}}^{\mathbf{H}}} \mathbf{r}. \tag{21}$$

This equation shows that the min-MMSE filter is the scaled version of the max-SNR filter. Hence they yield the same output SNR.

The minimum error of MMSE filter can also be written from (10):

$$\begin{aligned} \sigma_e^2 &= \frac{1}{(\mathbf{p^H R_v p} + \frac{1}{\sigma_s^2})} \\ &= \frac{\sigma_s^2}{\text{Max(SNR)}_{\text{out}} + 1} \end{aligned} \tag{22}$$

Here we have identified $\mathbf{p^H R_v p}\sigma_s^2$ as the maximum SNR for rank-1 stochastic signals, see (19).

The last equation can be written in the following simple form

$$\text{Max(SNR)}_{\text{out}} = \frac{1}{\text{norm-min-MMSE}} - 1 \tag{23}$$

where "norm-min-MMSE" is $\frac{\sigma_e^2}{\sigma_s^2}$ which is the normalized minimum mean square estimation error. The generalized version of similar results are available in [2, around eq.25].

4 Connection Between MMSE and Weighted Least Square Filters

The minimum mean square error filter always produces an output even when the number of observations is less than the number of unknowns (undertermined case). This is due to a-priori moment information about the observations and desired process.

As the a-priori information gets looser, that is less informing, the minimum mean square error filter approaches weighted least square (LS) filter. The weighted LS filter is known to be optimal for the estimation of non-random parameters which are observed with linear models under correlated noise, [3].

The convergence of (MMSE) → (weighted LS) can be noted from (10) by substituting $\mathbf{R_s} \rightarrow \infty\mathbf{I}$:

$$\widehat{\mathbf{s}} \;=\; \mathbf{W}_{\mathbf{MMSE}}^{\mathbf{H}}\mathbf{r} = \underbrace{(\mathbf{H^H R_v^{-1} H})^{-1}\mathbf{H^H R_v^{-1}}}_{\text{weighted LS}}\,\mathbf{r} \tag{24}$$

The last equation is the weighted least square solution of $\mathbf{r} = \mathbf{Hs}$ equation system with the weight matrix $\mathbf{R_v^{-1}}$. The mean square error covariance matrix becomes

$$\mathbf{R_e} = (\mathbf{H^H R_v^{-1} H})^{-1} \tag{25}$$

When $\mathbf{R_v}$ is taken as unit variance white noise, the classical least square solution emerges. In communications, the LS solution is also known as the zero-forcing solution.

References

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

[2] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2198–2206, 2001.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory* . Prentice Hall, 1993.