Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

H^∞ Bounds for the Recursive-Least-Squares Algorithm*

Babak Hassibi† and Thomas Kailath

Information Systems Laboratory
Stanford University, Stanford CA 94305

Abstract

We obtain upper and lower bounds for the H^∞ norm of the RLS (Recursive-Least-Squares) algorithm. The H^∞ norm may be regarded as the worst-case energy gain from the disturbances to the *prediction* errors, and is therefore a measure of the robustness of an algorithm to perturbations and model uncertainty. Our results allow one to compare the robustness of RLS compared to the LMS (Least-Mean-Squares) algorithm, which is known to minimize the H^∞ norm. Simulations are presented to show the behaviour of RLS relative to these bounds.

1 Introduction

In the spirit of recent work in robust control there has been growing interest in deterministic worst-case identification. In such problems one is confronted with the task of designing identification algorithms that have robust performance in the presence of unknown but bounded noise. Likewise it is required to analyze the worst-case behaviour of identification algorithms with respect to such disturbances. For an introduction to recent approaches in H^∞ and l_1 identification the reader is referred to [1,2,3,4,5,6] and the references therein.

Suppose we observe an output sequence $\{d_i\}$ that obeys the following model:

$$d_i = h_i w + v_i, \qquad i \geq 0 \tag{1.1}$$

where $h_i = [\; h_{i1} \quad h_{i2} \quad \dots \quad h_{in} \;]$ is a known $1 \times n$ input vector, w is an unknown $n \times 1$ weight vector that we intend to estimate, and $\{v_i\}$ is an unknown disturbance, which may also include modeling errors. We shall make no assumptions on the statistics or distribution of the noise sequence $\{v_i\}$ (such as whiteness, normal distributed, etc.).

Let $w_i = \mathcal{F}(d_0, d_1, \dots, d_i)$ denote the estimate of w using observations (*i.e.* input-output pairs $\{d_i, h_i\}$) from time 0 up to and including time i. The prediction error, defined as the difference between the uncorrupted output and the predicted output, will be therefore given by $e_i = h_i w - h_i w_{i-1}$. Any choice of an estimator $\mathcal{F}(d_0, d_1, \dots, d_i)$ will induce a transfer operator $T(\mathcal{F})$ that maps the the disturbances $\{w, v_i\}$ to the prediction errors $\{e_i\}$. (See Figure 1.) A robust estimator, $\mathcal{F}$, will be one for which if the disturbances are small (in some sense) then the prediction errors will be small. Likewise an estimator will not be robust if there exist small disturbances for which one may have large prediction errors. If one's measure of the size of the disturbances and prediction errors is energy, then the robustness of $\mathcal{F}$ will be measured by the H^∞ norm of $T(\mathcal{F})$, which is denoted by $\| T(\mathcal{F}) \|_\infty$ and is defined as

†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.

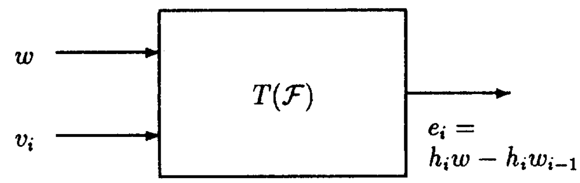

Figure 1.1: Transfer operator from the unknown disturbances $\{w - w_{-1}, v_i\}$ to the prediction errors $\{e_i\}$.

$$\|T(\mathcal{F})\|_\infty^2 = \sup_{w \neq 0, v \neq 0 \in h^2} \frac{\|e\|_2^2}{\mu^{-1}|w|^2 + \|v\|_2^2} \tag{1.2}$$

where h^2 is the space of all causal square-summable sequences, and $\|e\|_2^2$ and $\mu^{-1}|w|^2 + \|v\|_2^2$ are the prediction error and disturbance energies, respectively.

The H^∞ norm may thus be regarded as the maximum energy gain from input to ouput. Note that estimators with small H^∞ norm guarantee small prediction error energy over *all* possible disturbances of small energy. They are thus over conservative, which reflects in a more robust behaviour to disturbance variation.

As stated below, the celebrated LMS algorithm minimizes $\| T(\mathcal{F}) \|_\infty$ [7].

Theorem 1.1 (LMS Algorithm) *If the input vectors h_i are exciting and $0 < \mu < \inf_i \frac{1}{h_i h_i^*}$, then the "minimum" value of $\|T(\mathcal{F})\|_\infty$ in (1.2) is $\gamma_{opt} = 1$. In this case an optimal H^∞ estimator is given by the LMS algorithm with learning rate μ, viz.*

$$w_i = w_{i-1} + \mu h_i^*(d_i - h_i w_{i-1}), \quad w_{-1} = 0. \tag{1.3}$$

Another adaptive algorithm that is widely used, is the celebrated Recursive-Least-Squares (RLS) algorithm, given by

$$w_i = w_{i-1} + \frac{P_i h_i^*}{1 + h_i P_i h_i^*}, \quad w_{-1} = 0 \tag{1.4}$$

where P_i satisifies the Riccati recursion

$$P_{i+1} = P_i - \frac{P_i h_i^* h_i P_i}{1 + h_i P_i h_i^*}, \quad P_0 = \mu I. \tag{1.5}$$

The RLS algorithm is an *exact* least-squares solution that satisfies an H^2 criterion, and that enjoys certain well-known optimality properties under suitable stochastic assumptions about the exogenous noise. A natural question to ask is what is the performance of RLS if the above assumptions are violated? In other words how robust is the RLS algorithm to model uncertainties and lack of statistical information?

In order to answer this question, in this note we shall obtain upper and lower bounds on the H^∞ norm of the RLS algorithm. These bounds are of interest for several reasons. First they demonstrate that unlike the LMS algorithm whose H^∞ norm is unity (independent of the input-output data), the H^∞ norm of the RLS algorithm depends on the input-output data, and therefore RLS may be more robust or less robust with respect to different data sets. Moreover, the exact calculation of the H^∞ norm for RLS requires the calculation of the induced 2-norm of a linear *time-varying* operator, which in addition to being quite cumbersome, needs all the input-output data, which may not be available in real-time scenarios. The H^∞ bounds we obtain require only simple a priori knowledge of the data, and may therefore be used as a simple check to verify whether RLS has the desired robustness with respect to a given application.

2 H^∞ Bounds

The first set of bounds given are valid for the Kalman filter as well as for the RLS algorithm. However, for simplicity, we shall consider the RLS case only. The proofs are omitted due to lack of space.

Theorem 2.1 (Upper and Lower Bounds) *Denote the H^∞ norm of the RLS algorithm by $\| T(\mathcal{F}_{RLS}) \|_\infty$. Then*

$$\sqrt{r} - 1 \leq \| T(\mathcal{F}_{RLS}) \|_\infty \leq \sqrt{r} + 1.$$

where $r \triangleq \sup_i (1 + h_i P_i h_i^*)$.

In the RLS algorithm the P_i are given by $P_i = (\mu^{-1} I + \sum_{j=0}^{i-1} h_j^* h_j)^{-1}$. Therefore the P_i are a monotonically decreasing sequence of matrices. If we assume that the input vectors h_i have equal magnitude then we have the following result.

Corollary 2.1 (Constant Magnitude Inputs) *If the input vectors have constant magnitude $h_i h_i^* = h^2$, then $r = 1 + \mu h^2$, so that*

$$\sqrt{1 + \mu h^2} - 1 \leq \| T(\mathcal{F}_{RLS}) \|_\infty \leq \sqrt{1 + \mu h^2} + 1.$$

Corollary 2.2 *If $\bar{h}^2 = \sup_i h_i h_i^*$ and $\underline{h}^2 = \inf_i h_i h_i^*$, then*

$$\sqrt{1 + \mu \underline{h}^2} - 1 \leq \| T(\mathcal{F}_{RLS}) \|_\infty \leq \sqrt{1 + \mu \bar{h}^2} + 1.$$

It is possible to obtain tighter lower bounds for the special case of RLS by calculating the ratio $\frac{\|e\|_2^2}{\mu^{-1}|w|^2 + \|v\|_2^2}$ for a particular choice of disturbance w and $\{v_i\}$. The choice of disturbance $w = [\, 1 \ \ldots \ 1 \,]$ and $v_i = \frac{-1}{(1+i\mu\underline{h}^2)(1+(i+1)\mu\underline{h}^2)}$ yields the following result.

Theorem 2.2 (Lower Bound) *A lower bound for $\| T(\mathcal{F}_{RLS}) \|_\infty^2$ is*

$$\underline{h}^2 \frac{(\frac{1}{\underline{h}^2} + \sqrt{\mu})^2 S_2(\mu \underline{h}) - \frac{2}{\underline{h}^2}(\frac{1}{\underline{h}^2} + \sqrt{\mu}) S_3(\mu \underline{h}) + \frac{1}{\underline{h}^4} S_4(\mu \underline{h})}{1 - \frac{1}{\mu^2 \underline{h}^4}(1 + \frac{1}{\mu \underline{h}^2}) + \frac{2}{\mu^2 \underline{h}^4} S_2(\mu \underline{h})}$$

where $S_i(x) = \sum_{j=0}^{\infty} \frac{1}{(1+jx)^i}, \qquad i = 2, 3, 4.$

Corollaries 2.1 and 2.2 have an interesting interpretation: the RLS algorithm is less robust for large values of μ. Indeed we see that the H^∞ norm grows as $\sqrt{\mu}$. This is reminiscent of the robustness properties of LMS, where the learning rate μ had to be small enough to guarantee H^∞ optimality.

3 Example

In this section we shall consider a simple example where the h_i are scalars that randomly take on the values +1 and −1. Thus in this example $\bar{h}^2 = \underline{h}^2 = 1$. We have given the the H^∞ norm of RLS for $N = 50$ data points and for several different values of μ. The upper bound of Theorem 2.1 and the lower bound of Theorem 2.2 are also given.

μ	True value	Lower bound	Upper
0.9	1.62	1.24	2.38
2	1.75	1.55	2.73
5	2.29	2.28	3.45
10	3.22	3.22	4.31

Table 3.1: H^∞ norm of RLS for $N = 50$ data points as a function of μ. As can be seen, in this example, the lower bounds of Theorem 2.2 seem quite accurate for large μ.

References

[1] A.J. Helmicki, C.A. Jacobson and C.N. Nett. Control-Oriented System Identification: A Worst-Case/Deterministic Approach in H^∞. *IEEE Trans. on Automatic Control*, AC-36:1163–1176, 1991.

[2] B. Wahlberg and L. Ljung. Hard Frequency-Domain Model Error Bounds From Least-Squares Like Identification Techniques. *IEEE Trans. on Automatic Control*, AC-37:900-912, 1992.

[3] M.A. Dahleh, T.V. Theodosopoulos and J.N. Tsitsiklis. The Sample Complexity of Worst Case Identification of F.I.R. Linear Systems. *Systems and Control Letters*, March 1993.

[4] M. Milanese. Properties of Least-Squares Estimates in Set Membership Identification. Preprint, Politecnico di Torino, Italy, 1993.

[5] J.R. Partington and P.M. Makila. Analysis of Linear Methods for Robust Identification in l_1. Preprint, 1993.

[6] H. Akcay and P.P. Khargonekar. On the Worst-Case Performance of the Least-Squares Algorithm. *Proc. IEEE Conf. on Dec. and Contr.*, San Antonio, TX, pp. 2076-2081, 1993.

[7] B. Hassibi, A.H. Sayed and T. Kailath. LMS is H^∞ Optimal. *Proc. IEEE Conf. on Dec. and Contr.*, San Antonio, TX, pp. 74-80, 1993.

æ