EECS 225A Statistical Signal Processing Lecture 20 - 4/14/2020

Lecture 20: H^∞-Optimality of the LMS and Risk-Sensitive Estimation

Lecturer: Jiantao Jiao *Scribe: Shreya Ramachandran*

In this lecture, we will make connections between the LMS (least mean squares) algorithm, H^∞ estimation, and risk-sensitive estimation. We will show that the LMS algorithm is H^∞-optimal and a risk-averse algorithm. This is important in providing insight into the inherent robustness of the LMS algorithm versus the RLS (recursive least squares) algorithm, which is not H^∞- optimal.

1 Problem Setting for Adaptive Filtering [HSK99, Chapter 7]

Suppose we observe sequence $\{d_i\}$ for index $i \geq 0$, where

$$d_i = h_i w + v_i \tag{1}$$

Here, $d_i \in \mathbb{C}$ is the observation, $h_i \in \mathbb{C}^{1\times n}$ is a known input row vector, $w \in \mathbb{C}^{n\times 1}$ is an unknown weight vector (which will be estimated), and $v_i \in \mathbb{C}$ is noise. We assume that $\langle v_i, v_j\rangle = \delta_{ij}$ and $\langle v_i, w\rangle = 0$. The notation $\hat{w}_{|i}$ is used to denote the best estimate of the weight vector w given $\{d_j\}_{j=0}^{i}$. The estimate of the weight vector given no observations is $\hat{w}_{|-1} = 0$.

1.1 Connection to Linear Estimation Setting

This adaptive filter problem setting can be mapped to the general linear estimation problem setting in Lecture 18 with the following diagram.

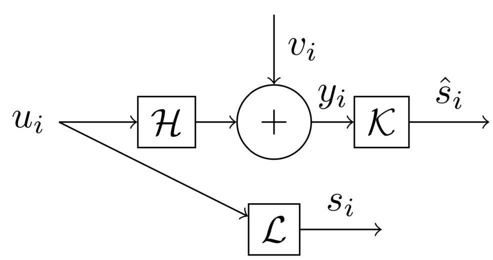


Figure 1: Decomposition of a Linear Estimation Problem (from Lecture 18)

Equating terms from the two problem settings, we have $y_i = d_i$, $s_i = h_i w$, and $u_i = \mu^{-\frac{1}{2}} w$, where the parameter μ is defined by $\langle w, w\rangle = \mu I$. The systems $\mathcal{H}$ and $\mathcal{L}$ multiply $(\mu^{\frac{1}{2}} h_i)$ to the input $u_i = \mu^{-\frac{1}{2}} w$, so that their outputs are exactly $h_i w$. Note that we need to guarantee that u, v are white noise with variance one to be consistent with lecture 18.

1.2 RLS Algorithm [HSK99]

In the RLS algorithm (or Wiener filter), the estimate of w at time i is the solution to the following least squares problem

$$\min_w \quad \mu^{-1}\|w\|^2 + \sum_{j=0}^{i} |d_j - h_j w|^2 \tag{2}$$

The recursive solution to this minimization is

$$\hat{w}_{|i} = \hat{w}_{|i-1} + k_{p,i}(d_i - h_i \hat{w}_{|i-1}) \tag{3}$$

where $k_{p,i} = \frac{P_i h_i^*}{1+h_i P_i h_i^*}$ is the predictive Kalman gain, and P_i satisfies the Riccati equation, that is,

$$P_{i+1} = P_i - \frac{P_i h_i^* h_i P_i}{1 + h_i P_i h_i^*}\ ,\ P_0 = \mu I \tag{4}$$

The RLS algorithm is H^2-optimal but not H^∞-optimal. It will be shown why RLS is not H^∞-optimal in the following section.

1.3 LMS Algorithm

In the case of LMS, the Kalman gain vector $k_{p,i}$ in Equation (4) is replaced with μh_i^* to give the following recursion

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \mu h_i^*(d_i - h_i \hat{w}_{|i-1}) \tag{5}$$

The LMS algorithm is not H^2-optimal, but is H^∞-optimal, which will be shown in the following section.

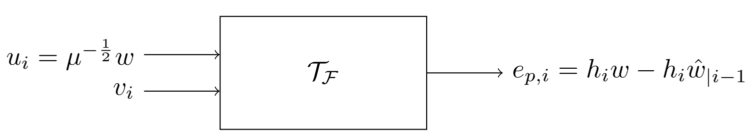
2 LMS is H^∞- optimal

We want to construct a linear filter $\mathcal{F}$ with inputs as the observations $d_0, d_1, ..., d_i$ and the parameters $(h_0, h_1, ..., h_i, \mu)$, and with output $\hat{w}_{|i}$, i.e.

$$\hat{w}_{|i} = \mathcal{F}(d_0, d_1, ..., d_i; h_0, h_1, ..., h_i, \mu) \tag{6}$$

Hence, $\mathcal{F}$ is constrained to be a causal filter.

For H^∞-optimality, we want to choose the filter $\mathcal{F}$ which minimizes the H^∞-norm of the transfer operator $\mathcal{T}_\mathcal{F}$, represented by the following diagram:



Note that $e_p = (e_{p,0}, e_{p,1}, ..., e_{p,i})^T \in \mathbb{C}^{i+1}$ and $v = (v_0, v_1, ..., v_i)^T \in \mathbb{C}^{i+1}$, but $w \in \mathbb{C}^n$ as defined in the problem setting.

As per the definition in Lecture 18, the H^∞ norm of $\mathcal{T}_\mathcal{F}$ is the following, which we define as $\gamma_{p,i}(\mathcal{T}_\mathcal{F})^2$.

$$\gamma_{p,i}(\mathcal{T}_\mathcal{F})^2 \triangleq \sup_{\substack{w,v: \\ \|w\|_2<\infty \\ \|v\|_2<\infty}} \frac{\|e_p\|_2^2}{\mu^{-1}\|w\|_2^2 + \|v\|_2^2} \tag{7}$$

Here, the numerator represents the output signal magnitude and the denominator represents the input signal magnitude. Hence, for H^∞-optimality, the filter $\mathcal{F}$ chosen in (6) must be the solution to the following optimization problem:

$$\inf_{\mathcal{F}} \sup_{\substack{w,v: \\ \|w\|_2<\infty \\ \|v\|_2<\infty}} \frac{\|e_p\|_2^2}{\mu^{-1}\|w\|_2^2 + \|v\|_2^2} = \inf_{\mathcal{F}} \gamma_{p,i}(\mathcal{T}_\mathcal{F})^2 \tag{8}$$

We define

$$\gamma_p \triangleq \lim_{i\to\infty} \inf_{\mathcal{F}} \gamma_{p,i}(\mathcal{T}_{\mathcal{F}}). \tag{9}$$

It is not a priori clear that the limit exist. We will show it exists under certain conditions.

We now show

Claim 1. γ_p *cannot be less than 1, i.e.* $\gamma_p \geq 1$.

Proof Let us consider a specific disturbance w, v where for each $i \geq 0$, $h_i w + v_i = d_i = h_i \hat{w}_{|-1} = 0$. Since all observations are 0, and each $\hat{w}_{|i}$ is only allowed to be *linear* estimator of $\{d_j\}_{j=0}^{i}$, it shows that $\hat{w}_{|i} = 0$ for all $i \geq 0$. This implies that

$$e_{p,i} = h_i w - h_i \hat{w}_{|i-1} = h_i w = -v_i \tag{10}$$

Therefore,

$$\begin{aligned} \frac{\|e_p\|_2^2}{\mu^{-1}\|w\|_2^2 + \|v\|_2^2} &= \frac{\|v\|_2^2}{\mu^{-1}\|w\|_2^2 + \|v\|_2^2} \\ &= \frac{\sum_{j=0}^{i} |h_j w|^2}{\mu^{-1}\|w\|_2^2 + \sum_{j=0}^{i} |h_j w|^2} \end{aligned} \tag{11}$$

Intuitively, we can see that as $i \to \infty$, this fraction will approach one if the term $\sum_{j=0}^{i} |h_j w|^2 \to \infty$, because $\mu^{-1}\|w\|_2^2$ remains constant with i. We will formalize the condition where $\sum_{j=0}^{i} |h_j w|^2 \to \infty$ as $i \to \infty$, in the Condition of Exciting Inputs.

Definition 2. *Inputs* $h_j \in \mathbb{C}^{1\times n}$ *satisfy the condition of exciting inputs if*

$$\lim_{N\to\infty} \sum_{j=0}^{N} h_j h_j^* = \infty$$

Continuing with the proof, we can expand the summation appearing in the numerator and denominator of (10) as

$$\sum_{j=0}^{i} |h_j w|^2 = \sum_{j=0}^{i} w^* h_j^* h_j w = w^* (\sum_{j=0}^{i} h_j^* h_j) w \tag{12}$$

Here, $(\sum_{j=0}^{i} h_j^* h_j) \in \mathbb{C}^{n\times n}$ is a matrix. If the inputs h_j satisfy the condition of exciting inputs, then $\mathrm{Tr}(\sum_{j=0}^{i} h_j^* h_j) \to \infty$ as $i \to \infty$. This implies that the sum of eigenvalues of this matrix, that is,

$$\sum_{k=1}^{n} \lambda_k (\sum_{j=0}^{i} h_j^* h_j) \to \infty$$

as $i \to \infty$. Since n is finite and all eigenvalues of $(\sum_{j=0}^{i} h_j^* h_j)$ are non-negative because it is a PSD matrix, then there must exist at least one eigenvalue $\to \infty$. If we choose w to be the eigenvector corresponding to this eigenvalue, then for every $i > 0$, there is an $\epsilon \geq 0$ such that,

$$\sum_{j=0}^{i} |h_j w|^2 \geq \frac{\|w\|_2^2}{\epsilon\mu} \tag{13}$$

Plugging this inequality into equation (11), we get that

$$\begin{aligned} \frac{\sum_{j=0}^{i} |h_j w|^2}{\mu^{-1}\|w\|_2^2 + \sum_{j=0}^{i} |h_j w|^2} &\geq \frac{\frac{\|w\|_2^2}{\epsilon\mu}}{\mu^{-1}\|w\|_2^2 + \frac{\|w\|_2^2}{\epsilon\mu}} \\ &= \frac{\frac{1}{\epsilon}}{1 + \frac{1}{\epsilon}} = \frac{1}{1+\epsilon} \geq 1 - \epsilon \end{aligned} \tag{14}$$

Replacing the left-hand side of this inequality using equation (10), it shows that for any $\epsilon > 0$, there exists some i large enough such that for any estimator $\mathcal{F}$,

$$\sup_{w,v} \frac{\|e_p\|_2^2}{\mu^{-1}\|w\|_2^2 + \|v\|_2^2} \geq 1 - \epsilon. \tag{15}$$

It shows that

$$\lim_{i\to\infty} \inf \inf_{\mathcal{F}} \gamma_{p,i}(\mathcal{T}_{\mathcal{F}}) \geq 1. \tag{16}$$

$\square$

Now that we have shown $\gamma_p \geq 1$, we will show that for the LMS algorithm $\gamma_{p,i}(\mathcal{T}_{\mathcal{F}}) \leq 1$ for any $i \geq 0$. Proving these two statements will prove that $\gamma_p = 1$, and hence, that the LMS Algorithm is H^∞-optimal.

Theorem 3. *If the condition of exciting inputs is satisfied, and $0 < \mu < \inf_i \frac{1}{h_i h_i^*}$, then the LMS algorithm with parameter μ achieves $\gamma_{p,i}(\mathcal{T}_{\mathcal{F}}) \leq 1$ for all $i \geq 0$.*

Proof Let $0 < \mu < \inf_i \frac{1}{h_i h_i^*}$ and h_j satisfy the condition of exciting inputs. From (5), we already know that the LMS recursion is the following

$$\hat{w}_{|j} = \hat{w}_{|j-1} + \mu h_j^*(d_j - h_j \hat{w}_{|j-1}) \tag{17}$$

Subtracting w from both sides, and multiplying both sides by $\mu^{-\frac{1}{2}}$,

$$\mu^{-\frac{1}{2}}(\hat{w}_{|j} - w) = \mu^{-\frac{1}{2}}(\hat{w}_{|j-1} - w + \mu h_j^*(d_j - h_j \hat{w}_{|j-1})) \tag{18}$$

We will use the notation $\tilde{w}_{|j} = w - \hat{w}_{|j}$. Then, the above equation can be written as

$$\mu^{-\frac{1}{2}}\tilde{w}_{|j} = \mu^{-\frac{1}{2}}(\tilde{w}_{|j-1} - \mu h_j^*(d_j - h_j \hat{w}_{|j-1})) \tag{19}$$

We also know from the observation model that

$$\begin{align} v_j &= d_j - h_j w \tag{20}\\ &= d_j - h_j(\tilde{w}_{|j-1} + \hat{w}_{|j-1}) \tag{21}\\ &= (d_j - h_j \hat{w}_{|j-1}) - h_j \tilde{w}_{|j-1} \tag{22}\end{align}$$

Squaring equations (19) and (22) respectively, we get (23) and (24) as follows

$$\begin{align} \mu^{-1}\|\tilde{w}_{|j}\|_2^2 &= \mu^{-1}\|\tilde{w}_{|j-1} - \mu h_j^*(d_j - h_j \hat{w}_{|j-1})\|_2^2 \tag{23}\\ |v_j|^2 &= |d_j - h_j \hat{w}_{|j-1} - h_j \tilde{w}_{|j-1}|^2 \tag{24}\end{align}$$

Subtracting (23) - (24) on both sides,

$$\begin{align} \mu^{-1}\|\tilde{w}_{|j}\|_2^2 - |v_j|^2 &= \mu^{-1}\|\tilde{w}_{|j-1} - \mu h_j^*(d_j - h_j \hat{w}_{|j-1})\|_2^2 - |d_j - h_j \hat{w}_{|j-1} - h_j \tilde{w}_{|j-1}|^2 \tag{25}\\ &= \mu^{-1}\|\tilde{w}_{|j-1}\|_2^2 - |h_j \tilde{w}_{|j-1}|^2 - (1 - \mu h_j h_j^*)|d_j - h_j \hat{w}_{|j-1}|^2 \tag{26}\end{align}$$

Since we have that $0 < \mu < \inf_i \frac{1}{h_i h_i^*}$, this implies that $(1 - \mu h_j h_j^*) \geq 0$, which then implies that $(1 - \mu h_j h_j^*)|d_j - h_j \hat{w}_{|j-1}|^2 \geq 0$. Therefore,

$$\mu^{-1}\|\tilde{w}_{|j}\|_2^2 - |v_j|^2 \leq \mu^{-1}\|\tilde{w}_{|j-1}\|_2^2 - |h_j \tilde{w}_{|j-1}|^2, \tag{27}$$

which implies

$$|h_j \tilde{w}_{|j-1}|^2 \leq \mu^{-1} \|\tilde{w}_{|j-1}\|_2^2 - \mu^{-1} \|\tilde{w}_{|j}\|_2^2 + |v_j|^2. \tag{28}$$

Since $|h_j \tilde{w}_{|j-1}|^2 = |e_{p,j}|^2$,

$$|e_{p,j}|^2 \leq \mu^{-1}(\|\tilde{w}_{|j-1}\|_2^2 - \|\tilde{w}_{|j}\|_2^2) + |v_j|^2 \tag{29}$$

Summing both sides from $j = 0$ to $j = i$,

$$\sum_{j=0}^{i} |e_{p,j}|^2 \leq \mu^{-1} \sum_{j=0}^{i} (\|\tilde{w}_{|j-1}\|_2^2 - \|\tilde{w}_{|j}\|_2^2) + \sum_{j=0}^{i} |v_j|^2 \tag{30}$$

We can use the idea of a telescoping sum to simplify the first term on the right-hand side of this inequality. The telescoping sum concept can be seen clearly with the following expansion

$$\sum_{j=0}^{i} (\|\tilde{w}_{|j-1}\|_2^2 - \|\tilde{w}_{|j}\|_2^2) = (\|\tilde{w}_{|-1}\|_2^2 - \|\tilde{w}_{|0}\|_2^2) + (\|\tilde{w}_{|0}\|_2^2 - \|\tilde{w}_{|1}\|_2^2) + ... + (\|\tilde{w}_{|i-1}\|_2^2 - \|\tilde{w}_{|i}\|_2^2) \tag{31}$$

$$= \|\tilde{w}_{|-1}\|_2^2 - \|\tilde{w}_{|i}\|_2^2 \tag{32}$$

Plugging this simplified summation back into (29),

$$\sum_{j=0}^{i} |e_{p,j}|^2 \leq \mu^{-1}(\|\tilde{w}_{|-1}\|_2^2 - \|\tilde{w}_{|i}\|_2^2) + \sum_{j=0}^{i} |v_j|^2 \tag{33}$$

$$= \mu^{-1}\|w\|_2^2 - \mu^{-1}\|\tilde{w}_{|i}\|_2^2 + \sum_{j=0}^{i} |v_j|^2 \tag{34}$$

$$\leq \mu^{-1}\|w\|_2^2 + \sum_{j=0}^{i} |v_j|^2 \tag{35}$$

since $\tilde{w}_{|-1} = w - \hat{w}_{|-1} = w$.
Hence, we have shown that the following inequality holds non-asymptotically

$$\frac{\sum_{j=0}^{i} |e_{p,j}|^2}{\mu^{-1}\|w\|_2^2 + \sum_{j=0}^{i} |v_j|^2} \leq 1, \tag{36}$$

which is the same as

$$\gamma_{p,i}(\mathcal{T}_{\mathcal{F}}) \leq 1 \tag{37}$$

for the LMS algorithm. □

Therefore, by Theorem 3, we know that the LMS algorithm achieves H^∞ norm at most 1, but for any algorithm it is at least 1 in the infimum limit. Hence the LMS algorithm is H^∞ optimal.

3 RLS is not H^∞-optimal

The numerical example in [HSK96] shows that the RLS is not H^∞ optimal even when $\mu < \inf_i \frac{1}{h_i h_i^*}$.

4 H^∞-estimation and Risk-Sensitive Estimation

In Kalman filter, the goal is to minimize the expectation of the predicted error $e_{p,j}$, that is,

$$\min \mathbb{E} \sum_{j=0}^{i} |e_{p,j}|^2 \tag{38}$$

We can generalize this to instead minimize the expectation of the following exponential function,

$$\min \frac{-2}{\theta} \log(\mathbb{E}[e^{\frac{-\theta}{2} \sum_{j=0}^{i} |e_{p,j}|^2}]) \tag{39}$$

Defining $C_p \triangleq \sum_{j=0}^{i} |e_{p,j}|^2$, then the above equation can be rewritten as,

$$\min \frac{-2}{\theta} \log(\mathbb{E}[e^{\frac{-\theta}{2} C_p}]) \tag{40}$$

When $\theta \to 0$, this minimization reduces to $\min \mathbb{E}[C_p]$. When $|\theta|$ is small, then a Taylor expansion can be performed to show that this problem reduces to

$$\min \mathbb{E}[C_p] - \frac{\theta}{4} \mathsf{Var}[C_p] + O(\theta^2) \tag{41}$$

which is a minimization of a weighted sum of the mean and the variance of c_p. There are two possible settings for this weighted sum problem:

- $\theta > 0$: This is known as the risk-seeking regime. The cost function decreases when $\mathsf{Var}[C_p]$ increases.
- $\theta < 0$: This is known as the risk-averse regime. The cost function decreases when $\mathsf{Var}[C_p]$ decreases.

It should be noted that to minimize $\mathbb{E}[C_p]$, we only require information about the 1st and 2nd order moments of C_p. However to minimize $\mathbb{E}[e^{\alpha C_p}]$, we usually require information about higher order moments, or complete distribution information (such as Gaussianity)

4.1 Risk-sensitive interpretation of LMS

Theorem 4. *Assume that the condition of exciting inputs is satisfied and that $0 < \mu < \inf_i \frac{1}{h_i h_i^*}$, then the LMS algorithm with parameter μ is the solution to the following minimization problem under joint Gaussian assumptions over causal linear filters:*

$$\min 2 \log(\mathbb{E}[e^{\frac{1}{2} C_p}]) \tag{42}$$

This implies that $\theta = -1$ for the LMS algorithm, and hence, it is a risk-averse algorithm. Furthermore, this demonstrates the inherent robustness behind the LMS algorithm versus the RLS algorithm. LMS penalizes the variance in the prediction error at an exponential scale, while RLS only penalizes this error at a linear scale.

For more on the connection between risk-sensitive estimation/control and H^∞ theory, we refer the readers to [SC08, Chapter 10] and [Whi90].

References

[HSK96] Babak Hassibi, Ali H Sayed, and Thomas Kailath. H/sup/spl infin//optimality of the lms algorithm. *IEEE Transactions on signal processing*, 44(2):267–280, 1996.

[HSK99] B. Hassibi, A.H. Sayed, and T. Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H2 and H-infinity Theories*, chapter 7. SIAM Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1999.

[SC08] Jason L Speyer and Walter H Chung. *Stochastic processes, estimation, and control*, volume 17. Siam, 2008.

[Whi90] Peter R Whittle. *Risk-sensitive optimal control*, volume 20. Wiley New York, 1990.