L_2-Sensitivity Minimization of 2-D Separable-Denominator State-Space Digital Filters Subject to L_2-Scaling Constraints

OSEMEKHIAN I. OMOIFO* and TAKAO HINAMOTO†
Hiroshima University
JAPAN

Abstract: - The problem of minimizing an L_2-sensitivity measure subject to L_2-norm dynamic-range scaling constraints for two-dimensional (2-D) separable-denominator digital filters is formulated. The constrained optimization problem is converted into an unconstrained optimization problem by using linear-algebraic techniques. Next, an efficient quasi-Newton algorithm is applied with closed-form formula for gradient evaluation to solve the unconstrained optimization problem. The optimal filter structure is then constructed by employing the resulting coordinate transformation matrix that minimizes the L_2-sensitivity measure subject to the scaling constraints. A numerical example is presented to illustrate the utility of the proposed technique.

Key-Words: - L_2-sensitivity minimization, L_2-scaling constraints, 2-D separable-denominator digital filters, quasi-Newton method

1 Introduction

In digital filter implementation, the synthesis of a state-space digital filter is known as the problem of obtaining a suitable set of state-space equations that describe a desired transfer function $H(z)$. However, the state-space equations corresponding to a transfer function $H(z)$ are not unique. Naturally, among the infinite number of state-space descriptions of $H(z)$, one may want to identify a state-space description that minimizes a suitable sensitivity measure. When realizing a fixed-point state-space description with finite word length (FWL) from a transfer function with infinite accuracy coefficients, the coefficients in the state-space description must be truncated or rounded to fit the FWL constraints. Since the quantization of the coefficients of the digital filter alters the characteristics of the designed digital filter, the sensitivity with respect to the coefficients of the digital filter is considered to be a measure of the influence of coefficient quantization. A number of sensitivity measures have been defined and applied to both one-dimensional (1-D) digital filters [1]–[7] and 2-D digital filters [8]–[16]. Presently, two main classes of techniques for constructing the state-space description that minimizes the coefficient sensitivity exist: mixed L_1/L_2-sensitivity minimization [1]–[4], [8]–[13] and pure L_2-sensitivity minimization [5]–[7], [13]–[16].

In this paper, we investigate the problem of minimizing an L_2-sensitivity measure subject to L_2-norm dynamic-range scaling constraints for 2-D separable-denominator digital filters. The L_2-norm dynamic-range scaling constraints are imposed on the synthesis since it is well known that the use of scaling constraints can be beneficial for suppressing overflow oscillations [17], [18]. This paper is organized as follows. In section 2, we present a standard definition for a pure L_2-sensitivity of a 2-D separable-denominator digital filter with respect to its realization coefficients and provide detailed analysis for this L_2-sensitivity measure. In section 3, we present our idea and develop a method for obtaining the optimal realization. In section 4, we illustrate the effectiveness of the proposed technique through a computer simulation. Finally, we provide some concluding remarks in section 5.

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

2 Sensitivity Analysis

A 2-D separable-denominator digital filter can be described by the Roesser local state-space (LSS) model [19], [20]

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

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 $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_4$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, and d are real constant matrices of appropriate dimensions. The LSS model is assumed to be asymptotically stable, separately locally controllable and separately locally observable [20]. The transfer function of the system in (1) is given by

$$\begin{aligned} H(z_1,z_2) &= \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 \end{bmatrix} \begin{bmatrix} z_1\boldsymbol{I}_m - \boldsymbol{A}_1 & -\boldsymbol{A}_2 \\ \boldsymbol{0} & z_2\boldsymbol{I}_n - \boldsymbol{A}_4 \end{bmatrix}^{-1} \\ &\quad \cdot \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} + d \\ &= \begin{bmatrix} 1 & \boldsymbol{c}_1(z_1\boldsymbol{I}_m - \boldsymbol{A}_1)^{-1} \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} d & \boldsymbol{c}_2 \\ \boldsymbol{b}_1 & \boldsymbol{A}_2 \end{bmatrix} \begin{bmatrix} 1 \\ (z_2\boldsymbol{I}_n - \boldsymbol{A}_4)^{-1}\boldsymbol{b}_2 \end{bmatrix}. \end{aligned} \tag{2}$$

The L_2-sensitivity of the system in (1) 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 $\boldsymbol{X}$, differentiable with respect to all the entries of $\boldsymbol{X}$. The sensitivity function of f with respect to $\boldsymbol{X}$ is defined as

$$\boldsymbol{S}_{\boldsymbol{X}} = \frac{\partial f}{\partial \boldsymbol{X}}, \qquad (\boldsymbol{S}_{\boldsymbol{X}})_{ij} = \frac{\partial f}{\partial x_{ij}} \tag{3}$$

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

Definition 2: 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_p-norm of $\boldsymbol{X}(z_1,z_2)$ is defined as

$$\|\boldsymbol{X}\|_p = \left[\frac{1}{(2\pi \mathrm{j})^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \|\boldsymbol{X}(z_1,z_2)\|_F^p \frac{\mathrm{d}z_1 \mathrm{d}z_2}{z_1 z_2} \right]^{\frac{1}{p}} \tag{4}$$

where $\|\boldsymbol{X}(z_1,z_2)\|_F$ is the Frobenius norm of the matrix $\boldsymbol{X}(z_1,z_2)$ defined by

$$\|\boldsymbol{X}(z_1,z_2)\|_F = \left[\sum_{p=1}^{m} \sum_{q=1}^{n} |x_{pq}(z_1,z_2)|^2 \right]^{\frac{1}{2}}.$$

From (2) and Definitions 1 and 2, the overall L_2-sensitivity measure for the filter in (1) is defined as

$$\begin{aligned} S &= \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{A}_1} \right\|_2^2 + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{A}_2} \right\|_2^2 \\ &\quad + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{A}_4} \right\|_2^2 + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{b}_1} \right\|_2^2 \\ &\quad + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{b}_2} \right\|_2^2 + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{c}_1^T} \right\|_2^2 \\ &\quad + \left\| \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{c}_2^T} \right\|_2^2 \end{aligned} \tag{5}$$

Defining

$$\begin{aligned} \boldsymbol{P}(z_2) &= (z_2\boldsymbol{I}_n - \boldsymbol{A}_4)^{-1}\boldsymbol{b}_2 \\ \boldsymbol{Q}(z_1) &= \boldsymbol{c}_1(z_1\boldsymbol{I}_m - \boldsymbol{A}_1)^{-1} \\ \boldsymbol{F}(z_1,z_2) &= (z_1\boldsymbol{I}_m - \boldsymbol{A}_1)^{-1}\left[\boldsymbol{b}_1 + \boldsymbol{A}_2\boldsymbol{P}(z_2)\right] \\ \boldsymbol{G}(z_1,z_2) &= \left[\boldsymbol{c}_2 + \boldsymbol{Q}(z_1)\boldsymbol{A}_2\right](z_2\boldsymbol{I}_n - \boldsymbol{A}_4)^{-1} \end{aligned}$$

it follows that

$$\begin{aligned} S &= \mathrm{tr}\left[\boldsymbol{M}_{\boldsymbol{A}_1} + \boldsymbol{M}_{\boldsymbol{A}_4} + \boldsymbol{W}^h + \boldsymbol{W}^v + \boldsymbol{K}^h + \boldsymbol{K}^v\right] \\ &\quad + \mathrm{tr}\left[\boldsymbol{W}^h\right] \mathrm{tr}\left[\boldsymbol{K}^v\right] \end{aligned} \tag{6}$$

where

$$\begin{aligned} \boldsymbol{M}_{\boldsymbol{A}_1} &= \frac{1}{(2\pi \mathrm{j})^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \left[\boldsymbol{F}(z_1^{-1},z_2^{-1})\boldsymbol{Q}(z_1^{-1})\right] \\ &\quad \cdot \left[\boldsymbol{Q}^*(z_1)\boldsymbol{F}^*(z_1,z_2)\right] \frac{\mathrm{d}z_1 \mathrm{d}z_2}{z_1 z_2} \\ \boldsymbol{M}_{\boldsymbol{A}_4} &= \frac{1}{(2\pi \mathrm{j})^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \left[\boldsymbol{G}^*(z_1,z_2)\boldsymbol{P}^*(z_2)\right] \\ &\quad \cdot \left[\boldsymbol{P}(z_2^{-1})\boldsymbol{G}(z_1^{-1},z_2^{-1})\right] \frac{\mathrm{d}z_1 \mathrm{d}z_2}{z_1 z_2} \end{aligned}$$

$$\boldsymbol{K}^v = \frac{1}{2\pi \mathrm{j}} \oint_{|z_2|=1} \boldsymbol{P}(z_2)\boldsymbol{P}^*(z_2) \frac{\mathrm{d}z_2}{z_2}$$
$$\boldsymbol{K}^h = \frac{1}{(2\pi \mathrm{j})^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \boldsymbol{F}(z_1,z_2)\boldsymbol{F}^*(z_1,z_2) \frac{\mathrm{d}z_1 \mathrm{d}z_2}{z_1 z_2}$$
$$\boldsymbol{W}^h = \frac{1}{2\pi \mathrm{j}} \oint_{|z_1|=1} \boldsymbol{Q}^*(z_1)\boldsymbol{Q}(z_1) \frac{\mathrm{d}z_1}{z_1}$$
$$\boldsymbol{W}^v = \frac{1}{(2\pi \mathrm{j})^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \boldsymbol{G}^*(z_1,z_2)\boldsymbol{G}(z_1,z_2) \frac{\mathrm{d}z_1 \mathrm{d}z_2}{z_1 z_2}.$$

Here, $\boldsymbol{K}^h$ and $\boldsymbol{K}^v$ ($\boldsymbol{W}^h$ and $\boldsymbol{W}^v$) are referred to as the horizontal and vertical local controllability (observability)

Gramians, respectively [20]. The sensitivity is independent of the state-space parameter d in (2) and therefore it is neglected here. It is easy to show that the L_2-sensitivity measure in (6) can be expressed as [16]

$$\begin{aligned} S &= \sum_{i=0}^{n} \sigma_i^v \mathrm{tr}\left[\boldsymbol{W}_i^h\right] + \sum_{j=0}^{m} \sigma_j^h \mathrm{tr}\left[\boldsymbol{K}_j^v\right] \\ &\quad + \mathrm{tr}\left[\boldsymbol{W}^h + \boldsymbol{W}^v + \boldsymbol{K}^h + \boldsymbol{K}^v\right] \\ &\quad + \mathrm{tr}\left[\boldsymbol{W}^h\right] \mathrm{tr}\left[\boldsymbol{K}^v\right] \end{aligned} \tag{7}$$

where all the Gramians can be obtained by solving the following Lyapunov equations

$$\begin{aligned} \begin{bmatrix} \boldsymbol{W}_i^h & * \\ * & * \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_1 & \tilde{\boldsymbol{u}}_i \boldsymbol{c}_1 \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_i^h & * \\ * & * \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \boldsymbol{A}_1 & \tilde{\boldsymbol{u}}_i \boldsymbol{c}_1 \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix}^T + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{K}_j^v & * \\ * & * \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_4 & \boldsymbol{0} \\ \boldsymbol{b}_2 \tilde{\boldsymbol{v}}_j^T & \boldsymbol{A}_4 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{K}_j^v & * \\ * & * \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \boldsymbol{A}_4 & \boldsymbol{0} \\ \boldsymbol{b}_2 \tilde{\boldsymbol{v}}_j^T & \boldsymbol{A}_4 \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix} \\ \boldsymbol{K}^v &= \boldsymbol{A}_4 \boldsymbol{K}^v \boldsymbol{A}_4^T + \boldsymbol{b}_2 \boldsymbol{b}_2^T \\ \boldsymbol{K}^h &= \boldsymbol{A}_1 \boldsymbol{K}^h \boldsymbol{A}_1^T + \boldsymbol{A}_2 \boldsymbol{K}^v \boldsymbol{A}_2^T + \boldsymbol{b}_1 \boldsymbol{b}_1^T \\ \boldsymbol{W}^h &= \boldsymbol{A}_1^T \boldsymbol{W}^h \boldsymbol{A}_1 + \boldsymbol{c}_1^T \boldsymbol{c}_1 \\ \boldsymbol{W}^v &= \boldsymbol{A}_4^T \boldsymbol{W}^v \boldsymbol{A}_4 + \boldsymbol{A}_2^T \boldsymbol{W}^h \boldsymbol{A}_2 + \boldsymbol{c}_2^T \boldsymbol{c}_2 \end{aligned}$$

$\tilde{\boldsymbol{u}}_i$ and $\tilde{\boldsymbol{v}}_j$ are obtained by performing eigenvalue-eigenvector decomposition on $\boldsymbol{W}^h$ and $\boldsymbol{K}^v$

$$\begin{aligned} \boldsymbol{K}^v &= \boldsymbol{U}\boldsymbol{\Sigma}^v \boldsymbol{U}^T = \sum_{i=1}^{n} \sigma_i^v \boldsymbol{u}_i \boldsymbol{u}_i^T \\ \boldsymbol{W}^h &= \boldsymbol{V}\boldsymbol{\Sigma}^h \boldsymbol{V}^T = \sum_{j=1}^{m} \sigma_j^h \boldsymbol{v}_j \boldsymbol{v}_j^T \end{aligned} \tag{8}$$

and then using the following relationship

$$\begin{aligned} \sigma_0^h &= \sigma_0^v = 1 \\ \tilde{\boldsymbol{u}}_0 &= \boldsymbol{b}_1, \;\; \tilde{\boldsymbol{u}}_i = \boldsymbol{A}_2 \boldsymbol{u}_i, \;\; i = 1, 2, \ldots, n \\ \tilde{\boldsymbol{v}}_0 &= \boldsymbol{c}_2^T, \;\; \tilde{\boldsymbol{v}}_j = \boldsymbol{A}_2^T \boldsymbol{v}_j, \;\; j = 1, 2, \ldots, m. \end{aligned}$$

to obtain $\tilde{\boldsymbol{u}}_i$ and $\tilde{\boldsymbol{v}}_j$

3 Sensitivity Minimization

If a coordinate transformation defined by

$$\begin{bmatrix} \overline{\boldsymbol{x}}^h(i,j) \\ \overline{\boldsymbol{x}}^v(i,j) \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}4 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}^h(i,j) \\ \boldsymbol{x}^v(i,j) \end{bmatrix} \tag{9}$$

is applied to the LSS model in (1), then the new realization is related to the original one by

$$\begin{aligned} \overline{\boldsymbol{A}}_1 &= \boldsymbol{T}_1^{-1} \boldsymbol{A}_1 \boldsymbol{T}_1, \;\; \overline{\boldsymbol{A}}_2 = \boldsymbol{T}_1^{-1} \boldsymbol{A}_2 \boldsymbol{T}_4 \\ \overline{\boldsymbol{A}}_4 &= \boldsymbol{T}_4^{-1} \boldsymbol{A}_4 \boldsymbol{T}_4 \\ \overline{\boldsymbol{b}}_1 &= \boldsymbol{T}_1^{-1} \boldsymbol{b}_1, \quad \overline{\boldsymbol{b}}_2 = \boldsymbol{T}_4^{-1} \boldsymbol{b}_2 \\ \overline{\boldsymbol{c}}_1 &= \boldsymbol{c}_1 \boldsymbol{T}_1, \qquad \overline{\boldsymbol{c}}_2 = \boldsymbol{c}_2 \boldsymbol{T}_4 \\ \overline{\boldsymbol{W}}^h &= \boldsymbol{T}_1^T \boldsymbol{W}^h \boldsymbol{T}_1, \qquad \overline{\boldsymbol{W}}^v = \boldsymbol{T}_4^T \boldsymbol{W}^v \boldsymbol{T}_4 \\ \overline{\boldsymbol{K}}^h &= \boldsymbol{T}_1^{-1} \boldsymbol{K}^h \boldsymbol{T}_1^{-T}, \quad \overline{\boldsymbol{K}}^v = \boldsymbol{T}_4^{-1} \boldsymbol{K}^v \boldsymbol{T}_4^{-T}. \end{aligned} \tag{10}$$

The L_2-sensitivity of the new realization is changed to

$$\begin{aligned} \overline{S} &= \sum_{i=0}^{n} \sigma_i^v \mathrm{tr}\left[\boldsymbol{T}_1^{-1} \boldsymbol{W}_i^h(\boldsymbol{T}_1) \boldsymbol{T}_1^{-T}\right] \\ &\quad + \sum_{j=0}^{m} \sigma_j^h \mathrm{tr}\left[\boldsymbol{T}_4^T \boldsymbol{K}_j^v(\boldsymbol{T}_4) \boldsymbol{T}_4\right] \\ &\quad + \mathrm{tr}\left[\overline{\boldsymbol{W}}^h + \overline{\boldsymbol{W}}^v + \overline{\boldsymbol{K}}^h + \overline{\boldsymbol{K}}^v\right] \\ &\quad + \mathrm{tr}\left[\overline{\boldsymbol{W}}^h\right] \mathrm{tr}\left[\overline{\boldsymbol{K}}^v\right]. \end{aligned} \tag{11}$$

Here $\boldsymbol{W}_i^h(\boldsymbol{T}_1)$ and $\boldsymbol{K}_j^v(\boldsymbol{T}_4)$ can be obtained by solving the Lyapunov equations

$$\begin{aligned} \begin{bmatrix} \boldsymbol{W}_i^h(\boldsymbol{T}_1) & * \\ * & * \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_1 & \tilde{\boldsymbol{u}}_i \boldsymbol{c}_1 \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_i^h(\boldsymbol{T}_1) & * \\ * & * \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \boldsymbol{A}_1 & \tilde{\boldsymbol{u}}_i \boldsymbol{c}_1 \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix}^T + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_1 \boldsymbol{T}_1^T \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{K}_j^v(\boldsymbol{T}_4) & * \\ * & * \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_4 & \boldsymbol{0} \\ \boldsymbol{b}_2 \tilde{\boldsymbol{v}}_j^T & \boldsymbol{A}_4 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{K}_j^v(\boldsymbol{T}_4) & * \\ * & * \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \boldsymbol{A}_4 & \boldsymbol{0} \\ \boldsymbol{b}_2 \tilde{\boldsymbol{v}}_j^T & \boldsymbol{A}_4 \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_4^{-T} \boldsymbol{T}_4^{-1} \end{bmatrix} \end{aligned}$$

Concerning the constraints for the minimization problem at hand, if the L_2-norm dynamic-range scaling constraints are imposed on the new state-variable vector

$$\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\boldsymbol{x}}^h(i,j) \\ \overline{\boldsymbol{x}}^v(i,j) \end{bmatrix} \tag{12}$$

then it is required that for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$

$$\begin{aligned} (\overline{\boldsymbol{K}}^h)_{jj} &= (\boldsymbol{T}_1^{-1} \boldsymbol{K}^h \boldsymbol{T}_1^{-T})_{jj} = 1 \\ (\overline{\boldsymbol{K}}^v)_{ii} &= (\boldsymbol{T}_4^{-1} \boldsymbol{K}^v \boldsymbol{T}_4^{-T})_{ii} = 1 \end{aligned} \tag{13}$$

The problem of L_2-sensitivity minimization subject to L_2-norm dynamic-range scaling constraints is now formulated as follows: *Given matrices* $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_4$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{c}_1$

and $\boldsymbol{c}_2$, obtain an $m \times m$ nonsingular matrix $\boldsymbol{T}_1$ and an $n \times n$ nonsingular matrix $\boldsymbol{T}_4$, which minimizes the sensitivity measure in (11) subject to the scaling constraints in (13).

Since the LSS model in (1) is assumed to be asymptotically stable, separately locally controllable and separately locally observable, the horizontal and vertical local controllability Gramians, $\boldsymbol{K}^h$ and $\boldsymbol{K}^v$ respectively, are symmetric and positive-definite. This implies that $(\boldsymbol{K}^h)^{1/2}$ and $(\boldsymbol{K}^v)^{1/2}$ satisfying

$$\boldsymbol{K}^h = (\boldsymbol{K}^h)^{1/2}(\boldsymbol{K}^h)^{1/2}$$
$$\boldsymbol{K}^v = (\boldsymbol{K}^v)^{1/2}(\boldsymbol{K}^v)^{1/2}$$

are also symmetric and positive-definite. Defining

$$\hat{\boldsymbol{T}}_1 = \boldsymbol{T}_1^T(\boldsymbol{K}^h)^{-1/2}, \ \hat{\boldsymbol{T}}_4 = \boldsymbol{T}_4^T(\boldsymbol{K}^v)^{-1/2} \tag{14}$$

the scaling constraints in (13) can be expressed as

$$\begin{aligned} (\hat{\boldsymbol{T}}_1^{-T}\hat{\boldsymbol{T}}_1^{-1})_{jj} &= 1, \ j = 1, 2, \cdots, m. \\ (\hat{\boldsymbol{T}}_4^{-T}\hat{\boldsymbol{T}}_4^{-1})_{ii} &= 1, \ i = 1, 2, \cdots, n. \end{aligned} \tag{15}$$

The constraints in (15) simply state that each column in $\hat{\boldsymbol{T}}_1^{-1}$ and $\hat{\boldsymbol{T}}_4^{-1}$ must be a unity vector. If matrices $\hat{\boldsymbol{T}}_1^{-1}$ and $\hat{\boldsymbol{T}}_4^{-1}$ are assumed to have the form

$$\begin{aligned} \hat{\boldsymbol{T}}_1^{-1} &= \left[\frac{\boldsymbol{t}_1^{(1)}}{\|\boldsymbol{t}_1^{(1)}\|}, \frac{\boldsymbol{t}_2^{(1)}}{\|\boldsymbol{t}_2^{(1)}\|}, \cdots, \frac{\boldsymbol{t}_m^{(1)}}{\|\boldsymbol{t}_m^{(1)}\|}\right] \\ \hat{\boldsymbol{T}}_4^{-1} &= \left[\frac{\boldsymbol{t}_1^{(4)}}{\|\boldsymbol{t}_1^{(4)}\|}, \frac{\boldsymbol{t}_2^{(4)}}{\|\boldsymbol{t}_2^{(4)}\|}, \cdots, \frac{\boldsymbol{t}_n^{(4)}}{\|\boldsymbol{t}_n^{(4)}\|}\right] \end{aligned} \tag{16}$$

then (15) is always satisfied. Using the coordinate transformations $\hat{\boldsymbol{T}}_1$ and $\hat{\boldsymbol{T}}_4$ in (14), with $\hat{\boldsymbol{T}}_1^{-1}$ and $\hat{\boldsymbol{T}}_4^{-1}$ specified in (16), the L_2-sensitivity measure in (11) becomes a function of matrices $\hat{\boldsymbol{T}}_1$ and $\hat{\boldsymbol{T}}_4$. If we denote this function by $J_o(\hat{\boldsymbol{T}})$, then it follows from (11) and (14) that

$$\begin{aligned} J_o(\hat{\boldsymbol{T}}) &= \sum_{i=0}^{n} \sigma_i^v \mathrm{tr}\left[\hat{\boldsymbol{T}}_1^{-T}\hat{\boldsymbol{W}}_i^h(\hat{\boldsymbol{T}}_1)\hat{\boldsymbol{T}}_1^{-1}\right] \\ &+ \sum_{j=0}^{m} \sigma_j^h \mathrm{tr}\left[\hat{\boldsymbol{T}}_4\hat{\boldsymbol{K}}_j^v(\hat{\boldsymbol{T}}_4)\hat{\boldsymbol{T}}_4^T\right] \\ &+ \mathrm{tr}\left[\hat{\boldsymbol{T}}_1\hat{\boldsymbol{W}}^h\hat{\boldsymbol{T}}_1^T + \hat{\boldsymbol{T}}_4\hat{\boldsymbol{W}}^v\hat{\boldsymbol{T}}_4^T\right] + m + n \\ &+ n \ \mathrm{tr}\left[\hat{\boldsymbol{T}}_1\hat{\boldsymbol{W}}^h\hat{\boldsymbol{T}}_1^T\right] \end{aligned} \tag{17}$$

where

$$\begin{aligned} \hat{\boldsymbol{W}}_i^h(\hat{\boldsymbol{T}}_1) &= \sum_{k=0}^{\infty} \hat{\boldsymbol{H}}_i^h(k)\hat{\boldsymbol{T}}_1^T\hat{\boldsymbol{T}}_1\hat{\boldsymbol{H}}_i^h(k)^T \\ \hat{\boldsymbol{K}}_j^v(\hat{\boldsymbol{T}}_4) &= \sum_{k=0}^{\infty} \hat{\boldsymbol{H}}_j^v(k)^T\hat{\boldsymbol{T}}_4^{-1}\hat{\boldsymbol{T}}_4^{-T}\hat{\boldsymbol{H}}_j^v(k) \\ \hat{\boldsymbol{H}}_i^h(k) &= (\boldsymbol{K}^h)^{-1/2}\boldsymbol{H}_i^h(k)(\boldsymbol{K}^h)^{1/2} \\ \boldsymbol{H}_i^h(k) &= \sum_{p=0}^{k} \boldsymbol{A}_1^p\tilde{\boldsymbol{u}}_i\boldsymbol{c}_1\boldsymbol{A}_1^{k-p} \\ \hat{\boldsymbol{H}}_j^v(k) &= (\boldsymbol{K}^v)^{-1/2}\boldsymbol{H}_j^v(k)(\boldsymbol{K}^v)^{1/2} \\ \boldsymbol{H}_j^v(k) &= \sum_{p=0}^{k} \boldsymbol{A}_4^p\boldsymbol{b}_2\tilde{\boldsymbol{v}_j^T}\boldsymbol{A}_4^{k-p} \\ \hat{\boldsymbol{W}}^h &= (\boldsymbol{K}^h)^{1/2}\boldsymbol{W}^h(\boldsymbol{K}^h)^{1/2} \\ \hat{\boldsymbol{W}}^v &= (\boldsymbol{K}^v)^{1/2}\boldsymbol{W}^v(\boldsymbol{K}^v)^{1/2} \end{aligned}$$

From the foregoing arguments, we can see that the problem of obtaining an $m \times m$ nonsingular matrix $\boldsymbol{T}_1$ and an $n \times n$ nonsingular matrix $\boldsymbol{T}_4$, which minimizes (11) subject to the scaling constraints in (15) can be converted into an unconstrained optimization problem of obtaining an $m \times m$ nonsingular matrix $\hat{\boldsymbol{T}}_1$ and an $n \times n$ nonsingular matrix $\hat{\boldsymbol{T}}_4$, which minimizes (17). To this end, we apply a quasi-Newton algorithm [21] to minimize (17) with respect to the matrices $\hat{\boldsymbol{T}}_1$ and $\hat{\boldsymbol{T}}_4$ given by (16). Let $\boldsymbol{x}$ be the column vector that collects the variables in matrices $\hat{\boldsymbol{T}}_1$ and $\hat{\boldsymbol{T}}_4$. Then $J_o(\hat{\boldsymbol{T}})$ is a function of $\boldsymbol{x}$, which we denote by $J(\boldsymbol{x})$. The optimization algorithm starts with a trivial initial point $\boldsymbol{x}_0$ obtained from an initial assignment $\hat{\boldsymbol{T}}_1 = \boldsymbol{I}_m$, $\hat{\boldsymbol{T}}_4 = \boldsymbol{I}_n$ Then, in the kth iteration a quasi-Newton algorithm updates the most recent point $\boldsymbol{x}_k$ to point $\boldsymbol{x}_{k+1}$ as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k \tag{18}$$

where

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

Here, $\nabla J(\boldsymbol{x})$ is the gradient of $J(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, and $\boldsymbol{S}_k$ is a positive-definite approximation of the inverse Hessian matrix of $J(\boldsymbol{x})$. The iteration process continues until

$$|J(\boldsymbol{x}_{k+1}) - J(\boldsymbol{x}_k)| < \varepsilon \tag{19}$$

where $\varepsilon > 0$ is a prescribed tolerance. If the iteration is terminated at step k, then $\boldsymbol{x}_k$ is taken to be the solution point.

4 Numerical Example

Consider a 2-D separable-denominator state-space digital filter specified by

$$\begin{aligned}
\boldsymbol{A}_{1o} &= \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ 0.599655 & -1.836929 & 2.173645 \end{bmatrix} \\
\boldsymbol{A}_{2o} &= \begin{bmatrix} 0.064564 & 0.033034 & 0.012881 \\ 0.091213 & 0.110512 & 0.102759 \\ 0.097256 & 0.151864 & 0.172460 \end{bmatrix} \\
\boldsymbol{A}_{4o} &= \begin{bmatrix} 0.0 & 0.0 & 0.564961 \\ 1.0 & 0.0 & -1.887939 \\ 0.0 & 1.0 & 2.280029 \end{bmatrix} \\
\boldsymbol{b}_{1o} &= \begin{bmatrix} 0.047053 & 0.062274 & 0.060436 \end{bmatrix}^T \\
\boldsymbol{b}_{2o} &= \begin{bmatrix} 1.0 & 0.0 & 0.0 \end{bmatrix}^T \\
\boldsymbol{c}_{1o} &= \begin{bmatrix} 1.0 & 0.0 & 0.0 \end{bmatrix} \\
\boldsymbol{c}_{2o} &= \begin{bmatrix} 0.016556 & 0.012550 & 0.008243 \end{bmatrix} \\
d &= 0.019421.
\end{aligned}$$

By computing all the Gramians from (7) and (8), the L_2-sensitivity measure S_o in (7) is found to be

$$S_o = 2.423893 \times 10^4.$$

To perform scaling so that (13) is satisfied, we apply a coordinate transformation matrix $\boldsymbol{T}_s = \boldsymbol{T}_{1s} \oplus \boldsymbol{T}_{4s}$ to the original linear system where $(\boldsymbol{T}_{1s})_{jj} = \sqrt{(\boldsymbol{K}^h)_{jj}}$ and $(\boldsymbol{T}_{4s})_{ii} = \sqrt{(\boldsymbol{K}^v)_{ii}}$. The L_2-sensitivity measure S is then found to be

$$S = 4.526079 \times 10^3.$$

When applying the quasi-Newton algorithm in (18) to the scaled realization, the L_2-sensitivity profile of the first 30 iterations is given in Table 1 and Fig. 1. From Table 1 and Fig. 1, it is observed that the algorithm practically converges with 19 iterations. After 19 iterations, matrices $\hat{\boldsymbol{T}}_1$ and $\hat{\boldsymbol{T}}_4$ are obtained as

$$\begin{aligned}
\hat{\boldsymbol{T}}_1 &= \begin{bmatrix} 1.124484 & -0.280394 & 0.255922 \\ -0.421171 & 1.025558 & 0.330664 \\ -0.529598 & -0.449919 & 0.764121 \end{bmatrix} \\
\hat{\boldsymbol{T}}_4 &= \begin{bmatrix} 1.354786 & -0.270031 & -0.026251 \\ -0.641918 & 1.123431 & 0.288145 \\ -0.366800 & -0.362196 & 0.924809 \end{bmatrix}.
\end{aligned}$$

or equivalently, matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_4$ are found as

$$\begin{aligned}
\boldsymbol{T}_1 &= \begin{bmatrix} 0.834682 & 0.277258 & -0.467950 \\ 0.536765 & 0.639090 & -0.165736 \\ 0.345994 & 0.706357 & 0.236653 \end{bmatrix} \\
\boldsymbol{T}_4 &= \begin{bmatrix} 1.113890 & -0.921189 & 0.336130 \\ -0.869618 & 0.928016 & -0.545490 \\ 0.733270 & -0.682433 & 0.691443 \end{bmatrix}.
\end{aligned}$$

The L_2-sensitivity measure $\overline{S}$ in is then found to be

$$\overline{S} = 1.010064 \times 10^2.$$

Table 1: L_2-sensitivity profile of first 30 iterations.

k	L_2-sensitivity	k	L_2-sensitivity
	2.423893×10^4	10	1.012574×10^2
0	4.526079×10^3	11	1.011758×10^2
1	1.232246×10^2	12	1.010602×10^2
2	1.112628×10^2	13	1.010325×10^2
3	1.082960×10^2	14	1.010271×10^2
4	1.046686×10^2	15	1.010148×10^2
5	1.039913×10^2	16	1.010109×10^2
6	1.032931×10^2	17	1.010074×10^2
7	1.019785×10^2	18	1.010065×10^2
8	1.014876×10^2	19	1.010064×10^2
9	1.013174×10^2	20	1.010064×10^2

5 Conclusion

The problem of minimizing an L_2-sensitivity measure subject to L_2-norm dynamic range scaling constraints has been investigated for 2-D separable-denominator digital filters. An efficient method has been developed by using the quasi-Newton algorithm and some matrix-theoretic techniques to develop a closed-form solution for the optimization problem. Our computer simulation results have demonstrated the effectiveness of the proposed technique.

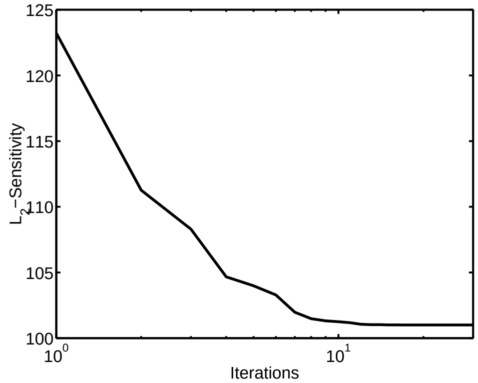

Figure 1: L_2-sensitivity profile of first 30 iterations.

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] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

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

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

[20] 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.

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