A Class of Algorithms for Identification in $\mathscr{H}_\infty$*

GUOXIANG GU† and PRAMOD P. KHARGONEKAR‡

Inverse fast Fourier transform, window functions, and the Nehari approximation theorem are used to generate a family of untuned and robustly convergent algorithms for frequency response based identification with guaranteed error bounds in the $\mathscr{H}_\infty$ norm.

Abstract—In this paper, a class of algorithms for the problem of system identification in $\mathscr{H}_\infty$ are investigated. These algorithms are characterized by a two-stage structure and involve a class of window functions. Some conditions in terms of properties of the window functions are derived, which guarantee robust convergence of the algorithms. Identification errors are analyzed for several common window functions. This leads to some insights into the trade-off between the error induced by approximation and that due to noise.

1. INTRODUCTION

RECENTLY, a problem "system identification in $\mathscr{H}_\infty$" has been formulated by Helmicki *et al.* (1990, 1991a) which is a concrete instance of an abstract worst case deterministic identification problem formulation by the same authors (1991b). In this approach, experimental data are taken to be noisy values of the frequency response of the system at a given set of frequencies. Roughly speaking, the problem is to find algorithms which map the experimental data into an identified model for which the worst case (with respect to the noise and the unknown system) identification error converges to zero in the $\mathscr{H}_\infty$ norm as the noise level goes to zero and the number of data points goes to infinity. This problem is motivated by the fact that it matches the framework and assumptions underlying modern robust control design techniques which give worst case stability and performance guarantees provided modeling uncertainty is quantified in terms of error bounds in the $\mathscr{H}_\infty$

norm. The interested reader is referred to Helmicki *et al.* (1990, 1991a) for a thorough and detailed discussion of this problem formulation.

Several nonlinear algorithms have been proposed, by Helmicki *et al.* (1990, 1991a), Gu and Khargonekar (1990), Partington (1992a, b) for solving the problem of $\mathscr{H}_\infty$ identification recently, which are robustly convergent. (See Section 2 for a precise definition.) Helmicki *et al.* (1990, 1991a) gave a robustly convergent algorithm based on spline interpolation. Their algorithm does converge robustly, but the order of the identified model is N^3 where N is the number of data points. In a recent paper (Gu and Khargonekar, 1990), we gave a robustly convergent algorithm based on the classical Cesaro sum technique where the order of the identified model was allowed to be less than N. We also gave a linear algorithm which does not converge robustly, but it was shown that the worst case error diverges very slowly. More recently, Partington (1992a) has given some nonlinear algorithms which are related to the one by Gu and Khargonekar (1990) and share similar properties. In addition, he has shown that there does not exist a linear robustly convergent algorithm for this specific problem of identification in $\mathscr{H}_\infty$. The paper by Mäkilä (1991) contains a different perspective on $\mathscr{H}_\infty$ identification.

In this paper, we analyze a class of nonlinear algorithms for the problem of identification in $\mathscr{H}_\infty$. This class includes the algorithms proposed in Helmicki *et al.* (1990, 1991a), Gu and Khargonekar (1990), Partington (1992a, b) as special cases. The algorithms in this class are characterized by a two-stage structure: the first stage involves taking the inverse discrete Fourier transform and multiplication by a suitable window function; the second stage involves finding the best analytic approximation using

Nehari's theorem (Nehari, 1957; Adamyan et al., 1971) to the function obtained at stage 1. For each choice of the window function, one gets an identification algorithm. Indeed, the choice of the window function is crucial and determines completely the properties of the resulting identification algorithm. [The use of window functions here is reminiscent of the digital filter design literature (Rabiner and Gold, 1975).]

The principal results of this paper give some general conditions on the window function which ensure that the resulting two-stage nonlinear algorithm is robustly convergent. Thus, one gets a large collection of robustly convergent algorithms for the problem of system identification in $\mathcal{H}_\infty$. We give results on the worst case identification error in terms of the properties of the window function.

It turns out that the worst case identification error can be broken into two parts: the error caused by noise and the error caused by partial data. (For reasons that will become clear later, we call the second component "error induced by approximation".) It is desirable to reduce both of these errors simultaneously. Our results indicate that this is in general not possible and there is a trade-off between these two types of errors. This trade-off between the error induced by approximation and that caused by noise for some specific window functions is analyzed. A parametrized window is proposed which incorporates this trade-off as a design parameter and therefore appears to be useful for system identification in $\mathcal{H}_\infty$. The identification algorithm generated by this window function is shown to be exponentially convergent. In a parallel and independent work, Partington (1992b) has employed the de la Vallee Poussin kernel to obtain an exponentially convergent algorithm.

System identification is a well established field of research with a large body of literature, see for example, Ljung (1987). The traditional approach to system identification is to take a stochastic problem formulation. In some recent papers, results on the resulting identification error have been obtained by Ljung and Yuan (1985), Goodwin et al. (1990), Goodwin and Salgado (1989). For deterministic worst case error bounds as used in robust control, the identification problem is more appropriately formulated in a deterministic setting. Recently, some papers that take a deterministic approach to the identification problem have appeared, see for example, Kosut et al. (1990), Krause et al. (1990), Krause and Khargonekar (1990), Lau et al. (1990), Parker and Bitmead (1987), Smith and Doyle (1990), Tse et al. (1991), Younce and Rohrs (1990) and the references cited there. Many of these papers deal with what may be broadly termed as "robust identification". We regard our paper as a contribution to the field of robust identification.

It is possible to formulate other identification problems with $\mathcal{H}_\infty$ norm as a measure of identification error. From this point of view, it is perhaps better to keep the terminology "identification in $\mathcal{H}_\infty$" for a generic class of identification problem in which $\mathcal{H}_\infty$ norm is used as a measure of identification error. At this time, no fixed and well established terminology is in place, and therefore, we will use the phrase identification in $\mathcal{H}_\infty$ to mean the specific problem described at the beginning of this introduction. It is hoped that as the field of robust identification develops further, a logical and coherent terminology will emerge.

As noted above, the experimental data for the problem treated in this paper are taken to be noisy values of the frequency response of system at a finite set of frequencies. And it is assumed that a deterministic bound on the noise is available. A relevant question here is how to obtain such experimental data. This issue has been discussed in great detail by Helmicki et al. (1990, 1991a) and some experimental plans are proposed to obtain such data. Also, in certain types of applications, it is common to perform sine-sweep experiments. From an engineering point of view, such noise bounds should be based on the characteristics of measuring instruments as well as knowledge of the physical system. It is felt that more research on this issue should be quite useful. Also, more research on relating other types of "real data" to these assumptions is necessary. The other alternative would be to modify the problem formulation.

In this paper, we will restrict our attention to single-input/single-output discrete-time linear time-invariant systems. Extensions to the multi-input/multi-output case are immediate. The rest of the paper is organized as follows. In Section 2, we describe the two-stage algorithm and give some preparatory results. The key conclusion of this section is that stage 1 is crucial in determining robust convergence as well as the worst case identification error. In Section 3, we give the main results of the paper. Here we establish some general conditions on the window function which guarantee robust convergence. We also give some results on worst case identification error. Some special classes of windows are analyzed in Section 4. Section 5 contains results of some numerical experiments.

2. TWO-STAGE NONLINEAR ALGORITHMS

In this section, we will define the $\mathcal{H}_\infty$ identification problem as formulated by Hel-

micki *et al.* (1990, 1991a), and then summarize briefly some existing results. Some preliminary results will also be given. The books by Hoffman (1962), Zygmund (1959) are excellent sources for the required mathematical background.

Let $\mathcal{H}_\infty$ 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 $\mathcal{H}_\infty$ norm of $\hat{f}$ is defined as

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

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

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

As is well known, $\mathcal{A}$ is a closed subalgebra of $\mathcal{H}_\infty$.

Let the transfer function $f(z) = \sum_0^\infty h_k z^k$ where $h_k, k = 0, 1, \ldots$ is the impulse response of the system. A causal l_2 BIBO-stable discrete-time shift-invariant system can be viewed as an element in $\mathcal{H}_\infty$. 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 also belongs to $\mathcal{A}$. Note that stability here corresponds to having no poles inside the closed unit disc.

The problem of identification in $\mathcal{H}_\infty$ formulated by Helmicki *et al.* (1990, 1991a) can now be described as follows.

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

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

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

where

$$B_N(\epsilon) := \{\hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_N) \in \mathscr{C}^N : |\hat{\eta}_k| \le \epsilon$$
$$\text{for } k = 1, 2, \ldots, N\}; \quad (2.4)$$

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

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

satisfies

$$\lim_{\epsilon \to 0, N \to \infty} e^N(\epsilon) = 0. \quad (2.6)$$

It is natural to require that the identified model be real rational. This is ensured by

requiring that (for n even) $E_{N/2+1+k}^N = [E_{N/2+1-k}^N]^*$ for $k = 1, 2, \ldots, N/2 - 1$ and E_1^N, $E_{N/2+1}^N$ being real. This is always true of real data. In this sense, one only has $N/2 + 1$ independent frequency response data points. The ball of noise $B_N(\epsilon)$ should thus be taken to be those complex numbers which satisfy the above complex conjugate symmetry.

The set $\mathcal{S}$ captures the prior information on the system being identified. An important example of $\mathcal{S}$ is the collection of all exponentially stable systems $\mathcal{H}(\mathcal{D}_\rho, M)$ as defined below

$$\mathcal{H}(\mathcal{D}_\rho, M) := \{\hat{f} : \hat{f} \text{ is analytic in } \mathcal{D}_\rho$$
$$\text{and } |\hat{f}| \le M, \forall z \in \mathcal{D}_\rho\},$$

where $\mathcal{D}_\rho = \{z : |z| < \rho\}$, and $\rho > 1$.

The noise bound in (2.4) is assumed to be independent of frequency. As discussed in Helmicki *et al.* (1990, 1991a), this assumption is without loss of generality. If the noise bound is frequency dependent, then one can use a weighting function to transform the original problem into the one where the noise bound is independent of frequency.

As in Helmicki *et al.* (1990, 1991a), we define robust convergence as follows.

Definition 1. An identification algorithm A_N is said to be convergent if the worst case identification error $e_N(\epsilon)$ satisfies (2.6). Further, if in addition the identification algorithm A_N does not depend on *a priori* information on $\mathcal{S}$ and is convergent, then A_N is said to be (untuned and) robustly convergent.

As mentioned in the introduction, several robustly convergent nonlinear algorithms have been proposed by Helmicki *et al.* (1990, 1991a), Gu and Khargonekar (1990), Partington (1992a, b). These algorithms share the following "two-stage" structure. At the first stage inverse discrete Fourier transform and a window function is 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 (Nehari, 1957; Adamyan *et al.*, 1971) is used to approximate this first stage approximation to obtain an analytic, i.e. stable, identified model.

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

$$\tilde{h}_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} E_{i+1}^N(\hat{h}, \hat{\eta}) e^{-jk(2i\pi/N)}, \quad j = \sqrt{-1}, \quad (2.7)$$

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

as follows:

$$\bar{h}_N(k) = \bar{h}_N(k \pm LN), \quad k, L \in \mathscr{I} \quad (2.8)$$

where $\mathscr{I}$ denotes the set of integers.

In this paper, we will study identification algorithms having the following structure.

Two-stage nonlinear algorithm

Stage 1. Let $n(N)$ be a given monotone nondecreasing function from positive integers to positive integers such that $\lim_{N \to \infty} n(N) = \infty$. Let $w_{n,k} : k = 0, \pm 1, \ldots, \pm n$ be a given sequence called a "window function" of real numbers. Define pre-identified model $\hat{h}_{pi}^{E;N}$ as

$$\hat{h}_{pi}^{E;N} = \sum_{k=-n}^{n} w_{n,k} \bar{h}_N(k) z^k. \quad (2.9)$$

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

$$\hat{h}_{id}^{N} := \arg\min \{ \|\hat{h}_{pi}^{E;N} - \hat{f}\|_\infty : \hat{f} \in \mathscr{H}_\infty \}. \quad (2.10)$$

(Whenever there is no possibility of confusion, to avoid cumbersome notation, we will not show explicitly the dependence of n on N.)

Clearly, the identification algorithm is completely characterized by the window function used in stage 1. Stage 2 involves solving the optimization problem defined in (2.10). Since the pre-identified model is not necessarily causal as it may have nonzero negative Fourier coefficients, stage 2 requires finding the best analytic approximation for the pre-identified model. This is the so-called Nehari best approximation problem which has been analyzed extensively in the mathematics literature. See, for example, Adamyan *et al.* (1971), Nehari (1957) and Young (1988) for a complete analysis and methods for solving this problem. Suffice it to say that there are well established algorithms for obtaining $\hat{f}$ given $\hat{h}_{pi}^{E;N}$ which solves (2.10). Moreover, since $\hat{h}_{pi}^{E;N}$ is rational (in fact, it is a trigonometric polynomial) the resulting $\hat{h}_{id}^{N}$ is also a rational function with no poles inside the closed unit disc.

Now the algorithms in Helmicki *et al.* (1990, 1991a), Gu and Khargonekar (1990), Partington (1992a, b) can be taken to be special cases of the above two-stage algorithm. In particular, the window function for the spline based identification is

$$w_{n,0} = 1, \quad w_{n,k} = \left(\frac{N}{k\pi}\right)^2 \left(\sin \frac{k\pi}{N}\right)^2,$$

$$k \neq 0, \quad \text{and} \quad w_{n,k} = 0 \quad \text{for } |k| > n. \quad (2.11)$$

The window function for Cesaro sum based identification proposed in Gu and Khargonekar

(1990) is

$$w_{n,k} = 1 - \frac{|k|}{n}, \quad \text{and} \quad w_{n,k} = 0 \quad \text{for } |k| \geq n \quad (2.12)$$

which is a triangular window. For the identification algorithm in Partington (1992a) which uses "second Bernstein procedure", the window function employed in stage 1 is of the following form

$$w_{n,k} = \cos (k\pi/(2n + 1)), \quad \text{and} \quad w_{n,k} = 0$$

$$\text{for } |k| > n. \quad (2.13)$$

It is seen that the only free parameter in the above two-stage nonlinear algorithm is the window function $w_{n,k}$ used in (2.9). It will be shortly shown that stage 1 plays a crucial role in determining robust convergence of the two-stage nonlinear algorithm. To make this precise, we introduce the following definition.

Definition 2. Stage 1 of the two-stage algorithm is said to be robustly convergent if

$$e_{pi}^{N}(\epsilon) := \sup \{ \|\bar{h} - \hat{h}_{pi}^{E;N}\|_\infty : \bar{\eta} \in B_N(\epsilon), \bar{h} \in \mathscr{S} \} \quad (2.14)$$

satisfies

$$\lim_{\epsilon \to 0, N \to \infty} e_{pi}^{N}(\epsilon) = 0. \quad (2.15)$$

We now have the following lemma that shows that the first stage of the algorithm indeed determines the properties of the identification algorithm. (This result is essentially contained in Helmicki *et al.*, 1990.)

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

$$e^{N}(\epsilon) \leq 2e_{pi}^{N}(\epsilon). \quad (2.16)$$

Proof. Using Nehari's best approximation result (Nehari, 1957), stage 2 of the nonlinear algorithm in (10) yields

$$\|\hat{h}_{id}^{N} - \hat{h}_{pi}^{E;N}\|_\infty = \|\hat{h}_{pi}^{E;N}\|_H \leq \|\hat{h}_{pi}^{E;N} - \bar{h}\|_\infty, \quad (2.17)$$

where $\|\hat{f}\|_H$ denotes the norm of the Hankel operator defined by $\hat{f}$. By the triangle inequality,

$$\|\hat{h}_{id}^{N} - \bar{h}\|_\infty \leq \|\hat{h}_{id}^{N} - \hat{h}_{pi}^{E;N}\|_\infty + \|\hat{h}_{pi}^{E;N} - \bar{h}\|_\infty$$

$$\leq 2 \|\hat{h}_{pi}^{E;N} - \bar{h}\|_\infty. \quad (2.18)$$

By taking supremum on both sides of the above inequality as in (2.5) and limits as in (2.6), we conclude that the robust convergence of stage 1 implies the robust convergence of the two-stage nonlinear algorithm as well as the worst case identification error inequality (2.16). $\square$

To analyze the identification error $e_{pi}^N(\epsilon)$ in stage 1, note that $\bar{h}_N(k)$ as in (2.7) can be written as

$$\bar{h}_N(k) = h_N(k) + \eta_N(k) \qquad (2.19)$$

where

$$h_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} h(e^{j(2i\pi/N)}) e^{-jk(2i\pi/N)}$$

and $$\eta_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\eta}_{i+1} e^{-jk(2i\pi/N)}. \qquad (2.20)$$

The pre-identified model in stage 1 of the nonlinear algorithm as in (2.9) can thus be written as

$$\hat{h}_{pi}^{E;N} = \hat{h}_{pi}^N + \hat{\eta}_{pi}^N \qquad (2.21)$$

where

$$\hat{h}_{pi}^N = \sum_{k=-n}^{n} w_{n,k} h_N(k) z^k$$

and $$\hat{\eta}_{pi}^N = \sum_{k=-n}^{n} w_{n,k} \eta_N(k) z^k \qquad (2.22)$$

with $h_N(k)$, $\eta_N(k)$ defined in (2.20). Hence, $\hat{h}_{pi}^N$ can be thought of as a pre-identified model in stage 1 for noise free case and $\hat{\eta}_{pi}^N$ as the effect of noise.

Lemma 2. Stage 1 of the identification algorithm is robustly convergent if and only if

(i) $\displaystyle \lim_{\epsilon \to 0, \, N \to \infty} \sup \{ \|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon) \} = 0$;

(ii) $\displaystyle \lim_{N \to \infty} \sup \{ \|\hat{h}_{pi}^N - h\|_\infty : h \in \mathcal{S} \} = 0$.

That is, stage 1 is robustly convergent if and only if $h = 0 \in \mathcal{S}$ can be identified robustly and for any system $h \in \mathcal{S}$, h can be identified exactly with $\epsilon = 0$.

Proof. By the triangle inequality and using (2.21)

$$\sup \{ \|\hat{h}_{pi}^{E;N} - h\|_\infty : \hat{\eta} \in B_N(\epsilon), \, h \in \mathcal{S} \}$$
$$\leq \sup \{ \|\hat{h}_{pi}^N - h\|_\infty : h \in \mathcal{S} \}$$
$$+ \sup \{ \|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon) \}. \qquad (2.23)$$

Hence, if (i) and (ii) are satisfied, then the robust convergence of stage 1 follows. Conversely, since $h = 0 \in \mathcal{S}$, the robust convergence of stage 1 implies that (i) is true. Moreover, by choosing $\hat{\eta}_k = 0$, robust convergence of stage 1 also implies that (ii) must hold which completes the proof. $\qquad \square$

Note that the worst case identification error $e_{pi}^N(\epsilon)$ is bounded by sum of two terms on the right hand side of (2.23). It seems appropriate to call the first term (on right hand side of (2.23))

$$\sup \{ \|\hat{h}_{pi}^N - h\|_\infty : h \in \mathcal{S} \}, \qquad (2.24)$$

worst case approximation error which corresponds to the noise free case (a number of people have worked on such approximation problem; see the survey paper by Henrici, 1979), and the second term [on the right hand side of (2.23)]

$$\sup \{ \|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon) \}, \qquad (2.25)$$

worst case noise error which corresponds to the pure noise case, i.e. $h = 0$.

Ideally, one would like to find an algorithm that simultaneously minimizes both the worst case approximation error and the worst case noise error. However, it will be shown by examining the algorithms proposed in Helmicki *et al.* (1990), Gu and Khargonekar (1990), Partington (1992a) in Section 4 that there appears to be a conflict between these two objectives. We would like to emphasize that it is not the purpose of this paper to compare different algorithms proposed in Helmicki *et al.* (1990), Gu and Khargonekar (1990), Partington (1992a). The objective here is to study the properties of the two-stage nonlinear algorithms within the framework of window functions.

In view of Lemmas 1 and 2, it is clear that the window function employed in stage 1 determines the robust convergence of the two-stage nonlinear algorithm as well as the worst case identification error. It turns out that different window functions used in (2.9) can be studied in terms of the triangular window as defined in (2.12) which will be shown in next section. The following result is taken from Gu and Khargonekar (1990).

Lemma 3. Let $w_{n,k}$ be the triangular window given by (2.12) with $n < N$. Then for any $h \in \mathcal{A}$,

$$\lim_{N > n \to \infty} \sup \{ \|\hat{h}_{pi}^{E;N} - h\|_\infty : \hat{\eta} \in B_N(\epsilon) \} = \epsilon. \qquad (2.26)$$

Thus, in the noise-free case, the pre-identified model $\hat{h}_{pi}^N$ converges to the true system transfer function h as $N \to \infty$ for the triangular window. However, this convergence need not guarantee robust convergence even in the noise free case, since we are interested in the worst case error measured in infinity norm taken over all $h \in \mathcal{S}$. Indeed it is easy to see that additional boundedness assumptions on $\mathcal{S}$ are needed in order to even expect the robust convergence of the two-stage nonlinear algorithm.

Define P_n to be the collection of all polynomials in z with degree no 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 from Zygmund (1959) that for each $\hat{h} \in \mathcal{A}$, there exists a $\hat{p}_n^* \in P_n$ such that

$$E_n(\hat{h}) = \|\hat{h} - \hat{p}_n^*\|_\infty$$

$$= \inf \{\|\hat{h} - \hat{p}_n\|_\infty : \hat{p}_n \in P_n\}. \quad (2.27)$$

Clearly, such a $\hat{p}_n^*$ is an optimal approximation of $\hat{h}$ in the set P_n and the optimal error as in (2.27) is a function of $\hat{h}$. Note that $E_n(\hat{h})$ is a monotone nonincreasing function of n.

Definition 3. A subset $\mathcal{S} \subseteq \mathcal{A}$ is called admissible, if

$$\lim_{m \to \infty} \delta_m = 0, \quad \text{where} \quad \delta_m := \sup \{E_m(\hat{h}) : \hat{h} \in \mathcal{S}\};$$

$$\text{and} \quad M_s := \sup \{\|\hat{h}\|_\infty : \hat{h} \in \mathcal{S}\} < \infty. \quad (2.28)$$

Roughly speaking, admissibility requires that the set of systems be bounded and be uniformly approximative by polynomials. It can be shown that admissibility is equivalent to requiring that $\mathcal{S}$ be totally bounded in $\mathcal{H}_\infty$. This concept is also related to some issues pertaining to metric complexity discussed by Zames (1979). We will assume that $\mathcal{S}$ is an admissible set in the remainder of this paper. It is noted that the collection of exponentially stable systems $\mathcal{S} := \mathcal{H}(\mathcal{D}_\rho, M)$ as discussed in Helmicki *et al.* (1990, 1991a) and Gu and Khargonekar (1990) is admissible. Other sets such as in Partington (1992a) having certain "smoothness" properties are also admissible.

It is shown in Helmicki *et al.* (1990, 1991a) that the worst case identification error can not be smaller than ϵ. Hence, the triangular window in (2.12) is in fact an optimal window in terms of minimizing error due to noise for identification. To distinguish the triangular window in (2.12) from other windows, we adopt the following notation for pre-identified model $\hat{h}_{pi}^{E;N}$ when the window function in (2.12) is used:

$$\hat{j}_{n,N}(z) := \hat{h}_{pi}^{E;N} = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right)\bar{h}_N(k)z^k. \quad (2.29)$$

The polynomial $\hat{j}_{n,N}$ is called the Jackson polynomial in Zygmund (1959, Vol. II, 22). We arrived at it in Gu and Khargonekar (1990) using the idea of Cesaro sums. It is noted that

$$\hat{j}_{n,N} = \frac{1}{n} \sum_{i=0}^{n-1} \hat{i}_i \quad (2.30)$$

where $\hat{i}_i$ is the partial sum

$$\hat{i}_i = \sum_{k=-i}^{i} \bar{h}_N(k)z^k \quad (2.31)$$

with $\bar{h}_N(k)$ defined as in (2.7).

3. CHARACTERIZATION OF WINDOW FUNCTIONS FOR ROBUST CONVERGENCE

In this section, we study robust convergence of the two-stage nonlinear algorithm. This analysis involves a study of the robust convergence of stage 1 of the two-stage algorithm in terms of properties of the window function. As in Zygmund (1959, Vol. I, p. 74), consider a doubly infinite matrix of complex numbers

$$M = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,n} & \cdots \\ a_{1,0} & a_{1,1} & \cdots & a_{1,n} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n,0} & a_{n,1} & \cdots & a_{n,n} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \quad (3.1)$$

With every sequence of complex numbers $s_0, s_1, s_2, \ldots$, we associate the sequence σ_n given by

$$\sigma_n := \sum_{k=0}^{\infty} a_{n,k}s_k, \quad (n = 0, 1, 2, \ldots), \quad (3.2)$$

provided the series on the right converges for all n. Suppose

$$N_n = \sum_{k=0}^{\infty} |a_{n,k}|, \quad \text{and} \quad A_n = \sum_{k=0}^{\infty} a_{n,k}, \quad (3.3)$$

exist and are finite for all n. The following result can be found in Zygmund (1959, Vol. I, 74).

Lemma 4. Let the sequence $a_{n,k}$ be as above and suppose

(i) $\lim_{n \to \infty} a_{n,k} = 0$ for $k = 0, 1, \ldots$;

(ii) the sequence N_n as in (3.3) is bounded; and

(iii) $\lim_{n \to \infty} A_n = 1$.

If the sequence s_k tends to a finite limit s as k goes to infinity, then σ_n tends to s as n approaches infinity.

For the window function $w_{n,k}$ used in the two-stage nonlinear algorithm, let

$$\Delta w_{n,k} := w_{n,k} - w_{n,k+1},$$

$$\text{and} \quad \Delta^2 w_{n,k} := \Delta w_{n,k} - \Delta w_{n,k+1}. \quad (3.4)$$

Suppose that the window function $w_{n,k}$ is even symmetric with respect to k, i.e. $w_{n,-k} = w_{n,k}$ for all n, k. Then, with above notation, the pre-identified model as in (2.9) can be written as

$$\hat{h}_{pi}^{E;N} = \sum_{k=-n}^{n} w_{n,k}\bar{h}_N(k)z^k = \sum_{k=0}^{n-1} \Delta w_{n,k}\hat{i}_k + w_{n,n}\hat{i}_n$$

$$= \sum_{k=0}^{n-2} (k+1)\Delta^2 w_{n,k}\hat{j}_{k+1,N}$$

$$+ n\Delta w_{n,n-1}\hat{j}_{n,N} + w_{n,n}\hat{i}_n, \quad (3.5)$$

by applying summation by part twice. The next

result concerns the worst case noise error as defined in (2.25).

Lemma 5. Suppose that the window function used in stage 1 of the nonlinear algorithm is even symmetric with respect to k. Then, there exists an absolute constant C_1 such that for all $N > n > 0$, the worst case noise error as defined in (2.25) satisfies

$$\sup \{ \|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon) \}$$

$$\leq \left(n \, |\Delta w_{n,n-1}| + \sum_{k=0}^{n-2} (k+1) \, |\Delta^2 w_{n,k}| \right) \epsilon$$

$$+ C_1 \, |w_{n,n}| \log (n)\epsilon. \tag{3.6}$$

Further, if $w_{n,n} = 0$ and $\Delta^2 w_{n,k}$ has the same sign as $\Delta w_{n,n-1}$ for $k = 0, 1, \ldots, n-2$, then the above error bound is attained for some $\hat{\eta} \in B_N(\epsilon)$.

Proof. As shown by Gu and Khargonekar (1990),

$$\sup \{ \|\hat{J}_{k,N}\|_\infty : \hat{\eta} \in B_N(\epsilon), \, \hat{h} = 0 \} \leq \epsilon, \tag{3.7}$$

for all possible $k < N$. Also, it follows from Zygmund (1959, Vol. II, p. 37) that there exists an absolute constant C_1 such that

$$\sup \{ \|\hat{I}_n\|_\infty : \hat{\eta} \in B_N(\epsilon), \, \hat{h} = 0 \} \leq C_1 \log (n)\epsilon. \tag{3.8}$$

Hence, the error bound in (3.6) follows from (3.5) by setting $\hat{h} = 0$ and taking norms of both sides. Further, if $w_{n,n} = 0$, then the upper bound on the right hand side of (3.7) is attained by taking $\hat{\eta}_i = \epsilon$ for all i which yields $\hat{J}_{k,N}(e^{j\omega}) = \epsilon$, for all $\omega \in [-\pi, \pi]$ and $k = 0, 1, \ldots, n-2$. Therefore, the right hand side of (3.6) is also attainable if $\Delta^2 w_{n,k}$ has the same sign as $\Delta w_{n,n-1}$ for all possible k which concludes the proof. $\square$

The next theorem is the main result of this section.

Theorem 1. Suppose that the window function $w_{n,k}$ is even symmetric with respect to k, i.e. $w_{n,k} = w_{n,-k}$ for all $n, k, n(N) < N/2$ and $\lim_{n \to \infty} w_{n,n} \log (n) = 0$. Then, stage 1 of the two-stage nonlinear algorithm is robustly convergent if the window function $w_{n,k}$ satisfies

(i) $\lim_{n \to \infty} \Delta^2 w_{n,k} = 0$ for $k = 0, 1, \ldots$;

(ii) $N_s := \lim \sup \left\{ N_n := n \, |\Delta w_{n,n-1}| \right.$

$$\left. + \sum_{k=0}^{n-2} (k+1) \, |\Delta^2 w_{n,k}| : n \geq 0 \right\} < \infty;$$

(iii) $\lim_{n \to \infty} w_{n,0} = 1$.

Consequently, if the above conditions hold, then the two-stage nonlinear identification algorithm is robustly convergent.

Proof. First note that Lemma 5, in conjunction with the hypotheses, implies that the worst case noise error as defined in (2.25) converges to zero as ϵ goes to zero and n goes to ∞. It now follows from Lemmas 1 and 2 that the robust convergence of the two-stage identification algorithm is guaranteed if we can establish the robust convergence of the worst case approximation error as defined in (2.24). We may thus set the noise level $\epsilon = 0$ for the remainder of the proof.

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

$$\hat{h} = \hat{p}_m^* + \xi_m \quad \text{with} \quad \|\xi_m\|_\infty = \|\hat{h} - \hat{p}_m^*\|_\infty = E_m(\hat{h}), \tag{3.9}$$

where $m > 0$ is an integer, and $\hat{p}_m^*$, $E_m(\hat{h})$ are defined as in (2.27) and (2.28). Hence, even in this noise free case, we may consider each $\hat{h} \in \mathcal{S}$ as a sum of $\hat{p}_m^*$ and a "noise" term $\xi_m = \hat{h} - \hat{p}_m^*$ with "noise level" $E_m(\hat{h})$. Let p_k be the coefficient of z^k in $\hat{p}_m^*$. Then,

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{p}_m^*(e^{j\omega}) e^{-jk\omega} \, d\omega. \tag{3.10}$$

Hence,

$$M_p(\hat{h}) := \sup \{ |p_k| : 0 \leq k \leq m \} \leq \|\hat{p}_m^*\|_\infty \leq 2 \|\hat{h}\|_\infty. \tag{3.11}$$

Now,

$$\|\hat{h} - \hat{h}_{pi}^N\|_\infty \leq E_m(\hat{h}) + \|\hat{p}_m^* - \hat{h}_{pi}^N\|_\infty. \tag{3.12}$$

Using the decomposition in (3.9), the inverse discrete Fourier transform in (2.20) can be written as

$$h_N(k) = p_N(k) + \zeta_N(k), \tag{3.13}$$

where

$$p_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{p}_m^*(W_N^i) W_N^{-ik} \quad \text{and}$$

$$\zeta_N(k) = \frac{1}{N} \sum_{i=0}^{N-1} \xi_m(W_N^i) W_N^{-ik}, \tag{3.14}$$

with $W_N = e^{j2\pi/N}$. It is noted that $\hat{p}_M^*$ is a polynomial in z of degree no larger than m. Hence, if $N > 2m$, the polynomial $\hat{p}_m^*$ can be recovered with inverse discrete Fourier transform. In particular, with $N > 2n \geq 2m$, we have

$$\hat{h}_{pi}^N - \hat{p}_m^* = \sum_{k=0}^{m} (w_{n,k} - 1)p_k z^k + \sum_{k=-n}^{n} w_{n,k}\zeta_N(k)z^k, \tag{3.15}$$

where p_k is the coefficient of z^k in $\hat{p}_m^*$. By applying Lemma 5 to ξ, the above can be bounded as

$$\|\hat{h}_{pi}^N - \hat{p}_m^*\|_\infty \le (N_n + C_1 |w_{n,n}| \log(n)) E_m(\hat{h})$$

$$+ \sum_{k=0}^m |1 - w_{n,k}| M_p(\hat{h}), \quad (3.16)$$

where $M_p(\hat{h})$ is defined in (3.11) and N_n is given in condition (ii) in the statement of the theorem.

To show the robust convergence of worst case approximation error, we need only to show that for any given $\delta > 0$, there exists an $N_1 > 0$, such that for all $N/2 > n \ge N_1$ (recall that N is the number of frequency response data points), the worst case approximation error as defined in (2.24) is less than δ. This is indeed true. Using the fact that $\mathcal{S}$ is admissible and definition of δ_m combined with (3.11) and (3.16), it follows that

$$\sup \{\|\hat{h}_{pi}^N - \hat{p}_m^*\|_\infty : \hat{h} \in \mathcal{S}\}$$

$$\le (N_n + C_1 w_{n,n} \log(n)) \delta_m$$

$$+ 2 \sum_{k=0}^m |1 - w_{n,k}| M_s. \quad (3.17)$$

Recall that δ_m is a non-negative monotonically nonincreasing function of m with limit 0. Therefore, there exists an $M_\delta > 0$ such that with $m = M_\delta$, $4(N_s + 1)\delta_m < \delta$. By the hypothesis $\lim_{n \to \infty} w_{n,n} \log(n) = 0$, we can find an $N_c > 0$ such that for all $n \ge N_c$, $C_1 w_{n,n} \log(n) \le N_s$ and $N_n < N_s + 1$. Moreover, in view of conditions (i) and (iii), an integer $N_w > 0$ can also be found such that for all $N \ge n \ge N_w$, $4(m + 1)M_s |1 - w_{n,k}| < \delta$ for $k = 0, 1, \ldots, m = M_\delta$. Set $N_1 = \max \{N_w, 1 + M_\delta, N_c\}$. Then for all N such that $N/2 > n \ge N_1$, it follows from (3.12) that the worst case approximation error satisfies

$$\sup \{\|\hat{h}_{pi}^N - \hat{h}\|_\infty : \hat{h} \in \mathcal{S}\}$$

$$\le \delta_m + \sup \{\|\hat{h}_{pi}^N - \hat{p}_m^*\|_\infty : \hat{h} \in \mathcal{S}\}$$

$$\le 2(N_s + 1)\delta_m + \delta/2 < \delta. \quad (3.18)$$

Thus, the worst case approximation error also converges to zero robustly. As we have already noted, the worst case noise error also converges to zero robustly as appropriate limits are taken. The proof is complete using Lemma 2.   □

It is noted that Theorem 1 shows only the sufficiency of conditions (i)–(iii). With some additional assumptions on set $\mathcal{S}$ and window function $w_{n,k}$, conditions (i)–(iii) also become necessary. For doing this, the following notion is useful.

Definition 4. A set $\mathcal{S}$ is called nontrivial if for

$k = 0, 1, 2, \ldots$, $\sup \{|h_k| : \hat{h} \in \mathcal{S}\} \ne 0$ where h_k is the kth component of the impulse response of $\hat{h}$.

Proposition 1. Suppose that the set $\mathcal{S}$ is nontrivial. Assume that the window function $w_{n,k}$ is even symmetric with $n < N/2$, and in addition, $n |\Delta w_{n,n-1}| \to 0$, $w_{n,n} \log(n) \to 0$, as $n \to \infty$ and there exists an $m > 0$ such that $\Delta^2 w_{n,k}$ has same sign $\forall k \ge m$ and all $n > 0$. Then, conditions (i)–(iii) are necessary for robust convergence of worst case identification error as defined in (2.14).

Proof. By Lemma 2, the robust convergence in (2.14) implies the robust convergence of the worst case approximation error as in (2.24) and the worst case noise error as in (2.25). The nontriviality assumption on $\mathcal{S}$ and the robust convergence of worst case approximation error imply that for each $\hat{h} \in \mathcal{S}$, the sequence $\hat{h}_N(k)$ as in (2.7) must converge to the true impulse response h_k as $N, n \to \infty$. It follows that for each k, $w_{n,k} \to 1$ as $n \to \infty$. Hence, conditions (i) and (iii) of Theorem 1 are true. Further, by setting noise $\hat{\eta}_k = \epsilon \forall k$, we have that

$$\hat{\eta}_{pi}^N = \sum_{k=0}^{n-2} (k + 1) \Delta^2 w_{n,k} \epsilon$$

$$+ (n \Delta w_{n,n-1} + C_1 w_{n,n} \log(n)) \epsilon, \quad (3.19)$$

with $\hat{\eta}_{pi}^N$ defined in (2.22). By the hypothesis on the window function, for $n - 2 \ge m$, we obtain that

$$\sup \{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\}$$

$$\ge \sum_{k=m}^{n-2} (k + 1) |\Delta^2 w_{n,k}| \epsilon - \sum_{k=0}^{m-1} (k + 1) |\Delta^2 w_{n,k}| \epsilon$$

$$- |(n \Delta \omega_{n,n-1} + C_1 w_{n,n} \log(n))| \epsilon. \quad (3.20)$$

This holds for all n. Therefore, taking limit as $N > 2n \to \infty$, and using the robust convergence of noise error we get that,

$$\sum_{k=0}^\infty (k + 1) |\Delta^2 w_{n,k}| < \infty. \quad (3.21)$$

This, along with the hypothesis, implies that condition (ii) of Theorem 1 is also necessary.   □

As an application of Theorem 1, we would like to analyze the window function (2.11) as used in Helmicki *et al.* (1990, 1991a) next. While it was assumed in Helmicki *et al.* (1990, 1991a) that $n \approx N^3$, we would like to investigate the problem of robust convergence when the window function (2.11) is used with $n \approx N/2$. Set $n(N)$ to be the largest integer no greater than $N/2$. Clearly, the window function in (2.11)

can be obtained by sampling the continuous function

$$w(t) = (\sin t/t)^2, \quad w_{n,k} = w(t)$$

with

$$t = \frac{2k\pi}{N}, \quad k \neq 0, \quad |k| \leq n(N).$$

It is rather easy to see that conditions (i) and (iii) of the Theorem 1 hold for this window function in (2.11). So, we only need to verify that condition (ii) holds in order to show the robust convergence of this algorithm. Now consider the twice differentiable function $w(t) = (\sin t/t)^2$. We have

$$\frac{d^2 w(t)}{dt^2} = 2\left(\frac{2\sin t - 2t\cos t - t^2 \sin t}{t^3}\right)^2$$
$$\times \frac{(\sin t)^2}{t^2} + 2\frac{(t\cos t - \sin t)^2}{t^4}. \quad (3.22)$$

It is not difficult to verify that,

$$\sup\left\{\left|t\frac{d^2 w(t)}{dt^2}\right| : t \in [0, \pi]\right\} = C_1 < \infty \quad (3.23)$$

so that

$$\frac{1}{\pi}\int_0^\pi \left|t\frac{d^2 w(t)}{dt^2}\right| dt \leq C_1 < \infty. \quad (3.24)$$

Some elementary calculations then show that the constant N_s can be bounded by $C_1 + C_2$ for some other constant C_2 since $n\Delta w_{n,n-1}$ goes to zero as $n \to \infty$. Hence, the robust convergence of the nonlinear algorithm proposed in Helmicki et al. (1990, 1991a) for the admissible set $\mathcal{S} = \mathcal{H}(\mathcal{D}_\rho, M)$ follows from Theorem 1 if $n \approx N/2$. This result is surprising since in Helmicki et al. (1990, 1991a), $n \approx N^3$ is needed in order to show the robust convergence when the window function (2.11) is used.

The above idea of sampling twice differentiable functions to arrive at the window function $w_{n,k}$ is quite common in the digital filtering literature. That is, suppose the window function can be written as $w_{n,k} = w(k\alpha/n)$ where $k \leq n$ and $w(t)$ is even symmetric and a twice differentiable function in $[-\alpha, \alpha]$ such that $w(0) = 1$, $w(\alpha) = 0$. Then $w_{n,n} \log(n) = 0$, and $n|\Delta w_{n,n-1}| \to \alpha |w'(\alpha)|$. In this case, it is not difficult to verify that the condition (ii) of Theorem 1 can be replaced by

$$N_\infty = \frac{1}{\alpha}\int_0^\alpha \left|t\frac{d^2 w(t)}{dt^2}\right| dt < \infty. \quad (3.25)$$

The above condition is often much easier to check than the condition (ii) of Theorem 1. As another application of this idea, consider a different window function which is related to an ideal low pass filter:

$$w_{n,k} = \frac{n}{k\pi}\sin\frac{k\pi}{n}, \quad \text{and} \quad w_{n,k} = 0 \quad \text{for } |k| > n.$$
$$(3.26)$$

The above window is obtained by sampling the function $w(t) = \sin t/t$ with $\alpha = \pi$. It is easy to show that

$$\frac{d^2 w(t)}{dt^2} = \frac{2\sin t - 2t\cos t - t^2 \sin t}{t^3}. \quad (3.27)$$

By using Taylor expansions for $\sin(t)$ and $\cos(t)$, we have that

$$\frac{2}{\pi}\int_0^\pi \left|\frac{\sin t - t\cos t}{t^2}\right| dt$$
$$= \frac{2}{\pi}\int_0^\pi \left|\sum_{k=1}^\infty (-1)^{k+1}\frac{2kt^{2k-1}}{(2k+1)!}\right| dt$$
$$\leq \frac{2}{3\pi}\int_0^\pi t \, dt = \frac{\pi}{3}. \quad (3.28)$$

Hence,

$$N_\infty = \frac{1}{\pi}\int_0^\pi \left|t\frac{d^2 w(t)}{dt^2}\right| dt$$
$$\leq \frac{1}{\pi}\int_0^\pi \left(2\left|\frac{\sin t - t\cos t}{t^2}\right| + |\sin t|\right) dt$$
$$\leq \frac{2}{\pi} + \frac{\pi}{3}. \quad (3.29)$$

Further, for the window function in (3.26), conditions (i) and (iii) of Theorem 1 are satisfied. We can thus conclude that the two-stage identification algorithm corresponding to the window function in (3.26) is robustly convergent.

4. ANALYSIS AND TRADE-OFF OF IDENTIFICATION ERRORS

In this section, we will analyze the identification error as well as the trade-offs between the approximation error and the noise error as defined in (2.24) and (2.25). For system identification in $\mathcal{H}_\infty$, the performance index of interest is the worst case identification error measured in the $\mathcal{H}_\infty$ norm. By Lemma 1, the worst case identification error is bounded by twice of the worst case identification error in the first stage of the nonlinear algorithm. It is thus important to investigate the identification error in stage 1 as defined in (2.23). We will first consider the noise error which is the worst case identification error for $h = 0$ case as discussed in Lemma 2.

The error bound in (3.6) is not very explicit. However, if additional information on the window function $w_{n,k}$ is available, the error

bound in (3.6) can be made more explicit. The following concepts of convex and concave windows are useful.

Definition 5. The window function $w_{n,k}$ is called convex at k if $\Delta^2 w_{n,k} \geq 0$, and called a convex function if $\Delta^2 w_{n,k} \geq 0$ for all $k \geq 0$. The window function $w_{n,k}$ is called concave at k, if $\Delta^2 w_{n,k} \leq 0$ and called a concave function if $\Delta^2 w_{n,k} \leq 0$ for all $k \geq 0$. The window function $w_{n,k}$ is called nonincreasing if $\Delta w_{n,k} \geq 0$ for all $k \geq 0$.

With this definition, we can state the following result.

Theorem 2. Let the window function $w_{n,k}$ used in (2.9) be even symmetric and satisfy $w_{n,n} = 0$ which is nonincreasing and $n < N$. The following error bounds hold.

(i) If $w_{n,k}$ is a convex function, then

$$\sup\{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\} \leq w_{n,0}\epsilon; \quad (4.1)$$

(ii) If $w_{n,k}$ is a concave function, then

$$\sup\{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\} \leq (2nw_{n,n-1} - w_{n,0})\epsilon; \quad (4.2)$$

(iii) If $w_{n,k}$ is convex for $k < m(<n)$ and concave for $k \geq m$, then

$$\sup\{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\} \leq (w_{n,0} - 2(m+1)w_{n,m} + 2mw_{n,m+1} + 2nw_{n,n-1})\epsilon; \quad (4.3)$$

(iv) If $w_{n,k}$ is concave for $k < m(<n)$ and convex for $k \geq m$, then

$$\sup\{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\}$$
$$\leq (2(m+1)w_{n,m} - 2mw_{n,m+1} - w_{n,0})\epsilon. \quad (4.4)$$

Further, the error bounds in (4.1) and (4.2) are attainable.

Proof. It is straightforward to verify the above bounds. Hence the details of the proof are omitted.

Remark 1. Using the results in Proposition 1, it is clear that for the window functions considered above, the three conditions in Theorem 1 are both necessary and sufficient for robust convergence of the two-stage nonlinear algorithm.

It is noted that the triangular window in (2.12) is a concave function. Further, $w_{n,n} = 0$ and $\Delta w_{n,k} = 1/n$ for $0 < k \leq n$. Hence, the worst case noise error as defined in (2.25) never exceeds ϵ. case noise error as defined in (2.25) never exceeds ϵ.

The results presented in Theorem 2 exclude a large class of window functions because of the restriction $w_{n,n} = 0$. For example, Hamming window is clearly excluded. In order to apply the

results in Theorem 2 for the case $w_{n,n} \neq 0$, we can simply add a term $w_{n,n}M_n$ into the error bounds given in Theorem 2, where $M_n = \sup\{\|\hat{L}_n\|_\infty : \hat{h} \in \mathcal{S}\} \leq C_1 \log(n)M_s$ with M_s as given in Definition 3. Although $w_{n,n} \neq 0$ implies that the corresponding window function is not guaranteed to lead to a robustly convergent nonlinear algorithm, it is possible to reduce the worst case identification error for fixed order n. The reason is that by allowing nonzero $w_{n,n}$, a smaller value of N_n as in condition (ii) of Theorem 1 might be achieved. This is especially true if the order n is not too large and hence $\log(n)$ is relatively small.

For example, let us consider the cosine-window as in (2.13), which is a concave function. In this case, although $w_{n,n} \neq 0$, as noted above we can still apply the results of Theorem 2. Since $w_{n,n} \log(n)$ converges to zero as $n \to \infty$, for n large enough we can neglect the contribution of the $w_{n,n}$ term. Thus, for n large enough, the worst case noise error as defined in (2.25) is approximately $(2n \cos((n-1)\pi/(2n+1)) - 1)\epsilon$ by (ii) of Theorem 2. In the limiting case $n \to \infty$, the noise error converges to $(\pi - 1)\epsilon$ which is more than twice as large as the noise error for triangular window. This worst case noise error is actually attainable in light of Lemma 5. Hence, if the noise level ϵ is large, the identification error will be large as observed in Partington (1992a). However, this should not be taken to mean that the cosine-window as in (2.13) is without merit. In fact, if the noise level ϵ is small, the cosine-window gives better performance than the triangular window.

Recall the definition of the set $\mathcal{H}(\mathcal{D}_\rho, M)$. It is a simple exercise to show that for any $\hat{h} \in \mathcal{H}(\mathcal{D}_\rho, M)$, its impulse response h_k satisfies the Cauchy estimate

$$|h_k| \leq M\rho^{-k}, \quad k \geq 0. \quad (4.5)$$

Theorem 3. Let $\mathcal{S} = \mathcal{H}(\mathcal{D}_\rho, M)$. Let the window function employed for the two-stage nonlinear algorithm be the cosine-window as given in (2.13) with $n < N$. Then the worst case identification error satisfies

$$e^N(\epsilon) \leq \frac{2M\rho}{\rho - 1}\left(2\rho^{-n} + \frac{(\rho + 1)\pi^2}{2(2n + 1)^2(\rho - 1)^2}\right)$$
$$+ 2\left(2n \cos\frac{(n-1)\pi}{2n+1} - 1\right)$$
$$+ C_1 \cos\frac{n\pi}{2n+1}\log(n)\right)\epsilon, \quad (4.6)$$

for some absolute constant C_1. Consequently, the identification algorithm obtained by taking the cosine window is robustly convergent.

Proof. We consider the worst case approximation error as defined in (2.24) first. Substituting $h_N(k)$ in (2.20) into $\hat{h}_{pi}^N$ as given in (2.22), we have

$$\|\hat{h}_{pi}^N - \hat{h}\|_\infty \leq \sum_{k=-n+1}^{n-1} \left| \cos\left(\frac{k\pi}{2n+1}\right) h_N(k) \right.$$
$$\left. - h_k \right| + \sum_{k=n}^{\infty} |h_k|, \qquad (4.7)$$

where $h_k = 0$ for $k < 0$ since $\hat{h} \in \mathcal{H}_\infty$. It follows from Henrici (1979), Gu *et al.* (1989) that

$$h_N(k) = \sum_{L=0}^{\infty} h_{LN+k}. \qquad (4.8)$$

By noting that $\cos(x) \leq 1$, we obtain

$$\|\hat{h}_{pi}^N - \hat{h}\|_\infty \leq \sum_{k=0}^{n-1} \left(1 - \cos\left(\frac{k\pi}{2n+1}\right)\right) |h_k|$$
$$+ 2 \sum_{k=n}^{\infty} |h_k|$$
$$\leq 2M \sum_{k=n}^{\infty} \rho^{-k}$$
$$+ 2M \sum_{k=0}^{n-1} \sin^2\left(\frac{k\pi/2}{2n+1}\right)\rho^{-k}. \qquad (4.9)$$

Since $\sin(\tfrac{1}{2}k\pi/(2n+1)) \leq \tfrac{1}{2}k\pi/(2n+1)$, it is now straightforward to obtain the bound for worst case approximation error

$$\sup \{\|\hat{h}_{pi}^N - \hat{h}\|_\infty : \hat{h} \in \mathcal{H}(\mathcal{D}_\rho, M)\}$$
$$\leq \frac{M\rho}{\rho-1} \left(2\rho^{-n} + \frac{(\rho+1)\pi^2}{2(2n+1)^2(\rho-1)^2}\right). \qquad (4.10)$$

To estimate the worst case noise error as defined in (2.25), we can use Lemma 5. Since the cosine window is a concave function, by (ii) of Theorem 2 we get

$$n |\Delta w_{n,n-1}| + \sum_{k=0}^{n-2} (k+1)| \Delta^2 w_{n,k}|$$
$$= 2n \cos\frac{(n-1)\pi}{2n+1} - 1. \qquad (4.11)$$

Hence, Lemma 5 implies that the worst case noise error at stage 1 is bounded by

$$\sup \{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\}$$
$$\leq \left(n \cos\frac{(n-1)\pi}{2n+1} - 1 + C_1 \cos\frac{n\pi}{2n+1} \log(n)\right)\epsilon \qquad (4.12)$$

for some absolute constant C_1. Since the worst case identification error at stage 1 is bounded by the sum of the worst case approximation error in (2.24) and the worst case noise error in (2.25), the error bound (4.6) is obtained using (4.10),

(4.12) and Lemma 1. The final conclusion follows by taking limit as $N > n \to \infty$. $\square$

It is shown by Gu and Khargonekar (1990) that the triangular window has a convergence rate $\mathcal{O}(1/n)$ for the worst case approximation error. The cosine-window actually achieves a faster convergence rate $\mathcal{O}(1/n^2)$. However, the faster convergence for worst case approximation error is obtained at the expense of increased worst case noise error.

Another widely used window is the one-sided rectangular window as studied by Gu and Misra (1991) and Parker and Bitmead (1987):

$$w_{n,k} = 1, \quad 0 \leq k \leq n \quad \text{and} \quad w_{n,k} = 0, \text{ otherwise.}$$
$$(4.13)$$

It is known that the worst case approximation error as defined in (2.24) converges to zero exponentially (Gu and Misra, 1991; Parker and Bitmead, 1987). However, the worst case noise error diverges in the order of $\log(n)$ (Gu and Misra, 1991).

In light of the above analysis for different window functions, we believe that it is in general not possible to minimize both the approximation error and the noise error as defined in (2.24) and (2.25) simultaneously. In fact, minimizing one seems to amplify the other. Therefore, the use of different window functions should reflect the trade-off between the approximation error and the noise error. Further research should be devoted to developing parametrized window functions which may offer such trade-off directly by choosing parameter values. Perhaps something similar to the Kaiser window for FIR filter design (Rabiner and Gold, 1975) may be very useful.

One such window is the mixture of one-sided rectangular window and triangular window as defined below:

$$w_{n,k} = 1, \quad 0 \leq k \leq 2m, \quad w_{n,k} = \frac{n+m-k}{n-m},$$

$$2m \leq k \leq n+m, \quad \text{and} \quad w_{n,k} = 1 + \frac{k}{n-m},$$
$$(4.14)$$

$$m - n \leq k \leq 0, \quad \text{and} \quad w_{n,k} = 0, \text{ otherwise,}$$

with $0 < m < n < N$. The above window is not even-symmetric with respect to k, and therefore does not fit directly into the framework of the results derived in the previous section. However, the shifted window $w_{n,k-m}$ is even-symmetric and satisfies all the properties of window functions required for the results derived earlier to apply. It is noted that if $m \to n$, the above window

function approaches the one-sided rectangular window which yields exponential convergence rate for the approximation error if $\mathcal{S} = \mathcal{H}(\mathcal{D}_\rho, M)$. On the other hand, if $m \to 0$, then it approaches the triangular window as in (2.12) whose noise error does not exceed ϵ (Gu and Khargonekar, 1990). Clearly, the value of parameter m shows the trade-off between the approximation error and the noise error. Moreover, the following error bound can be established.

Theorem 4. Let the window function be given as in (4.14) with $m < n < N$ and let $\mathcal{S} = \mathcal{H}(\mathcal{D}_\rho, M)$. Then the worst case identification error is bounded by

$$e^N(\epsilon) \leq \frac{2M}{\rho - 1}(\rho^{-(N-1)} + \rho^{-(N-n+m-1)} + 2\rho^{-2m})$$
$$+ \left(\frac{2(n+m)}{n-m}\right)\epsilon. \qquad (4.15)$$

Proof. The worst case approximation error as defined in (2.24) can be bounded by

$$\sup \{\|\hat{h}_{pi}^N - \hat{h}\|_\infty : \hat{h} \in \mathcal{S}\}$$
$$\leq \sum_{k=0}^{2m} |h_N(k) - h_k| + \sum_{k=m-n}^{-1} |h_N(k)|$$
$$+ \sum_{k=2m+1}^{n+m} |h_N(k)| + \sum_{k=2m+1}^{\infty} |h_k|, \quad (4.16)$$

where $h_N(k)$ is given as in (2.20). Using the exponential stability condition in (4.5) and equation (4.8), we obtain

$$\sup \{\|\hat{h}_{pi}^N - \hat{h}\|_\infty : \hat{h} \in \mathcal{S}\}$$
$$\leq \frac{M}{\rho - 1}(\rho^{-(N-1)} + \rho^{-(N-n+m-1)} + 2\rho^{-2m}).$$
$$(4.17)$$

To estimate the worst case noise error as defined in (2.25), we first shift the window function in (4.14) to obtain

$$\bar{w}_{n,k} := w_{n,k-m}. \qquad (4.18)$$

It is noted that the window function $\bar{w}_{n,k}$ is even-symmetric. Moreover, it is easy to show that the worst case noise error for $\bar{w}_{n,k}$ is the same as that for the window function in (4.14). Further, the shifted window function $\bar{w}_{n,k}$ is a concave function. Hence, using (ii) of Theorem 2, we obtain

$$\sup \{\|\hat{\eta}_{pi}^N\|_\infty : \hat{\eta} \in B_N(\epsilon)\} \leq \left(\frac{2n}{n-m} - 1\right)\epsilon$$
$$= \left(\frac{n+m}{n-m}\right)\epsilon. \quad (4.19)$$

The error bound in (4.15) is thus established in view of (2.23) and Lemma 1.   □

Now suppose we choose m in the window function (4.14) such that $m/n \approx \beta$ where $0 < \beta < 1$ is fixed. For each n, m is taken to be the largest integer no greater than $n\beta$. Then we see that the resulting window function leads to a robustly convergent algorithm for the problem of identification in $\mathcal{H}_\infty$ for the admissible set $\mathcal{H}(\mathcal{D}_\rho, M)$ with the important property that the approximation error (which is the first part of the right hand side of (4.15)) decays exponentially in n as n goes to infinity. Needless to say, since the noise error is always at least as large as ϵ, we can not expect it to decay with increasing n. Now as β becomes closer to one, the approximation error decreases, but the noise error amplification increases. Thus, β allows one to trade noise error versus approximation error.

It is noted that the window function as in (4.14) has an important property. Let $\hat{p}_{2m}^*$ be the best approximant to the true model $\hat{h}$ of degree no larger than $2m$ as defined in (2.27). Then, as in the proof of Theorem 1 with m replaced by $2m$,

$$\hat{h} = \hat{p}_{2m}^* + \hat{\zeta}_{2m}, \quad \text{with} \quad \hat{\zeta}_{2m} := \hat{h} - \hat{p}_{2m}^*. \quad (4.20)$$

Corollary 1. Let $\hat{h}$ be in $\mathcal{A}$, and let $E_{2m}(\hat{h})$ and $\hat{p}_{2m}^*$ be defined as in (2.27) with $m < n < N/2$. Let the window function be given as in (4.14). Then, the pre-identified model $\hat{h}_{pi}^N$ as given in (2.22) for the noise-free case satisfies

$$\|\hat{h}_{pi}^N - \hat{p}_{2m}^*\|_\infty \leq \left(\frac{n+m}{n-m}\right)E_{2m}(\hat{h}). \quad (4.21)$$

Proof. Take a fictitious "true model" to be $\hat{p}_{2m}^*$ and $\hat{\zeta}_{2m}$ as the noise. Then the result follows from (4.19).   □

The above result shows some advantages for using the window function as defined in (4.14). In particular, the approximation error will be close to the optimal value $E_{2m}(\hat{h})$. However, the near-optimality of approximation error is achieved by amplifying the noise error. The trade-off is thus clear in terms of the parameter m.

In many engineering applications, it is desirable to keep the order of the identified model small. It is interesting to notice that if the order of the identified model n is fixed, then one may use the above analysis to minimize the worst case identification error by choosing an appropriate m. This can be approached by minimizing the right hand side of (4.15) to obtain the optimal value of m. It should also be

pointed out that the straight line with nonzero slope in the window function (4.14) can be replaced by any smooth curve satisfying the condition (ii) of Theorem 1. However, the identification error may be difficult to analyze in this more general situation.

5. AN ILLUSTRATIVE EXAMPLE

To demonstrate the trade-off between worst case approximation error as in (2.24) and worst case noise error as in (2.25), we consider identification of the following transfer function

$$h = \frac{z^2 + 1}{z^2 + 2z + 2}. \tag{5.1}$$

It is assumed that the number of frequency response data available is $N = 128$ with noise level $\epsilon = 0.1$. At stage 1 of the nonlinear algorithm, the window function proposed in (4.14) is used with $n = 20$ and different values of m. Using identified models $\hat{h}_{id}^N$ at stage 2 of the nonlinear algorithm, we have plotted magnitude of the frequency response of $h - \hat{h}_{id}^N$ versus normalized frequencies for $m = 1$, $m = 4$ and $m = 10$ in Fig. 1. It is interesting to note that with $m = 1$, the approximation error dominates the identification error and with $m = 10$, the noise error dominates the identification error. Further, with $m = 4$, the identification error is smallest possible. We would like to point out that the identification error depicted in Fig. 1 is not necessarily the worst case error since we have used $\epsilon e^{j\pi\theta}$ with θ a random variable uniformly distributed on $[-1, 1]$ to generate the noise. However, the simulation does give an indication on the trade-off of worst case approximation error and noise error in terms of the parameter m.

We have also done numerical simulation of the identification algorithm for different window functions using the same transfer function as in (5.1) and the same noise model with $N = 128$
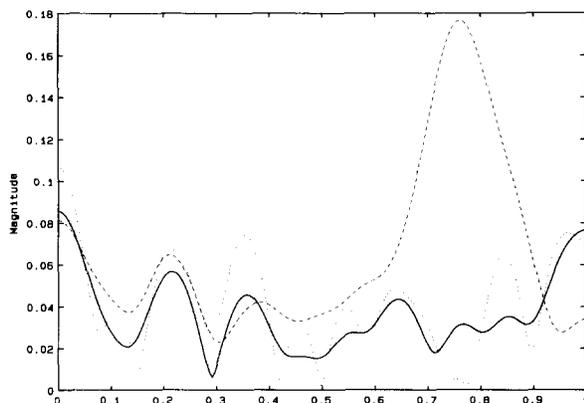

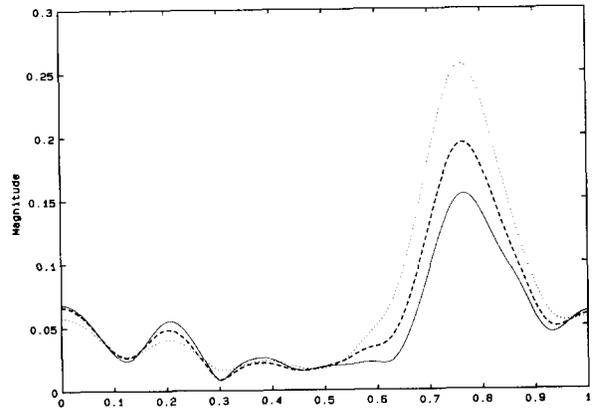FIG. 1. Magnitude response of identification errors. — $m = 4$, --- $m = 1$, ···· $m = 10$.

FIG. 2. Magnitude response of identification errors.

and $n = 20$. The first window function is the Hamming window

$$w_{n,k} = 0.54 + 0.46 \cos\left(\frac{k\pi}{n}\right), \quad \text{and}$$

$$w_{n,k} = 0 \quad \text{for } |k| > n. \tag{5.2}$$

The second window is a modification (N replaced by n) of (2.11)

$$w_{n,0} = 1, \quad w_{n,k} = \left(\frac{n}{k\pi}\right)^2 \left(\sin\frac{k\pi}{n}\right)^2,$$

$$k \neq 0, \quad \text{and} \quad w_{n,k} = 0 \quad \text{for } |k| \geq n. \tag{5.3}$$

The third window is as in (3.26). The magnitude frequency response of the identification error is plotted in Fig. 2 with a solid line for the window function in (3.26), dashed line for Hamming window and dotted line for the window function in (5.3). We find that the window function as in (3.26) achieves the smallest identification error. Since no trade-off between approximation error and noise error is offered with the three window functions above, as expected, the identification error for these windows is larger than the window function proposed in (4.14).

6. CONCLUSIONS

In this paper we have studied a class of algorithms for identification in $\mathcal{H}_\infty$ using window functions. A number of conditions for robust convergence are established. Further, the trade-off between different window functions in identification is investigated. The results reported in this paper may provide guidelines for choosing window functions when the two-stage nonlinear algorithm is used for identification in $\mathcal{H}_\infty$.

REFERENCES

Adamyan, V. M., D. Z. Arov and M. G. Krein (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem. *Math. USSR Sbornik*, **15**, 31–73.

Goodwin, G. C., B. Ninnes and M. E. Salgado (1990). Quantification of uncertainty in estimation. *Proc. 1990 American Control Conference*, San Diego, CA, pp. 2400–2405.

Goodwin, G. C. and M. E. Salgado (1989). A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *Int. J. Adaptive Control Signal Process*, **3**, 333–356.

Gu, G. and P. P. Khargonekar (1990). Linear and nonlinear algorithms for identification in $\mathcal{H}_\infty$ with error bounds. *IEEE Trans. Aut Control*, (to appear).

Gu, G., P. P. Khargonekar and E. B. Lee (1989). Approximation of infinite dimensional systems. *IEEE Trans. Aut. Control*, **34**, 610–618.

Gu, G. and P. Misra (1991). Modeling and identification of linear systems. *Proc. 1991 American Control Conf*, Boston, MA, pp. 687–692.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1990). Identification in $\mathcal{H}_\infty$: a robust convergent nonlinear algorithm. *Proc. 1990 American Control Conf.*, pp. 386–391.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1991a). Control-oriented system identification: A worst-case/deterministic approach in $\mathcal{H}_\infty$. *IEEE Trans. Aut. Control*, **36**, 1163–1176.

Helmicki, A. J., C. A. Jacobson, and C. N. Nett (1991b). Fundamentals of control-oriented system identification and their application for identification in $\mathcal{H}_\infty$. *Proc. 1991 American Control Conf.*, Boston, MA, pp. 89–99.

Henrici, P. (1979). Fast Fourier methods in computational complex analysis. *SIAM Rev.*, **21**, 481–527.

Hoffman, K. (1962). *Banach Spaces of Analytic Functions*. Prentice-Hall, Englewood Cliffs, NJ.

Kosut, R. L., M. Lau and S. Boyd (1990). Identification of systems with parametric and non-parametric uncertainty. *Proc. 1990 American Control Conf.* San Diego, CA, pp. 2412–2417.

Krause, J. M., P. P. Khargonekar and G. Stein (1990). Robust parameter adjustment with nonparametric weighted-ball-in-$\mathcal{H}_\infty$ uncertainty. *IEEE Trans. Aut. Control*, **AC-35**, 225–229.

Krause, J. M. and P. P. Khargonekar (1990). Parameter identification in the presence of non-parametric dynamic uncertainty. *Automatica*, **26**, 113–124.

Lau, M., R. L. Kosut and S. Boyd (1990). Parameter set identification of systems with uncertain nonparametric dynamics and disturbances. *Proc. 1990 IEEE Conf. on Decision and Control.* Honolulu, HI, pp. 3162–3167.

Ljung L. (1987). *Systems Identification, Theory for the User.* Prentice-Hall, Englewood Cliffs, NJ.

Ljung, L and Z.-D. Yuan (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Aut. Control*, **AC-30**, 514–530.

Mäkilä, P. M. (1990). Laguarre series approximation of infinite dimensional systems. *Automatica*, **26**, 985–995.

Mäkilä, P. M. (1991). On $\mathcal{H}_\infty$ identification of stable systems and optimal approximation. *Automatica*, **27**, 663–676.

Middleton, R. H., G. C. Goodwin and Y. Wang (1989). On the robustness of adaptive controllers using relative dead zones. *Automatica*, **25**, 889–896.

Nehari, Z. (1957). On bounded bilinear forms. *Ann. Math.*, **65**, 153–162.

Parker, P. J. and R. R. Bitmead (1987). Adaptive frequency response identification. *Proc. 28th IEEE Conf. on Decision and Control*, pp. 348–353.

Partington, J. R. (1992a). Robust identification in $\mathcal{H}_\infty$. *J. Math. Anal. Applic.* (to appear).

Partington, J. R. (1992b). Robust identification and interpolation in $\mathcal{H}_\infty$. *Int. J. Control*, (to appear).

Rabiner, L. R. and B. Gold (1975). *Theory and Applications of Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ.

Smith, R. and J. C. Doyle (1990). Towards a methodology for robust parameter identification. *Proc. 1990 American Control Conf.*, San Diego, CA, pp. 2394–2399.

Tse, D. N. C., M. A. Dahleh, and J. N. Tsitsiklis (1991). Optimal asymptotic identification under bounded disturbances. Preprint, LIDS, MIT, Cambridge, MA.

Younce, R. C. and C. E. Rohrs (1990). Identification with non-parametric uncertainty. *Proc. 1990 IEEE Conf. on Decision and Control*, Honolulu, HI, pp. 3154–3161.

Young, N. J. (1988). *An Introduction to Hilbert Space.* Cambridge University Press, Cambridge.

Zames, G. (1979). On the metric complexity of causal linear systems: ϵ-entropy and ϵ-dimension for continuous time. *IEEE Trans. Aut. Control*, **AC-24**, 222–230.

Zygmund, A. (1959). *Trigonometric Series.* Cambridge University Press, Cambridge.