Chapter 10

Robustness of Least-Squares Estimators

In the previous chapter we demonstrated that instantaneous-gradient-based adaptive filters, such as the LMS and normalized LMS algorithms, minimize the maximum energy gain from the disturbances to the (prediction and filtered) estimation errors. A natural question to ask is how do other adaptive algorithms compare to these (H^∞) "optimal" solutions? Needless to say, in the H^∞ approach to robust adaptive filtering, given any algorithm, this comparison is achieved by computing the corresponding maximum energy gain (or H^∞ norm) from the disturbances to the estimation errors.

In this chapter we shall study the robustness of the other important class of adaptive filters, namely, least-squares-based methods, such as the RLS algorithm, from the above point of view. We shall essentially obtain upper and lower bounds for the H^∞ norm of the RLS algorithm (in fact, more generally, of the Kalman filter) with respect to prediction and filtered errors. The main conclusion is that, unlike LMS and normalized LMS which do not allow for any amplification of the disturbances, the RLS algorithm does allow for such amplification. This fact can be especially pronounced in the prediction error case. Moreover, it is shown that the H^∞ norm for RLS is data-dependent, whereas for LMS and normalized LMS it was not so. [The H^∞ norm was simply unity.] The significance of the results will also discussed.

10.1 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 various identification algorithms with respect to such disturbances. For an introduction to recent approaches in H^∞ and l_1 identification the reader is referred to [AHN91, WL92, DTT93, AK93, Mil95, PM95] and the references therein.

In the previous chapter we showed that the LMS (normalized LMS) adaptive algorithm is H^∞-optimal in the sense that it minimizes the worst-case energy gain from the disturbances to the prediction (filtered) errors. This result confirmed the robustness of LMS (normalized LMS) with respect to model uncertainties and gave it a rigorous basis which was lacking. The celebrated recursive-least-squares (RLS) algorithm [Hay96], is an also widely used adaptive algorithm that enjoys certain well known optimality properties under suitable stochastic assumptions about the exogenous noise. In Chapter 9 we also compared the robustness of the RLS algorithm to LMS, and constructed a disturbance sequence of small energy for which RLS yielded a prediction error of large energy.

In the present chapter 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 it 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 two-norm of a linear time-varying operator, which can be quite cumbersome, and, in addition, needs all the input-output data, which may not be available in real-time scenarios. The H^∞ bounds we obtain only require 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.

A brief outline of the chapter is as follows. In Sec. 10.2 we give general upper and lower bounds for the H^∞ norm of the Kalman filter, with respect to prediction and filtered errors. The proofs of the upper bounds are given in Sec. 10.3 and are based on certain minimization properties of least-squares estimators. The proofs of the lower bounds are given in 10.4 and are essentially based on computing the energy gains for suitably chosen disturbances. Sec. 10.5 specializes the general results of Sec. 10.2 to the adaptive filtering problem, discusses its various implications and provides a simple example to illustrate the results. The chapter concludes with Sec. 10.6.

10.2 A General H^∞ Bound

In this section we shall derive general upper and lower bounds for the H^∞ norm of the Kalman filter with respect to prediction and filtered errors. These results will then be specialized to the adaptive filtering problem to obtain corresponding H^∞ norm bounds for the RLS algorithm.

To this end, consider, once more, the standard state-space model

$$\begin{cases} x_{i+1} = F_i x_i + G_i u_i, & x_0 \\ y_i = H_i x_i + v_i, & i \geq 0 \end{cases} \tag{10.2.1}$$

where $F_i \in \mathcal{C}^{n\times n}$, $G_i \in \mathcal{C}^{n\times m}$ and $H_i \in \mathcal{C}^{p\times n}$ are known matrices, x_0, $\{u_i\}$, and $\{v_i\}$ are unknown quantities and y_i is the measured output. Moreover, v_i can be regarded as measurement noise and u_i as process noise or driving disturbance.

Recall from Chapter 2 that the Kalman filter for computing the predicted estimates of the states, denoted by $\hat{x}_i$, (*i.e.*, $\hat{x}_i$ is the least-squares estimate of x_i, given $\{y_j, j < i\}$) is given by

$$\hat{x}_{i+1} = F_i\hat{x}_i + K_{p,i}(y_i - H_i\hat{x}_i), \tag{10.2.2}$$

where

$$K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1} \quad \text{and} \quad R_{e,i} = R_i + H_i P_i H_i^* \tag{10.2.3}$$

and where P_i satisfies the Riccati recursion,

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0. \tag{10.2.4}$$

[Note here that $\{Q_i, R_i\}$ and Π_0 are given positive definite weighting matrices.]

There is also a filtered form of the Kalman filter recursions for computing, $\hat{x}_{i|i}$, the least-squares estimate of x_i, given $\{y_j, j \leq i\}$, which is given below,

$$\hat{x}_{i+1|i+1} = F_i\hat{x}_{i|i} + K_{f,i+1}(y_{i+1} - H_{i+1}F_i\hat{x}_{i|i}), \tag{10.2.5}$$

where

$$K_{f,i} = P_iH_i^*R_{e,i}^{-1}. \tag{10.2.6}$$

Now using $\hat{x}_i$ and $\hat{x}_{i|i}$ the predicted and filtered estimation errors of the uncorrupted output, $s_i = H_ix_i$, are given by

$$e_{p,i} = H_ix_i - H_i\hat{x}_i \triangleq H_i\tilde{x}_i, \tag{10.2.7}$$

and

$$e_{f,i} = H_ix_i - H_i\hat{x}_{i|i} \triangleq H_i\tilde{x}_{i|i}. \tag{10.2.8}$$

Note that both these estimation errors are different from the innovations (the prediction errors for estimating y_i),

$$e_i = y_i - H_i\hat{x}_i. \tag{10.2.9}$$

Indeed it is straightforward to see that we have

$$e_{p,i} = e_i - v_i, \tag{10.2.10}$$

and

$$e_{f,i} = R_iR_{e,i}^{-1}e_i - v_i. \tag{10.2.11}$$

The latter equality is justified below.

$$\begin{aligned}
e_{f,i} &= H_ix_i - H_i\hat{x}_{i|i} \\
&= y_i - v_i - H_i(\hat{x}_i + P_iH_i^*R_{e,i}^{-1}e_i) \\
&= e_i - H_iP_iH_i^*R_{e,i}^{-1}e_i - v_i \\
&= (R_{e,i} - H_iP_iH_i^*)R_{e,i}^{-1}e_i - v_i = R_iR_{e,i}^{-1}e_i - v_i.
\end{aligned}$$

We shall have the occasion to make use of both identities (10.2.10) and (10.2.11).

We can now state the main result of this chapter.

Theorem 10.2.1 (Bounds for the H^∞ Norm of the Kalman Filter) *Consider the standard state-space model (10.2.1) and the predicted and filtered forms of the Kalman filter recursions, (10.2.2) and (10.2.5). Then for any N, we have the following results,*

$$(\sqrt{\bar{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{N}(H_ix_i - H_i\hat{x}_i)^* R_i^{-1}(H_ix_i - H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i} \leq (\sqrt{\bar{r}}+1)^2 \quad (10.2.12)$$

and

$$(\sqrt{1/\underline{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{N}(H_ix_i - H_i\hat{x}_{i|i})^* R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i} \leq (\sqrt{1/\underline{r}}+1)^2 \quad (10.2.13)$$

where we have defined,

$$\bar{r} = \sup_i \ \bar{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right) \quad (10.2.14)$$

and

$$\underline{r} = \inf_i \ \underline{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right) \quad (10.2.15)$$

and $\bar{\sigma}(A)$ and $\underline{\sigma}(A)$ denote the maximum and minimum singular values of the matrix A, respectively.

Remarks:

(i) Note that the quantities

$$\sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{N}(H_ix_i - H_i\hat{x}_i)^* R_i^{-1}(H_ix_i - H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i}$$

and

$$\sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{N}(H_ix_i - H_i\hat{x}_{i|i})^* R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i}$$

in the inequalities (10.2.12) and (10.2.13) are simply the maximum energy gains from the (normalized) disturbances $\{\Pi_0^{-1/2}x_0, \{Q_i^{-1/2}u_i, R_i^{-1/2}v_i\}_{i=0}^{N}\}$ to the (normalized) prediction and filtered estimation errors $\{R_i^{-1/2}e_{p,i}\}_{i=0}^{N}$ and $\{R_i^{-1/2}e_{f,i}\}_{i=0}^{N}$, respectively. Thus (10.2.12) and (10.2.13) yield upper and lower bounds on the H^∞ norm of the Kalman filter for prediction and filtered errors, respectively.

(ii) Note, moreover, that the upper and lower bounds on the H^∞ norms, as given by Theorem 10.2.1 are relatively tight (especially for large values of $\bar{r}$ and $1/\underline{r}$). Indeed the upper and lower bounds differ only by two, since,

$$(\sqrt{\bar{r}}+1)-(\sqrt{\bar{r}}-1)=2 \quad \text{and} \quad (\sqrt{1/\underline{r}}+1)-(\sqrt{1/\underline{r}}-1)=2.$$

(iii) Note that Theorem 10.2.1 bounds the H^∞ norm of the Kalman filter by quantities related to the maximum and minimum singular values of the normalized innovations variance, $R_i^{-1/2}R_{e,i}R_i^{-*/2}$. Intuitively, (10.2.12) suggests that the larger $R_{e,i}$ is, the larger $\bar{r}$ is, and hence the less robust the Kalman filter is with respect to prediction errors. On the other hand, (10.2.13) suggests that the smaller $R_{e,i}$ is, the larger $1/\underline{r}$ is, and hence the less robust the Kalman filter is with respect to filtered errors.

In particular, note that

$$R_i^{-1/2}R_{e,i}R_i^{-*/2}=I_p+R_i^{-1/2}H_iP_iH_i^*R_i^{-*/2}\geq I_p, \tag{10.2.16}$$

so that

$$\bar{r}\geq 1\geq 1/\underline{r}. \tag{10.2.17}$$

But using (10.2.13) this means that

$$\sup_{x_0,u,v\in h^2}\frac{\sum_{i=0}^{N}(H_ix_i-H_i\hat{x}_{i|i})^*R_i^{-1}(H_ix_i-H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0+\sum_{i=0}^{N}u_i^*Q_i^{-1}u_i+\sum_{i=0}^{N}v_i^*R_i^{-1}v_i}\leq 4, \tag{10.2.18}$$

which is a very explicit, and quite surprising, bound. Thus the Kalman filter guarantees that the energy gain from the disturbances to the filtered errors never exceeds four.

(iv) The bounds of Theorem 10.2.1 are true for any value of N, and, in fact, they are also true when the upper limits of the sums in (10.2.12) and (10.2.13) are infinite. In other words, it is true that,

$$(\sqrt{\bar{r}}-1)^2\leq\sup_{x_0,u,v\in h^2}\frac{\sum_{i=0}^{\infty}(H_ix_i-H_i\hat{x}_i)^*R_i^{-1}(H_ix_i-H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0+\sum_{i=0}^{\infty}u_i^*Q_i^{-1}u_i+\sum_{i=0}^{\infty}v_i^*R_i^{-1}v_i}\leq(\sqrt{\bar{r}}+1)^2 \tag{10.2.19}$$

and

$$(\sqrt{1/\underline{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{\infty}(H_ix_i - H_i\hat{x}_{i|i})^* R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{\infty} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{\infty} v_i^*R_i^{-1}v_i} \leq (\sqrt{1/\underline{r}}+1)^2 \tag{10.2.20}$$

where, as before

$$\bar{r} = \sup_i \ \bar{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right) \quad \text{and} \quad \underline{r} = \inf_i \ \underline{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right).$$

In particular, in the time-invariant (doubly) infinite-horizon case we have

$$(\sqrt{\bar{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=-\infty}^{\infty}(H_ix_i - H_i\hat{x}_i)^* R^{-1}(H_ix_i - H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=-\infty}^{\infty} u_i^*Q_i^{-1}u_i + \sum_{i=-\infty}^{\infty} v_i^*R_i^{-1}v_i} \leq (\sqrt{\bar{r}}+1)^2 \tag{10.2.21}$$

and

$$(\sqrt{1/\underline{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=-\infty}^{\infty}(Hx_i - H\hat{x}_{i|i})^* R^{-1}(Hx_i - H\hat{x}_{i|i})}{\sum_{i=-\infty}^{\infty} u_i^*Q^{-1}u_i + \sum_{i=-\infty}^{\infty} v_i^*R^{-1}v_i} \leq (\sqrt{1/\underline{r}}+1)^2 \tag{10.2.22}$$

where now

$$\bar{r} = \bar{\sigma}\left(R{-}1/2R_eR^{-*/2}\right) \quad \text{and} \quad \underline{r} = \underline{\sigma}\left(R^{-1/2}R_eR^{-*/2}\right), \tag{10.2.23}$$

and $R_e = R + HPH^*$, with P the unique positive semidefinite and stabilizing solution of the DARE

$$P = FPF^* + GQG^* - FPH^*(R + HPH^*)^{-1}HPF^*.$$

We will now proceed with the proof of Theorem 10.2.1.

10.3 Proof of the Upper Bounds

To prove the upper bounds of Theorem 10.2.1 we need the following three facts.

Lemma 10.3.1 (Minimization of a Quadratic Form) *We have*

$$\min_{x_0,\{u_i,v_i\}} \ x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i = \sum_{i=0}^{N} e_i^*R_{e,i}^{-1}e_i, \tag{10.3.1}$$

where the minimization is taken subject to the state-space constraints (10.2.1), and where $e_i = y_i - H_i\hat{x}_i$ *is the innovations.*

Proof: The lemma is simply a special case of Theorem 2.7.4.

■

Lemma 10.3.2 (Simple Inequality) *For any vectors a, b, and any matrix $M > 0$, we have*

$$(a+b)^*M(a+b) \geq (1-\frac{1}{\alpha})a^*Ma + (1-\alpha)b^*Mb, \qquad \forall \alpha > 0. \tag{10.3.2}$$

Proof: Follows from

$$\begin{aligned}(a+b)^*M(a+b) - (1-\frac{1}{\alpha})a^*Ma + (1-\alpha)b^*Mb &= \frac{1}{\alpha}a^*Ma + a^*Mb + b^*Ma + \alpha b^*Mb \\ &= (\frac{1}{\sqrt{\alpha}}a + \sqrt{\alpha}b)^*M(\frac{1}{\sqrt{\alpha}}a + \sqrt{\alpha}b) \\ &\geq 0.\end{aligned}$$

■

Lemma 10.3.3 (A Simple Minimization) *For all $\beta > 0$, we have*

$$\min_{\alpha>1} \frac{\alpha^2 + (\beta-1)\alpha}{\alpha-1} = (1+\sqrt{\beta})^2, \tag{10.3.3}$$

and

$$\arg\min_{\alpha>1} \frac{\alpha^2 + (\beta-1)\alpha}{\alpha-1} = 1+\sqrt{\beta}. \tag{10.3.4}$$

Proof: Readily verified via differentiation.

■

We shall first prove the upper bound in (10.2.12) for the prediction error case. To this end, define

$$J = x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i. \tag{10.3.5}$$

Now using Lemma 10.3.1 it is obvious that

$$J \geq \min_{x_0,\{u_i,v_i\}} J = \sum_{i=0}^{N} e_i^*R_{e,i}^{-1}e_i. \tag{10.3.6}$$

Thus we may write,

$$\begin{aligned}
J &\geq \sum_{i=0}^{N} e_i^* R_{e,i}^{-1} e_i \\
&= \sum_{i=0}^{N} e_i^* R_i^{-*/2} R_i^{*/2} R_{e,i}^{-1} R_i^{1/2} R_i^{-1/2} e_i \\
&\geq \sum_{i=0}^{N} \underline{\sigma}\left(R_i^{*/2} R_{e,i}^{-1} R_i^{1/2}\right) e_i^* R_i^{-1} e_i \\
&\geq \left[\inf_i \ \underline{\sigma}^2\left(R_i^{*/2} R_{e,i}^{-1} R_i^{1/2}\right)\right] \sum_{i=0}^{N} e_i^* R_i^{-1} e_i \\
&= \left[\sup_i \ \frac{1}{\bar{\sigma}^2\left(R_i^{-1/2} R_{e,i} R_i^{-*/2}\right)}\right] \sum_{i=0}^{N} e_i^* R_i^{-1} e_i \\
&= \frac{1}{\bar{r}} \sum_{i=0}^{N} e_i^* R_i^{-1} e_i \\
&= \frac{1}{\bar{r}} \sum_{i=0}^{N} (e_{p,i} + v_i)^* R_i^{-1} (e_{p,i} + v_i) \quad \text{using (10.2.10)}
\end{aligned}$$

Now using Lemma (10.3.2) with $a = e_{p,i}$, $b = v_i$ and $M = R_i^{-1}$, we have

$$x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i \geq \frac{1}{\bar{r}} \sum_{i=0}^{N} \left[(1-\alpha) v_i^* R_i^{-1} v_i + (1 - \frac{1}{\alpha}) e_{p,i}^* R_i^{-1} e_{p,i}\right], \tag{10.3.7}$$

for any $\alpha > 0$. Now rearranging terms, we can write

$$(1-\frac{1}{\alpha}) \sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i} \leq \bar{r} x_0^* \Pi_0^{-1} x_0 + \bar{r} \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \bar{r}(1 - \frac{1-\alpha}{\bar{r}}) \sum_{i=0}^{N} v_i^* R_i^{-1} v_i, \tag{10.3.8}$$

so that, assuming $\alpha > 1$, we have

$$\begin{aligned}
\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i} &\leq \frac{\bar{r}}{1-\frac{1}{\alpha}} \left[x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i\right] + \frac{\bar{r}}{1-\frac{1}{\alpha}} (1 - \frac{1-\alpha}{\bar{r}}) \sum_{i=0}^{N} v_i^* R_i^{-1} v_i \\
&= \frac{\bar{r}}{1-\frac{1}{\alpha}} \left[x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i\right] + \frac{\alpha^2 + \alpha(\bar{r}-1)}{\alpha - 1} \sum_{i=0}^{N} v_i^* R_i^{-1} v_i.
\end{aligned}$$

To obtain the "tightest" possible bound on $\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i}$, let us minimize, over $\alpha > 1$, the coefficient of $\sum_{i=0}^{N} v_i^* R_i^{-1} v_i$ on the RHS of the above inequality. But from

Lemma 10.3.3, we have

$$\min_{\alpha>1} \frac{\alpha^2+(\bar{r}-1)\alpha}{\alpha-1} = (1+\sqrt{\bar{r}})^2 \quad \text{and} \quad \arg\min_{\alpha>1} \frac{\alpha^2+(\bar{r}-1)\alpha}{\alpha-1} = 1+\sqrt{\bar{r}}. \quad (10.3.9)$$

Therefore

$$\begin{aligned}
\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i} &\le \frac{\bar{r}}{1-\frac{1}{1+\sqrt{\bar{r}}}}\left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i\right] + (1+\sqrt{\bar{r}})^2 \sum_{i=0}^{N} v_i^* R_i^{-1} v_i \\
&= \sqrt{\bar{r}}(1+\sqrt{\bar{r}})\left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i\right] + (1+\sqrt{\bar{r}})^2 \sum_{i=0}^{N} v_i^* R_i^{-1} v_i \\
&\le (1+\sqrt{\bar{r}})^2\left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i\right] + (1+\sqrt{\bar{r}})^2 \sum_{i=0}^{N} v_i^* R_i^{-1} v_i .
\end{aligned}$$

Therefore, we have

$$\frac{\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i}}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \le (1+\sqrt{\bar{r}})^2, \quad (10.3.10)$$

which is the desired result.

To prove the upper bound of (10.2.13) for the filtered estimation errors we need to proceed as follows.

$$\begin{aligned}
J &\ge \sum_{i=0}^{N} e_i^* R_{e,i}^{-1} e_i \\
&= \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1} R_{e,i} R_{e,i}^{-1} R_{e,i} R_i^{-1} (e_{f,i}+v_i) \quad \text{using (10.2.11)} \\
&= \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-*/2} R_i^{-1/2} R_{e,i} R_i^{-*/2} R_i^{-1/2} (e_{f,i}+v_i) \\
&\ge \sum_{i=0}^{N} \underline{\sigma}\left(R_i^{-1/2} R_{e,i} R_i^{-*/2}\right)(e_{f,i}+v_i)^* R_i^{-1}(e_{f,i}+v_i) \\
&\ge \left[\inf_i \, \underline{\sigma}^2\left(R_i^{-1/2} R_{e,i} R_i^{-*/2}\right)\right] \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1}(e_{f,i}+v_i) \\
&= \underline{r} \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1}(e_{f,i}+v_i).
\end{aligned}$$

Proceeding now with an argument similar to what was done in the predicted case, leads ro the desired result,

$$\frac{\sum_{i=0}^{N} e_{f,i}^* R_i^{-1} e_{f,i}}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \le (1+\sqrt{1/\underline{r}})^2. \quad (10.3.11)$$

■

10.4 Proof of the Lower Bounds

Perhaps the most general way of computing the a lower bound for the H^∞ norm of RLS, or any other algorithm for that matter, is to compute the energy gain for some particular choice of disturbance signal, $\{x_0, \{u_i, v_i\}_{i=0}^{N}\}$.[1] We shall presently see that the special choice of disturbance signal that yields the lower bound of Theorem 10.2.1 is that disturbance signal that minimizes the quadratic form

$$J = x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i, \tag{10.4.1}$$

subject to the state-space constraints (10.2.1). To facilitate the presentation of the proof it will be convenient to make the following definitions.

$$y \triangleq \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad u \triangleq \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{bmatrix}, \quad v \triangleq \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_N \end{bmatrix} \tag{10.4.2}$$

and

$$e \triangleq \begin{bmatrix} y_0 - H_0\hat{x}_0 \\ y_1 - H_1\hat{x}_1 \\ \vdots \\ y_N - H_N\hat{x}_N \end{bmatrix}, \quad e_p \triangleq \begin{bmatrix} H_0x_0 - H_0\hat{x}_0 \\ H_1x_1 - H_1\hat{x}_1 \\ \vdots \\ H_Nx_N - H_N\hat{x}_N \end{bmatrix}, \quad e_f \triangleq \begin{bmatrix} H_0x_0 - H_0\hat{x}_{0|0} \\ H_1x_1 - H_1\hat{x}_{1|1} \\ \vdots \\ H_Nx_N - H_N\hat{x}_{N|N} \end{bmatrix}. \tag{10.4.3}$$

Moreover, from Chapter 2 we know that the innovations, e, can be found via

$$e = L^{-1} y, \tag{10.4.4}$$

where L is a lower triangular matrix with unit diagonal that is given from the unique (block) LDU decomposition,

$$R_y = L R_e L^*, \tag{10.4.5}$$

[1] The resulting energy gain will be less than or equal to the maximum energy gain and hence a lower bound to the (squared) H^∞ norm.

where

$$R_e = \mathrm{diag}(R_{e,0}, R_{e,1}, \ldots, R_{e,N}), \tag{10.4.6}$$

and R_y, the output Gramian, is given by

$$R_y = \mathcal{O}\Pi_0\mathcal{O}^* + \Gamma Q \Gamma^* + R, \tag{10.4.7}$$

where

$$\mathcal{O} = \begin{bmatrix} H_0 \\ H_1 F_0 \\ \vdots \\ H_N F_{N_1} \ldots F_0 \end{bmatrix}, \tag{10.4.8}$$

is the observability map, and

$$\Gamma = \begin{bmatrix} 0 & & & & \\ H_1 G_0 & 0 & & & \\ H_2 F_1 G_0 & H_2 G_1 & 0 & & \\ H_3 F_2 F_1 G_0 & H_3 F_2 G_1 & H_3 G_2 & . & \\ . & . & . & . & . \end{bmatrix}, \tag{10.4.9}$$

is the impulse response matrix, and

$$Q \triangleq \mathrm{diag}(Q_0, Q_1, \ldots, Q_N) \quad \text{and} \quad R \triangleq \mathrm{diag}(R_0, R_1, \ldots, R_N). \tag{10.4.10}$$

Finally, we need to note the global relation,

$$y = \mathcal{O}x_0 + \Gamma u + v, \tag{10.4.11}$$

and, using (10.2.10) and (10.2.11), the relations

$$e_p = e - v = L^{-1}y - v, \tag{10.4.12}$$

and

$$e_f = RR_e^{-1}e - v = RR_e^{-1}L^{-1}y - v. \tag{10.4.13}$$

Now note that, using a completion of squares argument, we can write

$$\begin{aligned} J &= x_0^*\Pi_0^{-1}x_0 + u^*Q^{-1}u + (y - \mathcal{O}x_0 - \Gamma u)^* R^{-1}(y - \mathcal{O}x_0 - \Gamma u) \\ &= \begin{bmatrix} x_0^* - \hat{x}_{0|N}^* & u^* - \hat{u}_{|N}^* \end{bmatrix} \begin{bmatrix} \Pi_0^{-1} & \\ & Q^{-1} \end{bmatrix} \begin{bmatrix} x_0 - \hat{x}_{0|N} \\ u - \hat{u}_{|N} \end{bmatrix} + y^* R_y^{-1} y, \end{aligned}$$

where

$$\begin{bmatrix} \hat{x}_{0|N} \\ \hat{u}_{|N} \end{bmatrix} = \begin{bmatrix} \Pi_0 & \\ & Q \end{bmatrix} \begin{bmatrix} \mathcal{O}^* \\ \Gamma^* \end{bmatrix} R_y^{-1} y. \tag{10.4.14}$$

With the above choice of disturbance (for x_0 and u) we have

$$J = \hat{x}_{0|N}^* \Pi_0^{-1} \hat{x}_{0|N} + \hat{u}_{|N}^* Q^{-1} \hat{u}_{|N} + v^* R^{-1} v = y^* R_y^{-1} y, \tag{10.4.15}$$

and

$$\begin{aligned} e_p &= L^{-1}y - v && \text{using (10.4.12)} \\ &= L^{-1}y - (y - \mathcal{O}\hat{x}_{0|N} - \Gamma \hat{u}_{|N}) && \text{using (10.4.11)} \\ &= L^{-1}y - \left(y - (\mathcal{O}\Pi_0\mathcal{O}^* + \Gamma^* Q \Gamma) R_y^{-1} y\right) && \text{using (10.4.14)} \\ &= L^{-1}y - \left(y - (R_y - R) R_y^{-1} y\right) && \text{using (10.4.7)} \\ &= L^{-1}y - R R_y^{-1} y. \end{aligned}$$

[Note that we have not yet specified our choice of disturbance v, or equivalently, of observations y.]

Now if we define the notation, $\|a\|^2 = \sum_{i=0}^N a_i^* a_i$, we can write the maximum energy gain from disturbances to estimation errors as

$$\sup_{x_0,u,v} \frac{\sum_{i=0}^N (H_i x_i - H_i \hat{x}_i)^* R_i^{-1} (H_i x_i - H_i \hat{x}_i)}{x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^N u_i^* Q_i^{-1} u_i + \sum_{i=0}^N v_i^* R_i^{-1} v_i} = \sup_{x_0,u,v} \frac{\|R^{-1/2} e_p\|^2}{x_0^* \Pi_0^{-1} x_0 + \|Q^{-1/2} u\|^2 + \|R^{-1/2} v\|^2}.$$

Now, with our choice of disturbance signal, we can lower bound the maximum energy gain as follows,

$$\begin{aligned} \sup_{x_0,u,v} \frac{\|R^{-1/2} e_p\|^2}{x_0^* \Pi_0^{-1} x_0 + \|Q^{-1/2} u\|^2 + \|R^{-1/2} v\|^2} &\geq \sup_y \frac{\|R^{-1/2} e_p\|^2}{y^* R_y^{-1} y} \\ &= \sup_y \frac{\|R^{-1/2} (L^{-1} - R R_y^{-1}) y\|^2}{R_e^{-1/2} L^{-1} y} \\ &= \sup_x \frac{\|R^{-1/2} (L^{-1} - R R_y^{-1}) L R_e^{1/2} x\|^2}{\|x\|^2} \\ &= \sup_x \frac{\|(R^{-1/2} R_e^{1/2} - R^{*/2} L^{-*} R_e^{-*/2}) x\|^2}{\|x\|^2} \\ &= \left[\bar{\sigma}\left(R^{-1/2} R_e^{1/2} - R^{*/2} L^{-*} R_e^{-*/2}\right)\right]^2. \end{aligned}$$

Now the triangular inequality for the maximum singular value yields

$$\sup_{x_0,u,v} \frac{\|R^{-1/2} e_p\|^2}{x_0^* \Pi_0^{-1} x_0 + \|Q^{-1/2} u\|^2 + \|R^{-1/2} v\|^2} \geq \left[\bar{\sigma}(R^{-1/2} R_e^{1/2}) - \bar{\sigma}(R^{*/2} L^{-*} R_e^{-*/2})\right]^2, \tag{10.4.16}$$

Moreover, note that we have

$$\begin{aligned}
\bar{\sigma}(R^{-1/2}R_e^{1/2}) &= \sqrt{\bar{\sigma}(R^{-1/2}R_eR^{-*/2})} \\
&= \sqrt{\bar{\sigma}\left(\mathrm{diag}(R_0^{-1/2}R_{e,0}R_0^{-*/2},\ldots R_N^{-1/2}R_{e,N}R_N^{-*/2})\right)} \\
&= \sqrt{\sup_i \ \bar{\sigma}(R_i^{-1/2}R_{e,i}R_i^{-*/2})} \\
&= \sqrt{r}, \qquad (10.4.17)
\end{aligned}$$

and

$$\begin{aligned}
(R^{*/2}L^{-*}R_e^{-*/2})(R_e^{-1/2}L^{-1}R^{1/2}) &= R^{*/2}R_y^{-1}R^{1/2} \\
&= (I+R^{-1/2}\mathcal{O}\Pi_0\mathcal{O}^*R^{-*/2}+R^{-1/2}\Gamma Q\Gamma^*R^{-*/2})^{-1} \\
&\leq I,
\end{aligned}$$

so that

$$\bar{\sigma}(R^{*/2}L^{-*}R_e^{-*/2}) \leq 1. \qquad (10.4.18)$$

Combining Eqs. (10.4.16), (10.4.17) and (10.4.18) yields the desired result,

$$\sup_{x_0,u,v} \frac{\sum_{i=0}^N (H_ix_i-H_i\hat{x}_i)^*R_i^{-1}(H_ix_i-H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0+\sum_{i=0}^N u_i^*Q_i^{-1}u_i+\sum_{i=0}^N v_i^*R_i^{-1}v_i} \geq (\sqrt{r}-1)^2. \qquad (10.4.19)$$

The proof of the lower bound for filtered errors is very similar. The only difference is that here we need to compute e_f in terms of y (for our choice of disturbances as given by (10.4.14)). In this case

$$e_f = RR_e^{-1}e - v = RR_e^{-1}L^{-1}y - RR_y^{-1}y. \qquad (10.4.20)$$

Thus we now have

$$\begin{aligned}
\sup_{x_0,u,v} \frac{\|R^{-1/2}e_f\|^2}{x_0^*\Pi_0^{-1}x_0+\|Q^{-1/2}u\|^2+\|R^{-1/2}v\|^2} &\geq \sup_y \frac{\|R^{-1/2}e_f\|^2}{y^*R_y^{-1}y} \\
&= \sup_y \frac{\|R^{-1/2}(RR_e^{-1}L^{-1}-RR_y^{-1})y\|^2}{R_e^{-1/2}L^{-1}y} \\
&= \sup_x \frac{\|R^{-1/2}(RR_e^{-1}L^{-1}-RR_y^{-1})LR_e^{1/2}x\|^2}{\|x\|^2}
\end{aligned}$$

$$
\begin{aligned}
&= \sup_x \frac{\|(R^{*/2}R_e^{-*/2} - R^{*/2}L^{-*}R_e^{-*/2})x\|^2}{\|x\|^2} \\
&= \left[\bar{\sigma}\left(R^{*/2}R_e^{-*/2} - R^{*/2}L^{-*}R_e^{-*/2}\right)\right]^2 \\
&\geq \left[\bar{\sigma}(R^{*/2}R_e^{-*/2}) - \bar{\sigma}(R^{*/2}L^{-*}R_e^{-*/2})\right]^2 \\
&\geq \left[\sqrt{1/\underline{r}} - 1\right]^2,
\end{aligned}
$$

where in the last step we used the (readily verifiable) fact that

$$\bar{\sigma}(R^{*/2}R_e^{-*/2}) = \sqrt{1/\underline{r}}. \tag{10.4.21}$$

But this is our desired result.

■

10.5 RLS Adaptive Filtering

We are now in a position to specialize the result of Theorem 10.2.1 to the case of adaptive filtering. Recall that in adaptive filtering the model is given by

$$d_i = h_i^T w + v_i, \tag{10.5.1}$$

where d_i is the observation, $h_i^T = \begin{bmatrix} h_{i1} & h_{i2} & \dots & h_{in} \end{bmatrix}$ is a known $1 \times n$ input vector, w is the unknown weight vector that we intend to estimate, and v_i is an unknown disturbance signal.

As mentioned earlier in Secs. 1.2.1 and Sec. 9.5, the adaptive filtering problem is a special case of a state-space estimation problem (see the state-space model (10.2.1)) resulting from the parameters,

$$F_i = I_n \quad , \quad G_i = 0 \quad , \quad H_i = h_i^T \quad , \quad R_i = I_p. \tag{10.5.2}$$

The RLS (recursive-least-squares) algorithm is essentially the Kalman filter corresponding to a state-space model with the above system matrices. Thus the least-squares estimates $\hat{w}_{|i}$ (of the weight vector w, using the observations, $\{d_j, j \leq i\}$) obey the following recursions,

$$\hat{w}_{|i} = \hat{w}_{|i-1} + k_{p,i}(d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{\text{-}1} \tag{10.5.3}$$

where

$$k_{p,i} = \frac{P_i h_i}{1 + h_i^T P_i h_i}, \tag{10.5.4}$$

and P_i satisfies the Riccati recursion,

$$P_{i+1} = P_i - \frac{P_i h_i h_i^T P_i}{1 + h_i^T P_i h_i}, \qquad P_0 = \mu I. \tag{10.5.5}$$

It is also useful to remark that at each time instant, i, the above RLS algorithm solves the following least-squares problem,

$$\min_w \left[\mu^{-1} |w - \hat{w}_{|-1}|^2 + \sum_{j=0}^{i} |d_j - h_j^T w|^2 \right], \tag{10.5.6}$$

where $\mu^{-1}|w - w_{-1}|^2$ is a possible regularization term that reflects a priori knowledge as to how close w is to the initial estimate w_{-1}. The special case where $mu = \infty$, so that the first term in the cost function of (10.5.6) disappears, is referred to as a *pure* least-squares problem. We will shortly have more to say about such problems.

As before let us define the following prediction and filtered estimation errors,

$$e_{p,i} = h_i^T w - h_i^T \hat{w}_{|i-1}, \tag{10.5.7}$$

and

$$e_{f,i} = h_i^T w - h_i^T \hat{w}_{|i-1}. \tag{10.5.8}$$

It will also be useful to define the following smoothed estimation error (since we know from Sec. 3.4 that least-squares smoothers are H^∞ optimal),

$$e_{s,i} = h_i^T w - h_i^T \hat{w}_{|N}, \qquad i = 0, 1 \ldots N. \tag{10.5.9}$$

The following result is now immediate.

Theorem 10.5.1 (Bounds for the H^∞ Norm of the RLS Algorithm) *Consider the adaptive filtering model (10.5.1) and the least-squares estimates $\hat{w}_{|i}$, given by the RLS algorithm (10.5.3). Then for any N, we have the following results,*

$$(\sqrt{\bar{r}} - 1)^2 \leq \sup_{w, v \in h^2} \frac{\sum_{i=0}^{N} (h_i^T w - h_i^T \hat{w}_{|i-1})^* (h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w - \hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq (\sqrt{\bar{r}} + 1)^2 \tag{10.5.10}$$

$$(\sqrt{1/\underline{r}}-1)^2 \le \sup_{w,v\in h^2} \frac{\sum_{i=0}^N (h_i^T w - h_i^T \hat{w}_{|i})^*(h_i^T w - h_i^T \hat{w}_{|i})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^N v_i^* R_i^{-1} v_i} \le (\sqrt{1/\underline{r}}+1)^2 \quad (10.5.11)$$

and

$$\sup_{w,v\in h^2} \frac{\sum_{i=0}^N (h_i^T w - h_i^T \hat{w}_{|N})^*(h_i^T w - h_i^T \hat{w}_{|N})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^N v_i^* R_i^{-1} v_i} = 1 \quad (10.5.12)$$

where we have defined,

$$\bar{r} = \sup_i \left[1 + h_i P_i h_i^T\right], \quad (10.5.13)$$

and

$$\underline{r} = \inf_i \left[1 + h_i P_i h_i^T\right]. \quad (10.5.14)$$

Proof: The proof of the bounds (10.5.10) and (10.5.11) follow immediately from Theorem 10.2.1 when the system matrices are appropriately specialized. Eq. (10.5.12) follows from the H^∞ optimality of least-squares smoothers, as given by Sec. 3.4, and by the results of Chapter 9 on H^∞-optimal adaptive filters.

■

In the RLS algorithm the it is easy to solve (10.5.5) to obtain $P_i = (\mu^{-1}I + \sum_{j=0}^{i-1} h_j h_j^T)^{-1}$, which implies that the P_i are a monotonically decreasing sequence of matrices. If we assume that the input vectors h_i have equal magnitude (*i.e.* $h_i^T h_i =$ const.) then we have the following result.

Corollary 10.5.1 (Constant Magnitude Inputs) *If the input vectors have constant magnitude $h_i^T h_i = \bar{h}^2$ then*

$$(\sqrt{1+\mu\bar{h}^2}-1)^2 \le \sup_{w,v\in h^2} \frac{\sum_{i=0}^N (h_i^T w - h_i^T \hat{w}_{|i-1})^*(h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^N v_i^* R_i^{-1} v_i} \le (\sqrt{1+\mu\bar{h}^2}+1)^2. \quad (10.5.15)$$

The following is also straightforward.

Corollary 10.5.2 *If $\bar{h}^2 \triangleq \sup_i h_i^T h_i$ and $\underline{h}^2 \triangleq \inf_i h_i^T h_i$ then*

$$(\sqrt{1+\mu\underline{h}^2}-1)^2 \le \sup_{w,v\in h^2} \frac{\sum_{i=0}^N (h_i^T w - h_i^T \hat{w}_{|i-1})^*(h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^N v_i^* R_i^{-1} v_i} \le (\sqrt{1+\mu\bar{h}^2}+1)^2. \quad (10.5.16)$$

Finally, the following result for filtered errors is also immediate and follows readily from the expression $P_i = (\mu^{-1}I + \sum_{j=0}^{i-1} h_j h_j^T)^{-1}$.

Corollary 10.5.3 *We have*

$$\lim_{i\to\infty} \underline{r} = 1, \tag{10.5.17}$$

so that

$$\sup_{w,v\in h^2} \frac{\sum_{i=0}^{N}(h_i^T w - h_i^T \hat{w}_{|i})^*(h_i^T w - h_i^T \hat{w}_{|i})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq 4. \tag{10.5.18}$$

Remarks:

(i) Corollary 10.5.2 has an interesting interpretation: for large values of μ, the RLS algorithm is less robust with respect to prediction errors. In fact, we see that the (upper and lower bounds of the) H^∞ norm grows as $\sqrt{\mu}$. This is reminiscent of the robustness properties of LMS, where, as shown in Chapter 9, the learning rate μ had to be small enough to guarantee H^∞ optimality.

More importantly, Corollary 10.5.2 shows that the pure least-squares problem (corresponding to $\mu = \infty$) can be highly nonrobust with respect to prediction errors.

(ii) From Corollary 10.5.3, for filtered errors, the RLS algorithm yields,

$$\sup_{w,v\in h_2} \frac{\|e_f\|_2^2}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \|v\|_2^2} \leq 4,$$

Note that, as with the normalized LMS algorithm, the H^∞ norm does not depend on μ.

The above result shows that, for filtered errors, least-squares algorithms are at most four times worse than H^∞-optimal algorithms (where $\gamma_{opt} = 1$). This demonstrates an intermediate stage between the smoothed error case (which has access to all the observations, and where the H^∞ and H^2 optimal filters coincide) and the prediction error case (which does not have access to current observations, and where the performance of LMS and RLS can be drastically different).

10.5.1 An Alternative Lower Bound

As mentioned earlier, the most natural method for finding lower bounds to the H^∞ norm is to calculate the ratio $\frac{\|e\|_2^2}{\mu^{-1}|w-w_{-1}|^2+\|v\|_2^2}$ for a particular choice of disturbance $w-w_{-1}$ and $\{v_i\}$.

Since the RLS algorithm is considerably simpler than the Kalman filter (in its full generality), it is much easier to compute the energy gain for various disturbances in the RLS case than it is for the Kalman filter. The particular choice of disturbances,

$$w-w_{-1}=\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \quad \text{and} \quad v_i=\frac{-1}{(1+i\mu\underline{h}^2)(1+(i+1)\mu\underline{h}^2)} \tag{10.5.19}$$

where $\underline{h}^2=\inf_i h_i^T h_i$, leads to the following result.

Lemma 10.5.1 (Alternative Lower Bound) *Consider the adaptive filtering model (10.5.1) and the least-squares estimates $\hat{w}_{|i}$, given by the RLS algorithm (10.5.3). A lower bound for the maximum energy gain from the disturbances, $\{\mu^{-1/2}(w-\hat{w}_{|-1}), \{v_i\}_{i=0}^N\}$, to the prediction errors, $\{h_i^T w - h_i^T\hat{w}_{|i-1}\}_{i=0}^N$, is given by*

$$\underline{h}^2\frac{(\frac{1}{\underline{h}^2}+\sqrt{\mu})^2 S_2(\mu\underline{h})-\frac{2}{\underline{h}^2}(\frac{1}{\underline{h}^2}+\sqrt{\mu})S_3(\mu\underline{h})+\frac{1}{\underline{h}^4}S_4(\mu\underline{h})}{1-\frac{1}{\mu^2\underline{h}^4}(1+\frac{1}{\mu\underline{h}^2})+\frac{2}{\mu^2\underline{h}^4}S_2(\mu\underline{h})}, \tag{10.5.20}$$

where

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

Proof: The proof involves the algebraic computation of the energy gain and will be omitted for brevity. Similar results can be established for filtered errors.

∎

10.5.2 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 plotted the results of the H^∞ norm of RLS as a function of the number of iterations for different values of μ. The upper and lower bounds of Theorem 10.5.1 and Lemma 10.5.1 are also given.

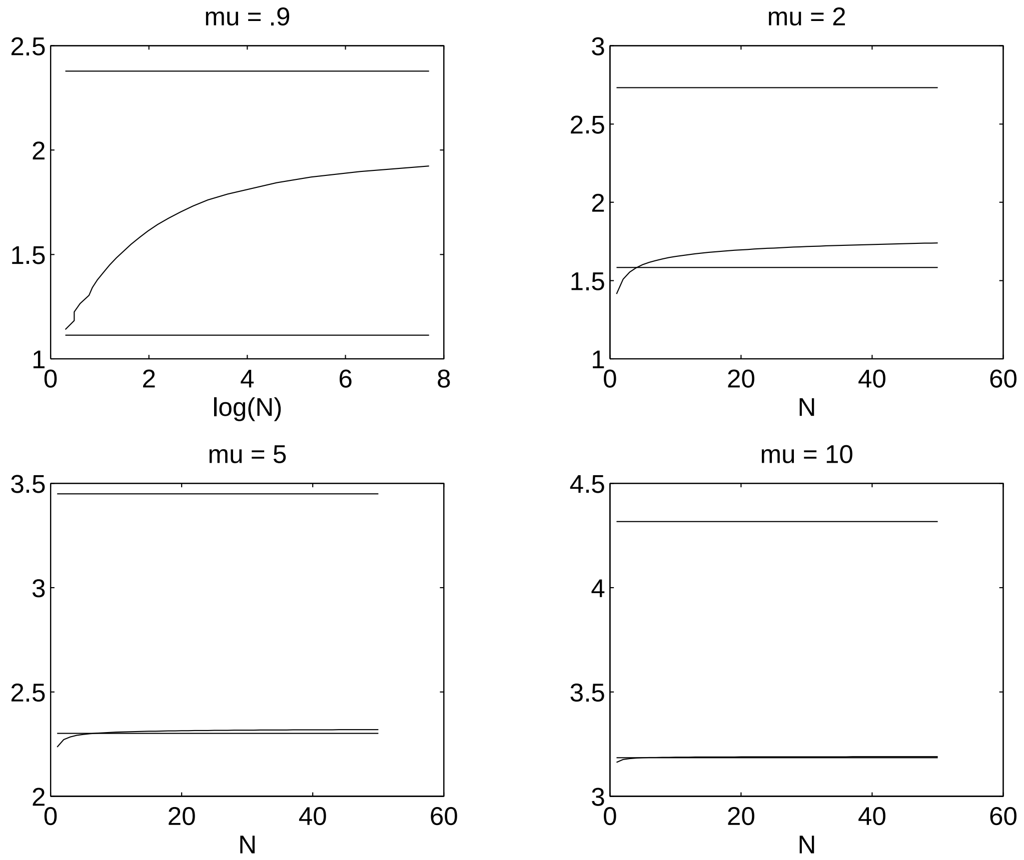

Figure 10.1: H^∞ norm of RLS as a function of the number of data points and as a function of μ. The upper and lower bounds of Theorem 10.5.1 and Lemma 10.5.1 are given by the horizontal lines. As can be seen from the figures, in this example, the lower bounds of Lemma 10.5.1 seem quite accurate for large μ.

10.6 Conclusion

In this chapter we obtained upper and lower bounds for the H^∞ norm of the RLS algorithm. These bounds may be used to study the robustness of RLS in different applications. Our results show that the H^∞ norm of RLS depends on the input-output data (viz. on the $\{h_i\}$), as opposed to the LMS and normalized LMS algorithms where the H^∞ norm is independent of the data. The bounds further show that, for prediction errors, the H^∞ norm of RLS grows as the square-root of μ (where $\mu^{-1}I$ is the regularization term in least-squares problems that reflects a priori knowledge of the weight vector), whereas for filtered errors, the H^∞ norm of RLS (and the Kalman filter) is bounded by two. Simulations show that the bounds are quite reasonable compared to the true computed H^∞ norms.