Mixed Least-Mean-Squares/H^{∞}-Optimal Adaptive Filtering*

Babak Hassibi[†] Thomas Kailath

Information Systems Laboratory
Stanford University, Stanford CA 94305

Abstract

In this paper we construct a so-called mixed least-mean-squares/H^{∞}-optimal (or mixed H^2/H^{∞}-optimal) algorithm for adaptive filtering. The resulting adaptive algorithm is nonlinear and requires $O(n^2)$ (where n is the number of filter weights) operations per iteration. Such mixed algorithms have the property of yielding the best average (least-mean-squares) performance over all algorithms that achieve a certain worst-case (H^{∞}-optimal) bound. They thus allow a tradeoff between average and worst-case performances and are most applicable in situations where the exact statistics and distributions of the underlying signals are not known. Simple simulations are also presented to compare the algorithm's behaviour with standard least-squares and H^{∞} adaptive filters.

1 Introduction

Classical methods in estimation theory (such as least-mean-squares, maximum-likelihood, and maximum entropy) and the more recent robust methods in estimation theory (such as H^{∞}) can be regarded as two extremes in terms of their requirements regarding the statistical properties of the exogenous signals, as well as in terms of their goals. In classical estimation methods optimality of the *average* (or expected) performance of the estimators, under some assumptions regarding the statistical nature of the signals, is the key issue and hence their performance heavily depends upon the validity of these assumptions. On the other hand, robust estimation methods, or so-called *minimax* estimation strategies, safeguard against the *worst-case* disturbances and therefore make no assumptions on the (statistical) nature of the signals.

Among the classical methods, the most widespread is the least-mean-squares (or H^2) estimation technique which (under certain statistical assumptions on the signals) minimizes the expected estimation error energy. However, in many applications, due to model uncertainties and lack of statistical information H^2 methods are not directly applicable and the behavior of such estimation schemes is uncertain. Recently, following some pioneering work in robust control theory [1], H^{∞} estimation theory has been developed to address such problems.

Adaptive filtering is currently widely used to cope with time variation of system parameters and lack of a priori statistical knowledge of the underlying signals. The adaptive filtering algorithms currently used fall into the following two general catagories: (i) least-squares algorithms, such as the recursive-least-squares (RLS) algorithm, that are H^2-optimal and have the best average performance, and (ii) gradient-based algorithms, such as the least-mean-squares (LMS) algorithm, that are H^{∞}-optimal (see [2]) and have the best worst-case performance.

The mixed estimation problem was introduced as a compromise between these two extreme point of views [3, 4, 5]. The mixed H^2/H^{∞} problem allows one to trade off between the best average performance of the H^2 estimator and the best guaranteed worst-case performance of the H^{∞} estimator. As a result, the optimal mixed H^2/H^{∞} estimators achieve the best average performance, not over the set of all estimators, but over a restricted set of estimators that achieve a certain worst-case performance bound. Unlike the H^2 and H^{∞} problems, the question of finding the optimal mixed estimator has been an open problem. In this paper, for the first

*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. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of the Advanced Research Projects Agency, or the US Government.

†**Contact author**: Information Systems Laboratory, Stanford University, Stanford CA 94305. Phone (415) 723-1538 Fax (415) 723-8473 E-mail: hassibi@rascals.stanford.edu

time, we shall show how to construct the optimal mixed least-mean-squares/H^∞ estimator for adaptive filtering. The resulting algorithm is nonlinear and requires $O(n^2)$ computations per iteration, which is the same order of comlexity required of least-squares adaptive filters.

2 H^2 and H^∞ Adaptive Filtering

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 $h_i^T = \begin{bmatrix} h_{i1} & h_{i2} & \dots & h_{in} \end{bmatrix}$ 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 modelling errors. Let $\hat{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ denote the estimate of w given the observations $\{d_j\}$ and $\{h_j\}$ from time 0 up to and including time i.

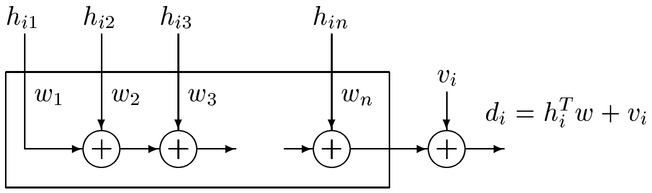

Figure 1: The model for adaptive filtering.

In this paper we will be interested in predicting the output of the filter, and therefore we define the output prediction error as

$$e_{p,i} \triangleq h_i^T w - h_i^T \hat{w}_{|i-1} = z_i - \hat{z}_i,$$

i.e., as the difference between the *uncorrupted* output $z_i \triangleq h_i^T w$ and $\hat{z}_i \triangleq h_i^T \hat{w}_{|i-1}$, the output predicted at time $i-1$. [We should remark that it is also possible to consider other forms of estimation error, such as filtered or smoothed errors, however, in this paper for brevity we shall focus only on prediction.]

2.1 The H^2 Approach

In the H^2 framework it is assumed that the unknown weight vector, w, and the additive disturbance, $\{v_i\}$, are random variables. In particular, it is assumed that they are zero-mean, uncorrelated (in the case of the $\{v_i\}$ temporally white) random variables with variances μI ($\mu > 0$) and unity, respectively. In this case we have the following problem.

Problem 1 (H^2 Adaptive Filtering) *Consider the linear model (1) and suppose that w and the $\{v_i\}$ are zero-mean, uncorrelated random variables with variances μI ($\mu > 0$) and unity, respectively. Find an H^2-optimal estimation strategy $\bar{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ that minimizes the expected prediction error energy*

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

for all i.

The solution is wellknown and is given by the RLS algorithm

$$\bar{w}_{|i} = \bar{w}_{|i-1} + \frac{P_i h_i}{1 + h_i^T P_i h_i}(d_i - h_i^T \bar{w}_{|i-1}), \quad \bar{w}_{|-1} = 0 \tag{3}$$

where P_i satisfies the (Riccati) recursion

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

2.2 The H^∞ Approach

Here we make no statistical assumptions on the unknown weight vector, w, and the additive disturbance, $\{v_i\}$. Note that any choice of estimation strategy $\mathcal{F}(\cdot)$ will induce a transfer operator from the disturbances $\{\mu^{-\frac{1}{2}} w, \{v_j\}_{j=0}^{i}\}$ to the output prediction errors $\{e_{p,j}\}_{j=0}^{i}$, that we shall denote by $T_{p,i}(\mathcal{F})$. See Figure 2.

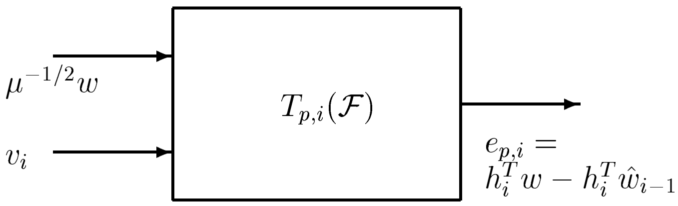

Figure 2: Transfer operator from disturbances to output prediction error.

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

Problem 2 (H^∞ Adaptive Filtering) *Consider the linear model (1). Find an H^∞-optimal estimation strategy $\hat{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ that minimizes* *the maximum energy gain of $T_{p,\infty}(\mathcal{F})$, and obtain the resulting*

$$\gamma_p^2 = \inf_{\mathcal{F}} \sup_{w,v \in h^2} \frac{\sum_{j=0}^{\infty} |e_{p,j}|^2}{\mu^{-1} w^T w + \sum_{j=0}^{\infty} |v_j|^2}, \quad (5)$$

where h^2 is the space of all causal square-summable sequences.

The above problem has been solved in [2] where it is shown that if the input vectors $\{h_i\}$ are such that $\lim_{i\to\infty} \sum_{j=0}^{i} h_i^T h_i = \infty$, and if μ is chosen (small enough) so that

$$\alpha_i \triangleq 1 - \mu h_i^T h_i > 0, \quad \text{for all } i, \quad (6)$$

then the min-max energy gain is

$$\gamma_p^2 = 1, \quad (7)$$

and one resulting H^∞-optimal filter is the LMS algorithm with learning rate μ, *i.e.*,

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \mu h_i (d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{|-1} = 0. \quad (8)$$

One interesting feature of the solution to Problem 2 is that the H^∞-optimal predictions of the uncorrupted output, which we have denoted by $\hat{z}_i$, are highly non-unique. In fact, in [2] it is shown that $\{\hat{z}_j\}$ is given by *any* sequence that satisfies the inequality,

$$\sum_{j=0}^{i-1} \alpha_j \xi_j^2 - \sum_{j=0}^{i} \frac{1}{\alpha_j} (\hat{z}_i - h_j^T \hat{w}_{|j-1})^2 \geq 0, \quad (9)$$

where we have defined

$$\xi_i \triangleq d_i - h_i^T \hat{w}_{|i-1} - \frac{\mu}{\alpha_i} (\hat{z}_i - h_i^T \hat{w}_{|i-1}), \quad (10)$$

and where now $\hat{w}_{|i}$ satisfies the recursion

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \mu h_i (d_i - \hat{z}_i), \quad \hat{w}_{|-1} = 0. \quad (11)$$

Note that in view of (6), $\alpha_j > 0$, so that the one obvious choice that guarantees (9) is $\hat{z}_j = h_j^T \hat{w}_{|j-1}$. But for this choice, (11) becomes simply the LMS algorithm.

3 Mixed Adaptive Filtering

Although H^∞-optimal estimators are highly robust with respect to disturbance variation, since they make no use of any, albeit incomplete, statistical information, they may be over conservative. The mixed least-mean-squares/H^∞-optimal approach is an attempt to alleviate this problem by exploiting the nonuniqueness of the H^∞ filters to improve some other aspect of the estimator besides robustness, namely its average performance. To be more specific, in mixed H^2/H^∞ estimation the goal is to come up with estimators that yield the smallest expected estimation error energy over all estimators that guarantee a certain worst-case (H^∞) bound. The problem may be formulated as follows.

Problem 3 (Mixed H^2/H^∞ Adaptive Filtering) *Consider the linear model (1) and suppose that the w and $\{v_j\}$ are independent zero-mean Gaussian random variables with variances μI and unity, respectively. Find an H^2/H^∞-optimal estimation strategy $\hat{z}_i = \mathcal{F}(h_0, h_1, \ldots, h_i; d_0, d_1, \ldots, d_{i-1})$ that minimizes the expected prediction error energy*

$$E \sum_{j=0}^{i} |e_{p,j}|^2 = E \sum_{j=0}^{i} |h_j^T w - \hat{z}_j|^2,$$

subject to the (optimal) H^∞ constraint

$$\sup_{w,v \in h^2} \frac{\sum_{j=0}^{i} |e_{p,j}|^2}{\mu^{-1} w^T w + \sum_{j=0}^{i} |v_j|^2} = 1,$$

for all i.

Remarks:

(i) The above problem formulation means that the resulting mixed adaptive filters will have the smallest mean-square estimation error over the set of all H^∞-optimal filters. They thus combine the average and worst-case performances of H^2 and H^∞ estimation, and in a sense yield the 'best of both worlds'.

(ii) Unlike Problem 1 where we only assumed knowledge of second-order statistics, here we have an additional Gaussian assumption. This is crucial, since the resulting solution is a nonlinear algorithm.

(iii) We have allowed $\hat{z}_i$ to be a function of h_i, since we are assuming that we know the input at time i and would like to predict the resulting output. [It is also possible to consider a problem where *all* of the input vectors (or regressor vectors) are known in advance, although the solution turns out to be considerably more complicated — see [6].]

Solution 1 (Solution to Problem 3) *The mixed least-mean-squares/H^∞-optimal predictions, $\hat{z}_i$, are found from the following optimization problem,*

$$\begin{cases} \min_{\hat{z}_i} \ (\hat{z}_i - h_i^T \bar{w}_{|i-1})^2 \\ \text{subject to } \ J_{i-1} - \frac{1}{\alpha_i} (\hat{z}_i - h_i^T \hat{w}_{|i-1})^2 \geq 0 \end{cases} \quad (12)$$

where $h_i^T \bar{w}_{|i-1}$ is the least-mean-squares prediction of the output, with $\bar{w}_{|i}$ satisfying the RLS algorithm (3), where $\hat{w}_{|i}$ satisfies the recursion (11), and where

$$J_i = J_{i-1} - \frac{1}{\alpha_i}(\hat{z}_i - h_i^T \hat{w}_{|i-1})^2 + \alpha_i \xi_i^2, \quad J_{-1} = 0 \quad (13)$$

with α_i and ξ_i defined via (6) and (10), respectively.

The above solution has an interesting structure and effectively combines the H^2 solution (3) and the H^∞ solution (11). In effect, the optimization problem (12) means that $\hat{z}_i$ tries to match the H^2 solution, $h_i^T \bar{w}_i$, as best as possible (in a least-squares sense) while satisfying a certain constraint.

The estimates $\{\hat{z}_j\}$ are, in general, nonlinear functions of the observations $\{d_j\}$, because of the nonlinear optimization step (12). (This nonlinearity is then propagated into the $\hat{w}_{|i}$ via (11).) The nonlinear optimization (12) is a convex quadratic program and can be readily solved using convex optimization techniques. In our application, however, we can actually solve it in closed-form. The result is given below.

Solution 2 (Mixed H^2/H^∞ Adaptive Filter) *Problem 3 has the following solution:*

(i) If

$$J_{i-1} - \frac{1}{\alpha_i}(h_i^T \bar{w}_{|i-1} - h_i^T \hat{w}_{|i-1})^2 \geq 0, \quad (14)$$

then

$$\hat{z}_i = h_i^T \bar{w}_{|i-1}. \quad (15)$$

(ii) Otherwise,

$$\hat{z}_i = \theta_i h_i^T \bar{w}_{|i-1} + (1 - \theta_i) h_i^T \hat{w}_{|i-1}, \quad (16)$$

where

$$\theta_i = \sqrt{\frac{\alpha_i J_{i-1}}{(h_i^T \bar{w}_{|i-1} - h_i^T \hat{w}_{|i-1})^2}} < 1 \quad (17)$$

with $\bar{w}_{|i-1}$, $\hat{w}_{|i-1}$, J_i and α_i as in Solution 1.

Remarks:

(i) The above solution shows, much more explicitly, the "mixed" nature of the H^2/H^∞ adaptive filter. Indeed, depending on the sign of the signal in (14) the desired estimate, $\hat{z}_i$, essentially *switches* between the H^2 estimate, $h_i^T \bar{w}_{|i-1}$, and the estimate of (16) which is a convex combination of the H^2 estimate and $h_i^T \hat{w}_{|i-1}$.

(ii) Despite being nonlinear, the major computational burden at each iteration of the algorithm is that of finding the least-mean-squares estimate, $\bar{w}_{|i}$. Thus the computational complexity is the same as the RLS algorithm, *i.e.*, $O(n^2)$ per iteration.

4 Example

To illustrate some properties of the mixed adaptive filter, and in order to compare its performance with standard H^2-optimal and H^∞-optimal algorithms, we shall now consider a very simple example. In this example we consider an adaptive filter with a single scalar weight and would like to use the past and current observations to predict the next output. In order to do so, we shall use the (H^2-optimal) RLS algorithm, the (H^∞-optimal) LMS algorithm and the mixed least-mean-squares/H^∞-optimal algorithm described above.

μ	0.1	0.2	0.5	0.8	0.9
LMS	1	1	1	1	1
RLS	1.39	1.73	2.15	2.37	2.43
Mixed	1	1	1	1	1

Table 1: *Maximum energy gains for the three filters for $N = 50$ (the number of observations) as a function of μ.*

Table 1 shows the maximum energy gain for each algorithm for $N = 50$ (the number of observations) as a function of μ. As can be seen, the mixed adaptive filter has optimal energy gain whereas RLS has a larger energy gain.

μ	0.1	0.2	0.5	0.8	0.9
LMS	2.88	5.80	16.9	33.5	41.0
RLS	1.83	2.49	3.52	4.15	4.33
Mixed	1.86	2.55	5.89	13.9	19.2

Table 2: *Expected prediction error energy for $N = 50$ (the number of observations) and white Gaussian unit variance disturbance as a function of μ.*

Table 2 shows the expected prediction error energy for $N = 50$ (the number of observations) and white Gaussian unit variance disturbance as a function of μ. As can be seen, the mixed adaptive filter shows significant average performance improvement over the LMS algorithm.

Figure 3 shows the prediction errors resulting from the worst-case RLS disturbance. As shown, the RLS prediction error is siginificantly larger than that of LMS and the mixed adaptive filter.

Figure 4 shows the prediction errors resulting from a white Gaussian disturbance with SNR = 10db ($\mu = 0.9$ and $N = 50$). As can be seen, the mixed adaptive filter shows significant improvement over the LMS algorithm.

Note, moreover, that RLS is the optimum adaptive filter for white Gaussian disturbances.

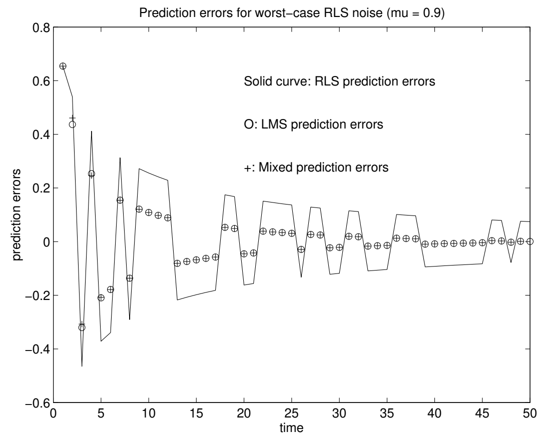

Figure 3: *Prediction errors for worst-case RLS disturbance ($\mu = 0.9$ and $N = 50$.)*

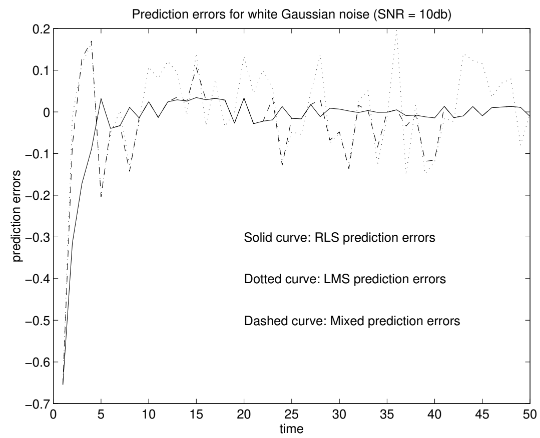

Figure 4: *Prediction errors for white Gaussian disturbance ($\mu = 0.9$ and $N = 50$ and SNR = 10db.)*

5 Conclusion

In this paper we have constructed a mixed least-mean-squares/H^∞-optimal algorithm for adaptive filtering that yields the best average performance over all adaptive filters satisfying an optimal worst-case bound. Finding such so-called mixed H^2/H^∞ optimal estimators had previously been an open problem. The adaptive algorithm developed here is nonlinear and requires $O(n^2)$ (where n is the number of filter weights) operations per iteration. It also allows one to study the tradeoff between average and worst-case performances and is most applicable in situations where (due to modeling errors and lack of a priori information) the exact statistics and distributions of the underlying signals are not known. We should also remark that it is possible to develop mixed least-mean-squares/H^∞-optimal estimators for a much more general class of problems, but for brevity we have confined ourselves here to adaptive filtering.

References

[1] G. Zames (1981). Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses, In *IEEE Trans. on Automatic Control*, AC-26, pp. 301-320.

[2] B. Hassibi, A. H. Sayed, and T. Kailath (1996). H^∞ optimality of the LMS algorithm. In *IEEE Trans. on Signal Processing*, SP-44, pp. 267-281.

[3] D. S. Bernstein and W. M. Haddad (1989). LQG control with an H^∞ performance bound: A Riccati equation approach. *IEEE Transactions on Automatic Control*, vol.34, no.3, pp. 293-305.

[4] P. P. Khargonekar and M. A. Rotea (1992). Mixed H^2/H^∞ filtering. In *Proc. IEEE Conference on Decision and Control*, Tucson, AZ, Dec. 1992, pp. 2299-304.

[5] K. Zhou, K. Glover, B. Bodenheimer and J. C. Doyle (1994). Mixed H^2 and H^∞ performance objectives I and II. *IEEE Transactions on Automatic Control*, vol.39, no.8, pp. 1564-1587.

[6] B. Hassibi and T. Kailath (1997). On nonlinear filters for mixed H^2/H^∞ estimation. Submitted to the 1997 American Control Conference.