Chapter 11

H^∞ Adaptive Filtering

The previous two chapters were more or less concerned with the study of adaptive filters from an H^∞ point of view. Chapter 9 studied the problems of output prediction and output filtered estimation in adaptive filtering, using an H^∞ approach, and demonstrated the H^∞ optimality of the LMS and normalized LMS algorithms with respect to these two estimation criteria. Chapter 10 studied the robustness of least-squares-based adaptive filters, such as the RLS algorithm, using the H^∞ framework. A reading of Chapters 9 and 10 indicates that there may be great promise in the interplay of adaptive filtering and H^∞ estimation theory.

To round up the story, in this chapter, we shall present a preliminary study of the design of adaptive filters using an H^∞ criterion. The strength of H^∞-optimal adaptive filters lies in the fact that they 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. Specifically, in this chapter we study the problem of prediction of the 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.

We should mention that the presentation of this chapter is deliberately brief. The main reasons are that the results given here are preliminary, and that we believe the area is worthy of much more scrutiny and study. Indeed, we have just scratched the surface. The main goal, therefore, has been to hint at the possibilities.

11.1 Introduction

Since its inception in the early 1960's [WH60, GWW61] adaptive filtering has been widely used to cope with time-variations of system parameters and lack of a priori knowledge of the statistical properties of the input data. This is in contrast to Wiener and Kalman filter theory which require a priori statistical information. Early successes of adaptive filtering were in channel equalization for digital communications [Luc65] and in adaptive antennas [How65, App66, Wid67]. Currently, recent advances in ASIC's (application-specific processors) and sensor technology have spurred an increasing range of applications from biomedical engineering to wireless communications, so that there is still great interest in the field.

Although there have been several successful applications of adaptive filtering, it is fair to say that most of the results are empirical and that there still is not an adequate theoretical basis for the analysis of the performance of these algorithms — limits of performance, sensitivity, optimality, etc. (The basic reason is that adaptive systems are inherently time-variant and nonlinear). Much work has been done in this area in control theory, especially in adaptive control, though many questions still remain.

Due to the similarity between the objectives of adaptive filtering (coping with time-variations, insensitivity to lack of statistical knowledge, etc.) and H^∞ estimation (robustness with respect to uncertainties in the underlying model and/or statistics), it is expected that there should be some connection between the two. Indeed, we have seen in Chapter that the celebrated LMS algorithm [WH60], which is widely used in adaptive filtering, is H^∞ optimal. This result gives more insight into the inherent robustness of the LMS algorithm and why it has found such wide applicability in such a diverse range of problems. As a matter of fact, we believe that the H^∞ approach yields a new way of looking at adaptive systems, and has several ramifications, as well as suggesting directions for further research.

In this chapter we further pursue the connections between adaptive filtering and H^∞ estimation. In particular, we consider predicting the complete filter weight vector which leads to a new adaptive algorithm. We also develop a host of H^∞ algorithms to deal with time-variations and non-stationary signals. Two of these algorithms are based on exponentially weighted and finite memory windows, respectively, and a third algorithm allows for general time variations in the underlying filter weight vector. The goal of this chapter is to outline the use of the H^∞ criterion in the design 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.

Important Remark

At this point we should make an important remark on the contribution of this chapter. As is wellknown, once an upper bound on the value of γ that ensures the existence of an H^∞ filter is given, the structure of the adaptive filters readily follow from the standard solution to the H^∞ estimation problem (see *e.g.*, Chapter 3). Therefore, since adaptive filtering is just a special case of state-space estimation, that has been solved (in its full generality) in the H^∞ framework, it appears that there is not much more to be done or said.

However, we should note that finding the optimum value of γ essentially amounts to finding the maximum singular value of a linear time-varying operator. 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 do not have the luxury of storing all the data and computing explicit bounds for γ using the aforementioned methods. 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. This is what is done in the results given below.

11.2 Full Weight Vector Estimation

As mentioned in earlier chapters, in adaptive filtering we assume that we observe an output sequence $\{d_i\}$ that obeys the following linear filter model

$$d_i = h_i^T w + v_i, \tag{11.2.1}$$

where $h_i^T = \begin{bmatrix} h_{i1} & h_{i2} & \dots & h_{in} \end{bmatrix}$ 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 modeling errors. Let $\hat{w}_{|i} = \mathcal{F}_i(d_0, d_1, \dots, d_i)$ denote the estimate of the weight vector given the inputs $\{h_j\}$ and the outputs $\{d_j\}$ from time 0 up to and including time i.

In H^∞ estimation the structure of the estimator depends on the linear combination of the state (in our case the linear combination of the weight vector) that we intend to estimate. This is as opposed to H^2 estimation where the estimate of any linear combination of the state is simply that linear combination of the state estimate.

In Chapters 9 and 10 our concern was prediction of the filter output, which corresponds to the linear combination $h_i^T w$. Although in many applications predicting the output of the filter suffices, in some applications (especially those concerned with identification) we are interested in estimating the weight vector itself. In this case we define the weight vector estimation error as

$$\tilde{w}_{|i} = w - \hat{w}_{|i}, \tag{11.2.2}$$

and we are concerned with the transfer operator $\mathcal{T}_{s,N}$, from the disturbances $\{w - \hat{w}_{|-1}, \{v_i\}_{i=0}^N\}$ to the weight estimation errors $\{\tilde{w}_{|i}\}_{i=0}^N$. The H^∞ full weight vector estimation problem may now be formulated as follows, where we have defined $\| \tilde{w} \|_2^2 = \sum_{i=0}^N \tilde{w}_{|i}^T \tilde{w}_{|i}$.

Problem 11.2.1 (Optimal Weight Estimation Problem) *Find an H^∞-optimal estimation strategy $\hat{w}_{|i} = \mathcal{F}_i(d_0, d_1, \dots, d_i)$ that minimizes $\|\mathcal{T}_{s,N}\|_\infty$, and obtain the resulting*

$$\gamma_s^2 = \inf_{\mathcal{F}} \ \|\mathcal{T}_s\|_\infty^2 = \inf_{\mathcal{F}} \ \sup_{w,v\in h^2} \ \frac{\|\tilde{w}\|_2^2}{\mu^{-1}|w - w_{-1}|^2 + \|v\|_2^2} \tag{11.2.3}$$

where $|w - w_{-1}|^2 = (w - w_{-1})^T(w - w_{-1})$*, and* μ *is a positive constant that reflects a priori knowledge as to how close* w *is to the initial guess* w_{-1}*.*

Before stating the solution to the above problem we need to define the sample covariance matrix of the input vectors $\{h_j\}_{j=0}^{i-1}$ as follows

$$R_i = \frac{1}{i}\sum_{j=0}^{i-1} h_j h_j^T. \tag{11.2.4}$$

Theorem 11.2.1 (Full Weight Estimator) *Consider the model (11.2.1), and suppose we wish to minimize the* H^∞ *norm of the transfer operator,* $\mathcal{T}_{s,N}$*, from the unknowns* $w - w_{-1}$ *and* $\{v_i\}_{i=0}^{N}$ *to the weight vector estimation errors* $\tilde{w}_{|i}$*. Then*

$$\gamma_s = \sup_i \sqrt{\frac{1}{\frac{1}{\mu i} + \underline{\sigma}(R_i)}}, \tag{11.2.5}$$

where $\underline{\sigma}(R_i)$ *denotes the minimum singular value of* R_i*. The central optimal* H^∞ *estimator is given by*

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \frac{P_i h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T \hat{w}_{|i-1}), \quad w_{-1} \tag{11.2.6}$$

where P_i *satisfies the recursion*

$$P_{i+1}^{-1} = P_i^{-1} + h_i h_i^T - \gamma_s^{-2} I, \qquad P_0^{-1} = \mu I. \tag{11.2.7}$$

Proof: The proof is a direct application of Theorem 3.2.1 on a posteriori state-space H^∞ estimators for the special state-space model

$$F_i = I \quad , \quad G_i = 0 \quad , \quad H_i = h_i^T \quad , \quad L_i = I_n \tag{11.2.8}$$

which corresponds to the adaptive filtering model (11.2.1). Note that we have taken $L_i = I_n$ since the desired combination of the weights is the identity matrix (*i.e.*, we want to estimate w itself). Now from Theorem 3.2.1 if an estimator of level γ_s exists, then the solution is given by (11.2.6) and (11.2.7). Therefore all that remains to be shown is the condition for the existence of the filter.

But according to Theorem 3.2.1 this condition is

$$P_{i+1}^{-1} = P_i^{-1} + h_i h_i^T - \gamma_s^{-2} I > 0, \quad i = 0, 1, \ldots N. \tag{11.2.9}$$

Now the above Riccati equation can be solved to yield,

$$\begin{aligned} P_{i+1}^{-1} &= \mu^{-1} I + \sum_{j=0}^{i} h_j h_j^T - (i+1)\gamma_s^{-2} I, \\ &= \mu^{-1} I + (i+1) R_{i+1} - (i+1)\gamma_s^{-2} I, \end{aligned}$$

so that the condition for $P_{i+1}^{-1} > 0$ becomes

$$R_{i+1} > \gamma_s^{-2} I - \frac{1}{\mu(i+1)} I,$$

or

$$\underline{\sigma}(R_{i+1}) > \gamma_s^{-2} - \frac{1}{\mu(i+1)}. \tag{11.2.10}$$

Rearranging the above inequality yields,

$$\gamma_s > \sqrt{\frac{1}{\frac{1}{\mu i} + \underline{\sigma}(R_i)}}, \quad i = 0, 1, \ldots N \tag{11.2.11}$$

from which we obtain the desired result (11.2.5).

■

Remarks:

(i) It is interesting to compare the algorithm of Theorem 11.2.1 with the RLS algorithm of Chapter 10. The only difference is that the covariance update in RLS can be written as $P_{i+1}^{-1} = P_i^{-1} + h_i h_i^T$. Due to this covariance update, P_i (and hence the gain vector) may approach zero for large i. However, in the algorithm of Theorem 11.2.1 the covariance update is more conservative and P_i does not tend to zero meaning that we always have a non-zero gain vector. This is similar to some ad-hoc schemes that are employed with RLS to guarantee that the gain vector does not go to zero (see [Hay96]).

(ii) It is interesting to note that $\underline{\sigma}(R_i)$, the minimum singular value of R_i, appears in the bound for γ_s. Indeed if $\underline{\sigma}(R_i) = 0$, then γ_s grows unbounded as time progresses to infinity. This makes perfect sense, since $\underline{\sigma}(R_i) = 0$ means that the sample covariance matrix $R_i = \frac{1}{i}\sum_{j=0}^{i-1} h_j h_j^T$ is singular, *i.e.*, that there are certain directions which the input vectors $\{h_i\}$ do not (persistently) excite. In such a case it will not be possible to estimate w along those directions, and hence γ_s will tend to infinity.

(iii) If the components of the input vectors $\{h_i\}$ are independent random variables with variance σ^2, then provided $\mu\sigma < 1$, for large i the algorithm of Theorem 11.2.1 reduces to

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \frac{\mu}{1-\mu\sigma}.\frac{h_i}{1+\frac{\mu}{1-\mu\sigma}h_i^T h_i}(d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{|-1} \tag{11.2.12}$$

which is the so-called normalized LMS algorithm with parameter $\frac{\mu}{1-\mu\sigma}$.

11.3 Time-Variation

The H^∞ adaptive filters developed in Chapter 9 and the previous section are robust with respect to model uncertainties and lack of statistical information on the exogenous signals. However, since we have assumed the underlying model to be stationary (we have taken the unknown weight vector as time-invariant) these algorithms may not have desirable tracking properties in the face of time variations.

In this section we shall incorporate provisions to cope for time variations in the underlying model. The methods used are similar to those used in the usual H^2 setting, and here we generalize them to the H^∞ setting. Although this can be done for both the output estimation case and the full weight vector case, for brevity we shall only consider the former.

11.3.1 Exponentially-Windowed Adaptive Filtering

One method to account for time-variations is to introduce a forgetting factor $0 < \lambda < 1$, and to exponentially weight the energies with this forgetting factor. In this manner the more recent data will be given larger weights than the earlier ones. This will allow the algorithms to track the time-variations of the underlying models. In particular, the prediction error and disturbance energies are computed as:

$$\sum_{j=0}^{N} \lambda^{-i}|e_{p,i}|^2 \quad \text{and} \quad \sum_{j=0}^{N} \lambda^{-i}|v_i|^2. \tag{11.3.1}$$

The forgetting factor $0 < \lambda < 1$ 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 $\mathcal{T}_{\lambda,N}(\mathcal{F})$ the transfer operator from the disturbances $\{\mu^{-\frac{1}{2}}(w - \hat{w}_{|-1}), \{\lambda^{-\frac{i}{2}}v_i\}_{i=0}^{N}\}$ to the prediction errors $\{\lambda^{-\frac{i}{2}}e_{p,i}\}_{i=M}^{i}$. The H^∞ output prediction problem with exponential weighting can thus be formulated as follows.

Problem 11.3.1 (Exponentially-Windowed Problem) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that minimizes $\|\mathcal{T}_{\lambda,N}\|_\infty$, and obtain the resulting*

$$\gamma_\lambda^2 = \inf_{\mathcal{F}} \ \|\mathcal{T}_{\lambda,N}\|_\infty^2 = \inf_{\mathcal{F}} \ \sup_{w,v\in h^2} \frac{\sum_{i=0}^{N} |e_{p,i}|^2\lambda^{-i}}{\mu^{-1}|w - w_{-1}|^2 + \sum_{i=0}^{N} |v_i|^2\lambda^{-i}} \tag{11.3.2}$$

where $|w - w_{-1}|^2 = (w - w_{-1})^T(w - \hat{w}_{|-1})$, and μ is a positive constant that reflects a priori knowledge as to how close w is to the initial guess $\hat{w}_{|-1}$.

Before giving the solution to the above problem we need the following definitions

$$\bar{h} \triangleq \sup_i h_i^T h_i \quad , \quad \underline{h} \triangleq \inf_i h_i^T h_i \tag{11.3.3}$$

and

$$R_{\lambda,i}\Delta{=}\lambda^i \sum_{j=0}^{i-1} \lambda^{-j} h_j h_j^T. \tag{11.3.4}$$

Theorem 11.3.1 (Exponentially-Windowed Algorithm) *Consider the model (11.2.1), and suppose we wish to minimize the H^∞ norm of the transfer operator $T_{\lambda,N}$ of Problem 11.3.1. Then*

$$\gamma_\lambda^2 \le \sup_i \frac{\bar{h} + \underline{\sigma}(R_{\lambda,i})}{\frac{\lambda^i}{\mu} + \underline{\sigma}(R_{\lambda,i})} \quad i = 0, 1, \ldots N \tag{11.3.5}$$

and the central H^∞-optimal estimator is given by

$$\hat{w}|i = \hat{w}_{|i-1} + \frac{\bar{P}_i h_i}{1 + h_i^T \bar{P}_i h_i}(d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{|-1} \tag{11.3.6}$$

where $\bar{P}_i$ satisfies the recursion

$$\bar{P}_{i+1}^{-1} = \lambda \bar{P}_i^{-1} + \lambda h_i h_i^T - \gamma_\lambda^{-2} h_{i+1} h_{i+1}^T, \qquad \bar{P}_0^{-1} = \mu^{-1} I - \gamma_\lambda^{-2} h_0 h_0^T. \tag{11.3.7}$$

Proof: The proof is a direct application of Theorem 3.2.2 on a priori H^∞ estimation as applied to a Krein space state-space model with

$$F_i = I \quad , \quad H_i = \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} \quad , \quad G_i = 0 \quad , \quad \Pi_0 = \mu I \quad , \quad R_i = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma_\lambda^2 \end{bmatrix} \lambda^i \tag{11.3.8}$$

and with $x_0 = w$. Therefore, using Theorem 3.2.2, assuming a solution exists, the solution is given by

$$\hat{x}_{i+1} = \hat{x}_i + \tilde{P}_i h_i (\lambda^i + h_i^T \tilde{P}_i h_i)^{-1} (d_i - h_i^T \hat{x}_i), \quad \hat{x}_0 = \hat{w}_{|-1} \tag{11.3.9}$$

or, since $x_i = w$, for all i

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \tilde{P}_i h_i (\lambda^i + h_i^T \tilde{P}_i h_i)^{-1} (d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{|-1} \tag{11.3.10}$$

where

$$\tilde{P}_i^{-1} = P_i^{-1} - \gamma_\lambda^{-2} \lambda^{-i} h_i h_i^T, \tag{11.3.11}$$

and where P_i satisfies the Riccati recursion

$$P_{i+1} = P_i - P_i \begin{bmatrix} h_i & h_i \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -\gamma_\lambda^2 \end{bmatrix} \lambda^i + \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} P_i \begin{bmatrix} h_i & h_i \end{bmatrix} \right\}^{-1} \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} P_i, \quad P_0 = \mu I. \tag{11.3.12}$$

Therefore we have

$$P_{i+1}^{-1} = P_i^{-1} + (1 - \gamma_\lambda^{-2}) \lambda^{-i} h_i h_i^T, \quad P_0^{-1} = \mu^{-1} I \tag{11.3.13}$$

so that

$$\tilde{P}_{i+1}^{-1} = \tilde{P}_i^{-1} + \lambda^{-i} h_i h_i^T - \gamma_\lambda^{-2} \lambda^{-i-1} h_{i+1} h_{i+1}^T, \quad \tilde{P}_0^{-1} = \mu^{-1} I - \gamma_\lambda^{-2} h_0 h_0^T. \tag{11.3.14}$$

Now if we define

$$\bar{P}_i = \lambda^{-i}\tilde{P}_i, \tag{11.3.15}$$

we see that the recursion for $\hat{w}_{|i}$ becomes

$$\hat{w}|i = \hat{w}_{|i-1} + \frac{\bar{P}_i h_i}{1+h_i^T\bar{P}_i h_i}(d_i - h_i^T\hat{w}_{|i-1}), \quad \hat{w}_{|-1} \tag{11.3.16}$$

which is precisely (11.3.6), and, moreover, $\bar{P}_i$ satisfies the recursion

$$\bar{P}_{i+1}^{-1} = \lambda\bar{P}_i^{-1} + \lambda h_i h_i^T - \gamma_\lambda^{-2}h_{i+1}h_{i+1}^T, \qquad \bar{P}_0^{-1} = \mu^{-1}I - \gamma_\lambda^{-2}h_0h_0^T. \tag{11.3.17}$$

which is precisely (11.3.7).

Therefore all that remains to be shown is the existence condition for the filter. But from Theorem 3.2.2 this existence condition is

$$\tilde{P}_i > 0, \quad i = 0,1,\dots N. \tag{11.3.18}$$

Now the recursion (11.3.14) can be solved for $\tilde{P}_i$ to yield

$$\begin{aligned}\tilde{P}_{i+1}^{-1} &= \mu^{-1}I + \sum_{j=0}^{i}(1-\gamma_\lambda^{-2})\lambda^{-j}h_jh_j^T - \gamma_\lambda^{-2}\lambda^{-i-1}h_{i+1}h_{i+1}^T \\ &= \mu^{-1}I + (1-\gamma_\lambda^{-2})\lambda^{-i-1}R_{\lambda,i+1} - \gamma_\lambda^{-2}\lambda^{-i-1}h_{i+1}h_{i+1}^T.\end{aligned}$$

Therefore the condition $\tilde{P}_{i+1} > 0$ implies

$$\mu^{-1}I + (1-\gamma_\lambda^{-2})\lambda^{-i-1}R_{\lambda,i+1} > \gamma_\lambda^{-2}\lambda^{-i-1}h_{i+1}h_{i+1}^T,$$

or, equivalently,

$$(1-\gamma_\lambda^{-2})R_{\lambda,i+1} > \gamma_\lambda^{-2}h_{i+1}h_{i+1}^T - \frac{\lambda^{i+1}}{\mu}I. \tag{11.3.19}$$

A sufficient condition for the above inequality is that

$$\begin{aligned}(1-\gamma_\lambda^{-2})\underline{\sigma}(R_{\lambda,i+1}) &> \gamma_\lambda^{-2}\bar{\sigma}(h_{i+1}h_{i+1}^T) - \frac{\lambda^{i+1}}{\mu} \\ &= \gamma_\lambda^{-2}\bar{h} - \frac{\lambda^{i+1}}{\mu}.\end{aligned}$$

Rearranging this last expressions leads to

$$\gamma_\lambda^2 > \frac{\bar{h} + \underline{\sigma}(R_{\lambda,i+1})}{\frac{\lambda^{i+1}}{\mu} + \underline{\sigma}(R_{\lambda,i+1})} \tag{11.3.20}$$

which yields the desired result (11.3.5).

∎

Remarks:

(i) It is interesting to note that $\underline{\sigma}(R_{\lambda,i})$, the minimum singular value of $R_{\lambda,i}$, appears in the bound for γ_λ. Moreover, it follows from (11.3.5) that if $\underline{\sigma}(R_{\lambda,i}) = 0$, then γ_λ grows unbounded as time progresses to infinity.

(ii) For large i the bound in (11.3.5) becomes

$$\gamma_\lambda^2 \leq 1 + \frac{\bar{h}}{\underline{\sigma}(R_{\lambda,i})}, \tag{11.3.21}$$

Note that for $\lambda = 1$, we have $\underline{\sigma}(R_{\lambda,i}) \to \infty$ so that we are reduced to the bound $\gamma = 1$ of H^∞ output prediction adaptive filtering (that was solved by LMS). Therefore the extra term $\frac{\bar{h}}{\underline{\sigma}(R_{\lambda,i})}$ represents the increase in the H^∞ norm we must incur due to the time-varying nature of the problem.

(iii) There is a similar exponentially-windowed a posteriori solution that can be constructed for filtered errors. However, for brevity we shall not do so here.

11.3.2 Finite-Memory Adaptive Filtering

Another approach for dealing with time-variations is the so-called sliding or finite-memory window. In this case one only considers data over a finite window of length L. Therefore as each new data is observed, the least recent data point is discarded so that we have a memory of constant size L. Due to the fact that old data is discarded, this method has the promise to cope with time variations in the underlying model.

In this framework, the prediction error and disturbance energies are computed as

$$\sum_{j=i-L+1}^{i} |e_{p,j}|^2 \quad \text{and} \quad \sum_{j=i-L+1}^{i} |v_j|^2, \tag{11.3.22}$$

respectively.

The H^∞ finite memory adaptive filtering problem may now be stated as follows.

Problem 11.3.2 (Finite-Memory Problem) *Find an H^∞-optimal estimation strategy $w_i = \mathcal{F}(d_0, d_1, \ldots, d_i)$ that achieves*

$$\gamma_L^2 = \inf_{\mathcal{F}} \sup_{w,v\in h^2} \frac{\sum_{j=i-L+1}^{i} |e_{p,j}|^2}{\mu^{-1}|w-w_{-1}|^2 + \sum_{j=i-L+1}^{i} |v_j|^2}, \quad i = 0, 1, \ldots N \tag{11.3.23}$$

where $|w-w_{-1}|^2 = (w-w_{-1})^T(w-w_{-1})$, and μ is a positive constant that reflects a priori knowledge as to how close w is to the initial guess w_{-1}.

To give the solution to the above problem we need to define the following finite-memory covariance matrix

$$R_i^L = \sum_{j=i-L+1}^{i} h_j h_j^T. \tag{11.3.24}$$

Theorem 11.3.2 (Finite-Memory Algorithm) *Consider the model (11.2.1), and suppose we would like to solve the adaptive problem 11.3.2. Then*

$$\gamma_L^2 \le \sup_i \frac{\bar{h} + \underline{\sigma}(R_i^L)}{\frac{1}{\mu} + \underline{\sigma}(R_i^L)}, \quad i = 0, 1, \ldots N \tag{11.3.25}$$

and the central optimal H^∞ estimator is given by the following equations

- *For downdating*

$$w_{i-1}^d = w_{i-1} + \frac{P_i^d h_{i-L}}{-1 + h_{i-L}^T P_i^d h_{i-L}}(d_{i-L} - h_{i-L}^T w_{i-1}) \tag{11.3.26}$$

  with

$$(P_i^d)^{-1} = P_i^{-1} - (1-\gamma_m^{-2}) h_{i-L} h_{i-L}^T.$$

- *For updating*

$$w_i = w_{i-1}^d + \frac{P_i^d h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T w_{i-1}^d) \tag{11.3.27}$$

  with

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

Proof: The proof is another application of the a priori H^∞ filtering result of Theorem 3.2.2. Since the proof is almost identical to the proof of Theorem 11.3.1 we shall omit it for brevity.

■

Remarks:

(i) Note, once more, that a minimum singular value, here $\underline{\sigma}(R_i^L)$ has entered the bound for γ_L. Now, however, if $\underline{\sigma}(R_i^L) = 0$, γ_L does not become unbounded. [Indeed it is bounded by $\mu\bar{h}$.

(ii) Note, from Theorem 11.3.2, 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 11.3.1 (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*

$$w_{i-1}^d = w_{i-1} - \mu h_{i-L}(d_{i-L} - h_{i-L}^T w_{i-1}) \tag{11.3.28}$$

for "downdating", and

$$w_i = w_{i-1}^d + \mu h_i(d_i - h_i^T w_{i-1}^d) \tag{11.3.29}$$

for "updating".

(iii) There is also a similar finite-memory a posteriori solution that can be constructed for filtered errors. However, for brevity we shall not do so here.

11.3.3 General Time-Variation

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

$$d_i = h_i^T x_i + v_i, \tag{11.3.30}$$

where $\{d_i\}$ is the observed output sequence, $\{h_i\}$ is the known input vector, $\{x_i\}$ is the unknown time-varying weight vector that we intend to estimate, and $\{v_i\}$

is an unknown disturbance that may include modeling errors. We shall denote by $\hat{x}_i = \mathcal{F}(d_0, d_1, \ldots, d_{i-1})$ the estimate of the weight vector x_i given observations $\{d_j\}$ from time 0 up to and including time $i-1$. The prediction error will therefore be

$$e_{p,i} = h_i^T x_i - h_i^T \hat{x}_i. \tag{11.3.31}$$

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

$$\delta x_i = x_{i+1} - x_i, \tag{11.3.32}$$

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 $\{\mu^{-\frac{1}{2}}(x_0 - \hat{x}_0), \{v_i\}_{i=0}^N, \{\delta x_i\}_{i=0}^N\}$ to the prediction errors $\{e_{p,i}\}_{i=0}^N$, that we shall denote by $\mathcal{T}_{g,N}(\mathcal{F})$. We are thus immediately led to the following problem.

Problem 11.3.3 (Time-Varying Problem) *Find an H^∞-optimal estimation strategy $\hat{x}_i = \mathcal{F}(d_0, d_1, \ldots, d_{i-1})$ that minimizes $\|\mathcal{T}_{g,N}\|_\infty$, and obtain the resulting*

$$\gamma_g^2 = \inf_{\mathcal{F}} \|\mathcal{T}_{g,N}\|_\infty^2 = \inf_{\mathcal{F}} \sup_{x_0, v, \tilde{x} \in h^2} \frac{\|e_p\|_2^2}{\mu^{-1}|x_0 - \hat{x}_0|^2 + \|v\|_2^2 + q^{-1}\|\delta x\|_2^2} \tag{11.3.33}$$

where $|x_0 - \hat{x}_0|^2 = (x_0 - \hat{x}_0)^T(x_0 - \hat{x}_0)$, and μ and q are positive constants that respectively reflect a priori knowledge as to how close x_0 is to the initial guess $\hat{x}_0$, and as to how rapid the weight vector x_i varies with time.

Note that for a filter that varies slowly with time q will typically be very small.

Theorem 11.3.3 (Time-Varying Algorithm) *Consider the model (11.3.30), and suppose we wish to minimize the H^∞ norm of the transfer operator $\mathcal{T}_{g,N}$ from the disturbances $\{x_0 - \hat{x}_0, \{v_i, \delta x_i\}_{i=0}^N\}$ to the prediction errors $\{e_{p,i}\}_{i=0}^N$. Then*

$$\gamma_q^2 \leq \sup_i \left[q + \frac{1}{h_i^T h_i}\right] h_{i+1}^T h_{i+1}, \quad i = 0, 1 \ldots N \tag{11.3.34}$$

and the central H^∞-optimal estimator is given by

$$\hat{x}_{i+1} = \hat{x}_i + \frac{\tilde{P}_i h_i}{1 + h_i^T \tilde{P}_i h_i}(d_i - h_i^T \hat{x}_i), \quad \hat{x}_0 \tag{11.3.35}$$

where $\tilde{P}_i^{-1} = P_i^{-1} - \gamma_g^{-2} h_i h_i^T$, *and*

$$P_{i+1} = \left[P_i^{-1} + (1-\gamma_g^{-2}) h_i h_i^T\right]^{-1} + qI, \quad P_0 = \mu I. \tag{11.3.36}$$

Proof: This is another application of the a priori H^∞ filter of Theorem 3.2.2 where now the Krein space model is,

$$F_i = I_n \ , \ G_i = I_n \ , \ H_i = \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} \ , \ \Pi_0 = \mu I_n \ , \ Q_i = qI_n \ , \ R_i = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma_g^2 \end{bmatrix}. \tag{11.3.37}$$

This implies that the central solution (assuming a solution of level γ exists) is given by (11.3.35), where $\tilde{P}_i^{-1} = P_i^{-1} - \gamma_g^{-2} h_i h_i^T$ and P_i satisfies the recursion

$$P_{i+1} = P_i - P_i \begin{bmatrix} h_i & h_i \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -\gamma_g^2 \end{bmatrix} + \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} P_i \begin{bmatrix} h_i & h_i \end{bmatrix} \right\}^{-1} \begin{bmatrix} h_i^T \\ h_i^T \end{bmatrix} P_i, \quad P_0 = \mu I. \tag{11.3.38}$$

Two consecutive applications of the matrix inversion lemma yield the recursions (11.3.36).

Therefore all that remains to be shown is the condition for existence. But according to Theorem 3.2.2 this is given by

$$\tilde{P}_{i+1}^{-1} = P_{i+1}^{-1} - \gamma_g^{-2} h_{i+1} h_{i+1}^T > 0, \quad i = 0, 1, \ldots N. \tag{11.3.39}$$

Now this last condition is equivalent to

$$\begin{aligned} \gamma_g^2 &> h_{i+1}^T P_{i+1} h_{i+1} \\ &= h_{i+1}^T \left[qI + \left(P_i^{-1} + (1-\gamma_g^{-2}) h_i h_i^T \right)^{-1} \right] h_{i+1} \\ &= q h_{i+1}^T h_{i+1} + h_{i+1}^T \left(\tilde{P}_i^{-1} + h_i h_i^T \right)^{-1} h_{i+1}. \end{aligned}$$

A "sufficient" condition fot the above to happen is that

$$\gamma_g^2 > q h_{i+1}^T h_{i+1} + h_{i+1}^T h_{i+1} \bar{\sigma} \left[(\tilde{P}_i^{-1} + h_i h_i^T)^{-1} \right]. \tag{11.3.40}$$

But

$$h_i^T(\tilde{P}_i^{-1} + h_i h_i^T)h_i = h_i^T \tilde{P}_i^{-1} h_i + (h_i^T h_i)^2 \leq (h_i^T h_i)^2,$$

since $\tilde{P}_i^{-1} > 0$ (by the assumption of the existence of an H^∞ filter). But the above inequality implies

$$\underline{\sigma}(\tilde{P}_i^{-1} + h_i h_i^T) \geq h_i^T h_i, \tag{11.3.41}$$

and therefore

$$\bar{\sigma}\left[(\tilde{P}_i^{-1} + h_i h_i^T)^{-1}\right] \leq \frac{1}{h_i^T h_i}. \tag{11.3.42}$$

This implies that a sufficient condition for the existence of a filter is that

$$\gamma_g^2 > q h_{i+1}^T h_{i+1} + \frac{h_{i+1}^T h_{i+1}}{h_i^T h_i}, \tag{11.3.43}$$

from which the desired result (11.3.34) follows.

■

11.4 Simulation Results

For brevity we shall only describe one typical simulation result here. To this end, consider the model (11.3.30) where the weight vector x_i is now a scalar. To reflect time-variation we chose $\delta x_i = .02$, and to reflect modeling error,

$$v_i = .1 * (h_i x_i)^3 + n_i, \tag{11.4.1}$$

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 Tables 11.1 and 11.2. The H^∞ algorithms considered were LMS and the algorithms of Theorems 11.3.1 and 11.3.3, 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. 11.3.1	Thm. 11.3.3
$\sum_{j=0}^{100} \lvert e_j \rvert^2$	1.48	1.54	1.48

Table 11.1: The H^∞ algorithms.

	RLS	λ-RLS	KF
$\sum_{j=0}^{100} \lvert e_j \rvert^2$	6.00	2.08	5.11

Table 11.2: The H^2 algorithms.

11.5 Conclusion

We close this chapter, and our study of H^∞ adaptive filtering, with a brief remark on the general relevance of the H^∞ approach to estimation.

It is important to 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 modeling errors, and where the statistics and/or distributions of the disturbances are not known (or are too expensive to obtain).