A Unified Approach for the Design of 2-D Digital Filters via Semidefinite Programming

W.-S. Lu, *Fellow, IEEE*

***Abstract*—This paper attempts to demonstrate that a modern optimization methodology known as semidefinite programming (SDP) can be served as the algorithmic core of a unified design tool for a variety of two-dimensional (2-D) digital filters. Representative SDP-based designs presented in the paper include minimax and weighted least-squares designs of FIR filters with continuous and discrete coefficients, and minimax design of stable separable-denominator IIR filters. Our studies are motivated by the fact that SDP as a subclass of convex programming can be solved efficiently using recently developed interior-point methods and, more importantly, constraints on amplitude/phase responses in certain frequency regions and on stability (for IIR filters), that are often encountered in many filter design problems, can be formulated in a natural way as linear matrix inequalities (LMI) which allow SDP to apply. Design examples for each class of filters are included to demonstrate that SDP-based methods can in many cases be useful in producing optimal or near-optimal 2-D filters with reduced computational complexity.**

***Index Terms*—Constrained optimization, digital filters, semidefinite programming, stability.**

I. Introduction

A NUMBER of methods for the design of two-dimensional (2-D) digital filters have been developed during the past three decades [1]–[25]. These design techniques are quite diverse, and often times a specific technique that works well for a certain type of filters may not be suitable for other classes of filters. It is therefore desirable to have a unified approach available that is based on a single analytical machinery to carry out a variety of designs which at the very least considerably simplifies code preparation for the designs. This paper attempts to demonstrate that a modern optimization methodology known as *semidefinite programming* (SDP) can be served as the algorithmic core of such a unified design tool for a wide variety of 2-D digital filters.

The purposes of the present paper are three-fold: 1) to demonstrate the ability of SDP as a unified design tool. Although, due to space limitation, only a few representative designs can be discussed in detail here, it would become immediately apparent after going through one design formulation or two that the popular weighted least-squares (WLS) and minimax designs of 2-D FIR and stable IIR filters that approximate arbitrarily given amplitude *and* phase responses can be accomplished in an SDP setting; 2) to show a new approach known as SDP relaxation can be used to design near-optimal 2-D FIR filters with discrete (such as sum-of-power-of-two) coefficients with considerably improved design efficiency compared with mixed-integer-programming-based design methods; and 3) to illustrate the suitability of SDP for filter design problems by demonstrating that a variety of constraints on amplitude/phase responses in certain frequency regions and on stability (for IIR filters), that are often encountered in many filter design problems, can be converted into linear matrix inequalities (LMI) which are basic components of SDP.

One-dimensional (1-D) FIR filter design problems in relation to convex optimization (SDP in particular) have been considered in [26], but paper [26] is mainly concerned with "magnitude response" of the filters, and IIR filters and 2-D filters are not the subject of study. Early version of several results presented in this paper was presented at recent ISCAS meetings [27]–[29].

The paper is organized as follows. In Section II, a brief overview of SDP is given. The overview outlines primal and dual SDP problems, and the necessary and sufficient conditions for both the primal and dual solutions to exist. Section III deals with 2-D FIR filters where two design problems, i.e., minimax design of linear-phase filters and weighted least-squares design of nonlinear-phase filters with equiripple passbands and peak-constrained stopbands (ERPPCS) are considered. In Section IV, the design of FIR filters with sum-of-power-of-two (SP2) coefficients is investigated using an SDP relaxation (SDPR) technique. In Section V, we demonstrate that separable-denominator IIR filters with guaranteed stability can be designed by using an alternating SDP method. Design examples for each class of filters are included to illustrate that SDP-based methods can in many cases be useful in producing optimal or near-optimal 2-D filters with reduced computational complexity.

In the rest of the paper, boldfaced quantities denote matrices and vectors, $\mathbf{I}$ represents identity matrix of appropriate dimension, $\mathrm{diag}\{a_1,\ldots,a_n\}$ denotes a diagonal matrix with $a_1,\ldots,a_n$ as its diagonal, ω_p and ω_a denote normalized passband and stopband edges, respectively, and Ω denotes the base frequency band, i.e., $\Omega=\{(\omega_1,\omega_2):-\pi\le\omega_1\le\pi,-\pi\le\omega_2\le\pi\}$.

II. A Brief Overview of SDP

SDP is a relatively new optimization methodology, that is primarily concerned with minimizing a linear or convex quadratic objective function subject to linear matrix inequality (LMI) type

Manuscript received May 1, 2001; revised January 10, 2002. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. This paper was recommended by Co-Guest Editor S. Basu.

The author is with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, B.C., Canada, V8W 3P6 (e-mail: wslu@ece.uvic.ca).

Publisher Item Identifier S 1057-7122(02)05608-8.

constraints that depend on design variables *affinely* [30]. An important class of SDP problems can be described as

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} && \text{(1a)}\\ \text{subject to:} \quad & \mathbf{F}(\mathbf{x}) \succeq \mathbf{0} && \text{(1b)}\\ & \mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{i=1}^{n} x_i\mathbf{F}_i && \text{(1c)} \end{aligned}$$

where $\mathbf{x} = [x_1 \ \cdots \ x_n]^T, \mathbf{F}_i \in R^{m\times n}(i = 0, \ldots, n)$ are given symmetric matrices and $\mathbf{F}(\mathbf{x}) \succeq \mathbf{0}$ denotes that $\mathbf{F}(\mathbf{x})$ is positive semidefinite at $\mathbf{x}$. Note that the constraint matrix $\mathbf{F}(\mathbf{x})$ in (1) is *affine* with respect to (w.r.t.) $\mathbf{x}$. SDP includes both linear and quadratic programming as special cases, and it represents a broad and important class of convex programming problems [30]. More importantly, many interior-point methods, which have proven efficient for linear programming, have recently been extended to SDP [30], [31]. Efficient and user-friendly software implementations of various SDP algorithms are available. In particular we mention the LMI Control Toolbox [32], SeDuMi [33], and SDPT3 Toolbox [34], all of which work with MATLAB.

Another class of SDP problems, that is also relevant to the subject of the present paper, can be described as

$$\begin{aligned} \text{minimize} \quad & \mathbf{C}\cdot\mathbf{X} && \text{(2a)}\\ \text{subject to:} \quad & \mathbf{A}_i\cdot\mathbf{X} = b_i, \qquad \text{for } i = 1, 2, \ldots, p && \text{(2b)}\\ & \mathbf{X} \succeq \mathbf{0} && \text{(2c)} \end{aligned}$$

where $\mathbf{C}, \mathbf{X}, \mathbf{A}_i$ are symmetric matrices and $\cdot$ denotes the matrix inner product defined by

$$\mathbf{A}\cdot\mathbf{B} = \text{trace}(\mathbf{AB})$$

for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$. We stress that essentially the two SDP problems in (2) and (1) are related to each other as *primal* and *dual* problems. To see this, note that if the problem in (2) is referred to as the primal SDP problem, then its dual is given by [31]

$$\begin{aligned} \text{maximize} \quad & \mathbf{b}^T\mathbf{y} && \text{(3a)}\\ \text{subject to:} \quad & \sum_{i=1}^{p} y_i\mathbf{A}_i + \mathbf{S} = \mathbf{C} && \text{(3b)}\\ & \mathbf{S} \succeq \mathbf{0} && \text{(3c)} \end{aligned}$$

where $\mathbf{y} = [y_1 \ \cdots \ y_p]^T$ and $\mathbf{b} = [b_1 \ \cdots \ b_p]^T$. The slack variable $\mathbf{S}$ in (3b) and (3c) can be eliminated to obtain an equivalence of (3) as

$$\begin{aligned} \text{maximize} \quad & \mathbf{b}^T\mathbf{y}\\ \text{subject to:} \quad & \mathbf{C} - \sum_{i=1}^{p} y_i\mathbf{A}_i \succeq \mathbf{0} \end{aligned}$$

which can, in turn, be converted to the minimization problem

$$\begin{aligned} \text{minimize} \quad & -\mathbf{b}^T\mathbf{y} && \text{(4a)}\\ \text{subject to:} \quad & \mathbf{C} - \sum_{i=1}^{p} y_i\mathbf{A}_i \succeq \mathbf{0} && \text{(4b)} \end{aligned}$$

Obviously, with substitutions $\mathbf{y} \longmapsto \mathbf{x}, -\mathbf{b} \longmapsto \mathbf{c}, \mathbf{C} \longmapsto \mathbf{F}_0$, and $-\mathbf{A}_i \longmapsto \mathbf{F}_i$, (4) becomes the SDP problem in (1).

The necessary and sufficient conditions for $\mathbf{X}$ and $(\mathbf{y}, \mathbf{S})$ to be the solutions of the SDP problem in (2) and (3), respectively, are the so-called Karush–Kuhn–Tucker (KKT) conditions given below

$$\begin{aligned} & \sum_{i=1}^{p} y_i\mathbf{A}_i + \mathbf{S} = \mathbf{C} && \text{(5a)}\\ & \mathbf{A}\cdot\mathbf{X} = b_i, \qquad \text{for } i = 1, \ldots, p && \text{(5b)}\\ & \mathbf{SX} = \mathbf{0} && \text{(5c)}\\ & \mathbf{X} \succeq \mathbf{0}, \quad \mathbf{S} \succeq \mathbf{0}. && \text{(5d)} \end{aligned}$$

Efficient primal-dual interior-point methods for solving (5) have been developed in the past several years [30], [31]. Another interior-point method, developed by Nemirovski and Gahinet [35], [36], generates iterates using orthogonal projection of certain positive define matrix onto a linear subspace characterized by the constraint operator. The algorithm has proven efficient and gained popularity through its MATLAB implementation as a core algorithm for the LMI toolbox [32].

III. Design of Nonrecursive Filters

Let the transfer function of a 2-D FIR digital filter be

$$H(z_1, z_2) = \sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} h_{ij}z_1^{-i}z_2^{-j} = \mathbf{z}_1^T\hat{\mathbf{H}}\mathbf{z}_2$$

where N_1 and N_2 are odd integers, $\mathbf{z}_1 = [1 \ z_1^{-1} \ \cdots \ z_1^{-(N_1-1)}]^T, \mathbf{z}_2 = [1 \ z_2^{-1} \ \cdots \ z_2^{-(N_2-1)}]^T$, and $\hat{\mathbf{H}} \in R^{N_1\times N_2}$. In what follows we describe SDP-based methods for the minimax design of linear phase FIR filters (in Section III-A) and weighted least-squares design of nonlinear phase FIR filters with equirriple passbands and peak-constrained stopbands (ERPPCS) in Section III-B.

A. Minimax Design of Linear-Phase FIR Filters

To derive a compact expression for the transfer function of a *linear phase* 2-D FIR filter, we partition $\hat{\mathbf{H}}$ as

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{h}_{12} & \mathbf{H}_{13}\\ \mathbf{h}_{21}^T & h_{22} & \mathbf{h}_{23}^T\\ \mathbf{H}_{31} & \mathbf{h}_{32} & \mathbf{H}_{33} \end{bmatrix}$$

where $\mathbf{H}_{11}, \mathbf{H}_{13}, \mathbf{H}_{31}, \mathbf{H}_{33} \in R^{n_1\times n_2}, \mathbf{h}_{12}, \ \mathbf{h}_{32} \in R^{n_1\times 1}, \mathbf{h}_{21}, \mathbf{h}_{23} \in R^{n_2\times 1}, h_{22} \in R, n_1 = (N_1 - 1)/2$, and $n_2 = (N_2 - 1)/2$. Now, if the FIR filter is assumed to have linear-phase response with normalized constant group delay (n_1, n_2), then

$$\begin{aligned} & \mathbf{H}_{13} = \text{flipud}(\mathbf{H}_{33}), \mathbf{H}_{31} = \text{flipir}(\mathbf{H}_{33})\\ & \mathbf{H}_{11} = \text{flipud}(\text{fliplr}\,\mathbf{H}_{33}))\\ & \mathbf{h}_{12} = \text{flipud}(\mathbf{h}_{32}) \quad \text{and} \quad \mathbf{h}_{21}^T = \text{fliplr}\,(\mathbf{h}_{23}^T) \end{aligned}$$

where flipud and fliplr represent the operations of flipping a matrix upside down and from left to right, respectively, and the frequency response of the filter is given by

$$H(\omega_1, \omega_2) = e^{-j(n_1\omega_1 + n_2\omega_2)}\mathbf{c}_1^T(\omega_1)\mathbf{H}\mathbf{c}_2(\omega_2)$$

where $\mathbf{c}_i(\omega_i) = [\mathbf{1} \ \cos\omega_i \ \cdots \ \cos n_i\omega_i]^T$ for $i = 1$ and 2, and

$$\mathbf{H} = \begin{bmatrix} h_{22} & 2\mathbf{h}_{23}^T\\ 2\mathbf{h}_{32} & 4\mathbf{H}_{33} \end{bmatrix}.$$

Consequently, the minimax design of a linear-phase 2-D FIR filter can be obtained as the solution of the optimization problem

$$\text{minimize}_{\mathbf{H}}\text{maximize}_{(\omega_1,\omega_2)\in\Omega}|e(\omega_1,\omega_2)| \tag{6}$$

where

$$e(\omega_1,\omega_2)=W(\omega_1,\omega_2)\left[\mathbf{c}_1^T(\omega_1)\mathbf{H}\mathbf{c}_2(\omega_2)-A_d(\omega_1,\omega_2)\right]$$

and $A_d(\omega_1,\omega_2)$ is the desired amplitude response. Evidently, problem in (6) is equivalent to

$$\text{minimize}\quad \delta \tag{7a}$$

$$\text{subject to:}\quad e^2(\omega_1,\omega_2)\le\delta \quad \text{for } (\omega_1,\omega_2)\in\Omega \tag{7b}$$

Let $\mathbf{h}$ be the column vector generated by stacking the columns of $\mathbf{H}$ from its first column to the last column, then

$$e(\omega_1,\omega_2)=\mathbf{h}^T\mathbf{c}_w(\omega_1,\omega_2)-A_{dw}(\omega_1,\omega_2) \tag{8}$$

where $c_w(\omega_1,\omega_2)$ is the column vector whose components are given by

$$c_{kn_1+i}(\omega_1,\omega_2)=\cos i\omega_1\cos k\omega_2 \quad \text{for } 0\le i\le n_1, 0\le k\le n_2$$

$$A_{dw}(\omega_1,\omega_2)=W(\omega_1,\omega_2)A_d(\omega_1,\omega_2)$$

Hence, the constraint in (7b) becomes

$$\delta-[(\mathbf{h}^T\mathbf{c}_w(\omega_1,\omega_2)-A_{dw}(\omega_1,\omega_2)]^2\ge 0 \quad \text{for } (\omega_1,\omega_2)\in\Omega \tag{9}$$

which is equivalent to (10) shown at the bottom of the page, for $(\omega_1,\omega_2)\in\Omega$. Note that matrix $\mathbf{\Gamma}(\omega_1,\omega_2)$ in (10) is *affine* w.r.t. design parameter $\mathbf{h}$ and the scalar auxiliary variable δ. A discretized version of the positive semidefinite condition in (10) is given by

$$\mathbf{F}(\mathbf{x})=\text{diag}\left\{\mathbf{\Gamma}(\omega_1^{(1)},\omega_2^{(1)}),\ldots,\mathbf{\Gamma}\left(\omega_1^{(M)},\omega_2^{(M)}\right)\right\}\succeq\mathbf{0} \tag{11}$$

where the set $\{(\omega_1^{(i)},\omega_2^{(i)}), 1\le i\le M\}$ forms a sufficiently dense grid in region Ω, and $\mathbf{x}=[\delta\ \mathbf{h}^T]^T$ is a vector of dimension $1+(n_1+1)(n_2+1)$. Taking the above analysis into account, a discretized version of the optimization problem in (7) can be formulated as

$$\text{minimize}\quad \mathbf{c}^T\mathbf{x} \tag{12a}$$

$$\text{subject to:}\quad \mathbf{F}(\mathbf{x})\succeq\mathbf{0} \tag{12b}$$

where $\mathbf{c}=[1\ 0\ \cdots\ 0]^T$ and $\mathbf{F}(\mathbf{x})$ is defined by (11) and (10). Since matrix $\mathbf{F}(\mathbf{x})$ is affine w.r.t. $\mathbf{x}$, (12) is an SDP problem.

B. Weighted Least-Squares Design of FIR Filters With ERPPCS

1) Objective Function: The design objective in this case is to determine a transfer function $H(z_1,z_2)$ that minimizes the weighted least-squares error

$$e_2=\int\int_\Omega W(\omega_1,\omega_2)|H(\omega_1,\omega_2)-H_d(\omega_1,\omega_2)|^2\,d\omega_1\,d\omega_2 \tag{13}$$

subject to certain constraints (to be specified shortly). Unlike the design problem in Section III-A, here the filter to be designed is *not* constrained to have linear phase response. Consequently the desired frequency response $H_d(\omega_1,\omega_2)$ is allowed to be general with arbitrary amplitude and phase responses. By writing

$$H(\omega_1,\omega_2)=[\mathbf{c}_1(\omega_1)-j\mathbf{s}_1(\omega_1)]^T\hat{\mathbf{H}}[(\mathbf{c}_2(\omega_2)-j\mathbf{s}_2(\omega_2)]$$

$$\mathbf{c}_i(\omega_i)=[1\quad \cos\omega_i\quad \cdots\quad \cos(N_i-1)\omega_i]^T\quad i=1,2$$

$$\mathbf{s}_i(\omega_i)=[0\quad \sin\omega_i\quad \cdots\quad \sin(N_i-1)\omega_i]^T\quad i=1,2$$

$$H_d(\omega_1,\omega_2)=H_r(\omega_1,\omega_2)-jH_i(\omega_1,\omega_2)$$

we compute

$$|H(\omega_1,\omega_2)-H_d(\omega_1,\omega_2)|^2=\left(\mathbf{c}_1^T\hat{\mathbf{H}}\mathbf{c}_2-\mathbf{s}_1^T\hat{\mathbf{H}}\mathbf{s}_2-H_r\right)^2+\left(\mathbf{c}_1^T\hat{\mathbf{H}}\mathbf{s}_2+\mathbf{s}_1^T\hat{\mathbf{H}}\mathbf{c}_2-H_i\right)^2.$$

Now, let $\mathbf{h}$ be the column vector formed by orderly stacking the columns of coefficient matrix $\hat{\mathbf{H}}$. Then, one can write

$$\mathbf{c}_1^T(\omega_1)\hat{\mathbf{H}}\mathbf{c}_2(\omega_2)=\mathbf{h}^T\mathbf{c}(\omega_1,\omega_2)$$

$$\mathbf{s}_1^T(\omega_1)\hat{\mathbf{H}}\mathbf{s}_2(\omega_2)=\mathbf{h}^T\mathbf{s}(\omega_1,\omega_2)$$

$$\mathbf{c}_1^T(\omega_1)\hat{\mathbf{H}}\mathbf{s}_2(\omega_2)=\mathbf{h}^T\mathbf{u}(\omega_1,\omega_2)$$

$$\mathbf{s}_1^T(\omega_1)\hat{\mathbf{H}}\mathbf{c}_2(\omega_2)=\mathbf{h}^T\mathbf{v}(\omega_1,\omega_2)$$

where the components of vectors $\mathbf{c},\mathbf{s},\mathbf{u}$ and $\mathbf{v}$ are given, respectively, by

$$c_{kN_1+i}(\omega_1,\omega_2)=\cos i\omega_1\cos k\omega_2$$

$$s_{kN_1+i}(\omega_1,\omega_2)=\sin i\omega_1\sin k\omega_2$$

$$u_{kN_1+i}(\omega_1,\omega_2)=\cos i\omega_1\sin k\omega_2$$

$$v_{kN_1+i}(\omega_1,\omega_2)=\sin i\omega_1\cos k\omega_2$$

for $0\le i\le N_1-1, 0\le k\le N_2-1$. Hence, we have

$$|H(\omega_1,\omega_2)-H_d(\omega_1,\omega_2)|^2=[\mathbf{h}^T(\mathbf{c}-\mathbf{s})-H_d]^2+[\mathbf{h}^T(\mathbf{u}+\mathbf{v})-H_i]^2 \tag{14}$$

which leads (13) to

$$e_2=\mathbf{h}^T\mathbf{Q}\mathbf{h}-2\mathbf{h}^T\mathbf{q}+\boldsymbol{\kappa} \tag{15a}$$

where

$$\mathbf{Q}=\int\int_\Omega W(\omega_1,\omega_2)\mathbf{Y}(\omega_1,\omega_2)\mathbf{Y}^T(\omega_1,\omega_2)\,d\omega_1\,d\omega_2 \tag{15b}$$

$$\mathbf{q}=\int\int_\Omega W(\omega_1,\omega_2)\mathbf{b}(\omega_1,\omega_2)\,d\omega_1\,d\omega_2 \tag{15c}$$

$$\kappa=\int\int_\Omega W(\omega_1,\omega_2)|H_d(\omega_1,\omega_2)|^2\,d\omega_1\,d\omega_2 \tag{15d}$$

with

$$\mathbf{Y}(\omega_1,\omega_2)=[\mathbf{c}(\omega_1,\omega_2)-\mathbf{s}(\omega_1,\omega_2)\mathbf{u}(\omega_1,\omega_2)+\mathbf{v}(\omega_1,\omega_2)]$$

$$\mathbf{b}(\omega_1,\omega_2)=H_r(\omega_1,\omega_2)[\mathbf{c}(\omega_1,\omega_2)-\mathbf{s}(\omega_1,\omega_2)]+H_i(\omega_1,\omega_2)[\mathbf{u}(\omega_1,\omega_2)+\mathbf{v}(\omega_1,\omega_2)].$$

For a fixed pair of frequencies (ω_1,ω_2) in Ω, the integrand in (15b) is a rank-one positive semidefinite matrix. Since set Ω

$$\mathbf{\Gamma}(\omega_1,\omega_2)=\begin{bmatrix}\delta & \mathbf{h}^T\mathbf{c}_w(\omega_1,\omega_2)-A_{dw}(\omega_1,\omega_2)\\ \mathbf{h}^T\mathbf{c}_w(\omega_1,\omega_2)-A_{dw}(\omega_1,\omega_2) & 1\end{bmatrix}\succeq\mathbf{0} \tag{10}$$

contains a rich variety of frequency pairs (ω_1, ω_2), the integral (sum) of these positive semidefinite matrices in (15b) always leads to a full-rank, hence positive definite, matrix $\mathbf{Q}$ in practically every filter design problem. Therefore, e_2 in (15a) is globally convex w.r.t. $\mathbf{h}$.

2) Constraints: To achieve ERPPCS, it is imposed that

$$|H(\omega_1,\omega_2) - H_d(\omega_1,\omega_2)|^2 \le \delta_p \quad \text{for } (\omega_1,\omega_2) \in \Omega_p \tag{16a}$$

and

$$|H(\omega_1,\omega_2)|^2 \le \delta_a \quad \text{for } (\omega_1,\omega_2) \in \Omega_a \tag{16b}$$

where $\Omega_p = \{(\omega_1^{(i)}, \omega_2^{(i)}), 1 \le i \le k_p\}$ and $\Omega_a = \{(\omega_1^{(i)}, \omega_2^{(i)}, k_p+1 \le i \le k_p+k_a\}$ are sets of dense grid points in the passbands and stopbands, respectively. Using (14), these constraints can be expressed as

$$[\mathbf{h}^T(\mathbf{c}-\mathbf{s}) - H_d]^2 + [\mathbf{h}^T(\mathbf{u}+\mathbf{v}) - H_i]^2 \le \delta_p \quad \text{for } (\omega_1,\omega_2) \in \Omega_p \tag{17a}$$

and

$$[\mathbf{h}^T(\mathbf{c}-\mathbf{s})]^2 + [\mathbf{h}^T(\mathbf{u}+\mathbf{v})]^2 \le \delta_a \quad \text{for } (\omega_1,\omega_2) \in \Omega_a. \tag{17b}$$

3) An SDP Formulation of the Design Problem: Let δ be an upper bound of e_2 in (15a), i.e.

$$e_2 = \mathbf{h}^T\mathbf{Q}\mathbf{h} - 2\mathbf{h}^T\mathbf{q} + \kappa \le \delta. \tag{18}$$

Minimizing e_2 then amounts to minimizing the bound δ in (18). This simply means that the weighted least-squares design with ERPPCS can be formulated as

$$\text{minimize} \quad \delta \tag{19a}$$

$$\text{subject to: constraints in } (18), (17a), (17b). \tag{19b}$$

Note that

$$\begin{aligned} e_2 &= \mathbf{h}^T\mathbf{Q}\mathbf{h} - 2\mathbf{h}^T\mathbf{q} + \kappa \\ &= \|\mathbf{Q}^{1/2}\mathbf{h} - \mathbf{Q}^{-1/2}\mathbf{q}\|^2 - (\|\mathbf{Q}^{-1/2}\mathbf{q}\|^2 - \kappa) \end{aligned}$$

thus (18) is equivalent to

$$\delta + \kappa_1 - \|\mathbf{Q}^{1/2}\mathbf{h} - \hat{\mathbf{q}}\|^2 \ge 0 \tag{20}$$

where $\hat{\mathbf{q}} = \mathbf{Q}^{-1/2}\mathbf{q}$ and $\kappa_1 = \|\hat{\mathbf{q}}\|^2 - \kappa$ are respectively a constant vector and a constant scalar. It is easy to verify that (20) holds if and only if

$$\mathbf{\Gamma}_0 = \begin{bmatrix} \delta + \kappa_1 & \mathbf{h}^T\mathbf{Q}^{1/2} - \hat{\mathbf{q}}^T \\ \mathbf{Q}^{1/2}\mathbf{h} - \hat{\mathbf{q}} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \tag{21}$$

Furthermore, the constraints in (17a) and (17b) are equivalent to

$$\hat{\mathbf{\Phi}}(\omega_1,\omega_2) = \begin{bmatrix} \delta_p & \beta_1 & \beta_2 \\ \beta_1 & 1 & 0 \\ \beta_2 & 0 & 1 \end{bmatrix} \succeq \mathbf{0}, \qquad \text{for } (\omega_1,\omega_2) \in \Omega_p \tag{22}$$

and

$$\hat{\mathbf{\Psi}}(\omega_1,\omega_2) = \begin{bmatrix} \delta_a & \gamma_1 & \gamma_2 \\ \gamma_1 & \mathbf{1} & 0 \\ \gamma_2 & 0 & 1 \end{bmatrix} \succeq \mathbf{0}, \qquad \text{for } (\omega_1,\omega_2) \in \Omega_a \tag{23}$$

respectively, where $\beta_1, \beta_2, \gamma_1$, and γ_2 are scalar functions of (ω_1, ω_2) defined by

$$\begin{aligned} \gamma_1(\omega_1,\omega_2) &= \mathbf{h}^T[\mathbf{c}(\omega_1,\omega_2) - \mathbf{s}(\omega_1,\omega_2)] \\ \gamma_2(\omega_1,\omega_2) &= \mathbf{h}^T[\mathbf{u}(\omega_1,\omega_2) + \mathbf{v}(\omega_1,\omega_2)] \\ \beta_1(\omega_1,\omega_2) &= \gamma_1(\omega_1,\omega_2) - H_d(\omega_1,\omega_2) \\ \beta_2(\omega_1,\omega_2) &= \gamma_2(\omega_1,\omega_2) - H_i(\omega_1,\omega_2) \end{aligned}$$

The linear matrix inequality (LMI) constraints in (22) and (23) can be expressed, respectively, in compact form as

$$\mathbf{\Phi} \succeq \mathbf{0} \tag{24a}$$

with

$$\mathbf{\Phi} = \operatorname{diag}\left\{\hat{\mathbf{\Phi}}\left(\omega_1^{(1)}, \omega_2^{(1)}\right), \ldots, \hat{\mathbf{\Phi}}\left(\omega_1^{(k_p)}, \omega_2^{(k_p)}\right)\right\} \tag{24b}$$

and

$$\mathbf{\Psi} \succeq \mathbf{0} \tag{25a}$$

$$\mathbf{\Psi} = \operatorname{diag}\left\{\hat{\mathbf{\Psi}}\left(\omega_1^{(k_p+1)}, \omega_2^{(k_p+1)}\right), \ldots, \times\hat{\mathbf{\Psi}}\left(\omega_1^{(k_p+k_a)}, \omega_2^{(k_p+k_a)}\right)\right\}. \tag{25b}$$

By defining an augmented vector

$$\mathbf{x} = \begin{bmatrix} \delta \\ \delta_p \\ \delta_a \\ \mathbf{h} \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{26}$$

it is quite clear that the design problem at hand can be formulated as the optimization problem

$$\text{minimize} \quad \mathbf{c}^T\mathbf{x} \tag{27a}$$

$$\text{subject to:} \quad \mathbf{F}(\mathbf{x}) \succeq \mathbf{0} \tag{27b}$$

with

$$\mathbf{F}(\mathbf{x}) = \operatorname{diag}\{\mathbf{\Gamma}_0, \mathbf{\Phi}, \mathbf{\Psi}\}. \tag{27c}$$

Since matrices $\mathbf{\Gamma}_0, \mathbf{\Phi}, \mathbf{\Psi}$ depend on variable vector $\mathbf{x}$ affinely, (27) is an SDP problem.

We conclude this section with a remark that by including the constraint bounds δ_p and δ_a as a part of the design parameters (see (26)), these bounds are *optimized* together with the filter coefficients for given filter orders, frequency weights, and passband/stopband regions. This is considered as one of the features of the proposed design formulation that distinguishes itself from most existing least-squares designs.

C. Examples

The method proposed in Section III-A was applied to design circularly symmetric (CS) lowpass and diamond-shaped linear-phase FIR filters with order $(N\,N) = (7,7), (11,11), \ldots$, and $(23,23)$. For the CS lowpass filters, the following parameters were used: $\omega_p = 0.425\pi, \omega_a = 0.575\pi, W(\omega_1,\omega_2) = 1$ in both passband and stopband. For the diamond shaped filters, the design parameters are $\omega_p = 0.8\pi, \omega_a = \pi$, and

$$W(\omega_1,\omega_2) = \begin{cases} 1, & \text{for } (\omega_1,\omega_2) \in \Omega_p \\ w, & \text{for } (\omega_1,\omega_2) \in \Omega_a \end{cases}$$

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH THE METHOD IN [14]: CIRCULARLY SYMMETRIC LOWPASS FILTERS

N	Maximum ripple		CPU time in seconds		Number of grid points	
	Passband	Stopband	Proposed method	Method in [14]	M_p	M_a
7	0.2026	0.2348	3.06	14.41	50	279
11	0.1247	0.1591	5.87	27.51	50	279
15	0.0822	0.1115	10.47	53.01	50	279
19	0.0549	0.0830	20.34	102.47	50	279
23	0.0397	0.0578	39.22	199.62	50	279

TABLE II
COMPARISON OF THE PROPOSED METHOD WITH THE METHOD IN [14]: DIAMOND-SHAPED LOWPASS FILTERS

N	Maximum ripple		CPU time in seconds		Number of grid points	
	Passband	Stopband	Proposed method	Method in [14]	M_p	M_a
7	0.2468	0.2477	1.07	9.94	25	81
11	0.1212	0.1293	1.44	17.05	25	81
15	0.0782	0.0794	2.08	31.82	25	81
19	0.0469	0.0487	7.13	92.41	36	121
23	0.0298	0.0319	20.24	276.74	49	169

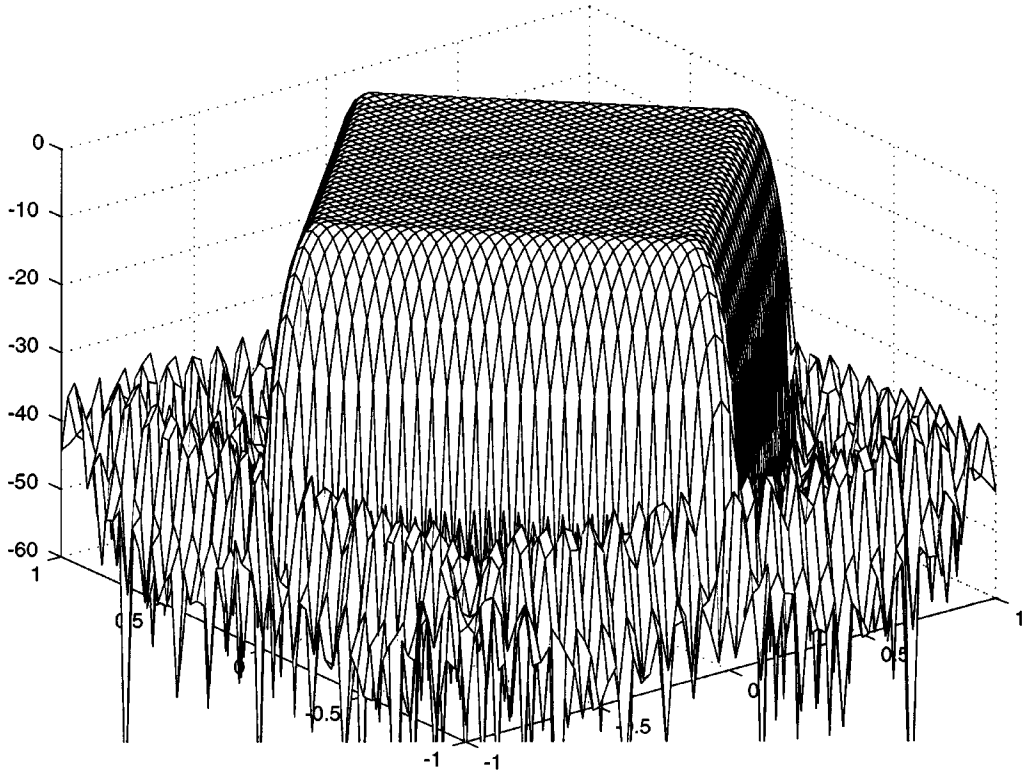

Fig. 1. Amplitude response (in decibels) of the diamond-shaped linear-phase FIR filter of order (23, 23).

where $w = 1$ for $N = 7$ and $11, w = 0.8$ for $N = 15, w = 0.6$ for $N = 19$, and $w = 0.4$ for $N = 23$. The maximum ripples in passband and stopband, and the CPU time used on a Pentium866 for a MATLAB implementation of the algorithm to accomplish the designs are given in Tables I and II Also listed in the tables are the number of grid points used in the passband (M_p) and stopband (M_a) with $M = M_p + M_a$. For comparison purposes, the same sets of filters were designed using the method in [14], which is based on some conjugate direction methods. As can be observed, both methods yield the same designs, but the SDP-based method required considerably less amount of computations (see Tables I and II). As a representative design, the amplitude response of the diamond-shaped FIR filter of order (23, 23) is shown in Fig. 1.

IV. DESIGN OF NONRECURSIVE FILTERS WITH DISCRETE COEFFICIENTS

The primary reason we are interested in digital filters with discrete coefficients, especially sum-of-power-of-two (SP2) coefficients, is that they admit fast implementation that requires no multiplications but simple superposition of shifted versions

of the input [37]–[43]. In this section, we describe an *SDP relaxation* (SDPR) algorithm for the design of 2-D FIR digital filters with SP2 coefficients. The method consists of four steps, namely: (i) design of a 2-D FIR filter with *continuous* coefficients that approximates a desired frequency response; (ii) formulation of the design of a 2-D FIR filter with SP2 coefficients (based on the continuous-coefficient filter obtained in Step (i)) as a $\{-1,1\}$-optimization problem; (iii) SDPR of the $\{-1,1\}$-optimization problem formulated in Step (ii); and (iv) derivation of a binary solution of the $\{-1,1\}$-optimization problem using the SDPR solution obtained from Step (iii). Since the design problem involved in Step (i) has largely been addressed in Section III, we shall focus on Steps (ii)–(iv) of the method in the rest of the section. Specifically, we assume that a linear-phase 2-D FIR transfer function $H_c(z_1,z_2)$ of order (N_1-1,N_2-1) with continuous coefficients has been designed by minimizing the WLS error e_2 in (13) for a desired frequency response $H_d(\omega_1,\omega_2)$, and that $H_c(\omega_1,\omega_2)=\mathbf{c}_1^T(\omega_1)\mathbf{H}_c\mathbf{c}_2(\omega_2)$ where $\mathbf{H}_c=\{h_{ij}^{(c)}\}$, $\mathbf{c}_i(\omega_i)=[1\ \ \cos\omega_i\ \ \cdots\ \ \cos n_i\omega_i]^T$ with $n_i=(N_i-1)/2$ and $i=1,2$.

A. Weighted Least-Squares $\{-1,1\}$-Optimization

Let the frequency response of linear-phase FIR filter with *discrete coefficients* be given by

$$H(\omega_1,\omega_2)=e^{-j(n_1\omega_1+n_2\omega_2)}\mathbf{c}_1^T(\omega_1)\mathbf{D}\mathbf{c}_2(\omega_2) \tag{28}$$

where $\mathbf{D}=\{d_{ij}\}$ and each d_{ij} is represented with m_{ij} power-of-two (P2) terms. Thus for a *budget* of M P2-terms, we have

$$\sum_{i=0}^{n_1}\sum_{j=0}^{n_2}m_{ij}=M. \tag{29}$$

Further, assume that the P2-terms used in the representation of $\{d_{ij}\}$ are constrained to within a given range, say, between 2^{-U} and 2^{-L} with $L<U$. Under these circumstances, each d_{ij} in $\mathbf{D}$ can be expressed as

$$d_{ij}=\sum_{k=1}^{m_{ij}}b_k^{(ij)}2^{-q_k^{(ij)}} \tag{30}$$

where $b_k^{(ij)}\in\{-1,1\}$ and $q_k^{(ij)}$ are integers between L and U.

Given a continuous-coefficient design $H_c(\omega_1,\omega_2)$, term budget $\{m_{ij}\}$, and range $\{L,U\}$, the *least* SP2 upper bound $\bar{d}_{ij}$ and *largest* SP2 lower bound $\underline{d}_{ij}$ for each continuous coefficient $h_{ij}^{(c)}$ can be readily determined such that $\underline{d}_{ij}\le h_{ij}^{(c)}\le\bar{d}_{ij}$ with both $\bar{d}_{ij}$ and $\underline{d}_{ij}$ being of the form (30), *and* in each open interval $(\underline{d}_{ij},\bar{d}_{ij})$, no SP2 numbers of form (30) exist for a given budget $\{m_{ij}\}$.

Having determined $\bar{d}_{ij}$ and $\underline{d}_{ij}$, an immediate *suboptimal* design of 2-D FIR filter with SP2 coefficients can be constructed as

$$H_s(\omega_1\ \ \omega_2)=e^{-j(n_1\omega_1+n_2\omega_2)}\mathbf{c}_1^T(\omega_1)\mathbf{D}_s\mathbf{c}_2(\omega_2) \tag{31a}$$

where $\mathbf{D}_s=\{d_{ij}^{(s)}\}$ with

$$d_{ij}^{(s)}=\begin{cases}\bar{d}_{ij}, & \text{if } \bar{d}_{ij}-h_{ij}^{(c)}\le h_{ij}^{(c)}-\underline{d}_{ij}\\ \underline{d}_{ij}, & \text{otherwise}\end{cases} \tag{31b}$$

In words, coefficients $\{d_{ij}^{(s)}\}$ are obtained as SP2 bounds of form (30) that are *closest* to $\{h_{ij}^{(c)}\}$. As will be demonstrated in Section IV-D, however, performance of this suboptimal design is often unsatisfactory. To obtain an improved design, we proceed by denoting the midpoint of each interval $[\underline{d}_{ij},\bar{d}_{ij}]$ as $d_{ij}^{(m)}=(\bar{d}_{ij}+\underline{d}_{ij})/2$ and a half of the interval length as $\delta_{ij}=(\bar{d}_{ij}-\underline{d}_{ij})/2$. The SP2 upper and lower bounds $\bar{d}_{ij}$ and $\underline{d}_{ij}$ can then be selected as $d_{ij}^{(m)}+b_{ij}\delta_{ij}$ with $b_{ij}=1$ and $b_{ij}=-1$, respectively. Hence the frequency response of discrete-coefficient FIR transfer function in (28) with

$$d_{ij}=d_{ij}^{(m)}+b_{ij}\delta_{ij} \tag{32}$$

can be expressed as

$$H(\omega_1\quad\omega_2)=e^{-j(n_1\omega_1+n_2\omega_2)}A(\omega_1,\omega_2)$$

where

$$A(\omega_1,\omega_2)=A_m(\omega_1,\omega_2)+\mathbf{c}_1^T(\omega_1)(\mathbf{B}\circ\boldsymbol{\Delta})\mathbf{c}_2(\omega_2) \tag{33}$$

where $A_m(\omega_1,\omega_2)=\mathbf{c}_1^T(\omega_1)\mathbf{D}_m\mathbf{c}_2(\omega_2)$ with $\mathbf{D}_m=\{d_{ij}^{(m)}\}$, and $\mathbf{B}\circ\boldsymbol{\Delta}$ denotes the pointwise product of $\mathbf{B}=\{b_{ij}\}$ and $\boldsymbol{\Delta}=\{\delta_{ij}\}$. Since the second term on the right-hand side of (33) is linear w.r.t. $\{b_{ij}\}$, we can write

$$A(\omega_1,\omega_2)=A_m(\omega_1,\omega_2)+\mathbf{b}^T\mathbf{v}(\omega_1,\omega_2,\boldsymbol{\Delta}) \tag{34}$$

where $\mathbf{b}=[b_{00}\ \ b_{10}\ \ \cdots\ \ b_{n_10}\ \ b_{01}\ \ \cdots\ \ b_{n_1n_2}]^T$ is obtained by orderly stacking the columns of $\mathbf{B}$, and $\mathbf{v}$ is a column vector of dimension $N=(n_1+1)(n_2+1)$ determined by $\mathbf{c}_1(\omega_1)$, $\mathbf{c}_2(\omega_2)$, and $\boldsymbol{\Delta}$. In the light of (34), the WLS objective function in (13) can be written as

$$e_2=\mathbf{b}^T\mathbf{Q}\mathbf{b}+2\mathbf{b}^T\mathbf{q}+\kappa \tag{35}$$

where

$$\mathbf{Q}=\iint_\Omega W(\omega_1,\omega_2)\mathbf{v}(\omega_1,\omega_2,\boldsymbol{\Delta})\mathbf{v}^T(\omega_1,\omega_2,\boldsymbol{\Delta})\times d\omega_1\,d\omega_2$$
$$\mathbf{q}=\iint_\Omega W(\omega_1,\omega_2)E(\omega_1,\omega_2)\mathbf{v}(\omega_1,\omega_2,\boldsymbol{\Delta})d\omega_1d\omega_2$$
$$E(\omega_1,\omega_2)=A_m(\omega_1,\omega_2)-A_d(\omega_1,\omega_2)$$

and a WLS design of $H(\omega_1,\omega_2)$ with SP2 coefficients can now be formulated as

$$\text{minimize}_{b_i\in\{-1,1\}}\mathbf{b}^T\mathbf{Q}\mathbf{b}+2\mathbf{b}^T\mathbf{q}. \tag{36}$$

Since the components of $\mathbf{Q}$ and $\mathbf{q}$ take real values while $\mathbf{b}$ assumes integer (binary) entries, (36) is a *mixed integer programming* (MIP) problem, which is a well-known NP-hard combinatorial optimization problem [43].

B. A Semidefinite Programming Relaxation of Problem (36)

The relevance of SDP to the design problem at hand lies in the fact that the MIP problem in (36) can be *relaxed* to an SDP problem so as to obtain an approximate solution which is of good quality and can be solved in polynomial time. Goemans and Williamson [44] was among the first to obtain an SDP relaxation solution of the MAX-CUT problem—a well-known

integer quadratic programming problem in graph theory. Following [44], SDP relaxations of various combinatorial optimization problems have been reported in graph optimization, network management and scheduling [31]. In what follows we present an SDP-relaxation-based solution to our design problem.

First, we replace design variable $\mathbf{b}$ in (36) by $\mathbf{x} = [x_1 \ \cdots \ x_N]^T$ and express the problem as

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{x}^T\mathbf{q} && \text{(37a)} \\ \text{subject to:} \quad & x_i^2 = 1 \quad \text{for } 1 \le i \le N && \text{(37b)} \end{aligned}$$

where constraints $x_i \in \{-1, 1\}$ have been expressed as the equality constraints in (37b). If we define matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, then the objective function in (37a) can be written as

$$\text{trace}(\hat{\mathbf{Q}}\hat{\mathbf{X}})$$

where

$$\hat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{q} \\ \mathbf{q}^T & 0 \end{bmatrix} \quad \hat{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & \mathbf{1} \end{bmatrix}. \tag{38}$$

Further notice that $\hat{\mathbf{X}}$ is positive semidefinite and that $\mathbf{X} = \mathbf{x}\mathbf{x}^T$ if and only if $\text{rank}(\hat{\mathbf{X}}) = 1$. Therefore, the problem in (37) is equivalent to

$$\begin{aligned} \text{minimize} \quad & \text{trace}(\hat{\mathbf{Q}}\hat{\mathbf{X}}) && \text{(39a)} \\ \text{subject to:} \ & \hat{\mathbf{X}} \succeq \mathbf{0} && \text{(39b)} \\ & \hat{X}_{ii} = 1 \quad \text{for } 1 \le i \le N+1 && \text{(39c)} \\ & \text{rank}(\hat{\mathbf{X}}) = 1. && \text{(39d)} \end{aligned}$$

An SDP relaxation of (39) is obtained by neglecting the rank constraint in (39d) while keeping the remaining two constraints, which leads to the minimization problem

$$\begin{aligned} \text{minimize} \quad & \text{trace}(\hat{\mathbf{Q}}\hat{\mathbf{X}}) && \text{(40a)} \\ \text{subject to:} \quad & \hat{\mathbf{X}} \succeq \mathbf{0} && \text{(40b)} \\ & \hat{X}_{ii} = 1 \quad \text{for } 1 \le i \le N+1. && \text{(40c)} \end{aligned}$$

If we define

$$\mathbf{A}_i = \text{diag}\{\underbrace{0, \ldots, 0, 1, 0, \ldots, 0}_{i}\} \tag{41}$$

then, the constraints in (40c) can be expressed as $\mathbf{A}_i \cdot \hat{\mathbf{X}} = 1$ for $i = 1, \ldots, N+1$, and the problem in (40) fits perfectly in with that of (2). Hence the problem in (40) is an SDP problem—a relaxation of the MIP problem in (37).

If we denote the minimum values of the objective functions in problems (39) and (40) by μ^* and ν^*, respectively, then because the feasible region of the problem in (39) is a subset of the feasible region of the problem in (40), we have $\nu^* \le \mu^*$. Furthermore, it has been shown [45] that if the diagonal elements of $\hat{\mathbf{Q}}$ are all zero and the off-diagonal elements of $\hat{\mathbf{Q}}$ are all nonnegative, then $\nu^* \ge 0.878\,56\mu^*$. Therefore, we have

$$0.878\,56\mu^* \le \nu^* \le \mu^*.$$

This indicates that the solution of the SDP problem in (40) under the above-mentioned conditions on $\hat{\mathbf{Q}}$ is a good approximation of the solution of the problem in (39). It should be stressed that for filter design problems, the conditions imposed on $\hat{\mathbf{Q}}$ are, in general, not satisfied, and a tight μ^*-dependent lower bound of ν^* has not been available. As will be demonstrated in Section IV-E, nevertheless, our design practice seems to indicate that suboptimal designs with satisfactory performance can be obtained through Goemans-Williamson SDP relaxation.

C. A Binary Solution of Problem (37)

Having solved the SDP problem in (39) for $\hat{\mathbf{X}}$, there are two approaches that can be used to generate a binary solution for the problem in (37). The first approach is based on the structure of $\hat{\mathbf{X}}$ in (38), i.e., if $\hat{\mathbf{X}}$ is the solution of (39), then $\mathbf{x}$ can be obtained from its last column immediately. This observation suggests a straightforward approach as follows. Let $\hat{\mathbf{X}}^*$ be the solution of (40), then, a suboptimal solution of (37) is obtained as

$$\mathbf{x} = \text{sign}(\hat{\mathbf{X}}^*(1 : N, N+1)) \tag{42}$$

where $\hat{\mathbf{X}}^*(1 : N, N+1)$ denotes the vector formed by the first N components of the last column of $\hat{\mathbf{X}}^*$.

At the cost of more computations, a better binary solution can be obtained using the eigen-decomposition of matrix $\hat{\mathbf{X}}^*$, i.e., $\hat{\mathbf{X}}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\hat{\mathbf{X}}^*$ on its diagonal in decreasing order. It is well known that an optimal rank-one approximation of $\hat{\mathbf{X}}^*$ in the 2-norm sense is given by $\lambda_1\mathbf{u}_1\mathbf{u}_1^T$, where λ_1 is the largest eigenvalue of $\hat{\mathbf{X}}^*$ and $\mathbf{u}_1$ is the eigenvector associated with λ_1. If we denote the vector formed by the first N components of $\mathbf{u}_1$ by $\tilde{\mathbf{u}}$ and the last component of $\mathbf{u}_1$ by u_{N+1}, i.e.

$$\mathbf{u}_1 = \begin{bmatrix} \tilde{\mathbf{u}} \\ u_{N+1} \end{bmatrix}$$

then, the optimal rank-one approximation of $\hat{\mathbf{X}}^*$ can be written as

$$\begin{aligned} \hat{\mathbf{X}}^* \approx \lambda_1\mathbf{u}_1\mathbf{u}_1^T &= \lambda_1 \begin{bmatrix} \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T & u_{N+1}\tilde{\mathbf{u}} \\ u_{N+1}\tilde{\mathbf{u}}^T & u_{N+1}^2 \end{bmatrix} \\ &= \frac{\lambda_1}{u_{N+1}^2} \begin{bmatrix} \tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_1^T & \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_1^T & 1 \end{bmatrix} \end{aligned}$$

where $\tilde{\mathbf{x}}_1 = \tilde{\mathbf{u}}/u_{N+1}$. Since $\lambda_1 > 0$, by comparing the above equation with $\hat{\mathbf{X}}$ in (38), we see that vector $\tilde{\mathbf{x}}_1$ is a reasonable approximation of $\mathbf{x}^*$ in terms of their signs. Therefore, a binary solution of the problem in (37) can be generated as

$$\mathbf{x} = \begin{cases} \text{sign}(\tilde{\mathbf{u}}), & \text{if } u_{N+1} > 0 \\ -\text{sign}(\tilde{\mathbf{u}}), & \text{if } u_{N+1} < 0 \end{cases} \tag{43}$$

D. An Efficient SDPR Solution via Duality

Although polynomial-time primal-dual interior-point algorithms [30], [31] can be applied to the SDP problem in (40), numerical difficulties may arise because of the involvement of a large number of variables even for filters of moderate order:

For a linear-phase 2-D filter of order (N_1-1, N_2-1), the dimension of vector $\mathbf{x}$ in (37) is $N=(N_1+1)(N_2+1)/4$, thus the number of variables in $\hat{\mathbf{X}}$ in the relaxed SDP problem (40) is $(N+1)N/2$. With $N_1=N_2=19$, for example, problem (40) involves as many as 5050 variables. It should be stressed that this "dimensionality" problem is quite common in SDP-relaxation based treatment of many combinational optimization problems [46]. In what follows, we present a method that first deals with the dual of the problem, which involve considerably less number of variables, and then converts the solution of the dual problem to that of the original problem.

It follows from Section II that the dual of (40) is given by

$$\text{minimize} \quad -\mathbf{b}^T\mathbf{y} \tag{44a}$$

$$\text{subject to:} \quad \hat{\mathbf{Q}}-\sum_{i=1}^{N+1} y_i\mathbf{A}_i \succeq \mathbf{0} \tag{44b}$$

where $\mathbf{y}=[y_1 \ \cdots \ y_{N+1}]^T, \mathbf{b}=[1 \ \cdots \ 1]^T, \hat{\mathbf{Q}}$ is defined in (38) and $\mathbf{A}_i$ are defined in (41). Note that the dual problem in (44) involves only $N+1$ variables. Interior-point methods, such as the projective methods proposed by Nemirovski and Gahinet [35], [36], that do not use a primal-dual solution setting, can be used to solve (44) efficiently.

To obtain the solution of the primal SDP problem in (40), we need a bit of analysis on the KKT conditions in (5). Let $\mathbf{y}^*$ be the solution of the dual problem (44), then the KKT conditions become

$$\sum_{i=1}^{N+1} y_i^*\mathbf{A}_i+\mathbf{S}^*=\hat{\mathbf{Q}} \tag{45a}$$

$$\mathbf{A}_i\cdot\hat{\mathbf{X}}^*=1 \quad \text{for } 1\le i\le N+1 \tag{45b}$$

$$\mathbf{S}^*\hat{\mathbf{X}}^*=\mathbf{0} \tag{45c}$$

$$\hat{\mathbf{X}}^*\succeq\mathbf{0}, \quad \mathbf{S}^*\succeq\mathbf{0}. \tag{45d}$$

From (45a), we have

$$\mathbf{S}^*=\hat{\mathbf{Q}}-\sum_{i=1}^{N+1} y_i^*\mathbf{A}_i. \tag{46}$$

Since $\mathbf{y}^*$ is obtained from an *interior-point iterative algorithm* (such as the projective method), it can only be a (good) approximate solution of (45) which is in the *interior* of the feasible region. Consequently, matrix $\mathbf{S}^*$ remains positive definite. This approximateness of $\mathbf{S}^*$ entails some modifications of (45c) and in this regard, we recall the concept of *central path* in the interior-point optimization theory. The central path is defined as a parameterized set $\{\mathbf{y}(\tau),\mathbf{S}(\tau),\hat{\mathbf{X}}(\tau), \text{for } \tau>0\}$ which satisfies the modified KKT conditions

$$\sum_{i=1}^{N+1} y_i(\tau)\mathbf{A}_i+\mathbf{S}(\tau)=\hat{\mathbf{Q}} \tag{47a}$$

$$\mathbf{A}_i\cdot\hat{\mathbf{X}}^*(\tau)=1 \quad \text{for } 1\le i\le N+1 \tag{47b}$$

$$\mathbf{S}(\tau)\hat{\mathbf{X}}(\tau)=\tau\mathbf{I} \tag{47c}$$

$$\mathbf{S}(\tau)\succ\mathbf{0} \quad \text{and} \quad \hat{\mathbf{X}}(\tau)\succ\mathbf{0}. \tag{47d}$$

Because of (47d), the entire central path lies in the interior of the feasible region and, as $\tau\to 0$, the path converges to the solution set $(\mathbf{y}^*,\mathbf{S}^*,\hat{\mathbf{X}}^*)$ of (45). This in conjunction with the fact (derived from (47c)) that

$$\hat{\mathbf{X}}(\tau)=\tau\mathbf{S}^{-1}(\tau) \tag{48}$$

suggests an approximate solution of (40) as

$$\hat{\mathbf{X}}=\tau\mathbf{S}^{*-1} \tag{49}$$

for some sufficiently small $\tau>0$, where $\mathbf{S}^*$ is obtained from (46). In order for $\hat{\mathbf{X}}$ in (49) to satisfy the equality conditions in (45b), $\hat{\mathbf{X}}$ is modified with a scaling matrix $\mathbf{\Pi}$ as

$$\hat{\mathbf{X}}=\mathbf{\Pi}(\mathbf{S}^*)^{-1}\mathbf{\Pi} \tag{50a}$$

where

$$\mathbf{\Pi}=\text{diag}\left\{\xi_1^{1/2},\ldots,\xi_{N+1}^{1/2}\right\} \tag{50b}$$

$$\xi_i=\text{the } i\text{th diagonal compnent of } (\mathbf{S}^*)^{-1}. \tag{50c}$$

Here, we pre- and post-multiply $\mathbf{\Pi}$ so that the modified $\hat{\mathbf{X}}$ in (50a) remains *symmetric* and positive definite. Note that because of the equality constraints (45b), the parameter τ in (49) has been absorbed by the scaling matrix $\mathbf{\Pi}$ in (50).

In summary, the steps for designing a linear-phase 2-D FIR filter with SP2 coefficients are as follows.

1) Design a linear-phase 2-D FIR filter with continuous coefficients $\mathbf{H}_c=\{h_{ij}^{(c)}\}$.
2) For a given term budget $\{m_{ij}\}$ and range $\{L,U\}$, find least SP2 upper bound $\bar{d}_{ij}$ and largest SP2 lower bound $\underline{d}_{ij}$ for each $h_{ij}^{(c)}$. Compute $d_{ij}^{(m)}=(\bar{d}_{ij}+\underline{d}_{ij})/2$ and $\delta_{ij}=(\bar{d}_{ij}-\underline{d}_{ij})/2$.
3) Compute matrix $\mathbf{Q}$ and vector $\mathbf{q}$ in (35) and form matrix $\hat{\mathbf{Q}}$ using (38).
4) Solving SDP problem (44) and denote its solution by $\mathbf{y}^*$.
5) Compute $\mathbf{S}^*$ using (46).
6) Compute $\hat{\mathbf{X}}$ using (50).
7) Obtain $\mathbf{x}$ using (42) or (43).
8) Set $\mathbf{b}=\mathbf{x}$ and compute the coefficient matrix $\mathbf{D}=\{d_{ij}\}$ using (32).

E. Design Examples

The SDP relaxation (SDPR) method described above was applied to design linear phase circularly symmetric lowpass 2-D FIR filters of order $(N-1,N-1)$ with SP2 coefficients for $N=3,5,7\ \ldots,31$. The prototype (continuous-coefficient) FIR filters were designed to minimize the WLS error in (13) with the following design parameters: $\omega_p=0.5\pi,\omega_a=0.7\pi,W\equiv 1$ in both passband and stopband, $L=0,U=12$, and the average number of P2 terms per coefficient was set to 2.0 for $3\le N\le 23$, 2.4 for $N=25,27$, and 2.6 for $N=29,31$. The designs obtained were compared with the suboptimal designs characterized by (31), the optimal designs with SP2 coefficients obtained

TABLE III
COMPARISONS OF VARIOUS DESIGNS

N	Design (31)	SDPR (error/CPU)	MIP (error/CPU)	WLS
3	0.1431	0.1375/0.03	0.1375/0.00	0.1339
5	0.0790	0.0769/0.06	0.0769/0.11	0.0768
7	0.0292	0.0281/0.11	0.0279/37.90	0.0272
9	0.0130	0.0095/0.33	0.0095/2.50e3	0.0094
11	0.0078	0.0061/0.54	0.0059/6.74e5	0.0056
13	0.0039	0.0017/1.21	—	0.0016
15	0.0017	0.0011/2.34	—	0.0010
17	0.8457e−3	0.4381e−3/4.59	—	0.4028e−3
19	0.3669e−3	0.2138e−3/7.53	—	0.1613e−3
21	0.2134e−3	0.1279e−3/12.85	—	0.1036e−3
23	0.1172e−3	0.0537e−3/23.92	—	0.0335e−3
25	0.5393e−4	0.2837e−4/48.61	—	0.2080e−4
27	0.3374e−4	0.1467e−4/97.45	—	0.0924e−4
29	0.3195e−4	0.0794e−4/181.33	—	0.0389e−4
31	0.2378e−4	0.0566e−4/303.87	—	0.0252e−4

using MIP, and the WLS designs with continuous coefficients. All the designs are performed using a Pentium866 where the algorithms involved were implemented using MATLAB 5. The results are summarized in Table III where the four types of designs are labeled as Design (31), SDPR, MIP, and WLS, respectively. It should be mentioned that for the SDPR designs, both (42) and (43) were examined to obtained the binary solutions $\mathbf{x}$ and, as expected, the rank-one approximation based approach (i.e., (43)) yielded consistently better designs. The results included in the SDPR column of the table were obtained using (43). For the SDPR and MIP designs, we also include the CPU time (in seconds) used (see the second figures in the SDPR and MIP columns of the table) as a measure of design efficiency. Due to the exceedingly long CPU time required by the MIP-based designs, we were only able to include the design data for five lower order designs. From the table it is observed that compared with the MIP-based designs the SDPR-based designs offered nearly optimal performance with considerably reduced computational complexity.

V. DESIGN OF RECURSIVE FILTERS

Our focus in this section is on weighted minimax design of stable 2-D IIR filters. Throughout the section the IIR filters are assumed to have *separable denominators*. This assumption simply imposes a constraint on the type of IIR filters being *quadrantally symmetric* [47]. Nevertheless this class of filters is broad enough to cover practically all types of IIR filters that have been found useful in image/video and other 2-D DSP applications.

A. Notation

Let the transfer function of a 2-D IIR digital filter be denoted by

$$H(z_1,z_2)=\frac{b(z_1,z_2)}{d(z_1,z_2)} \tag{51}$$

where

$$b(z_1,z_2)=\sum_{i=0}^{n}\sum_{k=0}^{m}b_{ik}z_1^{n-i}z_2^{m-k}$$

$$d(z_1,z_2)=f(z_1)g(z_2)$$

$$f(z_1)=z_1^{n-r_1}\sum_{i=0}^{r_1}f_iz_1^{r_1-i},\quad \text{with } f_0=1$$

$$g(z_2)=z_2^{m-r_2}\sum_{i=0}^{r_2}g_iz_2^{r_2-i},\quad \text{with } g_0=1$$

and r_1 and r_2 are integers with $0\le r_1\le n$ and $0\le r_2\le m$. This form of $d(z_1,z_2)$ is convenient to preserve a certain number of poles at the origin as it might be beneficial for the design of several types of digital filters [48]. The frequency response of the filter can now be written as

$$H(\omega_1,\omega_2)=\frac{b(\omega_1,\omega_2)}{d(\omega_1,\omega_2)} \tag{52}$$

where

$$b(\omega_1,\omega_2)=\sum_{i=0}^{n}\sum_{k=0}^{m}b_{ik}e^{-(i\omega_1+k\omega_2)j} \tag{53}$$

$$d(\omega_1,\omega_2)=\left(\sum_{i=0}^{r_1}f_ie^{-ji\omega_1}\right)\left(\sum_{i=0}^{r_2}g_ie^{-ji\omega_2}\right). \tag{54}$$

For the sake of description simplicity, we only consider the case $n=m$ and $r_1=r_2=r$. With straightforward modifications, our design algorithm can be applied to the cases where $n\neq m$ and $r_1\neq r_2$. The notation shown in the equation at the bottom of the page will be adopted in the rest of the section.

B. The Design Problem

The design problem considered here is to find a *stable* $H(z_1,z_2)$ that best approximates a given 2-D frequency response $H_d(\omega_1,\omega_2)$ in weighted minimax sense. Namely, $H(z_1,z_2)$ solves the constrained optimization problem

$$\text{minimize}_{\mathbf{f},\mathbf{g},\mathbf{b}}\quad \text{maximize}_{-\pi\le\omega_1,\omega_2\le\pi}|e(\omega_1,\omega_2)| \tag{55a}$$

$$\text{subject to:}\quad H(z_1,z_2)\text{ is stable} \tag{55b}$$

where

$$e(\omega_1,\omega_2) = W(\omega_1,\omega_2)[H(\omega_1,\omega_2) - H_d(\omega_1,\omega_2)]. \quad (55c)$$

C. Stability Constraints

The filter in (51) is stable if and only if $f(z_1)$ and $g(z_2)$ are stable 1-D polynomials, i.e., their zeros are strictly inside the unit circle. If we define the canonical matrices

$$\mathbf{D}_f = \begin{bmatrix} -f_1 & -f_2 & \cdots & -f_{r_1} \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \quad (56)$$

and

$$\mathbf{D}_g = \begin{bmatrix} -g_1 & -g_2 & \cdots & -g_{r_1} \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \quad (57)$$

then, an irreducible transfer function $b(z_1,z_2)/f(z_1)g(z_2)$ represents a stable IIR 2-D filter if and only if both $\mathbf{D}_f$ and $\mathbf{D}_g$ are stable, i.e., the maximum modulus of their eigenvalues is strictly less than 1. From the well known Lyapunov theory [49], it follows that $\mathbf{D}_f$ and $\mathbf{D}_g$ are stable if and only if there exist positive definite matrices $\mathbf{P}$ and $\mathbf{Q}$ such that

$$\mathbf{P} - \mathbf{D}_f^T \mathbf{P} \mathbf{D}_f \succ \mathbf{0} \quad (58)$$

and

$$\mathbf{Q} - \mathbf{D}_g^T \mathbf{Q} \mathbf{D}_g \succ \mathbf{0}. \quad (59)$$

It can be readily verified that (58) and (59) hold if and only if

$$\begin{bmatrix} \mathbf{P}^{-1} & \mathbf{D}_f \\ \mathbf{D}_f^T & \mathbf{P} \end{bmatrix} \succ \mathbf{0} \quad (60)$$

and

$$\begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{D}_g \\ \mathbf{D}_g^T & \mathbf{Q} \end{bmatrix} \succ \mathbf{0} \quad (61)$$

respectively. As will be seen in the next section, the LMI constraints in (60) and (61) are of convenience to use in an SDP-based design, because the design parameters in the denominator polynomials $f(z_1)$ and $g(z_2)$ appear *affinely*.

D. An Iterative SDP Formulation

First, we re-formulate the minimax problem in (55) as

$$\begin{aligned} \text{minimize} \quad & \delta && (62a)\\ \text{subject to:} \quad & |e(\omega_1,\omega_2)|^2 \le \delta && (62b)\\ & f(z_1) \neq 0, \quad \text{for } |z_1| \ge 1 && (62c)\\ & g(z_2) \neq 0, \quad \text{for } |z_2| \ge 1 && (62d) \end{aligned}$$

where the upper bound δ will be treated as an *auxiliary* design variable. Next we write the constraint in (62b) as

$$|e(\omega_1,\omega_2)|^2 = \frac{W^2}{|f|^2|g|^2}|b - fgH_d|^2 \le \delta \quad (63)$$

where for the simplicity of writing the dependence of functions W, b, f, g, and H_d on ω_1, ω_2 has been omitted. Note that the term $f_g H_d$ in (63) is *nonlinear* w.r.t. the design variables. This makes it difficult to convert the problem into a SDP problem because the square of the term $f_g H_d$ in magnitude is no longer *quadratic*. A remedy for this technical difficulty is to use an *alternating* iteration scheme as described below.

Suppose a stable pair, $\{f_0(z_1), g_0(z_2)\}$, has been chosen. For an iteration index $k \ge 1$, (63) suggests to first solve the following constrained problem for $b(z_1,z_2)$ and a stable $f(z_1)$:

$$\begin{aligned} \text{minimize} \quad & \delta && (64a)\\ \text{subject to:} \quad & \frac{W^2}{|f_{k-1}|^2|g_{k-1}|^2}|b - fg_{k-1}H_d|^2 \le \delta && (64b)\\ & f(z_1) \neq 0 \quad \text{for } |z_1| \ge 1 && (64c) \end{aligned}$$

Let the polynomial $f(z_1)$ obtained by solving (64) be denoted by $f_k(z_1)$. We seek to find a stable $g_k(z_2)$ and a $b_k(z_1,z_2)$ that solves the constrained problem

$$\begin{aligned} \text{minimize} \quad & \delta && (65a)\\ \text{subject to:} \quad & \frac{W^2}{|f_{k-1}|^2|g_{k-1}|^2}|b - f_k g H_d|^2 \le \delta && (65b)\\ & g(z_2) \neq 0 \quad \text{for } |z_2| \ge 1 && (65c) \end{aligned}$$

The iterations described above are similar in spirit to the Steiglitz-McBride (SM) scheme which finds applications in system identification and adaptive filtering [50]. The difference between (64), (65) and the SM scheme is that the SM scheme

---

$$\begin{aligned} \mathbf{b} &= [b_{00} \quad b_{10} \quad \cdots \quad b_{n0} \quad b_{01} \quad \cdots \quad b_{nn}]^T \\ \mathbf{f} &= [f_1 \quad f_2 \quad \cdots \quad f_r]^T \\ \mathbf{g} &= [g_1 \quad g_2 \quad \cdots \quad g_r]^T \\ \mathbf{c}(\omega_1,\omega_2) &= [1 \quad \cos\omega_1 \quad \cdots \quad \cos n\omega_1 \quad \cos\omega_2 \cdots \quad \cos(n\omega_1 + n\omega_2)]^T \\ \mathbf{s}(\omega_1,\omega_2) &= [0 \quad \sin\omega_1 \quad \cdots \quad \sin n\omega_1 \quad \sin\omega_2 \cdots \quad \sin(n\omega_1 + n\omega_2)]^T \\ \mathbf{c}_1(\omega_i) &= [\cos\omega_i \cdots \quad \cos r\omega_i]^T \quad \text{for } i = 1,2 \\ \mathbf{s}_1(\omega_i) &= [\sin\omega_i \cdots \quad \sin r\omega_i]^T \quad \text{for } i = 1,2 \\ H_d(\omega_1,\omega_2) &= H_r(\omega_1,\omega_2) - jH_i(\omega_1,\omega_2) \end{aligned}$$

iterates a least-squares objective function while each of (64) and (65) involves an iterative constraint in a minimax design.

With straightforward manipulations it can be verified that the constraint in (64b) can be expressed as a parameterized LMI constraint

$$\mathbf{\Phi}(\delta, \mathbf{x}, \omega_1, \omega_2) \succeq \mathbf{0} \tag{66}$$

where

$$\mathbf{\Phi}(\delta, \mathbf{x}, \omega_1, \omega_2) = \begin{bmatrix} \delta & a_1 & a_2 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{bmatrix}$$

with

$$\begin{aligned} a_1 &= \mathbf{x}^T \mathbf{c}_k - H_{a_1} \\ a_2 &= \mathbf{x}^T \mathbf{s}_k - H_{a_2} \\ \mathbf{x} &= \begin{bmatrix} \mathbf{b} \\ \mathbf{f} \end{bmatrix} \\ \mathbf{c}_k &= \begin{bmatrix} \mathbf{c}_w \\ \mathbf{u}_w \end{bmatrix} \\ \mathbf{s}_k &= \begin{bmatrix} \mathbf{s}_w \\ \mathbf{v}_w \end{bmatrix} \\ w_k &= \frac{W(\omega_1, \omega_2)}{|f_{k-1}(\omega_1) g_{k-1}(\omega_2)|} \\ \mathbf{c}_w &= w_k \mathbf{c}(\omega_1, \omega_2), \quad \mathbf{s}_w = w_k \mathbf{s}(\omega_1, \omega_2) \\ \mathbf{u}_w &= -\mathbf{c}_1(\omega_1) H_{a1} + \mathbf{s}_1(\omega_1) H_{a2} \\ \mathbf{v}_w &= -\mathbf{c}_1(\omega_1) H_{a2} - \mathbf{s}_1(\omega_1) H_{a1} \\ H_{a_1} &= w_k [H_r d_{k-1} - H_i e_{k-1}] \\ H_{a_2} &= w_k [H_r e_{k-1} + H_i d_{k-1}] \\ d_{k-1} &= 1 + \mathbf{g}_{k-1}^T \mathbf{c}_1(\omega_2) \\ e_{k-1} &= \mathbf{g}_{k-1}^T \mathbf{s}_1(\omega_2). \end{aligned}$$

We see that matrix $\mathbf{\Phi}(\delta, \mathbf{x}, \omega_1, \omega_2)$ depends on design parameters δ and $\mathbf{x}$ *affinely*, and the stability of $f_{k-1}(z)$ and $g_{k-1}(z)$ obtained from the $(k-1)$th iteration assures a well-defined weighting factor w_k.

Concerning the stability constraint in (64c), it follows from Section 5.C that for a stable $f_{k-1}(z_1)$ there exists a $\mathbf{P}_{k-1} \succ \mathbf{0}$ that satisfies the Lyapunov equation

$$\mathbf{P}_{k-1} - \mathbf{D}_{f_{k-1}}^T \mathbf{P}_{k-1} \mathbf{D}_{f_{k-1}} = \mathbf{I} \tag{67}$$

where $\mathbf{D}_{f_{k-1}}$ is the canonical matrix (56) with $-\mathbf{f}_{k-1}$ in its first row, and $\mathbf{I}$ is the $r \times r$ identify matrix. It now follows from (60) that a natural stability constraint for $f_k(z_1)$ is

$$\mathbf{S}_k(\mathbf{x}) = \begin{bmatrix} \mathbf{P}_{k-1}^{-1} - \tau \mathbf{I} & \mathbf{D}_f \\ \mathbf{D}_f^T & \mathbf{P}_{k-1} - \tau \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \tag{68}$$

where $\mathbf{D}_f$ is defined by (56) and $\tau > 0$ is a small scalar introduced to control the stability margin of $f_k(z_1)$. Since $\mathbf{S}_k(\mathbf{x})$ depends on $\mathbf{D}_f$ and hence $\mathbf{x}$ affinely, (68) is an LMI. Also notice that the positive define matrix $\mathbf{P}_{k-1}$ in (68) is obtained from (67), hence (68) is "constrained" by $\mathbf{D}_{f_{k-1}}$. As a result, (68) is a *sufficient* (but not necessary) constraint for the stability of $f_k(z_1)$.

At the k-th iteration the constrained optimization problem in (64) can now be formulated as

$$\text{minimize} \quad \hat{\mathbf{c}}^T \hat{\mathbf{x}} \tag{69a}$$

$$\text{subject to:} \quad \begin{bmatrix} \mathbf{\Phi}_k(\hat{\mathbf{x}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_k(\mathbf{x}) \end{bmatrix} \succeq \mathbf{0} \tag{69b}$$

where $\hat{\mathbf{x}}$ is the augmented variable defined by

$$\hat{\mathbf{x}} = \begin{bmatrix} \delta \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \delta \\ \mathbf{b} \\ \mathbf{f} \end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{69c}$$

and $\mathbf{\Phi}_k(\hat{\mathbf{x}}) \succeq \mathbf{0}$ is a discrete implementation of (66) on a set of frequencies $\{(\omega_1^{(m)}, \omega_2^{(m)})$, for $1 \le m \le M\}$ in a frequency region of interest

$$\mathbf{\Phi}_k(\hat{\mathbf{x}}) = \operatorname{diag}\left\{ \mathbf{\Phi}(\hat{\mathbf{x}}, \omega_1^{(1)}, \omega_2^{(1)}), \ldots, \mathbf{\Phi}\left(\hat{\mathbf{x}}, \omega_1^{(M)}, \omega_2^{(M)}\right) \right\}. \tag{69d}$$

Since both $\mathbf{\Phi}_k(\hat{\mathbf{x}})$ and $\mathbf{S}_k(\mathbf{x})$ depend on $\hat{\mathbf{x}}$ affinely, (69) is an SDP problem.

By a similar analysis, it can be verified that the companion optimization problem in (65) at the k-th iteration can be formulated as an SDP problem:

$$\text{minimize} \quad \hat{\mathbf{c}}^T \hat{\mathbf{y}} \tag{70a}$$

$$\text{subject to:} \quad \begin{bmatrix} \mathbf{\Psi}_k(\hat{\mathbf{y}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_k(\mathbf{y}) \end{bmatrix} \succeq \mathbf{0} \tag{70b}$$

where

$$\hat{\mathbf{y}} = \begin{bmatrix} \delta \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \delta \\ \mathbf{b} \\ \mathbf{g} \end{bmatrix} \tag{70c}$$

and matrices $\mathbf{\Psi}_k(\hat{\mathbf{y}})$ and $\mathbf{T}_k(\mathbf{y})$ are defined in a manner similar to that of $\mathbf{\Phi}_k(\hat{\mathbf{x}})$ and $\mathbf{S}_k(\mathbf{x})$, respectively.

E. The Algorithm

Given a desired frequency response $H_d(\omega_1, \omega_2)$, a weighting function $W(\omega_1, \omega_2)$, and filter order (n, r), one chooses a pair of convenient initial vectors $\mathbf{f}_0$ and $\mathbf{g}_0$ (for example, $\mathbf{f}_0 = \mathbf{0}$ and $\mathbf{g}_0 = \mathbf{0}$). Next, one solves the SDP problems in (69) and then (70) for $k = 1$, and evaluates $\varepsilon_k = \|\mathbf{x}_k - \mathbf{x}_{k-1}\| + \|\mathbf{y}_k - \mathbf{y}_{k-1}\|$. If ε_k is less then a prescribed tolerance ε, then the $\mathbf{b}_k, \mathbf{g}_k$ (from $\mathbf{y}_k$) and $\mathbf{f}_k$ (from $\mathbf{x}_k$) are deemed as the optimal solution for the design problem. Otherwise the algorithm proceeds by solving (69) and then (70) for $k = 2$, etc.

F. A Design Example

As an example, we applied the above algorithm to design a circular symmetric, lowpass IIR filter of order $(n, r) = (12, 8)$ with $\omega_p = 0.5\pi, \omega_a = 0.7\pi$, and linear phase response in the passband with group delay in both directions being 7.5 samples. A total of $M = 759$ grid points were used with 190 points in the passband and 569 points in the stopband. With $\varepsilon = 0.03, \tau = 10^{-12}, W(\omega_1, \omega_2) \equiv 1$ in the union of the passband and stopband regions and $W(\omega_1, \omega_2) \equiv 0$ elsewhere, it

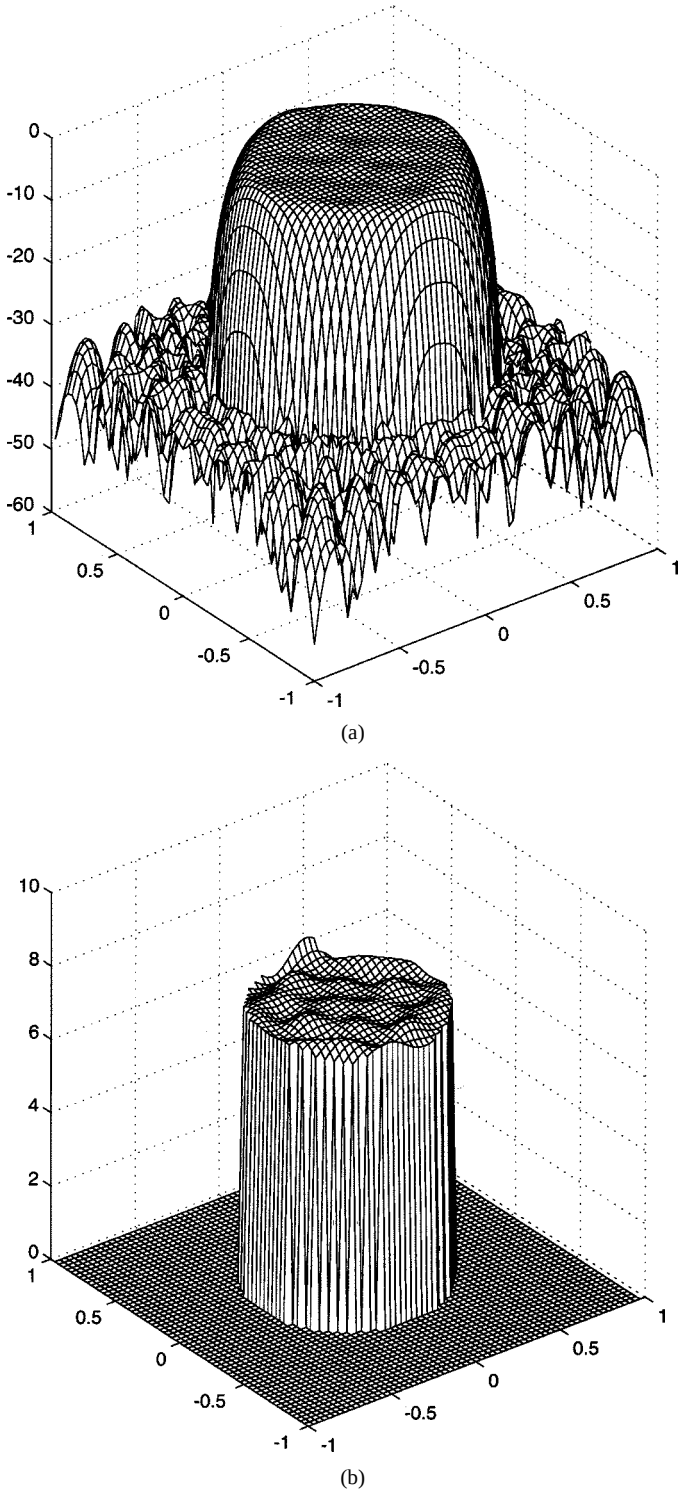

Fig. 2. (a) Amplitude responses (in decibels) of the IIR filter. (b) Group delay in passband.

took the proposed algorithm 16 iterations and 13.446 K seconds of CPU time to converge to a solution whose amplitude and phase (in passband) responses are depicted in Fig. 2. The maximum modulus of the poles of $b(z_1, z_2)/f(z_1)g(z_2)$ was 0.8238, and the maximum amplitude deviation in passband and stopband were 0.0315 and 0.0319, respectively. The maximum relative deviation in group delay in passband was 0.0886. The proposed method was compared with the SVD-based method [20]. The SVD-based method was applied to design a $(n, r) = (12, 12)$ lowpass circularly symmetric stable IIR filter with the same passband and stopband edges. The method started with designing a linear phase FIR filter of order (K, K) and a stable IIR filter of order (12, 12) was obtained by using the balanced approximation method. We tried a number of prototype FIR filters with K varying from 22 to 32, and the best result was obtained with $K = 26$ which yields a stable IIR filter whose maximum amplitude deviation in passband and stopband were 0.0578 and 0.0612, respectively, and the maximum relative deviation in group delay was 0.0861.

VI. Conclusion

We have attempted to show that SDP has the potential to serve as the optimization engine of a unified design tool for a wide range of 2-D digital filters. SDP is not only equipped with an ever-growing family of interior-point solution algorithms but offers the designer a natural setting in which many filter design problems can be formulated.

References

[1] T. S. Huang, "Two-dimensional windows," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 88–89, Mar. 1972.
[2] J. L. Shanks, S. Treitel, and J. H. Justice, "Stability and synthesis of two-dimensional recursive filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 115–128, June 1972.
[3] J. H. McClellan, "The design of two-dimensional filters by transformations," in *Proc. 7th Annual Conf. on Information Sciences and Systems*, 1973, pp. 247–251.
[4] G. A. Maria and M. M. Fahmy, "An L_p design technique for two-dimensional digital recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 15–21, Feb. 1974.
[5] J. M. Costa and A. N. Venetsanopoulos, "Design of circularly symmetric two-dimensional recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 432–443, Dec. 1974.
[6] N. A. Pendergrass, S. K. Mira, and E. I. Jury, "Spectral transformations for two-dimensional digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 26–35, Jan. 1976.
[7] R. M. Mersereau, W. F. G. Mecklenbräuker, and T. F. Quatieri, "McClellan transformations for two-dimensional filtering: I-design," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 405–413, July 1976.
[8] M. Ahmadi, A. G. Constantinides, and R. A. King, "Design technique for a class of stable two-dimensional recursive digital filters," in *Proc. IEEE Int. Conf. ASSP*, 1976, pp. 145–14.
[9] R. E. Twogood and S. K. Mitra, "Computer-aided design of separable two-dimensional digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 16–169, Feb. 1977.
[10] A. M. Ali, A. G. Constantinides, and R. A. King, "On 2-variable reactance functions for 2-dimensional recursive filters," *Electron. Lett.*, vol. 14, no. 1, pp. 12–13, 1978.
[11] D. M. Goodman, "A design technique for circularly symmetric lowpass filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 290–304, Aug. 1978.
[12] K. Hirano and J. K. Aggarwal, "Design of two-dimensional recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 106–1076, Dec. 1978.
[13] P. A. Ramamoorthy and L. T. Bruton, "Design of stable two-dimensional analog and digital filters with applications in image processing," *Int. J. Circuits Theory Appl.*, no. 4, pp. 229–246, 1979.
[14] C. Charalambous, "The performance of an algorithm on minimax design of two-dimensional linear phase FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 1016–1028, Oct. 1985.
[15] A. T. Chttera and G. A. Jullien, "Design of two-dimensional recursive digital filters using linear programming," *IEEE Trans. Circuits Syst.*, vol. CAS-29, pp. 817–826, Dec. 1982.
[16] T. Hinamoto and S. Maekawa, "Design of two-dimensional recursive digital filters using mirror image polynomials," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 750–758, Aug. 1986.

[17] A. Antoniou and W.-S. Lu, "Design of two-dimensional digital filters by using the singular value decomposition," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 1191–1198, Oct. 1987.
[18] W.-S. Lu, H.-P. Wang, and A. Antoniou, "Design of two-dimensional FIR digital filters by using the singular value decomposition," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 35–46, Jan. 1990.
[19] 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.
[20] W.-S. Lu, H.-P. Wang, and A. Antoniou, "Design of two-dimensional digital filters using singular value decomposition and balanced approximation method," *IEEE Trans. Signal Processing*, vol. 39, pp. 2253–2262, Oct. 1991.
[21] W.-P. Zhu, M. O. Ahmad, and M. N. S. Swamy, "An analytical approach for obtaining a closed-form solution to the least-square design problem of 2-D zero-phase FIR filters," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 796–807, Dec. 1994.
[22] ——, "Closed-form solution to the least-square design problem of 2-D linear-phase FIR filters," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 1032–1039, Dec. 1997.
[23] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[24] W.-S. Lu and A. Antoniou, *Two-Dimensional Digital Filters*. New York: Marcel-Dekker, 1992.
[25] B. A. Shenoi, *Magnitude and Delay Approximation of 1-D and 2-D Digital Filters*. Berlin, Germany: Springer Verlag, 1999.
[26] S. P. Wu, S. Boyd, and L. Vandenerghe, "FIR filter design via spectral factorization and convex optimization," in *Applied and Computational Control, Signal and Circuits*, B. Datta, Ed. Berlin, Germany: Birkhauser, 1998, ch. 5, pp. 215–245.
[27] W.-S. Lu and A. Antoniou, "Design of nonrecursive 2-D digital filters using semidefinite programming," in *Proc. ISCAS'99*, Orlando, FL, May 1999.
[28] W.-S. Lu, "Design of stable 2-D IIR digital filters using iterative semidefinite programming," in *Proc. ISCAS'00*, Geneva, Switzerland, May 2000.
[29] ——, "Design of 2-D FIR filters with power-of-two coefficients: A semidefinite programming relaxation approach," in *Proc. ISCAS'01*, Sydney, Australia, May 2001.
[30] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.
[31] *Handbook on Semidefinite Programming*, H. Wolkowicz, R. Saigal, and L. Vandenberghe, Eds., Kluwer, New York, 2000.
[32] P. Gahinet, A. Nemirovski, A. J. Laub, and M. Chilali, *Manual of LMI Control Toolbox*. Natick, MA: MathWorks, 1995.
[33] SeDuMi, Version 1.03, J. F. Sturm. (1999, Sept.). [Online]. Available: www2.unimaas.nl/~sturm/software/sedum:.html
[34] SDPT3 Version 2.1—A MATLAB Software for Semidefinite Programming, K. C. Toh, M. J. Tood, and R. H. Tütüncü. (1999, Sept.). [Online]. Available: www.math.ns.edu.sg/~mattohkc/
[35] A. Nemirovski and P. Gahinet, "The projective method for solving linear matrix inequalities," in *Proc. American Control Conf.*, Baltimore, MD, June 1994, pp. 840–844.
[36] P. Gahinet and A. Nemirovski, "The projective method for solving linear matrix inequalities," *Math. Prog.*, vol. 77, pp. 163–190, 1997.
[37] Y. C. Lim and S. R. Parker, "FIR filter design over a discrete power-of-two coefficient space," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 588–591, June 1983.
[38] S. C. Pei and S. B. Jaw, "Efficient design of 2-D multiplierless FIR filters by transformation," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 436–438, Apr. 1987.
[39] P. P. Vaidyanathan, "Efficient and multiplierless design of FIR filters with very sharp cutoff via maximally flat building blocks," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 236–244, Mar. 1985.
[40] P. Siohan and A. Benslimane, "Finite precision design of optimal linear phase 2-D FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 11–22, Jan. 1989.
[41] Y. C. Lim, "Design of discrete-coefficient-value linear phase FIR filters with optimum normalized peak ripple magnitude," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1480–1486, Dec. 1990.
[42] Y. C. Lim, R. Yang, D. Li, and J. Song, "Signed power-of-two (SPT) term allocation scheme for the design of digital filters," in *Proc. ISCAS*, Monterey, CA, June 1998.
[43] J. T. Yli-Kaakinen and T. A. Saramaki, "An Algorithm for the design of multiplierless approximately linear-phase lattice wave digital filters," in *Proc. ISCAS'2000*, vol. II, Geneva, June 2000, pp. 77–80.
[44] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problem using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.
[45] ——, ".878-approximation algorithm for MAX-CUT and MAX-2SAT," in *Proc. 26th ACM Symp. Theory of Computing*, 1994, pp. 422–431.
[46] C. Helmberg, *Semidefinite Programming for Combinational Optimization*. Berlin, Germany: Konrad-Zuse-Zentrum für Informationstechnik, Oct. 2000.
[47] P. K. Rajan and M. N. S. Swamy, "Quadrantal symmetry associated with two-dimensional digital transfer functions," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 340–343, June 1983.
[48] M. Lang, "Algorithms for constrained design of digital filters with arbitrary magnitude and phase responses," Ph.D. dissertation, Vienna Univ. of Technology, Vienna, Austria, June 1999.
[49] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1981.
[50] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker, 1995.

W.-S. Lu (S'81–M'85–SM'90–F'99) received the undergraduate degree in mathematics from Fudan University, Shanghai, China, in 1964, and the M.S. degree in electrical engineering, and Ph.D. degree in control science, both from University of Minnesota, Minneapolis, in 1983, and 1984, respectively.

He was a post-doctoral fellow at University of Victoria, Victoria, BC, Canada in 1985 and a visiting Assistant Professor at University of Minnesota in 1986. Since 1987, he has been with University of Victoria where he is currently a Professor. His teaching and research interests are in the areas of digital signal processing and numerical optimization. He is the co-author, with A. Antoniou, of *Two-Dimensional Digital Filters* (Marcel Dekker, 1992).

Dr. Lu was an Associate Editor of the *Canadian Journal of Electrical and Computer Engineering* in 1989, and the Editor of the same journal from 1990 to 1992. He served as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II from 1993 to 1995 and for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 1999 to 2001. He is currently an Associate Editor for the *International Journal of Multidimensional Systems and Signal Processing*, and a Fellow of the Engineering Institute of Canada.