The Nehari Shuffle: FIR(q) Filter Design with Guaranteed Error Bounds

Peter J. Kootsookos, Robert R. Bitmead, *Fellow, IEEE,* and Michael Green

Abstract—**This paper presents a new approach to the problem of designing a finite impulse response filter of specified length, q, which approximates in uniform frequency (L_∞) norm a given desired (possibly infinite impulse response) causal, stable filter transfer function. We derive an algorithm-independent lower bound on the achievable approximation error and then present an approximation method which involves the solution of a fixed number of all-pass (Nehari) extension problems and so is called the Nehari shuffle. Upper and lower bounds on the approximation error are derived for the algorithm. These bounds are calculable *a priori* so the length of filter required to satisfy a given maximum error can be found before designing the filter. Examples indicate that the method closely approaches the derived global lower bound. We compare the new method with the Preuss (complex Remez exchange) algorithm in some examples.**

I. FIR Filter Design

Finite impulse response (FIR) transfer functions are widely used because of their good numerical properties and the ease of implementation. Since magnitude response specifications may frequently be met with filters of lower order than FIR designs, infinite impulse response (IIR) filters are also used. This improved design performance of IIR filters occurs at the expense of other factors, such as group delay [1] and phase [2]–[4] specifications.

We present a method that is a direct approach to the problem of approximating a desired IIR transfer function by an FIR(q) design with an error criterion being the maximum magnitude of the error frequency response over all frequencies. Our algorithm has the following properties:

1) The algorithm approximates the magnitude and phase responses of the desired transfer function. As such, when an approximant that matches both phase and magnitude responses is required, our algorithm may be used.

The error criterion used is

$$\|E(\omega)\| = \max_{\omega \in (-\pi, \pi]} |G(e^{j\omega}) - \hat{G}(e^{j\omega})| \tag{1}$$

where G is the target or desired transfer function which we approximate by $\hat{G}$ in FIR(q).

Note that there are several "optimal" approaches to this problem, especially [1] and [5]. For a somewhat different approach involving an adaptive approach, see [6].

Manuscript received August 11, 1990; revised May 13, 1991.
The authors are with the Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, Canberra 2601, Australia.
IEEE Log Number 9201056.

2) Upper and lower bounds on the approximation error $\|E(\omega)\|$ are calculable *a priori*. Thus the filter length required to satisfy error constraints may be found before designing the filter. All of the approaches mentioned above [1], [5], [6] rely on *a posteriori* calculation of the approximation error; therefore the *a priori* calculation in our paper of both the global approximation error lower bound and the algorithm-dependent upper bound on the same quantity is highly significant.

3) The algorithm gives an exact solution and is guaranteed to terminate.

4) The algorithm is given in a state-space format and so is amenable to direct, numerically robust implementation.

5) The state-space algorithm description allows direct extension of the algorithm to the multi-input/multi-output (MIMO) case. Thus, problems such as model order reduction of quadrature mirror filters [7] may be approached using the algorithm presented here.

Due to the algorithm's extensive use of a concept called the Nehari extension (see Section II), we have named the algorithm the Nehari shuffle.

The remainder of the paper is organized as follows. Sections II and III contain definitions and some required background material. A statement of the FIR(q) approximation problem and a lower bound on the error associated with FIR(q) approximation is given in Section IV. Section V describes the Nehari shuffle. In Section VI we derive an upper bound on the approximation error of the algorithm and then, in Section VII, we give some examples using the algorithm which show that this bound is adhered to. The examples also show that, for larger filter lengths, the Nehari shuffle gives close to the minimum error. We also compare our algorithm with the Preuss algorithm [5]. Finally, Section VIII summarizes the results of the paper and points to possible areas of future research.

II. Nomenclature

In this paper, we use the following notation and definitions.

The real and complex numbers are denoted $\mathbb{R}$ and $\mathbb{C}$, respectively.

The transfer function $G(\sigma)$ with minimal state-space realization $G(\sigma) = D + C(\sigma I - A)^{-1}B$ will be written as $G(\sigma) = (A, B, C, D)$.

1053-587X/92$03.00 © 1992 IEEE

We denote by FIR(q) the set of all $G(z)$ which may be written

$$G(z) = \sum_{i=0}^{q-1} g_i z^{-i}$$

where $g_i \in \mathbb{C}$.

Given the stable discrete system $G(z) = (A, B, C, D)$ (where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{m \times n}$, and $D \in \mathbb{R}^{m \times p}$) then the controllability and observability gramians are given by

$$P = \sum_{k=0}^{\infty} A^k BB^* A^{*k} \tag{2}$$

$$Q = \sum_{k=0}^{\infty} A^{*k} C^* C A^k \tag{3}$$

where A^* denotes the Hermitian conjugate of A. A realization of $G(z)$ is called *balanced* [8] if $P = Q = \Sigma$ and Σ is diagonal. Standard software packages such as MATLAB and MATRIX_x provide function calls to balance a given nonbalanced realization.

The Hankel singular values of G (denoted $\sigma_i(G)$) are given by

$$\sigma_i(G) = \sqrt{\lambda_i(PQ)}, \qquad 1 \le i \le n$$

(where $\lambda_i(A)$ is the ith eigenvalue of A) and are also the singular values of the doubly infinite (but finite rank) Hankel matrix associated with G [9].

For simplicity, we shall assume that the $\sigma_i(G)$ are distinct (which is generally the case), and that

$$\bar{\sigma}(G) \triangleq \sigma_1(G) > \sigma_2(G) \cdots > \sigma_n(G) \triangleq \underline{\sigma}(G).$$

We define $\|G\|_H = \bar{\sigma}(G)$ to be the Hankel-norm of G. For an excellent treatise on the properties and theory of Hankel norms and Hankel singular values, see [9].

In Section V we shall describe the Nehari shuffle algorithm. To do so, we need to define some transfer function operations.

Let G be a stable transfer function (of McMillan degree N), which may be expressed as

$$G(z) = \sum_{i=0}^{q-1} g_i z^{-i} + z^{-(q-1)} \sum_{j=1}^{\infty} g_{j+q-1} z^{-j}$$
$$= G^h + z^{-(q-1)} G^t.$$

Then we use the following nomenclature:

$\mathcal{E}G$ Formation of the Nehari extension of G (see Section III).

$\mathcal{H}_q G$ Extraction of G^h from G.

$\mathfrak{I}_q G$ Extraction of G^t from G. If $G = (A, B, C, D)$ then

$$\mathfrak{I}_q G = (A, B, CA^{q-1}, 0). \tag{4}$$

Note that $\mathfrak{I}_q G$ is of McMillan degree N or less.

$\mathcal{R}G$ The reflection operator $\mathcal{R}$: $G(z) \mapsto G(z^{-1})$. If $G = (A, B, C, D)$ then

$$\mathcal{R}G = (A^{-1}, A^{-1}B, -CA^{-1}, D - CA^{-1}B). \tag{5}$$

$\mathcal{S}_q G$ The shift operation $z^{-(q-1)}G$. $\mathcal{S}_q^{-1}G$ denotes $z^{q-1}G$. Note that, on the unit circle (the frequency axis), $|\mathcal{S}_q G| = |\mathcal{S}_q^{-1} G| = |G|$; this identity is used frequently throughout the paper.

III. The Nehari Problem

The Nehari problem [10] may be stated as follows:

Given $G(z)$, a rational function analytic in $\{|z| > \rho, \rho < 1\}$ (i.e., possessing a power series in z^{-1} convergent on the unit circle), find $F(z)$, a rational function analytic in $\{|z| < r, r > 1\}$ (i.e., possessing a power series in z convergent on the unit circle) such that

$$\max_{\omega \in (-\pi, \pi]} |G(e^{j\omega}) - F(e^{j\omega})|$$

is minimized. The $F(z)$ so found is called the Nehari extension of $G(z)$.

The solution to this problem was provided by [11], and Glover[1] [9, theorem 6.3] gives an explicit state-space expression for the equivalent continuous time problem. We detail Glover's algorithm below.

The main points to note about $F(z)$ constructed below are:

1) $F(z)$ is the closest anticausal sequence to the causal $G(z)$.

2) If $G(z)$ is of McMillan degree N, then $F(z)$ is of degree $N - 1$.

3) For $G(z)$ of McMillan degree N,

$$\sigma_i(\mathcal{R}F) = \sigma_{i+1}(G), \qquad 1 \le i \le N - 1. \tag{6}$$

4) For optimal choice of F,

$$\min_F \max_\omega |G(e^{j\omega}) - F(e^{j\omega})| = \bar{\sigma}(G). \tag{7}$$

Glover's algorithm proceeds as follows:

1) Given $G(z)$ and its minimal state-space representation (A, B, C, D), we wish to obtain $F(z)$, the Nehari extension of G.

2) Balance (A, B, C, D) to obtain (A_b, B_b, C_b, D).

3) Form $\tilde{G}(s) = (\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$, the bilinear transformation of $G(z)$ using (see, e.g., [9]):

$$\tilde{A} = (I + A_b)^{-1}(A_b - I)$$
$$\tilde{B} = \sqrt{2}(I + A_b)^{-1}B_b$$
$$\tilde{C} = \sqrt{2}C_b(I + A_b)^{-1}$$
$$\tilde{D} = D - C_b(I + A_b)^{-1}B_b.$$

Since (A_b, B_b, C_b, D) was balanced, $\tilde{G}(s)$ has gramians

$$\tilde{P} = P = \text{diag}\,(\Sigma, \sigma_1(G))$$
$$\tilde{Q} = Q = \text{diag}\,(\Sigma, \sigma_1(G))$$

[1]Glover's paper deals with the more general problem of optimal Hankel-norm approximation where, given G, we wish to find that F_k with k stable poles such that $\max_{\omega \in (-\pi, \pi]} |G - F_k|$ is minimized. The Nehari problem solution assumes $k = 0$.

where $\Sigma = \text{diag}(\sigma_2(G), \sigma_3(G), \cdots, \sigma_n(G))$ and diag $(\cdot)$ indicates forming a diagonal matrix with the arguments on the diagonal.

4) Partition $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ conformably with $\tilde{P}$ and $\tilde{Q}$ as

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} \tag{8}$$

and

$$\tilde{C} = [\tilde{C}_1 \quad \tilde{C}_2]. \tag{9}$$

5) Form $\tilde{F}(s) = (\hat{\tilde{A}}, \hat{\tilde{B}}, \hat{\tilde{C}}, \hat{\tilde{D}})$:

$$\hat{\tilde{A}} = \Gamma^{-1}(\sigma_1 \tilde{A}_{11}^* + \Sigma \tilde{A}_{11} \Sigma - \sigma_1 \tilde{C}_1^* U \tilde{B}_1^*)$$

$$\hat{\tilde{B}} = \Gamma^{-1}(\Sigma \tilde{B}_1 + \sigma_1 \tilde{C}_1^* U)$$

$$\hat{\tilde{C}} = -\tilde{C}_1 \Sigma - \sigma_1 U \tilde{B}_1^*$$

$$\hat{\tilde{D}} = -\tilde{D} + \sigma_1 U$$

where U is a unitary matrix satisfying

$$\tilde{B}_2 = -\tilde{C}_2^* U$$

and

$$\Gamma \triangleq \Sigma^2 - \sigma_1^2 I.$$

6) Form $F(z) = (\hat{A}, \hat{B}, \hat{C}, \hat{D})$ the bilinear transformation of $\tilde{F}(s)$:

$$\hat{A} = (\hat{\tilde{A}} + I)(I - \hat{\tilde{A}})^{-1}$$

$$\hat{B} = \sqrt{2}(I - \hat{\tilde{A}})^{-1}\hat{\tilde{B}}$$

$$\hat{C} = \sqrt{2}\hat{\tilde{C}}(I - \hat{\tilde{A}})^{-1}$$

$$\hat{D} = \hat{\tilde{D}} + \hat{\tilde{C}}(I - \hat{\tilde{A}})^{-1}\hat{\tilde{B}}.$$

This $F(z)$ is then the Nehari extension of $G(z)$.

IV. The FIR(q) Approximation Problem

The FIR(q) approximation problem may be stated as follows:

Given $G_1(z)$, a rational transfer function analytic in $\{|z| > \rho, \rho < 1\}$ (i.e., possessing a power series in z^{-1} convergent on the unit circle), find $\hat{G}(z)$ such that $\hat{G} \in$ FIR(q) and

$$\max_{\omega \in (-\pi, \pi]} |G_1(e^{j\omega}) - \hat{G}(e^{j\omega})|$$

is minimized.

We present in the following lemma a lower bound on the accuracy with which a given transfer function may be approximated by an FIR(q) system.

Lemma 2: (Global Approximation Error Lower Bound): Given $G(z)$, a rational transfer function analytic in $\{|z| > \rho, \rho < 1\}$ expressed as

$$G(z) = \sum_{i=0}^{q-1} g_i z^{-i} + z^{-(q-1)} \sum_{j=1}^{\infty} g_{j+q-1} z^{-j}$$
$$= \mathcal{H}_q G(z) + \mathcal{S}_q \mathcal{T}_q G(z) \tag{10}$$

then

$$\min_{\hat{G} \in \text{FIR}(q)} \max_{\omega \in (-\pi, \pi]} |G(e^{j\omega}) - \hat{G}(e^{j\omega})| \geq \bar{\sigma}(\mathcal{T}_q G). \tag{11}$$

Proof: Write

$$|G(e^{j\omega}) - \hat{G}(e^{j\omega})| = |\mathcal{H}_q G(e^{j\omega}) + \mathcal{S}_q \mathcal{T}_q G(e^{j\omega}) - \hat{G}(e^{j\omega})|$$
$$= |\mathcal{S}_q \mathcal{T}_q G(e^{j\omega}) - \tilde{G}(e^{j\omega})|$$

where $\tilde{G} = \hat{G} - \mathcal{H}_q G$. Since $\mathcal{T}_q G$ is strictly causal and $\mathcal{S}_q^{-1}\tilde{G}$ is anticausal, we know (from the Nehari problem solution [9, theorem 6.1, 10]) that

$$\min_{\tilde{G}} \max_{\omega \in (-\pi, \pi]} |\mathcal{T}_q G(e^{j\omega}) - \mathcal{S}_q^{-1}\tilde{G}(e^{j\omega})| \geq \bar{\sigma}(\mathcal{T}_q G) \tag{12}$$

and the result follows directly from the Nehari theorem and that $|\mathcal{S}_q G| = |G|$ for $z = e^{j\omega}$. ■

Note that this is a global lower bound and is such regardless of the algorithm used to produce $\hat{G}$. Lemma 2 states nothing about whether or not the lower bound is achievable.

Note also that a consequence of this result is that if G is FIR(q), and a $q - 1$ coefficient approximant $\hat{G}$ is required, then the optimal choice of $\hat{G}$ is simply the $q - 1$ coefficient truncation of G; in this simple case the lower bound is certainly achievable.

We now present a result on the approximation error involved in approximation of G by $\mathcal{H}_q G$.

Lemma 3: (Truncation Approximation Error Upper Bound): Given $G(z)$, a rational transfer function (of McMillan degree N) analytic in $\{|z| > \rho, \rho < 1\}$ expressed as

$$G(z) = \sum_{i=0}^{q-1} g_i z^{-i} + z^{-(q-1)} \sum_{j=1}^{\infty} g_{j+q-1} z^{-j}$$
$$= \mathcal{H}_q (G(z) + \mathcal{S}_q \mathcal{T}_q G(z) \tag{13}$$

then

$$\max_{\omega \in (-\pi, \pi]} |G(e^{j\omega}) - \mathcal{H}_q G(e^{j\omega})| \leq 2 \sum_{i=1}^{N} \sigma_i(\mathcal{T}_q G). \tag{14}$$

Proof: We know (from the definition) that

$$\mathcal{S}_q \mathcal{T}_q G(z) = G(z) - \mathcal{H}_q G(z)$$

and also

$$|G(e^{j\omega})| = |z^n G(e^{j\omega})|$$

for all ω, $z = e^{j\omega}$ and all $n \in \mathbb{Z}$. So

$$|G(e^{j\omega}) - \mathcal{H}_q G(e^{j\omega})| = |\mathcal{S}_q \mathcal{T}_q G(z)| = |\mathcal{T}_q G(z)|.$$

From [12, theorem 2] we have that

$$\max_{\omega \in (-\pi, \pi]} |G(e^{j\omega})| \leq 2 \sum_{i=1}^{N} \sigma_i(G) \tag{15}$$

which, when applied to $\mathcal{T}_q G$, gives the required result. ■

It is computationally trivial to obtain $\mathcal{H}_q G$ from G and, in a sense, this result gives an upper bound on the FIR(q)

approximation error. This overbound is easily and exactly computable, as opposed to the effort required to calculate $\max_{\omega \in (-\pi, \pi]} |\mathfrak{T}_q G(e^{j\omega})|$ exactly. For any particular G, the overbound is not necessarily achieved.

V. Presenting the Nehari Shuffle

We now present our new algorithm for FIR(q) filter approximation.

To illustrate, consider Fig. 1. The impulse response of the IIR system, G_1, that we wish to approximate with an FIR(q) system $\hat{G}$, is shown as the solid line. The first approximant $\hat{G}_1$ is just the q-coefficient truncation of this.

To account for the tail of G_1 left out by $\hat{G}_1$, we Nehari extend $\mathfrak{T}_q G_1$. This extension is shown as the dotted line in Fig. 1. Since this extension "overshoots" the region in which we are interested ($n \in [0, q-1]$), we must extend its tail also. This second extension is indicated by the dashed line in the figure.

This procedure is repeated until the algorithm terminates. The "interleaving" of each successive Nehari extension leads to the algorithm being called the Nehari shuffle.

The algorithm proceeds as follows (using the definitions of Section II).

Given G, find $\hat{G} \in$ FIR(q).

1) Initialize $i = 1$, $\hat{G}_i = \mathcal{H}_q G$, and $G_i = G$.
2) Repeat

$$G_{i+1} = \mathcal{R}\mathcal{E}\mathfrak{T}_q G_i, \qquad \text{for all } i \tag{16}$$

$$\hat{G}_{i+1} = \hat{G}_i + \begin{cases} \mathcal{S}_q \mathcal{R}\mathcal{H}_q G_{i+1} & \text{for } i \text{ odd} \\ \mathcal{H}_q G_{i+1} & \text{for } i \text{ even} \end{cases}$$

$$i := i + 1. \tag{17}$$

Until $G_i = 0$.

So the final approximant is

$$\hat{G} = \hat{G}_{N+1} = \mathcal{H}_q G_1 + \mathcal{S}_q \mathcal{R}\mathcal{H}_q G_2 + \mathcal{H}_q G_3 + \cdots. \tag{18}$$

Note that the procedure above may be stopped after any number of steps rather than continued to completion. The penalty for doing this is an increased upper bound on the approximation error, as is shown in Section VI.

As an example of how the error

$$E_i(\omega) \triangleq |G(e^{j\omega}) - \hat{G}_i(e^{j\omega})|$$

changes as the algorithm proceeds (as i increments), consider Fig. 2. The three plots show $E_1(\omega)$ (solid line), $E_2(\omega)$ (dashed line), and $E_3(\omega)$ (dotted line). The transfer function being approximated is the low-pass filter of example 2 for $q = 8$.

To show that the termination condition is satisfied, we need the following theorem.

Theorem 4: (Termination of the Nehari Shuffle): Using the notation defined above,

$$G_{N+2} = (0, 0, 0, 0) = 0. \tag{19}$$

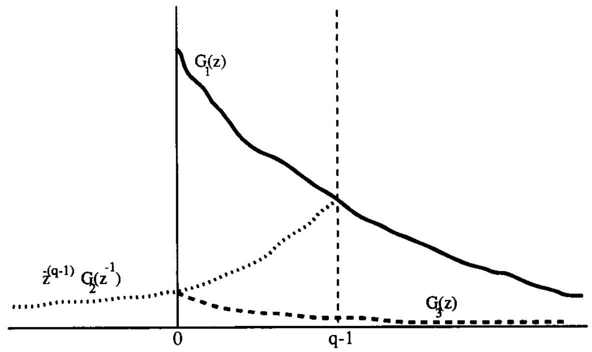


Fig. 1. Illustration of the Nehari shuffle algorithm.

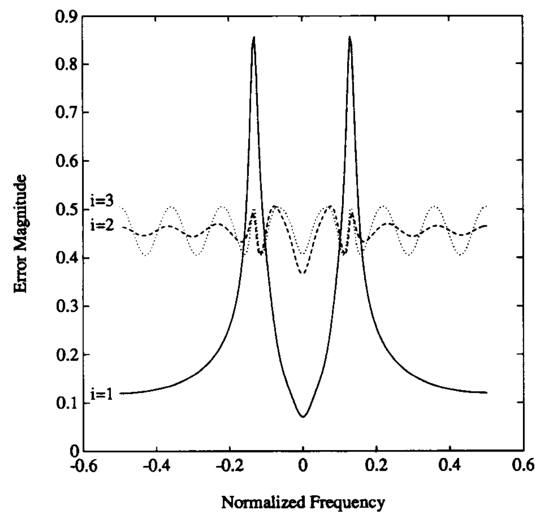


Fig. 2. $E_i(\omega)$ for $i = 1, 2,$ and 3.

Proof: We have that $G = G_1$ is of McMillan degree N.

Since formation of G_{i+1} in (16) involves the extension operator $\mathcal{E}$ (which reduces degree by 1) and the other operators are degree preserving, we have that, if G_i is of degree M, then G_{i+1} is of degree $M - 1$.

Hence, G_{N+1} is degree zero. This implies that $G_{N+1} = (0, 0, 0, D_{N+1})$ for which $G_{N+2} = (0, 0, 0, 0)$ and the algorithm terminates. ■

The Nehari shuffle algorithm thus terminates and the approximant arrived at is an exact answer, meaning that no iterative search/adaptation [1], [6] procedures are required.

VI. Error Bounds on the Nehari Shuffle

We give an upper bound on the approximation error in the following theorem.

Theorem 5: (Nehari Shuffle Approximation Error Upper Bound): Given $G(z)$, then the $\hat{G}(z) \in$ FIR(q) constructed as detailed in Section V has approximation error bounded *a priori* by

$$\max_{\omega\in(-\pi,\pi]} |G(e^{j\omega}) - \hat{G}(e^{j\omega})| \le \sum_{i=1}^{N} \sigma_i(\mathcal{J}_q G) \tag{20}$$

and *a posteriori* (after the filter has been designed) by

$$\max_{\omega\in(-\pi,\pi]} |G(e^{j\omega}) - \hat{G}(e^{j\omega})| \le \sum_{i=1}^{N} \bar{\sigma}(\mathcal{J}_q G_i). \tag{21}$$

In order to prove this theorem, we need the following result.

Lemma 6: (Hankel Singular Values of $\mathcal{J}_q G$*)*: Given a stable, causal transfer function of McMillan degree N, $G = \mathcal{H}_q G + \mathcal{S}_q \mathcal{J}_q G$ then $\sigma_i(\mathcal{J}_q G) \le \sigma_i(G)$ for $1 \le i \le N$.

Proof: From [9, theorem 7.2], we know that there is an optimal X_k, the extension of G with k stable modes $(0 \le k \le N-1)$, that satisfies

$$\min_{X_k} \max_{\omega\in(-\pi,\pi]} |G - X_k| = \sigma_{k+1}(G).$$

Thus we have

$$\begin{aligned}\min_{X_k} \max_{\omega\in(-\pi,\pi]} &|\mathcal{H}_q G + \mathcal{S}_q \mathcal{J}_q G - X_k|\\ &= \sigma_{k+1}(G)\\ &= \min_{X_k} \max_{\omega\in(-\pi,\pi]} |\mathcal{J}_q G + \mathcal{S}_q^{-1}\mathcal{H}_q G - \mathcal{S}_q^{-1} X_k|.\end{aligned}$$

Consider $\mathcal{S}_q^{-1}\mathcal{H}_q G - \mathcal{S}_q^{-1} X_k$. The first term is an FIR(q) system, shifted by $q-1$ samples to be made anticausal. The term $\mathcal{S}_q^{-1} X_k$ is a kth order system shifted so that the first $q-1$ terms of its impulse response are noncausal. Thus, the only causal, stable portion of $\mathcal{S}_q^{-1}\mathcal{H}_q G - \mathcal{S}_q^{-1} X_k$ is $\mathcal{J}_q X_k$ which is known to have less than or equal to k stable modes.

Now consider Y_k, the optimal extension of $\mathcal{J}_q G$ with k stable modes that satisfies

$$\min_{Y_k} \max_{\omega\in(-\pi,\pi]} |\mathcal{J}_q G - Y_k| = \sigma_{k+1}(\mathcal{J}_q G).$$

Because Y_k is optimal (for the class to which both Y_k and $\mathcal{S}_q^{-1}\mathcal{H}_q G - \mathcal{S}_q^{-1} X_k$ belong) we must have that

$$\begin{aligned}\min_{Y_k} \max_{\omega\in(-\pi,\pi]} &|\mathcal{J}_q G - Y_k|\\ &\le \min_{X_k} \max_{\omega\in(-\pi,\pi]} |\mathcal{J}_q G + \mathcal{S}_q^{-1}\mathcal{H}_q G - \mathcal{S}_q^{-1} X_k|\end{aligned}$$

which means

$$\sigma_{k+1}(\mathcal{J}_q G) \le \sigma_{k+1}(G)$$

for $0 \le k \le N-1$. ∎

We may now proceed to the proof of Theorem 5.

Proof: From (18)

$$\begin{aligned}\max_{\omega\in(-\pi,\pi]} &|G - \hat{G}|\\ &= \max_{\omega\in(-\pi,\pi]} |\overbrace{\mathcal{H}_q G_1 + \mathcal{S}_q \mathcal{J}_q G_1}^{G_1}\\ &\quad - \overbrace{(\mathcal{H}_q G_1 + \mathcal{S}_q \mathcal{R}\mathcal{H}_q G_2 + \mathcal{H}_q G_3 + \cdots)}^{\hat{G}}|\\ &= \max_{\omega\in(-\pi,\pi]} |\mathcal{S}_q \mathcal{J}_q G_1 - \overbrace{\mathcal{S}_q \mathcal{R}(G_2 - \mathcal{S}_q \mathcal{J}_q G_2)}^{\mathcal{S}_q \mathcal{R}\mathcal{H}_q G_2}\\ &\quad - G_3 + \mathcal{S}_q \mathcal{J}_q G_3 - \cdots|\\ &= \max_{\omega\in(-\pi,\pi]} |\mathcal{S}_q[\mathcal{J}_q G_1 - \mathcal{R} G_2] + [\mathcal{R}\mathcal{J}_q G_2 - G_3]\\ &\quad + \mathcal{S}_q[\mathcal{J}_q G_3 - \mathcal{R} G_4] + \cdots|\\ &= \max_{\omega\in(-\pi,\pi]} |\mathcal{S}_q[\mathcal{J}_q G_1 - \mathcal{E}\mathcal{J}_q G_1] + [\mathcal{R}\mathcal{J}_q G_2\\ &\quad - \mathcal{R}\mathcal{E}\mathcal{J}_q G_2] + \mathcal{S}_q[\mathcal{J}_q G_3 - \mathcal{E}\mathcal{J}_q G_3] + \cdots|\\ &\le \max_{\omega\in(-\pi,\pi]} \{|\mathcal{J}_q G_1 - \mathcal{E}\mathcal{J}_q G_1| + |\mathcal{R}\mathcal{J}_q G_2 - \mathcal{R}\mathcal{E}\mathcal{J}_q G_2|\\ &\quad + |\mathcal{J}_q G_3 - \mathcal{E}\mathcal{J}_q G_3| + \cdots\}\\ &= \sum_{i=1}^{N} \bar{\sigma}(\mathcal{J}_q G_i)\end{aligned} \tag{22}$$

which is result (21).

From Lemma 6 and (6)

$$\bar{\sigma}(\mathcal{J}_q G_2) \le \bar{\sigma}(G_2) = \sigma_2(\mathcal{J}_q G_1)$$

follows directly. Similarly,

$$\bar{\sigma}(\mathcal{J}_q G_3) \le \bar{\sigma}(G_3) = \sigma_2(\mathcal{J}_q G_2) \le \sigma_2(G_2) = \sigma_3(\mathcal{J}_q G_1).$$

Successive application of the lemma yields $(i > 1)$

$$\begin{aligned}\bar{\sigma}(\mathcal{J}_q G_i) &\le \bar{\sigma}(G_i)\\ &= \sigma_2(\mathcal{J}_q G_{i-1}) \le \cdots \le \sigma_{i-1}(F_1) = \sigma_i(\mathcal{J}_q G_1).\end{aligned}$$

The upper bound on the error (22) of the Nehari shuffle approximant then becomes

$$\max_{\omega\in(-\pi,\pi]} |G_1 - \hat{G}| \le \sum_{i=1}^{N} \sigma_i(\mathcal{J}_q G_1). \tag{23}$$

∎

VII. Examples

We consider two examples.

In the first example, we consider approximating a system with linear phase. The second example considers the case where the system we wish to approximate has highly nonlinear phase. In this second example, we compare the performance of our algorithm with that of Preuss [5].

In Figs. 3 and 4: 1) the solid line is the upper bound of (20); 2) the dashed line is the lower bound provided by (11); and 3) × points are the Nehari shuffle approximation errors.

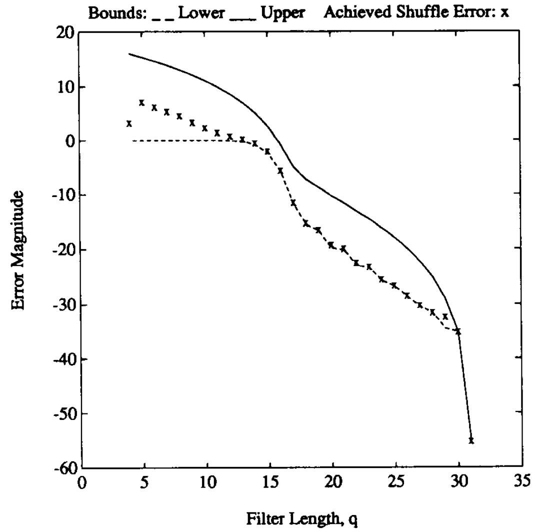


Fig. 3. Example 1—Chebyshev approximation errors (in decibels) versus length of approximant.

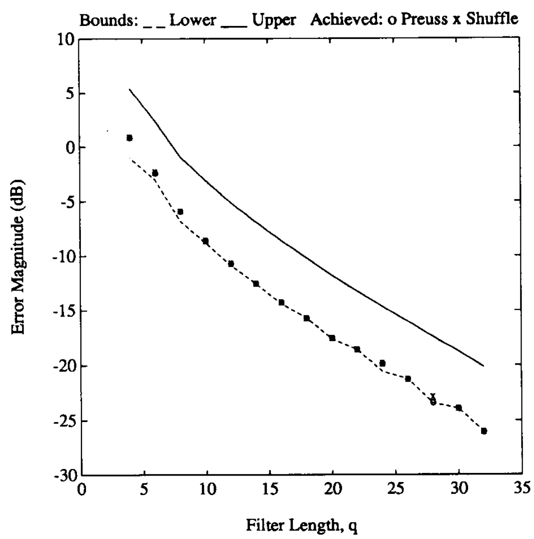


Fig. 4. Example 2—Chebyshev approximation errors (in decibels) versus length of approximant.

In Fig. 4 the ○ points show the approximation error achieved by the Preuss [5] algorithm.

A. Example 1: The Linear Phase Case

Take G_1 to be phase FIR(32) low-pass filter with frequency responses shown in Fig. 5.

If we select $q = 22$, then the resulting magnitude and phase responses of the approximant produced by the Nehari shuffle algorithm are depicted in Fig. 6. Note that, over the passband our approximant has approximately linear phase with the same phase rolloff rate as the original FIR(32) system.

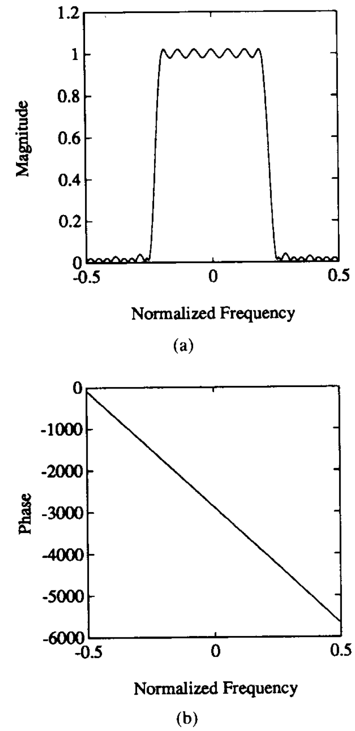


Fig. 5. Example 1—Original FIR(32) (a) magnitude and (b) phase responses.

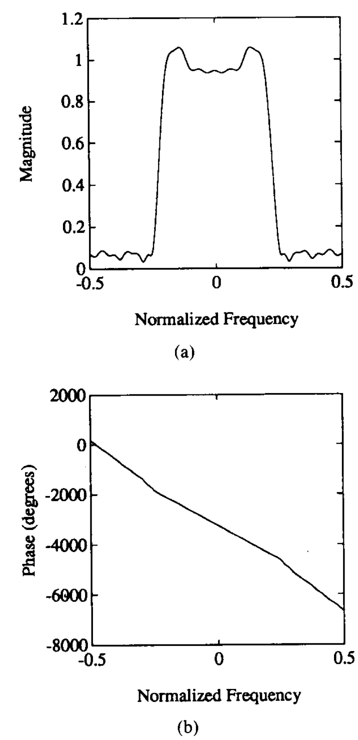


Fig. 6. Example 1—(a) Magnitude and (b) phase responses of Nehari shuffle FIR(22) approximant.

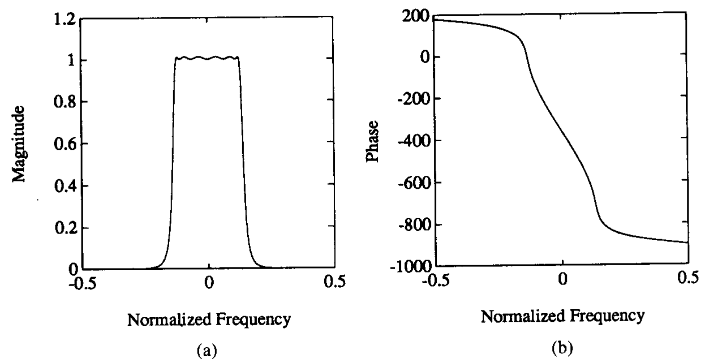


Fig. 7. Example 2—Original Chebyshev (a) magnitude and (b) phase responses.

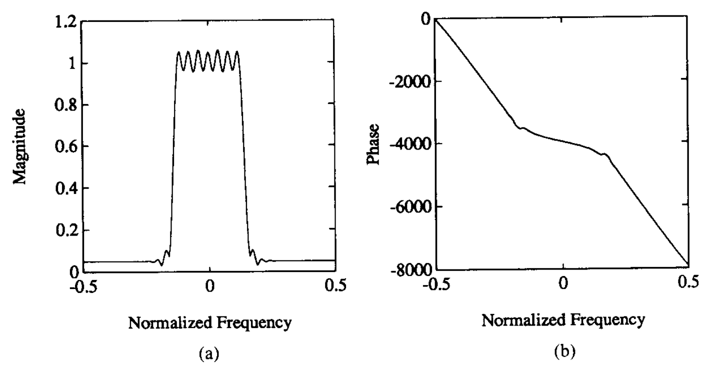


Fig. 8. Example 2—(a) Magnitude and (b) phase responses of Nehari shuffle FIR(32) approximant.

If the algorithm is then used to design filters of length 3 to 31, the resulting Chebyshev approximation errors

$$\|E(\omega)\| = \max_{\omega \in (-\pi, \pi]} |G(e^{j\omega}) - \hat{G}(e^{j\omega})|$$

are displayed in Fig. 3. Note that, for filters of any order, the lower and upper bounds are not exceeded and for filters with $q \geq 13$, the Nehari shuffle errors are very close to the lower bound.

B. Example 2: The Nonlinear Phase Case

Now take G_1 to be a sixth-order, 0.1-dB ripple low-pass Chebyshev filter with frequency responses shown in Fig. 7. If the Nehari shuffle algorithm is used to design an FIR(32) filter, then the resulting magnitude and phase responses are shown in Fig. 8.

If both the Nehari shuffle and Preuss (complex Remez [5]) algorithms were used to design filters of even length (from 4 to 32), the resulting Chebyshev approximation errors are displayed in Fig. 4.

The results show that the Nehari shuffle approximation errors, for all tested filter lengths, again lie within the derived bounds. Note also that the Preuss algorithm, while giving minor improvement over the Nehari shuffle approximant, does not improve the lower bound.

The number of floating-point operations actually required by each algorithm to perform the approximations is plotted in Fig. 9. Note that each algorithm takes a similar number of operations for the tested filter lengths. Both algorithms may be able to be implemented in a more ef-

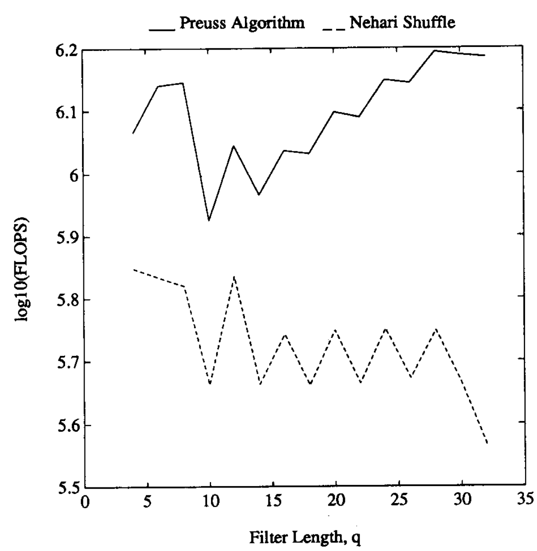


Fig. 9. Example 2—Floating-point operations required for each algorithm versus filter length q.

ficient manner than the state variable style chosen; since both were implemented using state variable descriptions, the comparison is still valid.

VIII. Conclusions

We have presented the Nehari shuffle, a novel approach to the design of FIR(q) filters. The method's main points are listed below.

1) It is an approach to the unconstrained FIR(q) approximation problem with an ∞-norm error criterion. The resulting FIR(q) filters are not constrained to have linear phase as in the Parks–McClellan algorithm. This allows FIR implementation of IIR transfer functions to within known (or calculable) error bounds.

2) Upper and lower bounds on the approximation error are calculated *a priori*, so the closeness of the designed FIR(q) filter to the given desired system may be selected.

3) The implementation method used gives an exact solution and is guaranteed to terminate. The Nehari extension solution method provided by Glover [9] involves only matrix manipulations.

4) Extension of the algorithm to the multi-input/multi-output case is direct. This may have particular application to quadrature mirror filter bank design, where current methods involve separate design of the individual filters [7].

Future Work: The present paper allows us to point to future research directions on the FIR(q) approximation problem:

1) The algorithm can be reworked to use

$$E_{\text{rel}}(\omega) = \max_{\omega \in (-\pi, \pi]} \left| \frac{G(e^{j\omega}) - \hat{G}(e^{j\omega})}{G(e^{j\omega})} \right|$$

as the error criterion [13]. Thus the relative error between G and $\hat{G}$ is the criterion, rather than the absolute error.

2) The use of general frequency weightings is a difficult problem and is yet to be resolved.

References

[1] X. Chen and T. W. Parks, "Design of FIR filters in the complex domain," *IEEE Trans. Acoust.*, Speech, Signal Processing, vol. ASSP-35, pp. 144–153, Feb. 1987.

[2] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[3] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (International Edition). Englewood Cliffs, NJ: Prentice-Hall, 1989.

[4] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. CT-19, no. 2, pp. 189–194, Mar. 1972.

[5] K. Preuss, "On the design of FIR filters by complex Chebychev approximation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 702–712, May 1989.

[6] N. N. Chit and J. S. Mason, "Complex Chebyshev approximation for FIR digital filters," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 49–54, Jan. 1991.

[7] P. P. Vaidyanathan, "Quadrature mirror filter banks, *M*-band extension and perfect-reconstruction techniques," *IEEE ASSP Mag.*, vol. 4, no. 3, pp. 4–20, July 1987.

[8] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.

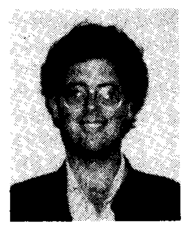
[9] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Contr.*, vol. 39, no. 6, pp. 1115–1193, 1984.

[10] Z. Nehari, "On bounded bilinear forms," *Ann. Math.*, vol. 65, no. 1, pp. 153–162, Jan. 1957.

[11] V. M. Adamjan, 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, no. 1, pp. 31–73, 1971.

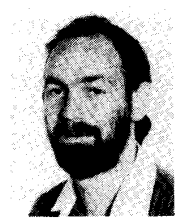
[12] U. M. Al-Saggaf and G. F. Franklin, "On model reduction," in *Proc. IEEE Conf. Decision Contr.*, Dec. 1986, pp. 1064–1069.

[13] K. Glover, "Multiplicative approximation of linear multivariable systems with L^∞ error bounds," in *Proc. ACC*, 1986.



Peter J. Kootsookos was born in Townsville, Australia, in 1965. He received the B.E. degree in 1986 and the M.Eng.Sc. degree in 1989, both in electrical engineering, from the University of Queensland, Australia. He is currently pursuing the Ph.D. degree in systems engineering at the Australian National University, Canberra, Australia.

During 1988 he worked at Mitec Ltd., a microwave engineering company. His research interests include FIR filter design, model reduction, spectrum estimation, and frequency estimation.



Robert R. Bitmead (S'76–M'79–M'86–F'91) was born in Sydney, Australia, in 1954. He received the B.Sc. degree in applied mathematics from the University of Sydney and the M.E. and Ph.D. degrees from the University of Newcastle, Australia, in 1976, 1977, and 1979, respectively.

He has held faculty positions at the University of Newcastle and James Cook University of North Queensland, and has held visiting positions at Cornell University, Ithaca, NY, and the University of Louvain, Belgium. He has been on the faculty of the Australian National University in Canberra since 1983, where he currently holds the position of Senior Fellow of Systems Engineering and is the Director of the Centre for Robust and Adaptive Systems. His research interests encompass adaptive systems in control, signal processing, and telecommunications; stochastic processes; and multivariable control.

Dr. Bitmead is currently a Subject Editor for the *International Journal of Adaptive Control and Signal Processing*, is an Associate Editor of *Automatica*, and has also served as Associate Editor of the IEEE Transactions on Automatic Control. He is currently the Chairman of the IFAC Technical Committee on Mathematics of Control.

Michael Green was born in Sydney, Australia, in 1961. He received the B.Sc. degree in applied mathematics from the University of New South Wales, Sydney, Australia, in 1984, and the Ph.D. degree in systems engineering from the Australian National University, Canberra, Australia, in 1987.

From 1987 to 1991 he was a postdoctoral Research Assistant and, subsequently, a Lecturer in the Department of Electrical Engineering, Imperial College, London, where he was also a member of the Interdisciplinary Research Centre in Process Simulation and Control. In 1991 he was appointed Senior Lecturer with the Interdisciplinary Engineering Program and the Department of Systems Engineering at the Australian National University, Canberra. His research interests include multivariable control, $\mathcal{H}_\infty$ control theory, and model reduction.