Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

H^∞-Optimality of H^2 Predictors[1]

Babak Hassibi[2] and Thomas Kailath

Information Systems Laboratory
Stanford University, Stanford CA 94305

Abstract

Given past observations of a process, $\{y_j, j < i\}$, suppose we are interested in constructing one-step-ahead predictors of y_i, denoted by $\hat{y}_{i|i-1}$. We show that, irrespective of whether the process $\{y_j\}$ is stationary or non-stationary, or whether it is scalar- or vector-valued, the H^2-optimal one-step-ahead predictor is also H^∞-optimal. Since the H^2 and H^∞ paradigms represent fundamentally different approaches to estimation and control, the estimators and controllers obtained from each formalism have often drastically different performances with respect to the other criterion. Our result, however, provides a non-trivial example of when the two formalisms lead to the same optimal design.

1 Introduction

The H^2 and H^∞ paradigms in estimation and control represent two extremes, both in terms of their requirements on the exogenous signals, and in terms of their objectives: one is stochastic, assumes statistical knowledge of the exogenous signals, and optimizes average performance, whereas the other is deterministic, makes no statistical assumptions on the signals, and optimizes the worst-case performance. For this reason, the estimators and controllers obtained from these two formalisms have often drastically different performances when measured with respect to the other criterion. This is especially true in control, where H^∞ theory was first developed to address the question of robustness, which could not be satisfactorily dealt with in the H^2 framework [1, 2], and, for example, in adaptive filtering, where the H^2- and H^∞-optimal solutions (RLS and LMS) are quite different [3].

Consider now the following one-step-ahead prediction problem: given the possibly vector-valued process $\{y_j\}$, construct estimates of y_i, using past observations of the process $\{y_j, j < i\}$, which we denote by $\hat{y}_{i|i-1}$.

The H^2 Case: Here we assume that the second-order statistics of $\{y_j\}$ are known. The mean-values can, without loss of generality, be taken to be zero, which means that in the stationary case, we know the covariance function $R_y(i) = Ey_j y^*_{j-i}$, or, equivalently, we know the z-spectral density

$$S_y(z) = \mathcal{Z}\{R_y(i)\} = \sum_{j=-\infty}^{\infty} R_y(j) z^{-j},$$

and that, in the nonstationary case, we know the two-dimensional covariance function

$$R_y(i,j) = Ey_i y^*_j.$$

In either case, the H^2 criterion requires us to minimize the mean-square prediction error:

$$E(y_i - \hat{y}_{i|i-1})^*(y_i - \hat{y}_{i|i-1}). \tag{1}$$

Now it is wellknown that H^2-optimal predictors result in a white prediction error process known as the *innovations* process, $e_i = y_i - \hat{y}_{i|i-1}$ (see *e.g.*, [4]). Thus, in the stationary case,

$$Ee_j e^*_{j-i} = R_e(i) = R_e\delta_i \quad \text{or} \quad S_e(z) = R_e$$

and, in the nonstationary case,

$$Ee_i e^*_j = R_e(i,j) = R_{e,i}\delta_{ij}.$$

The H^∞ Case: Here we assume that the process $\{y_j\}$ is generated by passing an *unknown* input sequence $\{u_j\}$ through a *known* causal linear system. In the stationary case, the known linear system is time-invariant, and can be represented by its transfer matrix, $H(z)$. In the nonstationary case, the known linear system is time-variant .In either case, the objective in the H^∞ approach is to minimize the worst-case energy gain from the unknown sequence $\{u_j\}$ to the prediction error sequence $\{y_j - \hat{y}_{j|j-1}\}$, *i.e.*, to minimize

$$\sup_{\{u_j\} \in l^2 - \{0\}} \frac{\|y - \hat{y}\|_2^2}{\|u\|_2^2}, \tag{2}$$

where $\|a\|_2^2 \triangleq \sum_{j=-\infty}^{\infty} a_j^* a_j$, and l^2 denotes the space of square-summable sequences.

[1] 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.

0-7803-4394-8/98 $10.00 © 1998 IEEE

The maximum energy gain in (2) can be regarded as the (squared) l^2-induced norm (or the (squared) H^∞ norm) of the transfer operator, $\mathcal{T}_\mathcal{K}$, that maps the unknown input $\{u_j\}$ to the prediction error sequence $\{y_j - \hat{y}_{j|j-1}\}$. In the time-invariant case, $\mathcal{T}_\mathcal{K}$ has a transfer matrix representation, $T_K(z)$, and

$$\sup_{\{u_j\}\in l^2-\{0\}} \frac{\|y-\hat{y}\|_2^2}{\|u\|_2^2} = \sup_{\omega\in[0,2\pi]} \bar{\sigma}\left(T_K(e^{j\omega})T_K^*(e^{j\omega})\right),$$

whereas in the time-invariant case $\mathcal{T}_\mathcal{K}$ can be represented by a (block) lower triangular matrix, and

$$\sup_{\{u_j\}\in l^2-\{0\}} \frac{\|y-\hat{y}\|_2^2}{\|u\|_2^2} = \bar{\sigma}\left(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*\right),$$

where $\bar{\sigma}(\cdot)$ represents the maximum singular value.

The H^2 and H^∞ approaches can be related in the following fashion: rather than assuming that the the process $\{y_j\}$ is a zero-mean stochastic process with known covariance function, we can equivalently assume that $\{y_j\}$ is generated by passing an unknown zero-mean unit variance process $\{u_j\}$ through a *known* causal linear system. In this framework the known covariance function of $\{y_j\}$ is determined by the linear system that generates $\{y_j\}$ from $\{u_j\}$. The mean-square-error (1), in the stationary case, is then given by

$$\frac{1}{2\pi}\int_0^{2\pi} \text{trace}\left[T_K(e^{j\omega})T_K^*(e^{j\omega})\right] d\omega,$$

and, in the (finite-horizon) non-stationary case,

$$E\left[\sum_{i=0}^{N}(y_i-\hat{y}_{i|i-1})^*(y_i-\hat{y}_{i|i-1})\right] = \text{trace}\left(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*\right).$$

The fact that H^2-optimal one-step predictors result in white prediction errors, means that the resulting H^2-optimal transfer operator $\mathcal{T}_{\mathcal{K}_2}\mathcal{T}_{\mathcal{K}_2}^*$ is (block) diagonal. Thus, in the stationary case,

$$T_{K_2}(e^{j\omega})T_{K_2}^*(e^{j\omega}) = R_e,$$

and in the (finite-horizon) non-stationary case,

$$\mathcal{T}_{\mathcal{K}_2}\mathcal{T}_{\mathcal{K}_2}^* = R_{e,0}\oplus R_{e,1}\oplus\ldots\oplus R_{e,N}.$$

This, of course, implies that, in the stationary case, the (squared) H^∞ norm of the H^2-optimal predictor is

$$\sup_{\omega\in[0,2\pi]} \bar{\sigma}\left(T_{K_2}(e^{j\omega})T_{K_2}^*(e^{j\omega})\right) = \bar{\sigma}\left(R_e\right),$$

since the error-spectrum $T_{K_2}(e^{j\omega})T_{K_2}^*(e^{j\omega})$ is frequency independent, and in the non-stationary case,

$$\bar{\sigma}\left(\mathcal{T}_{\mathcal{K}_2}\mathcal{T}_{\mathcal{K}_2}^*\right) = \max_{i=0,\ldots N} \bar{\sigma}\left(R_{e,i}\right),$$

since the error covariance $\mathcal{T}_{\mathcal{K}_2}\mathcal{T}_{\mathcal{K}_2}^*$ is (block) diagonal.

Scalar Stationary Processes: When $\{y_j\}$ is a scalar stationary process, the error spectrum $T_{K_2}(e^{j\omega})T_{K_2}^*(e^{j\omega}) = R_e$ is *flat*, since R_e is a scalar. Further reflection on this fact shows that the H^2-optimal predictor for a scalar stationary process must also be H^∞-optimal. To this end, suppose that it is not H^∞-optimal. Then there exists a predictor for which

$$\sup_{\omega\in[0,2\pi]} T_K(e^{j\omega})T_K^*(e^{j\omega}) < R_e.$$

But this clearly violates the H^2-optimality, since

$$\begin{aligned}\int_0^{2\pi} T_K(e^{j\omega})T_K^*(e^{j\omega})d\omega &< \int_0^{2\pi} R_e d\omega \\ &= \int_0^{2\pi} T_{K_2}(e^{j\omega})T_{K_2}^*(e^{j\omega})d\omega.\end{aligned}$$

[More intuitively, in the scalar case H^∞-optimal predictors minimize the *peak value* of $|T_K(e^{j\omega})|^2$, whereas H^2-optimal predictors minimize the *area under the curve* of $|T_K(e^{j\omega})|^2$. Clearly, if $|T_{K_2}(e^{j\omega})|^2$ is flat it is not possible to reduce its peak without further reducing its area under the curve.]

Vector-Valued and Nonstationary Processes: Unfortunately, the above simple argument cannot be extended to such processes. In the vector-valued case, since R_e is a matrix, rather than a scalar, it is not clear whether it possible to reduce $\sup_{\omega\in[0,2\pi]}\bar{\sigma}\left(T_K(e^{j\omega})\right)$ beyond $\bar{\sigma}(R_e)$, without having to reduce $\frac{1}{2\pi}\int_0^{2\pi}\text{trace}\left[T_K(e^{j\omega})T_K^*(e^{j\omega})\right]d\omega$ beyond $\frac{1}{2\pi}\int_0^{2\pi}\text{trace}\left[R_e\right]d\omega$.

In the non-stationary case the situation is even more confounded, since we must check whether it is possible to reduce $\bar{\sigma}\left(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*\right)$ beyond $\max_{i=0,\ldots N}\bar{\sigma}\left(R_{e,i}\right)$ without having to reduce trace $\left(\mathcal{T}_\mathcal{K}\mathcal{T}_\mathcal{K}^*\right)$ beyond $\sum_{i=0}^N \text{trace}[R_{e,i}]$.

Thus the question of whether the H^∞-optimality of H^2 predictors extends to such more general processes still remains. In the next sections we shall answer this question in the affirmative. However, to do so, we will need to delve more deeply into H^∞ theory, and especially into the structure of H^∞-optimal solutions [5, 6, 7].

2 The Stationary Case

A general estimation problem, in the stationary case, is shown in Fig. 2, where $\{u_i \in \mathcal{C}^m\}$ is an unknown input sequence, $\{v_i \in \mathcal{C}^p\}$ is an unknown additive disturbance sequence, $\{y_i \in \mathcal{C}^p\}$ is a known measurement sequence, and $H(z)$ and $L(z)$ are known causal and stable linear time-invariant systems. The goal is to construct the linear time-invariant $K(z)$ (called the estimator) to estimate the unobservable desired sequence $\{s_i \in \mathcal{C}^q\}$ from the observations $\{y_i\}$. The estimates are denoted by $\{\hat{s}_i\}$.

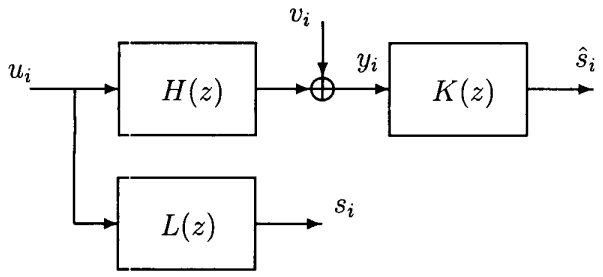

Figure 1: A stationary estimation problem.

The behavior of any estimator $K(z)$ can be captured by the induced transfer matrix, say $T_K(z)$, that maps the unknown disturbances $\{u_i\}$ and $\{R^{-1/2}v_i\}$, where $R = R^{1/2}R^{*/2} > 0$ is a weighting matrix, to the estimation errors $\{s_i - \hat{s}_i\}$. Now using Fig. 2, we readily see

$$T_K(z) = \begin{bmatrix} L(z) - K(z)H(z) & -K(z)R^{1/2} \end{bmatrix}.$$

In H^∞ estimation the goal is to choose $K(z)$ to minimize the H^∞ norm of $T_K(z)$.

Problem 1 (Optimal H^∞ Estimation Problem) *Find a causal estimator $K(z)$ that achieves*

$$\inf_{K(z)} \|T_K(z)\|_\infty \triangleq \gamma_{opt}. \tag{3}$$

To give an expression for γ_{opt} we need to introduce some notation. To this end, note that the causal and stable $H(z)$ and $L(z)$ have the following Laurent series expansions, analytic on and outside the unit circle:

$$H(z) = \sum_{j=0}^{\infty} H_j z^{-j} \quad \text{and} \quad L(z) = \sum_{j=0}^{\infty} L_j z^{-j}.$$

Now the input-output mappings

$$o_i = \sum_{j=0}^{\infty} H_j u_{i-j} \quad \text{and} \quad s_i = \sum_{j=0}^{\infty} L_j u_{i-j}$$

can be written in matrix notation as

$$\underbrace{\begin{bmatrix} \vdots \\ o_{-1} \\ o_0 \\ o_1 \\ \vdots \end{bmatrix}}_{o} = \underbrace{\begin{bmatrix} \ddots & & & & \\ \ddots & H_0 & & & \\ \dots & H_1 & H_0 & & \\ \dots & H_2 & H_1 & H_0 & \\ & \vdots & \vdots & \ddots & \ddots \end{bmatrix}}_{\triangleq \mathcal{H}} \underbrace{\begin{bmatrix} \vdots \\ u_{-1} \\ u_0 \\ u_1 \\ \vdots \end{bmatrix}}_{u},$$

with a similar expression $s = \mathcal{L}u$. If we further partition the input and output sequences $\{u_j\}$ and $\{o_j\}$ into the their past, $u_- \triangleq \{u_i, i < 0\}$ and $o_- \triangleq \{y_i, i < 0\}$, and present and future, $u_+ \triangleq \{u_i, i \geq 0\}$ and $o_+ \triangleq \{y_i, i \geq 0\}$, components, the operator $\mathcal{H}$ is also partitioned as follows:

$$\mathcal{H} = \left[\begin{array}{c|c} \mathcal{H}_- & 0 \\ \hline \mathcal{H}_H & \mathcal{H}_+ \end{array}\right],$$

where, for example,

$$\mathcal{H}_- \triangleq \begin{bmatrix} \ddots & & & \\ \ddots & H_0 & & \\ \dots & H_1 & H_0 & \\ \dots & H_2 & H_1 & H_0 \end{bmatrix}.$$

Similarly, also for $\mathcal{L}$:

$$\mathcal{L} = \left[\begin{array}{c|c} \mathcal{L}_- & 0 \\ \hline \mathcal{L}_H & \mathcal{L}_+ \end{array}\right]$$

The operators $\mathcal{H}_-$ and $\mathcal{L}_-$ map past inputs to past outputs, and are called *Toeplitz* operators, whereas the original doubly-infinite operators $\mathcal{H}$ and $\mathcal{L}$ are referred to as *Laurent* operators.

With these definitions, we have the following result.

Theorem 1 (Optimal H^∞ Norm) *The optimal H^∞ norm for Problem 1 is given by*

$$\begin{aligned} \gamma_{opt}^2 &= \bar{\sigma}\left(\mathcal{L}_-(I + \mathcal{H}_-^* \mathcal{R}_-^{-1} \mathcal{H}_-)^{-1} \mathcal{L}_-^*\right) \\ &= \bar{\sigma}\left(\mathcal{L}_-\mathcal{L}_-^* - \mathcal{L}_-\mathcal{H}_-^*(\mathcal{R}_- + \mathcal{H}_-\mathcal{H}_-^*)^{-1}\mathcal{H}_-\mathcal{L}_-^*\right), \end{aligned} \tag{4}$$

where we have defined $\mathcal{R}_- = \ldots \oplus R \oplus R$.

Proof: For a proof, see [7, 8]. ■

Note that formula (4) gives an expression for γ_{opt} in terms of the maximum singular value (or maximum spectral radius) of a certain combination of the Toeplitz operators, $\mathcal{H}_-$ and $\mathcal{L}_-$. However, it can be shown that the operator $\mathcal{L}_-(I + \mathcal{H}_-^* \mathcal{R}_-^{-1} \mathcal{H}_-)^{-1} \mathcal{L}_-^*$ is *not* in general Toeplitz, so that, except for some special cases, explicit frequency domain formulae for γ_{opt} cannot be found. However, expression (4) will indeed allow us to demonstrate that H^2 predictors are H^∞-optimal.

Finally, we should remark that the second of the two formulae in (4) does not require the invertibility of R, and so will be the main formula used in our proof.

2.1 One-Step-Ahead Prediction

The one-step-ahead prediction problem, in the stationary case, is depicted in Fig. 2.1. Here $H(z)$ is a $p \times m$ causal and stable transfer matrix that generates the

process $\{y_j\}$. The transfer matrix $z^{-1}H(z)$ clearly generates $\{y_{j-1}\}$. Thus the prediction problem becomes a special case of the estimation problem (3) with

$$L(z) \to H(z) \quad , \quad H(z) \to z^{-1}H(z)$$

and with $R = 0$, since there is no additive disturbance $\{v_i\}$ and so no penalty on the second block of $T_K(z)$.

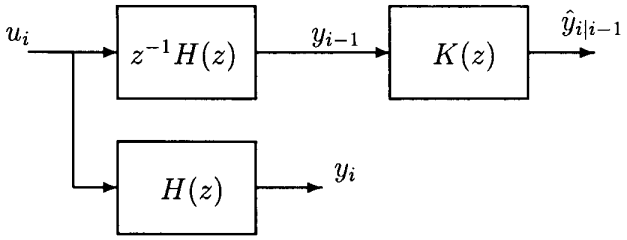

Figure 2: The stationary prediction problem.

Thus in the operator domain:

$$\mathcal{L}_- \to \mathcal{H}_- \quad \text{and} \quad \mathcal{H}_- \to \mathcal{Z}_p\mathcal{H}_-$$

where $\mathcal{Z}_p$ is the block lower triangular shift matrix,

$$\mathcal{Z}_p \triangleq \begin{bmatrix} \ddots & & & \\ \ddots & 0 & & \\ \cdots & I_p & 0 & \\ \cdots & 0 & I_p & 0 \end{bmatrix}.$$

Moreover, we should note that $\mathcal{Z}_p$ "commutes" with lower triangular Topelitz operators in the following sense:

$$\mathcal{Z}_p\mathcal{H}_- = \mathcal{H}_-\mathcal{Z}_m.$$

Finally, we will also use the relations

$$\mathcal{Z}_p\mathcal{Z}_p^* = I \quad \text{and} \quad \mathcal{Z}_p^*\mathcal{Z}_p = \begin{bmatrix} I & 0 \\ 0 & 0_{p\times p} \end{bmatrix}.$$

We thus have the following result, which clearly shows the H^∞-optimality of H^2 predictors for stationary processes.

Theorem 2 (H^∞-Optimal Predictor) *The optimal H^∞ norm for the problem*

$$\inf_{\text{causal } K(\cdot)} \left\| \begin{bmatrix} H(z) - K(z)z^{-1}H(z) & 0 \end{bmatrix} \right\|_\infty \triangleq \gamma_{opt}$$

is given by

$$\gamma_{opt}^2 = \bar{\sigma}\left(R_e\right), \tag{5}$$

where R_e, the "innovations variance" is found from the canonical spectral factorization,

$$H(z)H^*(z^{-*}) = M(z)R_eM^*(z^{-*}), \tag{6}$$

with $M(z)$ monic ($M(\infty) = I_p$) and $M(z)$ and $M^{-1}(z)$ analytic in $|z| \geq 1$.

Proof: The most straightforward proof is to compute γ_{opt} using (4). Since here $R = 0$, we shall use the second formula. Thus,

$$\begin{aligned} \gamma_{opt}^2 &= \bar{\sigma}\left[\mathcal{H}_-\mathcal{H}_-^* - \mathcal{H}_-\mathcal{H}_-^*\mathcal{Z}_p^*(0 + \mathcal{Z}_p\mathcal{H}_-\mathcal{H}_-^*\mathcal{Z}_p)^{-1}\mathcal{Z}_p\mathcal{H}_-\mathcal{H}_-^*\right] \\ &= \bar{\sigma}\left[\mathcal{H}_-\mathcal{H}_-^* - \mathcal{H}_-\mathcal{H}_-^*\mathcal{Z}_p^*\left(\mathcal{H}_-\mathcal{Z}_m\mathcal{Z}_m^*\mathcal{H}_-^*\right)^{-1}\mathcal{Z}_p^*\mathcal{H}_-\mathcal{H}_-^*\right] \\ &= \bar{\sigma}\left[\mathcal{H}_-\mathcal{H}_-^* - \mathcal{H}_-\mathcal{H}_-^*\mathcal{Z}_p^*\left(\mathcal{H}_-\mathcal{H}_-^*\right)^{-1}\mathcal{Z}_p^*\mathcal{H}_-\mathcal{H}_-^*\right]. \end{aligned}$$

To compute $(\mathcal{H}_-\mathcal{H}_-^*)^{-1}$, let us write the canonical spectral factorization (6) in operator form as

$$\left[\begin{array}{c|c} \mathcal{M}_- & 0 \\ \hline \mathcal{M}_H & \mathcal{M}_+ \end{array}\right]\left[\begin{array}{c|c} \mathcal{R}_{e-} & 0 \\ \hline 0 & \mathcal{R}_{e+} \end{array}\right]\left[\begin{array}{c|c} \mathcal{M}_-^* & \mathcal{M}_H^* \\ \hline 0 & \mathcal{M}_+^* \end{array}\right] = \left[\begin{array}{c|c} \mathcal{H}_- & 0 \\ \hline \mathcal{H}_H & \mathcal{H}_+ \end{array}\right]\left[\begin{array}{c|c} \mathcal{H}_-^* & \mathcal{H}_H^* \\ \hline 0 & \mathcal{H}_+^* \end{array}\right],$$

so that

$$\mathcal{M}_-\mathcal{R}_{e-}\mathcal{M}_-^* = \mathcal{H}_-\mathcal{H}_-^*.$$

Moreover, $\mathcal{M}_-$ is invertible, since $\mathcal{M}$ is causally invertible and

$$\mathcal{M}^{-1} = \left[\begin{array}{c|c} \mathcal{M}_-^{-1} & 0 \\ \hline -\mathcal{M}_+^{-1}\mathcal{M}_H\mathcal{M}_-^{-1} & \mathcal{M}_+^{-1} \end{array}\right].$$

Thus, defining $\Delta_- = \mathcal{M}_-\mathcal{R}_{e-}^{1/2}$, so that Δ_- is invertible and $\Delta_-\Delta_-^* = \mathcal{H}_-\mathcal{H}_-^*$, we may write

$$\begin{aligned} \gamma_{opt}^2 &= \bar{\sigma}\left[\Delta_-\Delta_-^* - \Delta_-\Delta_-^*\mathcal{Z}_p^*(\Delta_-\Delta_-^*)^{-1}\mathcal{Z}_p\Delta_-\Delta_-^*\right] \\ &= \bar{\sigma}\left[\Delta_-\Delta_-^* - \Delta_-\mathcal{Z}_p^*\Delta_-^*\Delta_-^{-*}\Delta_-^{-1}\Delta_-\mathcal{Z}_p\Delta_-^*\right] \\ &= \bar{\sigma}\left[\Delta_-\Delta_-^* - \Delta_-\mathcal{Z}_p^*\mathcal{Z}_p\Delta_-^*\right] \\ &= \bar{\sigma}\left[\Delta_-(I - \mathcal{Z}_p^*\mathcal{Z}_p)\Delta_-^*\right] \\ &= \bar{\sigma}\left[\Delta_-\begin{bmatrix} 0 & 0 \\ 0 & I_p \end{bmatrix}\Delta_-^*\right] \\ &= \bar{\sigma}\begin{bmatrix} 0 & 0 \\ 0 & \Delta_0\Delta_0^* \end{bmatrix} \\ &= \bar{\sigma}\left(\Delta_0\Delta_0^*\right), \end{aligned}$$

where

$$\Delta_- = \begin{bmatrix} \ddots & & & \\ \ddots & \Delta_0 & & \\ \cdots & \Delta_1 & \Delta_0 & \\ \cdots & \Delta_2 & \Delta_1 & \Delta_0 \end{bmatrix}.$$

But $\Delta(z) = M(z)R_e^{1/2}$, so that

$$\left(\Delta_0 + \Delta_1 z^{-1} + \ldots\right) = \left(I + M_1 z^{-1} + \ldots\right) R_e^{1/2},$$

since $M(z)$ is monic. Thus $\Delta_0 = R_e^{1/2}$ and

$$\gamma_{opt}^2 = \bar{\sigma}\left(R_e\right),$$

which is the desired result.

■

3 The Non-Stationary Case

A general estimation problem, in the finite-horizon nonstationary case, is shown in Fig. 3. Here, once again, $\{u_i \in \mathcal{C}^m\}_{i=0}^N$ is an unknown input sequence, $\{v_i \in \mathcal{C}^p\}_{i=0}^N$ is an unknown additive disturbance sequence, $\{y_i \in \mathcal{C}^p\}_{i=0}^N$ is a known measurement sequence, and $\mathcal{H}$ and $\mathcal{L}$ are known causal linear time-variant systems, that map their respective inputs to outputs according to the rules

$$\underbrace{\begin{bmatrix} o_0 \\ o_1 \\ \vdots \\ o_N \end{bmatrix}}_{o} = \underbrace{\begin{bmatrix} H_{0,0} & & & \\ H_{1,0} & H_{1,1} & & \\ \vdots & \ddots & \ddots & \\ H_{N,0} & \dots & H_{N,N-1} & H_{N,N} \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{bmatrix}}_{u},$$

and, similarly, $s = \mathcal{L}y$. The goal is to construct a causal linear time-variant estimator, $\mathcal{K}$, to estimate the unobservable desired sequence $\{s_i \in \mathcal{C}^q\}_{i=0}^N$ from the observations $\{y_i\}$. The estimates are denoted by $\{\hat{s}_i\}$.

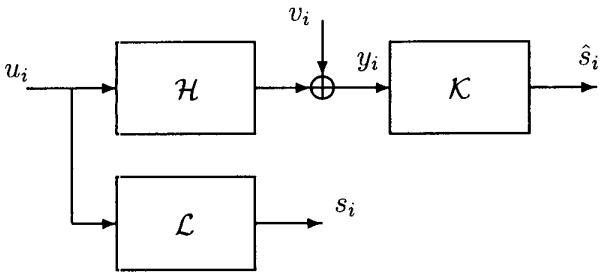

Figure 3: A non-stationary estimation problem.

As in the stationary case, the behavior of any estimator $\mathcal{K}$ can be captured by the induced transfer operator, say $\mathcal{T}_{\mathcal{K}}$, that maps the unknown disturbances $\{u_i\}$ and $\{R_i^{-1/2}v_i\}$, where $R_i = R_i^{1/2}R_i^{*/2} > 0$ is a weighting matrix, to the estimation errors $\{s_i - \hat{s}_i\}$. Using Fig. 3, it is straightforward to see that

$$\mathcal{T}_{\mathcal{K}} = \begin{bmatrix} \mathcal{L} - \mathcal{K}\mathcal{H} & -\mathcal{K}\mathcal{R} \end{bmatrix},$$

where we have defined $\mathcal{R} = R_0 \oplus R_1 \oplus \ldots \oplus R_N$. We thus have the following problem.

Problem 2 (Nonstationary H^∞ Estimation) *Find a causal estimator $\mathcal{K}$ that achieves*

$$\inf_{\text{causal } \mathcal{K}} \bar{\sigma}(\mathcal{T}_{\mathcal{K}}) \triangleq \gamma_{opt}. \tag{7}$$

In order to give an expression for γ_{opt}, consider the following leading submatrices of $\mathcal{H}$:

$$\mathcal{H}_{-}^{(i)} \triangleq \begin{bmatrix} H_{0,0} & & & \\ H_{1,0} & H_{1,1} & & \\ \vdots & \ddots & \ddots & \\ H_{i,0} & \dots & H_{i,i-1} & H_{i,i} \end{bmatrix},$$

and similarly $\mathcal{L}_{-}^{(i)}$ of $\mathcal{L}$, for $i = 0, 1, \ldots N$. Note that $\mathcal{H}_{-}^{(N)} = \mathcal{H}$ and $\mathcal{L}_{-}^{(N)} = \mathcal{L}$.

Theorem 3 (Optimal H^∞ Norm) *The optimal H^∞ norm for Problem 2 is given by*

$$\begin{aligned} \gamma_{opt}^2 &= \max_{i=0,1,\ldots N} \bar{\sigma}\left(\mathcal{L}_{-}^{(i)}\left(I + \mathcal{H}_{-}^{(i)*}(\mathcal{R}_{-}^{(i)})^{-1}\mathcal{H}_{-}^{(i)}\right)^{-1}\mathcal{L}_{-}^{(i)*}\right) \\ &= \max_{i=0,1,\ldots N} \bar{\sigma}\Big(\mathcal{L}_{-}^{(i)}\mathcal{L}_{-}^{(i)*} - \\ &\qquad \mathcal{L}_{-}^{(i)}\mathcal{H}_{-}^{(i)*}(\mathcal{R}_{-}^{(i)} + \mathcal{H}_{-}^{(i)}\mathcal{H}_{-}^{(i)*})^{-1}\mathcal{H}_{-}^{(i)}\mathcal{L}_{-}^{(i)*}\Big), \end{aligned} \tag{8}$$

where $\mathcal{R}_{-}^{(i)} = R_0 \oplus R_1 \oplus \ldots \oplus R_i$.

Proof: For the proof see [7, 8]. ■

Note that the above result is the time-variant counterpart of Theorem 1, since in the stationary case $\mathcal{L}$ and $\mathcal{H}$ are doubly-infinite Toeplitz matrices so that all their leading submatrices, at any time i, are equal to $\mathcal{L}_{-}$ and $\mathcal{H}_{-}$. Therefore instead of computing the maximum singular values for various time instants, as in (8), we need only compute it once, as in (4).

3.1 One-Step-Ahead Prediction

The one-step-ahead prediction problem, in the nonstationary case, is depicted in Fig. 3.1. Here $\mathcal{H}$ is a causal linear system that generates the process $\{y_j\}$. The linear system $\mathcal{Z}_p\mathcal{H}$ generates $\{y_{j-1}\}$, since $\mathcal{Z}_p$ denotes the lower-triangular shift matrix,

$$\mathcal{Z}_p \triangleq \begin{bmatrix} 0 & & & \\ I_p & 0 & & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & I_p & 0 \end{bmatrix}.$$

Thus the prediction problem becomes a special case of the estimation problem (7) with

$$\mathcal{L} \to \mathcal{H} \quad , \quad \mathcal{H} \to \mathcal{Z}_p\mathcal{H}$$

and with $\mathcal{R} = 0$.

Thus, the submatrices in Theorem 3 become:

$$\mathcal{L}_{-}^{(i)} \to \mathcal{H}_{-}^{(i)} \quad \text{and} \quad \mathcal{H}_{-}^{(i)} \to \mathcal{Z}_p^{(i)}\mathcal{H}_{-}^{(i)}.$$

Moreover, we should note that $\mathcal{Z}_p^{(i)}$ "commutes" with lower triangular Toeplitz matrices in the following sense:

$$\mathcal{Z}_p^{(i)}\mathcal{H}_{-}^{(i)} = \mathcal{H}_{-}^{(i)}\mathcal{Z}_m^{(i)}.$$

Finally, we will also use the relations

$$\mathcal{Z}_p^{(i)}\mathcal{Z}_p^{(i)*} = I \quad \text{and} \quad \mathcal{Z}_p^{(i)*}\mathcal{Z}_p^{(i)} = \begin{bmatrix} I & 0 \\ 0 & 0_{p\times p} \end{bmatrix}.$$

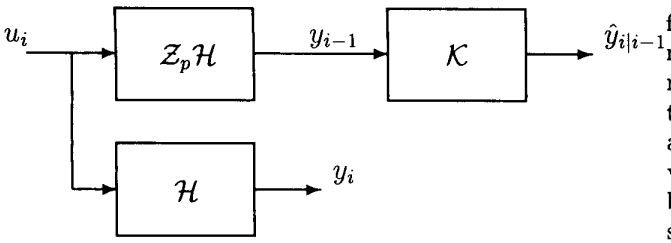

Figure 4: The nonstationary prediction problem.

We thus have the following result, which clearly shows the H^∞-optimality of H^2 predictors for nonstationary processes.

Theorem 4 (Nonstationary H^∞-Optimal Predictor) *The optimal H^∞ norm for the problem*

$$\inf_{\text{causal } \mathcal{K}} \bar{\sigma}\left(\begin{bmatrix} \mathcal{H} - \mathcal{K}\mathcal{Z}_p\mathcal{H} & 0 \end{bmatrix}\right) \triangleq \gamma_{opt} \qquad (9)$$

is given by

$$\gamma_{opt}^2 = \max_{i=0,1,\ldots N} \bar{\sigma}\left(R_{e_i}\right), \qquad (10)$$

where $R_e = R_{e,0} \oplus R_{e,1} \oplus \ldots \oplus R_{e,N}$, the "innovations variance", is found from the block LDU (lower-diagonal-upper) triangular factorization,

$$\mathcal{H}\mathcal{H}^* = \mathcal{M}R_e\mathcal{M}^*,$$

where $\mathcal{M}$ is a block lower triangular matrix with unit diagonal.

Proof: The proof is essentially identical to the proof of Theorem 2 and so is omitted for reasons of space. ■

4 Conclusion

We showed that H^2- and H^∞-optimal one-step-ahead predictors coincide. This fact is rather straightforward in the case of predicting a scalar stationary process. However, for vector-valued stationary processes and for nonstationary processes, the result is less obvious and the proof requires further tools from H^∞ theory that have been developed in [7]. [See also [5, 6].]

Since there are fundamental differences between the philosophies of the H^2 and H^∞ approaches to estimation and control, any estimator or controller obtained from one formalism has often drastically different performance with respect to the other criterion. However, our result gives a non-trivial example of when the two formalisms lead to the same optimal design. This represents, in a sense, the "best of both worlds", since the resulting predictors are optimal both from a stochastic (and hence expected-value) point of view, as well as from a deterministic (and hence worst-case) point of view. These results have also implications for problems beyond one-step-ahead prediction, and in [9] we have shown that for the important communications problem of decision-feedback equalization the H^2- and H^∞-optimal solutions are the same.

Finaly, we should mention that the above result is *not* true for predictors that predict more than one unit ahead in time. Nor is it true for predicting processes *related* to the observations. Consider for example the process,

$$y_i = s_i + v_i,$$

where $\{s_j\}$ and $\{v_j\}$ are independent (from both a stochastic and deterministic point of view) processes. Then, although the H^2- and H^∞-optimal predictors $\hat{y}_{i|i-1}$, of y_i given $\{y_j, j < i\}$ coincide, the H^2- and H^∞-optimal predictors $\hat{s}_{i|i-1}$, of s_i given $\{y_j, j < i\}$, are in general different.

References

[1] M. Green and D.J.N. Limebeer. *Linear Robust Control.* Prentice-Hall, Englewood Cliffs, NJ, 1995.

[2] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control.* Prentice Hall, Englewood Cliffs, N.J., 1996.

[3] 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.

[4] T. Kailath, A.H. Sayed, and B. Hassibi. *State Space Estimation.* Prentice-Hall, Englewood Cliffs, NJ, 1998.

[5] M. Verma and E.A. Jonckheere. l^∞ compensation with mixed sensitivity as a broadband matching problem. *Systems and Control Letters*, 4:125–129, 1984.

[6] A. Feintuch and B.A. Francis. Uniformly optimal control of linear systems. *Automatica*, 21:563–574, 1985.

[7] B. Hassibi and T. Kailath. On optimal solutions to two-block H^∞ problems. In *In the 1998 American Control Conference*, Philadelphia, PA, June 1998.

[8] B. Hassibi, A.H. Sayed, and T. Kailath. *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories.* SIAM Studies in Applied Mathematics, New York, 1998.

[9] B. Hassibi, A.E. Erdogan, and T. Kailath. Decision-feedback equalization with an H^∞ criterion. *Submitted to IEEE Transactions on Information Theory*, 1998.