Minimization of Frequency-Weighted l_2-Sensitivity Subject to l_2-Scaling Constraints for Two-Dimensional State-Space Digital Filters

Takao Hinamoto, Toru Oumi, Osemekhian I. Omoifo and Wu-Sheng Lu

***Abstract*— This paper investigates the problem of frequency-weighted l_2-sensitivity minimization subject to l_2-scaling constraints for two-dimensional (2-D) state-space digital filters described by the Roesser model. It is shown that the Fornasini-Marchesini second model can be imbedded in the Roesser model. Two iterative methods are developed to solve the constrained optimization problem encountered. The first iterative method introduces a Lagrange function and optimizes it using some matrix-theoretic techniques and an efficient bisection method. The second iterative method converts the problem into an unconstrained optimization formulation by using linear-algebraic techniques and solves it by applying an efficient quasi-Newton algorithm. The optimal filter structure with minimum frequency-weighted l_2-sensitivity and no overflow is then synthesized by an appropriate coordinate transformation. Case studies are presented to demonstrate the validity and effectiveness of the proposed techniques.**

***Index Terms*— 2-D digital filters, Roesser's model, Fornasini-Marchesini's second model, frequency-weighted l_2-sensitivity minimization, l_2-scaling constraints, no overflow, Lagrange function, bisection method, quasi-Newton method**

I. INTRODUCTION

It is well known that there exist infinitely many minimal state-space realizations for a given transfer function, and some inherent properties such as controllability, observability, stability, etc. are invariant within these realizations. However, performance measures such as coefficient sensitivity, output roundoff noise, overflow oscillations, etc. may be significantly varying among the realizations. Consider a transfer function with coefficients of infinite accuracy, which meets certain design specifications including stability. When the transfer function is implemented by a state-space model with a finite binary representation, truncation or rounding of the state-space model is required to satisfy the finite word length (FWL) constraints. As a result, the characteristics of the stable filter might be so altered that the filter may become unstable. This motivates the study of the coefficient sensitivity minimization problem. To date, several techniques have been reported for synthesizing the state-space descriptions with minimum coefficient sensitivity. The techniques can be divided into two main classes: those for l_1/l_2-mixed sensitivity minimization [1]-[5] and those for l_2-sensitivity minimization [6]-[11]. In [6]-[10], it has been argued that the sensitivity measure based on a sole l_2-norm is more natural and reasonable relative to the l_1/l_2-mixed sensitivity minimization. For 2-D state-space digital filters, the l_1/l_2-mixed sensitivity minimization problem [12]-[17] and l_2-sensitivity minimization problem [10],[17]-[20] have also been investigated. It has been realized that solutions for *frequency-weighted* sensitivity minimization would be of practical use as these solutions allow to emphasize or de-emphasize the filter's sensitivity in certain frequency regions of interest. Synthesis procedures of the optimal FWL 2-D filter structures that minimize the frequency-weighted sensitivity measure have been considered [15]-[18]. However, the minimization methods proposed in the above work do not impose constraints on the scaling of the design variables. As a result, elimination of overflow cannot be ensured. More recently, the minimization problem of l_2-sensitivity subject to l_2-scaling constraints has been explored for 1-D and a class of 2-D state-space digital filters [21]-[23]. It is well known that the use of scaling constraints can be beneficial for suppressing overflow [24],[25]. However, frequency-weighted sensitivity measure has not yet been considered in [21]-[23].

In this paper, we investigate the problem of minimizing a frequency-weighted l_2-sensitivity measure subject to l_2-scaling constraints for 2-D state-space digital filters described by the Roesser local state-space (LSS) model [26]. We then proceed by introducing an expression for evaluating the frequency-weighted l_2-sensitivity, and formulating the minimization problem for the frequency-weighted l_2-sensitivity measure subject to l_2-scaling constraints. Next, two iterative methods are developed for solving the constrained optimization problem. The first iterative method introduces a Lagrange function, and makes use of some matrix-theoretic techniques and an efficient bisection method. The second iterative method relies on a technique that converts the constrained optimization problem into an unconstrained optimization formulation and utilizes an efficient quasi-Newton method with closed-form formula for gradient evaluation. Finally, case studies are presented to demonstrate the validity and effectiveness of the proposed techniques.

One of the contributions made in this paper is to show that **either the Fornasini-Marchesini (FM) second LSS model [27] or its transposed-structure model [22],[28] can be imbedded in the Roesser model as a special case.** This justifies the use of the Roesser model in our studies. Another

Manuscript received mm dd, 2007; revised mm dd, 2007.

T. Hinamoto, T. Oumi and O. I. Omoifo are with the Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima 739-8527, Japan. (e-mail: {hinamoto,oumi,osei}@hiroshima-u.ac.jp, Phone:+81-82-424-7672, Fax:+81-82-422-7195)

W.-S. Lu is with the Depertment of Electrical and Computer Engineering, University of Victoria, Victoria, B.C, Canada, V8W 3P6. (e-mail: wslu@ece.uvic.ca, Phone:+1-250-721-8692, Fax:+1-250-721-6052)

contribution is that a bisection method is applied to obtain the Lagrange multipliers, which makes it possible to attain considerably faster convergence than the algorithm reported in [22]. Moreover, unlike [21] and [22], the present paper investigates a *frequency-weighted* l_2-sensitivity measure under l_2-scaling constraints. Although this extension is technically manageable, to the best of our knowledge, this is the first time a *frequency-weighted* l_2-sensitivity measure under l_2-scaling constraints is addressed in the l_2/l_2 framework for state-space digital filters.

Throughout the paper, $\boldsymbol{I}_n$ stands for the identity matrix of dimension $n \times n$, $\oplus$ is used to denote the direct sum of matrices, the transpose (conjugate transpose) of a matrix $\boldsymbol{A}$ is indicated by $\boldsymbol{A}^T$ ($\boldsymbol{A}^*$), and the trace and ith diagonal element of a square matrix $\boldsymbol{A}$ are denoted by tr[$\boldsymbol{A}$] and $(\boldsymbol{A})_{ii}$, respectively.

II. PROBLEM FORMULATION

2.1 System Models

Consider a stable, separately locally controllable and separately locally observable LSS model for 2-D recursive digital filters

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}^h(i+1,j)\\ \boldsymbol{x}^v(i,j+1)\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j)\\ \boldsymbol{x}^v(i,j)\end{bmatrix} + \begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{bmatrix} u(i,j)\\ y(i,j) &= \begin{bmatrix}\boldsymbol{c}_1 & \boldsymbol{c}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j)\\ \boldsymbol{x}^v(i,j)\end{bmatrix} + d\,u(i,j)\end{aligned} \tag{1}$$

which was originally proposed by Roesser [26],[29], where $\boldsymbol{x}^h(i,j)$ is an $m \times 1$ horizontal state vector, $\boldsymbol{x}^v(i,j)$ is an $n \times 1$ vertical state vector, $u(i,j)$ is a scalar input, $y(i,j)$ is a scalar output, and $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_3$, $\boldsymbol{A}_4$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, and d are $m \times m$, $m \times n$, $n \times m$, $n \times n$, $m \times 1$, $n \times 1$, $1 \times m$, $1 \times n$, and 1×1 real constant matrices, respectively. A block diagram of the LSS model in (39) is shown in Fig. 1. The

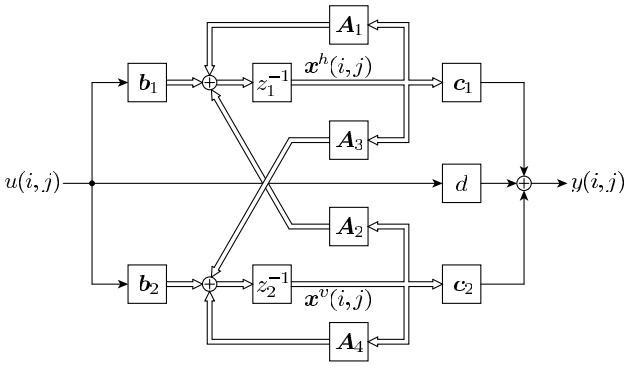

Fig. 1. The block diagram of the Roesser LSS model.

transfer function of the LSS model in (39) is given by

$$H(z_1, z_2) = \boldsymbol{c}(\boldsymbol{Z} - \boldsymbol{A})^{-1}\boldsymbol{b} + d \tag{2}$$

where $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are $(m+n) \times (m+n)$, $(m+n) \times 1$, and $1 \times (m+n)$ real constant matrices defined by

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix}\boldsymbol{c}_1 & \boldsymbol{c}_2\end{bmatrix},$$

respectively, and $\boldsymbol{Z} = z_1\boldsymbol{I}_m \oplus z_2\boldsymbol{I}_n$. For the sake of simplicity, the LSS model in (39) is denoted hereafter by $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{m,n}$.

Alternatively, an LSS model for a class of 2-D recursive digital filters can be described by [22],[28]

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}(i+1,j+1)\\ y(i,j)\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_1' & \boldsymbol{A}_2'\\ \boldsymbol{c}_1' & \boldsymbol{c}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{x}(i,j+1)\\ \boldsymbol{x}(i+1,j)\end{bmatrix}\\ &\quad + \begin{bmatrix}\boldsymbol{b}'\\ d\end{bmatrix} u(i,j)\end{aligned} \tag{3}$$

where $\boldsymbol{x}(i,j)$ is an $N \times 1$ local state vector, $u(i,j)$ is a scalar input, $y(i,j)$ is a scalar output, and $\boldsymbol{A}_1', \boldsymbol{A}_2', \boldsymbol{b}', \boldsymbol{c}_1', \boldsymbol{c}_2'$, and d are $N \times N$, $N \times N$, $N \times 1$, $1 \times N$, $1 \times N$, and 1×1 real constant matrices, respectively. The transfer function of the LSS model in (3) is given by

$$\begin{aligned}D(z_1, z_2) &= (z_1^{-1}\boldsymbol{c}_1' + z_2^{-1}\boldsymbol{c}_2')\\ &\quad \cdot \left(\boldsymbol{I}_n - z_1^{-1}\boldsymbol{A}_1' - z_2^{-1}\boldsymbol{A}_2'\right)^{-1}\boldsymbol{b}' + d.\end{aligned} \tag{4}$$

If we define

$$\boldsymbol{x}(i,j+1) = \boldsymbol{x}^h(i,j), \quad \boldsymbol{x}(i+1,j) = \boldsymbol{x}^v(i,j) \tag{5}$$

then the LSS model in (3) can then be imbedded in that of (39) as a special case as follows:

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}^h(i+1,j)\\ \boldsymbol{x}^v(i,j+1)\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_1' & \boldsymbol{A}_2'\\ \boldsymbol{A}_1' & \boldsymbol{A}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j)\\ \boldsymbol{x}^v(i,j)\end{bmatrix} + \begin{bmatrix}\boldsymbol{b}'\\ \boldsymbol{b}'\end{bmatrix} u(i,j)\\ y(i,j) &= \begin{bmatrix}\boldsymbol{c}_1' & \boldsymbol{c}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j)\\ \boldsymbol{x}^v(i,j)\end{bmatrix} + d\,u(i,j)\end{aligned} \tag{6}$$

where $m = n = N$. It is noted that $D(z_1, z_2)^T$ can be viewed as a transfer function of the FM second LSS model [27], which reveals that the LSS model of $D(z_1, z_2)^T$ can be realized by a transposed structure of that in (6). Therefore, the LSS model in (39) is more general than either the LSS model in (3) or the FM second LSS model [27] **(and vice versa with the same dimension, but increased number of coefficients [27])** . In addition, we note that the technique reported in [22] has merely treated the l_2-sensitivity minimization problem for the LSS model in (3) subject to l_2-scaling constraints. Recall that the total numbers of the coefficients in (39) and (3) are $(m+n)^2 + 2(m+n) + 1$ and $2N^2 + 3N + 1$, respectively. This means that the LSS model in (39) has less number of the coefficients than that of (3) when their local state vectors possess the same dimension, i.e., $m + n = N$. Under these circumstances, it is worthwhile to consider the more general problem of minimizing the *frequency-weighted* l_2-sensitivity for the LSS model in (39) subject to l_2-scaling constraints, because its solutions will make it possible to emphasize or de-emphasize the filter's sensitivity in certain frequency regions of interest, and because the LSS model in (39) owns less number of the coefficients in case $m + n = N$.

2.2 A Frequency-Weighted l_2-Sensitivity Measure

Suppose that the LSS model in (39) is implemented by FWL fixed-point arithmetic with a B-bit fractional representation, and is realized with coefficient matrices

$$\begin{aligned} \tilde{\boldsymbol{A}} &= \boldsymbol{A} + \Delta\boldsymbol{A}, \qquad \tilde{\boldsymbol{b}} = \boldsymbol{b} + \Delta\boldsymbol{b} \\ \tilde{\boldsymbol{c}} &= \boldsymbol{c} + \Delta\boldsymbol{c}, \qquad \tilde{d} = d + \Delta d \end{aligned} \tag{7}$$

where $\Delta\boldsymbol{A}$, $\Delta\boldsymbol{b}$, $\Delta\boldsymbol{c}$, and Δd stand for the quantization errors of the coefficient matrices. The transfer function of the FWL realization is then expressed as

$$\tilde{H}(z_1, z_2) = \tilde{\boldsymbol{c}}(\boldsymbol{Z} - \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{b}} + \tilde{d}. \tag{8}$$

Let $\{p_i, i = 1, 2, \cdots, M\}$ be the set of the ideal parameters of a realization and let $\{\tilde{p}_i, i = 1, 2, \cdots, M\}$ be its FWL version where $\tilde{p}_i = p_i + \Delta p_i$ with Δp_i indicating the corresponding parameter perturbation. If all Δp_i are sufficiently small in magnitude, then the first-order approximation of the Taylor series expansion yields

$$\begin{aligned} \Delta H(z_1, z_2) &= \tilde{H}(z_1, z_2) - H(z_1, z_2) \\ &\simeq \sum_{i=1}^{M} \frac{\partial H(z_1, z_2)}{\partial p_i} \Delta p_i. \end{aligned} \tag{9}$$

Obviously, smaller $\partial H(z_1, z_2)/\partial p_i$ for $i = 1, 2, \cdots, M$ yield smaller transfer function error $\Delta H(z_1, z_2)$. For a fixed-point implementation of B bits, the parameter perturbations can be considered to be independent random-variables uniformly distributed within the range $[-2^{-B-1},\ 2^{-B-1}]$. Under these circumstances, a measure of the transfer function error can statistically be defined as

$$\sigma^2_{\Delta H} = \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} E[|\Delta H(z_1, z_2)|^2] \frac{dz_1 dz_2}{z_1 z_2} \tag{10}$$

where $E(\cdot)$ denotes the ensemble-average operation. Since Δp_i's are uniformly-distributed independent random variables, it follows that

$$E[|\Delta H(z_1, z_2)|^2] = \sum_{i=1}^{M} \left| \frac{\partial H(z_1, z_2)}{\partial p_i} \right|^2 \sigma^2 \tag{11}$$

where

$$\sigma^2 = E[(\Delta p_i)^2] = \frac{1}{12} 2^{-2B}.$$

Thus, if the measure

$$S_o = \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \sum_{i=1}^{M} \left| \frac{\partial H(z_1, z_2)}{\partial p_i} \right|^2 \frac{dz_1 dz_2}{z_1 z_2} \tag{12}$$

is minimized then the minimum variance $\sigma^2_{\Delta H}$ can be attained because of the relation $\sigma^2_{\Delta H} = S_o \sigma^2$. The measure in (12) is referred to as an l_2-sensitivity measure.

The frequency-weighted l_2-sensitivity of the LSS model in (39) is defined as follows.

Definition 1: Let $\boldsymbol{X}$ be an $m \times n$ real matrix and let $f(\boldsymbol{X})$ be a scalar complex function of $\boldsymbol{X}$, that is differentiable with respect to all the entries of $\boldsymbol{X}$. The sensitivity function of $f(\boldsymbol{X})$ with respect to $\boldsymbol{X}$ is then defined as [5]

$$\boldsymbol{S}_{\boldsymbol{X}} = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}, \qquad (\boldsymbol{S}_{\boldsymbol{X}})_{ij} = \frac{\partial f(\boldsymbol{X})}{\partial x_{ij}} \tag{13}$$

where x_{ij} denotes the (i, j)th entry of matrix $\boldsymbol{X}$.

Definition 2: In order to take into account the sensitivity of the transfer function in a specified frequency band, or even at some discrete frequency points, the weighted sensitivity functions are defined as [5],[17]

$$\begin{aligned} \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{A}} &= W_A(z_1, z_2) \frac{\partial H(z_1, z_2)}{\partial \boldsymbol{A}} \\ \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{b}} &= W_B(z_1, z_2) \frac{\partial H(z_1, z_2)}{\partial \boldsymbol{b}} \\ \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{c}^T} &= W_C(z_1, z_2) \frac{\partial H(z_1, z_2)}{\partial \boldsymbol{c}^T} \end{aligned} \tag{14}$$

where $W_A(z_1, z_2)$, $W_B(z_1, z_2)$, and $W_C(z_1, z_2)$ are scalar, stable, causal functions of the complex variables z_1 and z_2.

Notice that δ in (14) is not meant to be a derivative operator, but rather a notation for defining the weighted parameter sensitivity.

Definition 3: Let $\boldsymbol{X}(z_1, z_2)$ be an $m \times n$ complex matrix valued function of the complex variables z_1 and z_2. The l_2 norm of $\boldsymbol{X}(z_1, z_2)$ is then defined by

$$\begin{aligned} &||\boldsymbol{X}(z_1, z_2)||_2 \\ &= \left(\text{tr} \left[\frac{1}{(2\pi j)^2} \oint_{\Gamma_1} \oint_{\Gamma_2} \boldsymbol{X}(z_1, z_2) \boldsymbol{X}^*(z_1, z_2) \frac{dz_1 dz_2}{z_1 z_2} \right] \right)^{\frac{1}{2}} \end{aligned} \tag{15}$$

where $j = \sqrt{-1}$ and $\Gamma_i = \{z_i : |z_i| = 1\}$ for $i = 1, 2$.

From (2) and Definitions 1-3, the overall frequency-weighted l_2-sensitivity measure for the LSS model in (39) can be defined as

$$\begin{aligned} S &= \left\| \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{A}} \right\|_2^2 + \left\| \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{b}} \right\|_2^2 + \left\| \frac{\delta H(z_1, z_2)}{\delta \boldsymbol{c}^T} \right\|_2^2 \\ &= \left\| W_A(z_1, z_2) [\boldsymbol{F}(z_1, z_2) \boldsymbol{G}(z_1, z_2)]^T \right\|_2^2 \\ &\quad + \left\| W_B(z_1, z_2) \boldsymbol{G}^T(z_1, z_2) \right\|_2^2 + \left\| W_C(z_1, z_2) \boldsymbol{F}(z_1, z_2) \right\|_2^2 \end{aligned} \tag{16}$$

where

$$\boldsymbol{F}(z_1, z_2) = (\boldsymbol{Z} - \boldsymbol{A})^{-1}\boldsymbol{b}, \quad \boldsymbol{G}(z_1, z_2) = \boldsymbol{c}\,(\boldsymbol{Z} - \boldsymbol{A})^{-1}.$$

It follows that the frequency-weighted l_2-sensitivity measure in (16) can be written as

$$S = \text{tr}[\boldsymbol{M}_A] + \text{tr}[\boldsymbol{W}_B] + \text{tr}[\boldsymbol{K}_C] \tag{17}$$

where $\boldsymbol{M}_A$, $\boldsymbol{W}_B$, and $\boldsymbol{K}_C$ are obtained by the following general expression:

$$\boldsymbol{X} = \frac{1}{(2\pi j)^2} \oint_{\Gamma_1} \oint_{\Gamma_2} \boldsymbol{Y}(z_1, z_2) \boldsymbol{Y}^*(z_1, z_2) \frac{dz_1 dz_2}{z_1 z_2}$$

with $\boldsymbol{Y}(z_1, z_2) = W_A(z_1, z_2)[\boldsymbol{F}(z_1, z_2)\boldsymbol{G}(z_1, z_2)]^T$ for $\boldsymbol{X} = \boldsymbol{M}_A$, $\boldsymbol{Y}(z_1, z_2) = W_B(z_1, z_2)\boldsymbol{G}^T(z_1, z_2)$ for $\boldsymbol{X} = \boldsymbol{W}_B$, and $\boldsymbol{Y}(z_1, z_2) = W_C(z_1, z_2)\boldsymbol{F}(z_1, z_2)$ for $\boldsymbol{X} = \boldsymbol{K}_C$.

2.3 Problem Formulation

Define a state-space coordinate transformation by [26],[29]

$$\begin{bmatrix} \overline{\boldsymbol{x}}^h(i, j) \\ \overline{\boldsymbol{x}}^v(i, j) \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_4^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^h(i, j) \\ \boldsymbol{x}^v(i, j) \end{bmatrix} \tag{18}$$

where $\boldsymbol{T}_1$ and $\boldsymbol{T}_4$ are $m\times m$ and $n\times n$ nonsingular matrices, respectively. New realizations can then be characterized as $(\overline{\boldsymbol{A}},\overline{\boldsymbol{b}},\overline{\boldsymbol{c}},d)_{m,n}$ with

$$\overline{\boldsymbol{A}}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T},\quad \overline{\boldsymbol{b}}=\boldsymbol{T}^{-1}\boldsymbol{b},\quad \overline{\boldsymbol{c}}=\boldsymbol{c}\boldsymbol{T} \tag{19}$$

where $\boldsymbol{T}=\boldsymbol{T}_1\oplus\boldsymbol{T}_4$. For a new realization, the frequency-weighted l_2-sensitivity measure in (17) is changed to

$$\begin{aligned}S(\boldsymbol{P}) &= \text{tr}[\boldsymbol{M}_A(\boldsymbol{P})\boldsymbol{P}]+\text{tr}[\boldsymbol{W}_B\boldsymbol{P}]+\text{tr}[\boldsymbol{K}_C\boldsymbol{P}^{-1}]\\ &= \text{tr}[\boldsymbol{N}_A(\boldsymbol{P})\boldsymbol{P}^{-1}]+\text{tr}[\boldsymbol{W}_B\boldsymbol{P}]+\text{tr}[\boldsymbol{K}_C\boldsymbol{P}^{-1}]\end{aligned} \tag{20}$$

with $\boldsymbol{P}=\boldsymbol{T}\boldsymbol{T}^T=\boldsymbol{P}_1\oplus\boldsymbol{P}_4$ and

$$\boldsymbol{M}_A(\boldsymbol{P})=\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{Y}(z_1,z_2)\boldsymbol{P}^{-1}\boldsymbol{Y}^*(z_1,z_2)\frac{dz_1dz_2}{z_1z_2}$$

$$\boldsymbol{N}_A(\boldsymbol{P})=\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{Y}^*(z_1,z_2)\boldsymbol{P}\boldsymbol{Y}(z_1,z_2)\frac{dz_1dz_2}{z_1z_2}$$

where $\boldsymbol{Y}(z_1,z_2)=W_A(z_1,z_2)[\boldsymbol{F}(z_1,z_2)\boldsymbol{G}(z_1,z_2)]^T$.

If l_2-scaling constraints are imposed on the horizontal and vertical state vectors $\overline{\boldsymbol{x}}^h(i,j)$ and $\overline{\boldsymbol{x}}^v(i,j)$, we require that [30]

$$\begin{aligned}(\overline{\boldsymbol{K}}_1)_{\xi\xi} &= (\boldsymbol{T}_1^{-1}\boldsymbol{K}_1\boldsymbol{T}_1^{-T})_{\xi\xi}=1 \quad \text{for} \quad \xi=1,2,\cdots,m\\ (\overline{\boldsymbol{K}}_4)_{\zeta\zeta} &= (\boldsymbol{T}_4^{-1}\boldsymbol{K}_4\boldsymbol{T}_4^{-T})_{\zeta\zeta}=1 \quad \text{for} \quad \zeta=1,2,\cdots,n\end{aligned} \tag{21}$$

where

$$\begin{aligned}\boldsymbol{K} &= \frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{F}(z_1,z_2)\boldsymbol{F}^*(z_1,z_2)\frac{dz_1dz_2}{z_1z_2}\\ &= \begin{bmatrix}\boldsymbol{K}_1 & \boldsymbol{K}_2\\ \boldsymbol{K}_3 & \boldsymbol{K}_4\end{bmatrix}\end{aligned}$$

is the local controllability Gramian for the LSS model in (39) with an $m\times m$ submatrix $\boldsymbol{K}_1$ and an $n\times n$ submatrix $\boldsymbol{K}_4$ along its diagonal [29].

Thus, the l_2-scaling constrained frequency-weighted l_2-sensitivity minimization problem can be formulated as follows: *Given matrices $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, obtain a block-diagonal nonsingular matrix $\boldsymbol{T}=\boldsymbol{T}_1\oplus\boldsymbol{T}_4$ which minimizes $S(\boldsymbol{P})$ in (20) subject to l_2-scaling constraints in (21).*

III. PROBLEM SOLUTION

3.1 A Constrained Optimization Method

Solving the optimization problem formulated above consists of several steps. First, we relax the problem of minimizing $S(\boldsymbol{P})$ in (20) subject to l_2-scaling constraints in (21) into the problem

$$\begin{aligned}&\text{minimize } S(\boldsymbol{P}) \text{ in (20)}\\ &\text{subject to } \text{tr}[\boldsymbol{K}_1\boldsymbol{P}_1^{-1}]=m \text{ and } \text{tr}[\boldsymbol{K}_4\boldsymbol{P}_4^{-1}]=n.\end{aligned} \tag{22}$$

If $\text{tr}[\boldsymbol{K}_1\boldsymbol{P}_1^{-1}]=m$ ($\text{tr}[\boldsymbol{K}_4\boldsymbol{P}_4^{-1}]=n$) is satisfied, then an $m\times m$ ($n\times n$) orthogonal matrix $\boldsymbol{U}_1$ ($\boldsymbol{U}_4$) matrix can always be constructed so that $\boldsymbol{T}_1=\boldsymbol{P}_1^{1/2}\boldsymbol{U}_1$ ($\boldsymbol{T}_4=\boldsymbol{P}_4^{1/2}\boldsymbol{U}_4$) satisfies l_2-scaling constraints in (21) [22]. This justifies the relaxation made in (22).

In order to solve problem (22) for $\boldsymbol{P}=\boldsymbol{P}_1\oplus\boldsymbol{P}_4$, we define the Lagrange function of the problem as

$$\begin{aligned}&J(\boldsymbol{P},\lambda_1,\lambda_4)\\ &\quad=\text{tr}[\boldsymbol{M}_1(\boldsymbol{P})\boldsymbol{P}_1]+\text{tr}[\boldsymbol{M}_4(\boldsymbol{P})\boldsymbol{P}_4]+\text{tr}[\boldsymbol{W}_{1B}\boldsymbol{P}_1]\\ &\qquad+\text{tr}[\boldsymbol{W}_{4B}\boldsymbol{P}_4]+\text{tr}[\boldsymbol{K}_{1C}\boldsymbol{P}_1^{-1}]+\text{tr}[\boldsymbol{K}_{4C}\boldsymbol{P}_4^{-1}]\\ &\qquad+\lambda_1(\text{tr}[\boldsymbol{K}_1\boldsymbol{P}_1^{-1}]-m)+\lambda_4(\text{tr}[\boldsymbol{K}_4\boldsymbol{P}_4^{-1}]-n)\end{aligned} \tag{23}$$

where λ_1 and λ_4 are the Lagrange multipliers, and

$$\boldsymbol{M}_A(\boldsymbol{P})=\begin{bmatrix}\boldsymbol{M}_1(\boldsymbol{P}) & \boldsymbol{M}_2(\boldsymbol{P})\\ \boldsymbol{M}_3(\boldsymbol{P}) & \boldsymbol{M}_4(\boldsymbol{P})\end{bmatrix}$$

$$\boldsymbol{W}_B=\begin{bmatrix}\boldsymbol{W}_{1B} & \boldsymbol{W}_{2B}\\ \boldsymbol{W}_{3B} & \boldsymbol{W}_{4B}\end{bmatrix},\quad \boldsymbol{K}_C=\begin{bmatrix}\boldsymbol{K}_{1C} & \boldsymbol{K}_{2C}\\ \boldsymbol{K}_{3C} & \boldsymbol{K}_{4C}\end{bmatrix}$$

with an $m\times m$ submatrix and an $n\times n$ submatrix along its diagonal for each matrix. Using the formula for evaluating matrix gradient [31, p.275]

$$\begin{aligned}\frac{\partial[\text{tr}(\boldsymbol{M}\boldsymbol{X})]}{\partial\boldsymbol{X}} &= \boldsymbol{M}^T\\ \frac{\partial[\text{tr}(\boldsymbol{M}\boldsymbol{X}^{-1})]}{\partial\boldsymbol{X}} &= -(\boldsymbol{X}^{-1}\boldsymbol{M}\boldsymbol{X}^{-1})^T\end{aligned} \tag{24}$$

we compute

$$\begin{aligned}\frac{\partial J(\boldsymbol{P},\lambda_1,\lambda_4)}{\partial\boldsymbol{P}_1} &= \boldsymbol{M}_1(\boldsymbol{P})-\boldsymbol{P}_1^{-1}\boldsymbol{N}_1(\boldsymbol{P})\boldsymbol{P}_1^{-1}+\boldsymbol{W}_{1B}\\ &\quad-\boldsymbol{P}_1^{-1}\boldsymbol{K}_{1C}\boldsymbol{P}_1^{-1}-\lambda_1\boldsymbol{P}_1^{-1}\boldsymbol{K}_1\boldsymbol{P}_1^{-1}\\ \frac{\partial J(\boldsymbol{P},\lambda_1,\lambda_4)}{\partial\boldsymbol{P}_4} &= \boldsymbol{M}_4(\boldsymbol{P})-\boldsymbol{P}_4^{-1}\boldsymbol{N}_4(\boldsymbol{P})\boldsymbol{P}_4^{-1}+\boldsymbol{W}_{4B}\\ &\quad-\boldsymbol{P}_4^{-1}\boldsymbol{K}_{4C}\boldsymbol{P}_4^{-1}-\lambda_4\boldsymbol{P}_4^{-1}\boldsymbol{K}_4\boldsymbol{P}_4^{-1}\end{aligned} \tag{25}$$

where

$$\boldsymbol{N}_A(\boldsymbol{P})=\begin{bmatrix}\boldsymbol{N}_1(\boldsymbol{P}) & \boldsymbol{N}_2(\boldsymbol{P})\\ \boldsymbol{N}_3(\boldsymbol{P}) & \boldsymbol{N}_4(\boldsymbol{P})\end{bmatrix}$$

with an $m\times m$ submatrix $\boldsymbol{N}_1(\boldsymbol{P})$ and an $n\times n$ submatrix $\boldsymbol{N}_4(\boldsymbol{P})$ along its diagonal. Setting $\partial J(\boldsymbol{P},\lambda_1,\lambda_4)/\partial\boldsymbol{P}_1=\boldsymbol{0}$ and $\partial J(\boldsymbol{P},\lambda_1,\lambda_4)/\partial\boldsymbol{P}_4=\boldsymbol{0}$, it follows that

$$\begin{aligned}\boldsymbol{P}_1\boldsymbol{F}_1(\boldsymbol{P})\boldsymbol{P}_1 &= \boldsymbol{G}_1(\boldsymbol{P},\lambda_1)\\ \boldsymbol{P}_4\boldsymbol{F}_4(\boldsymbol{P})\boldsymbol{P}_4 &= \boldsymbol{G}_4(\boldsymbol{P},\lambda_4)\end{aligned} \tag{26}$$

where

$$\begin{aligned}\boldsymbol{F}_1(\boldsymbol{P}) &= \boldsymbol{M}_1(\boldsymbol{P})+\boldsymbol{W}_{1B},\ \boldsymbol{F}_4(\boldsymbol{P})=\boldsymbol{M}_4(\boldsymbol{P})+\boldsymbol{W}_{4B}\\ \boldsymbol{G}_1(\boldsymbol{P},\lambda_1) &= \boldsymbol{N}_1(\boldsymbol{P})+\boldsymbol{K}_{1C}+\lambda_1\boldsymbol{K}_1\\ \boldsymbol{G}_4(\boldsymbol{P},\lambda_4) &= \boldsymbol{N}_4(\boldsymbol{P})+\boldsymbol{K}_{4C}+\lambda_4\boldsymbol{K}_4.\end{aligned}$$

The equations in (26) are highly nonlinear with respect to $\boldsymbol{P}_1$ and $\boldsymbol{P}_4$. Namely, $\boldsymbol{P}_i\boldsymbol{F}_i(\boldsymbol{P})\boldsymbol{P}_i$ for $i=1,4$ has a rational type R/D of nonlinearity, where the degree of the nonlinearity of R is $m+n+1$ and the degree of the nonlinearity of D is $m+n$, while $\boldsymbol{G}_i(\boldsymbol{P},\lambda_i)$ for $i=1,4$ depends on $\boldsymbol{P}$ linearly. An effective approach for solving these equations is to *relax* them into the following recursive second-order matrix equations:

$$\begin{aligned}\boldsymbol{P}_1^{(k+1)}\boldsymbol{F}_1(\boldsymbol{P}^{(k)})\boldsymbol{P}_1^{(k+1)} &= \boldsymbol{G}_1(\boldsymbol{P}^{(k)},\lambda_1^{(k+1)})\\ \boldsymbol{P}_4^{(k+1)}\boldsymbol{F}_4(\boldsymbol{P}^{(k)})\boldsymbol{P}_4^{(k+1)} &= \boldsymbol{G}_4(\boldsymbol{P}^{(k)},\lambda_4^{(k+1)})\end{aligned} \tag{27}$$

with initial condition $\boldsymbol{P}^{(0)} = \boldsymbol{P}_1^{(0)} \oplus \boldsymbol{P}_4^{(0)} = \boldsymbol{I}_{m+n}$. Noting that $\boldsymbol{P}\boldsymbol{W}\boldsymbol{P} = \boldsymbol{M}$ has the unique solution [5]

$$\boldsymbol{P} = \boldsymbol{W}^{-\frac{1}{2}}[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{M}\boldsymbol{W}^{\frac{1}{2}}]^{\frac{1}{2}}\boldsymbol{W}^{-\frac{1}{2}} \tag{28}$$

where $\boldsymbol{W} > 0$ and $\boldsymbol{M} \geq 0$ are symmetric, the unique solutions $\boldsymbol{P}_1^{(k+1)}$ and $\boldsymbol{P}_4^{(k+1)}$ of (27) are found to be

$$\begin{aligned}\boldsymbol{P}_1^{(k+1)} &= \boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{-\frac{1}{2}}[\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}} \\ &\quad \cdot \boldsymbol{G}_1(\boldsymbol{P}^{(k)}, \lambda_1^{(k+1)})\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}}]^{\frac{1}{2}}\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{-\frac{1}{2}} \\ \boldsymbol{P}_4^{(k+1)} &= \boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{-\frac{1}{2}}[\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}} \\ &\quad \cdot \boldsymbol{G}_4(\boldsymbol{P}^{(k)}, \lambda_4^{(k+1)})\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}}]^{\frac{1}{2}}\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{-\frac{1}{2}}.\end{aligned} \tag{29}$$

Here, the Lagrange multipliers $\lambda_1^{(k+1)}$ and $\lambda_4^{(k+1)}$ can be efficiently obtained using a bisection method [32] so that

$$\begin{aligned} f_1(\lambda_1^{(k+1)}) &= m - \text{tr}[\tilde{\boldsymbol{K}}_1^{(k)}\,\tilde{\boldsymbol{G}}_1^{(k)}(\lambda_1^{(k+1)})] = 0 \\ f_4(\lambda_4^{(k+1)}) &= n - \text{tr}[\tilde{\boldsymbol{K}}_4^{(k)}\,\tilde{\boldsymbol{G}}_4^{(k)}(\lambda_4^{(k+1)})] = 0 \end{aligned} \tag{30}$$

are satisfied where

$$\begin{aligned}\tilde{\boldsymbol{K}}_1^{(k)} &= \boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}}\boldsymbol{K}_1\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}} \\ \tilde{\boldsymbol{K}}_4^{(k)} &= \boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}}\boldsymbol{K}_4\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}} \\ \tilde{\boldsymbol{G}}_1^{(k)}(\lambda_1^{(k+1)}) &= [\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}}\boldsymbol{G}_1(\boldsymbol{P}^{(k)}, \lambda_1^{(k+1)})\boldsymbol{F}_1(\boldsymbol{P}^{(k)})^{\frac{1}{2}}]^{-\frac{1}{2}} \\ \tilde{\boldsymbol{G}}_4^{(k)}(\lambda_4^{(k+1)}) &= [\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}}\boldsymbol{G}_4(\boldsymbol{P}^{(k)}, \lambda_4^{(k+1)})\boldsymbol{F}_4(\boldsymbol{P}^{(k)})^{\frac{1}{2}}]^{-\frac{1}{2}}.\end{aligned}$$

The iteration process continues until

$$|J(\boldsymbol{P}^{(k)}, \lambda_1^{(k+1)}, \lambda_4^{(k+1)}) - J(\boldsymbol{P}^{(k-1)}, \lambda_1^{(k)}, \lambda_4^{(k)})| < \varepsilon \tag{31}$$

for a prescribed tolerance $\varepsilon > 0$. If the iteration is terminated at step k, then $\boldsymbol{P}^{(k)}$ is claimed to be a solution point.

It is noted that a straightforward extension of the iterative algorithm reported in [22] for updating λ_1 and λ_4 does not work well in this case, where we do not require $\text{tr}[\boldsymbol{K}\boldsymbol{P}^{-1}] = m+n$, but both $\text{tr}[\boldsymbol{K}_1\boldsymbol{P}_1^{-1}] = m$ and $\text{tr}[\boldsymbol{K}_4\boldsymbol{P}_4^{-1}] = n$. In addition, the bisection method offers an exponential convergence rate of $(1/2)^L$ where L is the number of iterations used. As such, accurate solutions of the Lagrange multipliers $\lambda_1^{(k+1)}$ and $\lambda_4^{(k+1)}$ in (30) can be identified with just a few iterations. In our simulation studies, the bisection method was found considerably faster than the iterative algorithm proposed in [22].

All the Gramians can be evaluated by truncating the corresponding infinite summations, see Appendix I for details.

3.2 An Unconstrained Optimization Method

By defining

$$\hat{\boldsymbol{T}} = \hat{\boldsymbol{T}}_1 \oplus \hat{\boldsymbol{T}}_4 = (\boldsymbol{T}_1 \oplus \boldsymbol{T}_4)^T(\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{-\frac{1}{2}}, \tag{32}$$

it follows that

$$\boldsymbol{T}_i^{-1}\boldsymbol{K}_i\boldsymbol{T}_i^{-T} = \hat{\boldsymbol{T}}_i^{-T}\hat{\boldsymbol{T}}_i^{-1} \quad \text{for} \quad i = 1, 4. \tag{33}$$

Thus, a convenient way to eliminate the l_2-scaling constraints in (21) is to choose $\hat{\boldsymbol{T}}_1^{-1}$ and $\hat{\boldsymbol{T}}_4^{-1}$ as

$$\begin{aligned}\hat{\boldsymbol{T}}_1^{-1} &= \left[\frac{\boldsymbol{t}_{11}}{||\boldsymbol{t}_{11}||}, \frac{\boldsymbol{t}_{12}}{||\boldsymbol{t}_{12}||}, \cdots, \frac{\boldsymbol{t}_{1m}}{||\boldsymbol{t}_{1m}||}\right] \\ \hat{\boldsymbol{T}}_4^{-1} &= \left[\frac{\boldsymbol{t}_{41}}{||\boldsymbol{t}_{41}||}, \frac{\boldsymbol{t}_{42}}{||\boldsymbol{t}_{42}||}, \cdots, \frac{\boldsymbol{t}_{4n}}{||\boldsymbol{t}_{4n}||}\right]\end{aligned} \tag{34}$$

where $\boldsymbol{t}_{1i}$ is an $m \times 1$ vector for $i = 1, 2, \cdots, m$ and $\boldsymbol{t}_{4j}$ is an $n \times 1$ vector for $j = 1, 2, \cdots, n$. With (34), all the diagonal elements of $\hat{\boldsymbol{T}}_1^{-T}\hat{\boldsymbol{T}}_1^{-1}$ and $\hat{\boldsymbol{T}}_4^{-T}\hat{\boldsymbol{T}}_4^{-1}$ are found to be unity. Taking (32) and (34) into account, we conclude that the coordinate transformation matrix $\boldsymbol{T}$ of the form

$$\boldsymbol{T} = \boldsymbol{T}_1 \oplus \boldsymbol{T}_4 = (\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{\frac{1}{2}}(\hat{\boldsymbol{T}}_1 \oplus \hat{\boldsymbol{T}}_4)^T \tag{35}$$

automatically satisfies the l_2-scaling constraints in (21). By substituting (35) into $S(\boldsymbol{P})$ in (20), the frequency-weighted l_2-sensitivity measure can be expressed as

$$J_o(\boldsymbol{x}) = \text{tr}[\hat{\boldsymbol{T}}\hat{\boldsymbol{M}}_A(\hat{\boldsymbol{P}})\hat{\boldsymbol{T}}^T] + \text{tr}[\hat{\boldsymbol{T}}\hat{\boldsymbol{W}}_B\hat{\boldsymbol{T}}^T] + \text{tr}[\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{K}}_C\hat{\boldsymbol{T}}^{-1}] \tag{36}$$

with

$$\begin{aligned}\boldsymbol{x} &= (\boldsymbol{t}_{11}^T, \boldsymbol{t}_{12}^T, \cdots, \boldsymbol{t}_{1m}^T, \boldsymbol{t}_{41}^T, \boldsymbol{t}_{42}^T, \cdots, \boldsymbol{t}_{4n}^T)^T \\ \hat{\boldsymbol{M}}_A(\hat{\boldsymbol{P}}) &= \frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\hat{\boldsymbol{Y}}(z_1, z_2)\hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{Y}}^*(z_1, z_2)\frac{dz_1 dz_2}{z_1 z_2}\end{aligned}$$

where $\hat{\boldsymbol{P}} = \hat{\boldsymbol{T}}^T\hat{\boldsymbol{T}}$ and

$$\begin{aligned}\hat{\boldsymbol{Y}}(z_1, z_2) &= (\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{\frac{1}{2}}\boldsymbol{Y}(z_1, z_2)(\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{-\frac{1}{2}} \\ \boldsymbol{Y}(z_1, z_2) &= W_A(z_1, z_2)[\boldsymbol{F}(z_1, z_2)\boldsymbol{G}(z_1, z_2)]^T \\ \hat{\boldsymbol{W}}_B &= (\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{\frac{1}{2}}\boldsymbol{W}_B(\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{\frac{1}{2}} \\ \hat{\boldsymbol{K}}_C &= (\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{-\frac{1}{2}}\boldsymbol{K}_C(\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{-\frac{1}{2}}.\end{aligned}$$

This shows that the problem of obtaining the block-diagonal nonsingular matrix $\boldsymbol{T} = \boldsymbol{T}_1 \oplus \boldsymbol{T}_4$ which minimizes $S(\boldsymbol{P})$ in (20) subject to the l_2-scaling constraints in (21) can be converted into an unconstrained optimization problem of obtaining an $(m^2+n^2) \times 1$ vector $\boldsymbol{x}$ which minimizes $J_o(\boldsymbol{x})$ in (36).

Applying a quasi-Newton algorithm to minimize $J_o(\boldsymbol{x})$ in (36), in the kth iteration the most recent point $\boldsymbol{x}_k$ is updated to point $\boldsymbol{x}_{k+1}$ as [33]

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k \tag{37}$$

where

$$\begin{aligned}\boldsymbol{d}_k &= -\boldsymbol{S}_k\nabla J_o(\boldsymbol{x}_k), \qquad \alpha_k = arg\ \min_{\alpha}\ J_o(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) \\ \boldsymbol{S}_{k+1} &= \boldsymbol{S}_k + \left(1 + \frac{\boldsymbol{\gamma}_k^T\boldsymbol{S}_k\boldsymbol{\gamma}_k}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k}\right)\frac{\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k} - \frac{\boldsymbol{\delta}_k\boldsymbol{\gamma}_k^T\boldsymbol{S}_k + \boldsymbol{S}_k\boldsymbol{\gamma}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k} \\ \boldsymbol{S}_0 &= \boldsymbol{I}_{m^2+n^2}, \qquad \boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \\ \boldsymbol{\gamma}_k &= \nabla J_o(\boldsymbol{x}_{k+1}) - \nabla J_o(\boldsymbol{x}_k).\end{aligned}$$

Here, $\nabla J_o(\boldsymbol{x})$ is the gradient of $J_o(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, and $\boldsymbol{S}_k$ is a positive-definite approximation of the inverse Hessian matrix of $J_o(\boldsymbol{x})$. The algorithm starts with a trivial initial point

$\boldsymbol{x}_0$ obtained from an initial assignment $\hat{\boldsymbol{T}} = \boldsymbol{I}_{m+n}$, and this iteration process continues until

$$|J_o(\boldsymbol{x}_{k+1}) - J_o(\boldsymbol{x}_k)| < \varepsilon \tag{38}$$

where $\varepsilon > 0$ is a prescribed tolerance.

The implementation of (37) requires the gradient of $J_o(\boldsymbol{x})$, which can be efficiently evaluated using closed-form expressions, see Appendix II for details.

IV. CASE STUDIES

In this section, we examine the proposed algorithms by applying them to a recursive 2-D state-space digital filter. In addition, the algorithms are applied to a 2-D filter utilized in [22] and the results are compared with those obtained by the method of [22].

Example 1: Consider a 2-D stable recursive digital filter realization $(\boldsymbol{A}^o, \boldsymbol{b}^o, \boldsymbol{c}^o, d)_{2,2}$ where [30, p. 971]

$$\boldsymbol{A}^o = \begin{bmatrix} 1.88899 & -0.91219 & -1.00000 & 0.00000 \\ 1.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.02771 & -0.02580 & 1.88899 & 1.00000 \\ -0.02580 & 0.02431 & -0.91219 & 0.00000 \end{bmatrix}$$

$$\boldsymbol{b}^o = \begin{bmatrix} 0.219089 & 0.000000 & -0.028889 & 0.091219 \end{bmatrix}^T$$

$$\boldsymbol{c}^o = \begin{bmatrix} 0.028889 & -0.091219 & -0.219089 & 0.000000 \end{bmatrix}$$

$$d = 0.08900.$$

After carrying out the l_2-scaling for the above realization with a diagonal coordinate matrix

$$\boldsymbol{T}^o = \text{diag}\{9.336421, 9.336414, 1.065102, 0.986642\},$$

we obtain the 2-D state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{2,2}$ characterized by

$$\boldsymbol{A} = \begin{bmatrix} 1.888990 & -0.912189 & -0.114080 & 0.000000 \\ 1.000001 & 0.000000 & 0.000000 & 0.000000 \\ 0.242899 & -0.226156 & 1.888990 & 0.926336 \\ -0.244141 & 0.230041 & -0.984729 & 0.000000 \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} 0.023466 & 0.000000 & -0.027123 & 0.092454 \end{bmatrix}^T$$

$$\boldsymbol{c} = \begin{bmatrix} 0.269720 & -0.851658 & -0.233352 & 0.000000 \end{bmatrix}$$

$$d = 0.08900$$

where $\boldsymbol{A} = \boldsymbol{T}^{o-1}\boldsymbol{A}^o\boldsymbol{T}^o$, $\boldsymbol{b} = \boldsymbol{T}^{o-1}\boldsymbol{b}^o$, and $\boldsymbol{c} = \boldsymbol{c}^o\boldsymbol{T}^o$. The 2-D state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{2,2}$ now satisfies the l_2-scaling constraints. The frequency-weighting functions used in this example were given by a 2-D nonrecursive low-pass digital filter with the following unit-sample response [34, p. 895]:

$$\begin{aligned} w_A(i,j) &= w_B(i,j) = w_C(i,j) \\ &= 0.256322 \exp[-0.103203\{(i-4)^2 + (j-4)^2\}] \end{aligned}$$

for $(0,0) \le (i,j) \le (20,20)$, and zero elsewhere.

Using (A.1) and (A.2) with truncation $(0,0) \le (i,j) \le (150,150)$ to evaluate the Gramians $\boldsymbol{K}_C$, $\boldsymbol{W}_B$, $\boldsymbol{M}_A$ and $\boldsymbol{K}$, it was found that

$$\boldsymbol{K}_C = \begin{bmatrix} 32.944701 & 32.414885 & 2.177594 & -2.390760 \\ 32.414885 & 32.944694 & 2.732580 & -2.852159 \\ 2.177594 & 2.732580 & 4.347875 & -4.136567 \\ -2.390760 & -2.852159 & -4.136567 & 4.056383 \end{bmatrix}$$

$$\boldsymbol{W}_B = 10^3 \begin{bmatrix} 0.429952 & -0.378923 & 0.215338 & 0.250313 \\ -0.378923 & 0.344207 & -0.219002 & -0.242021 \\ 0.215338 & -0.219002 & 3.257832 & 2.969311 \\ 0.250313 & -0.242021 & 2.969311 & 2.795534 \end{bmatrix}$$

$$\boldsymbol{M}_A = 10^5 \begin{bmatrix} 0.599216 & -0.523300 & 0.711632 & 0.788400 \\ -0.523300 & 0.466607 & -0.639356 & -0.707249 \\ 0.711632 & -0.639356 & 6.203669 & 5.638379 \\ 0.788400 & -0.707249 & 5.638379 & 5.322915 \end{bmatrix}$$

$$\boldsymbol{K} = \begin{bmatrix} 1.000000 & 0.978030 & 0.164886 & -0.167063 \\ 0.978030 & 1.000000 & 0.132847 & -0.133855 \\ 0.164886 & 0.132847 & 1.000000 & -0.985382 \\ -0.167063 & -0.133855 & -0.985382 & 1.000000 \end{bmatrix}$$

In what follows, we evaluate the frequency-weighted l_2-sensitivity of the 2-D state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{2,2}$. Later on, this sensitivity measure will be used as a benchmark in the examination of the performance of the proposed algorithms. Using (17), the frequency-weighted l_2-sensitivity of the LSS model $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{2,2}$ was found to be

$$S = 126.614237 \times 10^4.$$

As will be seen next, the proposed algorithms are able to deduce an equivalent state-space realization with much reduced frequency-weighted l_2-sensitivity.

4.1 Application of the Lagrange method

Choosing $\boldsymbol{P}^{(0)} = \boldsymbol{P}_1^{(0)} \oplus \boldsymbol{P}_4^{(0)} = \boldsymbol{I}_4$ in (29) as an initial estimate and a tolerance $\varepsilon = 10^{-8}$ in (31) as well as in the bisection method, it took the Lagrange-based algorithm 10 iterations to converge to the solution

$$\boldsymbol{P}^{opt} = \begin{bmatrix} 1.639466 & 1.715106 \\ 1.715106 & 1.828836 \end{bmatrix} \oplus \begin{bmatrix} 0.901558 & -0.915221 \\ -0.915221 & 0.962667 \end{bmatrix}$$

or equivalently,

$$\boldsymbol{T}^{opt} = \begin{bmatrix} 1.142108 & 0.578841 \\ 1.110711 & 0.771464 \end{bmatrix} \oplus \begin{bmatrix} 0.266022 & -0.911477 \\ -0.094157 & 0.976628 \end{bmatrix}.$$

The minimized frequency-weighted l_2-sensitivity measure in (23) corresponding to the above solution was found to be

$$J(\boldsymbol{P}^{opt}, \lambda_1, \lambda_4) = 4.076278 \times 10^4$$

with $\lambda_1 = -16880.585503$ and $\lambda_4 = 16962.901711$, and the optimal state-space filter structure $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_{2,2}$ (that

minimizes (20) subject to the l_2-scaling constraints in (21)) was synthesized by substituting matrix $\boldsymbol{T}^{opt}$ into (19) as

$$\overline{\boldsymbol{A}} = \begin{bmatrix} 0.930636 & -0.144497 & -0.098301 & 0.336809 \\ 0.140565 & 0.958354 & 0.141528 & -0.484919 \\ 0.024994 & -0.000749 & 0.958805 & 0.115614 \\ -0.021475 & 0.036943 & -0.175791 & 0.930185 \end{bmatrix}$$

$$\overline{\boldsymbol{b}} = \begin{bmatrix} 0.076010 & -0.109434 & 0.332103 & 0.126685 \end{bmatrix}^T$$

$$\overline{\boldsymbol{c}} = \begin{bmatrix} -0.637897 & -0.500899 & -0.062077 & 0.212695 \end{bmatrix}.$$

The profile of the frequency-weighted l_2-sensitivity measure $J(\boldsymbol{P}, \lambda_1, \lambda_4)$ and the profiles of the Lagrange multipliers λ_1 and λ_4 for the first 10 iterations are shown in Figs. 2 and 3, respectively.

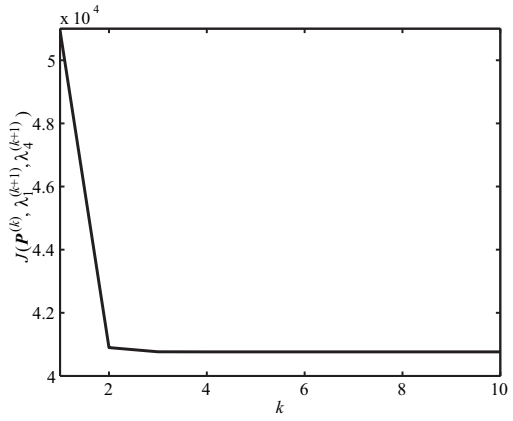

Fig. 2. Profile of $J(\boldsymbol{P}, \lambda_1, \lambda_4)$ during the first 10 iterations.

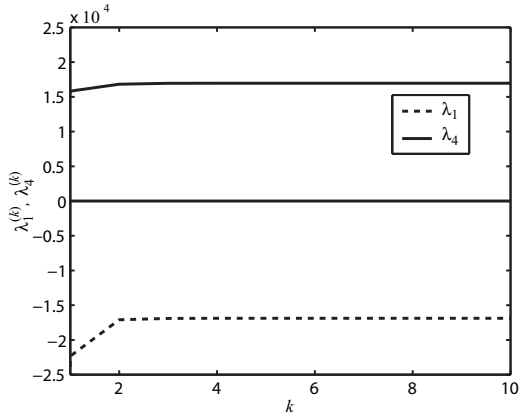

Fig. 3. Profiles of λ_1 and λ_4 during the first 10 iterations.

4.2 Application of the quasi-Newton method

By choosing $\hat{\boldsymbol{T}} = \boldsymbol{I}_2 \oplus \boldsymbol{I}_2$ (therefore $\boldsymbol{T} = (\boldsymbol{K}_1 \oplus \boldsymbol{K}_4)^{1/2}$ in (32)) as an initial estimate and a tolerance $\varepsilon = 10^{-8}$ in (38), the quasi-Newton algorithm took 16 iterations to converge to the solution

$$\hat{\boldsymbol{T}}^{opt} = \begin{bmatrix} 0.906814 & 0.694991 \\ -0.169732 & 1.129830 \end{bmatrix} \oplus \begin{bmatrix} 0.838552 & 0.582931 \\ -0.400572 & 0.939425 \end{bmatrix}$$

or equivalently,

$$\boldsymbol{T}^{opt} = \begin{bmatrix} 1.142108 & 0.578841 \\ 1.110711 & 0.771464 \end{bmatrix} \oplus \begin{bmatrix} 0.266022 & -0.911477 \\ -0.094157 & 0.976628 \end{bmatrix}$$

and the minimized frequency-weighted l_2-sensitivity was found to be

$$J_o(\hat{\boldsymbol{T}}^{opt}) = 4.076278 \times 10^4.$$

We see that the results obtained by this method are identical to those obtained by the Lagrange-based method. The profile of the l_2-sensitivity measure $J_o(\hat{\boldsymbol{T}})$ during the first 16 iterations is shown in Fig. 4.

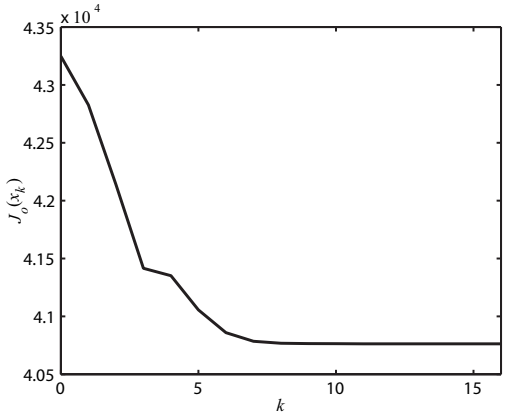

Fig. 4. Profile of $J_o(\hat{\boldsymbol{T}})$ during the first 16 iterations.

We now explain the effectiveness of the optimal realization that minimizes the frequency-weighted l_2-sensitivity subject to l_2-scaling constraints. The magnitude response of the original realization $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{2,2}$ is shown in Fig. 5, where the maximum value and the l_2-norm (which corresponds to a square root of the summation of squared values at 201×201 sampling points) were 11.975265 and 133.212501, respectively. When all coefficients in the original realization were rounded to power-of-two representation with 8 bits after binary point, the magnitude-response deviation between the original realization and that with rounded coefficients is shown in Fig. 6, where the maximum deviation and the l_2-norm were 4.141411 and 22.563131, respectively. Alternatively, when all coefficients in the optimal realization were rounded in the same manner, the magnitude-response deviation between the original realization and the optimal one with rounded coefficients is shown in Fig. 7, where the maximum deviation and the l_2-norm were 0.617833 and 2.323280, respectively. From these figures and data, it is observed that the coefficient sensitivity of the optimal realization is considerably lower than that of the original realization.

Example 2: Let a 2-D stable recursive digital filter realization be specified by $(\boldsymbol{A}^o, \boldsymbol{b}^o, \boldsymbol{c}^o, d)_{4,4}$ where

$$\boldsymbol{A}^o = \begin{bmatrix} \boldsymbol{A}_1^o & \boldsymbol{A}_2^o \\ \boldsymbol{A}_1^o & \boldsymbol{A}_2^o \end{bmatrix}, \quad \boldsymbol{b}^o = \begin{bmatrix} \boldsymbol{b}_1^o \\ \boldsymbol{b}_2^o \end{bmatrix}, \quad \boldsymbol{c}^o = \begin{bmatrix} \boldsymbol{c}_1^o & \boldsymbol{c}_2^o \end{bmatrix}$$

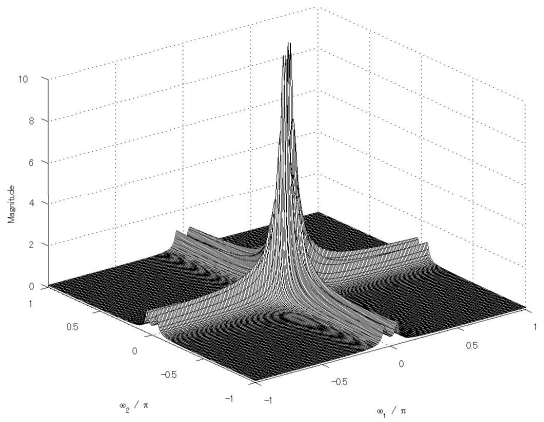

Fig. 5. Magnitude response of the original realization.

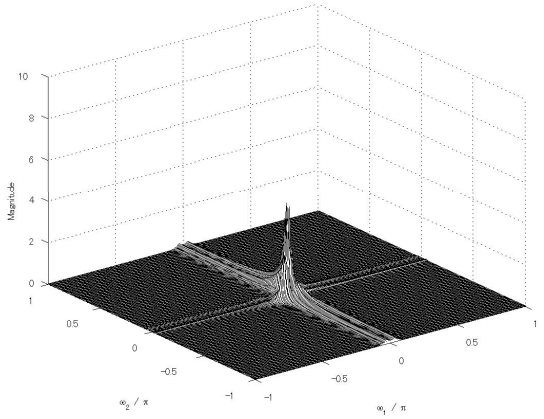

Fig. 6. Magnitude-response deviation between the original realization and that with rounded coefficients.

with

$$\boldsymbol{A}_1^o = \begin{bmatrix} 0.0 & 0.481228 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.510378 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.525287 \\ -0.031857 & 0.298663 & -0.808282 & 1.044600 \end{bmatrix}$$

$$\boldsymbol{A}_2^o = \begin{bmatrix} -0.226080 & 0.776837 & 0.024693 & -0.000933 \\ -0.843550 & 1.610400 & -0.309366 & 0.065898 \\ -1.260339 & 2.005100 & -0.453220 & 0.203118 \\ -1.121498 & 1.636435 & -0.590516 & 0.562890 \end{bmatrix}$$

$$\boldsymbol{b}_1^o = \boldsymbol{b}_2^o = [\ 0.0 \quad 0.0 \quad 0.0 \quad 0.198473\ ]^T$$

$$\boldsymbol{c}_1^o = [\ -0.567054 \quad 0.231913 \quad 0.197016 \quad 0.239932\ ]$$

$$\boldsymbol{c}_2^o = [\ 0.464344 \quad 0.441837 \quad -0.061100 \quad 0.105505\ ]$$

$$d = 0.009430.$$

This 2-D filter was obtained by imbedding the LSS model of Example 2 in [22] into the Roesser LSS model. By performing the l_2-scaling for the above realization with a diagonal coordinate matrix

$$\boldsymbol{T}^o = \text{diag}\{1.000001, 1.000002, 1.000003, 1.000003, 1.000001, 1.000002, 1.000003, 1.000003\}$$

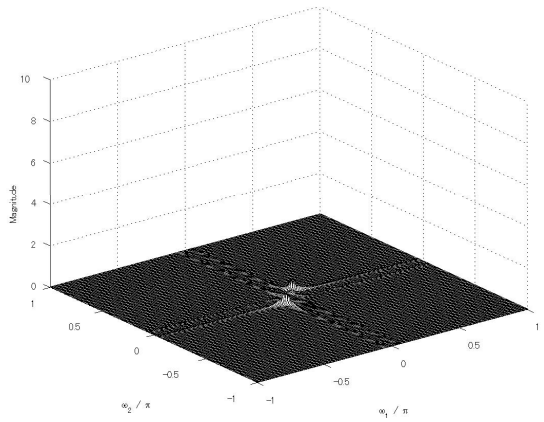

Fig. 7. Magnitude-response deviation between the original realization and the optimal one with rounded coefficients.

and then applying the same frequency-weighted functions as in Example 1 to the resulting realization $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_{4,4}$, the frequency-weighted l_2-sensitivity in (17) was found to be

$$S = 394.423680 \times 10^3$$

with truncation $(0,0) \le (i,j) \le (100,100)$ in (A.1) and (A.2).

4.3 Application of the Lagrange method

Choosing $\boldsymbol{P}^{(0)} = \boldsymbol{P}_1^{(0)} \oplus \boldsymbol{P}_4^{(0)} = \boldsymbol{I}_8$ in (29) as an initial estimate and a tolerance $\varepsilon = 10^{-8}$ in (31) as well as in the bisection method, it took the Lagrange-based algorithm 104 iterations to converge to the solution

$$\boldsymbol{P}^{opt} = \begin{bmatrix} 3.947574 & 3.057718 & 2.498923 & 2.123649 \\ 3.057718 & 2.599605 & 2.183384 & 1.861281 \\ 2.498923 & 2.183384 & 1.894674 & 1.650080 \\ 2.123649 & 1.861281 & 1.650080 & 1.483679 \end{bmatrix} \oplus \begin{bmatrix} 2.234666 & 1.993263 & 1.781601 & 1.653502 \\ 1.993263 & 1.814492 & 1.653747 & 1.534665 \\ 1.781601 & 1.653747 & 1.556000 & 1.464299 \\ 1.653502 & 1.534665 & 1.464299 & 1.432256 \end{bmatrix}$$

or equivalently,

$$\boldsymbol{T}^{opt} = \begin{bmatrix} 0.614712 & 1.232311 & 1.040088 & 0.984546 \\ 0.899544 & 0.797151 & 0.663356 & 0.845538 \\ 0.874957 & 0.504513 & 0.679077 & 0.642997 \\ 0.727174 & 0.269835 & 0.705983 & 0.619414 \end{bmatrix} \oplus \begin{bmatrix} 0.531664 & 0.880708 & 0.800816 & 0.731468 \\ 0.620875 & 0.737566 & 0.617178 & 0.709996 \\ 0.753633 & 0.564844 & 0.551426 & 0.604084 \\ 0.723352 & 0.347355 & 0.658078 & 0.596067 \end{bmatrix}.$$

The minimized frequency-weighted l_2-sensitivity measure in (23) corresponding to the above solution was found to be

$$J(\boldsymbol{P}^{opt}, \lambda_1, \lambda_4) = 4.670177 \times 10^3$$

with $\lambda_1 = 5.955188$ and $\lambda_4 = -14.111003$. The profile of the frequency-weighted l_2-sensitivity measure $J(\boldsymbol{P}, \lambda_1, \lambda_4)$ and the profiles of the Lagrange multipliers λ_1 and λ_4 for the first 104 iterations are shown in Figs. 8 and 9, respectively.

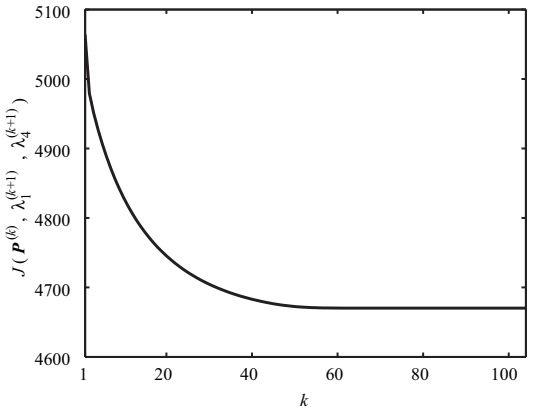

Fig. 8. Profile of $J(\boldsymbol{P}, \lambda_1, \lambda_4)$ during the first 104 iterations.

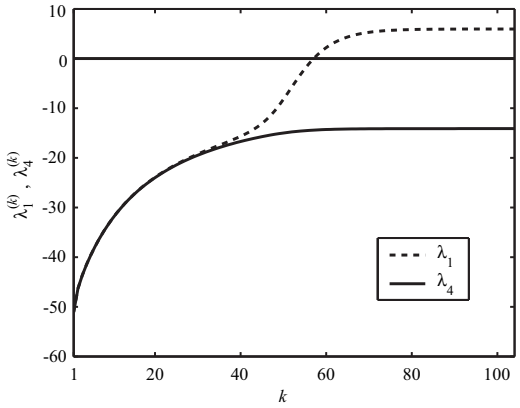

Fig. 9. Profiles of λ_1 and λ_4 during the first 104 iterations.

4.4 Application of the quasi-Newton method

By choosing $\hat{\boldsymbol{T}} = \hat{\boldsymbol{T}}_1 \oplus \hat{\boldsymbol{T}}_4 = \boldsymbol{I}_8$ in (37) as an initial estimate and a tolerance $\varepsilon = 10^{-8}$ in (38), the quasi-Newton algorithm took 54 iterations to converge to

$$\hat{\boldsymbol{T}}^{opt} = \begin{bmatrix} 3.056671 & -2.673365 & 0.575882 & -0.429287 \\ -0.331629 & 2.142411 & -0.401503 & -0.192081 \\ -2.530651 & 0.932586 & 0.553002 & -0.136935 \\ 1.754363 & -0.312582 & 0.624509 & 0.515370 \end{bmatrix} \oplus \begin{bmatrix} 1.307170 & -0.419919 & 0.045538 & -0.194118 \\ 0.762443 & 0.830435 & -0.297531 & 0.062104 \\ -0.405202 & 0.189220 & 0.976564 & -0.250656 \\ 1.071478 & -0.069804 & 0.315533 & 0.828727 \end{bmatrix}$$

or equivalently,

$$\boldsymbol{T}^{opt} = \begin{bmatrix} 0.690639 & 0.697890 & -0.986414 & 1.417930 \\ 0.205523 & 0.802226 & -0.589043 & 1.251725 \\ 0.091157 & 0.584768 & -0.335877 & 1.196490 \\ -0.023116 & 0.340604 & -0.305900 & 1.128518 \end{bmatrix} \oplus \begin{bmatrix} 0.595272 & 0.843931 & 0.159783 & 1.068904 \\ 0.406418 & 0.767684 & 0.282059 & 0.990157 \\ 0.270942 & 0.580437 & 0.379363 & 1.000879 \\ 0.148987 & 0.449337 & 0.230545 & 1.074708 \end{bmatrix}.$$

The minimized frequency-weighted l_2-sensitivity in (36) was found to be

$$J_o(\hat{\boldsymbol{T}}^{opt}) = 4.670177 \times 10^3$$

which is identical to the minimum value of the frequency-weighted l_2-sensitivity measure, obtained by the Lagrange-based method. The profile of the l_2-sensitivity measure $J_o(\hat{\boldsymbol{T}})$ during the first 54 iterations is shown in Fig. 10.

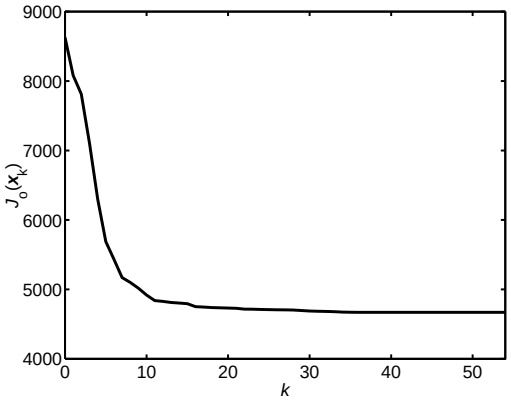

Fig. 10. Profile of $J_o(\hat{\boldsymbol{T}})$ during the first 54 iterations.

To compare the technique reported in [22] with the proposed ones, the optimal realization derived in Example 2 of [22] was imbedded in the Roesser LSS model by using (6). Then the frequency-weighted l_2-sensitivity of the resulting imbedded model was computed from (17) as

$$S = 5.564659 \times 10^3.$$

It is observed that this value is 1.192 times greater than the minimized frequency-weighted l_2-sensitivity obtained by the proposed techniques. Moreover, the proposed Lagrange-based method attains considerably faster convergence than that reported in [22], which required 2000 iterations for its convergence.

V. CONCLUSION

We have investigated the problem of minimizing the frequency-weighted l_2-sensitivity subject to l_2-scaling constraints for 2-D state-space digital filters described by the Roesser LSS model. It has been shown that the FM second LSS model can be imbedded in the Roesser LSS model as a special case. Two iterative methods have been developed to solve the problem at hand. The first iterative method is based on the introduction of a Lagrange function and makes use of an efficient bisection method. In our simulation studies, the bisection method was found to be considerably faster than the iterative method proposed in [22]. The second iterative method relies on the conversion of the constrained optimization problem into an unconstrained optimization formulation and utilizes an efficient quasi-Newton algorithm. The optimal state-space realization with minimum frequency-weighted l_2-sensitivity and no overflow has then been constructed by applying an appropriate coordinate transformation. Our computer simulation results have demonstrated the validity and effectiveness of the proposed techniques.

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}^h(i+1,j+1)\\ \boldsymbol{x}^v(i+1,j+1)\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j+1)\\ \boldsymbol{x}^v(i,j+1)\end{bmatrix}\\ &\quad+\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i+1,j)\\ \boldsymbol{x}^v(i+1,j)\end{bmatrix}\\ &\quad+\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{0}\end{bmatrix}u(i,j+1)+\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}_2\end{bmatrix}u(i+1,j)\\ y(i,j) &= \begin{bmatrix}\boldsymbol{c}_1 & \boldsymbol{c}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^h(i,j)\\ \boldsymbol{x}^v(i,j)\end{bmatrix}+d\,u(i,j)\end{aligned} \tag{39}$$

APPENDIX I

COMPUTATIONS OF GRAMIANS

The matrices $\boldsymbol{K}_C$, $\boldsymbol{W}_B$, and $\boldsymbol{M}_A$ can be computed using

$$\begin{aligned}\boldsymbol{K}_C &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{f}_C(i,j)\boldsymbol{f}_C^T(i,j)\\ \boldsymbol{W}_B &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{g}_B^T(i,j)\boldsymbol{g}_B(i,j)\\ \boldsymbol{M}_A &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{H}_A^T(i,j)\boldsymbol{H}_A(i,j)\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}\boldsymbol{A}^{(1,0)} &= \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{A}^{(0,1)} = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{bmatrix}\\ \boldsymbol{A}^{(0,0)} &= \boldsymbol{I}_{m+n},\ \boldsymbol{A}^{(-i,j)} = \boldsymbol{0}\,(i\geq 1),\ \boldsymbol{A}^{(i,-j)} = \boldsymbol{0}\,(j\geq 1)\\ \boldsymbol{A}^{(i,j)} &= \boldsymbol{A}^{(1,0)}\boldsymbol{A}^{(i-1,j)}+\boldsymbol{A}^{(0,1)}\boldsymbol{A}^{(i,j-1)}\\ &= \boldsymbol{A}^{(i-1,j)}\boldsymbol{A}^{(1,0)}+\boldsymbol{A}^{(i,j-1)}\boldsymbol{A}^{(0,1)},\quad (i,j)>(0,0)\\ \boldsymbol{f}\,(i,j) &= \boldsymbol{A}^{(i-1,j)}\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{0}\end{bmatrix}+\boldsymbol{A}^{(i,j-1)}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}_2\end{bmatrix}\\ \boldsymbol{g}\,(i,j) &= \boldsymbol{c}\boldsymbol{A}^{(i-1,j)}\begin{bmatrix}\boldsymbol{I}_m & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}+\boldsymbol{c}\boldsymbol{A}^{(i,j-1)}\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_n\end{bmatrix}\\ \boldsymbol{H}(i,j) &= \sum_{(0,0)\leq(k,r)<(i,j)}\sum \boldsymbol{f}\,(k,r)\boldsymbol{g}\,(i-k,j-r)\\ \boldsymbol{f}_C(i,j) &= \sum_{(0,0)\leq(k,r)<(i,j)}\sum w_C(k,r)\boldsymbol{f}(i-k,j-r)\\ \boldsymbol{g}_B(i,j) &= \sum_{(0,0)\leq(k,r)<(i,j)}\sum w_B(k,r)\boldsymbol{g}\,(i-k,j-r)\\ \boldsymbol{H}_A(i,j) &= \sum_{(0,0)\leq(k,r)<(i,j)}\sum w_A(k,r)\boldsymbol{H}\,(i-k,j-r)\end{aligned}$$

with partial ordering for integer pairs (i,j) [26, p. 2], and $w_A(k,r)$, $w_B(k,r)$, and $w_C(k,r)$ denoting the unit-sample responses of frequency-weighting functions $W_A(z_1,z_2)$, $W_B(z_1,z_2)$, and $W_C(z_1,z_2)$, respectively.

The local controllability Gramian $\boldsymbol{K}$ and the other Gramians $\boldsymbol{M}_A(\boldsymbol{P})$, $\boldsymbol{N}_A(\boldsymbol{P})$ and $\hat{\boldsymbol{M}}_A(\hat{\boldsymbol{P}})$ can be computed using

$$\begin{aligned}\boldsymbol{K} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{f}\,(i,j)\boldsymbol{f}^{\,T}(i,j)\\ \boldsymbol{M}_A(\boldsymbol{P}) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{H}_A^T(i,j)\,\boldsymbol{P}^{-1}\boldsymbol{H}_A(i,j)\\ \boldsymbol{N}_A(\boldsymbol{P}) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{H}_A(i,j)\,\boldsymbol{P}\boldsymbol{H}_A^T(i,j)\\ \hat{\boldsymbol{M}}_A(\hat{\boldsymbol{P}}) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\hat{\boldsymbol{H}}_A^T(i,j)\hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{H}}_A(i,j)\end{aligned} \tag{A.2}$$

where

$$\hat{\boldsymbol{H}}_A(i,j) = (\boldsymbol{K}_1\oplus\boldsymbol{K}_4)^{-\frac{1}{2}}\boldsymbol{H}_A(i,j)(\boldsymbol{K}_1\oplus\boldsymbol{K}_4)^{-\frac{1}{2}}.$$

APPENDIX II

GRADIENT EVALUATION OF $J_o(\boldsymbol{x})$

$$\begin{aligned}\frac{\partial J_o(\hat{\boldsymbol{T}})}{\partial t_{ij}} &= \lim_{\Delta\to 0}\frac{J_o(\hat{\boldsymbol{T}}_{ij})-J_o(\hat{\boldsymbol{T}})}{\Delta}\\ &= 2\beta_1-2\beta_2+2\beta_3-2\beta_4\end{aligned} \tag{A.3}$$

where $\hat{\boldsymbol{T}}_{ij}$ is the matrix obtained from $\hat{\boldsymbol{T}}=\hat{\boldsymbol{T}}_1\oplus\hat{\boldsymbol{T}}_4$ with a perturbed (i,j)th component, which is given by [35, p. 655]

$$\begin{aligned}\hat{\boldsymbol{T}}_{ij} &= \hat{\boldsymbol{T}}+\frac{\Delta\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}\boldsymbol{e}_j^T\hat{\boldsymbol{T}}}{1-\Delta\boldsymbol{e}_j^T\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}},\qquad \hat{\boldsymbol{T}}_{ij}^{-1} = \hat{\boldsymbol{T}}^{-1}-\Delta\boldsymbol{g}_{ij}\boldsymbol{e}_j^T\\ \boldsymbol{g}_{ij} &= \partial\left\{\frac{\boldsymbol{t}_j}{||\boldsymbol{t}_j||}\right\}/\partial t_{ij} = \frac{1}{||\boldsymbol{t}_j||^3}(t_{ij}\boldsymbol{t}_j-||\boldsymbol{t}_j||^2\boldsymbol{e}_i)\\ \beta_1 &= \boldsymbol{e}_j^T\hat{\boldsymbol{M}}_A(\hat{\boldsymbol{T}})\,\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}\\ \beta_2 &= \boldsymbol{e}_j^T\hat{\boldsymbol{T}}^{-T}\left[\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\hat{\boldsymbol{H}}_A(p,q)\hat{\boldsymbol{T}}^{\,T}\hat{\boldsymbol{T}}\hat{\boldsymbol{H}}_A^{\,T}(p,q)\right]\boldsymbol{g}_{ij}\\ \beta_3 &= \boldsymbol{e}_j^T\hat{\boldsymbol{T}}\,\hat{\boldsymbol{W}}_B\hat{\boldsymbol{T}}^{\,T}\hat{\boldsymbol{T}}\boldsymbol{g}_{ij},\qquad \beta_4 = \boldsymbol{e}_j^T\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{K}}_c\,\boldsymbol{g}_{ij}\end{aligned}$$

with

$$\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_{m+n}\} = \{\boldsymbol{t}_{11},\boldsymbol{t}_{12},\cdots,\boldsymbol{t}_{1m}\}\cup\{\boldsymbol{t}_{41},\boldsymbol{t}_{42},\cdots,\boldsymbol{t}_{4n}\}.$$

REFERENCES

[1] L. Thiele, "Design of sensitivity and round-off noise optimal state-space discrete systems," *Int. J. Circuit Theory Appl.*, vol. 12, pp. 39-46, Jan. 1984.

[2] V. Tavsanoglu and L. Thiele,"Optimal design of state-space digital filters by simultaneous minimization of sensitivity and roundoff noise," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 884-888, Oct. 1984.

[3] L. Thiele,"On the sensitivity of linear state-space systems," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 502-510, May 1986.

[4] M. Iwatsuki, M. Kawamata and T. Higuchi, "Statistical sensitivity and minimum sensitivity structures with fewer coefficients in discrete time linear systems," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 72-80, Jan. 1989.

[5] G. Li, B. D. O. Anderson, M. Gevers and J. E. Perkins, "Optimal FWL design of state-space digital systems with weighted sensitivity minimization and sparseness consideration," *IEEE Trans. Circuits Syst. I*, vol. 39, pp. 365-377, May 1992.

[6] W.-Y. Yan and J. B. Moore, "On L^2-sensitivity minimization of linear state-space systems," *IEEE Trans. Circuits Syst. I*, vol. 39, pp. 641-648, Aug. 1992.

[7] G. Li and M. Gevers, "Optimal synthetic FWL design of state-space digital filters", in *Proc. 1992 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, pp. 429-432.

[8] M. Gevers and G. Li, *Parameterizations in Control, Estimation and Filtering Problems: Accuracy Aspects*, Springer-Verlag, 1993.

[9] C. Xiao, "Improved L_2-sensitivity for state-space digital system," *IEEE Trans. Signal Processing*, vol. 45, pp. 837-840, Apr. 1997.

[10] T. Hinamoto, S. Yokoyama, T. Inoue, W. Zeng and W.-S. Lu, "Analysis and minimization of L_2-sensitivity for linear systems and two-dimensional state-space filters using general controllability and observability Gramians," *IEEE Trans. Circuits Syst. I*, vol. 49, pp. 1279-1289, Sept. 2002.

[11] S. Yamaki, M. Abe and M. Kawamata, "A closed form solution to L_2-sensitivity minimization of second-order state-space digital filters," in *Proc. 2006 IEEE Int. Symp. Circuits Syst*., pp. 5223-5226.

[12] M. Kawamata, T. Lin and T. Higuchi, "Minimization of sensitivity of 2-D state-space digital filters and its relation to 2-D balanced realizations," in *Proc. 1987 IEEE Int. Symp. Circuits Syst*., pp. 710-713.

[13] T. Hinamoto, T. Hamanaka and S. Maekawa, "Synthesis of 2-D state-space digital filters with low sensitivity based on the Fornasini-Marchesini model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, pp. 1587-1594, Sept. 1990.

[14] T. Hinamoto, T. Takao and M. Muneyasu, "Synthesis of 2-D separable-denominator digital filters with low sensitivity," *J. Franklin Institute*, vol. 329, pp. 1063-1080, 1992.

[15] T. Hinamoto and T. Takao, "Synthesis of 2-D state-space filter structures with low frequency-weighted sensitivity," *IEEE Trans. Circuits Syst. II*, vol. 39, pp. 646-651, Sept. 1992.

[16] T. Hinamoto and T. Takao, "Minimization of frequency-weighting sensitivity in 2-D systems based on the Fornasini-Marchesini second model, " in *1992 IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 401-404.

[17] T. Hinamoto, Y. Zempo, Y. Nishino and W.-S. Lu, "An analytical approach for the synthesis of two-dimensional state-space filter structures with minimum weighted sensitivity," *IEEE Trans. Circuits Syst. I*, vol. 46, pp. 1172-1183, Oct. 1999.

[18] G. Li, "On frequency weighted minimal L_2 sensitivity of 2-D systems using Fornasini-Marchesini LSS model", *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 642-646, July 1997.

[19] G. Li, "Two-dimensional system optimal realizations with L_2-sensitivity minimization," *IEEE Trans. Signal Processing*, vol. 46, pp. 809-813, Mar. 1998.

[20] T. Hinamoto and Y. Sugie, "L_2-sensitivity analysis and minimization of 2-D separable-denominator state-space digital filters," *IEEE Trans. Signal Processing*, vol. 50, pp. 3107-3114, Dec. 2002.

[21] T. Hinamoto, H. Ohnishi and W.-S. Lu, "Minimization of L_2-sensitivity for state-space digital filters subject to L_2-dynamic-range scaling constraints," *IEEE Trans. Circuits Syst.-II*, vol. 52, pp. 641-645, Oct. 2005.

[22] T. Hinamoto, K. Iwata and W.-S. Lu, "L_2-sensitivity Minimization of one- and two-dimensional state-space digital filters subject to L_2-scaling constraints," *IEEE Trans. Signal Processing*, vol. 54, pp. 1804-1812, May 2006.

[23] S. Yamaki, M. Abe and M. Kawamata, "A novel approach to L_2-sensitivity minimization of digital filters subject to L_2-scaling constraints," in *Proc. 2006 IEEE Int. Symp. Circuits Syst*., pp. 5219-5222.

[24] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed-point digital filters," *IEEE Trans. Circuits Syst*., vol. CAS-23, pp. 551-562, Sept. 1976.

[25] S. Y. Hwang, "Minimum uncorrelated unit noise in state-space digital filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 273-281, Aug. 1977.

[26] R. P. Roesser, "A discrete state-space model for linear image processing," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 1-10, Feb. 1975.

[27] E. Fornasini and G. Marchesini, "Doubly-indexed dynamical systems: State-space models and structural properties," *Math Syst. Theory*, vol. 12, pp. 59-72, 1978.

[28] T. Hinamoto, "A novel local state-space model for 2-D digital filters and its properties," in *Proc. 2001 IEEE Int. Symp. Circuits Syst*., vol.2, pp. 545-548.

[29] S. Kung, B. C. Levy, M. Morf and T. Kailath, "New results in 2-D systems theory, Part II: 2-D state-space model—Realization and the notions of controllability, observability, and minimality," *Proc*. IEEE, vol. 65, pp. 945-961, June 1977.

[30] W.-S. Lu and A. Antoniou, "Synthesis of 2-D state-space fixed-point digital filter structures with minimum roundoff noise," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 965-973, Oct. 1986.

[31] L. L. Scharf, Statistical Signal Processing, Reading, MA: Addison-Wesley, 1991.

[32] H. Togawa, Handbook of Numerical Methods, Saiensu-sha, Tokyo, 1992.

[33] R. Fletcher, Practical Methods of Optimization, 2nd ed., Wiley, New York, 1987.

[34] S. A. H. Aly and M. M. Fahmy, "Spatial-domain design of two-dimensional recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 892-901, Oct. 1980.

[35] T. Kailath, *Linear System*, Englewood Cliffs, N.J.: Prentice Hall, 1980.