

Robust convergence of two-stage nonlinear algorithms for identification in \mathcal{H}_∞ *

Guoxiang Gu

Department of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803, USA

Pramod P. Khargonekar

Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, USA

Yanlin Li

Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455, USA

Received 20 August 1991

Abstract: Two-stage nonlinear algorithms for the problem of identification in \mathcal{H}_∞ are analyzed in terms of the Fourier transform of the *window function*. Conditions characterizing the robust convergence of the two-stage nonlinear algorithm are derived. The sufficiency of the conditions holds for a general class of window functions while the necessity is proved for robust convergence of the first stage of the algorithm under the mild additional restriction that the window function is even symmetric. Some improved upper bounds for the worst case identification error are also obtained.

Keywords: Identification in \mathcal{H}_∞ ; robust identification; frequency response; fast Fourier transform; Nehari problem.

1. Introduction

An abstract framework for worst case system identification has been recently developed in [10]. A concrete problem which is a specific instance of this abstract framework is known as system identification in \mathcal{H}_∞ [6,7]. A conceptual description of this problem is as follows. Suppose the experimental data are taken to be the noisy values of the frequency response of the unknown stable system at a finite set of frequencies. The problem is to find algorithms which map the experimental data into an identified model for which the worst case identification error (with respect to the noise and the unknown system) converges to zero in the \mathcal{H}_∞ norm as the noise level goes to zero and the number of data points goes to infinity. Further, it is of interest to derive explicit upper bounds on the worst case identification error in terms of the \mathcal{H}_∞ norm on the error system. In contrast to the prevailing stochastic approach of the theory of identification [16], this is a worst case deterministic formulation of the identification problem. (There are other worst case deterministic formulations of the robust identification problem. See [12,14,13,15,21,23,27,29,28,30] and the references therein for some representative papers).

This problem of identification in \mathcal{H}_∞ has attracted much interest during the last couple of years and much progress has been made [1,6,7,8,9,10,4,5,24,20,25,26]. Several efficient algorithms [1,7,8,4,5,24,25] for solving the aforementioned identification problem have been developed. A common feature of these nonlinear algorithms is a ‘two-stage’ structure. Roughly speaking, in the first stage, one finds a good, possibly unstable, approximation to the given data, and in the second the Nehari theorem is used to find the identified model as the best stable approximation to this unstable approximation. A general class of

* Supported in part by National Science Foundation under grants no. ECS-9001371, ECS-9110636, Airforce Office of Scientific Research under contract no. AFOSR-90-0053, Army Research Office under grant no. DAAL03-90-G-0008.

two-stage nonlinear algorithms was analyzed in [5]. Each algorithm in this class is characterized by a particular ‘window function’ which multiplies the inverse discrete Fourier transform of the given experimental data. In [5] a sufficient condition (on the window function) was obtained that guarantees the robust convergence of the resulting identification algorithm. A significant advantage of the approach developed in [5] is that one gets not one but a whole family of robustly convergent identification algorithms. This includes the previous algorithms in [7,4,24,25] as special cases.

In this paper, we take a different approach to the analysis of the general two-stage identification algorithms. We analyze the performance of the identification algorithm in terms of the Fourier transform of the aforementioned window function. (This is in contrast to the approach taken in [5] where the analysis was directly in terms of the window function.) Here we establish conditions for the robust convergence of the two-stage nonlinear algorithm. These conditions are given in terms of properties of the Fourier transform of the window function. The sufficiency of the conditions holds for a general class of window functions while the necessity is proved for robust convergence of the first stage of the algorithm under the mild additional restriction that the window function is even symmetric. These results give tighter conditions than those in [5]. This approach also leads to improved error bounds, although they may be harder to compute in some cases.

2. Preliminary

Let \mathcal{H}_∞ denote the Hardy space of bounded analytic functions in the open unit disc \mathcal{D} . For any function $\hat{f} \in \mathcal{H}_\infty$, the \mathcal{H}_∞ norm of \hat{f} is defined as

$$\|\hat{f}\|_\infty := \text{esssup}\{|\hat{f}(z)| : z \in \mathcal{D}\}. \quad (2.1)$$

Let \mathcal{A} denote the subset of \mathcal{H}_∞ defined as

$$\mathcal{A} := \{\hat{f} : \hat{f} \in \mathcal{H}_\infty, \hat{f} \text{ is continuous on the unit circle}\}. \quad (2.2)$$

As is well known, \mathcal{A} is a closed subalgebra of \mathcal{H}_∞ . The transfer function of a linear, causal, ℓ_2 BIBO-stable discrete-time shift invariant system can be viewed as an element in \mathcal{H}_∞ . If such a transfer function has a continuous frequency response, then it is also in \mathcal{A} . In particular, if the system is ℓ_∞ BIBO-stable, then its transfer function belongs to \mathcal{A} . *Transfer functions and Z-transform are defined such that stability corresponds to having no poles inside the unit disc.*

The problem of identification in \mathcal{H}_∞ formulated by Helmicki, Jacobson and Nett [7] can be briefly described as follows.

Problem. Assume that the ‘true’ unknown system to be identified is a stable, linear, shift invariant discrete-time system with transfer function $\hat{h} \in \mathcal{S} \subseteq \mathcal{A}$. Here \mathcal{S} is a fixed subset of \mathcal{H}_∞ containing the zero element;

Given a finite number N of possibly noisy experimental frequency response data

$$E^N(\hat{h}, \hat{\eta}) := \left\{ E_k^N(\hat{h}, \hat{\eta}) := \hat{h}(e^{j2\pi(k-1)/N}) + \hat{\eta}_k, \hat{\eta} \in B_N(\varepsilon), 1 \leq k \leq N \right\}, \quad (2.3)$$

where

$$B_N(\varepsilon) := \{\hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_N) \in \mathcal{E}^N : |\hat{\eta}_k| \leq \varepsilon \text{ for } k = 1, 2, \dots, N\}; \quad (2.4)$$

Find an algorithm A_N which maps the given information $E^N(\hat{h}, \hat{\eta})$ into an identified model $\hat{h}_{\text{id}}^N \in \mathcal{H}_\infty$ in such a way that the worst case model error

$$e^N(\mathcal{S}, \varepsilon) := \sup\{\|\hat{h} - \hat{h}_{\text{id}}^N\|_\infty : \hat{\eta} \in B_N(\varepsilon), \hat{h} \in \mathcal{S}\} \quad (2.5)$$

satisfies

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} e^N(\mathcal{S}, \varepsilon) = 0. \quad (2.6)$$

In addition, derive explicit upper bound for worst case identification error $e^N(\varepsilon)$ in term of \mathcal{H}_∞ norm. It is noted that if the data comes from a real physical system, then it is required to satisfy the familiar complex conjugate symmetry condition. For example, for N even, we have that $E_{N/2+1+k}^N = \bar{E}_{N/2+1-k}^N$ for $1 \leq k \leq \frac{1}{2}N - 1$ and $E_1^N, E_{N/2+1}^N$ are real where \bar{c} denotes the complex conjugate of c .

An algorithm that satisfies the above requirements is called ‘convergent’, and is called ‘robustly convergent and (untuned)’ if in addition it does not depend on the *a priori* information contained in the set \mathcal{S} and the noise bound ε . It turns out that in order to have a solution to this problem, we need to make some boundedness assumption on the set \mathcal{S} of all possible systems. This implies a certain minimal amount of *a priori* information on the unknown system. This assumption is called *admissibility* in [5] and described below.

Let P_n be the collection of all polynomials in z with degree not larger than n . Then, $P_n \subseteq \mathcal{A}$. Each $\hat{p}_n \in P_n$ corresponds to a finite impulse response system. It is known [32] that for each $\hat{h} \in \mathcal{A}$, there exists a $\hat{p}_n^* \in P_n$ such that

$$E_n(\hat{h}) = \|\hat{h} - \hat{p}_n^*\|_\infty = \inf\{\|\hat{h} - \hat{p}_n\|_\infty : \hat{p}_n \in P_n\}. \quad (2.7)$$

Clearly, \hat{p}_n^* is the optimal approximation to \hat{h} in the set P_n and will be denoted as $\hat{p}_n^*(\hat{h})$. The optimal error as in (2.7) is also a function of \hat{h} . Note that $E_n(\hat{h})$ is a monotone non-increasing function of n .

Definition 2.1. A subset $\mathcal{S} \subseteq \mathcal{A}$ is called admissible, if

- (1) $\lim_{m \rightarrow \infty} \delta_m = 0$, where $\delta_m := \sup\{E_m(\hat{h}) : \hat{h} \in \mathcal{S}\}$;
- (2) $M_s := \sup\{\|\hat{h}\|_\infty : \hat{h} \in \mathcal{S}\} < \infty$.

Roughly speaking, admissibility requires that the set of systems be bounded and be uniformly approximable by polynomials. We will assume that \mathcal{S} is an admissible set in the remainder of this paper. It can be shown that admissibility is equivalent to total boundedness of \mathcal{S} . We will also assume in the remainder of this paper that the set \mathcal{S} is nontrivial as defined below.

Definition 2.2. A set \mathcal{S} is called nontrivial if for $k = 0, 1, 2, \dots$, $\sup\{|h_k| : \hat{h} \in \mathcal{S}\} \neq 0$ where h_k is the k -th component of the impulse response of \hat{h} .

As mentioned in the introduction, several robustly convergent nonlinear algorithms have been proposed in [7,4,24,25]. These algorithms share the following ‘two-stage’ structure. At the first stage the inverse discrete Fourier transform and a window function are used to arrive at a good, possibly nonanalytic (i.e., unstable) approximation to the given frequency response data. Then in the second stage, Nehari’s theorem [2,22] is used to approximate this first stage approximation to obtain an analytic, i.e., stable, identified model.

To be more specific, define the N -point inverse discrete Fourier transform of the given experimental data as

$$\tilde{h}_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} E_{i+1}^N(\hat{h}, \hat{\eta}) e^{-jk(2i\pi/N)}, \quad j = \sqrt{-1}, \quad (2.8)$$

where $k = 0, 1, 2, \dots, N - 1$. The sequence $\tilde{h}_N(k)$ can be extended into a periodic sequence as follows:

$$\tilde{h}_N(k) = \tilde{h}_N(k \pm LN), \quad k, L \in \mathcal{S}, \quad (2.9)$$

where \mathcal{S} denotes the set of integers.

In this paper, we will study identification algorithms having the following two-stage structure:

Two-Stage Nonlinear Algorithm. *Stage 1:* Let $n(N)$ be any given monotone non-decreasing function from positive integers to positive integers such that $\lim_{N \rightarrow \infty} n(N) = \infty$ (whenever there is no possibility of confusion, we will not show explicitly the dependence of n on N). Let $w_{n,k}$, $k = 0, \pm 1, \dots, \pm n$, be any given bounded sequence (window function) of real numbers. Define pre-identified model $\hat{h}_{\text{pi}}^{E;N}$ as

$$\hat{h}_{\text{pi}}^{E;N} = \sum_{k=-n}^n w_{n,k} \tilde{h}_N(k) z^k. \quad (2.10)$$

Stage 2: Take identified model \hat{h}_{id}^N as

$$\hat{h}_{\text{id}}^N := \operatorname{argmin} \left\{ \left\| \hat{h}_{\text{pi}}^{E;N} - \hat{f} \right\|_{\infty} : \hat{f} \in \mathcal{H}_{\infty} \right\}. \quad (2.11)$$

End.

It is clear that the above identification algorithm is completely determined by the window functions $w_{n,k}$ as in (2.10). Stage 2 involves solving an optimization problem to arrive at the identified model. This optimization problem is quite well known and is called the Nehari best approximation problem. There are many techniques for solving this problem. The book by Young [31] contains a nice expository treatment of this problem. The two-stage identification algorithm does not depend on \mathcal{S} or the noise bound ε . It should be noticed that the two-stage identification algorithm is a *nonlinear transformation of the given data* because of the Stage 2 involving the Nehari best approximation. It is necessary to consider a nonlinear algorithm since Partington has shown that no linear robustly convergent algorithm exists [24], although it is shown in [4] that some linear algorithms are still quite effective. In [5], some sufficient conditions on the window function $w_{n,k}$ have been established that ensure the robust convergence of the two-stage nonlinear algorithm. In this paper, we will derive necessary and sufficient conditions for the robust convergence of the two-stage identification algorithm in terms of the Fourier transform of the window function. The result obtained in this paper complement the results reported in [5].

Definition 2.3. Stage 1 of the two-stage algorithm is said to be convergent if

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ N, n \rightarrow \infty}} e_{\text{pi}}^N(\mathcal{S}, \varepsilon) = 0, \quad (2.12)$$

where

$$e_{\text{pi}}^N(\mathcal{S}, \varepsilon) := \sup \left\{ \left\| \hat{h} - \hat{h}_{\text{pi}}^{E;N} \right\|_{\infty} : \hat{\eta} \in B_N(\varepsilon), \hat{h} \in \mathcal{S} \right\}. \quad (2.13)$$

Stage 1 of the two-stage algorithm is called robustly convergent (and untuned) if it is convergent and does not depend on \mathcal{S} and ε .

The importance of stage 1 of the nonlinear algorithm is shown in following lemma.

Lemma 2.4. *The two-stage nonlinear algorithm is robustly convergent if Stage 1 is robustly convergent. Moreover*

$$e^N(\mathcal{S}, \varepsilon) \leq 2e_{\text{pi}}^N(\mathcal{S}, \varepsilon). \quad (2.14)$$

The above bound was implicitly used in [7] to derive robustly convergent algorithm based on linear spline approximation, and a proof of the above lemma can be found in [5]. To analyze the identification error $e_{\text{pi}}^N(\varepsilon)$ in Stage 1, we follow the procedure developed in [5] and write $\tilde{h}_N(k)$ in (2.8) as

$$\tilde{h}_N(k) = h_N(k) + \eta_N(k) \quad (2.15)$$

where

$$h_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{h}(e^{j2i\pi/N}) e^{-jk(2i\pi/N)} \quad \text{and} \quad \eta_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\eta}_{i+1} e^{-jk(2i\pi/N)}. \quad (2.16)$$

The pre-identified model in Stage 1 of the nonlinear algorithm can thus be written as

$$\hat{h}_{\text{pi}}^{E;N} = \hat{h}_{\text{pi}}^N + \hat{\eta}_{\text{pi}}^N \quad (2.17)$$

where

$$\hat{h}_{\text{pi}}^N = \sum_{k=-n}^n w_{n,k} h_N(k) z^k \quad \text{and} \quad \hat{\eta}_{\text{pi}}^N = \sum_{k=-n}^n w_{n,k} \eta_N(k) z^k \quad (2.18)$$

with $h_N(k)$, $\eta_N(k)$ defined as in (2.16). Hence, \hat{h}_{pi}^N can be thought of as a pre-identified model in Stage 1 for noise free case and $\hat{\eta}_{\text{pi}}^N$ as that due to noise. The following lemma is also taken from [5].

Lemma 2.5. *Stage 1 of the identification algorithm is robustly convergent if and only if*

- (i) $\lim_{\substack{\varepsilon \rightarrow 0 \\ N, n \rightarrow \infty}} e_N^{\text{noise}}(\varepsilon) := \lim_{\substack{\varepsilon \rightarrow 0 \\ N, n \rightarrow \infty}} \sup \left\{ \|\hat{\eta}_{\text{pi}}^N\|_{\infty} : \hat{\eta} \in B_N(\varepsilon) \right\} = 0;$
- (ii) $\lim_{N, n \rightarrow \infty} e_N^{\text{app}}(\mathcal{S}) := \lim_{N, n \rightarrow \infty} \sup \left\{ \|\hat{h}_{\text{pi}}^N - \hat{h}\|_{\infty} : \hat{h} \in \mathcal{S} \right\} = 0.$

That is, State 1 is robustly convergent if and only if $\hat{h} = 0 \in \mathcal{S}$ can be identified robustly and for any model $\hat{h} \in \mathcal{S}$, it can be identified exactly as $N, n \rightarrow \infty$ with noise absent.

In the above lemma, $e_N^{\text{noise}}(\varepsilon)$ stands for *the worst case noise error* while e_N^{app} stands for *the worst case approximation error*. Lemmas 2.4 and 2.5 imply that although the identification algorithm considered in this paper is nonlinear, *the resulting identification error admits linear analysis*.

3. Main result

In this section, we will establish necessary and sufficient conditions on the Fourier transform of the window function which ensures the robust convergence of the first stage of the nonlinear algorithm. For each window function $w_{n,k}$, define

$$K_n(\omega) = \sum_{k=-n}^n w_{n,k} e^{jk\omega}. \quad (3.1)$$

The main result of this paper is the following theorem.

Theorem 3.1. *Consider the two-stage nonlinear identification algorithm described in the previous section. Suppose that the set \mathcal{S} is admissible. Let $K_n(\omega)$ be given in (3.1) with $N > 2n$. The two-stage nonlinear algorithm is robustly convergent if*

- (1) $\lim_{N > 2n \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} K_n(2i\pi/N) e^{-jk(2i\pi/N)} = 1 \quad \text{for } k = 0, \pm 1, \pm 2, \dots;$
- (2) $M_w = \sup \left\{ \frac{1}{N} \sum_{i=0}^{N-1} |K_n(\omega - 2i\pi/N)| : \omega \in [0, 2\pi/N], N > 2n > 0 \right\} < \infty.$

Conversely, suppose the set \mathcal{S} is admissible and nontrivial, and window function is evensymmetric, i.e., $w_{n,k} = w_{n,-k}$ for all k, n . If the first stage of the two-stage nonlinear algorithm is robustly convergent then conditions (1) and (2) are satisfied.

Proof. Since multiplication in the time domain is equivalent to (circular) convolution in the frequency domain, it can be easily verified that the pre-identified model is given by

$$\hat{h}_{\text{pi}}^{E;N}(e^{j\omega}) = \hat{h}_{\text{pi}}^N(e^{j\omega}) + \hat{\eta}_{\text{pi}}^N(e^{j\omega}) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{h}(e^{j2i\pi/N}) K_n(\omega_i) + \frac{1}{N} \sum_{i=0}^{N-1} \hat{\eta}_{i+1} K_n(\omega_i), \quad (3.2)$$

with $\hat{h}_{\text{pi}}^N, \hat{\eta}_{\text{pi}}^N$ as in (2.18) and $\omega_i := \omega - 2i\pi/N$. To show that the two conditions above are sufficient for robust convergence, we need to show that with the conditions given, the worst case identification error as defined in (2.13) converges to zero.

Indeed, using condition (2), we have that the worst case noise error $e_N^{\text{noise}}(\varepsilon)$ as defined in Lemma 2.5 satisfies

$$e_N^{\text{noise}}(\varepsilon) \leq \sup \left\{ \left\| \frac{1}{N} \sum_{i=0}^{N-1} \hat{\eta}_{i+1} K_n(\omega - 2i\pi/N) \right\| : \omega \in [0, 2\pi/N], \hat{\eta} \in B_N(\varepsilon) \right\} \leq M_w \varepsilon. \quad (3.3)$$

Thus, the worst case noise error converges to zero, i.e.,

$$\lim_{\varepsilon \rightarrow 0} e_N^{\text{noise}}(\varepsilon) = 0.$$

Next consider the worst case approximation error $e_N^{\text{app}}(\mathcal{S})$. Let $\delta > 0$ be given. Since $N > 2n$, well known properties of the discrete Fourier transform (DFT) yield the equality

$$w_{n,k} = \frac{1}{N} \sum_{i=0}^{N-1} K_n\left(\frac{2i\pi}{N}\right) W_N^{-ik}, \quad W_N = e^{j2\pi/N}, \quad (3.4)$$

for $k = 0, \pm 1, \pm 2, \dots, \pm n$. Condition (1) implies that for $k = 0, \pm 1, \pm 2, \dots$,

$$\lim_{N > 2n \rightarrow \infty} w_{n,k} = 1. \quad (3.5)$$

Now, for each function $\hat{h} \in \mathcal{S}$, we have that

$$\hat{h} = \hat{p}_m^*(\hat{h}) + \hat{\zeta}_m \quad \text{with} \quad \|\hat{\zeta}_m\|_\infty = \|\hat{h} - \hat{p}_m^*(\hat{h})\|_\infty = E_m(\hat{h}), \quad (3.6)$$

where $m > 0$ is any integer, and $\hat{p}_m^*(\hat{h}), E_m(\hat{h})$ are defined as in (2.7). Hence, even in this noise free case, we may still consider each $\hat{h} \in \mathcal{S}$ as the sum of polynomial $\hat{p}_m^*(\hat{h})$ and a ‘noise’ term $\hat{\zeta}_m = \hat{h} - \hat{p}_m^*(\hat{h})$ with ‘noise level’ $E_m(\hat{h})$. Let p_k be the coefficient of z^k in $\hat{p}_m^*(\hat{h})$. Then,

$$M_p(\hat{h}) := \sup\{|p_k| : 0 \leq k \leq m\} \leq \|\hat{p}_m^*(\hat{h})\|_\infty \leq 2 \|\hat{h}\|_\infty. \quad (3.7)$$

Using the decomposition in (3.6), it is clear that

$$\|\hat{h} - \hat{h}_{\text{pi}}^N\|_\infty \leq E_m(\hat{h}) + \|\hat{p}_m^*(\hat{h}) - \hat{h}_{\text{pi}}^N\|_\infty. \quad (3.8)$$

Since $\hat{p}_m^*(\hat{h})$ is a polynomial of degree no larger than m , we have that with $N > 2n > 2m$, (3.6) in combination with the linearity of the stage 1 yields

$$\hat{h}_{\text{pi}}^N = \sum_{k=0}^m w_{n,k} p_k z^k + \frac{1}{N} \sum_{i=0}^{N-1} \hat{\zeta}_m(e^{j2i\pi/N}) K_n(\omega - 2i\pi/N), \quad (3.9)$$

where p_k is the coefficient of $\hat{p}_m^*(\hat{h})$. Hence, by (3.7) and the condition (2), we further obtain

$$\|\hat{h} - \hat{h}_{pi}^N\|_\infty \leq (1 + M_w)E_m(\hat{h}) + M_p(\hat{h}) \sum_{k=0}^m |1 - w_{n,k}|. \quad (3.10)$$

Since \mathcal{S} is admissible, there exists a positive integer M_1 such that $(1 + M_w)\delta_m < \frac{1}{2}\delta$ for all $m \geq M_1$. By the admissibility of \mathcal{S} and (3.5), there exists an $M_2 > M_1$ such that for all $\frac{1}{2}N > n \geq M_2$, we have

$$\sup \left\{ M_p(\hat{h}) \sum_{k=0}^{M_1} |1 - w_{n,k}| : \hat{h} \in \mathcal{S} \right\} \leq 2M_s \sum_{k=0}^{M_1} |1 - w_{n,k}| \leq \frac{1}{2}\delta. \quad (3.11)$$

Hence, by setting $m = M_1$ in (3.10) and taking $N > 2n \geq 2M_2$, the worst case approximation error at the first stage is bounded as

$$e_N^{\text{app}}(\mathcal{S}) = \sup \left\{ \|\hat{h}_{pi}^N - \hat{h}\|_\infty : \hat{h} \in \mathcal{S} \right\} \leq \frac{1}{2}\delta + \frac{1}{2}\delta = \delta. \quad (3.12)$$

Since $\delta > 0$ is arbitrary, the worst case approximation error e_N^{app} at the first stage also converges to zero. By combining this with (3.3) and using Lemma 2.5, we have established the sufficiency of the conditions (1) and (2).

We show next that the conditions (1) and (2) are necessary for robust convergence at the first stage of the nonlinear identification algorithm. Since the set \mathcal{S} is nontrivial, the robust convergence of the first stage implies that for each $k \geq 0$, $\lim_{n \rightarrow \infty} w_{n,k} = 1$. (See also the proof of Proposition (3.5) in [5].) Now taking limit $N > 2n \rightarrow \infty$ in equation (3.4), we get that the condition (1) holds.

To prove the necessity of the condition (2), set $\hat{h} = 0$ in (3.2). Now using the facts that $K_n(\omega)$ is periodic with period 2π and that the noise $\hat{\eta}_{i+1}$ satisfies the conjugate symmetry property, we have

$$\begin{aligned} \hat{\eta}_{pi}^N(e^{j\omega}) &= \frac{1}{N} \left(\hat{\eta}_1 K_n(\omega) + \bar{\hat{\eta}}_{N/2+1} K_n(\omega + \pi) \right) \\ &\quad + \frac{1}{N} \sum_{i=1}^{N/2-1} \left(\hat{\eta}_{i+1} K_n\left(\omega - \frac{2i\pi}{N}\right) + \bar{\hat{\eta}}_{i+1} K_n\left(\omega + \frac{2i\pi}{N}\right) \right). \end{aligned} \quad (3.13)$$

We set next

$$c_{n,N}(\omega) = \frac{1}{N} (|K_n(\omega)| + |K_n(\omega + \pi)|). \quad (3.14)$$

Since the window function is even symmetric, the kernel $K_n(\omega)$ is a real function of ω . Let $\hat{\eta}_1 = \text{sgn}[K_n(\omega)]\varepsilon$ and $\hat{\eta}_{N/2+1} = \text{sgn}[K_n(\omega + \pi)]\varepsilon$. Here sgn is the sign function, i.e., $\text{sgn}[x] = 1$ if $x > 0$ and -1 if $x < 0$ and 0 if $x = 0$. Then $\hat{\eta}_1$ and $\hat{\eta}_{N/2+1}$ are real as required. For the rest of noise data, if we set

$$\hat{\eta}_{i+1} = \text{sgn} \left[K_n\left(\omega - \frac{2i\pi}{N}\right) + K_n\left(\omega + \frac{2i\pi}{N}\right) \right] \varepsilon,$$

for $1 \leq i \leq \frac{1}{2}N - 1$, then (3.13) yields

$$\hat{\eta}_{pi}^N(e^{j\omega}) = c_{n,N}(\omega)\varepsilon + \frac{\varepsilon}{N} \sum_{i=1}^{N/2-1} \left| K_n\left(\omega - \frac{2i\pi}{N}\right) + K_n\left(\omega + \frac{2i\pi}{N}\right) \right|.$$

On the other hand, if we set

$$\hat{\eta}_{i+1} = j \text{sgn} \left[K_n\left(\omega - \frac{2i\pi}{N}\right) - K_n\left(\omega + \frac{2i\pi}{N}\right) \right] \varepsilon, \quad j = \sqrt{-1},$$

for $1 \leq i \leq \frac{1}{2}N - 1$, then (3.13) yields

$$\hat{\eta}_{pi}^N(e^{j\omega}) = c_{n,N}(\omega)\varepsilon + \frac{j\varepsilon}{N} \sum_{i=1}^{N/2-1} \left| K_n\left(\omega - \frac{2i\pi}{N}\right) - K_n\left(\omega + \frac{2i\pi}{N}\right) \right|.$$

Hence, using the fact that $|a + b| + |a - b| \geq |a| + |b|$ for any real numbers a and b , the worst case noise error at first stage of the identification algorithm satisfies

$$e_N^{\text{noise}}(\varepsilon) \geq \frac{c_{n,N}(\omega)\varepsilon}{2} + \frac{\varepsilon}{2N} \sum_{i=1}^{N/2-1} \left(\left| K_n\left(\omega - \frac{2i\pi}{N}\right) \right| + \left| K_n\left(\omega + \frac{2i\pi}{N}\right) \right| \right).$$

Now suppose M_w is not finite. Then there exists a sequence of frequencies ω_κ and integers N_κ, n_κ satisfying $N_\kappa > 2n_\kappa, \kappa = 1, 2, \dots$, such that

$$\lim_{\kappa \rightarrow \infty} \left\{ c_{n_\kappa, N_\kappa}(\omega_\kappa) + \frac{1}{N_\kappa} \sum_{i=1}^{N_\kappa/2-1} \left(\left| K_{n_\kappa}\left(\omega_\kappa - \frac{2i\pi}{N_\kappa}\right) \right| + \left| K_{n_\kappa}\left(\omega_\kappa + \frac{2i\pi}{N_\kappa}\right) \right| \right) \right\} = \infty.$$

It follows that for any $\varepsilon > 0$,

$$\lim_{\kappa \rightarrow \infty} e_{N_\kappa}^{\text{noise}}(\varepsilon) = \infty$$

which contradicts the robust convergence of the first stage. \square

If we compare Theorem 3.1 with Theorem 3.3 in [5], it is not difficult to see that the condition (1) of Theorem 3.1 is equivalent to conditions (i) and (iii) of Theorem 3.3 in [5]. However, the condition (2) here is weaker than the condition (ii) in [5] in general. In fact, if the window function is even symmetric with respect to k , i.e., $w_{n,-k} = w_{n,k}$, and $w_{n,n} = 0$, then the following inequality can be shown:

$$M_w \leq \sup \left\{ n \left| \Delta w_{n,n-1} \right| + \sum_{k=0}^{n-2} (k+1) \left| \Delta^2 w_{n,k} \right| : n > 0 \right\}, \tag{3.15}$$

where M_w is as in condition (2) while the right hand side is from the condition (ii) of Theorem 3.3 in [5] and Δ denotes difference operator. In Theorem 3.3 of [5], condition (ii) requires that the right hand side of (3.15) be finite.

We would like to point out that although the conditions in Theorem 3.1 are tighter than those in [5], it may be more difficult to compute the upper bound M_w as in condition (2) above, except some special cases which is contrast to the *time domain* conditions as in [5] where error bounds can be easily established for many commonly used window functions.

We will now apply the result in Theorem 3.1 to a parametrized window function proposed in [5]. We being with the triangular window

$$t_{n,k} = 1 - \frac{|k|}{n}, \quad |k| < n, \quad w_{n,k} = 0, \quad |k| \geq n. \tag{3.16}$$

This window function is analyzed in [4] to study the convergence of the Cesaro mean based two-stage nonlinear algorithm for identification in \mathcal{X}_∞ . The Fourier transform of this window function is given by

$$F_n(\omega) = \sum_{k=1-n}^{n-1} \left(1 - \frac{|k|}{n} \right) e^{jk\omega} = \frac{1}{n} \left(\frac{\sin(\frac{1}{2}n\omega)}{\sin(\frac{1}{2}\omega)} \right)^2 \tag{3.17}$$

which is called the Fejér kernel. The key properties of the Fejér kernel are summarized in the following lemma.

Lemma 3.2. Let $F_n(\omega)$ be defined as in (3.17). Then:

- (1) $F_n(\omega) \geq 0$.
- (2) $(1/2\pi) \int_{-\pi}^{\pi} F_n(\omega) d\omega = 1$
- (3) For any positive integer $L \geq n$,

$$\frac{1}{L} \sum_{k=0}^{L-1} F_n\left(\omega - \frac{2k\pi}{L}\right) = 1, \quad \forall \omega \in [0, 2\pi/L]. \quad (3.18)$$

See [11] for a proof of the above lemma.

The window function we are going to analyze is the trapezoidal window

$$a_{n,k} = \begin{cases} 1, & 0 \leq k \leq 2m, \\ \frac{n+m-k}{n-m}, & 2m \leq k \leq n+m, \\ 1 + \frac{k}{n-m}, & m-n \leq k \leq 0, \\ 0, & \text{elsewhere,} \end{cases} \quad (3.19)$$

with $m < n$. Although the trapezoidal window does not satisfy the even symmetry property, the shifted window $w_{n,k} := a_{n,k+m}$ does and Theorem 3.1 still applies to the trapezoidal window (see also [5]). It is shown in [5] that with the above window, the worst case noise error $e_N(\hat{\eta})$ as defined in Lemma 2.5 is bounded by

$$e_N^{\text{noise}} \leq \left(\frac{n+m}{n-m}\right) \varepsilon. \quad (3.20)$$

Using the result in Theorem 3.1, we obtain:

Corollary 3.3. Let the window function be given as in (3.19) with $m < n < N - m$. Then the worst case noise error as defined in Lemma 2.5 can be bounded by

$$e_N^{\text{noise}} \leq \varepsilon \sqrt{\frac{n+m}{n-m}}. \quad (3.21)$$

(Note that the above bound is strictly smaller than the one as reported in [5].)

Proof. Since the worst case noise error as defined in Lemma 2.5 is shift invariant, we need only to consider following window function:

$$w_{n,k} = a_{n,k+m} \quad (3.22)$$

which is even-symmetric with respect to k . It is noted that the window function above can be written as the difference of two triangular window functions,

$$w_{n,k} = \frac{n}{n-m} t_{n,k} - \frac{m}{n-m} t_{m,k}, \quad (3.23)$$

where $t_{n,k}$ is given in (3.16). Using trigonometric identities, the Fourier transform of the above window can be written as

$$|K_n(\omega)| = \left| \frac{nF_n(\omega) - mF_m(\omega)}{n-m} \right| = \sqrt{\left(\frac{n+m}{n-m}\right) F_{n+m}(\omega) F_{n-m}(\omega)}, \quad (3.24)$$

where F_k is the k -th order Fejér kernel. The fact that F_k is nonnegative implies that

$$\sqrt{F_{n+m}(\omega)F_{n-m}(\omega)} \leq \frac{1}{2}(F_{n-m}(\omega) + F_{n-m}(\omega)). \quad (3.25)$$

Using (3.3), we have

$$e_N(\hat{\eta}) \leq \frac{\varepsilon}{2} \sqrt{\frac{n+m}{n-m}} \sup_{\omega} \left\{ \frac{1}{N} \sum_{i=0}^{N-1} \left(F_{n+m} \left(\omega - \frac{2i\pi}{N} \right) + F_{n-m} \left(\omega - \frac{2i\pi}{N} \right) \right) \right\}. \quad (3.26)$$

Hence, by Lemma 3.2, the error estimate as in (3.21) is easily obtained. \square

Corollary 3.4. *Let the window function be given as in (3.19) with $m < n < N - m$. Let $E_{2m}(\hat{h})$ and δ_{2m} be defined similarly as in (2.7) and Definition 2.1. Then, the worst case identification error is bounded by*

$$e^N(\mathcal{S}, \varepsilon) \leq 2 \sqrt{\frac{n+m}{n-m}} (\delta_{2m} + \varepsilon) + 2\delta_{2m} \quad (3.27)$$

where $e^N(\mathcal{S}, \varepsilon)$ is defined as in (2.5).

Proof. For every $\hat{h} \in \mathcal{S}$, there exists a $\hat{p}_{2m}^* \in \mathcal{P}_{2m}$ where \mathcal{P}_{2m} is the collection of all polynomials with degree not larger than $2m$ such that

$$\hat{h} = \hat{p}_{2m}^*(\hat{h}) + \hat{\zeta}_{2m} \quad \text{and} \quad \|\hat{\zeta}_{2m}\|_{\infty} = E_{2m}(\hat{h}). \quad (3.28)$$

Now take the ‘true’ system to be $\hat{p}_{2m}^*(\hat{h})$, and the noise bound to be $\delta_{2m} + \varepsilon$. Then the ‘worst case approximation error’ is zero. Using Corollary 3.3, it now follows that the ‘worst case noise error’ is no larger than

$$\sqrt{\frac{n+m}{n-m}} (\delta_{2m} + \varepsilon). \quad (3.29)$$

Further, by noting that

$$\|\hat{h} - \hat{p}_{2m}^*(\hat{h})\|_{\infty} = E_{2m}(\hat{h}) \leq \delta_{2m}, \quad (3.30)$$

the error bound in (3.27) follows by using Lemma 2.4. \square

The above result indicates that if the window function (3.19) is used in two-stage nonlinear algorithm, the computation of the worst case identification error bound requires the computation of δ_{2m} which is in fact a minimax functional optimization problem (see Definition 2.1). It is noted that in many cases, an upper bound on δ_{2m} can be easily estimated.

References

- [1] H. Akçay, G. Gu and P.P. Khargonekar, A class of algorithms for identification in \mathcal{R}_{∞} : continuous-time case, *IEEE Trans. Automat. Control* (Nov. 1992, to appear).
- [2] V.M. Adamyán, D.Z. Arov and M.G. Krein, Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur–Takagi problem, *Math. USSR Sbornik* **15** (1971) 31–73.
- [3] R.E. Edwards, *Fourier Series, A Modern Introduction* (Holt, Rinehart and Winston, New York, 1967).
- [4] G. Gu and P.P. Khargonekar, Linear and nonlinear algorithms for identification in \mathcal{R}_{∞} with error bounds, to appear in *IEEE Trans. Automatic Control*. Abridged version in *Proceedings of the American Control Conference* (1991) 64–69.
- [5] G. Gu and P.P. Khargonekar, A class of algorithms for system identification in \mathcal{R}_{∞} , to appear in *Automatica*. Abridged version in *Proceedings of the 30th IEEE Conference on Decision and Control* (1991) 634–639.
- [6] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Identification in \mathcal{R}_{∞} : A robustly convergent nonlinear algorithm, Presented at the 1989 American Control Conference and the 1989 International Symposium on Mathematical Theory of Networks and Systems, *Proceedings of the American Control Conference* (1990) 386–391.

- [7] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Control-oriented system identification: A worst-case/deterministic approach in \mathcal{H}_∞ , *IEEE Trans. Automat. Control* **36** (1991) 1163–1176.
- [8] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Identification in \mathcal{H}_∞ : linear algorithms, *Proceedings of the American Control Conference* (1990) 2418–2423.
- [9] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Identification in \mathcal{H}_∞ : the continuous-time case, *Proceedings of the American Control Conference* (1990) 1893–1898. To appear in the *IEEE Trans. Automat. Control*.
- [10] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Fundamentals of control oriented system identification and their application for identification in \mathcal{H}_∞ , *Proceedings of the American Control Conference* (1990) 89–99.
- [11] K. Hoffman, *Banach Spaces of Analytic Functions* (Prentice-Hall, Englewood Cliffs, NJ, 1962).
- [12] R.L. Kosut, M. Lau and S. Boyd, Identification of systems with parametric and nonparametric uncertainty, *Proceedings of the American Control Conference* (1990) 2412–2417.
- [13] J.M. Krause and P.P. Khargonekar, Parameter identification in the presence of nonparametric dynamic uncertainty, *Automatica* **26** (1990) 113–124.
- [14] J.M. Krause, P.P. Khargonekar and G. Stein, Robust parameter adjustment with nonparametric weighted-ball-in- \mathcal{H}_∞ uncertainty, *IEEE Transactions on Automatic Control* **35** (1990) 225–229.
- [15] M. Lau, R.L. Kosut and S. Boyd, Parameter set identification of systems with uncertain nonparametric dynamics and disturbances, *Proceedings of the 29th IEEE Conference on Decision and Control* (1990) 3162–3167.
- [16] L. Ljung, *System Identification, Theory for the User* (Prentice-Hall, Englewood Cliffs, NJ, 1987.)
- [17] L. Ljung and Z.-D. Yuan, Asymptotic properties of black-box identification of transfer functions, *IEEE Trans. Automat. Control* **30** (1985) 514–530.
- [18] P.M. Mäkilä, On Laguerre methods and \mathcal{H}_∞ identification of continuous-time systems, to appear in the *Internat. J. Control*.
- [19] P.M. Mäkilä, On \mathcal{H}_∞ identification of stable systems and optimal approximation, to appear in *Automatica*.
- [20] P.M. Mäkilä and J.R. Partington, Robust approximation and identification in \mathcal{H}_∞ , *Proceedings of the American Control Conference* (1991) 70–76.
- [21] M. Milanese, R. Tempo and A. Vicino, Strongly optimal algorithms and optimal information in estimation problems, *Journal of Complexity* **2** (1986) 78–94.
- [22] Z. Nehari, On bounded bilinear forms, *Ann. of Math.* **65** (1957) 153–162.
- [23] P.J. Parker and R.R. Bitmead, Adaptive frequency response identification, *Proceedings of the 28th IEEE Conference on Decision and Control* (1987) 348–353.
- [24] J.R. Partington, Robust identification and interpolation in \mathcal{H}_∞ , to appear in the *Internat. J. Control*.
- [25] J.R. Partington, Robust identification in \mathcal{H}_∞ , to appear in the *J. Math. Anal. Appl.*
- [26] J.R. Partington and P.M. Mäkilä, Robust identification of stabilizable systems, Preprint, University of Leeds, 1991, in: *Proceedings of the 30th IEEE Conference on Decision and Control* (1991) 629–633.
- [27] R. Smith and J.C. Doyle, Towards a methodology for robust parameter identification, *Proceedings of the American Control Conference* (1990) 2394–2399.
- [28] R. Tempo and G. Wasilkowski, Maximum likelihood estimators and worst case optimal algorithms for system identification, *Systems and Control Letters* **10** (1988) 265–270.
- [29] D.N.C. Tse, M.A. Dahleh and J.N. Tsitsiklis, Optimal asymptotic identification under bounded disturbances, Preprint, LIDS, MIT, Cambridge, MA. Abridged version in *Proceedings of the American Control Conference* (1991) 1786–1787.
- [30] R.C. Younce and C.E. Rohrs, Identification with non-parametric uncertainty, *Proceedings of the 29th IEEE Conference on Decision and Control* (1990) 3154–3161.
- [31] N.J. Young, *An Introduction to Hilbert Space* (Cambridge University Press, Cambridge, 1988).
- [32] A. Zygmund, *Trigonometric Series* (Cambridge University Press, Cambridge, 1959).