MIMO MMSE Esitmation And SNR

November 21, 2021

1 Scalar Channel

We start from simple scalar case. We want to show an important formula

$$\frac{\sigma_x^2}{\text{MMSE}} - 1 = \text{SNR} \tag{1}$$

Consider a scalar channel relating the input x and output y by

$$y = hx + w$$

where h is a known scalar (that's why it called scalar channel) and x, w are zero-mean uncorrelated random scalars. They are all complex numbers.

First we find the SNR. By definition

$$\text{SNR} = \frac{\mathcal{E}\{|hx|^2\}}{\mathcal{E}\{|w|^2\}} = \frac{|h|^2\sigma_x^2}{\sigma_w^2} = \rho|h|^2 \tag{2}$$

where we have defined the transmit SNR $\rho = \sigma_x^2/\sigma_w^2$.

For LMMSE estimation, we are seeking an estimate of x, $\hat{x} = cy$, to minimize the mean square error $\mathcal{E}|\hat{x} - x|^2$. In other word, we are decomposing x into

$$x = cy + z \tag{3}$$

and trying to find the c to minimize σ_z^2.

Theorem 1 (Scalar version of orthogonality principle)**.** *Given random variables x and y, we decompose x into $x = cy + z$ for some random z. If $\mathcal{E}zy^* = 0$, then for any a we have*

$$\mathcal{E}|ay + z|^2 \geq \sigma_z^2$$

Proof. Since y and z are orthogonal (uncorrelated)

$$\mathcal{E}|ay+z|^2 = |a|^2\sigma_y^2 + \sigma_z^2$$

□

So we can apply orthogonality principle by taking correlation with y on both sides of equation 3

$$\mathcal{E}xy^* = c\sigma_y^2$$

Hence we get

$$c = \frac{\sigma_{xy}}{\sigma_y^2}$$

Now we can calculate the MMSE

$$\begin{aligned}
\mathcal{E}|cy-x|^2 &= \mathcal{E}(cy-x)(cy-x)^* \\
&= |c|^2\sigma_y^2 + \sigma_x^2 - c\sigma_{yx} - c^*\sigma_{xy}
\end{aligned}$$

Note that

$$\sigma_{xy}^* = (\mathcal{E}xy^*)^* = \mathcal{E}yx^* = \sigma_{yx}$$

we can simplify MMSE to be

$$\mathcal{E}|cy-x|^2 = \sigma_x^2 - \sigma_{xy}\sigma_{yx}/\sigma_y^2$$

Also note that

$$\sigma_y^2 = |h|^2\sigma_x^2 + \sigma_w^2$$
$$\sigma_{xy} = h^*\sigma_x^2$$

we can further simplify MMSE

$$\begin{aligned}
\mathcal{E}|cy-x|^2 &= \sigma_x^2 - \frac{\sigma_x^4|h|^2}{\sigma_x^2|h|^2+\sigma_w^2} \\
&= \frac{\sigma_x^2\sigma_w^2}{\sigma_x^2|h|^2+\sigma_w^2} \\
&= \frac{\sigma_x^2}{\rho|h|^2+1} \qquad (4)
\end{aligned}$$

Compare equations 2 and 4, we get the relation between MMSE and SNR for scalar channel as in equation 1.

2 Vector Channel

Before we study the LMMSE of vector channel, we first take a look at SNR. In scalar channel, it is obvious that linear estimators do not affect SNR, but it is not the case in vector channel.

2.1 Matched Filter

Consider a vector channel

$$\boldsymbol{y} = \boldsymbol{h}x + \boldsymbol{w}$$

where $\boldsymbol{h}$ is a known vector, x is a zero-mean random scalar, $\boldsymbol{w}$ is a zero-mean random vector uncorrelated with x.

We are still looking for a linear estimate of x

$$\hat{x} = \langle \boldsymbol{y}, \boldsymbol{c} \rangle = \boldsymbol{c}^* \boldsymbol{h} x + \boldsymbol{c}^* \boldsymbol{w} \tag{5}$$

for some vector $\boldsymbol{c}$.

The SNR can be defined by

$$\gamma = \frac{\mathcal{E}||\boldsymbol{c}^* \boldsymbol{h} x||^2}{\mathcal{E}||\boldsymbol{c}^* \boldsymbol{w}||^2} = \frac{||\boldsymbol{c}^* \boldsymbol{h}||^2 \sigma_x^2}{\boldsymbol{c}^* \boldsymbol{K_w} \boldsymbol{c}} \tag{6}$$

where $\boldsymbol{K_w}$ is the covariance matrix of $\boldsymbol{w}$.

The choice of $\boldsymbol{c}$ affects γ. To maximize γ, we define $\boldsymbol{d} = \boldsymbol{K_w}^{1/2} \boldsymbol{c}$ (the square root of a positive definite matrix is well-defined) and $\boldsymbol{h}' = \boldsymbol{K_w}^{-1/2} \boldsymbol{h}$ and substitute into equation 6, then we get

$$\begin{aligned}
\gamma &= \frac{||\boldsymbol{d}^* \boldsymbol{h}'||^2 \sigma_x^2}{||\boldsymbol{d}||^2} \\
&\leq \frac{||\boldsymbol{d}||^2 ||\boldsymbol{h}'||^2 \sigma_x^2}{||\boldsymbol{d}||^2} \qquad (7) \\
&= ||\boldsymbol{h}'||^2 \sigma_x^2 \\
&= \boldsymbol{h}^* \boldsymbol{K_w}^{-1} \boldsymbol{h} \sigma_x^2 \qquad (8)
\end{aligned}$$

Equation 8 does not depend on $\boldsymbol{c}$ and hence gives us the upper bound of γ. 7 is by Cauchy-Riemann inequality with the equality if $\boldsymbol{d}$ is parallel to $\boldsymbol{h}'$, that is, there exists a complex scalar λ such that

$$\boldsymbol{c} = \lambda \boldsymbol{K_w}^{-1} \boldsymbol{h} \tag{9}$$

Therefore, any linear estimator that achieves maximum SNR must satisfy equation 9. This class of linear estimators has a name called matched filter.

Let's take a look at what is going on in the matched filter. Substitute equation 9 into 5 (WLOG, set $\lambda = 1$) and we get

$$\hat{x} = \boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h} x + \boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{w}$$

The noise $\boldsymbol{w}$ is first whitened by $\boldsymbol{K}_{\boldsymbol{w}}^{-1}$ and then projected along $\boldsymbol{h}$. This has a good geometric interpretation. If the noise were white, i.e., $\boldsymbol{K}_{\boldsymbol{w}} = \boldsymbol{I}$, then along any direction its energy stays the same, and the optimal linear estimator is of course by projecting the received signal onto the direction of the channel vector.

2.2 LMMSE For Vector Channel

There are two approaches to get the LMMSE estimator of vector channel. We will show that the results coincide.

2.2.1 Scalarized By Matched Filter

We can first scalarize the vector channel by taking inner product with some $\boldsymbol{c}$, and get MMSE($\boldsymbol{c}$) and SNR($\boldsymbol{c}$) by LMMSE for scalar channel. By equation 1, the minimal MMSE($\boldsymbol{c}$) over $\boldsymbol{c}$ can be found by making SNR($\boldsymbol{c}$) maximal, and the optimal $\boldsymbol{c}$ is simply the matched filter.

Choose an arbitrary $\boldsymbol{c}$, we get a scalar channel

$$y' = h'x + w'$$

where $y' = \langle \boldsymbol{y}, \boldsymbol{c} \rangle$, $h' = \langle \boldsymbol{h}, \boldsymbol{c} \rangle$, $w' = \langle \boldsymbol{w}, \boldsymbol{c} \rangle$. The LMMSE estimator for this scalar channel is

$$\lambda = \frac{\sigma_{xy'}}{\sigma_{y'}^2} = \frac{\sigma_x^2 h'^*}{\sigma_x^2 h' h'^* + \sigma_{w'}^2} = \frac{\sigma_x^2 \boldsymbol{h}^* \boldsymbol{c}}{\boldsymbol{c}^* (\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}) \boldsymbol{c}}$$

We can choose an arbitrary matched filter from class 9 such as

$$\boldsymbol{c} = \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}$$

Hence the LMMSE estimator of vector channel is

$$\boldsymbol{a} = \lambda^* \boldsymbol{c} = \frac{\sigma_x^2 \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}\boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}}{\boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \left(\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}\right) \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}} \tag{10}$$

where the conjugation of λ is due to the fact that $\lambda\langle \boldsymbol{y}, \boldsymbol{c}\rangle = \langle \boldsymbol{y}, \lambda^* \boldsymbol{c}\rangle$.

Since we are using matched filter, the resulting SNR is the same with equation 8. Also, as we are dealing with scalar channel, by equation 1, the achieved LMMSE is

$$\text{MMSE} = \frac{\sigma_x^2}{\gamma + 1} = \frac{\sigma_x^2}{\sigma_x^2 \boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h} + 1}$$

2.2.2 Direct LMMSE For Vector Channel

Theorem 2 (Vector version of orthogonality principle)**.** *Given random scalar x and vector $\boldsymbol{y}$, we decompose x into*

$$x = \langle \boldsymbol{y}, \boldsymbol{a}\rangle + z \tag{11}$$

for some random z. If $\mathcal{E} z \boldsymbol{y}^* = \boldsymbol{0}$, then for any $\boldsymbol{c}$ we have

$$\mathcal{E}|\langle \boldsymbol{y}, \boldsymbol{c}\rangle + z|^2 \geq \sigma_z^2$$

Proof. Since $\boldsymbol{y}$ and z are uncorrelated, we have

$$\mathcal{E}|\langle \boldsymbol{y}, \boldsymbol{c}\rangle + z|^2 = \boldsymbol{c}^* \boldsymbol{K}_{\boldsymbol{y}} \boldsymbol{c} + \sigma_z^2$$

it follows from the positive semi-definiteness of $\boldsymbol{K}_{\boldsymbol{y}}$. □

To find the optimal $\boldsymbol{a}$, we correlate $\boldsymbol{y}$ with equation 11 and get

$$\mathcal{E} x \boldsymbol{y}^* = \boldsymbol{a}^* \boldsymbol{K}_{\boldsymbol{y}}$$

Hence we have

$$\boldsymbol{a} = \boldsymbol{K}_{\boldsymbol{y}}^{-1} \mathcal{E} \boldsymbol{y} x^* = \sigma_x^2 \left(\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}\right)^{-1} \boldsymbol{h} \tag{12}$$

Next we will show that equation 10 and 12 coincide.

Note that this is equivalent to justifying

$$\left(\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}\right) \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}\boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h} = \boldsymbol{h}\boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \left(\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}\right) \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{h}$$

This equation is due to the fact that $\left(\sigma_x^2 \boldsymbol{h}\boldsymbol{h}^* + \boldsymbol{K}_{\boldsymbol{w}}\right) \boldsymbol{K}_{\boldsymbol{w}}^{-1}$ and $\boldsymbol{h}\boldsymbol{h}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1}$ commute, which is not hard to see by expanding the former.

3 Matrix Channel

A matrix channel is a MIMO channel given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \tag{13}$$

where $\boldsymbol{x}$ and $\boldsymbol{w}$ are zero-mean uncorrelated random vectors and are not necessarily of the same dimension, and $\boldsymbol{H}$ is a matrix.

Again we will show two approaches to obtain LMMSE estimator, and prove that the two results coincide.

3.1 Matched Filter On Each Stream

Equation 13 is in fact the sum of vector channels. To see this, assume the dimensions of x and y are M and N respectively. Then it can be rewritten as

$$\boldsymbol{y} = \sum_{i=1}^{M} \boldsymbol{h}_i x^i + \boldsymbol{w}$$

And hence the channel seen from the k-th stream is

$$\boldsymbol{y} = \boldsymbol{h}_k x^k + \left(\sum_{i \neq k} \boldsymbol{h}_i x^i + \boldsymbol{w} \right)$$

where the parenthesized noise term consists of pure noise plus interference from other streams. We denote it with $\boldsymbol{w}_k$. The LMMSE for vector channels can be applied on the k-th stream

$$\begin{aligned}
\boldsymbol{a}_k &= \sigma_{x_k}^2 \left(\sigma_{x_k}^2 \boldsymbol{h}_k \boldsymbol{h}_k^* + \boldsymbol{K}_{\boldsymbol{w}_k} \right)^{-1} \boldsymbol{h}_k \\
&= \sigma_{x_k}^2 \left(\sigma_{x_k}^2 \boldsymbol{h}_k \boldsymbol{h}_k^* + \sum_{(i,j) \neq (k,k)} \sigma_{x_i x_j} \boldsymbol{h}_i \boldsymbol{h}_j^* + \boldsymbol{K}_{\boldsymbol{w}} \right)^{-1} \boldsymbol{h}_k \\
&= \sigma_{x_k}^2 \left(\sum_{i,j} \sigma_{x_i x_j} \boldsymbol{h}_i \boldsymbol{h}_j^* + \boldsymbol{K}_{\boldsymbol{w}} \right)^{-1} \boldsymbol{h}_k \\
&= \sigma_{x_k}^2 \left(\boldsymbol{H} \boldsymbol{K}_{\boldsymbol{x}} \boldsymbol{H}^* + \boldsymbol{K}_{\boldsymbol{w}} \right)^{-1} \boldsymbol{h}_k
\end{aligned} \tag{14}$$

By equation 8, the SNR of stream k is

$$\gamma_k = \sigma_{x_k}^2 \boldsymbol{h}_k^* \left(\sum_{(i,j) \neq (k,k)} \sigma_{x_i x_j} \boldsymbol{h}_i \boldsymbol{h}_j^* + \boldsymbol{K}_{\boldsymbol{w}} \right)^{-1} \boldsymbol{h}_k \tag{15}$$

3.2 Direct LMMSE For Matrix Channel

Theorem 3 (Matrix version of orthogonality principle)**.** *Given zero-mean random vectors* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *which are not necessarily of the same dimension, we decompose* $\boldsymbol{x}$ *into*

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{z} \tag{16}$$

for some random vector $\boldsymbol{z}$*. If* $\mathcal{E}\boldsymbol{z}\boldsymbol{y}^* = \boldsymbol{0}$*, then for any* $\boldsymbol{G}$ *(dimension compatible with* x *and* y*) we have*

$$Cov(\boldsymbol{G}\boldsymbol{y} + \boldsymbol{z}) \succeq \boldsymbol{K_z}$$

where $Cov(\cdot)$ *denotes taking covariance matrix,* $\succeq$ *is the partial ordering induced by positive semi-definite cone.*

Proof. Since $\boldsymbol{z}$ and $\boldsymbol{y}$ are uncorrelated, we have

$$\mathrm{Cov}(\boldsymbol{G}\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{G}\boldsymbol{K_y}\boldsymbol{G}^* + \boldsymbol{K_z}$$

The proof follows from the positive semi-definiteness of $\boldsymbol{G}\boldsymbol{K_y}\boldsymbol{G}^*$. □

We take correlation with $\boldsymbol{y}$ on both sides of equation 16 to get

$$\boldsymbol{A} = \mathcal{E}\boldsymbol{x}\boldsymbol{y}^*\boldsymbol{K_y}^{-1} \tag{17}$$

substitute with equation 13

$$\boldsymbol{A} = \boldsymbol{K_x}\boldsymbol{H}^*\left(\boldsymbol{H}\boldsymbol{K_x}\boldsymbol{H^*} + \boldsymbol{K_w}\right)^{-1} \tag{18}$$

Claim 1 (Matrix LMMSE)**.** *The LMMSE of the matrix channel 13 is*

$$MMSE = \boldsymbol{K_x} - \mathcal{E}\boldsymbol{x}\boldsymbol{y}^*\boldsymbol{K_y}^{-1}\mathcal{E}\boldsymbol{y}\boldsymbol{x}^* \tag{19}$$

Proof. Simply substitute equation 17 into $\mathrm{Cov}(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{y})$. □

We can expand equation 19 into

$$\mathrm{MMSE} = \boldsymbol{K_x} - \boldsymbol{K_x}\boldsymbol{H}^*\left(\boldsymbol{H}\boldsymbol{K_x}\boldsymbol{H}^* + \boldsymbol{K_w}\right)^{-1}\boldsymbol{H}\boldsymbol{K_x}$$

Apply matrix inversion identity (without proof)

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

with $A = \boldsymbol{K}_{\boldsymbol{x}}^{-1}$, $B = \boldsymbol{H}^*$, $C = \boldsymbol{K}_{\boldsymbol{w}}^{-1}$, $D = \boldsymbol{H}$, and we arrive at

$$\mathrm{MMSE} = \left(\boldsymbol{K}_{\boldsymbol{x}}^{-1} + \boldsymbol{H}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{H}\right)^{-1}$$

We have to admit that this simplification is tricky.

The MMSE for k-th stream is by definition

$$\mathcal{E}\left|[\boldsymbol{x}]^k - [\boldsymbol{A}\boldsymbol{y}]^k\right|^2 = [\mathrm{Cov}(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{y})]_k^k$$

Therefore the SNR of k-th stream is

$$\gamma_k = \frac{\sigma_{x_k}^2}{\left[\left(\boldsymbol{K}_{\boldsymbol{x}}^{-1} + \boldsymbol{H}^* \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{H}\right)^{-1}\right]_k^k} - 1 \tag{20}$$

Next we will show the equivalence between two approaches. Let $\boldsymbol{e}_k$ be the k-th standard basis vector, then the estimate of x_k is

$$\begin{aligned}
\hat{x}_k &= \boldsymbol{e}_k^* \hat{\boldsymbol{x}} \\
&= \boldsymbol{e}_k^* \boldsymbol{A}\boldsymbol{y} \\
&= \boldsymbol{e}_k^* \mathcal{E}\boldsymbol{x}\boldsymbol{y}^* \boldsymbol{K}_{\boldsymbol{y}}^{-1} \boldsymbol{y} \\
&= \mathcal{E}\boldsymbol{e}_k^* \boldsymbol{x}\boldsymbol{y}^* \boldsymbol{K}_{\boldsymbol{y}}^{-1} \boldsymbol{y} \\
&= \mathcal{E}x_k \boldsymbol{y}^* \boldsymbol{K}_{\boldsymbol{y}}^{-1} \boldsymbol{y} \\
&= \sigma_{x_k}^2 \boldsymbol{h}_k^* \left(\boldsymbol{H}\boldsymbol{K}_{\boldsymbol{x}}\boldsymbol{H}^* + \boldsymbol{K}_{\boldsymbol{w}}\right)^{-1} \boldsymbol{y} \qquad (21)
\end{aligned}$$

Compared to equation 14, equation 21 is exactly

$$\hat{x}_k = \boldsymbol{a}_k^* \boldsymbol{y}$$

This implies that taking MMSE with respect to positive semi-definite partial ordering is equivalent to taking MMSE for each stream, which further justifies the coincidence between equation 20 and 15, even though the two γ_k's look far away from each other.