H^∞ ADAPTIVE FILTERING

Babak Hassibi, *Thomas Kailath*

Information Systems Laboratory
Stanford University, Stanford CA 94305

ABSTRACT

H^∞ optimal estimators guarantee the smallest possible estimation error energy over all possible disturbances of fixed energy, and are therefore robust with respect to model uncertainties and lack of statistical information on the exogenous signals. We have recently shown that if prediction error is considered, then the celebrated LMS adaptive filtering algorithm is H^∞ optimal. In this paper we consider prediction of the filter weight vector itself, and for the purpose of coping with time-variations, exponentially weighted, finite-memory and time-varying adaptive filtering. This results in some new adaptive filtering algorithms that may be useful in uncertain and non-stationary environments. Simulation results are given to demonstrate the feasiblity of the algorithms and to compare them with well-known H^2 (or least-squares based) adaptive filters.

1. INTRODUCTION

In contrast to Wiener and Kalman filter theory which require a priori statistical information of the input data, adaptive filtering has been widely used to cope with time variation of system parameters and lack of such a priori knowledge. Recently, following some pioneering work in robust control theory (see *e.g.* [1]), there has been an increasing interest in minimax estimation (see [4, 5, 6] and the references therein) with the belief that the resulting so-called H^∞ algorithms will be more robust and less sensitive to model uncertainties and parameter variations.

The similarity between the objectives of adaptive filtering and H^∞ estimation leads one to suspect some connection between the two. Indeed it turns out (see [7]) that the celebrated LMS algorithm [2], which is widely used in adaptive filtering, is H^∞ optimal. This result gives more insight into the inherent robustness of LMS and why it has found wide applicability in such a diverse range of problems.

In this paper we further pursue the connections between adaptive filtering and H^∞ estimation by considering algorithms for the prediction of the complete filter weight vector, and by developing a host of H^∞ algorithms to deal with time-variations and non-stationary signals. The goal of this paper is to outline the use of the H^∞ criterion in the design of adaptive filter algorithms. There are, no doubt, a wide variety of other H^∞ adaptive algorithms (not considered here) that could be worthy of further scrutiny.

This research was supported by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Office of Scientific Research under Contract F49620-93-1-0085.

2. ROBUSTNESS AND H^∞ ESTIMATION

H^2-optimal (i.e. least-squares based) estimators, such as the RLS algorithm or Kalman filter, are maximum-likelihood and minimize the expected prediction error energy, if we assume disturbances that are "independent zero-mean Gaussian random variables". However, the question that begs itself is what the performance of such estimators will be if the assumptions on the disturbances are violated, or if there are modelling errors in our model so that the disturbances must include the modelling errors? In other words

- *is it possible that* **small** *disturbances and modelling errors may lead to* **large** *estimation errors?*

Obviously, a nonrobust algorithm would be one for which the above is true, and a robust algorithm would be one for which small disturbances lead to small estimation errors. More explicitly, in the adaptive filtering problem, where we assume an FIR model, the *true* model may be IIR, but we neglect the tail of the filter response since its components are small. However, unless one uses a robust estimation algorithm, it is conceivable that this small modelling error may result in large estimation errors.

The problem of robust estimation is thus an important one, and the H^∞ estimation formulation is an attempt at addressing it. The idea is to come up with estimators that minimize (or in the suboptimal case, bound) the maximum energy gain from the disturbances to the estimation errors. This will guarantee that if the disturbances are small (in energy) then the estimation errors will be as small as possible (in energy), *no matter what the disturbances are.* In other words the maximum energy gain is minimized over *all possible* disturbances. The robustness of the H^∞ estimators arises from this fact. Since they make no assumption about the disturbances, they have to accomodate for all conceivable disturbances, and are thus over-conservative.

The following definition implies that the H^∞ norm may be regarded as a maximum *energy gain.*

Definition 1 (The H^∞ Norm) *Let h_2 denote the vector space of square-summable complex-valued causal sequences with inner product $<\{f_k\},\{g_k\}> = \sum_{k=0}^{\infty} f_k^* g_k$, where $*$ denotes complex conjugation. Let T be a transfer operator that maps an input sequence $\{u_i\}$ to an output sequence $\{y_i\}$. Then the H^∞ norm of T is defined as*

$$\|T\|_\infty = \sup_{u\neq 0, u\in h_2} \frac{\|y\|_2}{\|u\|_2}$$

where the notation $\|u\|_2$ denotes the h_2-norm of the causal sequence $\{u_k\}$, viz., $\|u\|_2^2 = \sum_{k=0}^{\infty} u_k^* u_k$.

2.1. Problem Formulations

In adaptive filtering we assume that we observe an output sequence $\{d_i\}$ that obeys the following linear filter model

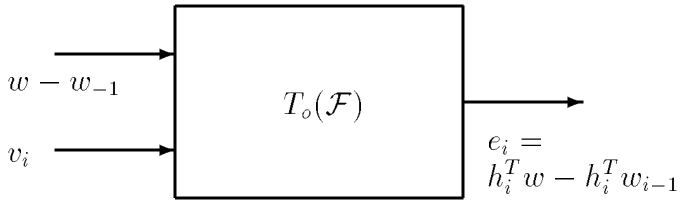
$$d_i = h_i^T w + v_i, \tag{1}$$

where $h_i^T = \left[\begin{array}{cccc} h_{i1} & h_{i2} & \ldots & h_{in} \end{array}\right]$ is a known input vector, w is the unknown filter weight vector that we intend to estimate, and $\{v_i\}$ is an unknown disturbance sequence that may include modelling errors. Let $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ denote the estimate of w given the observations $\{d_j\}$ and $\{h_j\}$ from time 0 up to and including time i.

Since we are first interested in predicting the output of the filter, we define the output prediction error as

$$e_i = h_i^T w - h_i^T w_{i-1},$$

i.e., as the difference between the *uncorrupted* output $h_i^T w$ and $h_i^T w_{i-1}$, the output predicted at time $i-1$. Any choice of estimation strategy $\mathcal{F}(.)$ will induce a transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(w - w_{-1}), \{v_j\}_{j=0}^i\}$ (where w_{-1} is an initial estimate of the weight vector w and μ is a positive constant reflecting a priori knowledge of how close w is to w_{-1}) to the output prediction errors $\{e_j\}_{j=0}^i$, that we shall denote by $T_{o,i}(\mathcal{F})$. See Figure 1.



æ

Figure 1: Transfer operator from disturbances to output prediction error.

In the H^∞ framework, robustness is ensured by minimizing the maximum energy gain from the disturbances to the estimation errors. This leads to the following problem.

Problem 1 (Output Prediction) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that minimizes $\|T_{o,i}(\mathcal{F})\|_\infty$, and obtain the resulting*

$$\begin{aligned} \gamma_o^2 &= \inf_{\mathcal{F}} \;\; \|T_{o,i}(\mathcal{F})\|_\infty^2 \\ &= \inf_{\mathcal{F}} \;\; \sup_{w,v \in h_2} \;\; \frac{\sum_{j=0}^i |e_j|^2}{\mu^{-1}|w-w_{-1}|^2 + \sum_{j=0}^i |v_j|^2}. \end{aligned} \tag{2}$$

In some applications (*e.g.*, in system identification) one is interested in estimating the weight vector itself. In such cases we need to define the weight prediction error $\tilde{w}_i = w - w_{i-1}$. As before, any choice of estimator $\mathcal{F}(.)$ will induce a transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(w - w_{-1}), \{v_j\}_{j=0}^i\}$ to the weight prediction errors $\{\tilde{w}_j\}_{j=0}^i$. This transfer operator we designate by $T_{s,i}(\mathcal{F})$.

Problem 2 (Weight Prediction) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that minimizes $\|T_{s,i}(\mathcal{F})\|_\infty$, and obtain the resulting*

$$\gamma_s^2 = \inf_{\mathcal{F}} \|T_{s,i}(\mathcal{F})\|_\infty^2.$$

In the above two problems we have asumed that the weight vector w is constant in time. However, in many applications we need to cope with time-variations in w itself. One approach for such non-stationary situations is to use a so-called *exponential window*. The exponential window gives (exponentially) larger weight to the more recent data, whereby we may be able to compensate for a time-varying w. In particular, the prediction error and disturbance energies will be computed as:

$$\sum_{j=0}^i \lambda^{-j}|e_j|^2 \quad \text{and} \quad \sum_{j=0}^i \lambda^{-j}|v_j|^2, \tag{3}$$

where $0 < \lambda < 1$ is the so-called forgetting factor that is chosen based upon a priori knowledge of how fast the weight vector varies with time.

Now for any choice of estimator $\mathcal{F}$, we shall denote by $T_{\lambda,i}(\mathcal{F})$ the transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(w - w_{-1}), \{\lambda^{-\frac{j}{2}} v_j\}_{j=0}^i\}$ to the prediction errors $\{\lambda^{-\frac{j}{2}} e_j\}_{j=0}^i$. We are thus lead to the following problem.

Problem 3 (Exponential Weights) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that minimizes $\|T_{\lambda,i}(\mathcal{F})\|_\infty$, and obtain the resulting*

$$\begin{aligned} \gamma_\lambda^2 &= \inf_{\mathcal{F}} \;\; \|T_{\lambda,i}(\mathcal{F})\|_\infty^2 \\ &= \inf_{\mathcal{F}} \;\; \sup_{w,v \in h_2} \;\; \frac{\sum_{j=0}^i \lambda^{-j}|e_j|^2}{\mu^{-1}|w-w_{-1}|^2 + \sum_{j=0}^i \lambda^{-j}|v_j|^2} \end{aligned} \tag{4}$$

Another approach for dealing with time-variations is the so-called sliding or finite-memory window. In this case one only considers the last L data points. Thus the prediction error and disturbance energies are computed as

$$\sum_{j=i-L+1}^i |e_j|^2 \quad \text{and} \quad \sum_{j=i-L+1}^i |v_j|^2, \tag{5}$$

respectively.

Defining by $T_{L,i}(\mathcal{F})$, the transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(w - w_{-1}), \{v_j\}_{j=i-L+1}^i\}$ to the prediction errors $\{e_j\}_{j=i-L+1}^i$, we have the following problem.

Problem 4 (Finite Memory Pr.) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that minimizes $\|T_{L,i}(\mathcal{F})\|_\infty$, and obtain the resulting*

$$\gamma_L^2 = \inf_{\mathcal{F}} \|T_{L,i}(\mathcal{F})\|_\infty^2.$$

2.2. Solutions

Finding the optimum value of γ in the preceding problems essentialy amounts to finding the maximum singular value of a linear time-varying operator. Upper bounds on γ can be found by checking for the positivity of the solution of a certain time-varying discrete-time Riccati recursion. Although both approaches can be used in principle, they require knowledge of *all* the input data vectors $\{h_i\}$.

Since in adaptive filtering problems we are given, and are forced to process, the data in real time, we cannot store all the data and use the aforementioned methods to compute bounds for γ. Therefore the main effort in H^∞ adaptive filtering is to obtain bounds on γ that use simple a priori knowledge of the $\{h_i\}$ and not their explicit values.

Once such bounds are found, the adaptive filters readily follow from the standard solution to the H^∞ estimation problem (see *e.g.* [4]). In what follows we shall call the input vectors $\{h_i\}$ *exciting* if,

$$\lim_{N\to\infty}\sum_{i=0}^{N} h_i^T h_i = \infty.$$

Moreover, we shall define

$$\bar{h} = \sup_i h_i^T h_i \quad , \quad \underline{h} = \inf_i h_i^T h_i$$

and

$$R_i = \frac{1}{i}\sum_{j=0}^{i-1} h_j h_j^T \quad , \quad R_i^L = \sum_{j=i-L+1}^{i} h_j h_j^T .$$

Theorem 1 (Solution to Problem 1) *If the input vectors h_i are exciting and*

$$\mu\bar{h} < 1, \tag{6}$$

then

$$\gamma_o = 1. \tag{7}$$

If this is the case, an optimal H^∞ estimator is given by the LMS algorithm with learning rate μ, viz.

$$w_i = w_{i-1} + \mu h_i (d_i - h_i^T w_{i-1}) \qquad , \qquad w_{-1} \tag{8}$$

Note that, according to Theorem 1, the LMS algorithm guarantees that the energy of the prediction errors will never exceed the energy of the disturbances.

Note also that, via (6), Theorem 1 gives an upper bound on the learning rate μ that guarantess the H^∞ optimality of LMS. This is in accordance with the well-known fact that LMS behaves poorly if the learning rate is chosen too large.

Theorem 2 (Solution to Problem 2)

$$\gamma_s = \inf_i \sqrt{\frac{1}{\frac{1}{\mu(i+1)} + \frac{i}{i+1}\bar{\sigma}^2(R_i)}}, \tag{9}$$

where $\bar{\sigma}(R_i)$ denotes the maximum singular value of R_i. An optimal H^∞ estimator is given by

$$w_i = w_{i-1} + \frac{P_i h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T w_{i-1}), \qquad w_{-1} \tag{10}$$

where P_i satisfies the recursion

$$P_{i+1}^{-1} = P_i^{-1} + h_i h_i^T - \gamma_s^{-2} I, \tag{11}$$

initialized with $P_0^{-1} = (\mu^{-1} - \gamma_s^{-2})I$.

Comparing the algorithm of Theorem 2 with the RLS algorithm [3], we note that the only difference is in the covariance update which, due to the subtraction of the diagonal matrix $\gamma_s^{-2}I$, is more conservative than that of RLS. This ensures that P_i, and hence the gain vector in Theorem 2, do not tend to zero, and is reminiscent of some ad-hoc schemes that are employed with RLS to guarantee that the gain vector does not go to zero (see [3]).

Theorem 3 (Solution to Problem 3)

$$\gamma_\lambda^2 \leq max\ (\ \mu\bar{h}\ ,\ 1 + \frac{1-\lambda}{\lambda}.\frac{\bar{h}}{\underline{h}}\ ). \tag{12}$$

An H^∞ estimator is given by (10), where now P_i satisfies

$$P_{i+1}^{-1} = \lambda P_i^{-1} + \lambda h_i h_i^T - \gamma_\lambda^{-2} h_{i+1} h_{i+1}^T, \tag{13}$$

initialized with $P_0^{-1} = \mu^{-1} I - \gamma_\lambda^{-2} h_0 h_0^T$.

Note that if $\mu\bar{h} < 1$ (in accordance with (6)) then

$$\gamma_\lambda^2 \leq 1 + \frac{1-\lambda}{\lambda}.\frac{\bar{h}}{\underline{h}}.$$

The second term in the above expression shows the deviation from the optimum value of $\gamma = 1$, that was obtained in Theorem 1, and that we must pay for because of the time-variation in the weight vector w.

Theorem 4 (Solution to Problem 4)

$$\gamma_L^2 \leq \sup_i \frac{\bar{h} + \bar{\sigma}(R_i^L)}{\frac{1}{\mu} + \bar{\sigma}(R_i^L)}. \tag{14}$$

An H^∞ estimator is given by the following equations

$$w_{i-1}^d = w_{i-1} + \frac{P_i^d h_{i-L}}{-1 + h_{i-L}^T P_i^d h_{i-L}}(d_{i-L} - h_{i-L}^T w_{i-1}) \tag{15}$$

for "downdating", with

$$(P_i^d)^{-1} = P_i^{-1} - (1 - \gamma_L^{-2}) h_{i-L} h_{i-L}^T, \tag{16}$$

and

$$w_i = w_{i-1}^d + \frac{P_i h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T w_{i-1}^d) \tag{17}$$

for "updating", with

$$P_{i+1}^{-1} = (P_i^d)^{-1} + (1 - \gamma_L^{-2}) h_{i+1} h_{i+1}^T. \tag{18}$$

Note, from Theorem 4, that if $\mu\bar{h} < 1$ then $\gamma_L < 1$, and that if $\mu\bar{h} > 1$ then $\gamma_L > 1$. However, the case $\mu\bar{h} = 1$ deserves special attention since it leads to the following LMS-type finite-memory algorithm.

Corollary 1 (Finite Memory LMS) *Suppose that $\mu\bar{h} = 1$. Then $\gamma_L = 1$, and an H^∞ optimal estimator is given by the following LMS-type algorithm*

$$w_{i-1}^d = w_{i-1} - \mu h_{i-L}(d_{i-L} - h_{i-L}^T w_{i-1}) \tag{19}$$

for "downdating", and

$$w_i = w_{i-1}^d + \mu h_i (d_i - h_i^T w_{i-1}^d) \tag{20}$$

for "updating".

2.3. General Time-Variation

In this section we shall consider a time-varying filter model of the form

$$d_i = h_i^T x_i + v_i, \tag{21}$$

where the only difference with (1) is that $\{x_i\}$, the weight vector we intend to estimate, is time-varying. As before, we shall denote by $\hat{x}_i = \mathcal{F}(d_0, d_1, \ldots, d_{i-1})$ the prediction of the weight vector x_i, and define the output prediction error as

$$e_i = h_i^T x_i - h_i^T \hat{x}_i.$$

Note that since the time variation in the weight vector x_i,

$$\delta x_i = x_{i+1} - x_i$$

is *unknown*, we shall consider it as a disturbance. Thus for every choice of estimator $\mathcal{F}$ we will have a transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(x_0 - \hat{x}_0), \{v_j\}_{j=0}^{i}, q^{-\frac{1}{2}}\{\delta x_j\}_{j=0}^{i}\}$ (where q is a positive constant that reflects a priori knowledge of how rapidly the weight vector x_i varies with time) to the prediction errors $\{e_j\}_{j=0}^{i}$, that we shall denote by $T_{g,i}(\mathcal{F})$. We are thus led to the following problem.

Problem 5 (Time-Variation) *Find an H^∞-optimal estimation strategy $\hat{x}_i = \mathcal{F}(d_0, d_1, \ldots, d_{i-1})$ that minimizes $\|T_{g,i}(\mathcal{F})\|_\infty$, and obtain the resulting*

$$\gamma_g^2 = \inf_{\mathcal{F}} \quad \|T_{g,i}(\mathcal{F})\|_\infty^2$$

$$= \inf_{\mathcal{F}} \quad \sup_{x_0, v, \delta x \in h_2} \quad \frac{\|e\|_2^2}{\mu^{-1}|x_0 - \hat{x}_0|^2 + \|v\|_2^2 + q^{-1}\|\delta x\|_2^2}. \tag{22}$$

We have the following solution to the above problem.

Theorem 5 (Solution to Problem 5)

$$\gamma_g^2 \leq 1 + q\bar{h}. \tag{23}$$

An H^∞ estimator is given by

$$\hat{x}_{i+1} = \hat{x}_i + \frac{P_i h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T \hat{x}_i), \qquad \hat{x}_0 \tag{24}$$

where

$$P_i^{-1} = \tilde{P}_i^{-1} - \gamma_g^{-2} h_i h_i^T, \tag{25}$$

and $\tilde{P}_i$ satisfies the recursion

$$\tilde{P}_{i+1} = \left[\tilde{P}_i^{-1} + (1 - \gamma_g^{-2}) h_i h_i^T\right]^{-1} + qI, \tag{26}$$

initialized with $\tilde{P}_0 = \mu I$.

Note, as before, that the second term in the bound $\gamma_g \leq 1 + q\bar{h}$, reflects the deviation from the optimum value (of Theorem 1) that we must incur due to the time-variation in the weight vector x_i.

3. SIMULATION RESULTS

Due to lack of space we shall only describe one typical simulation result here. To this end, consider the model (21) where the weight vector x_i is now a scalar. To reflect time-variation we chose $\delta x_i = .02$, and to reflect both noise and modelling error,

$$v_i = .1 * (h_i x_i)^3 + n_i,$$

where n_i is a zero-mean Gaussian random variable with variance $\sigma^2 = .04$. We chose $x_0 = -1$ and considered 100 time samples so that $x_{100} = 1$. We predicted the output of the filter using various H^∞ and H^2 adaptive algorithms and computed the prediction error energy for each. The resulting prediction error energies were averaged over 50 independent runs, and the results are given in Table 1. The H^∞ algorithms considered were LMS and the algorithms of Theorems 3 and 5, and the H^2 algorithms were RLS, exponentially-weighted RLS (denoted by λ-RLS) and the Kalman filter (denoted by KF). Note that the prediction error energies for the H^∞ algorithms are virtually identical, and that although the exponentially-weighted RLS algorithm performs significantly better than RLS and the Kalman filter, it does not perform as well as the H^∞ algorithms. (The parameters used in this simulation were $\mu = .9$, $\lambda = .9$ and $q = .0004$.)

	LMS	Thm. 3	Thm. 5
$\sum_{j=0}^{100} \|e_j\|^2$	1.48	1.54	1.48
	RLS	λ-RLS	KF
$\sum_{j=0}^{100} \|e_j\|^2$	6.00	2.08	5.11

Table 1: The H^∞ and H^2 algorithms.

4. CONCLUSION

In closing, we should note that if one has a priori knowledge of the underlying statistics and distributions of the signals, one is always best served by considering algorithms that are specifically tuned for the situation at hand. On the other hand, if one does not have such a priori knowledge and uses an algorithm that makes specific assumptions about the disturbances, then the algorithm may perform poorly if these assumptions are not met. H^∞ optimal algorithms will therefore be most applicable in uncertain environments where there may be modelling errors, and where the statistics and/or distributions of the disturbances are not known (or are too expensive to obtain).

5. REFERENCES

[1] G. Zames. Feedback optimal sensitivity: model preference transformation, multiplicative seminorms and approximate inverses. *IEEE Trans. on Automatic Control*, AC-26:301–320, 1981.

[2] B. Widrow and M. E. Hoff, Jr. Adaptive switching circuits. *IRE WESCON Conv. Rec.*, pages 96–104, 1960. Pt. 4.

[3] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ, second edition, 1991.

[4] U. Shaked and Y. Theodor. $H^\infty-$optimal estimation: A tutorial. In *Proc. IEEE Conference on Decision and Control*, pages 2278–2286, Tucson, AZ, Dec. 1992.

[5] P.P. Khargonekar and K. M. Nagpal. Filtering and smoothing in an $H^\infty-$ setting. *IEEE Trans. on Automatic Control*, AC-36:151–166, 1991.

[6] T. Basar. Optimum performance levels for minimax filters, predictors and smoothers. *Systems and Control Letters*, 16:309–317, 1991.

[7] B. Hassibi, A. H. Sayed, and T. Kailath. LMS is H^∞ Optimal. In *Proc. 32nd IEEE Conference on Decision and Control*, pp. 74-80, San Antonio, TX, Dec. 1993.