STOCHASTIC H^∞ IDENTIFICATION: AN ITERATIVELY WEIGHTED LEAST SQUARES ALGORITHM

by

Sundeep Rangan and Wei Ren

Memorandum No. UCB/ERL M94/25

11 April 1994

COVER PAGE

STOCHASTIC H^∞ IDENTIFICATION: AN ITERATIVELY WEIGHTED LEAST SQUARES ALGORITHM

by

Sundeep Rangan and Wei Ren

Memorandum No. UCB/ERL M94/25

11 April 1994

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

STOCHASTIC H^∞ IDENTIFICATION: AN ITERATIVELY WEIGHTED LEAST SQUARES ALGORITHM

by

Sundeep Rangan and Wei Ren

Memorandum No. UCB/ERL M94/25

11 April 1994

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Stochastic $\mathbf{H}^\infty$ Identification: An Iteratively Weighted Least Squares Algorithm

Sundeep Rangan and Wei Ren

Department of Electrical Engineering and Computer Science
University of California at Berkeley, Berkeley, CA 94720
srangan@eecs.berkeley.edu, ren@eecs.berkeley.edu
fax: (510) 642-1341

Abstract

This paper presents an algorithm for $\mathbf{H}^\infty$ SISO system identification using a stochastic noise model and a constrained model set. The objective is to estimate a model within a fixed model set closest in the $\mathbf{H}^\infty$ norm to the unknown plant. Since the plant is unknown, the optimal $\mathbf{H}^\infty$ approximation cannot be exactly determined. Instead, we propose to select the model that minimizes the supremum norm of a non-parametric estimate of the plant to model error. It is shown by analysis and simulation that this strategy can produce model estimates closer in the $\mathbf{H}^\infty$ norm to the plant than either a simple least square identification scheme or a two-stage procedure where a non-parametric estimate is first formed, and then $\mathbf{H}^\infty$ model reduction is subsequently employed to satisfy the model constraints.

Moreover, by establishing a connection between a minimax problem and a sequence of weighted least square problems, it is shown that the proposed estimate can be obtained via a computationally attractive and conceptually simple iteratively weighted least square (IWLS) procedure. The IWLS algorithm is based on a sequence of classical parametric weighted least square output error identification problems, where the weighting filter is updated to asymptotically achieve the $\mathbf{H}^\infty$ criteria. The local (global) convergence of the method can be guaranteed whenever the least square output error minimizations converge locally (globally). When the model set is linearly parametrized, the IWLS procedure can be implemented in a recursive, online manner.

Under standard mild assumptions on the plant, bounds on the true plant to model error, which are probabilistic with respect to the noise and worst-case with respect to the plant, are derived. These $\mathbf{H}^\infty$ plant-to-model error is shown to converge to its minimum in mean square.

1 Introduction

System identification from random noise-corrupted time-domain data is central to control and signal processing. However, while identification within the stochastic framework has been extensively studied with least-square ($\mathbf{H}^2$) criteria, the demands of robust control have recently generated an interest in identification with $\mathbf{H}^\infty$ objectives. The traditional least square criteria is well-suited to model estimation in the presence of random noise since the noise can be "correlated out" and the resulting optimization problems are amenable to numerical computation. However, since robust control design is more closely related to the $\mathbf{H}^\infty$ norm, identification within the $\mathbf{H}^\infty$ framework is better suited to control-oriented system identification and has recently been receiving much attention [1]-[7], [11], [14].

A deterministic worst-case $\mathbf{H}^\infty$ identification problem was formulated and examined in [4]-[6] and further analyzed in [1], [2] and [11]. In this formulation, the identification is to be done from frequency response measurements at a finite number of frequencies with known, bounded errors with the object of obtaining an identification algorithm satisfying a certain asymptotic requirement on the worst-case $\mathbf{H}^\infty$ plant to model error. A robustly convergent untuned non-linear algorithm and a tuned linear algorithm were developed using polynomial interpolation. The disadvantages with this approach is that hard tight bounds on the frequency response data are difficult to obtain, and that the order of the models can be very high.

Another line of research [3, 7, 14] has focused on analyzing the performance of classical least square output error methods and estimating a frequency domain confidence band on the estimated transfer function. In [14] and [7], hard bounds are obtained assuming bounded deterministic noise. In [3], the noise model and the unmodeled dynamics are considered stochastic, and a probabilistic confidence band on the estimate is established. Moreover, a method for estimating the unmodeled dynamic statistics is also produced, and based on these statistics a model order can be determined. However, although frequency-domain confidence bounds can be established, simple least square methods produce estimates that are close in the $\mathbf{H}^2$, and not necessarily in the $\mathbf{H}^\infty$ sense. Moreover, the least-square analysis has been concerned with estimating error bounds but not minimizing them.

To attempt to overcome some of the difficulties with both these methods, we consider in this paper the system identification problem with a $\mathbf{H}^\infty$-type criteria, but with a constrained model set and a stochastic, as opposed to deterministic, noise model. The problem considered is: given a set of possible plant models and a finite input-output data record corrupted by random noise, to estimate the model within the model set closest to the plant in the $\mathbf{H}^\infty$ sense. If the plant were known, then the problem reduces to finding the member of the model set closest in the $\mathbf{H}^\infty$ norm to a known plant, which is precisely a problem of deterministic $\mathbf{H}^\infty$ model reduction. However, since the plant is unknown, only some estimate of the closest model within the model set can be made. The quality of such an estimate is to be evaluated on its statistical performance with respect to the noise and worst-case performance with respect to the plant.

This problem formulation differs in two important respects with the previous formulations. Firstly, the model set in fixed. The motivation for fixing the model set is that, at least in the deterministic algorithms, to obtain robust convergence the model order tends to grow rapidly. For example, in the algorithm in [6], the model order must grow at a rate greater than the square of the number of data points. It is argued in [4] that this is too large for practical purposes and a method for model reduction is proposed. This results in a two-stage procedure of high-order identification followed by model reduction (note that here, the term *two-stage* does not denote two stages in the sense of $\mathbf{L}^\infty$ interpolation, followed by a Nehari problem). Similarly in the stochastic setting, a natural $\mathbf{H}^\infty$ identification procedure, which we refer to as *empirical* $\mathbf{H}^\infty$ *identification (EHI)*, consists of first obtaining a high-order or non-parametric estimate, and subsequently conducting $\mathbf{H}^\infty$ model reduction to satisfy the model constraints. We show in this paper, that this scheme is in fact statistically inefficient, since large a bias in the high-order identification stage and eliminated during model reduction. Thus, placing constraints on the model order from the beginning can provide definite statistical advantages.

The second main feature of the proposed problem formulation is the noise model is stochastic as opposed to deterministic, which can potentially result in a less conservative estimate of the plant to model error. Of course, in the presence of unbounded noise, hard bounds on the plant to model estimate cannot be obtained. However, under mild assumptions on the plant, similar to those made in the deterministic case, we can provide a frequency-domain confidence band for the estimate and establish a stochastic robust convergence property where $\mathbf{H}^\infty$ plant-to-model error converges to its minimum.

To estimate the closet model in the $\mathbf{H}^\infty$ norm to the plant, we propose that the model be chosen to minimize the supremum of the plant to model error as estimated by an empirical transfer function estimate (ETFE) [10]. The ETFE is a classical, non-parametric model estimate given by the quotient of the input-output cross-spectrum and the input spectrum, which is then smoothed to reduce the effects of noise. We show that this identification scheme provides a better estimate in the $\mathbf{H}^\infty$ sense of the optimal model than either a simple least square approach or empirical $\mathbf{H}^\infty$ identification strategy. Moreover, by establishing a connection between a minimax problem and sequence of weighted least square problems, it is shown that the estimate can be computed via an iterative sequence of classical weighted least square identification problems.

In the iteratively weighted least square scheme (IWLS), the estimate is formed by a sequence of weighted least square output error estimates, where the weighting filter is updated by a simple multiplicative rule to asymptotically achieve the $\mathbf{H}^\infty$ criteria. The IWLS procedure is shown to converge to a local (global) minimum of the desired objective when the individual least square output error minimization converges to a local (global) minimum. In the special case where the plant model is linearly parametrized, the algorithm may be implemented online and recursively, either by repeated complex minimum deviation problems or through an IWLS scheme.

The outline of this paper is as follows. In section 2, we precisely state the problem formulation as the minimization of the smoothed ETFE errors. In section 3, probabilistic bounds on the ETFE and true plant to model errors are given. In section 4, the proposed problem is compared analytically against the LS and EHI procedures. A general IWLS procedure for minimax problems is introduced in section 5. The procedure is implemented in section 6 to achieve the identification algorithm. The special case where the model set is linearly parametrized is considered in section 7. In section 8, a numerical simulation is given. Proofs of the results are given in the Appendix.

2 Problem Formulation

Let G be the unknown plant, which is known a priori to belong to a given set $\mathcal{P} \subset \mathbf{H}^\infty$, where $\mathbf{H}^\infty$ denotes the set of causal, stable, discrete-time linear systems. Suppose that u is the input, v is a wide-sense stationary zero mean output noise process with known statistics and y is the noise-corrupted output

$$y = Gu + v. \tag{1}$$

We observe L input-output data values

$$u^L = (u_0, \ldots, u_{L-1})^T \tag{2}$$

$$y^L = (y_0, \ldots, y_{L-1})^T. \tag{3}$$

Let $\mathcal{M} \subset \mathbf{H}^\infty$ be the set of possible models. Typically, $\mathcal{P}$ will be a large set, while $\mathcal{M}$ will be a small subset of plants with a simple structure. Let $W_u \in \mathbf{H}^\infty$ be an uncertainty weight, where W_u is larger at frequencies where greater accuracy is required. Ideally we would like to find

$$\hat{G}_{opt} = \arg \min_{H \in \mathcal{M}} \|W_u(G - H)\|_\infty \tag{4}$$

Of course, $\hat{G}_{opt}$ cannot be computed since G is unknown. We thus consider an approximate objective as follows. The approximate objective will be based on M frequency points uniformly spaced on the unit circle. Define the sampled weight $W_u^M \in \mathbf{C}^M$ ($\mathbf{C}^M$ and $\mathbf{R}^M$ will denote the complex and real M-dimensional vector spaces respectively), by $W_u^M(k) = W_u(2\pi k/M)$ for $k = 0, \ldots, M-1$. Our objective is to find a model, $\hat{G} \in \mathcal{M}$, which minimizes the weighted supremum norm of a smoothed empirical transfer function estimate (ETFE) of the residuals

$$\hat{G} = \arg \min_{H \in \mathcal{M}} \|W_u^M \cdot \mathrm{ETFE}(y^L - Hu^L, u^L, W, M)\|_\infty \tag{5}$$

where $W \in \mathbf{L}^\infty$ ($\mathbf{L}^\infty$ denotes the space of measurable, essentially bounded functions on the unit circle) is a given smoothing window function, $\mathrm{ETFE}(y^L - Hu^L, u^L, W, M)$ is the smoothed M point ETFE, and the dot $(\cdot)$ denotes componentwise multiplication. The precise definition of the ETFE is given in [10] and will be repeated in the next section since our notation is slightly different. The objective criteria (5) is an approximation of the desired objective (4). The closeness of this approximation will be discussed below, but to the extent that the two objectives agree, solving problem (5) approximates finding a $\hat{G} \in \mathcal{M}$ closest to G in the $\mathbf{H}^\infty$ sense. It is in this sense, that we regard the problem as having an $\mathbf{H}^\infty$-type criteria.

3 ETFE Minimization as an Objective Criteria

We briefly repeat the definition of the ETFE here, both for completeness and to introduce our notation. For any time sequence x, let

$$x^L = (x_0, \cdots, x_L)^T \tag{6}$$

and let X^L denotes its DFT

$$X^L(k) = \frac{1}{\sqrt{L}} \sum_{t=0}^{L-1} x_t e^{-j\omega_0 kt} \qquad k = 0, \ldots, L-1 \tag{7}$$

where $\omega_0 = 2\pi/L$. Also, if $Y \in \mathbf{L}^\infty$, let $Y^L \in \mathbf{C}^L$ be the sampled function

$$Y^L(k) = Y(e^{j\omega_0 k}) \qquad k = 0, \ldots, L-1 \tag{8}$$

For any X^L, $Y^L \in \mathbf{C}^L$, their convolution product will be denoted by $X^L \star Y^L$, and the pointwise product by $X^L \cdot Y^L$. Decimation by n will be denoted by the map $D_n : \mathbf{C}^{Ln} \rightarrow \mathbf{C}^L$

$$D_n(X^{Ln})(k) = X^{Ln}(kn) \qquad k = 0, \ldots, L-1 \tag{9}$$

Now suppose $H \in \mathbf{H}^\infty$ and

$$z = Hx + w \tag{10}$$

for some input x and noise process w. Given z^L and x^L, a positive window function $W \in \mathbf{L}^\infty$ and a number of frequency points M, such that $L = nM$ for some $n > 0$, the smoothed ETFE of H is

$$\mathrm{ETFE}(z^L, x^L, W, M) = D_n \left\{ \frac{W^L \star (Z^L \cdot X^{L*})}{W^L \star (X^L \cdot X^{L*})} \right\} \tag{11}$$

where X^{L*} is the conjugate of X^L. The ETFE is a smoothed quotient of the input-output cross spectrum and the input spectrum, where the smoothing has been done to reduce the effects of the noise. The amount of smoothing, determined by the window, is based on a variance versus bias trade-off and is discussed in [10].

For $C > 0$ and $0 \leq \rho < 1$, define the subset of $\mathbf{H}^\infty$

$$\mathbf{H}^\infty(C, \rho) = \{H \in \mathbf{H}^\infty \mid H \text{ is analytic in } D_\rho \text{ and } |H(z)| \leq C, \forall z \in D_\rho\} \tag{12}$$

where $D_\rho = \{z \in \mathbf{C} \mid |z| \geq \rho\}$. For establishing error bounds on the ETFE, we will make only the assumptions that the plant $G \in \mathbf{H}^\infty(C, \rho)$ for some $C \geq 0$ and ρ with $0 \leq \rho < 1$. This is the identical assumption as made in the deterministic literature [4]-[7], and does not require any elaborate structural knowledge of G. Under this assumption, the following theorem uses bounds given in [10] to estimate the accuracy of the ETFE and establishes a stochastic-type robust convergence of the identification strategy of using ETFE minimization. The proof is given in the Appendix.

Theorem 1 *Let $C > 0$, $0 \leq \rho < 1$, $G \in \mathbf{H}^\infty(C, \rho)$ be the unknown plant and $\mathcal{M} \subseteq \mathbf{H}^\infty(C, \rho)$ be the model set. Denote the input, output and noise by u, y and v respectively as related by (1). Assume that u and v are quasi-stationary, uncorrelated and have continuous spectrums Φ_u and Φ_v.*

Fix an uncertainty weight $W_u \in \mathbf{H}^\infty$, with $W_u(e^{j\omega})$ continuously differentiable in ω. For each data length L, pick a number of frequency points M_L and a smoothing window $W_L \in \mathbf{L}^\infty$ assumed to be positive, even about $\omega = 0$ and normalized

$$\int_{-\pi}^{\pi} W_L(e^{j\omega}) d\omega = 1. \tag{13}$$

Define the estimate sequence

$$\hat{G}_L = \arg \min_{H \in \mathcal{M}} \|W_u^{M_L} \cdot ETFE(y^L - Hu^L, u^L, W_L, M_L)\|_\infty. \tag{14}$$

Then

(a) For any $L > 0$, there is a $\mathbf{C}^{M_L}$-valued zero mean random vector, X_L, such that for any $H \in \mathbf{H}^\infty$,

$$ETFE(Hu^L, u^L, W_L, M_L) = H^{M_L} + \mu_L(H) + X_L \tag{15}$$

where H^{M_L} is the true sampled transfer function (8), $\mu_L(H) \in \mathbf{C}^{M_L}$ is the ETFE bias, which is deterministic and given by

$$\mu_L(H, k) = S_L R(H, \omega_k) + \mathbf{O}(P_L) + \mathbf{O}(1/\sqrt{L}) \tag{16}$$

where $k \in \{0, \ldots, M_L - 1\}$, $\omega_k = 2\pi k/M_L$ and

$$S_L = \int_{-\pi}^{\pi} \omega^2 W_L(e^{j\omega}) d\omega \tag{17}$$

$$P_L = \int_{-\pi}^{\pi} |\omega|^3 W_L(e^{j\omega}) d\omega \tag{18}$$

$$R(H, \omega) = \frac{1}{2} H''(e^{j\omega}) + H'(e^{j\omega}) \frac{\Phi'_u(\omega)}{\Phi_u(\omega)}, \tag{19}$$

and X_L is ETFE variance, with

$$\mathbf{E}(|X_L(k)|^2) = \frac{\overline{W}_L \Phi_v(\omega_k)}{L \Phi_u(\omega_k)} + \mathbf{o}(\overline{W}_L/N). \tag{20}$$

Here $\mathbf{E}(\cdot)$ denotes expectation and

$$\overline{W}_L = 2\pi \int_{-\pi}^{\pi} W_L^2(e^{j\omega}) d\omega. \tag{21}$$

The prime ($'$) notation denotes differentiation with respect to ω.

(b) The plant to model error, $G - \hat{G}_L$, at the frequencies ω_k can be estimated by

$$|G(e^{j\omega_k}) - \hat{G}_L(e^{j\omega_k})| \leq |ETFE(y^L - \hat{G}_L u^L, u^L, W_L, M_L)(k)| + 2|\overline{\mu}_L(k)| + |X_L(k)| \tag{22}$$

where

$$\overline{\mu}_L(k) = S_L \overline{R}(\omega_k) + \mathbf{O}(P_L) + \mathbf{O}(1/\sqrt{L}) \tag{23}$$

$$\overline{R}(\omega) = \frac{C\rho^2}{(1-\rho)^2} + \frac{C|\Phi'_u(\omega)|\rho}{\Phi_u(\omega)(1-\rho)}. \tag{24}$$

(c) The effect of using finite frequencies in the objective criteria can be bounded by

$$\|W_u(G-H)\|_\infty \leq \|W_u^{M_L} \cdot (G^{M_L} - H^{M_L})\|_\infty + \beta/M_L \tag{25}$$

where $H \in \mathbf{H}^\infty(C, \rho)$ *and*

$$\beta = 2C(\|W_u'\|_\infty + \rho\|W_u\|_\infty/(1-\rho)). \tag{26}$$

(d) Suppose that the input, u, is sufficiently exciting so that

$$\int_{-\pi}^{\pi} \frac{|W_u(e^{j\omega})|^2 \Phi_v(\omega)}{\Phi_u(\omega)} d\omega < \infty. \tag{27}$$

If, as $L \to \infty$, $M_L \to \infty$, *and the windows* W_L *are selected so that* $M_L \overline{W}_L / L$, S_L *and* P_L $\to 0$ *then*

$$\|W_u(G-\hat{G}_L)\|_\infty \to \min_{H \in \mathcal{M}} \|W_u(G-H)\|_\infty \text{ in mean square.} \tag{28}$$

Moreover, the number of frequency points, M_L, *and the ETFE smoothing windows,* W_L, *can be selected, independent of any of the given data, so that* S_L, P_L *and* $\overline{W}_L M_L / L$ $\to 0$ *and* $M_L \to \infty$.

4 Comparison with LS and Empirical $\mathbf{H}^\infty$ Identification Procedures

In this section, we evaluate two alternative strategies to the proposed ETFE minimization criteria for estimating the closest model in the $\mathbf{H}^\infty$ norm to the unknown plant. The first is to minimize a standard LS output error criteria, and the second is to use an empirical $\mathbf{H}^\infty$ identification procedure. We will show that ETFE minimization criteria has certain advantages over both these procedures.

Suppose we are given a model set $\mathcal{M}$ and input-output data u^L and y^L. The least square (LS) output error method selects the model based on the criteria

$$\hat{G}_{LS} = \arg \min_{H \in \mathcal{M}} \sum_{t=0}^{L-1} (y_t - Hu_t)^2 \tag{29}$$

If the output y is given by (1), and u and v have spectrums Φ_u and Φ_v, then, as shown in [10],

$$\hat{G}_{LS} \approx \arg \min_{H \in \mathcal{M}} \int_{-\pi}^{\pi} |G(e^{j\omega}) - H(e^{j\omega})|^2 \frac{\Phi_u(\omega)}{\Phi_v(\omega)} d\omega. \tag{30}$$

Thus, as the data length increases, and the above approximation becomes tighter, the LS estimate, $\hat{G}_{LS}$, will converge to the closest model in the $\mathbf{H}^2$ norm to the plant, instead of the $\mathbf{H}^\infty$ norm as required. We will show in simulations below that this deviation can be substantial, and hence a simple LS output error criteria fails to solve the given $\mathbf{H}^\infty$ problem.

A second approach, which we will refer to as empirical $\mathbf{H}^\infty$ identification (EHI), is to first determine a high-order or non-parametric estimate, and then employ deterministic $\mathbf{H}^\infty$ model reduction to satisfy the model constraints. We will show here that this two-stage strategy, while a natural solution to the problem, is in fact statistically inefficient in comparison with the proposed direct criteria (5).

To analytically compare the EHI and the direct criteria, consider an EHI scheme where we obtain an M-point non-parametric ETFE estimate, F^M, and then use model reduction to obtain the final estimate $\hat{G}_{EHI}$. That is, if $\mathcal{M}$ is the model set, u^L and y^L the input-output data, $W \in \mathbf{L}^\infty$ the ETFE smoothing window and $W_u \in \mathbf{H}^\infty$ the uncertainty weight, then

$$\begin{aligned} F^M &= \mathrm{ETFE}(y^L, u^L, W, M) && (31) \\ \hat{G}_{EHI} &= \arg \min_{H \in \mathcal{M}} \|W_u^M \cdot (H^M - F^M)\|_\infty. && (32) \end{aligned}$$

Now consider an estimate, $\hat{G}_{dir}$ based on directly minimizing the criteria (5)

$$\hat{G}_{dir} = \arg \min_{H \in \mathcal{M}} \|W_u^M \cdot \mathrm{ETFE}(y^L - Hu^L, u^L, W, M)\|_\infty. \tag{33}$$

Using the notation in Theorem 1(a)

$$\begin{aligned} \hat{G}_{EHI} &= \arg \min_{H \in \mathcal{M}} \|W_u^M \cdot (H^M - G^M + \mu(G) + X_L)\|_\infty && (34) \\ \hat{G}_{dir} &= \arg \min_{H \in \mathcal{M}} \|W_u^M \cdot (H^M - G^M + \mu(G - H) + X_L)\|_\infty && (35) \end{aligned}$$

where G is the true plant. Since the desired objective criteria is

$$\min_{H \in \mathcal{M}} \|W_u(G - H)\|_\infty \approx \min_{H \in \mathcal{M}} \|W_u^M(G^M - H^M)\|_\infty, \tag{36}$$

the size of the bias terms, $\mu(G)$ and $\mu(G-H)$, and the variance, X_L, in (34) and (35) reflect the quality of the estimate. The EHI and direct estimates have the same variance. However, since $\hat{G}_{dir}$ is an estimate of G, $G - \hat{G}_{dir}$ will be small in comparison to G. By (16), $\mu(X)$ is proportional to X, and therefore the EHI bias $\mu(G)$ will in general be much larger than the direct method bias $\mu(G - H)$ for $H = \hat{G}_{dir}$. Using different windowing, the EHI bias can be decreased, but only at the expense of a higher variance. Such conclusions will be demonstrated in simulations below.

5 General IWLS Minimization

Before considering solving the minimization problem (5), we consider in this section a more general supremum norm problem and show that it can be solved as a sequence of weighted least square problems. It should be stressed that these algorithms are not intended as a computationally efficient way of solving the minimax problem, but rather a method for posing the minimax problem as a iteratively WLS problem, which is what is required in the following section.

Let $f : \mathbf{R}^m \rightarrow \mathbf{C}^M$ be an arbitrary function, for which the solution to the supremum norm problem

$$\theta^* = \arg \min_{\theta \in \mathbf{R}^m} \|f(\theta)\|_\infty \tag{37}$$

is required. We will show that this supremum norm problem is equivalent to an iterative sequence of weighted least square (WLS) problems. A WLS least square problem is, given a positive weight, $q \in \mathbf{R}^M$, minimize

$$\theta_{LS}(q) = \arg \min_{\theta \in \mathbf{R}^m} \|q \cdot f(\theta)\|_2^2 \tag{38}$$

Suppose that this WLS problem can be solved for every q. Our intent is to show that we can construct a sequence of positive weights, q_k such that $\theta_{LS}(q_k) \rightarrow \theta^*$ where θ^* solves (37). That is, asymptotically the WLS problems converge to the supremum norm problem.

To construct such a weight sequence, a natural recursive rule to update the weights would be

$$\tilde{q}_{k+1}^2 = q_k^2 \cdot ((1-t) + t|f(\theta_k)|^\alpha) \tag{39}$$

$$q_{k+1} = \tilde{q}_{k+1}/\|\tilde{q}_{k+1}\|_2 \tag{40}$$

where $\theta_k = \theta_{LS}(q_k)$, $t \in (0,1]$ is a step size, α is some positive constant, and the powers are to be done componentwise. The second step is simply for normalization. The rationale for this update is that, by multiplying by $(1-t) + t|f(\theta_k)|^\alpha$ the weights are increased at the places where the $f(\theta_k)$ is large with the hope of obtaining a minimum uniform norm.

The convergence of such a procedure will be analyzed later, but first we state the algorithm more precisely.

Algorithm 2 (General IWLS) *Let $f : \mathbf{R}^m \rightarrow \mathbf{C}^M$, and $\theta_{LS} : \mathbf{R}^M \rightarrow \mathbf{R}^m$ be the LS minimizer. Fix a step size $t \in (0,1]$, and weighting exponent $\alpha > 0$.*

1. *Initialize counter and weight: $k = 1$, $q_1 = (1,\ldots,1)^T/\sqrt{M} \in \mathbf{R}^M$.*
2. *Perform WLS minimization: $\theta_k = \theta_{LS}(q_k)$*
3. *Update the weights by (39) and (40).*
4. *$k = k+1$. Go to step 2*

This procedure uses a fixed step-size t, although in practice an adaptive step-size routine, such as the Armijo descent routine [12] would be used. For simplicity, we omit the implementation of this feature.

We may now state the convergence of the algorithm which is our main result on the equivalence between the IWLS and the supremum norm problems.

Theorem 3 *Let $f : \mathbf{R}^m \rightarrow \mathbf{C}^M$ and $\theta_{LS} : \mathbf{R}^M \rightarrow \mathbf{R}^m$ be a local minimization of the WLS problem for f. That is, for every positive $q \in \mathbf{R}^M$, there exists a U_q, an open neighborhood of $\theta_{LS}(q)$, such that*

$$\theta_{LS}(q) = \arg \min_{\theta \in U_q} \|q \cdot f(\theta)\|_2^2 \tag{41}$$

Let $\alpha > 0$.

(a) If $|f|^2$ and θ_{LS} are continuously differentiable, then for sufficiently small $t > 0$, q_k and θ_k as produced by Algorithm 2 will converge. If we define the Lagrangian

$$L_1(q,\theta) = \sum_{i=1}^{M} q_i^2 |f_i(\theta)|^2 / \sum_{i=1}^{M} q_i^2 \tag{42}$$

then $L_1(q_k, \theta_k)$ increases monotonically.

(b) If $|f|$ is componentwise convex, and θ_{LS} is continuous, then q_k and θ_k from Algorithm 2 will converge with $\alpha = 1$ and the step size parameter $t = 1$. If we define the Lagrangian

$$L_2(q,\theta) = \sum_{i=1}^{M} q_i^2 |f_i(\theta)| / \sum_{i=1}^{M} q_i^2 \tag{43}$$

then $L_2(q_k, \theta_k)$ increases monotonically.

(c) In either case (a) or (b), if $q_k \to q^*$ and $\theta_k \to \theta^*$, then θ^* is a local minimum of the supremum norm problem

$$\theta^* = \arg \min_{\theta \in U_{q^*}} \|f(\theta)\|_\infty \tag{44}$$

In particular, if θ_{LS} is a global minimizer of the WLS problem, then θ^* solves (37). In case (a) or (b), if $\theta \in U_{q^*}$ then,

$$L_1(q_k, \theta_k) \leq \min_{\theta \in U_q} \|f(\theta)\|_\infty^2 \leq \|f(\theta_k)\|_\infty^2 \qquad \textit{case (a)} \tag{45}$$

$$L_2(q_k, \theta_k) \leq \min_{\theta \in \mathbf{R}^m} \|f(\theta)\|_\infty \leq \|f(\theta_k)\|_\infty \qquad \textit{case (b).} \tag{46}$$

These bounds become tighter as $k \to \infty$, and therefore, $\|f(\theta_k)\|_\infty^2 - L_1(q_k, \theta_k) < \epsilon$ or $\|f(\theta_k)\|_\infty - L_2(q_k, \theta_k) < \epsilon$ for an $\epsilon > 0$ can be used as a stopping criteria for the algorithm.

The IWLS procedure is in fact a type of Lagrangian method where the problem is solved indirectly by updating Lagrange parameters (in this case q_k) to solve a dual maximization problem.

6 IWLS System Identification

Consider the identification problem (5), and suppose the model set can be parametrized by a map $\hat{G} : \mathbf{R}^m \rightarrow \mathcal{M}$, $\theta \mapsto \hat{G}(\theta)$. Let

$$f(\theta) = W_u^M \cdot \mathrm{ETFE}(y^L - \hat{G}(\theta)u^L, u^L, W, M) \tag{47}$$

so that our objective is to minimize

$$\theta^* = \arg \min_{\theta \in \mathbf{R}^m} \|f(\theta)\|_\infty \tag{48}$$

which is an estimate of the desired objective (4) since

$$\|f(\theta)\|_\infty \approx \|W_u(G - \hat{G}(\theta))\|_\infty \tag{49}$$

The problem is now in the form (37) and can be posed as a sequence of WLS problems (38) as in the previous section. The rationale to convert the problem to an IWLS form, is that each WLS minimization can be approximately implemented as a classical weighted output error minimization problem, and the weighting filter can be updated in frequency domain with a simple multiplicative rule.

To implement the IWLS procedure in this manner, first consider the WLS minimization in step 2 of Algorithm 2 and let e denote the output errors

$$e(\theta) = y - \hat{G}(\theta)u \tag{50}$$

Suppose q_k is the weight at the k^{th} iteration of the IWLS algorithm, and $H_k \in \mathbf{H}^\infty$ is any filter such that

$$|H_k(e^{j\omega_l})|^2 = |q_{k,l} W_u(e^{j\omega_l})|^2 \Phi_u^{-1}(\omega_l) \tag{51}$$

where $\omega_l = 2\pi l/M$, $l = 0, \ldots, M-1$ and Φ_u is the power spectrum of the input u, which we assume to be quasi-stationary in the sense of [10]. Then the model estimate θ_k is given by

$$\begin{aligned}
\theta_k &= \arg \min_{\theta \in \mathbf{R}^m} \|q_k \cdot f(\theta)\|_2^2 \\
&= \arg \min_{\theta \in \mathbf{R}^m} \|q_k \cdot W_u^M \cdot \mathrm{ETFE}(y^L - \hat{G}(\theta)u^L, u^L, W, M)\|_2^2
\end{aligned}$$

$$
\begin{aligned}
&\approx \arg \min_{\theta\in\mathbf{R}^m} \sum_{l=0}^{M-1} |q_{k,l} W_u(e^{j\omega_l})(G(e^{j\omega_l}) - \hat{G}(\theta, e^{j\omega_l}))|^2 \\
&= \arg \min_{\theta\in\mathbf{R}^m} \sum_{l=0}^{M-1} |H_k(e^{j\omega_l})(G(e^{j\omega_l}) - \hat{G}(\theta, e^{j\omega_l}))|^2 \Phi_u(\omega_l) + |H_k(e^{j\omega_l})|^2 \Phi_v(\omega) \quad (52) \\
&\approx \arg \min_{\theta\in\mathbf{R}^m} \frac{1}{L} \sum_{t=0}^{L-1} |H_k(z) e_t(\theta)|^2 \quad (53)
\end{aligned}
$$

This is a precisely a weighted output error minimization criteria, which is classical and extensively studied. Note that in (52) the second term could be added since it does not depend on θ.

To implement the weight update in step 3 in Algorithm 2, it is not necessary to compute the weights, q_k, and then compute the filter H_k from equation (51). Combining (39), (40) and (51) we see that it suffices to find a filter $H_k \in \mathbf{H}^\infty$ such that $|H_k^M| = h_k$ where the sequence $h_k \in \mathbf{R}^M$ is updated by

$$
\begin{aligned}
\tilde{h}_{k+1}^2 &= h_k^2 \cdot ((1-t) + tE_k^\alpha) \quad (54) \\
h_{k+1} &= \tilde{h}_{k+1} / \|\tilde{h}_{k+1}\|_2 \quad (55)
\end{aligned}
$$

where

$$E_k = |W_u^M \cdot \mathrm{ETFE}(y^L - \hat{G}(\theta_k) u^L, u^L, W, M)| \quad (56)$$

We thus have the following approximate implementation of an IWLS identification algorithm.

Algorithm 4 (Offline IWLS system identification) *Let u^L and y^L be the given input-output data, $W_u \in \mathbf{H}^\infty$ be an uncertainty weight, $W \in \mathbf{L}^\infty$ be the smoothing window and $\hat{G} : \mathbf{R}^m \rightarrow \mathcal{M}$ be the model set parametrization. Let $M > 0$, $0 < t < 1$ and $\alpha = 1$ or 2.*

1. *Initialize counter and weight: $k = 1$, $h_1 = (1, \ldots, 1)^T / \sqrt{M} \in \mathbf{R}^M$.*

2. *Construct weighting filter: Find a $H_k \in \mathbf{H}^\infty$ with*

$$|H_k(e^{j\omega\pi i/M})| = h_i \qquad i = 0, \ldots, M-1 \quad (57)$$

3. Minimize the WLS output errors:

$$\theta_k = \arg\ \min \frac{1}{L} \sum_{t=0}^{L-1} |H(z)(y_t - \hat{G}(\theta)u_t)|^2 \tag{58}$$

4. Compute error estimate, E_k by (56).

5. Update weight by equations (54) and (55).

6. $k = k + 1$. Go to step 2

The output error minimization in (58) is a classical problem in system identification and adaptive control, and discussed in detail in [9].

7 Online, Recursive ETFE Minimization for Linearly Parametrized Model Sets

In the case where the model set is parametrized linearly we can implement the IWLS procedure in a recursive online, algorithm. Suppose

$$\hat{G}(\theta) = \sum_{i=1}^{m} \theta_i G_i \tag{59}$$

where $G_i \in \mathbf{H}^\infty$ are basis functions of the model set. In the case of an FIR model set $G_i(z) = z^{-i+1}$

For an online implementation of the IWLS procedure, suppose that we receive the input-output data in L length data blocks, and after each block we are to update the model estimate based on all the blocks currently received. Denote the data record at the k^{th} time instant by

$$u_k^L = (u_{k,0}, \ldots, u_{k,L-1})^T \tag{60}$$

$$y_k^L = (y_{k,0}, \ldots, y_{k,L-1})^T \tag{61}$$

It is known that averaging of ETFE estimates from separate data blocks asymptotically eliminates the effects of noise, and thus a natural identification criteria in the online case is

$$\theta_k = \arg \min_{\theta \in \mathbf{R}^m} \|f_k(\theta)\|_\infty \tag{62}$$

where

$$f_k(\theta) = \frac{1}{k} \sum_{l=1}^{k} W_u^M \cdot \mathrm{ETFE}(y_l^L - \hat{G}(\theta) u_l^L, u_l^L, W, M) \tag{63}$$

To evaluate f_k recursively, we note that due to the linearity of the ETFE

$$\mathrm{ETFE}(y_k^L - \hat{G}(\theta) u_k^L, u_k^L, W, M) = G_{0,k} - \sum_{i=1}^{m} G_{i,k}\theta_i = G_{0,k} - A_k \theta \tag{64}$$

where

$$G_{0,k} = \mathrm{ETFE}(y_k^L, u_k^L, W, M) \qquad G_{i,k} = \mathrm{ETFE}(G_i u_k^L, u_k^L, W, M) \tag{65}$$

$$A_k = (G_{1,k} \mid \cdots \mid G_{m,k}) \tag{66}$$

Therefore f_k is linear and given by

$$f_k(\theta) = W_u^M \cdot (\overline{G}_{0,k} - \overline{A}_k \theta) \tag{67}$$

where

$$\overline{G}_{0,k} = \frac{1}{k} \sum_{l=1}^{k} G_{0,l} \qquad \overline{A}_k = \frac{1}{k} \sum_{l=1}^{k} A_l \tag{68}$$

Consequently, the online problem (62) is a linear minimum deviation problem which can be solved with convex programming. Moreover, the parameters of the problem, $\overline{G}_{0,k}$ and $\overline{A}_k$ can be computed recursively

$$\overline{G}_{0,k} = \overline{G}_{0,k-1} + \frac{1}{k}(G_{0,k} - \overline{G}_{0,k-1}) \qquad \overline{A}_k = \overline{A}_{k-1} + \frac{1}{k}(A_k - \overline{A}_{k-1}) \tag{69}$$

Thus, to recursively compute the estimate θ_k, after each data block, the parameters can be updated by equation (69) and the new estimate obtained by solving the minimum deviation problem (62).

However, such a method requires solving a convex programming problem at each step, and can be avoided by a recursive, online IWLS method. The online procedure is simply done as in the offline case except that the parameters are updated with new data as the optimization is performed. Let n_l denote the number of WLS iterations to be performed between receiving each data block. Then we have the following recursive IWLS algorithm.

Algorithm 5 (Online recursive IWLS) *Let the model set be parametrized by (59), and let u_k^L, y_k^L denote the data block sequence as given in (60). Fix the number of frequency points M, the uncertainty weight $W_u \in \mathbf{H}^\infty$ and the ETFE window $W \in \mathbf{L}^\infty$. Let $n_l > 0$ denote the number of WLS iterations to be performed per data block.*

1. *Initialize parameters:* $k = 1$, $q_{1,1} = (1, \ldots, 1)/\sqrt{M} \in \mathbf{R}^M$, $\overline{G}_{0,0} = 0$, $\overline{A}_0 = 0$.

2. *Obtain next data block* u_k^L, y_k^L.

3. *Compute the parameters* $\overline{G}_{0,k}$ *and* $\overline{A}_k$ *from equations (65), (66) and (69).*

4. *Generate the estimate* θ_k *via an IWLS scheme.*

   for $l = 1$ to n_l

   $\theta_{k,l} = \arg\ \min_{\theta \in \mathbf{R}^L} \|q_{k,l} \cdot f_k(\theta)\|_2^2$

   $\tilde{q}_{k,l+1} = q_{k,l} \cdot |f_k(\theta_{k,l})|^{1/2}$

   $q_{k,l+1} = \tilde{q}_{k,l+1} / \|\tilde{q}_{k,l+1}\|_2$

   end

   $q_{k+1,1} = q_{k,n_l+1}$, $\theta_k = \theta_{k,n_l}$

5. $k = k + 1$. *Go to step 2*

To track time-varying plants, one can use an exponential forgetting factor in (69).

8 Simulation Example

For simulation, we considered a lightly-damped plant as typical in $\mathbf{H}^\infty$ applications. We used a sixth order IIR plant

$$G(z) = 1/A(z) \tag{70}$$

where $A(z)$ is monic with zeros at

$$\{0.85e^{\pm 0.35\pi j},\ 0.7,\ -0.5,\ -0.3,\ 0.2\} \tag{71}$$

The model set, $\mathcal{M}$, was chosen as ninth order FIR filters, and we took $M = 128$, $L = 512$, $W_u = 1$ and W^M to be a $2L/M$ length Hamming window.

The optimal filter $\hat{G}_{opt}$ given in (4) was computed using the deterministic IWLS algorithm, Algorithm 2, and found to be

$$\begin{aligned} \hat{G}_{opt}(z) &= 0.9995 - 0.1283z^{-1} - 0.4824z^{-2} - 0.6090z^{-3} - 0.0544z^{-4} \\ &+ 0.2922z^{-5} + 0.2400z^{-6} + 0.0335z^{-7} - 0.4813z^{-8} - 0.1560z^{-9} \end{aligned} \tag{72}$$

with $\|\hat{G}_{opt} - G\|_\infty = 0.3405$. The plant and optimal FIR model are shown in figure 1. Shown in figure 2 is the plant-model error and the final weight q^*. It can be seen that weight has been adjusted so that the final error is flattened, and in this case, the minimal error is all-pass.

To implement the identification algorithm, Algorithm 4, a 64-tap FIR filter was used to do the weighting. Random Gaussian white noise was used to generate both the input u and the output noise v. Three noise levels, as measured by the signal-to-noise ratio (SNR), were used: SNR = 18, 24 and 30 dB. In the case where the input and noise are white noise, the SNR can be calculated by

$$SNR = 20\log(\|G\|_2\sigma_u/\sigma_v) \tag{73}$$

where σ_u and σ_v are the input and noise standard deviations. The convergence of the IWLS identification algorithm is shown in figure 3 where it can be seen that the final value is almost attained after fifty iterations, and a satisfactory estimate after ten to fifteen.

Finally, we compared the IWLS identification procedure against LS and EHI identification methods. The LS estimate was obtained by minimizing the output error criteria (29). The EHI model was determined by computing the ETFE estimate of the plant, and then finding the model that best fits the ETFE.

The result of the comparison is shown in table 1. For accuracy, at each noise level, the experiment was repeated fifty times and the values shown in the table are the average values.

Table 1: IWLS and MR estimation comparison.

SNR	Optimal	Avg. $\mathbf{H}^\infty$ estimate error				
	error	IWLS	EHI	LS	EHI-IWLS ratio	LS-IWLS ratio
(dB)	ϵ_{opt}	ϵ_{IWLS}	ϵ_{EHI}	ϵ_{LS}	$\epsilon_{EHI}/\epsilon_{IWLS}$	$\epsilon_{LS}/\epsilon_{IWLS}$
30	0.341	0.397	0.597	0.562	1.50	1.41
24	0.341	0.433	0.618	0.623	1.42	1.43
18	0.341	0.598	0.635	0.742	1.06	1.24

The optimal error is the minimum $\mathbf{H}^\infty$ plant to model error over the FIR model set. All errors quoted are in the $\mathbf{H}^\infty$ norm. For example,

$$\epsilon_{opt} = \|\hat{G}_{opt} - G\|_\infty \tag{74}$$

It can be seen that IWLS algorithm performed better than both the simple LS and EHI schemes. As expected, the improvement in the performance of the IWLS over the EHI is larger at lower noise levels where the bias component dominates the error. A typical plant-model error spectra for the IWLS and EHI estimates is shown in figure 4.

Conclusions

We have proposed a new strategy for $\mathbf{H}^\infty$ identification with a stochastic noise model and a fixed model set, where the model is chosen to minimize the supremum norm of a non-parametric estimate of the plant to model error. Probabilistic bounds are established on the true plant to model error. It is shown by analysis and simulation that minimizing such an objective is preferable both to an empirical $\mathbf{H}^\infty$ identification strategy or to a simple least-square procedure.

By establishing an equivalence of a minimax problem with a sequence of iteratively weighted LS problems, we have shown that the proposed $\mathbf{H}^\infty$ identification estimate can be computed by a sequence of classical weighted least square output error problems, resulting in an conceptually simple and computationally attractive iteratively weighted least square

(IWLS) algorithm. Our simulation experience shows that the IWLS procedure generally converges rapidly, usually in about 10-15 iterations. Moreover, in the case when the model set can be linearly parametrized, the IWLS procedure may be implemented in a recursive, online manner.

Because of the mild assumptions on the noise and plant, a natural and conceptually simple problem formulation and computational attractiveness, we believe that IWLS identification algorithm has much to offer to meet the demands of practical robust control design.

Further work however is necessary. The current bounds on the bias error and the uniform norm of the variance are quite conservative and it would be desirable to find tighter estimates. A more general problem is how soft error bounds that result from stochastic identification are to be incorporated in robust control. Also useful would be exploring tuning the algorithm to use a priori information on the plant and extensions to the MIMO case.

Acknowledgement

The authors are grateful to Dr. Kameshwar Poolla for valuable comments and useful discussions. The authors would also like to thank Ms. Chau Cong for preliminary simulations and Dr. Jian Zhou and Mr. Karim Nassir-Toussi for useful comments.

Appendix: Proofs

Proof of Theorem 1.

(a) These bounds are given in [10, pg.156] with a slight change in notation. It is also shown there that the variance, X_L, depends only on the noise, input and windowing and not H.

(b) From [5, Fact 5.3], for all $\omega \in \mathbf{R}$ and any $Y \in \mathbf{H}^\infty(C,\rho)$, the derivatives of Y are bounded by

$$|Y^{(k)}(e^{j\omega})| \leq Ck!\rho^k/(1-\rho)^k. \tag{75}$$

Thus, for any $H \in \mathbf{H}^\infty(C,\rho)$, we see $|R(H)| \leq \overline{R}$ and $|\mu(H)| \leq \overline{\mu}$. Also, from (16),

$|\mu(G-\hat{G}_L)| \leq |\mu(G)| + |\mu(\hat{G}_L)| \leq 2\overline{\mu}$. Since $y - \hat{G}_L u = (G - \hat{G}_L)u + v$, we can substitute $H = G - \hat{G}_L$ in (15) to obtain (22).

(c) Suppose $M > 0$, $\omega_k = 2\pi k/M$, and $Y \in \mathbf{L}^\infty$ is continuously differentiable. By approximating any ω with some ω_k and using Taylor's theorem, we have

$$|\|Y\|_\infty - \|Y^M\|_\infty| \leq \frac{2}{M}\|Y'\|_\infty. \tag{76}$$

For any $L > 0$ and $H \in \mathcal{M}$, taking $Y = W_u(G-H)$ in (76), using the product rule to evaluate the derivative and employing the bound (75), one obtains

$$|\|W_u(G-H)\|_\infty - \|W_u^{M_L} \cdot (G^{M_L} - H^{M_L})\|_\infty| = \leq beta/M_L \tag{77}$$

where β is defined in (26). Taking $H = \hat{G}_L$ gives (25).

(d) Now suppose the limits in statement (b) are satisfied, and define

$$\hat{E}_L(H) = \mathrm{ETFE}(y^L - Hu^L, u^L, W_L, M_L) \qquad H \in \mathcal{M}. \tag{78}$$

Taking $\hat{G}_{opt}$ as defined in (4) and using (14), (15), (77) and the fact that $|\mu(G-H)| \leq 2\overline{\mu}$ for all $H \in \mathbf{H}^\infty(C,\rho)$

$$\begin{aligned}
\|W_u(G-\hat{G}_{opt})\|_\infty &\geq \|W_u^{M_L}(G^{M_L} - \hat{G}_{opt}^{M_L})\|_\infty - \beta/M_L \\
&\geq \|W_u^{M_L} \cdot \hat{E}_L(\hat{G}_{opt})\|_\infty - 2\|W_u\|_\infty \overline{\mu}_L - \|W_u^{M_L} \cdot X_L\|_\infty - \beta/M_L \\
&\geq \|W_u^{M_L} \cdot \hat{E}_L(\hat{G}_L)\|_\infty - 2\|W_u\|_\infty \overline{\mu}_L - \|W_u^{M_L} \cdot X_L\|_\infty - \beta/M_L \\
&\geq \|W_u(G-\hat{G}_L)\|_\infty - 4\|W_u\|_\infty \overline{\mu}_L - 2\|W_u^{M_L} \cdot X_L\|_\infty - 2\beta/M_L. \quad (79)
\end{aligned}$$

Since S_L and $P_L \to 0$, and $M_L \to \infty$, $\overline{\mu}_L$ and $\beta/M_L \to 0$. Therefore, it suffices to show

$$\mathbf{E}\|W_u^{M_L} \cdot X_L\|_\infty^2 \to 0 \tag{80}$$

in order to show $\|W_u(G-\hat{G}_L)\|_\infty \to \inf_{H\in\mathcal{M}} \|W_u(G-H)\|_\infty$ in mean square.

Since $\overline{W_L}/L \to 0$, we will ignore the $\mathbf{o}(\overline{W}_L/L)$ term in (20) since it is dominated by the $\overline{W}_L \Phi_v(\omega)/L\Phi_u(\omega)$ term.

$$\mathbf{E}\|W_u^{M_L} \cdot X_L\|_\infty^2 \;\leq\; \mathbf{E}\|W_u^{M_L} \cdot X_L\|_2^2 = \sum_{k=0}^{M_L-1} \mathbf{E}|W_u^{M_L}(k) \cdot X_L(k)|^2$$

$$= \sum_{k=0}^{M_L-1} \frac{\overline{W}_L |W_u(e^{j\omega_k})|^2 \Phi_v(\omega_k)}{\Phi_u(\omega_k) L\epsilon^2}$$

$$\approx \frac{2\pi \overline{W}_L M_L}{L\epsilon^2} \int_{-\pi}^{\pi} \frac{|W_u(e^{j\omega})|^2 \Phi_v(\omega)}{\Phi_u(\omega)} d\omega \to 0 \tag{81}$$

since $\overline{W}_L M_L / L \to 0$ and the integral is finite. This establishes the convergence in mean square.

To obtain the limits, take, for example $M_L = \log(L)$, It is shown in [10] that if W_L is a Hamming window whose time domain impulse response has length $2\gamma_L + 1$, then

$$S_L \approx \frac{12}{\gamma_L^2} \qquad \overline{W_L} \approx 0.75\gamma_L \qquad \lim_{\gamma_L \to \infty} P_L = 0 \tag{82}$$

Choosing $\gamma_L = L^{1/2}$, $S_L \to 0$, $\overline{W}_L M_L / L \to 0$, $P_L \to 0$ and $M_L \to \infty$ as required. □

Proof of Theorem 3. Since the weight update, Lagrangians, WLS optimization and the supremum norm $\|f(\theta)\|_\infty$ depend only on $|f(\theta)|$, we can replace f with $|f|$ and assume f is componentwise non-negative.

(a) Denote the state space by

$$Q = \{q \in \mathbf{R}^M \mid q_i \geq 0 \text{ for } i = 1, \ldots, M, \text{ and } \|q\|_2 = 1\} \tag{83}$$

and for $t \in (0,1)$, define the maps on Q

$$U_t(q) = q \cdot [(1-t) + t f(\theta_{LS}(q))^\alpha]^{1/2} \tag{84}$$

$$S_t(q) = U_t q / \|U_t q\|_2. \tag{85}$$

Thus, if t is the step-size, then S_t is the transition map (ie. $S_t q_k = q_{k+1}$). Let $V_1(q) = L_1(q, \theta_{LS}(q))$, and note that $V_1(S_t q) = V_1(U_t q)$ for any $t \in (0,1)$ and $q \in Q$ since $S_t q$ is a constant multiple of $U_t q$ and therefore $\theta_{LS}(U_t q) = \theta_{LS}(S_t q)$ and $L_1(U_t q, \theta_{LS}(U_t q)) = L_1(S_t q, \theta_{LS}(S_t q))$. We will show that for sufficiently small t, $V_1(q_k)$ is an increasing Lyapanov function.

Fix a $q \in Q$, and let

$$\Delta = \frac{\partial U_t(q)^2}{\partial t} = q^2 \cdot (f(\theta_{LS}(q))^\alpha - 1)) \tag{86}$$

Since $\theta_{LS}(q)$ is a local minimum of $L_1(q,\theta)$ with respect to θ, we have

$$\left.\frac{\partial L_1(q,\theta)}{\partial \theta}\right|_{\theta=\theta_{LS}(q)} = 0. \tag{87}$$

Writing f_i for $f_i(\theta_{LS}(q))$, we have

$$\begin{aligned}
\|q\|_2^4 \left.\frac{\partial V_1(S_t q)}{\partial t}\right|_{t=0} &= \|q\|_2^4 \left.\frac{\partial V_1(U_t q)}{\partial t}\right|_{t=0} &(88)\\
&= \|q\|_2^4 \left.\frac{\partial L_1(q',\theta_{LS}(q'))}{\partial q'^2}\right|_{q'=q} \Delta &(89)\\
&= \sum_{i,j}(\Delta_i f_i^2 q_j^2 - \Delta_j f_i^2 q_i^2) &(90)\\
&= \sum_{i,j} q_i^2 q_j^2 \left(f_i^2(f_i^\alpha - 1) - f_i^2(f_j^\alpha - 1)\right) &(91)\\
&= \frac{1}{2}\sum_{i,j} q_i^2 q_j^2 \left(f_i^{2+\alpha} - f_i^2 f_j^\alpha - f_i^\alpha f_j^2 + f_j^{2+\alpha}\right). &(92)
\end{aligned}$$

Define $h(q)$ by

$$h(q) = \left.\frac{\partial V_1(S_t q)}{\partial t}\right|_{t=0} = \frac{1}{2\|q\|_2^4}\sum_{i,j=1}^{M} q_i^2 q_j^2 (f_i^2 - f_j^2)(f_i^\alpha - f_j^\alpha) \tag{93}$$

Now, for any $s > 0$, the map $x \mapsto x^s$ is strictly increasing for $x \geq 0$. Thus, for any $x, y \geq 0$, $x^2 - y^2$ and $x^\alpha - y^\alpha$ are either both non-positive or both non-negative, and therefore $(x^2 - y^2)(x^\alpha - y^\alpha) \geq 0$. Thus, $(f_i^2 - f_j^2)(f_i^\alpha - f_j^\alpha) \geq 0$, and consequently $h(q) \geq 0$.

Moreover, $h(q) = 0$ implies that $f_i = f_j$ for all i and j with q_i and $q_j \neq 0$. In such a case, $U_t q$ is a constant multiple of q and therefore $S_t q = q$.

Now fix a $\beta \in (0,1)$ and define the function $T : Q \to [0,1]$ by

$$T(q) = \sup\{t \in (0,1) \mid V_1(S_s q) \geq V_1(q) + \beta h(q) s,\ \forall s,\ 0 \leq s \leq t\}. \tag{94}$$

If $h(q) > 0$, then, since $h(q)$ is the derivative of $V_1(S_t q)$ at $t = 0$ and $\beta < 1$, $T(q) > 0$. If $h(q) = 0$, then $S_t(q) = q$ and therefore, $T(q) = 1 > 0$. Since θ_{LS} and f are continuous, so is T. Thus, since Q is compact, there exists a $t_0 \in (0,1)$ such that $T(q) \geq t_0$ for all $q \in Q$. Taking t_0 as the stepsize we have, for all $q \in Q$

$$V_1(S_{t_0} q) \geq V_1(q) + \beta t_0 h(q). \tag{95}$$

Therefore, $V_1(q)$ monotonically increases; and, since Q is compact, V_1 is bounded above. By LaSalle's principle, $q_k \rightarrow E$ where

$$E = \{q \in Q \mid V_1(S_{t_0}q) = V_1(q)\} = \{q \in Q \mid h(q) = 0\}. \tag{96}$$

But $h(q) = 0$ implies $S_{t_0}q = q$, and thus E is a set of fixed points. We have shown q_k converges to the set of fixed points, E. We must now show it converges to a particular fixed point within E.

For $q \in E$, $h(q) = 0$ implies there is an $m(q) > 0$ such that for every $i = 1, \ldots, M$, either $q_i = 0$ or $f_i(\theta_{LS}(q)) = m(q)$. We will show $m(q)$ is locally constant. Let $q \in E$, and $I = \{i \mid q_i \neq 0\}$. Let V be an open neighborhood of q such that, for all $q' \in V$, $q'_i \neq 0$ for $i \in I$, and $\theta_{LS}(q') \in U_q$. Let $q' \in E \cap V$. Then $q_i = q'_i = 0$ for $i \in I^c$, and for $i \in I$, $f_i(\theta_{LS}(q)) = m(q)$ and $f_i(\theta_{LS}(q')) = m(q')$. Using (41), and $\|q\|_2 = 1$,

$$m(q')^2 = \sum_{i \in I} q_i^2 f(\theta_{LS}(q'))^2 = \sum_i q_i^2 f(\theta_{LS}(q'))^2 \geq \sum_i q_i^2 f(\theta_{LS}(q))^2 = m(q)^2 \tag{97}$$

Similarly, $m(q') \leq m(q)$ and therefore $m(q)$ is locally constant. Therefore, we can partition E into subsets on which $m(q)$ is constant, such that the subsets are separable by disjoint neighborhoods. Since q_k converges to a set of fixed points, $q_{k+1} - q_k \rightarrow 0$, and thus q_k converges to some subset $E_1 \subseteq E$, where $m(q) = m$ for all $q \in E_1$ and some $m > 0$.

For $i = 1, \ldots, M$, define the closed sets

$$F_i = \{q \in E_1 \mid q_i = 0\} \qquad G_i = \{q \in E_1 \mid f_i(\theta_{LS}(q)) = m\} \tag{98}$$

so that $E_1 = F_i \cup G_i$. Since $V_1(q_k)$ converges and $\|q_k\|_2 = 1$, let $q_{k,i} f_i(\theta_k) \rightarrow g_i$ for $i = 1, \ldots, M$. If $g_i > 0$, then, since $f(\theta_{LS}(q))$ is bounded above, $\liminf q_k > 0$. Therefore, $q_k \rightarrow G_i$, which implies $f_i(\theta_k) \rightarrow m$ and $q_{k,i} \rightarrow g_i/m$. Now suppose $g_i = 0$ and $\epsilon > 0$. Find a $\delta > 0$, such that $\delta < \epsilon(m - \epsilon)$ and an $N > 0$, such that $k > N$ implies $q_{k,i} f_i(\theta_k) < \delta$ and either $q_{k,i} < \epsilon$ or $|f_i(\theta_k) - m| \leq \epsilon$. For $k > N$, if $q_{k,i} > \epsilon$, then $|f_i(\theta_k) - m| < \epsilon$ and

$$\delta > q_{k,i} f_i(\theta_k) \geq \epsilon(m - \epsilon) \geq \delta \tag{99}$$

which is impossible. Thus, $q_{k,i} \leq \epsilon$ for $k > N$, and consequently $q_{k,i} \rightarrow 0$. Therefore, if either $g_i = 0$ or $g_i > 0$, $q_{k,i}$ converges and thus so does q_k.

(b) In this case, the update is

$$q_{k+1} = q_k \cdot f(\theta_k)^{1/2} / \|q_k \cdot f(\theta_k)^{1/2}\|_2 \tag{100}$$

Fix some $q \in \mathbf{R}^M$ with positive components and use primed ($'$) coordinates to denote the values after one update. Then

$$L_2(q', \theta') = \frac{\sum_i q_i'^2 f_i'}{\sum_i q_i'^2} = \frac{\sum_i q_i^2 f_i f_i'}{\sum_i q_i^2 f_i} \tag{101}$$

Defining $M^{-1} = 2(\sum_i q_i^2 f_i)(\sum_i q_i^2)$ we have

$$L_2(q', \theta') - L_2(q, \theta) = 2M \sum_{i,j} q_i^2 q_j^2 (f_i f_i' - f_i f_j) \tag{102}$$

By the convexity of $f(\theta)$ in θ, if D denotes any subgradient with respect to θ

$$\sum_i q_i^2 f_i f_i' \geq \sum_i q_i^2 f_i^2 + (\theta' - \theta)^T \sum_i q_i^2 f_i D f_i \tag{103}$$

$$= \sum_i q_i^2 f_i^2 + \frac{1}{2}(\theta' - \theta)^T D \left(\sum_i q_i^2 f_i^2 \right) \tag{104}$$

$$= \sum_i q_i^2 f_i^2 \tag{105}$$

where the last equality holds since $\sum_i q_i^2 f_i^2(\theta)$ is minimized in θ and thus, there is a subgradient for which this is zero.

Substituting this in (102) gives

$$L_2(q', \theta') - L_2(q, \theta) \geq 2M \sum_{i,j} q_i^2 q_j^2 (f_i f_i - f_i f_j) = M \sum_{i,j} q_i^2 q_j^2 (f_i - f_j)^2 \tag{106}$$

and now the proof follows similar to part (a).

(c) Suppose $q_k \to q^*$. By continuity of θ_{LS} and f, θ_k and $f(\theta_k)$ converge, so let $\theta_k \to \theta^*$ and $f_k \to f^* = f(\theta^*)$. Define

$$I = \{ i \mid f_i^* = \|f^*\|_\infty,\ 1 \leq i \leq M \} \tag{107}$$

Take any $i \in I$ and $j \in I^c$, and let $a_{l,k} = ((1-t) + t f_l(\theta_k)^\alpha)$. Since $f_j(\theta_k) \to f_j^* < \|f^*\|_\infty$ and $f_i(\theta_k) \to \|f^*\|_\infty$, we have $a_{j,k}/a_{i,k} \to 0$. Thus,

$$\frac{q_{j,k}^2}{q_{i,k}^2} = \frac{q_{j,1}^2}{q_{i,1}^2} \frac{a_{j,1}}{a_{i,1}} \cdots \frac{a_{j,k-1}}{a_{i,k-1}} \to 0 \tag{108}$$

as $k \rightarrow \infty$. Since $q_{i,k}$ is bounded, $q_j^* = 0$. Consequently, $f_i^* = \|f^*\|_\infty$ for all i with $q_i^* \neq 0$.

Since θ^* is a local minima of the least-square function

$$\theta^* = \arg \min_{\theta \in U_{q^*}} \|q^* \cdot f(\theta)\|_2^2, \tag{109}$$

for all $\theta \in U_{q^*}$

$$\sum_{i \in I} q_i^{*2} f_i^2(\theta) = \sum_i q_i^{*2} f_i^2(\theta) = \sum_i q_i^{*2} f_i^2(\theta^*) = \sum_{i \in I} q_i^{*2} f_i^2(\theta^*) \tag{110}$$

Thus, for at least some $i \in I$, $f_i(\theta) \geq f_i(\theta^*) = \|f^*\|_\infty$. Therefore $\|f(\theta)\|_\infty \geq \|f(\theta^*)\|_\infty$ and (41) is shown.

For the inequality (45), the right hand inequality is clear. Since the Lyapanov function in (42) are increasing,

$$L_1(q_k, \theta_k) \leq L_1(q^*, \theta^*) = \frac{\sum_j q_j^{*2} f_j^{*2}}{\sum_j q_j^{*2}} = \frac{\sum_{j \in I} q_j^{*2} f_j^{*2}}{\sum_{j \in I} q_j^{*2}} = \|f^*\|_\infty^2 \tag{111}$$

where the second last equality holds since $q_j = 0$ for all $j \in I^c$. Also,

$$\lim_{k \rightarrow \infty} L_1(q_k, \theta_k) = L_1(q^*, \theta^*) = \|f^*\|_\infty^2 = \lim_{k \rightarrow \infty} \|f(q_k)\|_\infty^2 \tag{112}$$

so $\|f(q_k)\|_\infty^2 - L_1(q_k, \theta_k) \rightarrow 0$ The inequality (46) is proved similarly. □

References

[1] G. Gu , P. P. Khargonekar, "Linear and Non-linear Algorithms for Identification in $\mathbf{H}^\infty$ with Error Bounds," *IEEE Trans. Auto. Contr.*, vol.37, no.7, pp.953-963, July 1992.

[2] G. Gu , P. P. Khargonekar, "A Class of Algorithms for Identification in $\mathbf{H}^\infty$," *Automatica*, vol. 28, pp.299-312, 1992.

[3] G. C. Goodwin, M. Gevers, B. Ninness, "Quantifying the Error in Estimated Transfer Functions with Application to Model Order Selection," *IEEE Trans. Auto. Contr.*, vol.37, no.7, pp.913-928, July 1992.

[4] A. J. Helmicki, C. A. Jacobson, C. N. Nett, "Identification in $\mathbf{H}^{\infty}$: A Robustly Convergent Non-Linear Algorithm," in *Proc. 1990 ACC*, San Diego, CA, 1990, pp.386-391.

[5] A. J. Helmicki, C. A. Jacobson, C. N. Nett, "Control Oriented System Identification: A Worst Case/Deterministic Approach in $\mathbf{H}^{\infty}$". *IEEE Trans. Auto. Contr.*, vol.36, no.10, pp. 1163-1176, Oct. 1991.

[6] A. J. Helmicki, C. A. Jacobson, C. N. Nett, "Fundamentals of Control-Oriented System Identification and their Application for Identification in $\mathbf{H}^{\infty}$," in in *Proc. 1991 ACC*, Boston, MA, 1991, pp.88-99.

[7] A. J. Helmicki, C. A. Jacobson, C. N. Nett, "Least Square Methods for $\mathbf{H}^{\infty}$ Control-Oriented System Identification," *IEEE Trans. Auto. Contr.*, vol.38, no.5, pp. 819-826, May 1993.

[8] H. Kushner, D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Springer-Verlag, New York, 1978.

[9] L. Ljung, T. Söderströrm, *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA, 1983.

[10] L. Ljung, *System Identification: Theory for the User.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[11] J. R. Parkington, "Robust Identification and Interpolation in $\mathbf{H}^{\infty}$," *J. Math. Anal. and Appl.,* vol. 166, pp.428-441, 1992.

[12] E. Polak, *Computational Methods in Optimization: a Unified Approach.* Mathematics in Engineering and Science, v.77. Academic Press, New York, 1971.

[13] D. E. Rivera, J. F. Pollard, C. E. Garciá. "Control-Relevant Prefiltering: A Systematic Design Approach and Case Study," *IEEE Trans. Auto. Contr.*, vol.37, no.7, pp.964-974, July 1992.

[14] B. Wahlberg, L. Ljung, "Hard Frequency-Domain Model Error Bounds from Least-Squares Like Identification Techniques," *IEEE Trans. Auto. Contr.*, vol.37, no.7, pp.900-912, July 1992.

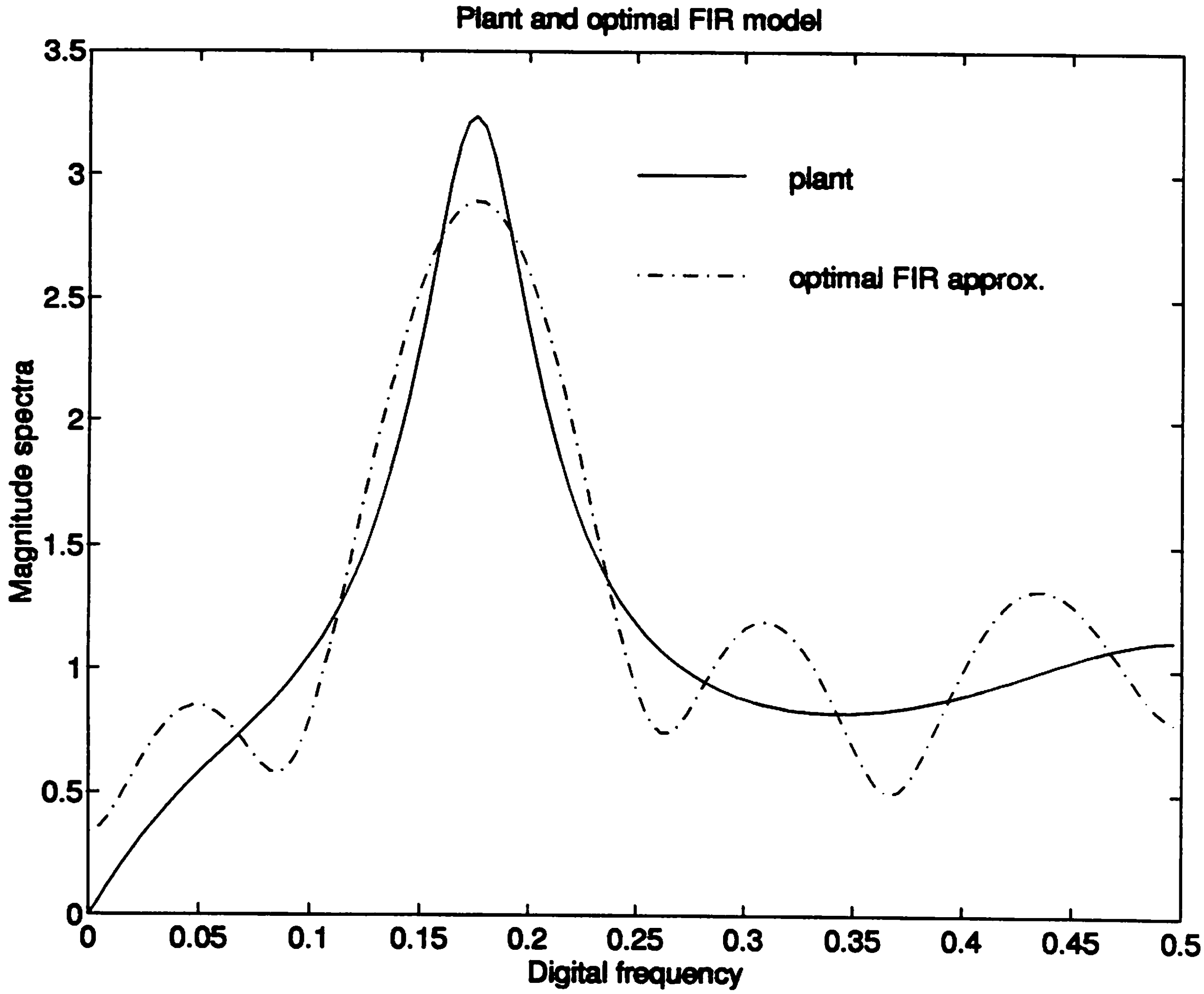

Figure 1: Plant and optimal FIR model approximation magnitude spectra

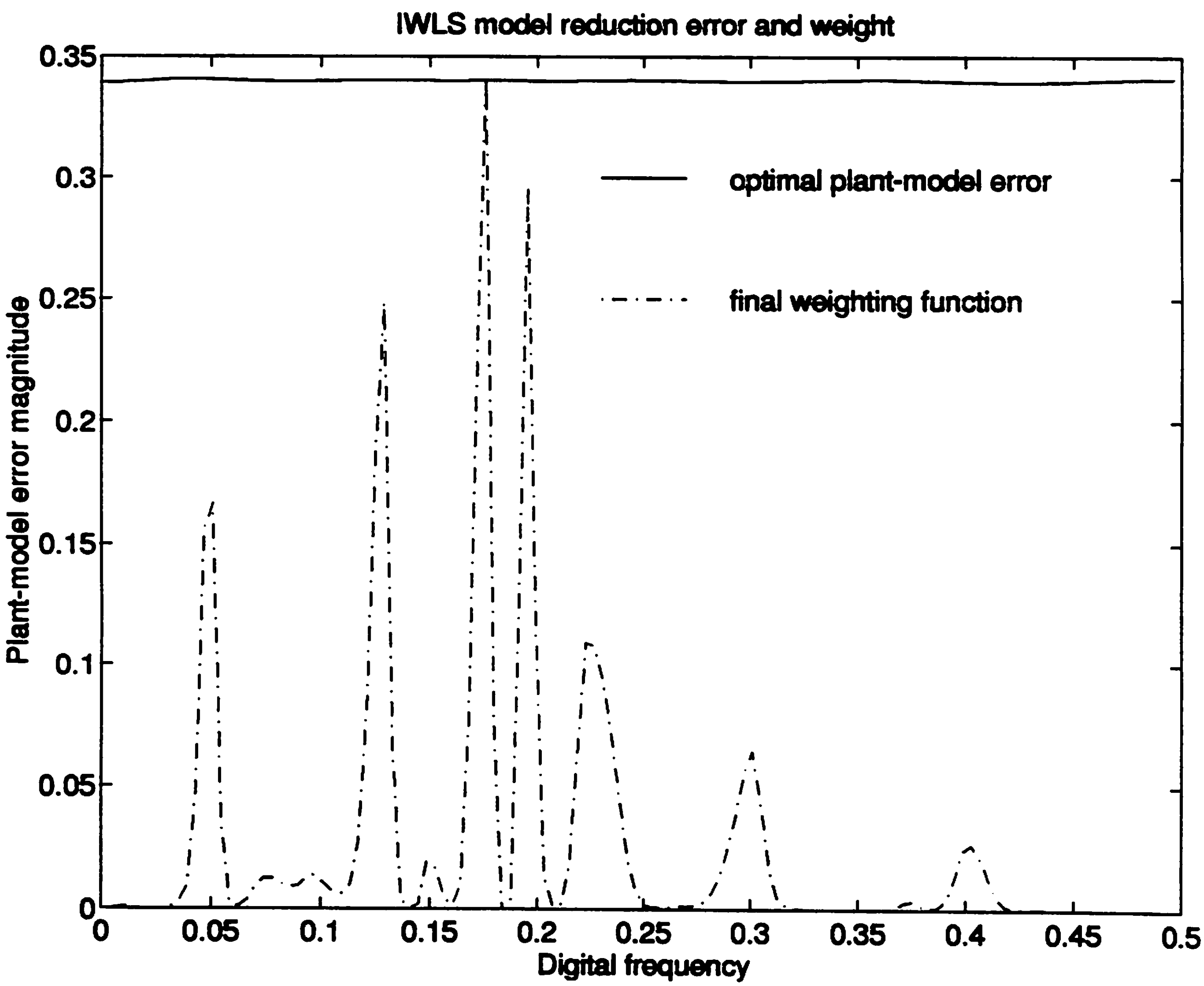

Figure 2: Minimum error and optimal weight magnitude

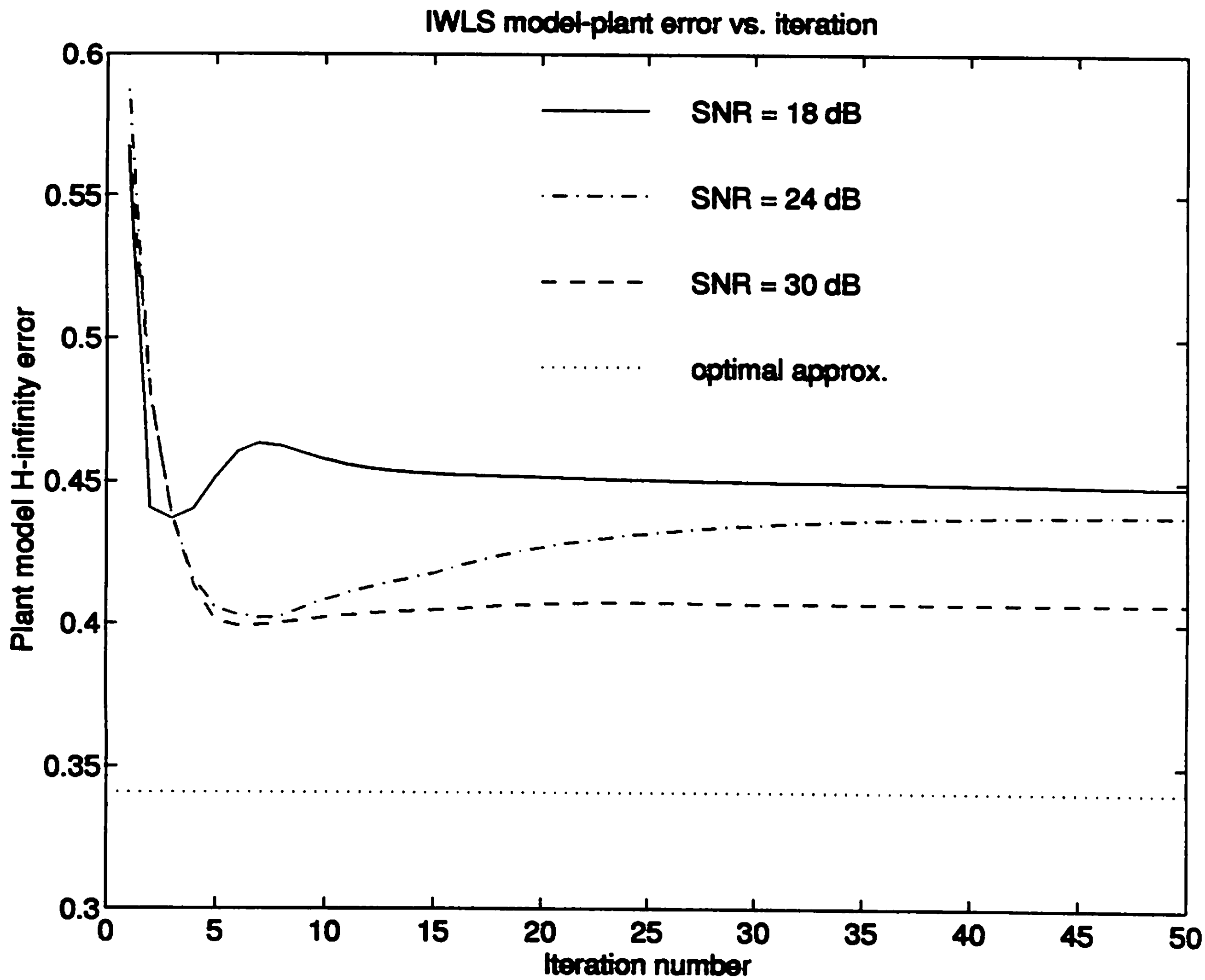

Figure 3: $\mathbf{H}^{\infty}$ plant model error vs. iteration in IWLS algorithm

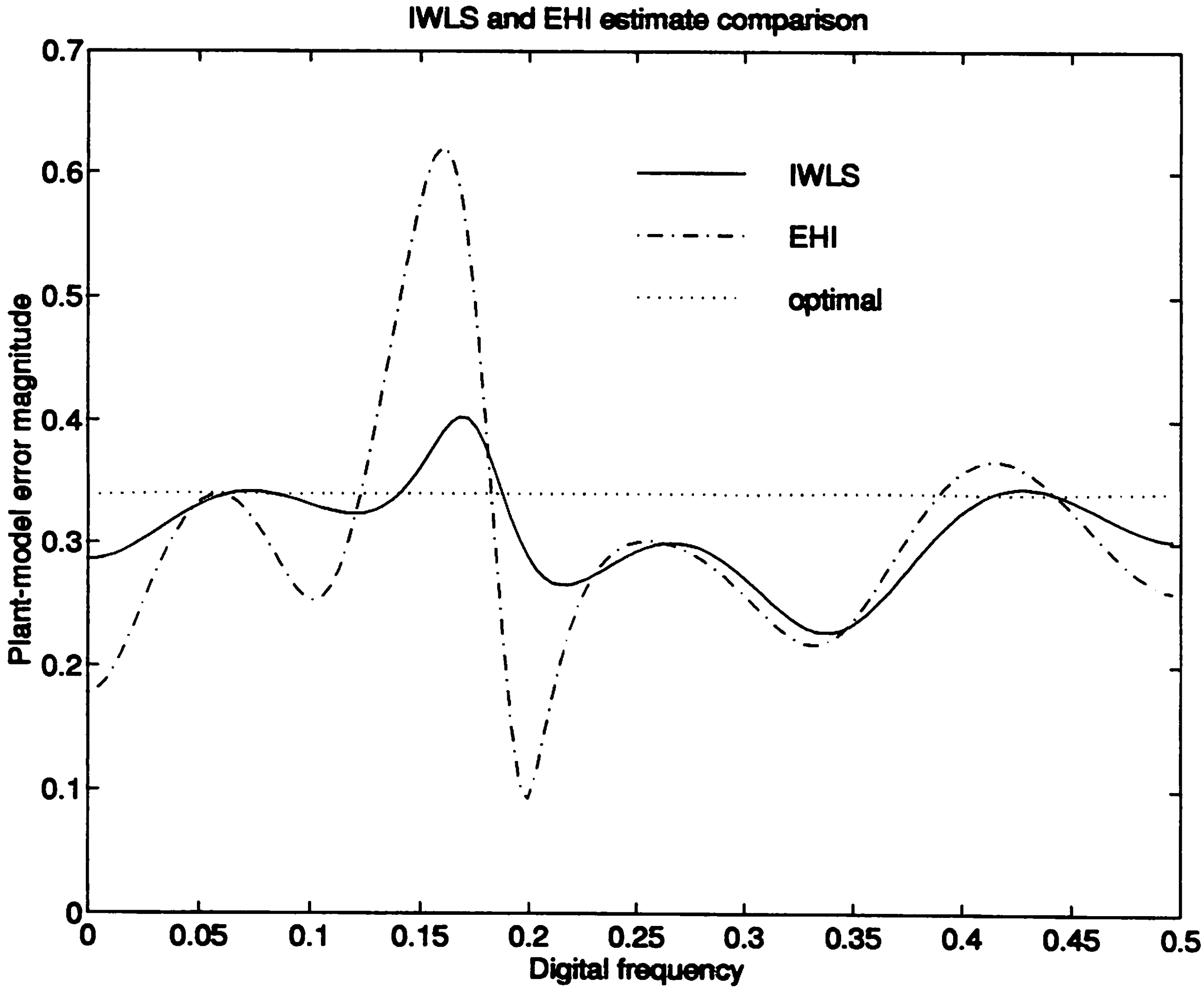

Figure 4: Typical plant-model error magnitude for IWLS and ~~MR~~ EHI estimates. SNR = 22 dB.