CrossMark

A fast design algorithm for elliptic-error and phase-error constrained LS 2-D FIR filters

Xiaoying Hong · Xiaoping Lai · Ruijie Zhao

Abstract Two-dimensional (2-D) nonlinear-phase finite impulse response (FIR) filters have found many applications in signal processing and communication systems. This paper considers the elliptic-error and phase-error constrained least-squares design of 2-D nonlinear-phase FIR filters, and develops a matrix-based algorithm to solve the design problem directly for the filter's coefficient matrix rather than vectorizing it first as in the conventional methods. The matrix-based algorithm makes the design to consume much less design time than existing algorithms. Design examples and comparisons with existing methods demonstrate the effectiveness and high efficiency of the proposed design method.

Keywords 2-D nonlinear-phase FIR filters · Elliptic-error and phase-error constrained least-squares design · Matrix-based algorithm

1 Introduction

Optimal design of two-dimensional (2-D) nonlinear-phase finite impulse response (FIR) filters has attracted much attention of many researchers because they are more desirable than linear-phase FIR filters in many applications, e.g., the low group-delay FIR filters (Lu 2002; Lu and Hinamoto 2006) and fractional delay FIR filters (Deng and Lu 2000; Laakso et al. 1996; Shyu et al. 2009) in image and video processing, prototype FIR filters in 2-D filter banks (Dam et al. 2005; de Haan et al. 2003; Yan and Ikehara 1999) and allpass FIR

X. Hong (✉) · R. Zhao
School of Mechanical, Electrical and Information Engineering, Shandong University,
Weihai 264209, China
e-mail: hongxiaoyingwh@sohu.com

R. Zhao
e-mail: zhao_rj@163.com

X. Lai
Institute of Information and Control, Hangzhou Dianzi University, Hangzhou 310018, China
e-mail: laixp@hdu.edu.cn

filters in phase compensation (Zhu et al. 1999). In addition, optimal designs of 2-D infinite impulse response (IIR) filters are often transformed into convex subproblems in the same or similar formulations as 2-D nonlinear-phase FIR filter designs (Chottera and Jullien 1982; Dumitrescu 2005; Lu et al. 1998).

In constrained least-square (CLS) and minimax designs of 2-D nonlinear-phase FIR filters (Lu 2002; Lu and Hinamoto 2006), constraints are usually imposed merely on the filter's frequency response error. The resultant maximum magnitude error and phase error are about the same value. In order to obtain smaller magnitude error than phase error, Lai (2009) imposes constraints simultaneously on the frequency response error and phase error. The phase-error constrained minimax design in Lai et al. (2009) constrains the frequency response using elliptic-error and phase-error constraints, resulting in smaller magnitude error than those obtained by using frequency-response-error and phase-error constraints.

The impulse response coefficients of 2-D FIR filters are naturally arranged in matrices. Most existing algorithms for optimal designs of 2-D FIR filters, say, those in Algazi et al. (1986), Dumitrescu (2006), Gislason et al. (1993), Lai (2009), Lai et al. (2009), Lang et al. (1996), Lu (2002), Lu and Hinamoto (2006, 2011), Rusu and Dumitrescu (2012), rearrange the coefficient matrices into vectors and then solve the design problems using algorithms for one-dimensional (1-D) FIR filters. This leads to high computational complexities of the design algorithms and much design time for large size filters.

There are also several works (Ahmad and Wang 1989; Aravena and Gu 1996; Gu and Aravena 1994; Hong et al. 2013; Zhao and Lai 2011, 2013; Zhu et al. 1997, 1999) that formulate the design problems in terms of the coefficient matrices and develop algorithms to directly solve for the coefficient matrices rather than vectorizing them first, leading to lower computational complexities than corresponding algorithms vectorizing the coefficient matrices. The algorithms in Ahmad and Wang (1989), Aravena and Gu (1996), Gu and Aravena (1994), Zhao and Lai (2011, 2013), Zhu et al. (1997, 1999) only consider (weighted) least-squares designs of 2-D FIR filters without any other constraints. The algorithm in Hong et al. (2013) considers the constrained least-squares (CLS) design of 2-D FIR filters but is only applicable to the design of exact linear-phase filters.

This paper considers the elliptic-error and phase-error constrained least-squares (EPCLS) design of 2-D FIR filters. The problem can be solved by the Goldfarb–Idnani (GI) based algorithm in Lai (2009) after minor modifications. However, the algorithm in Lai (2009) is a vectorized one and thus has high computational complexity. To reduce the complexity, a matrix-based algorithm is developed in this paper to directly solve the EPCLS problem for the co-efficient matrix. Through design examples, the presented method is shown to be much more efficient than competing methods.

The rest of this paper is organized as follows. In Sect. 2, the EPCLS design of 2-D nonlinear-phase FIR filters is formulated as a semi-infinite convex quadratic programming (QP) problem in terms of the coefficient matrix. In Sect. 3, a matrix-based version of the GI-based algorithm in Lai (2009) is developed for the semi-infinite convex QP problem resulting from the EPCLS design. Design examples and comparisons with existing algorithms are given in Sect. 4. Conclusions of this paper are drawn in Sect. 5.

2 Matrix-based formulation for EPCLS design problem

The frequency response of a 2-D FIR filter with real impulse response coefficients $\{h(n_1, n_2) : n_1 = 0, 1, \ldots, N_1 - 1; n_2 = 0, 1, \ldots, N_2 - 1\}$ can be expressed as

$$G(\omega_1, \omega_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} h(n_1, \, n_2) e^{-i(n_1\omega_1 + n_2\omega_2)}, \tag{1}$$

where ω_1 and ω_2 are the horizontal and vertical frequencies, and i is the imaginary unit.

By introducing a function vector of ω,

$$\varphi(\omega, N) = [1, \, e^{-i\omega}, \, \ldots, \, e^{-i(N-1)\omega}]^H, \tag{2}$$

where the symbol "H" in the superscript denotes the conjugate transpose operation, the frequency response can be rewritten in a matrix form as

$$G(\omega_1, \omega_2, \mathbf{A}) = \varphi^H(\omega_1, N_1)\mathbf{A}\varphi(-\omega_2, N_2), \tag{3}$$

where $\mathbf{A}$ is an $N_1 \times N_2$ real matrix whose elements are the impulse response coefficients, i.e., $\mathbf{A} = [h(n_1, n_2)]$, $n_1 = 0, 1, \ldots, N_1 - 1$; $n_2 = 0, 1, \ldots, N_2 - 1$.

Denote by $D(\omega_1, \omega_2) = \bar{D}(\omega_1, \omega_2)e^{i\beta(\omega_1, \omega_2)}$ the desired complex frequency response, where $\bar{D}(\omega_1, \omega_2)$ and $\beta(\omega_1, \omega_2)$ are respectively the magnitude and phase of $D(\omega_1, \omega_2)$. The approximation of $D(\omega_1, \omega_2)$ by $G(\omega_1, \omega_2)$ is defined on the frequency plane $\Omega = \{(\omega_1, \omega_2)| - \pi \le \omega_1 \le, \pi, | - \pi \le \omega_2 \le \pi\}$. In practical designs, we discretize the continuous frequency plane Ω by a sufficiently dense rectangular frequency grid given by $\Pi = \{(\omega_{1j}, \omega_{2k})|\omega_{1j} = 2\pi(j-1)/(M_1-1) - \pi, \omega_{2k} = 2\pi(k-1)/(M_2-1) - \pi \text{ with } (j, k) \in F\}$, where $F = \{(j, k)|j = 1, 2, \ldots, M_1; k = 1, 2, \ldots, M_2\}$. Then the frequency response approximation error is defined as

$$E(\omega_1, \omega_2, \mathbf{A}) \equiv G(\omega_1, \omega_2, \mathbf{A}) - D(\omega_1, \omega_2), \quad \forall(\omega_1, \omega_2) \in \Pi, \tag{4}$$

and the magnitude approximation error and phase approximation error are represented by $E_m(\omega_1, \omega_2, \mathbf{A}) = |G(\omega_1, \omega_2, \mathbf{A})| - \bar{D}(\omega_1, \omega_2)$ and $E_p(\omega_1, \omega_2, \mathbf{A}) = \phi(\omega_1, \omega_2, \mathbf{A}) - \beta(\omega_1, \omega_2)$, where $\phi(\omega_1, \omega_2, \mathbf{A})$ is the phase of $G(\omega_1, \omega_2, \mathbf{A})$.

We further define a transformed error by

$$\bar{E}(\omega_1, \, \omega_2, \mathbf{A}) = e^{-i\beta(\omega_1, \omega_2)}E(\omega_1, \, \omega_2, \mathbf{A}), \tag{5}$$

and an elliptic error as

$$\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A}) = \text{Re}\left[\bar{E}(\omega_1, \omega_2, \mathbf{A})\right] + i\sigma^{-1}\text{Im}\left[\bar{E}(\omega_1, \omega_2, \mathbf{A})\right], \tag{6}$$

where $\sigma \ge 1$ is a prescribed model parameter. When $\sigma > 1$, the boundary of the constraint

$$\left|\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A})\right| \le \rho(\omega_1, \omega_2), \quad \forall(\omega_1, \omega_2) \tag{7}$$

for any $\rho(\omega_1, \omega_2) > 0$ is an ellipse, and thus the name elliptic error for $\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A})$. When $\sigma = 1$, $\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A})$ reduces to the circular error $\bar{E}(\omega_1, \omega_2, \mathbf{A})$ and the elliptic error constraint (7) reduces to the following circular error constraint:

$$|\bar{E}(\omega_1, \, \omega_2, \mathbf{A}) \le \rho(\omega_1, \omega_2). \tag{8}$$

In this paper, we consider the EPCLS design of the 2-D nonlinear-phase FIR filters. That is, we minimize the energy of the frequency response error and impose constraints simultaneously on the elliptic error and the phase error as follows:

$$\min_{\mathbf{A}} f(\mathbf{A}) = \sum\sum |E(\omega_1, \, \omega_2, \mathbf{A})|^2, \quad \forall(\omega_1, \omega_2) \in \Pi, \tag{9a}$$

$$\text{s.t.} : |\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A})| \le \rho(\omega_1, \omega_2), \quad \forall(\omega_1, \omega_2) \in \Pi_1, \tag{9b}$$

$$|\bar{E}(\omega_1, \omega_2, \mathbf{A})| \le \rho(\omega_1, \omega_2), \quad \forall(\omega_1, \omega_2) \in \Pi_2, \tag{9c}$$

$$|E_p(\omega_1, \omega_2, \mathbf{A})| \le \gamma(\omega_1, \omega_2), \quad \forall(\omega_1, \omega_2) \in \Pi_1, \tag{9d}$$

where Π_1 and Π_2 are two subsets of the passband and stopband frequencies on Π, "s.t." stands for "subject to", and $0 < \gamma(\omega_1, \omega 2) < \pi/2$. The above EPCLS problem (9) reduces to the CPCLS problem considered in Lai (2009) when $\sigma = 1$.

It is obvious that, $\{\mathbf{A}||\bar{E}(\omega_1, \omega_2, \mathbf{A})| \leq \rho(\omega_1, \omega_2)\} \subset \{\mathbf{A}||\bar{E}_\sigma(\omega_1, \omega_2, \mathbf{A})| \leq \rho(\omega_1, \omega_2)\}$ for any $(\omega_1, \omega_2) \in \Pi_1$. This implies that the feasible domain of the CPCLS problem in Lai (2009) is a subset of the feasible domain of EPCLS problem (9). Thus, solving the EPCLS problem (9) will result in better filters than solving the CPCLS problem in Lai (2009).

By introducing the following notation:

$$\mathbf{p}_j = \varphi(\omega_{1j}, N_1)(j = 1, 2, \cdots, M_1), \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{M1}], \tag{10a}$$

$$\mathbf{q}_k = \varphi(-\omega_{2k}, N_2)(k = 1, 2, \cdots, M_2), \mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{M2}], \tag{10b}$$

$$d_{jk} = D(\omega_{1j}, \omega_{2k})(j, k) \in F, \mathbf{D} = (d_{j \times k})_{M1 \times M2}, \tag{10c}$$

the frequency response $G(\omega_1, \omega_2, \mathbf{A})$ and cost function $f(\mathbf{A})$ can be rewritten as

$$G(\omega_{1j}, \omega_{2k}, \mathbf{A}) \equiv g_{jk}(\mathbf{A}) = \mathbf{p}_j^H \mathbf{A} \mathbf{q}_k, \quad \forall (j, k) \in F, \tag{11a}$$

$$f(\mathbf{A}) = \text{tr}\left\{\left(\mathbf{P}^H \mathbf{A} \mathbf{Q} - \mathbf{D}\right)^H \left(\mathbf{P}^H \mathbf{A} \mathbf{Q} - \mathbf{D}\right)\right\}, \tag{11b}$$

where $\text{tr}\{\bullet\}$ denotes the trace operation. It can be proved that $f(\mathbf{A})$ is convex when $M_1 \geq N_1$ and $M_2 \geq N_2$ (Zhu et al. 1999).

We further introduce $\rho_{jk} = \rho(\omega_{1j}, \omega_{2k})$, $\gamma_{jk} = \gamma(\omega_{1j}, \omega_{2k})$, $\bar{d}_{jk} = \bar{D}(\omega_{1j}, \omega_{2k})$, $\beta_{jk} = \beta(\omega_{1j}, \omega_{2k})$, $F_1 = \{(j, k)|(\omega_{1j}, \omega_{2k}) \in \Pi_1\}$, $F_2 = \{(j, k)|(\omega_{1j}, \omega_{2k}) \in \Pi_2\}$. Then, the elliptic-error constraints (9b), the circular-error constraints (9c) and the phase-error constraints (9d) can be transformed respectively into

$$\left(\cos(\alpha)\cos(\beta_{jk}) - \sigma^{-1}\sin(\alpha)\sin(\beta_{jk})\right) g_{jk}^R(\mathbf{A})$$
$$+ \left(\cos(\alpha)\sin(\beta_{jk}) + \sigma^{-1}\sin(\alpha)\cos(\beta_{jk})\right) g_{jk}^I(\mathbf{A}) \leq \rho_{jk} + \cos(\alpha)\bar{d}_{jk},$$
$$\forall \alpha \in [0, 2\pi), (j, k) \in F_1, \tag{12a}$$

$$\left(\cos(\alpha)\cos(\beta_{jk}) - \sin(\alpha)\sin(\beta_{jk})\right) g_{jk}^R(\mathbf{A})$$
$$+ \left(\cos(\alpha)\sin(\beta_{jk}) + \sin(\alpha)\cos(\beta_{jk})\right) g_{jk}^I(\mathbf{A}) \leq \rho_{jk}, \quad \forall \alpha \in [0, 2\pi), (j, k) \in F_2, \tag{12b}$$

$$\left(\cos(\alpha)\cos(\beta_{jk}) - \sin(\alpha)\sin(\beta_{jk})\right) g_{jk}^R(\mathbf{A}) + \left(\cos(\alpha)\sin(\beta_{jk}) + \sin(\alpha)\cos(\beta_{jk})\right)$$
$$g_{jk}^I(\mathbf{A}) \leq 0, \quad \forall \alpha \in \{\pm(\pi/2 + \gamma_{jk})\}, (j, k) \in F_1. \tag{12c}$$

where $g_{jk}^R(\mathbf{A})$ and $g_{jk}^I(\mathbf{A})$ are the real and imaginary parts of $g_{jk}(\mathbf{A})$ respectively, and α is an angle parameter. While the constraints in (12a) (12b) are linear but of infinite number because of the continuity of parameter α, the constraints in (12c) are linear and of finite number.

We now introduce a quadruplet (α, j, k, l) to index the constraints in (12), where $l \in \{1, 2, 3\}$, $(j, k) \in F_1$ and $\alpha \in [0, 2\pi)$ when $l = 1$, $(j, k) \in F_2$ and $\alpha \in [0, 2\pi)$ when $l = 2$, and $(j, k) \in F_1$ and $\alpha \in \{\pm(\pi/2 + \gamma_{jk})\}$ when $l = 3$. Let

$$\Theta \equiv [0, 2\pi), \quad \Xi_{jk} \equiv \{\pm(\pi/2 + \gamma_{jk})\}, \tag{13}$$

$$I_1 = \{(\alpha, j, k, l)|\alpha \in \Theta, (j, k) \in F_1, l = 1\}, \tag{14a}$$

$$I_2 = \{(\alpha, j, k, l)|\alpha \in \Theta, (j, k) \in F_2, l = 2\}, \tag{14b}$$

$$I_3 = \{(\alpha, j, k, l)|\alpha \in \Xi_{jk}, (j, k) \in F_1, l = 3\}. \tag{14c}$$

Then, the matrix-based EPCLS design problem (9) can be rewritten as

$$\min_{\mathbf{A}} f(\mathbf{A}) = \text{tr}\left\{ \left(\mathbf{P}^H \mathbf{A} \mathbf{Q} - \mathbf{D} \right)^H \left(\mathbf{P}^H \mathbf{A} \mathbf{Q} - \mathbf{D} \right) \right\}, \tag{15a}$$

$$\text{s.t. } \Gamma_{jkl}(\alpha) g_{jk}^R(\mathbf{A}) + \bar{\Gamma}_{jkl}(\alpha) g_{jk}^I(\mathbf{A}) \leq \theta_{jkl}(\alpha), \quad \forall (\alpha, j, k, l) \in I_1 \cup I_2 \cup I_3, \tag{15b}$$

where

$$\Gamma_{jkl}(\alpha) = \begin{cases} \cos(\alpha)\cos(\beta_{jk}) - \sigma^{-1}\sin(\alpha)\sin(\beta_{jk}), & (\alpha, j, k, l) \in I_1, \\ \cos(\alpha)\cos(\beta_{jk}) - \sin(\alpha)\sin(\beta_{jk}), & (\alpha, j, k, l) \in I_2 \cup I_3, \end{cases} \tag{15c}$$

$$\bar{\Gamma}_{jkl}(\alpha) = \begin{cases} \cos(\alpha)\sin(\beta_{jk}) + \sigma^{-1}\sin(\alpha)\cos(\beta_{jk}), & (\alpha, j, k, l) \in I_1, \\ \cos(\alpha)\sin(\beta_{jk}) + \sin(\alpha)\cos(\beta_{jk}), & (\alpha, j, k, l) \in I_2 \cup I_3, \end{cases} \tag{15d}$$

$$\theta_{jkl}(\alpha) = \begin{cases} \rho_{jk} + \cos(\alpha)\bar{d}_{jk}, & (\alpha, j, k, l) \in I_1, \\ \rho_{jk}, & (\alpha, j, k, l) \in I_2, \\ 0, & (\alpha, j, k, l) \in I_3. \end{cases} \tag{15e}$$

Problem (15) is a linearly constrained semi-infinite convex QP problem because the index set I_1 and I_2 are two infinite sets.

3 Matrix-based algorithm for linearly constrained semi-infinite convex QP problem

3.1 The matrix-based EPCLS-GI algorithm

In Sect. 2, the EPCLS design of 2-D FIR filters has been expressed as a matrix-based linearly constrained semi-infinite convex QP problem. In this section, an improved GI algorithm referred to as the matrix-based EPCLS-GI algorithm will be developed to solve the problem.

Like the original GI algorithm (Goldfarb and Idnani 1983), the matrix-based EPCLS-GI algorithm belongs to an active method that maintains an active set of constraints which are satisfied as equalities. We assume in the ℓth iteration, the active set is $J(\ell) = \{(\alpha_m, j_m, k_m, l_m), m = 1, 2, \ldots, t\}$. The solution matrix of problem (15a) subject only to the equality constraints indexed by the active set $J(\ell)$ is $\mathbf{A}(\ell)$, and the corresponding Lagrangian multiplier vector is $\boldsymbol{\lambda}(\ell)$.

To proceed, we first identify the most violated constraint among the infinite set of constraints in (12), or, (15) at the current solution matrix $\mathbf{A}(\ell)$. By introducing the following notation:

$$\bar{g}_{jk}(\mathbf{A}(\ell)) = \begin{cases} \text{Re}[e^{-i\beta_{jk}} g_{jk}(\mathbf{A}(\ell))] + i\sigma^{-1}\text{Im}[e^{-i\beta_{jk}} g_{jk}(\mathbf{A}(\ell))], & \text{for } (j,k) \in F_1, \\ \text{Re}[e^{-i\beta_{jk}} g_{jk}(\mathbf{A}(\ell))] + \text{Im}[e^{-i\beta_{jk}} g_{jk}(\mathbf{A}(\ell))], & \text{for } (j,k) \in F_2, \end{cases} \tag{16a}$$

the most violated constraint among the infinite set of constraints (15) with $(\alpha, j, k, l) \in I_1 \bigcup I_2$, if any, is chosen as the one indexed by $(\alpha_u, j_u, k_u, l_u)$ where

$$(j_u, k_u) = \arg \max_{(j,k) \in F_1 \cup F_2} \left\{ \left| \bar{g}_{jk}(\mathbf{A}(\ell)) - \bar{d}_{jk} \right| - \rho_{jk} \right\}, \tag{16b}$$

$$\alpha_u = \tan^{-1}\left(\frac{\text{Im}\left(\left| \bar{g}_{j_u k_u}(\mathbf{A}(\ell)) - \bar{d}_{j_u k_u} \right| \right)}{\text{Re}\left(\left| \bar{g}_{j_u k_u}(\mathbf{A}(\ell)) - \bar{d}_{j_u k_u} \right| \right)} \right), \tag{16c}$$

$$l_u = \begin{cases} 1, & \text{for } (j_u, k_u) \in F_1, \\ 2, & \text{for } (j_u, k_u) \in F_2. \end{cases} \tag{16d}$$

In the complex plane of $\bar{G}(\omega_{1j}, \omega_{2k}) \equiv \bar{g}_{jk}(\mathbf{A}(\ell))$, the most violated constraint is the one whose boundary is a tangent of the circle $|\bar{G}(\omega_{1j}, \omega_{2k}) - \bar{d}_{jk}| = \rho_{jk}$, the tangency point of which is the perpendicular projection of $\bar{G}(\omega_{1j}, \omega_{2k})$ on the circle along the radial direction with angle α_u (see, e.g., Lai 2009).

Among the finite set of constraints (15) with $(\alpha, j, k, l) \in I_3$, the most violated constraint, if any, is identified by the index quadruplet $(\alpha_v, j_v, k_v, l_v)$ with

$$(\alpha_v, j_v, k_v) = \arg \max_{\alpha \in \Xi (j,k) \in F_1} \{\Gamma_{jk3}(\alpha) g_{jk}^R(\mathbf{A}(\ell)) + \bar{\Gamma}_{jk3}(\alpha) g_{jk}^I(\mathbf{A}(\ell))\}, \quad (17a)$$

$$l_v = 3. \quad (17b)$$

Then, with Eqs. (16) and (17), the most violated constraint at the current iterative point $\mathbf{A}(\ell)$ among all constraints in (12) can be determined by Step 2 of the matrix-based EPCLS-GI algorithm given later in this section.

Suppose now we are going to add the most violated constraint to the active set $J(\ell)$ as the $(t + 1)$th active constraint to obtain the new active set J^+. Denote by $\mathbf{A}^+$ the new solution matrix to problem (15a) subject to the equality constraints indexed by J^+ and by $\boldsymbol{\lambda}^+$ the corresponding new multiplier vector. If we let

$$\boldsymbol{\lambda}^+ = \begin{bmatrix} \boldsymbol{\lambda}(\ell) \\ 0 \end{bmatrix} + w \begin{bmatrix} -\mathbf{r} \\ 1 \end{bmatrix}, \quad (18)$$

$$\mathbf{A}^+ = \mathbf{A}(\ell) + w\mathbf{Z}, \quad (19)$$

where the $[-\mathbf{r}^H, 1]^H$ is the step direction vector in the dual space, i.e., the direction from the old multiplier vector $[\boldsymbol{\lambda}(\ell)^H, 0]^H$ to the new multiplier vector $\boldsymbol{\lambda}^+$, $\mathbf{Z}$ is the step direction matrix in the primal space, i.e., the direction from the old solution matrix $\mathbf{A}(\ell)$ to the new solution matrix $\mathbf{A}^+$, and $w > 0$ is a hypothetical common step length in the dual and primal spaces. We have the following theorem regarding the primal and dual step directions. The proof of the theorem is given in the "Appendix".

Theorem 1 *The vector $\mathbf{r}$ and matrix $\mathbf{Z}$ in the dual and primal step directions can be given by*

$$\mathbf{r} = \mathbf{S}^{-1}\xi, \quad (20)$$

$$\mathbf{Z} = \sum_{m=1}^{t} r_m \mathbf{R}_m - \mathbf{R}_{t+1}, \quad (21)$$

where r_m represents the mth component of $\mathbf{r}$, and

$$\mathbf{S} = (S_{mn})_{t \times t}, \text{ where } S_{mn} = \Gamma_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^R(\mathbf{R}_n)$$
$$+ \bar{\Gamma}_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^I(\mathbf{R}_n), m, n = 1, 2, \ldots, t, \quad (22a)$$

$$\xi = (\xi_1, \xi_2, \ldots, \xi_t)^H, \text{ where } \xi_m = \Gamma_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^R(\mathbf{R}_{t+1})$$
$$+ \bar{\Gamma}_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^I(\mathbf{R}_{t+1}), m = 1, \ldots, t, \quad (22b)$$

$$\mathbf{R}_n = \left(\mathbf{PP}^H\right)^{-1} \left[\Gamma_{j_n k_n l_n}(\alpha_n) \text{Re}\left(\mathbf{p}_{j_n} \mathbf{q}_{k_n}^H\right) - \bar{\Gamma}_{j_n k_n l_n}(\alpha_n) \text{Im}\left(\mathbf{p}_{j_n} \mathbf{q}_{k_n}^H\right)\right] \left(\mathbf{QQ}^H\right)^{-1},$$
$$n = 1, \ldots, t + 1. \quad (22c)$$

After the primal and dual step directions are determined, we may compute the maximum dual step length w_1 and the minimum primal step length w_2. Then we can update the active set $J(\ell)$, solution matrix $\mathbf{A}(\ell)$ and multiplier vector $\boldsymbol{\lambda}(\ell)$ using the primal and dual step

directions and the primal and dual step lengths. The procedure details for determination of the step lengths and that for calculation of the active set, new solution matrix and new multiplier vector are described in Steps 4 and 5 of the following matrix-based EPCLS-GI algorithm for problem (15).

Step 1. Prepare the data matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{D}$ by (10), compute the solution to the unconstrained LS problem (15a) by

$$\mathbf{A}_0 = \left(\mathbf{PP}^H\right)^{-1} \mathbf{PDQ}^H \left(\mathbf{QQ}^H\right), \tag{23}$$

and initialize the algorithm by $\mathbf{S} = []$, $\mathbf{A}(0) = \mathbf{A}_0$, $\boldsymbol{\lambda}(0) = []$, $t = 0$, and $\ell = 0$.

Step 2. Identify the most violated constraint.

(a) Compute $(\alpha_u, j_u, k_u, l_u)$ and $(\alpha_v, j_v, k_v, l_v)$ by (16), (17), and then let

$$E_c^{mv} = \left|\bar{g}_{j_u k_u}(\mathbf{A}(\ell)) - \bar{d}_{j_u k_u}\right| - \rho_{j_u k_u}, \tag{24a}$$

$$E_p^{mv} = \Gamma_{j_v k_v 3}(\alpha_v) g_{j_v k_v}^R(\mathbf{A}(\ell)) + \bar{\Gamma}_{j_v k_v 3}(\alpha_v) g_{j_v k_v}^I(\mathbf{A}(\ell)). \tag{24b}$$

(b) If $E_c^{mv} \leq \varepsilon_1 \rho_{j_u k_u}$ and $E_p^{mv} \leq \varepsilon_2$ (where $\varepsilon_1 > 0$ is a tolerance of the relative complex error, $\varepsilon_2 > 0$ is a small positive number) which means no constraints are violated, then terminate the algorithm and $\mathbf{A}(\ell)$ is the solution to problem (15).

(c) If $E_p^{mv} \geq E_c^{mv}$, choose the constraint indexed by $(\alpha_v, j_v, k_v, l_v)$ as the most violated one. Let $(\alpha_{t+1}, j_{t+1}, k_{t+1}, l_{t+1}) = (\alpha_v, j_v, k_v, l_v)$, and compute $\mathbf{R}_{t+1}$ and ξ_ms$(m = 1, 2, \ldots, t)$ by (22c) and (22b).

(d) If $E_p^{mv} < E_c^{mv}$, choose the constraint indexed by $(\alpha_u, j_u, k_u, l_u)$ as the most violated one. Let $(\alpha_{t+1}, j_{t+1}, k_{t+1}, l_{t+1}) = (\alpha_u, j_u, k_u, l_u)$, and compute $\mathbf{R}_{t+1}$ and ξ_ms$(m = 1, 2, \ldots, t)$ by (22c) and (22b).

Step 3. Determine the step directions. Let $\mathbf{r}=[]$ if $t = 0$; Or compute $\mathbf{r}$ by (20) otherwise. Compute $\mathbf{Z}$ by (21).

Step 4. Compute the step lengths.

(a) Compute the maximum dual step length w_1 within which the dual feasibility would not be violated. If $\mathbf{r} \leq 0$ or $t=0$, let $w_1 = \infty$; Otherwise, let

$$\bar{m} = \arg \min_{\{m=1,\ldots,t|r_m>0\}} \left\{\frac{\lambda_m(\ell)}{r_m}\right\} \text{ and } w_1 = \frac{\lambda_{\bar{m}}(\ell)}{r_{\bar{m}}}, \tag{25}$$

where $\bar{m}$ is actually the subscript of the multiplier that would change sign earliest when the multiplier vector changes along the dual step direction.

(b) Compute the minimum primal step length w_2 with which the most violated constraint would become feasible. If $\|\mathbf{Z}\| = 0$, let $w_2 = \infty$; Otherwise, let

$$w_2 = \frac{\Gamma_{j_{t+1} k_{t+1} l_{t+1}}(\alpha_{t+1}) g_{j_{t+1} k_{t+1}}^R(\mathbf{A}(\ell)) + \bar{\Gamma}_{j_{t+1} k_{t+1} l_{t+1}}(\alpha_{t+1}) g_{j_{t+1} k_{t+1}}^I(\mathbf{A}(\ell)) - \theta_{j_{t+1} k_{t+1} l_{t+1}}(\alpha_{t+1})}{\Gamma_{j_{t+1} k_{t+1} l_{t+1}}(\alpha_{t+1}) g_{j_{t+1} k_{t+1}}^R(\mathbf{Z}) + \bar{\Gamma}_{j_{t+1} k_{t+1} l_{t+1}}(\alpha_{t+1}) g_{j_{t+1} k_{t+1}}^I(\mathbf{Z})}. \tag{26}$$

(c) Let $w = \min(w_1, w_2)$.

Step 5. Update the active set, and compute the new solution matrix and new multiplier vector.

(a) If $w = \infty$, problem (15) is infeasible, terminate the algorithm.

(b) If $w = w_1$, delete the $\bar{m}$th element of $\boldsymbol{\xi}$, and the $\bar{m}$ th column and $\bar{m}$th row of $\mathbf{S}$; Update $\mathbf{R}_m$s by $\mathbf{R}_m \leftarrow \mathbf{R}_{m+1}$ for $m = \bar{m}, \bar{m} + 1, \ldots, t$; Update the multiplier vector

by firstly computing $\boldsymbol{\lambda}(\ell + 1) = \boldsymbol{\lambda}^+$ with (18) and then deleting its $\bar{m}$th element; If $w_2 < \infty$, update the solution matrix by $\mathbf{A}(\ell + 1) = \mathbf{A}^+$ with (19); Update the active constraints by $(\alpha_m, j_m, k_m, l_m) \leftarrow (\alpha_{m+1}, j_{m+1}, k_{m+1}, l_{m+1})$ for $m = \bar{m}, \bar{m} + 1, \ldots, t$; Let $t \leftarrow t - 1$, $\ell \leftarrow \ell + 1$, and go back to Step 3.

(c) If $w = w_2$, update the multiplier vector and solution matrix by $\boldsymbol{\lambda}(\ell + 1) = \boldsymbol{\lambda}^+$ and $\mathbf{A}(\ell + 1) = \mathbf{A}^+$ using (18) and (19); Let $t \leftarrow t + 1$, $\ell \leftarrow \ell + 1$; Compute $\mathbf{S}$ by (22a) and go back to Step 2.

Remark 1 Since we use the perpendicular projection method to identify the most violated constraint and the GI primal–dual method to update the active set, solution matrix and multiplier vector, theoretically, without taking into account the numerical precision, the matrix-based EPCLS-GI algorithm has the same convergence properties, say, the same iteration number, as the vectorized CPCLS-GI algorithm in Lai (2009).

Remark 2 When the coefficient vector of a newly added (most violated) constraint obtained in Step 2 is linearly dependent on those of the constraints in the active set, it can be proved that $\mathbf{R}_{t+1} = \sum_{m=1}^{t} r_m \mathbf{R}_m$, and thus $\mathbf{Z} = 0$ from (21). In that case, the $\bar{m}$ th constraint will be deleted from the active set until the direction matrix $\mathbf{Z}$ becomes nonzero. This ensures the non-singularity of matrix $\mathbf{S}$.

Remark 3 Since the matrix $\mathbf{S}$ in each iterate is actually obtained either by adding to or by deleting from $\mathbf{S}$ in the previous iterate one column and one row, we can use the methods presented in Lemmas 1 and 2 of Hong et al. (2013) to update the inverse matrix $\mathbf{S}^{-1}$.

3.2 Complexity analysis and comparison

The computational complexity of the matrix-based EPCLS-GI algorithm is dominated by Steps 2, 3, 4 and 5. We assume the frequency grid lengths are $M_1 = L_1 N_1$ and $M_2 = L_2 N_2$, where L_1 and L_2 are the grid density constants. Then in each iterate, the multiplication number in Step 2 is $2M_1 N_1 N_2 + 4M_1 M_2 N_2 + 12M_1 M_2 + N_1^2 N_2 + N_1 N_2^2 + 2t N_1 N_2 + 8N_1 N_2 + 4N_2 t + 2t + 4N_2 + 3 = [(2L_1 + 1)N_1 + (4L_1 L_2 + 1)N_2 + 2t + 12L_1 L_2 + 8]N_1 N_2 + 4N_2 t + 2t + 4N_2 + 3$, that in Step 3 is $t^2 + t N_1 N_2$, that in Step 4 is $4N_1 N_2 + 8N_2 + t$, and that in Step 5 is $3N_1 N_2 + 2t^2 + 3t + 4N_2 + 3$ for adding a constraint and $N_1 N_2 + t^2$ for deleting a constraint, respectively.

The number t of the Lagrangian multipliers varies with iteration. We assume problem (15), or equivalently, the EPCLS problem (9), is feasible. Then, the value of t changes from zero to some $t_{max} \leq N_1 N_2$, where t_{max} depends on the tightness of the problem constraints. When t is very small, say, $t \leq \max\{N_1, N_2\}$, the complexity of the matrix-based EPCLS-GI algorithm for one iterate is $O(\max\{N_1, N_2\}N_1 N_2)$. When $t \approx N_1 N_2$, the complexity for one iterate is $O(N_1^2 N_2^2)$. Therefore, the complexity of the matrix-based EPCLS-GI algorithm for one iterate is between $O(\max\{N_1, N_2\}N_1 N_2)$ and $O(N_1^2 N_2^2)$.

For comparison, the complexity of the vectorized CPCLS-GI algorithm (Lai 2009) for the EPCLS design problem in one iterate is between $O(N_1^2 N_2^2)$ when $t \leq \max\{N_1, N_2\}$ and $O(N_1^3 N_2^3)$ when $t \approx N_1 N_2$. Since the matrix-based EPCLS-GI and vectorized CPCLS-GI algorithms consume the same number of iterations (Remark 1), the complexity of the matrix-based EPCLS-GI algorithm is smaller than the vectorized CPCLS-GI algorithm.

4 Design examples and comparisons

Two design examples are provided in this section. While Example 1 is mainly for the efficiency comparison between the matrix-based EPCLS-GI algorithm and the vectorized CPCLS-GI

algorithm in Lai (2009) and the SDP method in Lu (2002), Example 2 is provided to show the better performance of the filter obtained by the EPCLS design than that by the CPCLS design. All algorithms are implemented in MATLAB 7.11.0.584 (R2010b). All designs are conducted on a personal computer with an Intel(R) Core(TM) i7-4510U CPU @ 2.00 GHz and an 8.0 GB memory. We pointed out here that minor changes have been made to the original SDP method in Lu (2002) with merely circular complex-error constraints since the design problem in this paper has not only elliptic-error constraints but also phase-error constraints. Minor changes has also been made to the CPCLS-GI algorithm to meet the elliptic-error constrains. Moreover, the SDP method is implemented in this paper using the SDPT3 (Tütüncü et al. 2003) in CVX (Grant and Boyd 2013). Without loss of generality, we assume the 2-D FIR filters are of square size, i.e., $N_1 = N_2 = N$. The size of the frequency grid is taken as $M_1 = M_2 = 8N$. The phase of the desired complex frequency response is defined by

$$\beta(\omega_1, \omega_2) = -i(\tau_1 \omega_1 + \tau_2 \omega_2), \tag{27}$$

where τ_1 and τ_2 are the desired group delays along the ω_1 and ω_2 directions.

Both of the matrix-based EPCLS-GI and vectorized CPCLS-GI algorithms use the same stopping conditions, i.e., they return feasible solutions when $E_c^{mv} \leq \varepsilon_1 \rho_{i_u k_u}$ and $E_p^{mv} \leq \varepsilon_2$ with the same tolerance parameters $\varepsilon_1 = 10^{-5}$ and $\varepsilon_2 = 10^{-12}$.

Example 1 The first example considers the EPCLS design of circular lowpass filters with a passband $\Omega_p = \{(\omega_1, \omega_2) | \sqrt{\omega_1^2 + \omega_2^2} \leq \omega_p\}$ and a stopband $\Omega_s = \{(\omega_1, \omega_2) | \sqrt{\omega_1^2 + \omega_2^2} \geq \omega_s, |\omega_1| \leq \pi, |\omega_2| \leq \pi\}$, where $\omega_p = 0.4\pi$ and $\omega_s = 0.6\pi$. The desired group delays are τ_1 and τ_2 and the desired magnitude response is described by

$$\bar{D}(\omega_1, \omega_2) = \begin{cases} 1, & \omega \in \Omega_p, \\ \frac{\omega_s - \sqrt{\omega_1^2 + \omega_2^2}}{\omega_s - \omega_p}, & \omega \in \Omega \setminus (\Omega_p \cup \Omega_s), \\ 0, & \omega \in \Omega_s. \end{cases} \tag{28}$$

Such filters of size $N \times N$ with N ranging from 9 to 31 have been designed using the matrix-based EPCLS-GI algorithm, the vectorized CPCLS-GI algorithm and the SDP method. Table 1 gives a comparison of the three methods in terms of the mean squared complex error (MSCE) of the filter and the CPU time and iteration number of the algorithm under the specified maximum magnitude error (MME) (i.e., ρ), maximum phase error MPE (i.e., γ), group delays τ_1 and τ_2, and the model parameter τ. It is seen from Table 1 that, under the same N, ρ, γ, τ_1, τ_2 and σ, the matrix-based EPCLS-GI, vectorized CPCLS-GI and SDP methods obtain almost the same MSCE, but the matrix-based EPCLS-GI algorithm consumes much less CPU time than other two methods. For the designs with the same specifications, the matrix-based EPCLS-GI takes the same number of iterations as the vectorized CPCLS-GI. Figure 1 draws the magnitude response and passband phase error of the 31×31 circular filter with $\rho = 0.01$, $\gamma = 0.002$, $\tau_1 = \tau_2 = 13.5$ and $\sigma = 3$.

Filters with the same size $N = 21$ and group delays $\tau_1 = 8$ and $\tau_2 = 9$ but different ρs and γs have also been designed using the matrix-based EPCLS-GI and vectorized CPCLS-GI algorithms under $\sigma = 5$. Table 2 gives a comparison of the design results. It is seen that the matrix-based EPCLS-GI is much more efficient than the vectorized CPCLS-GI in all designs.

Example 2 The second example considers the design of fan filters with a passband $\Omega_p = \{(\omega_1, \omega_2) | 0 \leq \omega_1 \leq \omega_2 \tan \chi, 0 \leq \omega_2 \leq \pi\}$ and a stopband $\Omega_s = \{(\omega_1, \omega_2) | \phi \leq \omega_1 \leq \pi, 0 \leq \omega_2 \leq (\omega_1 - \phi) \tan \chi\}$, where $\chi = \pi/6$, and $\phi = 0.1\pi$ is a positive number used to

Table 1 Efficiency comparison of the matrix-based EPCLS-GI with the vectorized CPCLS-GI and SDP using SDPT3 in CVX in Example 1

Filter specifications and model parameter σ	Design algorithm	MSCE (10^{-4})	CPU time(s)[a]/iterates
$N = 9, \rho = 0.09, \gamma = 0.04,$	Matrix-based EPCLS-GI	23.604	0.156/105
$\tau_1 = \tau_2 = 3, \sigma = 1$	vectorized CPCLS-GI	23.604	0.454/105
	SDP using SDPT3 in CVX	23.604	27.44/30
$N = 13, \rho = 0.04, \gamma = 0.02,$	Matrix-based EPCLS-GI	4.7362	0.672/325
$\tau_1 = \tau_2 = 5, \sigma = 1.5$	vectorized CPCLS-GI	4.7362	3.031/325
	SDP using SDPT3 in CVX	4.7362	233.5/33
$N = 15, \rho = 0.035, \gamma = 0.02,$	Matrix-based EPCLS-GI	3.9803	1.25/501
$\tau_1 = \tau_2 = 6, \sigma = 2$	vectorized CPCLS-GI	3.9803	6.785/501
	SDP using SDPT3 in CVX	3.9804	524.4/32
$N = 21, \rho = 0.013, \gamma = 0.004,$	Matrix-based EPCLS-GI	3.0317	3.86/877
$\tau_1 = \tau_2 = 9, \sigma = 2$	vectorized CPCLS-GI	3.0317	37.11/877
	SDP using SDPT3 in CVX	3.0317	3,726/32
$N = 31, \rho = 0.01, \gamma = 0.002,$	Matrix-based EPCLS-GI	1.5069	33.37/2,807
$\tau_1 = \tau_2 = 13.5, \sigma = 3$	vectorized CPCLS-GI	1.5069	527.4/2,807
	SDP using SDPT3 in CVX	1.5069	45,764/36

[a] CPU time for the SDP method using SDPT3 in CVX does not include the time of parsing the problem and formatting it in the input format of SDPT3
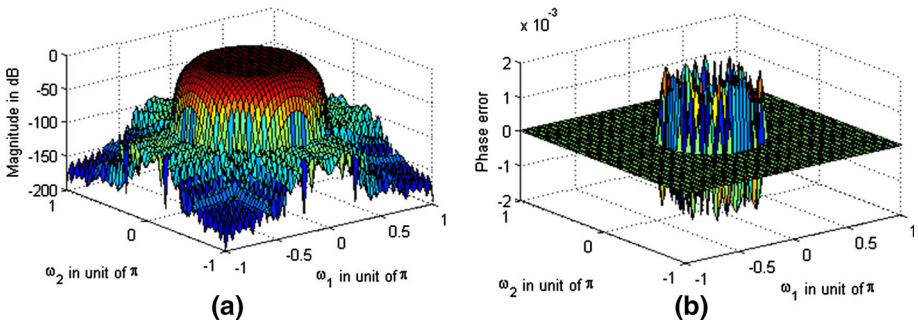

Fig. 1 **a** Magnitude response and **b** passband phase error of the 31×31 circular filter obtained by the matrix-based EPCLS-GI with $\rho = 0.01, \gamma = 0.002, \tau_1 = \tau_2 = 13.5$ and $\sigma = 3$

control the width of transition band. The desired group delays are τ_1 and τ_2 and the desired magnitude response is taken as

$$\bar{D}(\omega_1, \omega_2) = \begin{cases} 1, & \omega \in \Omega_p, \\ \frac{\phi - (|\omega_1| - |\omega_2| \tan \chi)}{\phi}, & \omega \in \Omega \backslash (\Omega_p \cup \Omega_s), \\ 0, & \omega \in \Omega_s. \end{cases} \quad (29)$$

Several such fan filters of size 25×25 with the same γ, ρ, τ_1 and τ_2 have been designed using the EPCLS and CPCLS criterions. Table 3 gives the results for the CPCLS design and EPCLS designs with various σs, where the CPCLS design is actually the EPCLS design under $\sigma = 1$. Results show that the MSCEs of the resultant EPCLS filters, decreasing with

Table 2 Efficiency comparison of the matrix-based EPCLS-GI and the vectorized CPCLS-GI algorithms for filters with $N = 21$, $\tau_1 = 8$ and $\tau_2 = 9$ in Example 1

Specifications of the filter	Matrix-based EPCLS-GI ($\sigma = 5$)		Vectorized CPCLS-GI ($\sigma = 5$)	
($N = 21$, $\tau_1 = 8$ and $\tau_2 = 9$)	MSCE (10^{-4})	Time(s)/iterates	MSCE (10^{-4})	Time(s)/iterates
$\rho = 0.017$, $\gamma = 0.006$	2.2954	2.891/633	2.2954	29.49/633
$\rho = 0.015$, $\gamma = 0.006$	2.8031	4.125/902	2.8031	38.80/902
$\rho = 0.015$, $\gamma = 0.005$	2.8418	4.312/933	2.8418	40.42/933
$\rho = 0.013$, $\gamma = 0.005$	3.7772	6.425/1,413	3.7772	58.96/1,413

Table 3 Comparison of the filter's MSCE obtained by the CPCLS design and EPCLS designs with various σs in Example 2

Filter specifications	Design criterion	MSCE (10^{-4})	Time(s)/iterates of matrix-based EPCLS
	CPCLS (EPCLS with $\sigma = 1$)	5.9071	8.141/929
	EPCLS with $\sigma = 2$	5.7663	6.235/719
	EPCLS with $\sigma = 5$	5.6964	5.703/641
$N = 25$, $\rho = 0.051$, $\gamma = 0.03$,	EPCLS with $\sigma = 10$	5.6841	5.531/592
$\tau_1 = \tau_2 = 11$	EPCLS with $\sigma = 20$	5.6810	5.359/579
	EPCLS with $\sigma = 50$	5.6801	4.906/560
	EPCLS with $\sigma = 100$	5.6800	3.406/560
	EPCLS with $\sigma = 200$	5.6800	3.390/560

σ, are all smaller than that of the CPCLS filter. That is to say, the EPCLS designs obtain better filters than the CPCLS design. It is interesting that the iteration number and CPU time of the matrix-based EPCLS algorithm also decrease with σ. This is because the domain of the elliptic-error and phase-error constraints becomes larger and thus the constraints become easier to meet when σ gets larger. Figure 2 draws the magnitude response and passband phase error of the 25×25 fan filter with $\rho = 0.051$, $\gamma = 0.03$, $\tau_1 = \tau_2 = 11$ and $\sigma = 5$.

We also design several such fan filters of size 25×25 with minimal MME using a combination of the sequential CLS (SCLS) method (see, e.g., Hong et al. 2013)) and the matrix-based EPCLS-GI algorithm with various σs. Table 4 shows the design results. Under the same MPE, the minimal MME decreases with σ, with a sacrifice of possible increase in MSCE. This implies that an EPCLS design may obtain filters with smaller MME than a CPCLS design.

5 Conclusions

A matrix-based GI algorithm has been developed for the elliptic-error and phase-error constrained least-squares design of 2-D FIR filters with arbitrarily specified frequency responses. The developed matrix-based EPCLS-GI algorithm is a matrix-based extension of the vectorized CPCLS-GI algorithm in Lai (2009) but has a lower computational complexity than it. Design examples and comparisons demonstrate that the matrix-based EPCLS algorithm consumes much less CPU time than the vectorized CPCLS-GI algorithm and the SDP method,
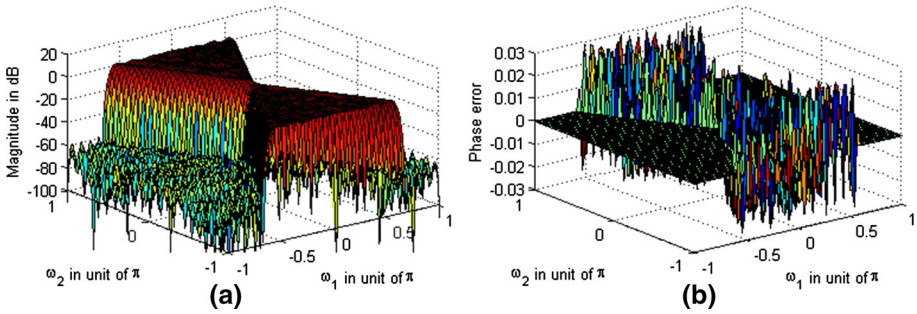
Fig. 2 **a** Magnitude response and **b** passband phase error of the 25×25 fan filter obtained by the matrix-based EPCLS-GI with $\rho = 0.051$, $\gamma = 0.03$, $\tau_1 = \tau_2 = 11$ and $\sigma = 5$

Table 4 Comparison of the filter's minimal MME that can be obtained by the CPCLS design and EPCLS designs with various σs in Example 2

Filter specifications	Design criterion	Minimal MME	MPE	MSCE (10^{-4})
	CPCLS (EPCLS with $\sigma = 1$)	0.05001527	0.04	10.69142
	EPCLS with $\sigma = 2$	0.04966005	0.04	10.75726
	EPCLS with $\sigma = 4$	0.04938960	0.04	10.96028
$N = 25$, $\gamma = 0.04$, $\tau_1 = \tau_2 = 11$	EPCLS with $\sigma = 8$	0.04926968	0.04	11.64738
	EPCLS with $\sigma = 16$	0.04923793	0.04	11.98660
	EPCLS with $\sigma = 32$	0.04923181	0.04	11.78357

and the EPCLS design may obtain better filters in terms of smaller MSCE and/or MME than the CPCLS design.

Appendix

Proof of Theorem 1 Using the Lagrangian method, we can obtain the following equations for the multiplier vector $\boldsymbol{\lambda}(\ell)$ and the solution matrix $\mathbf{A}(\ell)$ of problem (15a) subject to the equality constraints indexed by the active set $J(\ell)$,

$$\mathbf{S}\boldsymbol{\lambda}(\ell) = \mathbf{b}, \ \mathbf{A}(\ell) = \mathbf{A}_0 - \sum_{m=1}^{t} \lambda_m(\ell) R_m, \tag{30}$$

where $\lambda_m(\ell)$ is the mth multiplier in $\boldsymbol{\lambda}(\ell)$, $\mathbf{S}$ and $\mathbf{R}_n$ are defined in (22a) and (22c), and $\mathbf{b} = [b_m]_{t \times 1}$ with

$$b_m = \Gamma_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^R(\mathbf{A}_0) + \bar{\Gamma}_{j_m k_m l_m}(\alpha_m) g_{j_m k_m}^I(\mathbf{A}_0)$$
$$- \theta_{j_m k_m l_m}(\alpha_m), m = 1, 2, \ldots, t. \tag{31}$$

If the $(t + 1)$th equality constraint is added into the active set, the matrix $\mathbf{S}$ and vector $\mathbf{b}$ become respectively into

$$\mathbf{S}^+ = \begin{bmatrix} \mathbf{S} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^H & \xi_{t+1} \end{bmatrix}, \mathbf{b}^+ = [\mathbf{b}^H, b_{t+1}]^H, \tag{32}$$

where ξ is defined in (22b), b_{t+1} and ξ_{t+1} are defined by (31) and (22b) for $m = t + 1$. Then, by (30) and from (18), (19) and (32), we have

$$\mathbf{S}^+\boldsymbol{\lambda}^+ = \begin{bmatrix} \mathbf{S} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^H & \xi_{t+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}(\ell) - w\mathbf{r} \\ w \end{bmatrix} = \mathbf{b}^+ = \begin{bmatrix} \mathbf{b} \\ b_{t+1} \end{bmatrix},$$

$$\mathbf{A}^+ = \mathbf{A}_0 - \sum_{m=1}^{t+1} \lambda_m^+ \mathbf{R}_m = A_0 - \sum_{m=1}^{t} \lambda_m(\ell)\mathbf{R}_m + w\mathbf{Z}, \tag{33}$$

where λ_m^+ is the mth element of $\boldsymbol{\lambda}^+$. Equations (20) and (21) immediately follow from (18), (19) and (33). □

References

Ahmad, M. O., & Wang, J. D. (1989). An analytical least square solution to the design problem of two-dimensional FIR filters with quadrantally symmetric and antisymmetric frequency response. *IEEE Transactions on Circuits and Systems*, *36*(7), 968–979.

Algazi, V. R., SUK, M., & Rim, C.-S. (1986). Design of almost minimax FIR filters in one and two dimensions by WLS techniques. *IEEE Transactions on Circuits and Systems*, *CAS–33*(6), 590–596.

Aravena, J. L., & Gu, G. (1996). Weighted least mean square design of 2-D FIR digital filters: The general case. *IEEE Transactions on Signal Processing*, *44*(10), 2568–2578.

Chottera, A. T., & Jullien, G. A. (1982). Design of two-dimensional recursive digital filters using linear programming. *IEEE Transactions on Circuits and Systems*, *CAS–29*(12), 817–826.

Dam, H. H., Nordholm, S., & Cantoni, A. (2005). Uniform FIR filterbank optimization with group delay specifications. *IEEE Transactions on Signal Processing*, *53*(11), 4249–4260.

de Haan, J. M., Grbic, N., Claesson, I., & Nordholm, S. (2003). Filterbank design for subband adaptive microphone arrays. *IEEE Transactions on Speech Audio Process*, *11*(1), 14–23.

Deng, T. B., & Lu, W.-S. (2000). Weighted least-squares method for designing variable fractional delay 2-D FIR digital filters. *IEEE Transactions on Circuits and Systems - II*, *47*(2), 114–124.

Dumitrescu, B. (2005). Optimization of two-dimensional IIR filters with nonseparable and separable denominator. *IEEE Transactions on Signal Processing*, *53*(5), 1768–1777.

Dumitrescu, B. (2006). Trigonometric polynomials positive on frequency domains and applications to 2-D FIR filter design. *IEEE Transactions on Signal Processing*, *54*(11), 4282–4292.

Gislason, E., Johansen, M., Ersboll, B. K., & Jacobsen, S. K. (1993). Three different criteria for the design of two-dimensional zero phase FIR digital filters. *IEEE Transactions on Signal Processing*, *41*(10), 3070–3074.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, *27*, 1–33.

Grant, M., & Boyd, S. (2013). *CVX: Matlab software for disciplined convex programming, version 2.0 beta*. http://cvxr.com/cvx

Gu, G., & Aravena, J. L. (1994). Weighted least mean square design of 2-D FIR digital filters. *IEEE Transactions on Signal Processing*, *42*(11), 3178–3187.

Hong, X. Y., Lai, X. P., & Zhao, R. J. (2013). Matrix-based algorithms for constrained least-squares and minimax designs of 2-D linear-phase FIR filters. *IEEE Transactions on Signal Processing*, *61*(14), 3620–3631.

Laakso, T. I., Valimaki, V., Karjalainen, M., & Laine, U. K. (1996). Splitting the unit delay: Tools for fractional delay filter design. *IEEE Signal Processing Magazine*, *13*, 30–60.

Lai, X. P. (2009). Optimal design of nonlinear-phase FIR filters with prescribed phase error. *IEEE Transactions on Signal Processing*, *57*, 3399–3410.

Lai, X. P., Wang, J. Z., & Xu, Z. (2009). A new minimax design of two-dimensional FIR filters with reduced group delay. In *Proceedings on 7th Asian control conference* (pp. 610–614).

Lang, M. C., Selesnick, I. W., & Burrus, C. S. (1996). Constrained least squares design of 2-D FIR filters. *IEEE Transactions on Signal Processing*, *44*(5), 1234–1241.

Lu, W. S. (2002). A unified approach for the design of 2-D digital filters via semidefinite programming. *IEEE Transactions on Circuits and Systems - I*, *49*(6), 814–826.

Lu, W. S., & Hinamoto, T. (2006). A second-order cone programming approach for minimax design of 2-D FIR filters with low group delay. In *Proceedings IEEE international symposium on circuits and systems* (pp. 2521–2524).

Lu, W. S., & Hinamoto, T. (2011). Two-dimensional digital filters with sparse coefficients. *Multidimensional Systems and Signal Processing*, *22*(1–3), 173–189.

Lu, W. S., Pei, S. C., & Tseng, C. C. (1998). A weighted least-squares method for the design of stable 1-D and 2-D IIR digital filters. *IEEE Transactions on Signal Processing*, *46*(1), 1–10.

Rusu, C., & Dumitrescu, B. (2012). Iterative reweighted l1 design of sparse FIR filters. *Signal Processing*, *92*(4), 905–911.

Shyu, J. J., Pei, S. C., & Huang, Y. D. (2009). Two-dimensional Farrow structure and the design of variable fractional delay 2-D FIR digital filters. *IEEE Transactions on Circuits and Systems - I*, *56*(2), 395–404.

Tütüncü, R. H., Toh, K. C., & Todd, M. J. (2003). Solving semidefinite–quadratic–linear programs using SDPT3. *Mathematical Programming, Series B*, *95*, 189–217.

Yan, H., & Ikehara, M. (1999). Two-dimensional perfect reconstruction FIR filter banks with triangular supports. *IEEE Transactions on Signal Processing*, *47*(4), 1003–1009.

Zhao, R. J., & Lai, X. P. (2011). A fast matrix iterative technique for the WLS design of 2-D quadrantally symmetric FIR filters. *Multidimensional Systems and Signal Processing*, *22*(4), 303–317.

Zhao, R. J., & Lai, X. P. (2013). Efficient 2-D based algorithms for WLS design of 2-D FIR filters with arbitrary weighting functions. *Multi-dimensional Systems and Signal Processing*, *24*(3), 417–434.

Zhu, W. P., Ahmad, M. O., & Swamy, M. N. S. (1997). A closed-form solution to the least-square design problem of 2-D linear-phase FIR filters. *IEEE Transactions on Circuits Systems - II*, *44*(12), 1032–1039.

Zhu, W. P., Ahmad, M. O., & Swamy, M. N. S. (1999). A least-squares design approach for 2-D FIR filters with arbitrary frequency response. *IEEE Transactions Circuits and Systems - II*, *46*(8), 1027–1034.

Xiaoying Hong received the B.S. and M.S. degree from Inner Mongolia University of Science & Technology, Baotou, China, in 2001 and 2004, respectively. She has been with Shandong University, Weihai, China, since 2004, where she became a Lecturer in 2006. She is currently a Ph.D. candidate in signal and information processing of Shandong University. Her main research interests include digital filter design, optimization algorithm, and fuzzy control system.

Xiaoping Lai received the B.S. and M.S. degrees in physics from Shandong University in June 1985 and June 1988, and the Ph.D. degree in applied mathematics from Shandong University in June 2000. From 1988 to 2008, he was with Shandong University, China, where he became a professor in 2001, and was the Director of the Department of Information Science and Control Engineering from 2003 to 2004 and the Dean of the School of Information Engineering from 2004 to 2007. Since 2008, he has been with Hangzhou Dianzi University, China, as a distinguished professor of Control Science and Engineering, where he became the Dean of the Automation School in 2010. Dr. Lai has served as an Associate Editor of Multidimensional Systems and Signal Processing since June 2014. His general research interests lie in the areas of optimization methods and applications in digital signal processing and control systems. His current research focuses on optimal designs of one- and multidimensional digital filters and filter banks, as well as compressive sensing.

Ruijie Zhao received the B.S., M.S., and Ph.D. degrees from Shandong University, China, in 2001, 2004, and 2012, respectively. He has been with Shandong University, Weihai, China, since 2004, where he became an Associate Professor in 2013. His main research interests include digital filter design, optimization algorithm, and approximation of functions.