H^∞ Bounds for Least-Squares Estimators *

Babak Hassibi and Thomas Kailath

Information Systems Laboratory

Stanford University

September 25, 1997

Abstract

In this paper we obtain upper and lower bounds for the H^∞ norm of the Kalman filter and RLS algorithm, with respect to prediction and filtered errors. These bounds can be used to study the robustness properties of such estimators. One main conclusion is that, unlike H^∞-optimal estimators which do not allow for any amplification of the disturbances, the least-squares estimators do allow for such amplification. This fact can be especially pronounced in the prediction error case, whereas in the filtered error case the energy amplification is at most four. Moreover, it is shown that the H^∞ norm for RLS is data-dependent, whereas for LMS and normalized LMS the H^∞ norm is simply unity.

1 Introduction

Since its inception in the early 1960's, the Kalman filter (and the closely related recursive-least-squares (RLS) algorithm) has played a central role in estimation theory and adaptive filtering. Recently, on the other hand, there has been growing interest in (so-called) H^∞ estimation, with the belief that the resulting H^∞-optimal estimators will be more robust with respect to disturbance variation and lack of statistical knowledge of the exogenous signals. Therefore a natural question to ask is what the robustness properties of the Kalman filter and RLS algorithm are within the H^∞ framework.

In an attempt to address this question, in this paper we obtain upper and lower bounds on the H^∞ norm of the Kalman filter and RLS algorithm, with respect to prediction and filtered errors. These

*This work was supported in part by DARPA through the Department of Air Force under contract F49620-95-1-0525-P00001 and by the Joint Service Electronics Program at Stanford under contract DAAH04-94-G-0058-P00003.

bounds are also of interest for several other reasons. First, they demonstrate that unlike the LMS algorithm whose H^∞ norm is unity (independent of the input-output data) [1], 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 (and for the Kalman filter) requires the calculation of the induced two-norm of a linear time-variant 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 (or the Kalman filter) has the desired robustness with respect to a given application.

A brief outline of the paper is as follows. In Sec. 2 we give general upper and lower bounds for the H^∞ norm of the Kalman filter. The proofs of the upper bounds are given in Sec. 3 and are based on certain minimization properties of least-squares estimators. The proofs of the lower bounds are given in Sec. 4 and are essentially based on computing the energy gains for suitably chosen disturbances. Sec. 5 specializes the general results of Sec. 2 to the adaptive filtering problem and discusses its various implications. The paper concludes with Sec. 6.

2 A General H^∞ Bound

Consider the possibly time-variant state-space model

$$\begin{cases} x_{i+1} &= F_i x_i + G_i u_i, \qquad x_0 \\ y_i &= H_i x_i + v_i, \qquad i \geq 0 \end{cases} \tag{2.1}$$

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.

It is wellknown 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

$$\hat{x}_{i+1} = F_i \hat{x}_i + K_{p,i}(y_i - H_i \hat{x}_i), \tag{2.2}$$

where $K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1}$ and $R_{e,i} = R_i + H_i P_i H_i^*$ and where P_i satisfies the Riccati recursion,

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0. \tag{2.3}$$

[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,

$$\hat{x}_{i+1|i+1} = F_i\hat{x}_{i|i} + K_{f,i+1}(y_{i+1} - H_{i+1}F_i\hat{x}_{i|i}), \tag{2.4}$$

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

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 defined as

$$e_{p,i} = H_ix_i - H_i\hat{x}_i \stackrel{\Delta}{=} H_i\tilde{x}_i \quad \text{and} \quad e_{f,i} = H_ix_i - H_i\hat{x}_{i|i} \stackrel{\Delta}{=} H_i\tilde{x}_{i|i}. \tag{2.5}$$

Note that both these estimation errors are different from the innovations (the prediction errors for estimating y_i), $e_i \stackrel{\Delta}{=} y_i - H_i\hat{x}_i$. Indeed it is straightforward to see that we have

$$e_{p,i} = e_i - v_i \quad \text{and} \quad e_{f,i} = R_iR_{e,i}^{-1}e_i - v_i. \tag{2.6}$$

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 the identities in (2.6).

We can now state the main result of this paper.

Theorem 1 (Bounds for the H^∞ Norm of the Kalman Filter) *Consider the standard state-space model (2.1) and the predicted and filtered forms of the Kalman filter recursions, (2.2) and (2.4). Then for any N, we have the following results,*

$$(\sqrt{\bar{r}} - 1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^N (H_ix_i - H_i\hat{x}_i)^*R_i^{-1}(H_ix_i - H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^N u_i^*Q_i^{-1}u_i + \sum_{i=0}^N v_i^*R_i^{-1}v_i} \leq (\sqrt{\bar{r}} + 1)^2 \tag{2.7}$$

and

$$(\sqrt{1/\underline{r}} - 1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^N (H_ix_i - H_i\hat{x}_{i|i})^*R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^N u_i^*Q_i^{-1}u_i + \sum_{i=0}^N v_i^*R_i^{-1}v_i} \leq (\sqrt{1/\underline{r}} + 1)^2 \tag{2.8}$$

where h^2 denotes the space of square-summable causal sequences, and where we have defined,

$$\bar{r} = \sup_i \ \bar{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right) \quad \textit{and} \quad \underline{r} = \inf_i \ \underline{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right) \tag{2.9}$$

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

Remarks:

(i) Note that the ratios in the inequalities (2.7) and (2.8) 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 (2.7) and (2.8) 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 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 1 bounds the H^∞ norm of the Kalman filter by quantities related to the maximum and minimum singular values of the normalized innovations variance, $R_i^{-1/2}R_{e,i}R_i^{-*/2}$. In particular, note that

$$R_i^{-1/2}R_{e,i}R_i^{-*/2} = I_p + R_i^{-1/2}H_iP_iH_i^*R_i^{-*/2} \geq I_p, \tag{2.10}$$

so that

$$\bar{r} \geq \underline{r} \geq 1 \geq 1/\underline{r}. \tag{2.11}$$

But using (2.8) this means that

$$\sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^N (H_ix_i - H_i\hat{x}_{i|i})^* R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^N u_i^*Q_i^{-1}u_i + \sum_{i=0}^N v_i^*R_i^{-1}v_i} \leq 4, \tag{2.12}$$

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.

This should be compared with the optimal energy gain $\gamma_{opt}^2 \approx 1$ obtained from an H^∞-optimal estimator [2]. Thus, for filtered errors, and from an H^∞ point of view, H^2-optimal estimators have a performance (roughly) four times worse than H^∞-optimal estimators.

(iv) The bounds of Theorem 1 are true for any value of N, and, in fact, they are also true when the upper limits of the sums in (2.7) and (2.8) are infinite. In other words, it is true that,

$$(\sqrt{\bar{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^\infty (H_ix_i - H_i\hat{x}_i)^* R_i^{-1}(H_ix_i - H_i\hat{x}_i)}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^\infty u_i^*Q_i^{-1}u_i + \sum_{i=0}^\infty v_i^*R_i^{-1}v_i} \leq (\sqrt{\bar{r}}+1)^2 \tag{2.13}$$

and

$$(\sqrt{1/\underline{r}}-1)^2 \leq \sup_{x_0,u,v\in h^2} \frac{\sum_{i=0}^{\infty}(H_ix_i - H_i\hat{x}_{i|i})^* R_i^{-1}(H_ix_i - H_i\hat{x}_{i|i})}{x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{\infty} u_i^* Q_i^{-1}u_i + \sum_{i=0}^{\infty} v_i^* R_i^{-1} v_i} \leq (\sqrt{1/\underline{r}}+1)^2 \tag{2.14}$$

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 infinite-horizon case we have

$$(\sqrt{\bar{r}}-1)^2 \leq \sup_{u,v\in l^2} \frac{\sum_{i=-\infty}^{\infty}(H_ix_i - H_i\hat{x}_i)^* R^{-1}(H_ix_i - H_i\hat{x}_i)}{\sum_{i=-\infty}^{\infty} u_i^* Q_i^{-1}u_i + \sum_{i=-\infty}^{\infty} v_i^* R_i^{-1} v_i} \leq (\sqrt{\bar{r}}+1)^2 \tag{2.15}$$

and

$$(\sqrt{1/\underline{r}}-1)^2 \leq \sup_{u,v\in l^2} \frac{\sum_{i=-\infty}^{\infty}(Hx_i - H\hat{x}_{i|i})^* R^{-1}(Hx_i - H\hat{x}_{i|i})}{\sum_{i=-\infty}^{\infty} u_i^* Q^{-1}u_i + \sum_{i=-\infty}^{\infty} v_i^* R^{-1} v_i} \leq (\sqrt{1/\underline{r}}+1)^2 \tag{2.16}$$

where l^2 is the space of square-summable sequences and now

$$\bar{r} = \bar{\sigma}\left(R^{-1/2}R_eR^{-*/2}\right) \quad \text{and} \quad \underline{r} = \underline{\sigma}\left(R^{-1/2}R_eR^{-*/2}\right), \tag{2.17}$$

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

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

3 Proof of the Upper Bounds

To prove the upper bounds of Theorem 1 we need the following three facts.

Lemma 1 (Minimization of a Quadratic Form) *We have*

$$\min_{x_0,\{u_i,v_i\}} \left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N} u_i^*Q_i^{-1}u_i + \sum_{i=0}^{N} v_i^*R_i^{-1}v_i\right] = \sum_{i=0}^{N} e_i^*R_{e,i}^{-1}e_i, \tag{3.1}$$

where the minimization is taken subject to the state-space constraints (2.1), and where $e_i = y_i - H_i\hat{x}_i$ *is the innovations.*

Proof: This is a wellknown result. One proof can be found in [3].

■

Lemma 2 (Simple Inequality) *For any vectors a, b, and any matrix $M > 0$, we have*

$$(a+b)^*M(a+b) \geq (1-\frac{1}{\alpha})a^*Ma + (1-\alpha)b^*Mb, \qquad \forall \alpha > 0. \tag{3.2}$$

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

$$\min_{\alpha>1}\frac{\alpha^2+(\beta-1)\alpha}{\alpha-1}=(1+\sqrt{\beta})^2 \quad \textit{and} \quad \arg\min_{\alpha>1}\frac{\alpha^2+(\beta-1)\alpha}{\alpha-1}=1+\sqrt{\beta}. \tag{3.3}$$

Proof: Readily verified via differentiation.

■

We shall first prove the upper bound in (2.7) for the prediction error case. To this end, define

$$J \triangleq x_0^*\Pi_0^{-1}x_0+\sum_{i=0}^{N}u_i^*Q_i^{-1}u_i+\sum_{i=0}^{N}v_i^*R_i^{-1}v_i. \tag{3.4}$$

Now using Lemma 1 it is obvious that

$$J \geq \min_{x_0,\{u_i,v_i\}} J=\sum_{i=0}^{N}e_i^*R_{e,i}^{-1}e_i. \tag{3.5}$$

Thus we may write,

$$\begin{aligned} J\geq\sum_{i=0}^{N}e_i^*R_{e,i}^{-1}e_i &= \sum_{i=0}^{N}e_i^*R_i^{-*/2}R_i^{*/2}R_{e,i}^{-1}R_i^{1/2}R_i^{-1/2}e_i \\ &\geq \sum_{i=0}^{N}\underline{\sigma}\left(R_i^{*/2}R_{e,i}^{-1}R_i^{1/2}\right)e_i^*R_i^{-1}e_i \\ &\geq \left[\inf_i\ \underline{\sigma}\left(R_i^{*/2}R_{e,i}^{-1}R_i^{1/2}\right)\right]\sum_{i=0}^{N}e_i^*R_i^{-1}e_i \\ &= \left[\frac{1}{\sup_i\ \bar{\sigma}\left(R_i^{-1/2}R_{e,i}R_i^{-*/2}\right)}\right]\sum_{i=0}^{N}e_i^*R_i^{-1}e_i \\ &= \frac{1}{\bar{r}}\sum_{i=0}^{N}e_i^*R_i^{-1}e_i \\ &= \frac{1}{\bar{r}}\sum_{i=0}^{N}(e_{p,i}+v_i)^*R_i^{-1}(e_{p,i}+v_i) \quad \text{using (2.6)} \end{aligned} \tag{3.6}$$

Now using Lemma 2 with $a=e_{p,i}$, $b=v_i$ and $M=R_i^{-1}$, we have

$$x_0^*\Pi_0^{-1}x_0+\sum_{i=0}^{N}u_i^*Q_i^{-1}u_i+\sum_{i=0}^{N}v_i^*R_i^{-1}v_i\geq\frac{1}{\bar{r}}\sum_{i=0}^{N}\left[(1-\alpha)v_i^*R_i^{-1}v_i+(1-\frac{1}{\alpha})e_{p,i}^*R_i^{-1}e_{p,i}\right], \tag{3.7}$$

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

$$(1-\frac{1}{\alpha})\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i} \leq \bar{r} x_0^* \Pi_0^{-1} x_0 + \bar{r}\sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \bar{r}(1-\frac{1-\alpha}{\bar{r}})\sum_{i=0}^{N} v_i^* R_i^{-1} v_i, \tag{3.8}$$

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 3, we have

$$\min_{\alpha>1} \frac{\alpha^2+(\bar{r}-1)\alpha}{\alpha-1} = (1+\sqrt{\bar{r}})^2 \quad \text{and} \quad \arg\min_{\alpha>1} \frac{\alpha^2+(\bar{r}-1)\alpha}{\alpha-1} = 1+\sqrt{\bar{r}}. \tag{3.9}$$

Therefore

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

Therefore, we have

$$\frac{\sum_{i=0}^{N} e_{p,i}^* R_i^{-1} e_{p,i}}{x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq (1+\sqrt{\bar{r}})^2, \tag{3.10}$$

for all $\{x_0, \{u_i, v_i\}_{i=0}^{N}\}$,which is the desired result.

To prove the upper bound of (2.8), for the filtered estimation errors, we need to proceed as follows.

$$\begin{aligned}
J \geq \sum_{i=0}^{N} e_i^* R_{e,i}^{-1} e_i &= \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1} R_{e,i} R_{e,i}^{-1} R_{e,i} R_i^{-1} (e_{f,i}+v_i) \quad \text{using (2.6)} \\
&= \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-*/2} R_i^{-1/2} R_{e,i} R_i^{-*/2} R_i^{-1/2} (e_{f,i}+v_i) \\
&\geq \sum_{i=0}^{N} \underline{\sigma}\left(R_i^{-1/2} R_{e,i} R_i^{-*/2}\right) (e_{f,i}+v_i)^* R_i^{-1} (e_{f,i}+v_i) \\
&\geq \left[\inf_i \ \underline{\sigma}\left(R_i^{-1/2} R_{e,i} R_i^{-*/2}\right)\right] \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1} (e_{f,i}+v_i) \\
&= \underline{r} \sum_{i=0}^{N} (e_{f,i}+v_i)^* R_i^{-1} (e_{f,i}+v_i).
\end{aligned} \tag{3.11}$$

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

4 Proof of the Lower Bounds

Perhaps the most general way of computing a lower bound for the H^∞ norm of the Kalman filter, or any other algorithm for that matter, is to compute the energy gain for some particular choice of disturbance signal, $\{x_0, \{u_i, v_i\}_{i=0}^N\}$. We shall presently see that the special choice of disturbance signal that yields the lower bound of Theorem 1 is that disturbance signal that minimizes the quadratic form J of (3.4), subject to the state-space constraints (2.1). To facilitate the presentation of the proof it will be convenient to introduce the following lemmas which are the straightforward counterparts of Lemmas 2 and 3, and whose proofs will therefore be omitted.

Lemma 4 (Simple Inequality) *For any vectors a, b, and any matrix $M > 0$, we have*

$$(a+b)^* M (a+b) \leq (1+\frac{1}{\alpha}) a^* M a + (1+\alpha) b^* M b, \qquad \forall \alpha > 0. \tag{4.1}$$

Lemma 5 (A Simple Maximization) *For all $\beta > 1$, we have*

$$\max_{\alpha>0} \frac{-\alpha^2 + (\beta-1)\alpha}{\alpha+1} = (\sqrt{\beta}-1)^2 \quad \textit{and} \quad \arg\max_{\alpha>0} \frac{-\alpha^2+(\beta-1)\alpha}{\alpha+1} = \sqrt{\beta}-1. \tag{4.2}$$

We shall first prove the lower bound in (2.7) for the prediction error case. To this end, for a given sequence of observations $\{y_i\}_{i=0}^N$, let $\{\bar{x}_0, \{\bar{u}_i, \bar{v}_i\} i = 0^N, \}$ denote the values that minimize the quadratic form J in (3.4). We may thus write

$$\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i + \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i = \sum_{i=0}^N e_i^* R_{e,i}^{-1} e_i. \tag{4.3}$$

Since the $\{y_i\}_{i=0}^N$, and hence the $\{e_i\}_{i=0}^N$, are arbitrary, we can always choose them such that the lower bound in (3.6) is achieved. Denoting this choice by $\{\bar{e}_i\}_{i=0}^N$, we may write

$$\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i + \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i = \frac{1}{\bar{r}} \sum_{i=0}^N (\bar{e}_{p,i} + \bar{v}_i)^* R_i^{-1} (\bar{e}_{p,i} + \bar{v}_i). \tag{4.4}$$

Now using Lemma 4 with $a = \bar{e}_{p,i}$, $b = \bar{v}_i$ and $M = R_i^{-1}$, we have

$$\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i + \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i \leq \frac{1}{\bar{r}} \sum_{i=0}^N \left[(1+\alpha) \bar{v}_i^* R_i^{-1} \bar{v}_i + (1+\frac{1}{\alpha}) \bar{e}_{p,i}^* R_i^{-1} \bar{e}_{p,i} \right], \tag{4.5}$$

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

$$\begin{aligned}
\sum_{i=0}^N \bar{e}_{p,i}^* R_i^{-1} \bar{e}_{p,i} \quad &\geq \quad \frac{\bar{r}}{1+\frac{1}{\alpha}} \left[\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i \right] + \frac{\bar{r}}{1+\frac{1}{\alpha}} (1 - \frac{1+\alpha}{\bar{r}}) \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i \\
&= \quad \frac{\bar{r}}{1+\frac{1}{\alpha}} \left[\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i \right] + \frac{-\alpha^2 + \alpha(\bar{r}-1)}{\alpha+1} \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i.
\end{aligned} \tag{4.6}$$

To obtain the "tightest" possible bound on $\sum_{i=0}^N \bar{e}_{p,i}^* R_i^{-1} \bar{e}_{p,i}$, let us maximize, over $\alpha > 0$, the coefficient of $\sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i$ on the RHS of the above inequality. But from Lemma 5, we have

$$\max_{\alpha>0} \frac{-\alpha^2 + (\bar{r}-1)\alpha}{\alpha+1} = (\sqrt{\bar{r}}-1)^2 \quad \text{and} \quad \arg\max_{\alpha>0} \frac{-\alpha^2 + (\bar{r}-1)\alpha}{\alpha+1} = \sqrt{\bar{r}}-1. \tag{4.7}$$

Inserting this value of α into the inequality (4.6) leads to

$$\begin{aligned}
\sum_{i=0}^N \bar{e}_{p,i}^* R_i^{-1} \bar{e}_{p,i} &\geq \frac{\bar{r}}{1+\frac{1}{\sqrt{\bar{r}}-1}} \left[\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i \right] + (\sqrt{\bar{r}}-1)^2 \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i \\
&= \sqrt{\bar{r}}(\sqrt{\bar{r}}-1) \left[\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i \right] + (\sqrt{\bar{r}}-1)^2 \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i \\
&\geq (\sqrt{\bar{r}}-1)^2 \left[\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i \right] + (\sqrt{\bar{r}}-1)^2 \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i.
\end{aligned}$$

Therefore

$$\frac{\sum_{i=0}^N \bar{e}_{p,i}^* R_i^{-1} \bar{e}_{p,i}}{\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i + \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i} \leq (\sqrt{\bar{r}}-1)^2, \tag{4.8}$$

which shows that the bound $(\sqrt{\bar{r}}-1)^2$ is achievable by the appropriate choice of $\{\bar{x}_0, \{\bar{u}_i, \bar{v}_i\}_{i=0}^N\}$. But this is the desired result.

The proof of the lower bound of (2.8) for the filtered estimation errors is very similar. One needs only to note that the lower bound in (3.11) is achievable, *i.e.*, that there exists $\{\bar{x}_0, \{\bar{u}_i, \bar{v}_i\}_{i=0}^N\}$ such that

$$\bar{x}_0^* \Pi_0^{-1} \bar{x}_0 + \sum_{i=0}^N \bar{u}_i^* Q_i^{-1} \bar{u}_i + \sum_{i=0}^N \bar{v}_i^* R_i^{-1} \bar{v}_i = \underline{r} \sum_{i=0}^N (\bar{e}_{f,i} + \bar{v}_i)^* R_i^{-1} (\bar{e}_{f,i} + \bar{v}_i). \tag{4.9}$$

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

5 RLS Adaptive Filtering

We are now in a position to specialize the result of Theorem 1 to the case of adaptive filtering. In adaptive filtering the model is given by

$$d_i = h_i^T w + v_i, \tag{5.1}$$

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.

The adaptive filtering problem can be regarded as a special case of a state-space estimation problem resulting from the parameters,

$$F_i = I_n \quad , \quad G_i = 0 \quad , \quad H_i = h_i^T \quad , \quad R_i = I_p. \tag{5.2}$$

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,

$$\hat{w}_{|i} = \hat{w}_{|i-1} + k_{p,i}(d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{-1} \tag{5.3}$$

where $k_{p,i} = \frac{P_i h_i}{1+h_i^T P_i h_i}$, and P_i satisfies the Riccati recursion,

$$P_{i+1} = P_i - \frac{P_i h_i h_i^T P_i}{1 + h_i^T P_i h_i}, \qquad P_0 = \mu I. \tag{5.4}$$

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

$$\min_w \left[\mu^{-1} |w - \hat{w}_{|-1}|^2 + \sum_{j=0}^{i} |d_j - h_j^T w|^2 \right], \tag{5.5}$$

where $\mu^{-1}|w - \hat{w}_{-1}|^2$ is a possible regularization term that reflects a priori knowledge as to how close w is to the initial estimate $\hat{w}_{-1}$. The special case where $\mu = \infty$, so that the first term in the cost function of (5.5) 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, $e_{p,i} = h_i^T w - h_i^T \hat{w}_{|i-1}$ and $e_{f,i} = h_i^T w - h_i^T \hat{w}_{|i-1}$. The following result is now immediate.

Theorem 2 (Bounds for the H^∞ Norm of the RLS Algorithm) *Consider the adaptive filtering model (5.1) and the least-squares estimates $\hat{w}_{|i}$, given by the RLS algorithm (5.3). Then for any N, we have the following results,*

$$(\sqrt{\bar{r}} - 1)^2 \leq \sup_{w,v \in h^2} \frac{\sum_{i=0}^{N} (h_i^T w - h_i^T \hat{w}_{|i-1})^* (h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w - \hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* v_i} \leq (\sqrt{\bar{r}} + 1)^2 \tag{5.6}$$

and

$$(\sqrt{1/\underline{r}} - 1)^2 \leq \sup_{w,v \in h^2} \frac{\sum_{i=0}^{N} (h_i^T w - h_i^T \hat{w}_{|i})^* (h_i^T w - h_i^T \hat{w}_{|i})}{\mu^{-1}|w - \hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* v_i} \leq (\sqrt{1/\underline{r}} + 1)^2 \tag{5.7}$$

where we have defined,

$$\bar{r} = \sup_i \; \left[1 + h_i P_i h_i^T\right] \quad \textit{and} \quad \underline{r} = \inf_i \; \left[1 + h_i P_i h_i^T\right]. \tag{5.8}$$

In the RLS algorithm it is easy to solve (5.4) 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.* $h_i^T h_i = \text{const.}$) then we have the following result.

Corollary 1 (Constant Magnitude Inputs) *If the input vectors have constant magnitude* $h_i^T h_i = \bar{h}^2$ *then*

$$(\sqrt{1+\mu\bar{h}^2}-1)^2 \leq \sup_{w,v\in h^2} \frac{\sum_{i=0}^{N}(h_i^T w - h_i^T \hat{w}_{|i-1})^*(h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq (\sqrt{1+\mu\bar{h}^2}+1)^2. \tag{5.9}$$

The following is also straightforward.

Corollary 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*

$$(\sqrt{1+\mu\underline{h}^2}-1)^2 \leq \sup_{w,v\in h^2} \frac{\sum_{i=0}^{N}(h_i^T w - h_i^T \hat{w}_{|i-1})^*(h_i^T w - h_i^T \hat{w}_{|i-1})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq (\sqrt{1+\mu\bar{h}^2}+1)^2. \tag{5.10}$$

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 3 *We have*

$$\lim_{i\to\infty} \underline{r} = 1, \tag{5.11}$$

so that

$$\sup_{w,v\in h^2} \frac{\sum_{i=0}^{N}(h_i^T w - h_i^T \hat{w}_{|i})^*(h_i^T w - h_i^T \hat{w}_{|i})}{\mu^{-1}|w-\hat{w}_{|-1}|^2 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \leq 4. \tag{5.12}$$

Remarks:

(i) Corollary 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 [1], the learning rate μ had to be small enough to guarantee H^∞ optimality.

More importantly, Corollary 2 shows that the pure least-squares problem (corresponding to $\mu = \infty$) can be highly nonrobust with respect to prediction errors.

(ii) From Corollary 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 [1].

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).

6 Conclusion

In this paper we obtained upper and lower bounds for the H^∞ norm of the Kalman filter and RLS algorithm. These bounds may be used to study the robustness of these algorithms 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 [1]. 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.

References

[1] B. Hassibi, A.H. Sayed, and T. Kailath. H^∞ optimality of the LMS algorithm. *IEEE Transactions on Signal Processing*, 44(2):267–280, Feb. 1996.

[2] B. Hassibi and T. Kailath. H^∞ filtering of signals in additive noise. *Submitted to IEEE Transactions on Automatic Control*, 1997.

[3] B. Hassibi, A.H. Sayed, and T. Kailath. Linear estimation in Krein spaces - Part I: Theory. *IEEE Transactions on Automatic Control*, 41(1):18–33, Jan. 1996.

[4] T. Kailath, A.H. Sayed, and B. Hassibi. *State Space Estimation*. Prentice-Hall, Englewood Cliffs, NJ, 1997.