[5] S. Bhat and D. Bernstein, "Finite-time stability of continuous autonomous systems," SIAM J. Control Optim., to be published.
[6] V. T. Haimo, "Finite time controllers," *SIAM J. Control Optim.*, vol. 24, no. 4, pp. 760–770, 1986.
[7] H. Hermes, "Homogeneous coordinates and continuous asymptotically stabilizing feedback controls," in *Differential Equations, Stability and Control*, S. Elaydi, Ed. New York: Marcel Dekker, 1991, pp. 249–260.
[8] Y. Hong, Y. S. Xu, and J. Huang, "Finite time control for robot manipulators," *Syst. Control Lett.*, submitted for publication.
[9] M. R. James, "Finite time observer design by probabilistic-variational methods," *SIAM J. Control Optim.*, vol. 30, pp. 795–807, 1992.
[10] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1980.
[11] M. Kawski, "Stabilization of nonlinear systems in the plane," *Syst. Control Lett.*, vol. 12, no. 2, pp. 169–175, 1989.
[12] E. R. Rang, "Isochrone families for second order systems," *IEEE Trans Automat. Contr.*, vol. AC-8, pp. 64–65, 1963.
[13] L. Rosier, "Homogeneous Lyapunov function for homogeneous continuous vector field," *Syst. Control Lett.*, vol. 19, no. 4, pp. 467–473, 1992.
[14] E. P. Ryan, "Singular optimal controls for second-order saturating system," *Int. J. Control*, vol. 36, no. 4, pp. 549–564, 1979.

H^∞ Bounds for Least-Squares Estimators

Babak Hassibi and Thomas Kaliath

***Abstract*—In this note, we obtain upper and lower bounds for the H^∞ norm of the Kalman filter and the recursive-least-squares (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 least-mean-squares (LMS) algorithms and normalized LMS, the H^∞ norm is simply unity.**

***Index Terms*—Estimation, H^∞, least-squares, robustness.**

I. Introduction

Since its inception in the early 1960s, 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 initial attempt to address this question, in this note we obtain upper and lower bounds on the H^∞ norm of the Kalman filter and RLS algorithm, with respect to the prediction and filtered errors of the uncorrupted output of a linear time-variant system.[1] These bounds are also of interest for several other reasons. First, they demonstrate that unlike the least-mean-squares (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 Section II, we give general upper and lower bounds for the H^∞ norm of the Kalman filter. The proofs of the upper bounds are given in Section III and are based on certain minimization properties of least-squares estimators. The proofs of the lower bounds are given in Section IV and are essentially based on computing the energy gains for suitably chosen disturbances. Section V specializes the general results of Section II to the adaptive filtering problem and discusses its various implications. The paper concludes with Section VI.

II. A General H^∞ Bound

Consider the possibly time-variant state-space model

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

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. We shall be interested in estimating the uncorrupted output, $s_i = H_i x_i$.

It is well known 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

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

[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 $K_{f,i} = P_i H_i^* R_{e,i}^{-1}$.

Now using $\hat{x}_i$ and $\hat{x}_{i|i}$, the predicted and filtered estimation errors of the uncorrupted output, $s_i = H_i x_i$, are defined as

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

and

$$e_{f,i} = H_i x_i - H_i \hat{x}_{i|i} \triangleq H_i \tilde{x}_{i|i}. \tag{2.5}$$

[1]We should stress that these bounds are not for the problem of parameter estimation, for which causality is not an issue and for which the H^2 and H^∞ solutions coincide.

Manuscript received October 9, 1997; revised June 3, 2000. Recommended by Associate Editor H. Katayama. This work was supported in part by the U.S. Air Force through under Contract F49620-95-1-0525-P00001 and in part by the Joint Service Electronics Program at Stanford University, Stanford, CA, under Contract DAAH04-94-G-0058-P00003.

B. Hassibi is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: hassibi@caltech.edu).

T. Kaliath is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 0018-9286(01)01021-2.

0018–9286/01$10.00 © 2001 IEEE

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

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

The latter equality is justified as follows:

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

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:

$$\begin{aligned} &\left(\sqrt{\overline{r}} - 1\right)^2 \\ &\le \sup_{x_0, u, v \in h^2} \frac{\sum_{i=0}^{N} (H_i x_i - H_i \hat{x}_i)^* R_i^{-1} (H_i x_i - H_i \hat{x}_i)}{x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i} \\ &\le \left(\sqrt{\overline{r}} + 1\right)^2 \end{aligned} \tag{2.7}$$

and

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

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

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

and

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

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

Remarks:

1) Note that the ratios in (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.

2) 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 $\overline{r}$ and $1/\underline{r}$). Indeed the upper and lower bounds differ only by two, since

$$\left(\sqrt{\overline{r}} + 1\right) - \left(\sqrt{\overline{r}} - 1\right) = 2$$

and

$$\left(\sqrt{1/\underline{r}} + 1\right) - \left(\sqrt{1/\underline{r}} - 1\right) = 2.$$

3) 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

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

so that

$$\overline{r} \ge \underline{r} \ge 1 \ge 1/\underline{r}. \tag{2.11}$$

However, using (2.8), this means that

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

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. 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 performances can be drastically different).

4) 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

$$\begin{aligned} &\left(\sqrt{\overline{r}} - 1\right)^2 \\ &\le \sup_{x_0, u, v \in h^2} \frac{\sum_{i=0}^{\infty} (H_i x_i - H_i \hat{x}_i)^* R_i^{-1} (H_i x_i - H_i \hat{x}_i)}{x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{\infty} u_i^* Q_i^{-1} u_i + \sum_{i=0}^{\infty} v_i^* R_i^{-1} v_i} \\ &\le \left(\sqrt{\overline{r}} + 1\right)^2 \end{aligned} \tag{2.13}$$

and

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

where, as before

$$\overline{r} = \sup_i \overline{\sigma}(R_i^{-1/2} R_{e,i} R_i^{-*/2})$$

and

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

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

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

and

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

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

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

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

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

III. 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 $e_i = y_i - H_i\hat{x}_i$ are the innovations.

Proof: This is a well-known result. One proof can be found in [3]. ∎

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

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

Proof: Follows from

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

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

$$\min_{\alpha>1}\frac{\alpha^2+(\beta-1)\alpha}{\alpha-1} = \left(1+\sqrt{\beta}\right)^2$$

and

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

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

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

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

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

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

$$\begin{aligned}\left(1-\frac{1}{\alpha}\right)&\sum_{i=0}^{N}e_{p,\,i}^*R_i^{-1}e_{p,\,i}\\ &\leq \overline{r}x_0^*\Pi_0^{-1}x_0 + \overline{r}\sum_{i=0}^{N}u_i^*Q_i^{-1}u_i + \overline{r}\left(1-\frac{1-\alpha}{\overline{r}}\right)\sum_{i=0}^{N}\\ &\quad\cdot v_i^*R_i^{-1}v_i,\end{aligned} \tag{3.8}$$

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

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

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 right-hand side (RHS) of the above inequality. However, from Lemma 3, we have

$$\min_{\alpha>1} \frac{\alpha^2+(\overline{r}-1)\alpha}{\alpha-1} = (1+\sqrt{\overline{r}})^2$$

and

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

Therefore,

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

Therefore, we have

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

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:

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

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

IV. 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

$$(a+b)^* M (a+b) \le \left(1+\frac{1}{\alpha}\right) a^* M a + (1+\alpha) b^* M b \quad \forall\, \alpha > 0. \tag{4.1}$$

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

$$\max_{\alpha>0} \frac{-\alpha^2+(\beta-1)\alpha}{\alpha+1} = \left(\sqrt{\beta}-1\right)^2$$

and

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

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 $\{\overline{x}_0,\ \{\overline{u}_i,\ \overline{v}_i\}_{i=0}^N\}$, denote the values that minimize the quadratic form J in (3.4). Thus, we may write

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

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.[2] Denoting this choice by $\{\overline{e}_i\}_{i=0}^N$, we may write

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

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

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

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

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

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

$$\max_{\alpha > 0} \frac{-\alpha^2 + (\overline{r} - 1)\alpha}{\alpha + 1} = \left(\sqrt{\overline{r}} - 1\right)^2$$

and

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

Inserting this value of α into (4.6) leads to

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

Therefore,

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

Thus, $\left(\sqrt{\overline{r}} - 1\right)^2$ is a lower bound on the energy gain, which is our 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 $\{\overline{x}_0, \{\overline{u}_i, \overline{v}_i\}_{i=0}^N\}$ such that

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

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

V. RLS Adaptive Filtering

We are now in a position to specialize the result of Theorem 1 to the case of adaptive filtering with no parameter drift. The corresponding model is given by

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

where d_i is the observation, $h_i^T = [\, h_{i1} \quad h_{i2} \quad \ldots \quad h_{in} \,]$ is a known $1 \times n$ input vector, w is an unknown weight vector, and v_i is an unknown disturbance signal. We should note that it is also possible to consider more general models, e.g., ones for which the weight vector w drifts with time, but this would require more space than is permitted here.

The above model can be considered a state-space model with the parameters,

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

The RLS algorithm is essentially the corresponding Kalman filter. 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:

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

where $k_{p,i} = P_i h_i / (1 + h_i^T P_i h_i)$, 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.

[2]This can be done by choosing all e_i's as zero, except for the one that achieves $\inf_i \underline{\sigma}(R_i^{*/2} R_{e,i}^{-1} R_i^{1/2})$, which we take as being the corresponding singular vector.

As before, we define the prediction and filtered estimation errors as $e_{p,i} = h_i^T w - h_i^T \hat{w}_{|i-1}$ 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:

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

and

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

where we have defined

$$\overline{r} = \sup_i [1 + h_i P_i h_i^T] \quad \text{and} \quad \underline{r} = \inf_i [1 + h_i P_i h_i^T]. \tag{5.8}$$

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 $h_i^T h_i = \overline{h}^2$, then,

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

Remark: Corollary 1 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 $\sqrt{\mu}$. The lower bound, in fact, is quite nontrivial. 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.

VI. Conclusion

In this note, 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 (i.e., 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.

Finally, we should mention that robustness is only one desirable aspect of an estimator, and, thus, comparisons between various estimators should not be made based on robustness alone. In this regard, it would be useful to obtain H^∞ bounds for the Kalman filter for more general estimation problems, i.e., when the desired signal is any arbitrary linear combination of the state, and not just the uncorrupted output. The techniques employed in this note may prove to be useful in this more general setting as well.

References

[1] B. Hassibi, A. H. Sayed, and T. Kailath, "H^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 267–280, Feb. 1996.

[2] B. Hassibi and T. Kailath, "H^∞ filtering of signals in additive noise," IEEE Trans. Automat. Contr., 1997, submitted for publication.

[3] B. Hassibi, A. H. Sayed, and T. Kailath, "Linear estimation in Krein spaces—Part I: Theory," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 18–33, Jan. 1996.

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

On Formal Power Series Representations for Uncertain Systems

Carolyn Beck

***Abstract*—The concept of minimality as developed for uncertain and multidimensional systems represented by linear fractional transformations (LFTs) is related to realization theory results for formal power series. We discuss the relationship between the notions of minimality for LFT and series realizations, and present a method for obtaining one type of minimal realization from the opposing type. An extension of an existing minimality result for formal power series to the multi-input–multi-output (MIMO) case is also presented.**

***Index Terms*—Formal power series, minimality, uncertain and multidimensional systems.**

I. Introduction

In this note, we relate the notion of minimality developed for linear fractional transformation (LFT) representations of uncertain systems and certain classes of multidimensional systems, discussed fully in [1]–[3], to realization theory results for formal power series (FPS). LFTs on structured sets provide a convenient and general framework for representing and manipulating models of not only uncertain and

Manuscript received October 12, 1999; revised July 10, 2000. Recommended by Associate Editor T. Iwasaki. This work was supported in part by the National Science Foundationunder Grant ECS9733043, and in part by the Army Research Office and the Electric Power Research Institute under Grant DAAG55-98-R-RA08.

The author is with the Department of General Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: beck3@uiuc.edu).

Publisher Item Identifier S 0018-9286(01)01024-8.