Proof of Two Conclusions Associated with Linear Minimum Mean Square Estimation by Matrix Inverse Lemma[1]

Jianping Zheng

State Key Lab of ISN, Xidian University, Xi'an, 710071, P. R. China

jpzheng@xidian.edu.cn

May 6, 2017

Linear minimum mean square Error (LMMSE) estimation is a classic estimate algorithm. Here, the matrix inverse lemma is applied to proof two associated conclusions. The first is that the joint LMMSE estimation is information theoretically optimal in linear Gaussian channels. The second is that the LMMSE estimation is equivalent to the estimation consisting of noise whitening and match filter (NW-MF) in linear Gaussian channels.

I. LMMSE for Linear Gaussian Channels

Consider the linear Gaussian channel presented by

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{w} \tag{1}$$

Here, $\mathbf{x}$ is an N-dimensional state vector, $\mathbf{y}$ is a M-dimensional ($M \geq N$) observation vector, $\mathbf{H}$ is observation matrix, and $\mathbf{w}$ is a M-dimensional additive Gaussian white noise (AWGN) vector with $\mathbf{w} \sim \mathcal{CN}\left(\mathbf{0}, N_0\mathbf{I}_M\right)$.

Given the first two moments of $\mathbf{x}$ and $\mathbf{y}$, the LMMSE estimation of $\mathbf{x}$ can be expressed by, from [1],

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_x + \mathbf{K}_{xy}\mathbf{K}_y^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_y\right) \tag{2}$$

Here, $\boldsymbol{\mu}_x = \mathrm{E}\left[\mathbf{x}\right]$ and $\boldsymbol{\mu}_y = \mathrm{E}\left[\mathbf{y}\right]$ are the expectation values of $\mathbf{x}$ and $\mathbf{y}$, respectively,

---

[1] Citation:

Jianping Zheng, Proof of Two Conclusions Associated with Linear Minimum Mean Square Estimation by Matrix Inverse Lemma, *Lecture Note,* http://web.xidian.edu.cn/jpzheng/teach.html

$\mathbf{K}_{xy} = \mathrm{E}\left[\left(\mathbf{x} - \boldsymbol{\mu}_x \right) \left(\mathbf{y} - \boldsymbol{\mu}_y \right)^H \right]$ is the cross-correlation matrix of **x** and **y**, and $\mathbf{K}_{y} = \mathrm{E}\left[\left(\mathbf{y} - \boldsymbol{\mu}_y \right) \left(\mathbf{y} - \boldsymbol{\mu}_y \right)^H \right]$ is the autocorrelation matrix of **y**.

II. Matrix Inverse Lemma

Lemma 1 (**matrix inverse lemma** [2]) For matrices **A**, **B**, **C** and **D** with proper sizes, it has

$$\left(\mathbf{A} + \mathbf{BCD} \right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B} \right)^{-1} \mathbf{DA}^{-1} \tag{3}$$

if **A** is invertible. A simplified version of this matrix inverse lemma is

$$\left(\mathbf{A} + c\mathbf{bb}^H \right)^{-1} = \mathbf{A}^{-1} - \frac{c\mathbf{A}^{-1}\mathbf{bb}^H\mathbf{A}^{-1}}{1 + c\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \tag{4}$$

where **b** is a vector, c is a scalar, and $\|\mathbf{b}\|$ is the Euclidean norm of **b**.

III. The Optimality of Joint LMMSE

The information theoretic optimality of LMMSE with successive decoding in the linear Gaussian channel has been discussed in several scenarios [3]-[6]. However, the direct proof of the optimality of joint LMMSE detection has not been reported, to the best of our knowledge. Here, a proof is given by utilizing the relation of differential entropy and the determinant of autocorrelation matrix and the matrix inverse lemma. In this proof, the linear Gaussian multiple-input multiple-output (MIMO) is taken as the practical example to facilitate the presentation.

Theorem 1: The Joint LMMSE estimation is information theoretically optimal for linear Gaussian channels.

Proof:

Take the linear Gaussian MIMO channel as example. The **x**, **y** and **H** in (1) can be interpreted as the transmit signal, receive signal, and MIMO channel matrix, respectively. The mutual information of **x** and **y** conditioned on **H** is, from [7],

$$I\left(\mathbf{x}; \mathbf{y} | \mathbf{H} \right) = \log \left| \mathbf{I}_M + \frac{1}{N_0} \mathbf{HK}_x\mathbf{H}^H \right| \tag{5}$$

with $|\cdot|$ denoting the determinant of the matrix argument.

On the other hand, as shown in Fig. 1, consider the concatenated system by the linear

Gaussian channel and LMMSE estimator. The input and the output variables are $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively. Therefore, the proof of Theorem 1 can be replaced by the proof of $I(\mathbf{x};\mathbf{y}|\mathbf{H}) = I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{H})$, i.e., the LMMSE estimator is information-looseness.

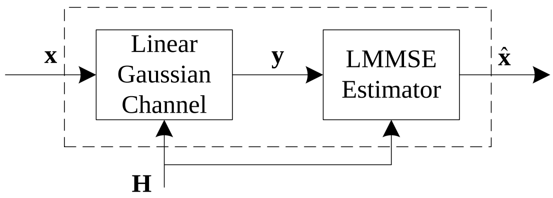

Fig. 1 LMMSE estimator in fixed Gaussian MIMO channel.

Define the estimation error as

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}} \tag{6}$$

Then, $I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{H})$ can be computed by

$$\begin{aligned} I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{H}) &= h(\mathbf{x}|\mathbf{H}) - h(\mathbf{x}|\hat{\mathbf{x}},\mathbf{H}) \\ &= h(\mathbf{x}) - h(\hat{\mathbf{x}}+\tilde{\mathbf{x}}|\hat{\mathbf{x}}) \\ &= h(\mathbf{x}) - h(\tilde{\mathbf{x}}) \\ &= N\log\left(\pi e|\mathbf{K}_x|^{1/N}\right) - N_t\log\left(\pi e|\mathbf{K}_{\tilde{x}}|^{1/N}\right) \\ &= \log\frac{|\mathbf{K}_x|}{|\mathbf{K}_{\tilde{x}}|} \end{aligned} \tag{7}$$

Here $h(\cdot)$ denotes the differential entropy of the argument, $\mathbf{K}_x$ and $\mathbf{K}_{\tilde{x}}$ are the autocorrelation matrices of $\mathbf{x}$ and $\tilde{\mathbf{x}}$, respectively. The fourth line follows from that the differential entropy per complexity dimension is

$$\frac{h(\mathbf{x})}{N} = \log\left(\pi e|\mathbf{K}_x|^{1/N}\right) \qquad \frac{h(\tilde{\mathbf{x}})}{N} = \log\left(\pi e|\mathbf{K}_{\tilde{x}}|^{1/N}\right) \tag{8}$$

In the LMMSE estimation, from (1), it has

$$\mathbf{K}_{\tilde{x}} = \mathbf{K}_x - \mathbf{K}_{xy}\mathbf{K}_y^{-1}\mathbf{K}_{yx} \tag{9}$$

with

$$\mathbf{K}_{xy} = \mathbf{K}_x\mathbf{H}^H = \mathbf{K}_{yx}^H \tag{10}$$

$$\mathbf{K}_y = \mathbf{H}\mathbf{K}_x\mathbf{H}^H + N_0\mathbf{I}_M \tag{11}$$

Then, (7) can be further computed by

$$
\begin{aligned}
I\left(\mathbf{x};\hat{\mathbf{x}}\middle|\mathbf{H}\right) &= \log\frac{|\mathbf{K}_x|}{|\mathbf{K}_{\tilde{x}}|} = \log\frac{|\mathbf{K}_x|}{\left|\mathbf{K}_x - \mathbf{K}_{xy}\mathbf{K}_y^{-1}\mathbf{K}_{yx}\right|} \\
&= \log\frac{|\mathbf{K}_x|}{\left|\mathbf{K}_x - \mathbf{K}_x\mathbf{H}^H\left(\mathbf{H}\mathbf{K}_x\mathbf{H}^H + N_0\mathbf{I}_M\right)^{-1}\mathbf{H}\mathbf{K}_x\right|} \\
&= \log\frac{|\mathbf{K}_x|}{|\mathbf{K}_x|\left|\mathbf{I}_N - \mathbf{H}^H\left(\mathbf{H}\mathbf{K}_x\mathbf{H}^H + N_0\mathbf{I}_M\right)^{-1}\mathbf{H}\mathbf{K}_x\right|} \\
&= \log\frac{|\mathbf{K}_x|}{|\mathbf{K}_x|\left|\mathbf{I}_N - \mathbf{H}^H\left(\mathbf{H}\mathbf{K}_x^{1/2}\mathbf{K}_x^{1/2}\mathbf{H}^H + N_0\mathbf{I}_M\right)^{-1}\mathbf{H}\mathbf{K}_x^{1/2}\mathbf{K}_x^{1/2}\right|} \\
&\overset{\text{(i)}}{=} \log\frac{1}{\left|\mathbf{I}_N - N_0^{-1}\mathbf{K}_x^{1/2}\mathbf{H}^H\left(N_0^{-1}\mathbf{H}\mathbf{K}_x^{1/2}\mathbf{K}_x^{1/2}\mathbf{H}^H + \mathbf{I}_M\right)^{-1}\mathbf{H}\mathbf{K}_x^{1/2}\right|} \\
&\overset{\text{(ii)}}{=} \log\frac{1}{\left|\left(\mathbf{I}_N + N_0^{-1}\mathbf{K}_x^{1/2}\mathbf{H}^H\mathbf{H}\mathbf{K}_x^{1/2}\right)^{-1}\right|} \\
&\overset{\text{(iii)}}{=} \log\left|\mathbf{I}_N + N_0^{-1}\mathbf{K}_x^{1/2}\mathbf{H}^H\mathbf{H}\mathbf{K}_x^{1/2}\right| \\
&\overset{\text{(iv)}}{=} \log\left|\mathbf{I}_M + N_0^{-1}\mathbf{H}\mathbf{K}_x^{1/2}\mathbf{K}_x^{1/2}\mathbf{H}^H\right| \\
&= \log\left|\mathbf{I}_M + N_0^{-1}\mathbf{H}\mathbf{K}_x\mathbf{H}^H\right| = I\left(\mathbf{x};\mathbf{y}\middle|\mathbf{H}\right)
\end{aligned} \tag{12}
$$

Here, (i) and (iv) follow from $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$, (ii) is from the matrix inverse lemma (3) with $\mathbf{A}=\mathbf{I}_N$, $\mathbf{B}=N_0^{-1}\mathbf{K}_x^{1/2}\mathbf{H}^H$, $\mathbf{C}=\mathbf{I}_M$ and $\mathbf{D}=\mathbf{H}\mathbf{K}_x^{1/2}$, and (iii) is from $\left|\mathbf{A}^{-1}\right| = |\mathbf{A}|^{-1}$. □

IV. The Equivalence of LMMSE and NW-MF

Define x_k as the k-th entry of the N-dimensional state vector $\mathbf{x}$, i.e., $\mathbf{x}=(x_1, x_2, \ldots, x_N)^T$. Consider the LMMSE estimation of x_k. It is more convenient to rewrite (1) as

$$
\begin{aligned}
\mathbf{y} &= \mathbf{h}_k x_k + \sum_{j\neq k}\mathbf{h}_j x_j + \mathbf{w} \\
&= \mathbf{h}_k x_k + \mathbf{z}
\end{aligned} \tag{13}
$$

with $\mathbf{z} = \sum_{j\neq k}\mathbf{h}_j x_j + \mathbf{w}$. Without loss of generality, $\mathbf{x}$ is assumed to be zero-mean, i.e., $\boldsymbol{\mu}_x = \mathbf{0}$. Define the autocorrelation matrix of $\mathbf{z}$ by

$$
\mathbf{K}_z = \mathrm{E}\left(\left(\mathbf{z}-\boldsymbol{\mu}_z\right)\left(\mathbf{z}-\boldsymbol{\mu}_z\right)^H\right) = \mathrm{E}\left[\mathbf{z}\mathbf{z}^H\right] = \sum_{j\neq k} P_j\mathbf{h}_j\mathbf{h}_j^H + N_0\mathbf{I}_M \tag{14}
$$

with $P_j = \mathrm{E}\left[\left|x_j\right|^2\right]$.

In [7], the LMMSE estimate of x_k is realized by the NW-MF described as follows. First, the NW is performed as

$$
\mathbf{y}' = \mathbf{K}_z^{-1/2}\mathbf{y} = \mathbf{K}_z^{-1/2}\mathbf{h}_k x_k + \mathbf{K}_z^{-1/2}\mathbf{z} \tag{15}
$$

Then, a MF is used as

$$\begin{aligned}\bar{\mathbf{y}} &= \left(\mathbf{K}_z^{-1/2}\mathbf{h}_k\right)^H \mathbf{y}' = \left(\mathbf{K}_z^{-1/2}\mathbf{h}_k\right)^H \mathbf{K}_z^{-1/2}\mathbf{h}_k x_k + \left(\mathbf{K}_z^{-1/2}\mathbf{h}_k\right)^H \mathbf{K}_z^{-1/2}\mathbf{z} \\ &= \mathbf{h}_k^H \mathbf{K}_z^{-1}\mathbf{h}_k x_k + \mathbf{h}_k^H \mathbf{K}_z^{-1}\mathbf{z}\end{aligned} \tag{16}$$

From (15) and (16), the NW-MF estimate vector is $\mathbf{v}_{\text{NW-MF}}^H = \mathbf{h}_k^H \mathbf{K}_z^{-1}$ since $\bar{\mathbf{y}} = \mathbf{v}^H \mathbf{h}_k^H \mathbf{K}_z^{-1}$. The declaration that NW-MF is equivalent to LMMSE is based on the relation between signal-to-noise-ratio (SNR) and mean square error (MSE) (Exercise 8.18, [7]). Here, this conclusion is presented as Theorem 2 and an alternative proof is given.

Theorem 2: The NW-MF estimation is equivalent to the LMMSE estimation in the linear Gaussian channel.

Proof:

With the zero-mean assumption, the standard LMMSE estimate vector is $\mathbf{v}_{\text{LMMSE}}^H = \mathbf{K}_{x_k y}\mathbf{K}_y^{-1}$ from (2). Therefore, the proof of Theorem 2 is equivalent to the proof that $\mathbf{v}_{\text{LMMSE}}^H = \alpha \mathbf{v}_{\text{NW-MF}}^H$ with a scalar α .

Note that

$$\mathbf{K}_{x_k y} = \mathrm{E}\left[\left(x_k - \mu_{x_k}\right)\left(\mathbf{y} - \boldsymbol{\mu}_y\right)^H\right] = \mathrm{E}\left[x_k \left(\mathbf{h}_k x_k + \mathbf{z}\right)^H\right] = P_k \mathbf{h}_k^H \tag{17}$$

$$\begin{aligned}\mathbf{K}_y &= \mathrm{E}\left[\left(\mathbf{y} - \boldsymbol{\mu}_y\right)\left(\mathbf{y} - \boldsymbol{\mu}_y\right)^H\right] = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] \\ &= \mathrm{E}\left[\left(\mathbf{h}_k x_k + \mathbf{z}\right)\left(\mathbf{h}_k x_k + \mathbf{z}\right)^H\right] = P_k \mathbf{h}_k \mathbf{h}_k^H + \mathbf{K}_z\end{aligned} \tag{18}$$

It has

$$\mathbf{v}_{\text{LMMSE}}^H = \mathbf{K}_{x_k y}\mathbf{K}_y^{-1} = P_k \mathbf{h}_k^H \left(P_k \mathbf{h}_k \mathbf{h}_k^H + \mathbf{K}_z\right)^{-1} \tag{19}$$

Using the matrix inverse lemma (4), it has

$$\left(P_k \mathbf{h}_k \mathbf{h}_k^H + \mathbf{K}_z\right)^{-1} = \mathbf{K}_z^{-1} - \frac{P_k \mathbf{K}_z^{-1}\mathbf{h}_k \mathbf{h}_k^H \mathbf{K}_z^{-1}}{1 + P_k \left\|\mathbf{h}_k\right\|^2} \tag{20}$$

Then,

$$\begin{aligned}\mathbf{v}_{\text{LMMSE}}^H &= P_k \mathbf{h}_k^H \left(\mathbf{K}_z^{-1} - \frac{P_k \mathbf{K}_z^{-1}\mathbf{h}_k \mathbf{h}_k^H \mathbf{K}_z^{-1}}{1 + P_k \left\|\mathbf{h}_k\right\|^2}\right) = P_k \mathbf{h}_k^H \mathbf{K}_z^{-1} - \frac{P_k^2 \mathbf{h}_k^H \mathbf{K}_z^{-1}\mathbf{h}_k \mathbf{h}_k^H \mathbf{K}_z^{-1}}{1 + P_k \left\|\mathbf{h}_k\right\|^2} \\ &= \left(P_k - \frac{P_k^2 \mathbf{h}_k^H \mathbf{K}_z^{-1}\mathbf{h}_k}{1 + P_k \left\|\mathbf{h}_k\right\|^2}\right)\mathbf{h}_k^H \mathbf{K}_z^{-1} = \alpha \mathbf{h}_k^H \mathbf{K}_z^{-1} = \alpha \mathbf{v}_{\text{NW-MF}}^H\end{aligned} \tag{21}$$

with

$$\alpha = P_k - \frac{P_k^2 \mathbf{h}_k^H \mathbf{K}_z^{-1} \mathbf{h}_k}{1 + P_k \|\mathbf{h}_k\|^2} \tag{22}$$

□

VI. Conclusions

Applying the matrix inverse lemma, two conclusions associated with LMMSE were proved. Specifically, first, the information theoretical optimality of LMMSE in the linear Gaussian channel was proved based on the relation of the differential entropy with the determinant of the autocorrelation matrix. Second, an alternative proof of the equivalence of NW-MF and LMMSE was given based on the linear relation of the two estimation vectors.

References

[1] Van Trees, H. L. *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

[2] Nicholas J. Higham, *Accuracy and Stability of Numerical Algorithms (2nd ed.)*, PA: SIAM, 2002.

[3] G. D. Forney, Jr., "Shannon meets wiener II: On MMSE estimation in successive decoding schemes," in *Proceedings of 2004 Allerton Conference*, Monticello, IL, Oct. 2004.

[4] J. M. Cioffi and G. D. Forney, Jr., "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing (A. Paulraj et al., eds.)*, pp. 79-127.

[5] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding-Part I: Equalization results; Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2581-2604, Oct. 1995.

[6] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 173-187, Jan. 2005.

[7] D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.