EECS 225A Statistical Signal Processing Lecture 18 - 3/31/2020

Lecture 18: H^2 and H^∞ Estimation Theory

Lecturer: Jiantao Jiao *Scribe: Milo Webster*

In this lecture we explore H^2 and H^∞ estimation. These topics are important because they provide a unified view towards the optimal linear estimation theory and adaptive filtering theory. For example, it shows the recursive least squares solution is H^2 but not H^∞ optimal, and the Least Mean Square (LMS) algorithm is H^∞ but not H^2 optimal. The materials in this lecture are mainly based on [HSK99, Chapter 10].

Note: "H" stands for Hardy Space

1 Background

We are used to using certain norms to quantify the magnitude of a signal. For example, the ℓ^2 norm of a **signal** $v(t)$ would be defined as:

$$\|v(t)\|_2 = \sqrt{\sum_t \mathrm{Tr}(v(t)v(t)^*)}$$

Similarly we can consider **system** norms, some of which we introduce below.

1.1 A family of linear estimation problems

The following diagram captures a fairly general setting of linear estimation.

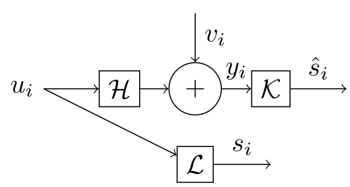

Figure 1: Decomposition of a Linear Estimation Problem

In this diagram we have the following signals and systems. It captures two cases: in one case, all signals are finite-dimensional, and the corresponding systems are matrices; in the other case, all signals are WSS processes, and all systems are LTI systems with the system symbol denoting its transfer function.

1. u_i: some *unobserved* input stochastic process
2. $\mathcal{H}$: a *known* causal linear system
3. $\mathcal{L}$: a *known* causal linear system
4. v_i: additive white noise
5. $\mathcal{K}$: the transfer function (could be either causal, or non-causal) that estimates s_i using the observations $\{y_j\}$, the estimate is denoted by $\hat{s}_i$
6. $\hat{s}_i$: our estimate of s_i
7. s_i: quantity to be estimated

Now, in order to cast this problem in terms of H^2, H^∞ theory, we must define system that maps u_i and v_i to the final $\hat{s}_i$.

1.2 Derivation of Linear System Representing this Estimate

We now define a linear system $\mathcal{T}_\mathcal{K}$ that performs the following map, where $\tilde{s}_i$ is the estimation error at time step i:

$$\begin{pmatrix} u \\ v \end{pmatrix} \longrightarrow \tilde{s}_i \triangleq s_i - \hat{s}_i \tag{1}$$

Now we derive $\mathcal{T}_\mathcal{K}$'s transfer function:

$$\begin{aligned}
\tilde{s} &= s - \hat{s} && (2)\\
&= \mathcal{L}u - \mathcal{K}(\mathcal{H}u + v) && (3)\\
&= (\mathcal{L} - \mathcal{K}\mathcal{H})u - \mathcal{K}v && (4)\\
&= \begin{bmatrix} \mathcal{L} - \mathcal{K}\mathcal{H} & -\mathcal{K} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} && (5)\\
&= \mathcal{T}_\mathcal{K} \begin{bmatrix} u \\ v \end{bmatrix} && (6)
\end{aligned}$$

where

$$\mathcal{T}_\mathcal{K} = \begin{bmatrix} \mathcal{L} - \mathcal{K}\mathcal{H} & -\mathcal{K} \end{bmatrix} \tag{7}$$

Since $\tilde{s}$ represents the estimation error, and the inputs are fixed with known first and second order statistics, the filter $\mathcal{K}$ being a good estimator should be equivalent to saying that the *norm* of the system $\mathcal{T}_\mathcal{K}$ is *small* in some sense. What norm shall we consider here?

1.3 $\mathcal{H}^2$ Norm of System $\mathcal{T}_\mathcal{K}$

A natural target is to see what norm the usual optimal linear estimation theory corresponds to. We show that it corresponds to the $\mathcal{H}^2$ norm of the system $\mathcal{T}_\mathcal{K}$. We distinguish two cases: finite-horizon and infinite horizon.

1.3.1 Finite-horizon

In this case we have that u,v,y are finite dimensional vectors, and $\mathcal{T}_\mathcal{K}$ is a finite matrix, and we define its H^2 norm as its Frobenius Norm:

$$\|\mathcal{T}_\mathcal{K}\|_{\mathcal{H}^2} \triangleq \|\mathcal{T}_\mathcal{K}\|_F = (\mathrm{Tr}(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*))^{\frac{1}{2}} = (\sum_{i,j} \|\mathcal{T}_{\mathcal{K},ij}\|_F^2)^{\frac{1}{2}}, \tag{8}$$

here $\mathcal{T}_{\mathcal{K},ij}$ is the block (i,j) of matrix $\mathcal{T}_\mathcal{K}$. In other words, $\mathcal{T}_{\mathcal{K},ij}$ maps $[u_j, v_j]^T$ to $\tilde{s}_i$, and $\|\mathcal{T}_{\mathcal{K},ij}\|_F^2 = \mathrm{Tr}(\mathcal{T}_{\mathcal{K},ij}\mathcal{T}_{\mathcal{K},ij}^*)$.

1.3.2 Infinite-horizon, time-invariant

Here, all of the systems described become LTI systems and all the processes become WSS processes, and we have that:

$$\|\mathcal{T}_\mathcal{K}\|_{H^2} \triangleq (\frac{1}{2\pi}\int_{-\pi}^{\pi} \left\|\mathcal{T}_\mathcal{K}(e^{j\omega})\right\|_F^2 d\omega)^{\frac{1}{2}} \tag{9}$$

where

$$\mathcal{T}_{\mathcal{K}}(e^{j\omega}) \triangleq \sum_{i=-\infty}^{+\infty} \mathcal{T}_{\mathcal{K},i} e^{-j\omega i} \tag{10}$$

And by Parseval's Identity, we can also write the H^2 norm as:

$$\|\mathcal{T}_{\mathcal{K}}\|_{H^2} = (\sum_{i=-\infty}^{+\infty} \|\mathcal{T}_{\mathcal{K},i}\|_F^2)^{\frac{1}{2}} \tag{11}$$

In the infinite-horizon case what we are doing is first computing the DTFT, then for every ω with compute the square of the Frobenius norm and take the average. That is where the name H^2 comes from.

2 H^2 Estimation

The widespread use of H^2 theory is mainly due to the facts that optimal H^2 problem has a simple closed-form solution and as we show below, it can be interpreted as optimal linear estimation under squared error loss function. In particular, we look at the following optimization problem:

$$\min_{\mathcal{K}} \|\mathcal{T}_{\mathcal{K}}\|_{H^2} \tag{12}$$

We first show the H^2 norm of $\mathcal{T}_{\mathcal{K}}$ is exactly the mean squared error in estimating s_i.

Theorem 1. *[HSK99, Lemma 10.3.1] Given white noise processes u_i, v_j (i.e. $\mathbb{E}[\begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_j^* & v_j^* \end{bmatrix}] = \langle \begin{pmatrix} u_i \\ v_i \end{pmatrix}, \begin{pmatrix} u_j \\ v_j \end{pmatrix} \rangle = I\delta_{ij}$). Then,*

1. *Finite-horizon case*

$$\|\mathcal{T}_{\mathcal{K}}\|_{H^2}^2 = \mathbb{E}[\sum_{i=0}^{N} |\tilde{s}_i|^2] \tag{13}$$

2. *Infinite-horizon case*

$$\|\mathcal{T}_{\mathcal{K}}\|_{H^2}^2 = \mathbb{E}[|\tilde{s}_i|^2] \tag{14}$$

This theorem beautifully shows that, for any estimator $\mathcal{K}$, if we compute the corresponding H^2 norm of our transfer function $\mathcal{T}_{\mathcal{K}}$, then in both finite and infinite horizon cases, this captures the expected estimation error corresponding to a squared loss function.

Proof First we consider the finite-horizon case (though the infinite-horizon is very similar):

$$\sum_{i=0}^{N} |\tilde{s}_i|^2 = \begin{pmatrix} u^* & v^* \end{pmatrix} \mathcal{T}_{\mathcal{K}}^* \mathcal{T}_{\mathcal{K}} \begin{pmatrix} u \\ v \end{pmatrix} = \mathrm{Tr}(\mathcal{T}_{\mathcal{K}}^* \mathcal{T}_{\mathcal{K}} \begin{pmatrix} u \\ v \end{pmatrix} \begin{pmatrix} u^* & v^* \end{pmatrix}) \tag{15}$$

Taking expectations on both sides, we have

$$\mathbb{E}[\sum_{i=0}^{N} |\tilde{s}_i|^2] = \mathrm{Tr}(\mathcal{T}_{\mathcal{K}}^* \mathcal{T}_{\mathcal{K}}) = \|\mathcal{T}_{\mathcal{K}}\|_{H^2}^2 \tag{16}$$

□

After showing the connection between H^2 estimation, we already know the answer to the optimal H^2 estimation problem since it is exactly Wiener filter. We would like to offer an algebraically different proof for the *non-causal* case here to obtain a crisp formula in this case, since it will be useful for later discussions on H^∞ estimation.

We can verify algebraically

$$\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^* = (\mathcal{K} - \mathcal{L}\mathcal{H}^*(I+\mathcal{H}\mathcal{H}^*)^{-1})(I+\mathcal{H}\mathcal{H}^*)(\mathcal{K} - \mathcal{L}\mathcal{H}^*(I+\mathcal{H}\mathcal{H}^*)^{-1})^* + \mathcal{L}(I+\mathcal{H}^*\mathcal{H})^{-1}\mathcal{L}^* \tag{17}$$

If we denote

$$\mathcal{K}_0 = \mathcal{L}\mathcal{H}^*(I+\mathcal{H}\mathcal{H}^*)^{-1} \tag{18}$$

then it shows for any $\mathcal{K}$

$$\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^* \geq \mathcal{T}_{\mathcal{K}_0}\mathcal{T}_{\mathcal{K}_0}^* \tag{19}$$

since the first term on the RHS is positive semidefinite. Note that this inequality means that for *any* disturbance u, v and *transfer* operators $\mathcal{K}$, the operator $\mathcal{K}_0$ results in the smallest estimation error energy *pointwise*, not only in expectation! Clearly, $\mathcal{K}_0$ also minimizes the trace of $\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*$, showing that it is a solution to the H^2 estimation problem. In fact, it is *the* solution due to the uniqueness of optimal linear estimator investigated in lecture 2.

3 H^∞ Estimation

Now we address the matter of robustness using H^∞ estimation theory. Consider the following:

$$\frac{\|\tilde{s}\|_2^2}{\|u\|_2^2\,\|v\|_2^2} = \frac{\left\|\mathcal{T}_\mathcal{K}\begin{pmatrix}u\\v\end{pmatrix}\right\|_2^2}{\|u\|_2^2 + \|v\|_2^2} \tag{20}$$

Our goal is to take the supremum of this quantity over all signals u,v, and then the infemum over transfer function $\mathcal{K}$, in order to find the worst case energy gain. We care about this worst case energy gain, because this quantity contains the noise in our system:

$$\inf_{\mathcal{K}} \sup_{u:\|u\|_2<\infty, v:\|v\|_2<\infty} \frac{\left\|\mathcal{T}_\mathcal{K}\begin{pmatrix}u\\v\end{pmatrix}\right\|_2^2}{\|u\|_2^2 + \|v\|_2^2} \tag{21}$$

Now we precisely define the H^∞ norm:

Definition 2. *The H^∞ norm of any LTI system or matrix is defined as*

$$\|\mathcal{T}\|_{H^\infty} \triangleq \sup_{x:\|x\|_2<\infty, x\neq 0} \frac{\|\mathcal{T}x\|_2}{\|x\|_2} \tag{22}$$

To investigate this definition we have the following theorem:

Theorem 3. *H^∞ norm of $\mathcal{T}$*

1. *In the finite-horizon case*

$$\|\mathcal{T}\|_{H^\infty} = \sigma_{\max}(\mathcal{T}) \tag{23}$$

(the largest singular value of $\mathcal{T}$)

2. In the infinite-horizon case

$$\|\mathcal{T}\|_{H^\infty} = \sup_{\omega \in (-\pi,\pi)} \sigma_{\max}[\mathcal{T}(e^{j\omega})] \tag{24}$$

In general, H^∞ estimation is hard, however the non-causal case is easy to obtain. Recall that previously we said that the optimal $\mathcal{K}_0$ is given by $\mathcal{L}\mathcal{H}^*(I+\mathcal{H}\mathcal{H}^*)^{-1}$. We also said that for any $\mathcal{K}$,

$$\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^* \geq \mathcal{T}_{\mathcal{K}_0}\mathcal{T}_{\mathcal{K}_0}^* \tag{25}$$

Unlike for H^2 estimation where we take the trace, here we take the largest eigenvalue of $\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*$, which directly implies that

$$\|\mathcal{T}_{\mathcal{K}_0}\|_{H^\infty}^2 = \sigma_{\max}(\mathcal{T}_{\mathcal{K}_0}\mathcal{T}_{\mathcal{K}_0}^*) \leq \sigma_{\max}(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*) = \|\mathcal{T}_\mathcal{K}\|_{H^\infty}^2 \tag{26}$$

So it means $\mathcal{K}_0$ is also optimal for H^∞ estimation. However, one key observation for H^∞ estimation is that in general the solution is *not* unique. Indeed, if we denote

$$\gamma_s = \sqrt{\sigma_{\max}(\mathcal{T}_{\mathcal{K}_0}\mathcal{T}_{\mathcal{K}_0}^*)} = \sqrt{\sigma_{\max}(\mathcal{L}(I+\mathcal{H}^*\mathcal{H})^{-1}\mathcal{L}^*)}, \tag{27}$$

then $\mathcal{K}$ is H^∞ optimal if and only if $\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^* \leq \gamma_s^2 I$. As we show below, many filters $\mathcal{K}$ satisfy this relationship.

Theorem 4. *[HSK99, Theorem 10.4.1] All H^∞ estimators are given by*

$$\mathcal{K} = \mathcal{L}\mathcal{H}^*(I+\mathcal{H}\mathcal{H}^*)^{-1} + \Delta S(I+\mathcal{H}\mathcal{H}^*)^{-\frac{1}{2}} \tag{28}$$

Where the matrix Δ is given by

$$\Delta\Delta^* = \gamma_s^2 I - \mathcal{T}_{\mathcal{K}_0}\mathcal{T}_{\mathcal{K}_0}^* \tag{29}$$

And the matrix S is any contractive operator. Precisely,

$$SS^* \leq I \tag{30}$$

We must again emphasize that an estimator being H^∞ optimal does in no way guarantee or imply that it is a *good* estimator. The solution to H^∞ estimation is not unique and in fact will include solutions with high H^2 error.

References

[HSK99] Babak Hassibi, Ali H Sayed, and Thomas Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*, volume 16. Siam, 1999.