Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

From IIR to FIR Digital MIMO Models: A Constructive Hankel Norm Approximation Method

Li Chai†§, Jingxin Zhang§, Cisheng Zhang‡ and Edoardo Mosca*

§ Department of Electrical and Computer Systems Engineering
Monash University, Clayton, VIC3800, Australia

†Department of Automation, Hangzhou Dianzi University
Hang Zhou, 310018, P.R. China

‡School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

*Dipartimento di Sistemi e Informatica, Universityá di Firenze, Italy

Abstract— **This paper presents a constructive method to (sub)optimal finite impulse response (FIR) approximation of a given infinite impulse response (IIR) MIMO model. The method minimizes the Hankel norm of approximation error by using the explicit solution of norm-preserve dilation problem. It has the advantage over the existing methods that it provides an explicitly constructive solution and allows the trade-off between the Chebyshev and least square criteria. The lower and upper bounds on the H_2 and H_∞ norms of approximation error are given. The algorithm for approximating non-causal IIR filters by causal FIR filters is also formulated and solved. The effectiveness and properties of the proposed algorithm are demonstrated through two examples.**

Keywords: Hankel-norm; FIR approximation; norm-preserve dilation; mixed norm design; IIR filters

I. INTRODUCTION

Finite impulse response (FIR) models have the advantage of intrinsically stable properties and easy implementation, thus are more preferred to infinite impulse response (IIR) models [1], [2]. However, we often end up with IIR models in system and signal modelling, filter and controller design, etc [1], [2], [3]. Therefore, effective methods are required to approximate an IIR model by FIR model. Generally, the approximation problem can be stated as follows:

Given $G(z)$, a stable rational transfer matrix, find

$$F(z) = f_0 + f_1 z^{-1} + \cdots + f_{m-1} z^{-m+1}$$

such that the norm of the error $\|G(z) - F(z)\|$ is minimized, where $||\cdot||$ could be different norms corresponding to different design criteria.

The early methods to the approximation use direct truncation of impulse response that minimizes the least-square error criterion, or equivalently the H_2 error norm $\|G(z) - F(z)\|_2$ [2]. In [4], [5], [6], the minimum Chebyshev error criterion, or equivalently, the H_∞ error norm $\|G(z) - F(z)\|_\infty$ is used. In [5], [6], a method called Nehari Shuffle is proposed and upper and lower bound on the approximation error are derived. However, the Nehari Shuffle doesn't provide the optimal solutions with respect to H_∞ norm. A direct H_∞ norm optimization approach is given by the powerful tool of linear matrix inequalities (LMIs) [4].

As pointed out in [7], the least square criterion is appropriate if the input signal is narrow-band, and Chebyshev criterion is appropriate if the input signal is wide-band and distributed approximately uniformly in the frequency. Thus, there are situations where neither the Chebyshev criterion nor the least square criterion is appropriate, and where we call for alternative design methods with trade-off between least square and Chebyshev criteria [7], [**?**].

In this paper, the Hankel norm of the error is chosen to be minimized. Hankel-norm approximation is extensively used in model reduction after the remarkable work of Glover [8], [9]. However, the problem here is different from that of [8], which is to find a lower order IIR model for a given high order IIR model. The resulting method of this paper has the following advantages.

- It is developed for multi-input multi-output (MIMO) models directly, and allows the tradeoff between H_2 error norm and H_∞ error norm.
- The design algorithm is constructive, and only involves algebraic manipulations, therefore no iteration and convex optimization program (as LMIs) are needed.
- No need to carry out balanced realization and truncation as [6].
- Lower and Upper bounds on H^2 norm and H_∞ norm of the error system are provided.

II. PRELIMINARY

This section introduces the notations and some preliminary results used in the sequel. Let $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex numbers respectively. For a matrix X, let X^* denote its complex conjugate transpose, $\lambda(X)$ its eigenvalue, and $\sigma(X)$ its singular value. Denote the spectrum norm of X as $||X|| = (\bar{\lambda}(X^*X))^{\frac{1}{2}}$, where $\bar{\lambda}$ denotes the largest eigenvalue of X. For a positive definite matrix X, we use

$X^{\frac{1}{2}}$ to denote its Hermitian square root, that is, $X^{\frac{1}{2}}X^{\frac{1}{2}} = X$ and $(X^{\frac{1}{2}})^* = X^{\frac{1}{2}}$.

A. Spaces, norms and Hankel Operators

Definition 1: Given a causal transfer matrix $G(z) \in \mathbb{C}^{q\times p}$, (A, B, C, D) is called a state space realization if $G(z) = D + C(zI - A)^{-1}B$, where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{q\times n}$ and $D \in \mathbb{R}^{q\times p}$.

Definition 2: For a stable transfer matrix with state space realization $G(z) = D + C(zI - A)^{-1}B$, the controllability and observability Gramian, denoted by P and Q, is defined by $P = \sum_{k=0}^{\infty} A^k BB^* A^{*k}$ and $Q = \sum_{k=0}^{\infty} A^{*k} C^* CA^k$.

It is well known that P and Q can be computed from the following Lyapunov equations respectively

$$APA^* - P + B^*B = 0 \quad (1)$$
$$A^*QA - Q + C^*C = 0. \quad (2)$$

The realization is minimal if P and Q are nonsingular.

For causal and stable $G(z) = \sum_{k=0}^{\infty} g_k z^{-k}$ with minimal state space realization $D+C(zI-A)^{-1}B$, where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{q\times n}$ and $D \in \mathbb{R}^{q\times p}$, the Hankel operator of G, denoted by Γ_G, is defined as

$$\Gamma_G = \begin{bmatrix} g_1 & g_2 & g_3 & \cdots \\ g_2 & g_3 & g_4 & \cdots \\ g_3 & g_4 & g_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The Hankel singular values of $G(z)$, denoted by $\sigma_i(\Gamma_G)$, $i = 1, \ldots, n$ are the ith singular values of Γ_G. The Hankel norm of G, denoted by $||\Gamma_G||$ is defined to be the largest singular value of Γ_G, i.e. $||\Gamma_G|| = \sigma_1(\Gamma_G)$. The following can be used to compute the Hankel-norm of a transfer matrix, see [11], [8], [10] for detail.

Lemma 1: For the above $G(z)$, we have

$$\sigma_i^2(\Gamma_G) = \lambda_i(QP) = \lambda_i(Q^{\frac{1}{2}}PQ^{\frac{1}{2}}) = \lambda_i(P^{\frac{1}{2}}QP^{\frac{1}{2}})$$
$$||\Gamma_G||^2 = \bar{\lambda}(QP) = \bar{\lambda}(Q^{\frac{1}{2}}PQ^{\frac{1}{2}}) = \bar{\lambda}(P^{\frac{1}{2}}QP^{\frac{1}{2}})$$

where P and Q are controllability and observerbility granmians respectively.

B. Norm-preserve Dilations

Consider the block matrix $\begin{bmatrix} X & R \\ S & T \end{bmatrix}$, where X, R, S and T are matrices of compatible dimensions, and denote

$$\alpha(X) = \left\| \begin{bmatrix} X & R \\ S & T \end{bmatrix} \right\|.$$

The norm-preserve dilation problem is to find X such that $\alpha(X)$ is minimized for given matrices R, S, and T. Denote

$$\gamma_0 = \min_X \alpha(X). \quad (3)$$

The following results play a very important role in our development [12].

Lemma 2: The minimum γ_0 in (3) is given by

$$\gamma_0 = \max\left\{ \left\| \begin{bmatrix} S & T \end{bmatrix} \right\|, \left\| \begin{bmatrix} R \\ T \end{bmatrix} \right\| \right\}.$$

Moreover, assume $\gamma \geq \gamma_0$, then the solution set X such that $\alpha(X) \leq \gamma$ can be characterized by

$$X = -YT^*Z + \gamma(I - YY^*)^{1/2}W(I - Z^*Z)^{1/2} \quad (4)$$

where W is an arbitrary contraction ($||W|| \leq 1$) and Y and Z are contractions satisfying

$$R = Y(\gamma^2 I - T^*T)^{1/2} \quad (5)$$
$$S = (\gamma^2 I - TT^*)^{1/2}Z. \quad (6)$$

The following lemma gives a more explicit formula when $||T|| < \gamma$.

Lemma 3: Assume that $\gamma \geq \gamma_0$ and $||T|| < \gamma$. Then the solution set X such that $\alpha(X) \leq \gamma$ can be characterized by

$$X = -R(\gamma^2 I - T^*T)^{-1}T^*S + \gamma[I - R(\gamma^2 I - T^*T)^{-1}R^*]^{\frac{1}{2}} W[I - S^*(\gamma^2 I - TT^*)^{-1}S]^{\frac{1}{2}}. \quad (7)$$

The norm-preserve dilation problem is solved independently by Parrot and Davis et. al. For more detail, please refer to [12].

III. HANKEL-NORM FIR APPROXIMATION

In this section, an algorithm is developed to solve the (sub)optimal Hankel-norm FIR approximation of a given IIR model. First, we present a basic theorem from which the approximation can be converted to a matrix norm-preserving dilation problem. Then a constructive algorithm is developed step by step. Finally, some properties of the resulting FIR approximation are discussed and the bounds on error norms are given.

The problem to be considered in this section is as follows. Given an IIR model $G(z) = D + C(zI - A)^{-1}B \in \mathbb{C}^{q\times p}$, find an $(m-1)$th order FIR model

$$F(z) = f_0 + f_1 z^{-1} + \cdots + f_{m-1}z^{-m+1}$$

that minimizes $||\Gamma_E||$, the Hankel norm of the approximation error $E(z) = z^{-1}(G(z) - F(z))$. The reason we put a delay term z^{-1} in $E(z)$ is due to the fact that the Hankel norm of a system is unrelated to the feed-through term. The relation of Hankel norm, H_2 norm and H_∞ norm are given in the following lemma

Lemma 4: Let $E(z) = \sum_{i=1}^{\infty} e_i z^{-i}$ satisfy $||\Gamma_E|| < \infty$. Then we have

$$\|E(z)\|_2 \leq ||\Gamma_E|| \leq \|E(z)\|_\infty \leq 2\sum_{i=1}^{N} \sigma_i(\Gamma_E)$$

where $\sigma_i(\Gamma_E)$ is the ith singular value of $E(z)$ and N=rank(Γ_E) = McMillan degree of $E(z)$.

The first two inequalities are shown by Theorem 4.2 of [13] and the last inequality is shown in [10], [6].

Lemma 4 tells us that the Hankel-norm can be seen as the tradeoff of H_2 norm and H_∞ norm. The following theorem is important to develop our algorithm. A similar result is given in [14]. Here our proof is more direct.

Theorem 1: For $G(z) = D + C(zI - A)^{-1}B$, define $H(z) = z^{-1}G(z)$. Then we have

$$||\Gamma_H|| = \left\| \begin{array}{cc} D & CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right\|$$

where P and Q are solutions of Lyapunov equations (1) and (2) respectively.

Proof: Note that $H(z)$ can be written in the state space form as $H(z) = \bar{C}(zI - \bar{A})^{-1}\bar{B}$, where

$$\bar{A} = \left[\begin{array}{cc} A & 0 \\ C & 0 \end{array} \right], \ \bar{B} = \left[\begin{array}{c} B \\ D \end{array} \right], \ \bar{C} = \left[\begin{array}{cc} 0 & I \end{array} \right].$$

Then we get the Lyapunov equations for $H(z)$ as follows

$$\begin{aligned} \bar{A}\bar{P}\bar{A}^* - \bar{P} + \bar{B}^*\bar{B} &= 0 \qquad (8) \\ \bar{A}^*\bar{Q}\bar{A} - \bar{Q} + \bar{C}\bar{C}^* &= 0. \qquad (9) \end{aligned}$$

Then we have

$$\bar{P} = \left[\begin{array}{cc} P & APC^* + BD^* \\ CPA^* + DB^* & CPC^* + DD^* \end{array} \right].$$

and

$$\bar{Q} = \left[\begin{array}{cc} Q & 0 \\ 0 & I \end{array} \right].$$

From direct matrix manipulation, we have

$$\bar{Q}^{\frac{1}{2}}\bar{P}\bar{Q}^{\frac{1}{2}} = \left[\begin{array}{cc} Q^{\frac{1}{2}}AP^{\frac{1}{2}} & Q^{\frac{1}{2}}B \\ CP^{\frac{1}{2}} & D \end{array} \right] \left[\begin{array}{cc} Q^{\frac{1}{2}}AP^{\frac{1}{2}} & Q^{\frac{1}{2}}B \\ CP^{\frac{1}{2}} & D \end{array} \right]^*.$$

Then it follows from Lemma 1 that

$$||\Gamma_H|| = \sqrt{\bar{\lambda}\left(\bar{Q}^{\frac{1}{2}}\bar{P}\bar{Q}^{\frac{1}{2}}\right)} = \left\| \begin{array}{cc} D & CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right\|.$$

■

Theorem 2: Given a transfer matrix $G(z) = C(zI - A)^{-1}B \in \mathbb{C}^{q\times p}$ with $||\Gamma_G|| = \gamma_0$, define $H(z) = z^{-1}(D + G(z))$ for a matrix D $in\mathbb{R}^{q\times p}$. Then we have
(i) $||\Gamma_H|| \geq \gamma_0$ for any D.
(ii) There exist D's such that $||\Gamma_H|| = \gamma_0$, and all such D's can be characterized by

$$D = -YP^{\frac{1}{2}}A^*Q^{\frac{1}{2}}Z + \gamma_0(I - YY^*)^{\frac{1}{2}}W(I - Z^*Z)^{\frac{1}{2}} \quad (10)$$

where $||W|| \leq 1$ and Y and Z are contractions satisfying

$$\begin{aligned} CP^{\frac{1}{2}} &= Y\left(\gamma_0^2 I - P^{\frac{1}{2}}A^*QAP^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ Q^{\frac{1}{2}}B &= \left(\gamma_0^2 I - P^{\frac{1}{2}}A^*QAP^{\frac{1}{2}}\right)^{\frac{1}{2}} Z. \end{aligned}$$

(iii) For any $\gamma > \gamma_0$, all D's such that $||\Gamma_H|| \leq \gamma$ are given by

$$D = \alpha + \beta \qquad (11)$$

where $||W|| \leq 1$, and

$$\begin{aligned} \alpha &= -C(\gamma^2P^{-1} - A^*QA)^{-1}A^*QB \qquad (12) \\ \beta &= \gamma\left[I - C(\gamma^2P^{-1} - A^*QA)^{-1}C^*\right]^{1/2} \\ &\quad W\left[I - B^*(\gamma^2Q^{-1} - APA^*)^{-1}B\right]^{1/2}. \qquad (13) \end{aligned}$$

Proof: We know from Theorem 1 that

$$||\Gamma_H|| = \left\| \begin{array}{cc} D & CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right\|.$$

Note that

$$\begin{aligned} &\left[\begin{array}{cc} Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right] \left[\begin{array}{cc} Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right]^* \\ &\quad = Q^{\frac{1}{2}}(BB^* + APA^*)Q^{\frac{1}{2}} = Q^{\frac{1}{2}}PQ^{\frac{1}{2}} \qquad (14) \\ &\left[\begin{array}{cc} P^{\frac{1}{2}}C^* & P^{\frac{1}{2}}A^*Q^{\frac{1}{2}} \end{array} \right] \left[\begin{array}{c} CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right] \\ &\quad = P^{\frac{1}{2}}(C^*C + A^*QA)P^{\frac{1}{2}} = P^{\frac{1}{2}}QP^{\frac{1}{2}}. \qquad (15) \end{aligned}$$

By Lemma 1 and (14-15), we have

$$\begin{aligned} &\left\|\left[\begin{array}{cc} Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right]\right\|^2 = \bar{\lambda}(Q^{\frac{1}{2}}PQ^{\frac{1}{2}}) = \gamma_0^2 \\ &\quad = \bar{\lambda}(P^{\frac{1}{2}}QP^{\frac{1}{2}}) = \left\|\left[\begin{array}{c} CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right]\right\|^2. \qquad (16) \end{aligned}$$

Then it follows from Lemma 2 that

$$\min_D ||\Gamma_H|| = \max\left\{ \left\|\left[\begin{array}{cc} Q^{\frac{1}{2}}B & Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right]\right\|, \left\|\left[\begin{array}{c} CP^{\frac{1}{2}} \\ Q^{\frac{1}{2}}AP^{\frac{1}{2}} \end{array} \right]\right\| \right\} = \gamma_0.$$

Moreover, there exists D such that $||\Gamma_H|| = ||\Gamma_G||$. Substituting $R = CP^{\frac{1}{2}}$, $S = Q^{\frac{1}{2}}B$ and $T = Q^{\frac{1}{2}}AP^{\frac{1}{2}}$ into the formula (4), we can get (10) after some direct algebraic manipulations. This completes the proof of (i) and (ii).
Obviously we have $\left\|Q^{\frac{1}{2}}AP^{\frac{1}{2}}\right\| \leq \gamma_0 < \gamma$. Substituting $R = CP^{\frac{1}{2}}$, $S = Q^{\frac{1}{2}}B$ and $T = Q^{\frac{1}{2}}AP^{\frac{1}{2}}$ into equation (7), we have $D = \alpha + \beta$. ■

We are now ready to present the main result of this section. Before presentation, we recall the following well known fact [6]: a causal transfer matrix $G(z) = D + C(zI - A)^{-1}B$ can be written in the form $G(z) = G_1(z) + z^{-m+1}G_m(z)$, where

$$G_1(z) = \sum_{i=0}^{m-1} g_i z^{-i} \qquad (17)$$

with g_i being the first m impulse responses of $G(z)$, and $G_m(z)$ is a strictly proper (rational function of z) transfer matrix.

Theorem 3: For a stable and causal transfer matrix $G(z) = G_1(z) + z^{-m+1}G_m(z)$ with $||\Gamma_{G_m}|| = \gamma_0$, and $\mathbf{e} = [e_0 \quad e_1 \quad \cdots \quad e_{m-1}]$, define

$$E(z) = z^{-1}\left(\sum_{i=0}^{m-1} e_i z^{-i} + z^{-m+1}G_m(z)\right).$$

Here $e_i \in \mathbb{R}^{q\times p}$. Then we have
(i) $||\Gamma_E|| \geq \gamma_0$ for any $\mathbf{e}$.
(ii) For any $\gamma \geq \gamma_0$, a particular $\mathbf{e}$ can be constructed explicitly such that $||\Gamma_E|| \leq \gamma$.

Proof: Denote $E_m(z) = G_m(z)$ and

$$E_{m-i-1}(z) = z^{-1}(e_{m-i-1} + E_{m-i}(z)) \tag{18}$$

for $i = 0, \ldots, m-1$. Using part (i) of Theorem 2 recursively, we have

$$||\Gamma_{E_0}|| \geq ||\Gamma_{E_1}|| \geq \cdots \geq ||\Gamma_{E_m}||$$

for any $\mathbf{e}$. Obviously, $E_0(z) = E(z)$. Therefore

$$||\Gamma_E|| \geq ||\Gamma_{E_m}|| = ||\Gamma_{G_m}|| = \gamma_0 \tag{19}$$

for any $\mathbf{e}$. We will prove part (ii) in a constructive manner by showing that e_{m-i-1} can be computed if e_{m-i} is obtained. Now assume that a state-space realization for $E_{m-i}(z)$ is given by

$$E_{m-i}(z) = C_{m-i}(zI - A_{m-i})^{-1}B_{m-i}. \tag{20}$$

Then the controllability and observability Gramians P_{m-i} and Q_{m-i} can be computed from equations (1) and (2) respectively. Since $\gamma \geq \gamma_0$, it then follows from Theorem 2 that there exists e_{m-i-1} such that

$$||\Gamma_{E_{m-i-1}}|| \leq \gamma$$

where $E_{m-i-1}(z)$ is defined by (18). Moreover, if $\gamma > \gamma_0$, then those e_{m-i-1} are given by

$$e_{m-i-1} = \alpha_{m-i} + \beta_{m-i} \tag{21}$$

where

$$\begin{aligned}\alpha_{m-i} &= -C_{m-i}P_{m-i}\\ &\cdot(\gamma^2 I - A^*_{m-i}Q_{m-i}A_{m-i}P_{m-i})^{-1}A^*_{m-i}Q_{m-i}B_{m-i}\end{aligned} \tag{22}$$

$$\begin{aligned}&\beta_{m-i} =\\ &\gamma[I - C_{m-i}P_{m-i}(\gamma^2 I - A^*_{m-i}Q_{m-i}A_{m-i}P_{m-i})^{-1}C^*_{m-i}]^{\frac{1}{2}}\\ &W_{m-i}[I - B^*_{m-i}Q_{m-i}(\gamma^2 I-\\ &A_{m-i}P_{m-i}A^*_{m-i}Q_{m-i})^{-1/2}B_{m-i}]^{\frac{1}{2}}\end{aligned} \tag{23}$$

and $|W_{m-i}| \leq 1$. Following the same line as the proof of Theorem 1, it is easy to check that a state space realization for $E_{m-i-1}(z)$ is given by

$$E_{m-i-1}(z) = C_{m-i-1}(zI - A_{m-i-1})^{-1}B_{m-i-1} \tag{24}$$

where $A_{m-i-1} = \begin{bmatrix} A_{m-i} & 0 \\ C_{m-i} & 0 \end{bmatrix}$, $B_{m-i-1} = \begin{bmatrix} B_{m-i} \\ e_{m-i} \end{bmatrix}$ and $C_{m-i-1} = \begin{bmatrix} 0 & I \end{bmatrix}$. The controllability and observability Gramians P_{m-i-1} and Q_{m-i-1} for the state space realization (24) are as follows

$$P_{m-i} = \begin{bmatrix} P_{m-i+1} & A_{m-i+1}P_{m-i+1}C^*_{m-i+1} + B_{m-i+1}e^*_{m-i} \\ C_{m-i+1}P_{m-i+1}A^*_{m-i+1} + e_{m-i}B^*_{m-i+1} & C_{m-i+1}P_{m-i+1}C^*_{m-i+1} + e_{m-i}e^*_{m-i} \end{bmatrix} \tag{25}$$

$$Q_{m-i} = \begin{bmatrix} Q_{m-1} & 0 \\ 0 & I \end{bmatrix}. \tag{26}$$

The proof is then completed by noting that we can now compute e_{m-i-2} by Theorem 2 again. ∎

The proof of Theorem 3 provides us an algorithm to compute the m-tap ($m-1$th order) suboptimal Hankel-norm FIR approximation of a given IIR filter $G(z) = D + C(zI - A)^{-1}B$. This algorithm is summarized below.

Algorithm 1

1) Set $G_m(z) = C_m(zI - A_m)^{-1}B_m$, where $C_m = CA^{m-1}$, $A_m = A$ and $B_m = B$.

2) Obtain P_m and Q_m by solving the Lyapunov equations (1) and (2) and compute $P_m^{\frac{1}{2}}$ and $Q_m^{\frac{1}{2}}$.

3) Compute the Hankel norm of $G_m(z)$ by any of the following equations

$$||\Gamma_{G_m}|| = \bar{\lambda}(Q_m P_m) = \bar{\lambda}(Q_m^{\frac{1}{2}}P_m Q_m^{\frac{1}{2}}) = \bar{\lambda}(P_m^{\frac{1}{2}}Q_m P_m^{\frac{1}{2}})$$

4) Obtain e_{m-1} by the following equation

$$e_{m-1} = -C_m P_m(\gamma^2 I - A^*_m Q_m A_m P_m)^{-1}A^*_m Q_m B_m. \tag{27}$$

5) Obtain a state space realization of $E_{m-1}(z)$ from (24) and obtain P_{m-1} and Q_{m-1}.

6) Repeat step 3) and 4) to find $e_{m-2}, \ldots, e_0$.

7) The optimal Hankel-norm approximant $F(z)$ is then given by

$$\begin{aligned}F(z) &:= \sum_{i=0}^{m-1} f_i z^{-i}\\ &= G_1(z) - \sum_{i=0}^{m-1} e_i z^{-i} = \sum_{i=0}^{m-1}(g_i - e_i)z^{-i}.\end{aligned}$$

Remark 1: Algorithm 1 only gives the solution in the case of $\gamma > \gamma_0$. We can also get the optimal solution by the equation (10) although it is not explicit. Actually, the optimal solution can be obtained from Algorithm 1 by simply using pseudo-inverse if $\gamma^2 I - A^*_{m-i}Q_{m-i}A_{m-i}P_{m-i}$ is singular.

The following Corollary gives the lower and upper bounds on the H_2 and H_∞ norms of the approximation error of the above algorithm.

Corollary 1: For $G(z) = G_1(z) + z^{-m+1}G_m(z)$, let $F(z)$ be obtained by algorithm 1. Then the following holds for the approximation error $E(z) = G(z) - F(z)$.

$$\begin{aligned}||\Gamma_{G_m}|| &\leq \|E(z)\|_\infty \leq 2\sum_{i=1}^{N}\sigma_i(\Gamma_E)\\ \|G_m(z)\|_2 &\leq \|E(z)\|_2 \leq ||\Gamma_{G_m}||.\end{aligned}$$

Proof: It is easy to see from the proof of Theorem 3 that $E(z) = z^{-1}E_0(z)$. The above inequalities then follow from Lemma 4 and the fact that $\|E(z)\|_2 = \|E_0(z)\|_2$ and $\|E(z)\|_\infty = \|E_0(z)\|_\infty$. ∎

Corollary 1 tells us that the upper bound on the H_∞ norm of approximation error is $2\sum_{i=1}^{N}\sigma_i(\Gamma_E)$. Actually we can achieve a tighter upper bound simply by another choice of f_0. The result is as follows, see [8], [10] for details.

Corollary 2: For $G(z) = G_1(z) + z^{-m+1}G_m(z)$, let $f_1 \cdots f_{m-1}$ be chosen as in algorithm 1. If f_0 is chosen such that $\|G(z) - F(z)\|_\infty$ is minimized, then we have

$$\|G(z) - F(z)\|_\infty \le \sum_{i=1}^{N} \sigma_i(\Gamma_E).$$

IV. APPROXIMATION FOR NON-CAUSAL SYSTEMS

There are situations to approximate a noncausal filters by causal filters [6], [15]. In this section, we will consider the problem of approximating a noncausal $G(z)$ by a causal FIR filter. First we consider the anticausal case. For an anticausal IIR filter $G(z) = \sum_{k=1}^{\infty} g_{-k}z^k$, denote the reverse operator $\mathcal{R}$, as $\mathcal{R}G(z) = \sum_{k=1}^{\infty} g_{-k}z^{-k}$. Then the Hankel norm and ith singular value of $G(z)$ are defined as $||\Gamma_G|| = ||\Gamma_{\mathcal{R}G}||$ and $\sigma_i(G(z)) = \sigma_i(\mathcal{R}G(z))$.

The problem is stated as follows: given an anticausal IIR filter $G(z) = \sum_{k=1}^{\infty} g_{-k}z^k$, find an FIR filter $F(z) = \sum_{k=0}^{m-1} f_k z^{-k}$ such that $\|\Gamma_E\|$ is minimized, where $E(z) = z^m(G(z) - F(z))$. Since we have developed the algorithm for causal IIR filters, the idea here is to convert the anticausal approximation problem to an equivalent causal problem.

Theorem 4: Given an anticausal IIR filter $G(z) = \sum_{k=1}^{\infty} g_{-k}z^k$, and $\gamma \ge \|\Gamma_G\|$. For

$$\mathbf{f} = [\ f_0 \quad f_1 \quad \cdots \quad f_{m-1}\ ],$$

define $F_1(z) = \sum_{k=0}^{m-1} f_k z^{-k}$ and $F_2(z) = \sum_{k=1}^{m} f_{m-k}z^{-k}$. Then $\|\Gamma_{E_1}\| \le \gamma$ if and only if $\|\Gamma_{E_2}\| \le \gamma$, where $E_1(z) = z^m(G(z) - F_1(z))$ and $E_2(z) = z^{-m}\mathcal{R}G(z) - F_2(z)$.
Proof. Note that

$$\begin{aligned} E_1(z) &= z^{m-1}(G(z) - F(z)) \\ &= -\sum_{k=1}^{m} f_{m-k}z^k + z^m \sum_{k=1}^{\infty} g_{-k}z^k. \end{aligned}$$

Then

$$\begin{aligned} \mathcal{R}E_1(z) &= -\sum_{k=1}^{m} f_{m-k}z^{-k} + z^{-m} \sum_{k=1}^{\infty} g_{-k}z^{-k} \\ &= -F_2(z) + z^{-m}\mathcal{R}G(z) = E_2(z). \end{aligned}$$

Therefore $\|\Gamma_{E_1}\| \le \gamma$ if only if $\|\Gamma_{E_2}\| \le \gamma$.

Theorem 4 shows that Algorithm 1 can be revised to find the (sub)optimal Hankel-norm FIR approximant for a given anticausal IIR filter. For completeness, We summarize it as Algorithm 2 below.

Algorithm 2

1) Set $G_{m+1}(z) = C_{m+1}(zI - A_{m+1})^{-1}B_{m+1}$, where $C_{m+1} = C$, $A_{m+1} = A$ and $B_{m+1} = B$.
2) Use Algorithm 1 to find $e_m, \ldots, e_1$.
3) Set the approximant as $F(z) = \sum_{k=0}^{m-1} e_{m-k}z^{-k}$.

Similar to Corollary 1, we have

Corollary 3: For $G(z) = G(z) = \sum_{k=0}^{\infty} g_{-k}z^k$, let $F(z)$ be obtained by Algorithm 2 and $E(z) = G(z) - F(z)$. Then we have

$$\begin{aligned} ||\Gamma_G|| &\le \|E(z)\|_\infty \le 2\sum_{i=1}^{N} \sigma_i(\Gamma_E) \\ \|G(z)\|_2 &\le \|E(z)\|_2 \le ||\Gamma_G||. \end{aligned}$$

Proof. The results are obvious from Corollary 1 by noting that the reverse operator preserve the norms.

So far, we have developed algorithm to obtain the Hankel-norm FIR approximation of causal and anticausal IIR filters respectively. For a general noncausal IIR filter, we have the following algorithm using the treatment of [6], [15].

Algorithm 3.

1. Decompose the noncausal $G(z)$ into its causal and anticausal (both stable) components: $G(z) = G_c(z) + G_a(z)$.
2. Using algorithm 1 to get the Hankel-norm FIR approximation $F_1(z)$ to $G_c(z)$.
3. Using algorithm 2 to get the Hankel-norm FIR approximation $F_2(z)$ to $G_a(z)$.
4. Set the approximating as $F(z) = F_1(z) + F_2(z)$.

Corollary 4: For $G(z) = G_c(z) + G_a(z)$, let $F(z)$ be obtained by Algorithm 3 and $E_1(z) = G_c(z) - F_1(z)$ and $E_2(z) = G_a(z) - F_2(z)$. Then we have

$$\begin{aligned} &\max\{||\Gamma_{G_a}||, ||\Gamma_{G_{cm}}||\} \le \|G(z) - F(z)\|_\infty \\ &\quad \le 2\sum_{i=1}^{N_1} \sigma_i(\Gamma_{E_1}) + 2\sum_{i=1}^{N_2} \sigma_i(\Gamma_{E_2}) \\ &\|G_a(z)\|_2^2 + \|G_{cm}(z)\|_2^2 \le \|G(z) - F(z)\|_2^2 \\ &\quad \le ||\Gamma_{G_a}||^2 + ||\Gamma_{G_{cm}}||^2. \end{aligned}$$

Proof. Note that

$$\begin{aligned} \max\{||E_1(z)||_\infty, ||E_2(z)||_\infty\} &\le \|G(z) - F(z)\|_\infty \\ &\le ||E_1(z)||_\infty + ||E_2(z)||_\infty \\ \|G(z) - F(z)\|_2^2 &= ||E_1(z)||_2^2 + ||E_2(z)||_2^2. \end{aligned}$$

The results then follow directly from Corollary 1 and Corollary 3.

V. COMPUTATION EXAMPLES

In this section, two examples from system modelling and filter design are given to illustrate our algorithms.

Example 1: Given below is a 6th order IIR model $G(z)$. This is the model of spindle vibration we obtained at a hot steel rolling mill for prediction and reduction of mechanical failure [16]. The model is non-minimum phase and has a pole very close to unit circle. Hence, it is prone to numerical error and not suitable for DSP implementation. To overcome this implementation difficulty, an FIR approximation is required.

$$G(z) = \frac{-0.1242z^5 + 0.1581z^4 + 0.5273z^3}{z^6 - 1.095z^5 + 1.299z^4 - 1.113z^3} \frac{+0.2154z^2 - 0.0647z^1 + 0.6889}{+1.028z^2 - 0.6043z + 0.426}$$

As shown in Figure 1, the model's frequency response spikes at about $\omega = 0.6, 1.4, 2.1$. These spikes, particularly those

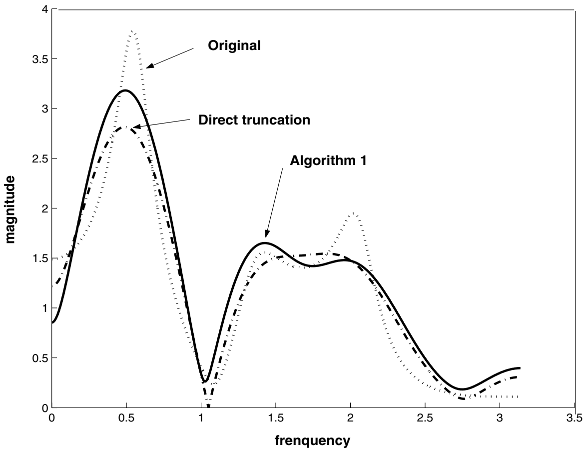

Fig. 1. Comparison of frequency response for $m = 12$

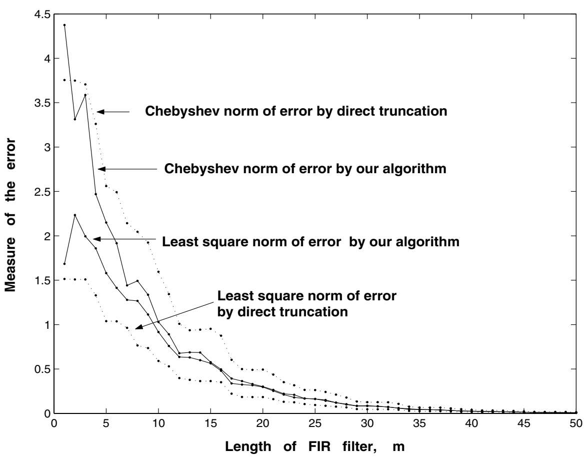

Fig. 2. Bounds of H_2 and H_∞ norms for error systems for Example 1

two at $\omega = 0.6, 1.4$ cause mechanical damage to the spindle [16]. Thus, for this particular application, we need an FIR approximation that better captures these two spikes. Now we use Algorithm 1 to find an FIR approximation of the model with length $m = 12$.

Figure 1 compares the frequency responses of the original IIR model and those of the 12-length FIR approximations obtained by Algorithm 1 and by direct truncation of impulse response. We can see from the figure that the FIR approximation of Algorithm 1 better captures the frequency spikes at $\omega = 0.6, 1.4$, whereas that of direct truncation tends to smooth out these spikes. Compared with the IIR model, the 12-length FIR approximation of Algorithm 1 has the same arithmetic complexity and very similar responses in the frequency range $0 \leq \omega \leq 1.75$ that is critical to the application. But it is numerically more robust since its intrinsic stability.

Figure 2 compares the H_∞ and H_2 norms of approximation errors achievable by Algorithm 1 with those of direct truncation. As can be seen from the figure, the H_∞ error norms are above the H_2 error norms, and the H_∞ (H_2) error norm achievable by Algorithm 1 is below (above) that of direct truncation. These agree with the analysis of Corollary 2, and demonstrate that Algorithm 1 truly provides a trade off between the H_∞ and H_2 approximation criteria.

VI. CONCLUSION

A constructive method is presented to obtain the optimal FIR Hankel norm approximation for a given IIR model. This method can provide a trade-off design between the worse case Chebyshev criterion and the least square criterion. Lower and upper bounds on the H_2 and H_∞ error norms are provided for the Hankel norm approximate. The effectiveness and properties of the proposed algorithm are demonstrated through a computation example. The algorithm can be extended to MIMO systems directly which may provide potential application to filter banks design and MIMO system modelling.

Acknowledgments: This research was supported in part by the National Natural Science Foundation of China under grant 60304011 and 60434020 and in part by the Australian Research Council under grant DP03430457.

REFERENCES

[1] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, McGraw-Hill, Irwin, 2nd edition, 2001.

[2] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 2nd edition, 1989.

[3] L. Ljung, *System Identification, Theory for the Users*, Prentice-Hall, 2nd edition, 1999.

[4] Y. Yamamoto, B.D.O. Anderson, M. Nagahara, and Y. Koyanagi, "Optimizing FIR approximation for discrete-time IIR filters," *IEEE Signal Processing Letters*, vol. 10, no. 9, pp. 273–276, September 2003.

[5] P. J. Kootsookos, R. R. Bitmead, and M. Green, "The Nehari shuffle: FIR(q) filter design with guaranteed error bounds," *IEEE Trans. Signal processing*, vol. 40, pp. 1876–1883, August 1992.

[6] P. J. Kootsookos and R. R. Bitmead, "The Nehari shuffle and Minimax FIR filter design," in *Control and Dynamic Systems*, vol. 64, pp. 239–298. Orlando, FL: Academic Press, 1994.

[7] J. W. Adams, "FIR digital filters with least-squares stopbands subject to peak-gain constraints," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 376–388, April 1991.

[8] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their l^∞-error bounds," *Int. J. Control*, vol. 39, pp. 1115–1193, June 1984.

[9] G. Obinata and B. D. O. Aderson, *Model Reduction for Control System Design*, Berli, Germany: Spring-Verlag, 2001.

[10] K. Zhou, J. C. Doyle, and K. Glover, *Robust Optimal Control*, Prentice-Hall, 1996.

[11] C. Foias, A. E. Frazho, I. Gohberg, and M. A. Kaashoek, *Metric Constrained Interpolation, Commutant Lifting and Systems*, Birkhäuser, 1998.

[12] C. Davis, W. M. Kahan, and H. F. Weinberger, "Norm-preserving dilitions and their applications to optimal error bounds," *SIAM Journal on Numerial Analysis*, vol. 19, no. 3, pp. 445–469, June 1982.

[13] C. K. Chui and G. Chen, *Discrete H^∞ Optimization: With Application in Signal Processing and Control Systems*, Springer-Verlag, 2nd edition, 1997.

[14] T. T. Georgiou and P. P. Khargonekar. "A constructive algorithm for sensitility optimization of periodic systems". *SIAM J. Control and Optimization*, 25:334–340, 1987.

[15] G. Gu and B. A. Shenoi, "A novel approach to the synthesis of recursive digital filters with linear phase," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 602–612, June 1991.

[16] K. Zhang, J. Zhang, and S. Nahavandi, "Mechanical failure reduction of hot steel rolling mills via modeling and control," in *The 4th Asian Control Conference*, Singapore, 2002, pp. 1630–1635.