Error Analysis and Numerical Stabilization of the Fast $H_\infty$ Filter

Tomonori KATSUMATA†, Kiyoshi NISHIYAMA(α), Members, and Katsuaki SATOH††, Nonmember

SUMMARY  The fast $H_\infty$ filter is developed by one of the authors, and its practical use in industries is expected. This paper derives a linear propagation model of numerical errors in the recursive variables of the fast $H_\infty$ filter, and then theoretically analyzes the stability of the filter. Based on the analyzed results, a numerical stabilization method of the fast $H_\infty$ filter is proposed with the error feedback control in the backward prediction. Also, the effectiveness of the stabilization method is verified using numerical examples.

key words: fast $H_\infty$ filter, error analysis, numerical stabilization, system identification

1. Introduction

System identification is an experimental approach to the modeling of an unknown system, and is of great importance in the fields of signal processing, communication, and automatic control [1]. In many cases, the unknown system is modeled as an adaptive filter whose finite impulse response (FIR) are adjusted with $N$ taps so that its output will perfectly match that of the unknown system except for observation noise. The most celebrated adaptive schemes for the FIR filter are the least mean square (LMS) and recursive least squares (RLS) algorithms [1]–[6]. The performance of such adaptive algorithms can be quantified by examining a number of characteristics, such as the a) accuracy of the obtained solution, b) convergence speed, c) tracking ability, d) computational complexity, and e) robustness to round off error accumulation [3]. Although the LMS algorithm is very simple, with a computational complexity of $O(N)$, and is relatively robust to numerical errors, it suffers from slow convergence, especially when applied to highly correlated excitation signals such as speech. In contrast, the RLS algorithm with $O(N^2)$ provides very fast convergence regardless of the characteristics of the input signals. Usually, the RLS algorithm and its fast versions use a forgetting factor $\rho$ to improve the tracking performance for time-varying systems. However, the value of $\rho$ has been typically determined empirically, without any evidence of optimality. In our previous work, this open problem is solved using the framework of $H_\infty$ optimization [7], [8]. The resultant $H_\infty$ filter, the hyper $H_\infty$ filter, enables the forgetting factor $\rho$ to be optimally determined by means of the so-called $\gamma$-iteration. Expectably, the hyper $H_\infty$ filter has possessed the robustness to disturbances (especially near-end speech) and the extremely high convergence speed. The hyper $H_\infty$ filter is applicable to the estimation problem of an infinite impulse response (IIR) filter as well as an FIR filter. However, the IIR filter often becomes numerically unstable when implemented with a finite precision. So, the FIR filter is widely used in practical applications, especially in communication and acoustic systems. Furthermore, the FIR filter is much favorable to deriving a fast algorithm of the adaptive filter which requires a shifting property for the input covariance matrix. Indeed, a fast algorithm of the hyper $H_\infty$ filter for FIR systems, the fast $H_\infty$ filter, has also been successfully developed, providing a computational complexity of $O(N)$. However, the numerical behavior of the fast $H_\infty$ filter has not been analyzed yet.

In this paper, we derive a linear propagation model of numerical errors in the recursive variables of the fast $H_\infty$ filter, and then theoretically analyze the stability of the filter. Based on the analyzed results, we propose a numerical stabilization method of the fast $H_\infty$ filter with the error feedback control in the backward prediction. Also, the effectiveness of the stabilization method is verified using numerical examples.

The remainder of this paper is organized as follows. Section 2 presents the hyper $H_\infty$ filter. Section 3 shortly explains the fast $H_\infty$ filter. In Sect. 4, an analysis of numerical errors appeared in the fast $H_\infty$ filter is given. Based on the results of Sect. 4, we propose a numerical stabilization method of the fast $H_\infty$ filter in Sect. 5. In Sect. 6, numerical examples are given to verify the effectiveness of the proposed method. Finally, we present our conclusions in Sect. 7.

2. The Hyper $H_\infty$ Filter

The hyper $H_\infty$ filter, which was originally proposed in [7], [8], can simultaneously optimize both the forgetting factor and the robustness to disturbances through the so-called $\gamma$-iteration. This filter, which differs from the ordinary $H_\infty$ Filters [9], is effective for tracking time-varying systems with unknown dynamics of the state vector $x_k$ [10], and furthermore a fast algorithm of the filter is successfully derived as seen in the next section.

Theorem 1: (The Hyper $H_\infty$ Filter) For the following $N$-
dimensional state-space model with system noise \( w_k \) and observation noise \( v_k \):

\[
x_{k+1} = x_k + G_k w_k, \quad w_k, x_k \in \mathcal{R}^N (1)
\]

\[
y_k = H_k x_k + v_k, \quad y_k, v_k \in \mathcal{R} (2)
\]

\[
z_k = H_k x_k, \quad z_k \in \mathcal{R}, \quad H_k \in \mathcal{R}^{1 \times N} (3)
\]

one possible level-\( \gamma_f \) hyper \( H_\infty \) filter to achieve

\[
\sup_{x_0} \left( \sum_{i=0}^{k} ||e_{f,i}||^2 / \rho \right) < \gamma_f^2
\]

\[
||x_0 - \hat{x}_0||^2 + \sum_{i=0}^{k} ||w_i||^2 + \sum_{i=0}^{k} ||v_i||^2 / \rho < \gamma_f^2
\]

is represented by

\[
\hat{z}_{ik} = H_k \hat{x}_{ik}
\]

\[
\hat{x}_{ik} = \hat{x}_{i-1,k} + K_{i,k}(y_k - H_k \hat{x}_{i-1,k})
\]

\[
K_{i,k} = \Sigma_{ik-1} - \Sigma_{ik-1} R_k^{-1} C_k \Sigma_{ik-1}
\]

\[
\Sigma_{k+1} = \Sigma_{ik}/\rho
\]

where \( ^T \) denotes the transpose, \( \cdot^{-1} \) the inverse, and

\[
e_{f,i} = \hat{z}_{i0} - H_k x_{i0}, \quad \hat{z}_{i0} = \hat{x}_0, \quad \Sigma_{i0} = \Sigma_0
\]

\[
R_{c,k} = R + C_k \Sigma_{ik-1} C_k^T
\]

\[
R = \begin{bmatrix} \rho & 0 \\ 0 & -\rho \gamma_f^2 \end{bmatrix}, \quad C_k = \begin{bmatrix} H_k \\ H_k \end{bmatrix}
\]

0 < \( \rho = 1 - \chi(\gamma_f) \leq 1 \), \( \gamma_f > 1 \)

in which \( \chi(\gamma_f) \) is a monotonically decreasing scalar function of \( \gamma_f \) such that \( \chi(1) = 1 \) and \( \chi(\infty) = 0 \), and the driving matrix \( G_k \) is generated by

\[
G_k G_k^T = \chi(\gamma_f) \Sigma_{k+1,k} = \chi(\gamma_f) \Sigma_{ik,k}.
\]

For the filter to exist, the parameter \( \gamma_f \) must satisfy

\[
\Sigma_{ik} = \Sigma_{ik-1} + 1 - \gamma_f^2 / \rho H_k^T H_k > 0, \quad i = 0, \ldots, k.
\]

**Proof**: See [7], [8].

If \( \gamma_f \) is chosen to be as small as the existence condition allows, the \( H_\infty \) filter with the optimal \( G_k \) (or \( \rho \)) will be able to very quickly track the changed state vector \( x_k \). This is because the sum of the squared error \( ||e_{f,i}||^2 \) must be subject to the inequality of (4), which prevents the maximum energy gain from growing to a value in excess of \( \gamma_f^2 \).

Note that the algorithm of the hyper \( H_\infty \) filter in the limit of \( \gamma_f = \infty \) coincides with that of the Kalman filter with \( G_k = 0 \).

### 3. The Fast \( H_\infty \) Filter

The computational complexity of the hyper \( H_\infty \) filter is \( O(N^2) \), making it difficult to implement this filter in real-time applications. To overcome this difficulty, a fast \( H_\infty \) filter, whose complexity is \( O(N) \), has been derived for an FIR system with input \( u_k \) using a shifting property of the observation matrix such that \( H_{k+1} = [u_{k+1}, H_k(1), H_k(2), \ldots, H_k(N-1)] \) [7], [8].

**Theorem 2**: (The Fast \( H_\infty \) Filter) When the observation matrix \( H_k \) has the shifting property, the hyper \( H_\infty \) filter for the \( N \)-dimensional state-space model (1)–(3) with \( \Sigma_{i0} = \varepsilon_0 I, \varepsilon_0 > 0 \) can be recursively performed at a computational complexity of \( O(N) \) per iteration as

\[
\hat{x}_{ik} = \hat{x}_{i-1,k} + \hat{K}_{i,k}(y_k - H_k \hat{x}_{i-1,k})
\]

\[
\hat{K}_{i,k} = \frac{K_i(:,1)}{1 + \gamma_f^2 H_k \hat{K}_i(:,1)} \in \mathcal{R}^{N \times 1}
\]

where the gain matrix \( K_i \) is recursively calculated as

\[
K_k = m_k - \frac{D_k \mu_k}{1 - \mu_k W_k}
\]

\[
D_k = \frac{D_{k-1} - m_k W_k}{1 - \mu_k W_k}
\]

\[
\eta_k = c_k - N + C_k D_{k-1}
\]

\[
S_k = \rho S_{k-1} + \mu_k \hat{e}_{k-1}^T \hat{W}_k \hat{e}_k, \quad \hat{e}_k = c_k + C_k - A_k
\]

\[
\hat{e}_k = c_k + C_k - A_k n_{k-1}
\]

where \( C_k = \begin{bmatrix} H_k \\ H_k \end{bmatrix}, \quad W = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \rho \in \mathcal{R}^{2 \times 1} \)

and \( \epsilon_k \in \mathcal{R}^{2 \times 1} \) is the first row of \( C_k = [c_k, \ldots, c_k - N + 1] \), assuming that \( c_k - i = 0 \) for \( i < 0 \). Also, \( \rho \) and \( \gamma_f \) are chosen as \( 0 < \rho \leq 1 - \chi(\gamma_f) \leq 1 \), and \( \gamma_f > 1 \). The recursions are initialized with \( K_0 = 0_{N \times 2}, \quad A_0 = 0, \quad S_0 = \frac{1}{\epsilon_0}, \quad D_0 = 0, \quad \hat{e}_{00} = 0 \) where \( \varepsilon_0 \) is set to be a relatively large positive number, \( 0 \) denotes the zero vector, and \( 0_{n \times m} \) the \( n \times m \) zero matrix.

For the filter achieving (4) to exist, the value of \( \gamma_f \) must be chosen so as to satisfy the following scalar existence condition:

\[
-\rho \hat{e}_i + \rho \gamma_f^2 > 0, \quad i = 0, \ldots, k
\]

where \( \gamma_f \) and \( \hat{e}_i \) are defined by

\[
\rho = 1 - \gamma_f^2, \quad \hat{e}_i = \frac{\rho H_i K_{si}}{1 - H_i K_{ti}}
\]

respectively.

**Proof**: The key idea to derivation of the fast \( H_\infty \) filter is based on the following equations:

\[
\hat{Q}_k K_k = C_k^T, \quad \hat{Q}_k := \sum_{i=0}^{k} \rho^{k-i} C_i^T W C_i
\]

\[
\hat{Q}_k = C_k^T, \quad \hat{Q}_k := \sum_{i=0}^{k} \rho^{k-i} C_i^T W C_i
\]
where

\[ \mathbf{\tilde{C}}_k := [c_k, \ldots, c_{k-N}] = [C_k, c_{k-N}] \]  

(21)

The details of the derivation are described in [7] and [8]. □

4. Error Analysis of the Fast H∞ Filter

When the gain matrix \( \mathbf{K}_k \) is initialized with \( \mathbf{K}_0 = \mathbf{0}_{N \times 2} \), the numerical stability of the fast H∞ filter is mainly governed by the following recursive variables of the algorithm:

\[
\begin{align*}
\mathbf{A}_k &= \mathbf{A}_{k-1} - \mathbf{K}_{k-1} \mathbf{W}_k \hat{e}_k \\
\mathbf{S}_k &= \rho \mathbf{S}_{k-1} + \mathbf{e}_k^T \mathbf{W}_k \mathbf{e}_k \\
\mathbf{D}_k &= (\mathbf{D}_{k-1} - m_k \mathbf{W}_k \eta_k)(1 - \mu_k \mathbf{W}_k \eta_k)^{-1}
\end{align*}
\]

(22)

where

\[
\begin{align*}
\mathbf{K}_k &= \mathbf{m}_k - \mathbf{D}_k \mathbf{\mu}_k \\
\mathbf{m}_k &= \left[ \begin{array}{c} S_k^{-1} \mathbf{e}_k^T \\ \mathbf{K}_{k-1} + \mathbf{A}_k S_k^{-1} \mathbf{e}_k^T \end{array} \right] \\
\mathbf{\epsilon}_k &= \mathbf{c}_k + \mathbf{C}_{k-1} \mathbf{A}_{k-1} \mathbf{e}_k, \quad \mathbf{\epsilon}_k^* = \mathbf{\epsilon}_k + \mathbf{C}_{k-1} \mathbf{A}_k \\
\eta_k &= \mathbf{c}_{k-N} + \mathbf{C}_k \mathbf{D}_{k-1}.
\end{align*}
\]

(23)

Then, the recursions of the forward linear prediction coefficient \( \mathbf{A}_k \), the power of the forward prediction error \( \mathbf{S}_k \), and the backward linear prediction coefficient \( \mathbf{D}_k \) can be represented by the following nonlinear system:

\[
\begin{align*}
\mathbf{\Theta}_k &= \mathbf{f}_k(\mathbf{\Theta}_{k-1}), \quad \mathbf{\Theta}_k := \begin{bmatrix} \mathbf{A}_k \\ \mathbf{S}_k \\ \mathbf{D}_k \end{bmatrix}
\end{align*}
\]

Linearizing the nonlinear system, we obtain a propagation model of numerical errors \( \mathbf{d}\mathbf{\Theta}_k \) of \( \mathbf{\Theta}_k \) as

\[
\begin{align*}
\mathbf{dA}_k &= \mathbf{\hat{A}}_k - \mathbf{A}_k, \quad \mathbf{dS}_k := \mathbf{\hat{S}}_k - \mathbf{S}_k \\
\mathbf{dD}_k &= \mathbf{\hat{D}}_k - \mathbf{D}_k \\
\mathbf{\hat{F}}_k := \frac{\partial \mathbf{f}_k}{\partial \mathbf{\Theta}_{k-1}} &= \begin{bmatrix} \mathbf{I} - \mathbf{K}_{k-1} \mathbf{W} \mathbf{C}_{k-1} & \mathbf{0} & \mathbf{0}_{N \times N} \\
\mathbf{F}_{k}^{31} & \mathbf{F}_{k}^{32} & \mathbf{F}_{k}^{33} \end{bmatrix} \\
\mathbf{F}_{k}^{33} &= \frac{\mathbf{I} - \mathbf{m}_k \mathbf{W} \eta_k}{\beta_k} \\
\beta_k &= 1 - \mu_k \mathbf{W} \eta_k
\end{align*}
\]

(24)

(25)

(26)

(27)

(28)

Assuming the \( \{c_k\} \) is statistically stationary, i.e., \( \mathbf{Q}_{k-1} \approx \mathbf{Q}_{k-2} \), the 1-1 block matrix \( \mathbf{F}_k, \mathbf{F}_k^{11} \), is approximated using

\[
\mathbf{Q}_k = \rho \mathbf{Q}_{k-1} + \mathbf{C}_k^T \mathbf{W} \mathbf{C}_k
\]

\[
\begin{align*}
\mathbf{I} - \mathbf{K}_{k-1} \mathbf{W} \mathbf{C}_{k-1} &= \mathbf{I} - \mathbf{Q}_{k-1}^T \mathbf{Q}_{k-1} - \rho \mathbf{Q}_{k-2} \\
&= \mathbf{I} - \mathbf{Q}_{k-1}^T (\mathbf{Q}_{k-1} - \rho \mathbf{Q}_{k-2}) \\
& \approx \rho \mathbf{I}.
\end{align*}
\]

(29)

Namely, the eigenvalues of \( \mathbf{F}_k^{11} \) are almost less than one when \( \rho < 1 \). This implies that the recursion of \( \mathbf{dA}_k \) is stable. Similarly, we obtain

\[
\mathbf{dS}_k = \rho \mathbf{dS}_{k-1} + \mathbf{F}_{k}^{31} \mathbf{dA}_{k-1} \]

(30)

which is stable because of \( \rho < 1 \) and the stability of \( \mathbf{dA}_k \). Therefore, the recursive variable \( \mathbf{S}_k \) is also numerically stable.

However, the updating of \( \mathbf{dD}_k \) is recursively performed by

\[
\mathbf{dD}_k = \mathbf{F}_{k}^{33} \mathbf{dD}_{k-1} + \mathbf{F}_{k}^{32} \mathbf{dS}_{k-1} + \mathbf{F}_{k}^{31} \mathbf{dA}_{k-1}
\]

(31)

whose dynamics is given by

\[
\begin{align*}
\mathbf{F}_{k}^{33} &= \frac{1}{\beta_k} \left\{ \left( \mathbf{I} - \mathbf{m}_k \mathbf{W} \eta_k \right) \mathbf{\mu}_k \mathbf{W} \eta_k \right\} \\
&= \frac{1}{\beta_k} \left\{ \left( \mathbf{I} + (\mathbf{D}_k \mathbf{\mu}_k - \mathbf{m}_k \mathbf{W} \eta_k) \mathbf{\eta}_k \mathbf{W} \eta_k \right) \right\} \\
&= \frac{\mathbf{I} - \mathbf{K}_k \mathbf{W} \mathbf{C}_k}{\beta_k} \\
& \approx \frac{\rho}{\beta_k} \mathbf{I}.
\end{align*}
\]

(32)

The last equation means that if \( \beta_k > \rho \) for any \( k \), the backward prediction is stable. However, the stable condition may not be guaranteed under finite precision. This is the reason why the fast H∞ filter causes the numerical instability.

5. Numerical Stabilization of the Fast H∞ Filter

5.1 Preliminaries

Definition 1: The forward and backward prediction errors are newly defined by

\[
\hat{\mathbf{e}}_k := \mathbf{c}_k + \mathbf{C}_{k-1} \mathbf{A}_{k-1}, \quad \hat{\mathbf{e}}_k := \mathbf{c}_k, \quad \tilde{\mathbf{e}}_k := \mathbf{c}_k - \mathbf{C}_k \mathbf{D}_{k-1}.
\]

(33)

Similarly, the forward and backward estimation errors and their powers are defined by

\[
\mathbf{e}_k := \mathbf{c}_k + \mathbf{C}_{k-1} \mathbf{A}_k, \quad \mathbf{\epsilon}_k := \mathbf{c}_{k-N} + \mathbf{C}_k \mathbf{D}_k
\]

(34)

and

\[
\begin{align*}
\mathbf{\tilde{F}}_k := \sum_{i=0}^{k} \rho^{k-i} \mathbf{e}_i^T \mathbf{W} \mathbf{e}_i, \quad \mathbf{e}_i := \mathbf{c}_i + \mathbf{C}_{i-1} \mathbf{A}_k
\end{align*}
\]
Proposition 1: The backward estimation error \( \mathbf{e}_k \) is written with its own power \( \mathbf{S}_k \) as
\[
\mathbf{e}_k = \mathbf{S}_k \mathbf{\mu}_k^T.
\]

Proof: The inverse of \( \mathbf{Q}_k \) partitioned with \( \mathbf{\bar{C}}_j = [C, c_{r-N}] \) can be expressed with the inverses of its diagonal blocks as
\[
\mathbf{Q}_k^{-1} = \begin{bmatrix}
\mathbf{I} & \mathbf{0} \\
\mathbf{0} & (\mathbf{q}_k^T - \mathbf{L}_k^T \mathbf{Q}_k^{-1} \mathbf{L}_k)^{-1}
\end{bmatrix}
\]

using the matrix inversion for partitioned matrices lemma. Here, \( \mathbf{L}_k \) and \( \mathbf{q}_k \) are defined, respectively, by
\[
\mathbf{L}_k := \sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_{i-N}, \quad \mathbf{q}_k := \sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_{i-N}.
\]

On the other hand, minimizing \( \mathbf{S}_k \) with respect to \( \mathbf{D}_k \) provides
\[
\sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_{i-N} \mathbf{D}_k = - \sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_{i-N} \mathbf{Q}_k \mathbf{D}_k = -\mathbf{L}_k.
\]

Then, since the power of the backward estimation error is expressed as
\[
\mathbf{S}_k = \mathbf{q}_k^T + 2 \mathbf{D}_k^T \mathbf{L}_k + \mathbf{D}_k^T \mathbf{Q}_k \mathbf{D}_k = \mathbf{q}_k^T + \mathbf{D}_k^T \mathbf{L}_k
\]
\[
= \mathbf{q}_k^T + \mathbf{L}_k \mathbf{Q}_k^{-1} \mathbf{L}_k
\]
we can rewrite (37) as
\[
\mathbf{Q}_k^{-1} \mathbf{C}_k^T = \begin{bmatrix}
\mathbf{Q}_k^{-1} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}
\end{bmatrix}
\]
\[
\mathbf{Q}_k^{-1} = \begin{bmatrix}
\mathbf{I} & \mathbf{0} \\
\mathbf{0} & (\mathbf{q}_k^T - \mathbf{L}_k^T \mathbf{Q}_k^{-1} \mathbf{L}_k)^{-1}
\end{bmatrix}
\]
\[
\mathbf{K}_k = \begin{bmatrix}
\mathbf{K}_k & \mathbf{D}_k \\
\mathbf{0} & \mathbf{1}
\end{bmatrix}
\]
whose transpose completes the proof of Proposition 1. \( \square \)

Proposition 2: The powers of the forward and backward estimation errors are expressed by
\[
\mathbf{S}_k = \frac{\det(\mathbf{Q}_k)}{\det(\mathbf{Q}_{k-1})}, \quad \mathbf{S}_k = \frac{\det(\mathbf{Q}_k)}{\det(\mathbf{Q}_{k-1})}
\]
where \( \det(\cdot) \) denotes the determinant of matrix.

Proof: Taking the determinant of both sides of the following identity equation for \( \mathbf{\bar{Q}}_k = \begin{bmatrix}
\mathbf{q}_k^T & \mathbf{t}_k^T \\
\mathbf{t}_k & \mathbf{Q}_{k-1}
\end{bmatrix} \) partitioned with \( \mathbf{\bar{C}}_j = [C, c_{r-N}] \):
\[
\begin{bmatrix}
\mathbf{q}_k^T & \mathbf{t}_k^T \\
\mathbf{t}_k & \mathbf{Q}_{k-1}
\end{bmatrix} = \begin{bmatrix}
\mathbf{I} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}
\end{bmatrix}
\]
\[
\begin{bmatrix}
\mathbf{q}_k^T & \mathbf{t}_k^T \\
\mathbf{t}_k & \mathbf{Q}_{k-1}
\end{bmatrix} = \begin{bmatrix}
\mathbf{q}_k^T & \mathbf{t}_k^T \\
\mathbf{t}_k & \mathbf{Q}_{k-1}
\end{bmatrix}
\]
we have
\[
\det(\mathbf{Q}_k) \det(\mathbf{Q}_{k-1}) = \mathbf{q}_k^T - \mathbf{t}_k^T \mathbf{Q}_{k-1} \mathbf{t}_k
\]
in which \( \mathbf{t}_k \) and \( \mathbf{q}_k^T \) are defined, respectively, by
\[
\mathbf{t}_k := \sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_i, \quad \mathbf{q}_k^T := \sum_{i=0}^{k} \mathbf{\rho}^k \mathbf{c}_i^T \mathbf{W}_i \mathbf{c}_i.
\]
Here the properties of determinant, \( \det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}) \),
\[
\text{det}\left(\begin{bmatrix}
A & 0 \\
0 & B
\end{bmatrix}\right) = \text{det}(A) \text{det}(B),
\]
are used for the calculation of (46).

On the other hand, when \( S_k \) is minimized by \( A_k \), we have
\[
Q_{k-1}A_k = -t_k \quad (A_k = -Q_{k-1}^{-1}t_k). \tag{48}
\]
Substituting (48) to the right-hand side of (46) provides
\[
det(\tilde{Q}_k) \text{det}(Q_{k-1}^{-1}) = q_k^f + t_k^T A_k. \tag{49}
\]
Furthermore, using
\[
S_k = \sum_{i=0}^{k} \rho^{k-i} (c_i + C_{i-1} A_k)^T W (c_i + C_{i-1} A_k)
\]
\[
= \sum_{i=0}^{k} \rho^{k-i} c_i^T W c_i + 2A_k^T \sum_{i=0}^{k} \rho^{k-i} C_{i-1}^T W c_i
\]
\[+ A_k^T \left( \sum_{i=0}^{k} \rho^{k-i} C_{i-1}^T W C_{i-1} \right) A_k
\]
\[= \sum_{i=0}^{k} \rho^{k-i} c_i^T W c_i + 2A_k^T t_k + A_k^T Q_{k-1} A_k
\]
\[= q_k^f + t_k^T A_k \tag{50}
\]
we can rewrite (49) as
\[
det(\tilde{Q}_k) \text{det}(Q_{k-1}^{-1}) = S_k. \tag{51}
\]
Then, recalling \( \det(A^{-1}) = 1/\det(A) \), we have
\[
S_k = \frac{\text{det}(\tilde{Q}_k)}{\text{det}(Q_{k-1})}. \tag{52}
\]
Similarly, the power of backward estimation error is given by
\[
S^w = \frac{\text{det}(\tilde{Q}_k)}{\text{det}(Q_k)}. \tag{53}
\]

**Proposition 3:** The backward prediction error \( e_k \) is expressed with the backward estimation error \( f_k \) as
\[
e_k = 1 - \frac{\gamma_f^2}{\rho} \zeta_k f_k \tag{54}
\]
where
\[
\zeta_k = H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2). \tag{55}
\]

**Proof:** The backward linear prediction coefficient \( D_k \) minimizing \( S^w \) satisfies
\[
Q_k D_k = -\bar{L}_k. \tag{56}
\]
Hence, substituting the following equations:
\[
Q_k = \rho Q_{k-1} + C_k^T W C_k
\]
to (56) and using
\[
K_k := \rho^{-1} Q_k^{-1} C_k^T \tag{58}
\]
we have
\[
(\rho Q_{k-1} + C_k^T W C_k) D_k = -\rho t_{k-1} - C_k^T W c_{k-N}
\]
\[
\rho Q_{k-1} D_k + C_k^T W (c_{k-N} + C_k D_k) = -\rho t_{k-1}
\]
\[
\rho Q_{k-1} (D_k + \rho^{-1} Q_k^{-1} C_k^T W c_{k-N}) = -\rho t_{k-1}
\]
\[
Q_{k-1} (D_k + \bar{K}_k W f_{k-N}) = -\bar{L}_{k-1}. \tag{59}
\]
Then, recalling \( Q_{k-1} D_k = -\bar{L}_{k-1} \), we find
\[
D_k = D_{k-1} - \bar{K}_k W f_{k-N}. \tag{60}
\]
Similarly, arranging (56) at time \( k - 1 \) as
\[
(\rho^{-1} Q_k - \rho^{-1} C_k^T W c_{k-N}) D_{k-1}
\]
\[
= -\rho^{-1} L_k - \rho^{-1} C_k^T W c_{k-N}
\]
\[
\rho^{-1} Q_{k-1} D_{k-1} - \rho^{-1} C_k^T W (c_{k-N} + C_k D_{k-1})
\]
\[
= -\rho^{-1} t_{k-1}
\]
\[
Q_{k-1} (D_{k-1} - C_k^T W f_{k-N}) = -\bar{L}_{k-1}
\]
and comparing the last equation with (56), we obtain
\[
D_k = D_{k-1} - \bar{K}_k W f_{k-N}, \quad \bar{K}_k = Q_k^{-1} C_k^T. \tag{62}
\]
On the other hand, the recursion of \( P_k^{-1} := Q_k \):
\[
P_k^{-1} = \rho P_{k-1}^{-1} + C_k^T W C_k
\]
\[
= \rho P_{k-1}^{-1} + (1 - \gamma_f^2) H_k^T H_k
\]
allows the Riccati equation in the hyper \( H_{\infty} \) filter to reduce to
\[
H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2)
\]
\[
= \left( H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2) \right) H_k P_{k-1} + \rho/(1 - \gamma_f^2)
\]
\[
\zeta_k = H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2). \tag{63}
\]

Multiplying both sides of (63) by \( H_k^T \), we have the following relationship between \( K_k (\cdot, 1) \) and \( \bar{K}_k (\cdot, 1) \):
\[
K_k (\cdot, 1) = K_k (\cdot, 1) \begin{bmatrix}
1 & -H_k P_{k-1} H_k^T \\
H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2)
\end{bmatrix}
\]
\[
= \bar{K}_k (\cdot, 1) \rho/(1 - \gamma_f^2) \tag{64}
\]

\( L_k = \rho L_{k-1} + C_k^T W c_{k-N} \)
where
\[ \zeta_k = H_k P_{k-1} H_k^T + \rho/(1 - \gamma_f^2). \] (65)

Besides, comparing (60) with (62) in terms of \( D_k - D_{k-1} \), we find
\[ \tilde{K}_k W \tilde{e}_k = K_k W \tilde{e}_k. \] (66)

Then, substituting (64) into (66) provides
\[ \tilde{K}_k W \tilde{e}_k = \rho/(1 - \gamma_f^2) \tilde{K}_k W \tilde{e}_k. \] (67)

From this, through some calculations, we can obtain
\[ \tilde{\epsilon}_k = \frac{(1 - \gamma_f^2) \zeta_k}{\rho}. \] (68)

Note that \( K_k(\cdot, 1) = K_k(\cdot, 2), \ K_k(\cdot, 1) = K_k(\cdot, 2), \ \tilde{e}_k(1, 1) = \tilde{e}_k(2, 1) \), and \( \tilde{\epsilon}_k(1, 1) = \tilde{\epsilon}_k(2, 1) \).

**Proposition 4:** The backward prediction error \( \tilde{\epsilon}_k \) is expressed with the power of the forward estimation error \( \bar{S}_k \)
as
\[ \tilde{\epsilon}_k = \rho^{-N} \bar{S}_k \mu_k^T. \] (69)

**Proof:** From Proposition 1, the backward estimation error \( \epsilon_\kappa \) can be written as
\[ \epsilon_\kappa = \bar{S}_k \mu_k^T. \] (70)

Starting with this, we derive an alternative expression of the backward prediction error \( \tilde{\epsilon}_k \).

First, we arrange the Riccati equation of (63) as
\[ P_k = \left( P_{k-1} - P_{k-1} H_k^T H_k \right) + \frac{\rho}{1 - \gamma_f^2} H_k P_{k-1} \right)/\rho \]
\[ = \left( P_{k-1} - P_{k-1} H_k^T H_k \right)/\rho \] (71)

where
\[ \zeta_k = H_k P_{k-1} H_k^T + \frac{\rho}{1 - \gamma_f^2}. \] (72)

Multiplying both sides of (71) by \( P_{k-1}^{-1} \) from the right-hand side provides
\[ P_k P_{k-1}^{-1} = \left( I - \frac{P_{k-1} H_k^T H_k}{\zeta_k} \right)/\rho. \] (73)

Next, taking the determinant of both sides of (73), from \( \det(AB) = \det(A) \det(B) \), \( \det(I + AB) = \det(I + BA) \), and (72), we have
\[ \det(P_k) \det(P_{k-1}^{-1}) = \det\left\{ I - \frac{P_{k-1} H_k^T H_k}{\zeta_k} \right\} \rho^{-N} \]
\[ = \det\left\{ 1 - \frac{H_k P_{k-1} H_k^T}{\zeta_k} \right\} \rho^{-N} \]
\[ = \det\left\{ 1 - \frac{\rho}{1 - \gamma_f^2} \right\} \rho^{-N} \]
\[ = \rho^{-N} \det\left\{ \frac{\rho}{\zeta_k(1 - \gamma_f^2)} \right\} \rho^{-N} \]
\[ = \rho^{-N} \det\left\{ \frac{\rho}{\zeta_k(1 - \gamma_f^2)} \right\} = \rho^{-N+1}, \] (74)

The last equation of (74) leads to the following relationship:
\[ \frac{(1 - \gamma_f^2)}{\rho^{-N+1}} \zeta_k = \frac{\det(Q_k')}{\det(Q_k^{-1})} = \frac{\det(Q_k)}{\det(Q_k^{-1})}. \] (75)

On the other hand, recalling Proposition 2, namely
\[ \bar{S}_k = \frac{\det(Q_k)}{\det(Q_k^{-1})}, \ S_k = \frac{\det(Q_k)}{\det(Q_k^{-1})}. \] (76)

we can express (75) as
\[ \frac{(1 - \gamma_f^2)}{\rho^{-N+1}} \zeta_k = \frac{\bar{S}_k}{S_k}. \] (77)

Then, arranging (77) for \( S_k \), we have
\[ S_k = \frac{\rho^{-N+1}}{(1 - \gamma_f^2)} \bar{S}_k. \] (78)

Last, using (78) with Propositions 1 and 3, we obtain an alternative expression of the backward prediction error \( \tilde{\epsilon}_k \) as
\[ \tilde{\epsilon}_k = \frac{1 - \gamma_f^2}{\rho} \tilde{\epsilon}_k = \frac{1 - \gamma_f^2}{\rho} \bar{S}_k \mu_k^T \]
\[ = \frac{1 - \gamma_f^2}{\rho} \frac{\rho^{-N+1}}{(1 - \gamma_f^2)} \bar{S}_k \mu_k^T = \rho^{-N} \bar{S}_k \mu_k^T. \]

**Proposition 5:** The minimum forward estimation error power \( \bar{S}_k \) is equal to \( S_k \) in (16).

**Proof:** Recalling (50), we obtain
\[ \bar{S}_k = \sum_{i=0}^{k} \rho^{i} \epsilon_i^T W \epsilon_i + A_i^T t_k. \] (79)

Then, using (16) and (48), we can arrange (79) as
\[ \bar{S}_k = \sum_{i=0}^{k} \rho^{i} \epsilon_i^T W \epsilon_i + (A_{k-i} - K_{k-i} W \tilde{e}_k)^T t_k \]
\[ = \sum_{i=0}^{k} \rho^{i} \epsilon_i^T W \epsilon_i + c_k^T W c_k + \rho A_{k-1}^T t_{k-1} \]
\[ + A_{k-1}^T C_{k-1} W c_k - \tilde{e}_k^T W K_{k-1} t_k \]
\[ = \sum_{i=0}^{k} \rho^{i} \epsilon_i^T W \epsilon_i + (A_{k-i} - K_{k-i} W \tilde{e}_k)^T t_k \]
\[ = \sum_{i=0}^{k} \rho^{i} \epsilon_i^T W \epsilon_i + (A_{k-i} - K_{k-i} W \tilde{e}_k)^T t_k \]
\[ J(\hat{D}_{k-1}) \]
\[ = \sum_{i=0}^{M} \rho^{i}[((c_{i-N} + c_{i-1})\hat{D}_{k-1}) - (c_{i-N} + c_{i-1}D_{k-1})]^T \]
\[ \times W((c_{i-N} + c_{i-1}\hat{D}_{k-1}) - (c_{i-N} + c_{i-1}D_{k-1})) \]
\[ + \kappa((c_{i-N} + c_{i-1}\hat{D}_{k-1}) - \rho^{-1}S_{k}^{T}e_{k-1})^T \]
\[ \times W((c_{i-N} + c_{i-1}\hat{D}_{k-1}) - \rho^{-1}S_{k}^{T}e_{k-1}) \]
\[ = (D_{k-1} - \hat{D}_{k-1})^T Q_{k}(D_{k-1} - \hat{D}_{k-1}) \]
\[ + \kappa\tilde{e}_{k}^{T} W \tilde{e}_{k}, \quad \kappa \leq 0 \] (81)

where
\[ \tilde{e}_{k} := \hat{e}_{k} - \rho^{-1}S_{k}e_{k-1}^{T}, \quad \hat{e}_{k} := c_{k-N} + c_{k}\hat{D}_{k-1}. \] (82)

Setting the derivative of \( J(\hat{D}_{k-1}) \) to 0 gives
\[ \frac{\partial J(\hat{D}_{k-1})}{\partial \hat{D}_{k-1}} = 2Q_{k}(D_{k-1} - \hat{D}_{k-1}) + 2\kappa C_{k}^{T} W \tilde{e}_{k} \]
\[ = 0 \] (83)

which leads to
\[ \hat{D}_{k-1} = D_{k-1} - \kappa Q_{k}^{-1} C_{k}^{T} W \tilde{e}_{k} \]
\[ = D_{k-1} - \kappa K_{k} W \tilde{e}_{k}. \] (84)

Then, replacing \( D_{k-1} \) in (62) with the corrected \( \hat{D}_{k-1} \) and using (84), we can also correct \( D_{k} \) as
\[ D_{k} = D_{k-1} - K_{k} W \hat{e}_{k} \]
\[ = (D_{k-1} - \kappa K_{k} W \tilde{e}_{k}) - K_{k} W \hat{e}_{k} \]
\[ = D_{k-1} - K_{k} W (\hat{e}_{k} + \kappa \tilde{e}_{k}) \]
\[ = D_{k-1} - K_{k} W (\eta_{k} + \kappa (\hat{e}_{k} - \rho^{-1}S_{k}^{T}e_{k-1})). \] (85)

Note that (84), nonexplicit form of \( \hat{D}_{k-1} \), works well to concentrate the corrections in the second term of the right-hand side of the last equation in (85), and \( \tilde{e}_{k} \) when \( D_{k-1} \) is replaced with \( \hat{D}_{k-1} \) is equal to \( \hat{e}_{k} \).

For simplicity, assuming that \( \hat{e}_{k} \approx \eta_{k} \), i.e., the required correction is small, we obtain
\[ D_{k} = D_{k-1} - K_{k} W \hat{e}_{k} \] (86)

from (85) where
\[ \hat{e}_{k} = \eta_{k} + \kappa (\eta_{k} - \rho^{-1}S_{k}^{T}e_{k-1}). \] (87)

Furthermore, substituting \( K_{k} = m_{k} - D_{k} \mu_{k} \) to (86), we obtain a new feasible update equation of \( D_{k} \) as
\[ D_{k} = D_{k-1} - (m_{k} - D_{k} \mu_{k}) W \hat{e}_{k} \]
\[ D_{k} = D_{k-1} - m_{k} W \hat{e}_{k} + D_{k} \mu_{k} W \hat{e}_{k} \]
\[ D_{k}(1 - \mu_{k} W \hat{e}_{k}) = D_{k-1} - m_{k} W \hat{e}_{k} \]
\[ D_{k} = \frac{D_{k-1} - m_{k} W \hat{e}_{k}}{1 - \mu_{k} W \hat{e}_{k}}. \] (88)

Namely, the numerical stabilization of the recursive variable \( D_{k} \) is accomplished by merely replacing \( \eta_{k} \) in (22) with \( \hat{e}_{k} \). This \( \hat{e}_{k} \) includes the backward prediction errors computed in two different ways, \( \eta_{k} \) and \( \rho^{-1}S_{k}^{T}e_{k-1} \), as
\[ \hat{e}_{k} = \eta_{k} + \kappa (\eta_{k} - \rho^{-1}S_{k}^{T}e_{k-1}) \] (89)

which will yield a feedback mechanism to influence the error propagation dynamics.

Then, the quantity \( \beta_{k} \) playing the important role in the backward linear prediction is replaced with \( \tilde{\beta}_{k} \) as
\[ \tilde{\beta}_{k} = 1 - \mu_{k} W \hat{e}_{k} \]
\[ = \beta_{k} - \kappa \mu_{k} W (\eta_{k} - \rho^{-1}S_{k}^{T}e_{k-1}). \] (90)

A statistical analysis suggests that \( \tilde{\beta}_{k} \) is more apart from zero than \( \beta_{k} \) (Appendix B). Consequently, the numerical error, which causes the instability, contributes to preventing \( \tilde{\beta}_{k} \) from being close to zero, rescuing the recursion of \( D_{k} \) from the worst scenario.

Hence, it is expected that the feedback control of (89) for the backward prediction error \( \eta_{k} \) improves the numerical stability of the fast \( H_{\infty} \) filter since the error feedback control prevents \( \tilde{\beta}_{k} \) from approaching to zero. The modified filter is referred to as the stabilized fast \( H_{\infty} \) filter (SFHF) hereafter.

6. Numerical Examples

The performance of the proposed stabilization method for the fast \( H_{\infty} \) filter is evaluated using identification of a finite
impulse response (FIR) system such as widely used in echo cancellers. Also, the forgetting factor is chosen as \( \rho = 1 - \gamma^2 \) for simplicity, and all calculations are executed using MATLAB 5.3.

As an unknown system to be identified, we consider an echo path whose impulse response \( \{h_i\} \) consists of \( \{0.0, 0.008, -0.012, 0.064, 0.013, -0.052, -0.007, 0.039, 0.011, 0.0, -0.002, -0.009 -0.016, -0.013, -0.001, 0.004, 0.015, 0.013 0.007, 0.0, -0.001, -0.002, -0.001, 0.0 \} \) for \( i < 24 \) and zero otherwise (\( 24 \leq i < N \)). The observed echo is given by

\[
y_k = \sum_{i=0}^{N-1} h_i u_{k-i} + v_k, \quad k = 1, 2, \ldots, L
\]

where \( v_k \) is a stationary, white Gaussian noise with zero mean and standard deviation \( \sigma_v = 1.0 \times 10^{-4} \). \( N \) denotes the length of the adjustable impulse response (tap number), and \( L \) stands for the length of observation data. Note that the state-space model of the observed echo \( y_k \) becomes time-varying due to \( H_k = [u_k, u_{k-1}, \ldots, u_{k-(N-1)}] \) with a shifting property such that \( H_{k+1} = [u_{k+1}, H_k(1), H_k(2), \ldots, H_k(N-1)] \), provided that \( x_k = [h_0, h_1, \ldots, h_{N-1}]^T \). The received signal (tap inputs) \( \{u_i\} \) is generated by the following autoregressive (AR) model:

\[
u_k = \alpha_1 u_{k-1} + \alpha_2 u_{k-2} + w'_k
\]

where \( \alpha_1 = 0.7, \alpha_2 = 0.1 \), and \( w'_k \) is a stationary, white Gaussian noise with zero mean and standard deviation \( \sigma_{w'} = 0.04 \).

Figure 1 demonstrates the differences in stability between the fast \( H_\infty \) filter and the stabilized fast \( H_\infty \) filter (\( \kappa = 1.0 \)) using the squared norm of tap error vector \( \|\vec{x}_{k+1} - \bar{x}_k\|^2 \) as a measure, where \( N = 64, \gamma_f = 2.2, \lambda(\gamma_f) = \gamma_f^2, \) and \( \epsilon_0 = 1.0 \).

Figure 2 shows the distribution of the maximum eigenvalues of \( F^{33}_k \) for a period of 1 to \( L \), i.e., the histogram of \( \max_i ||\lambda_i(F^{33}_k)|| \) for each filter, where \( \lambda_i \) denotes the eigenvalue, and \( F^{33}_k \) the error transition matrix of the backward linear prediction coefficient. Furthermore, the maximum eigenvalue of the averaged transition matrix \( \vec{F}^{33} = 1/L \sum_{k=1}^{L} F^{33}_k \), \( \max_i ||\lambda_i(F^{33}_k)|| \), is listed in Table 1. From Fig. 2 and Table 1, we see that the error feedback control drastically decreases the eigenvalues of the error transition matrix \( F^{33}_k \), and succeeds to numerically stabilize the fast \( H_\infty \) filter.

Fortunately, the modification of the backward prediction error \( \vec{h}_k \) has not given a significant effect on the overall performance of the fast \( H_\infty \) filter although the optimality of \( D_k \) is different from the original one, and an additional computational burden for the modification was also negligible.

### 7. Conclusions

In this paper, we have derived a linear propagation model of numerical errors in the recursive variables of the fast \( H_\infty \) filter, and then analyzed the stability of the filter using the first-order error propagation model. Additionally, based on the analyzed results, we have proposed a numerical stabilization method of the fast \( H_\infty \) filter with the error feedback control in the backward prediction. The effectiveness of the stabilization method was verified using numerical examples.

Also, in our recent experiments, the stabilized fast \( H_\infty \) filter has succeeded in continuously running without restart on a DSP for days.

Further analysis of the behavior of \( \vec{h}_k \) will be one of our
future works.

This work was partly supported by the Ministry of Education, Science, Sports and Culture in Japan, Grant-in-Aid for Scientific Research (C) 20560410.

References


Appendix A: Derivation of the Error Transition Matrix

Each block entry of the error transition matrix \( F_k \) is derived as follows:

\[
F_{k}^{11} = \frac{\partial A_k}{\partial A_{k-1}^T} = \frac{\partial (A_{k-1} - K_{k-1} W c_k)}{\partial A_{k-1}^T} = \frac{\partial A_{k-1}}{\partial A_{k-1}^T} - \frac{\partial (K_{k-1} W c_k)}{\partial A_{k-1}^T} = (I - K_{k-1} W C_{k-1}) \tag{A-1}
\]

\[
F_{k}^{22} = \frac{\partial S_k}{\partial S_{k-1}} = \frac{\partial (\rho S_{k-1})}{\partial S_{k-1}} + \frac{\partial (\rho W e_k)}{\partial S_{k-1}} = \rho \frac{\partial S_{k-1}}{\partial S_{k-1}} = \rho \tag{A-2}
\]

\[
F_{k}^{33} = \frac{\partial D_k}{\partial D_{k-1}^T} = \frac{\partial}{\partial D_{k-1}^T} \left( D_{k-1} - m_k W \eta_k \right) \beta_k^{-1}
\]

\[
= \frac{\partial (D_{k-1} - m_k W \eta_k)}{\partial D_{k-1}^T} \beta_k^{-1} + (D_{k-1} - m_k W \eta_k) \frac{\partial \beta_k^{-1}}{\partial D_{k-1}^T} = (I - m_k W C_k) \beta_k^{-1} + (D_{k-1} - m_k W \eta_k) \frac{\partial \beta_k^{-1}}{\partial D_{k-1}^T} \frac{\partial \beta_k}{\partial D_{k-1}^T} = (I - m_k W C_k) \beta_k^{-1} + (D_{k-1} - m_k W \eta_k) \frac{\partial \beta_k^{-1}}{\partial \beta_k} \frac{\partial \beta_k}{\partial D_{k-1}^T} = (I - m_k W C_k) \beta_k^{-1} + (D_{k-1} - m_k W \eta_k) \frac{\partial \beta_k^{-1}}{\partial \beta_k} \frac{\partial \beta_k}{\partial D_{k-1}^T} = (I - m_k W C_k) \beta_k^{-1} + (D_{k-1} - m_k W \eta_k) \frac{\partial \beta_k^{-1}}{\partial \beta_k} \frac{\partial \beta_k}{\partial D_{k-1}^T}
\]

where

\[
\beta_k = 1 - \mu_k W \eta_k, \quad \eta_k = e_{k-N} + C_k D_{k-1} \tag{A-4}
\]

Also, noting

\[
S_k = \rho S_{k-1} + [c_k + C_{k-1} A_{k-1} - C_{k-1} K_{k-1} W (c_k + C_{k-1} A_{k-1})] \times W (c_k + C_{k-1} A_{k-1}) + [c_k - C_{k-1} K_{k-1} W c_k + C_{k-1} (I - K_{k-1} W C_{k-1}) A_{k-1})] \times W (c_k + C_{k-1} A_{k-1}) \tag{A-5}
\]

we can obtain the explicit form of \( F_{k}^{21} \) as

\[
F_{k}^{21} = \frac{\partial S_k}{\partial A_{k-1}^T} = (c_k + C_{k-1} A_{k-1}) \times W (c_k + C_{k-1} A_{k-1}) + \rho [c_k - C_{k-1} K_{k-1} W c_k + C_{k-1} (I - K_{k-1} W C_{k-1}) A_{k-1})] \times W (c_k + C_{k-1} A_{k-1}) \tag{A-6}
\]

Similarly, we can calculate \( F_{k}^{31} = \frac{\partial D_k}{\partial A_{k-1}^T} \) and \( F_{k}^{32} = \frac{\partial D_k}{\partial \beta_k} \) and \( F_{k}^{33} = \frac{\partial D_k}{\partial D_{k-1}^T} \).

Appendix B: Behavior Analysis of \( \beta_k \)

When the backward and forward estimations are sufficiently performed, \( \eta_k \) and \( e_k \) almost become roundoff errors. So, suppose that \( \eta_k \) and \( e_k \) are zero-mean random vectors being mutually independent. Indeed, these properties were also observed experimentally in our examples. Then, using (23), \( S_k > 0 \) and \( 1 - \gamma_f^2 > 0 \), we have

\[
E[\beta_k] = E[\beta_k] + \kappa E[\mu_k W \eta_k] + \kappa \rho^{-N} E[S_t \mu_k W \mu_k^T] = E[\beta_k] + \kappa \rho^{-N} E[S_t \mu_k W \mu_k^T]
\]
= E{β_k} + κp^{-N}(1 - γ_f^2)E{S_kμ_k(1, 1)^2}
≥ E{β_k}

where E{·} denotes expectation. Note that μ_kWμ_k^T = (1 - γ_f^2)μ_k(1, 1)^2 holds due to μ_k(1, 1) = μ_k(1, 2), which stems from C_k = [H_k^T, H_k^T]^T.

This implies that β̃_k is more apart from zero than β_k due to E{β̃_k} > 0. Indeed, in our examples, β̃_k takes the value more than 1, whereas β_k almost takes the value in the range of 0 and 1.

Tomonori Katsumata was born in Chiba, Japan in 1973. He received the B.E. and M.E. degrees from Yamanashi University in 1997 and 2003, respectively. He is currently pursuing the doctoral degree at Iwate University. His research interests are in system identification and echo canceller.

Kiyoshi Nishiyama was born in Tokyo, Japan in 1957. He received the M.E. degree in electrical engineering from Chiba University in 1985 and the degree of Dr.Eng. from Tokyo Institute of Technology in 1991. He joined the Department of Computer and Information Sciences, Iwate University in 1998. He is currently a Professor. His research interests are in estimation, optimization, and machine learning.

Katsuaki Satoh was born in Tokyo, Japan in 1947. He graduated from Yamanashi University. From 1970 to 1991 he worked for the Acoustic Research Lab. of Matsushita Electric Co., Ltd. Since 1991, he has been the President of Amenity Research Institute Co., Ltd.