REDUCTION OF L_2-SENSITIVITY FOR THREE-DIMENSIONAL SEPARABLE-DENOMINATOR DIGITAL FILTERS

Takao Hinamoto, Osamu Tanaka and Masayoshi Nakamoto

Graduate School of Engineering, Hiroshima University
Higashi-Hiroshima 739-8527, Japan
phone:+81-82-424-7672, fax:+81-82-422-7195
email:hinamoto@ieee.org

Wu-Sheng Lu

Dept. of Elect. & Comput. Engineering,
University of Victoria
Victoria, B.C., V8W 3P6, Canada
phone:+1-250-721-8692, fax:+1-250-721-6052
email: wslu@ece.uvic.ca

ABSTRACT

The problem of reducing the deviation from a desired transfer function caused by the coefficient quantization errors is investigated for a three-dimensional (3-D) separable in denominator digital filter. To begin with, a 3-D transfer function with separable denominator is represented with the cascade connection of three one-dimensional (1-D) transfer functions by applying a minimal decomposition technique, and the multi-input multi-output (MIMO) 1-D transfer function located in the middle of the cascade connection is realized by a minimal state-space model. Next, the l_2-sensitivity of the state-space model is analyzed, and the minimization problem of the l_2-sensitivity subject to l_2-scaling constraints is formulated. This problem is then converted into an unconstrained optimization problem by using linear-algebraic techniques, and an efficient quasi-Newton algorithm is applied to solve it. A numerical example is presented to illustrate the validity and effectiveness of the proposed technique.

1. INTRODUCTION

It is of practical significance in many applications to construct a filter structure such that the coefficient sensitivity of a digital filter is minimum or nearly minimum in a certain sense. Due to finite word length (FWL) effects caused by coefficient truncation or rounding, poor sensitivity may lead to degradation of the transfer characteristics in a FWL implementation of the digital filter. For instance, the characteristics of an originally stable filter might be so altered that the filter may become unstable. This motivates the study of the coefficient sensitivity minimization problem for digital filters. The minimization of the coefficient sensitivity for two-dimensional (2-D) state-space digital filters has been investigated extensively [1]-[7]. Zilouchian and Carroll have explored a coefficient sensitivity bound in 2-D state-space digital filters [1]. Several techniques have been proposed for synthesizing 2-D state-space filter structures with minimum coefficient sensitivity [2]-[7]. Some of these techniques evaluate the sensitivity by using a mixture of l_1/l_2 norms [1]-[16], while the others rely on the use of a pure l_2 norm [5]-[7]. In [6], the weighted-sensitivity minimization of 2-D state-space digital filters has been considered in both cases of a mixture of L_1/L_2 norms and a pure L_2 norm. Note that the l_2-sensitivity minimization is more natural and reasonable than the conventional l_1/l_2 mixed sensitivity minimization, but it is technically more challenging [7]. A technique has also been presented for constructing three-dimensional (3-D) separable-denominator (SD) state-space digital filters that minimize the l_2-sensitivity [8]. More recently, the minimization problem of l_2-sensitivity subject to l_2-scaling constraints has been treated for 2-D state-space digital filters [9],[10]. It is known that the use of scaling constraints can be beneficial for suppressing overflow [11]. In addition, considerable research interest has been observed in the design of multidimensional recursive digital filters [12]-[16]. This is partially due to the potential applications in such fields as video processing, seismic signal processing, magnetic data processing, and biomedical tomography, etc.

This paper investigates the realization problem of a 3-D SD digital filter which reduces the l_2-sensitivity subject to l_2-scaling constraints. A 3-D transfer function with separable denominator is decomposed into three one-dimensional (1-D) transfer functions with a cascade connection, and the MIMO 1-D transfer function in the middle of the cascade connection is described by minimal state-space realization. Then, the minimization problem of l_2-sensitivity subject to l_2-scaling constraints for the state-space model is converted into an unconstrained optimization formulation, and an iterative technique is developed for constructing the optimal state-space model of the MIMO 1-D system. A numerical example is presented to illustrate the validity and effectiveness of the proposed technique.

2. PROBLEM STATEMENT

Consider a stable 3-D separable-denominator digital filter is described by

$$H(z_1,z_2,z_3)=\frac{N(z_1,z_2,z_3)}{D_1(z_1)D_2(z_2)D_3(z_3)} \tag{1}$$

where

$$\begin{aligned}
N(z_1,z_2,z_3)&=\sum_{i=0}^{N_1}\sum_{j=0}^{N_2}\sum_{k=0}^{N_3}a_{ijk}z_1^{-i}z_2^{-j}z_3^{-k}\\
D_1(z_1)&=1+b_{11}z_1^{-1}+\cdots+b_{1N_1}z_1^{-N_1}\\
D_2(z_2)&=1+b_{21}z_2^{-1}+\cdots+b_{2N_2}z_2^{-N_2}\\
D_3(z_3)&=1+b_{31}z_3^{-1}+\cdots+b_{3N_3}z_3^{-N_3}.
\end{aligned}$$

The 3-D transfer function in (1) can be decomposed into three 1-D systems as

$$H(z_1,z_2,z_3)=\frac{\boldsymbol{Z}_1^T}{D_1(z_1)}\boldsymbol{H}_2(z_2)\frac{\boldsymbol{Z}_3}{D_3(z_3)} \tag{2}$$

where

$$\begin{aligned}
\boldsymbol{Z}_1 &= (1, z_1^{-1},\cdots,z_1^{-N_1})^T\\
\boldsymbol{Z}_3 &= (1, z_3^{-1},\cdots,z_3^{-N_3})^T\\
\boldsymbol{H}_2(z_2) &= \frac{\boldsymbol{\Delta}_0+\boldsymbol{\Delta}_1 z_2^{-1}+\cdots+\boldsymbol{\Delta}_{N_2} z_2^{-N_2}}{D_2(z_2)}\\
\boldsymbol{\Delta}_m &= \begin{bmatrix} a_{0m0} & a_{0m1} & \cdots & a_{0mN_3}\\ a_{1m0} & a_{1m1} & \cdots & a_{1mN_3}\\ \vdots & \vdots & \ddots & \vdots\\ a_{N_1m0} & a_{N_1m1} & \cdots & a_{N_1mN_3}\end{bmatrix}\\
& m=0,1,\cdots,N_2.
\end{aligned}$$

The above 1-D transfer function $\boldsymbol{H}_2(z_2)$ with (N_3+1) inputs and (N_1+1) outputs can be realized by the minimal state-space model $(\boldsymbol{A}_2,\boldsymbol{B}_2,\boldsymbol{C}_2,\boldsymbol{\Delta}_0)_p$:

$$\begin{aligned}\boldsymbol{x}(k+1) &= \boldsymbol{A}_2\boldsymbol{x}(k)+\boldsymbol{B}_2\boldsymbol{u}(k)\\ \boldsymbol{y}(k) &= \boldsymbol{C}_2\boldsymbol{x}(k)+\boldsymbol{\Delta}_0\boldsymbol{u}(k)\end{aligned} \tag{3}$$

where $\boldsymbol{x}(k)$ is a $p\times 1$ state-variable vector, $\boldsymbol{u}(k)$ is an $(N_3+1)\times 1$ input vector, $\boldsymbol{y}(k)$ is an $(N_1+1)\times 1$ output vector, and $\boldsymbol{A}_2,\boldsymbol{B}_2,\boldsymbol{C}_2$ and $\boldsymbol{\Delta}_0$ are real constant matrices of appropriate dimensions. The transfer function of the linear system in (3) is given by

$$\boldsymbol{H}_2(z_2)=\boldsymbol{C}_2(z_2\boldsymbol{I}_p-\boldsymbol{A}_2)^{-1}\boldsymbol{B}_2+\boldsymbol{\Delta}_0. \tag{4}$$

The l_2-sensitivities of a 3-D digital filter in (2) with respect to coefficient matrices $\boldsymbol{A}_2$, $\boldsymbol{B}_2$, and $\boldsymbol{C}_2$ are 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 $f(\boldsymbol{X})$ with respect to $\boldsymbol{X}$ is then defined as

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}=\begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial x_{11}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}}\\ \frac{\partial f(\boldsymbol{X})}{\partial x_{21}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}} & \frac{\partial f(\boldsymbol{X})}{\partial x_{m2}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}}\end{bmatrix} \tag{5}$$

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

Definition 2: Let $\boldsymbol{X}(z_1,z_2,z_3)$ be an $m\times n$ complex matrix-valued function of complex variables z_1, z_2 and z_3. Let $x_{pq}(z_1,z_2,z_3)$ be the (p,q)th entry of $\boldsymbol{X}(z_1,z_2,z_3)$. Then the l_2-norm of $\boldsymbol{X}(z_1,z_2,z_3)$ is defined as

$$\begin{aligned}&\|\boldsymbol{X}(z_1,z_2,z_3)\|_2\\ &=\left(\mathrm{tr}\left[\frac{1}{(2\pi j)^3}\oint_{|z_1|=1}\oint_{|z_2|=1}\oint_{|z_3|=1}\boldsymbol{X}(z_1,z_2,z_3)\right.\right.\\ &\qquad\left.\left.\cdot\boldsymbol{X}^*(z_1,z_2,z_3)\frac{dz_1}{z_1}\frac{dz_2}{z_2}\frac{dz_3}{z_3}\right]\right)^{\frac{1}{2}}.\end{aligned} \tag{6}$$

From (2), (4) and Definitions 1 and 2, the l_2-sensitivity of the transfer function $H(z_1,z_2,z_3)$ with respect to the coefficient matrices $\boldsymbol{A}_2$, $\boldsymbol{B}_2$, and $\boldsymbol{C}_2$ is evaluated by

$$\begin{aligned}S &= \left\|\frac{\partial H(z_1,z_2,z_3)}{\partial \boldsymbol{A}_2}\right\|_2^2+\left\|\frac{\partial H(z_1,z_2,z_3)}{\partial \boldsymbol{B}_2}\right\|_2^2\\ &\quad+\left\|\frac{\partial H(z_1,z_2,z_3)}{\partial \boldsymbol{C}_2}\right\|_2^2\\ &=\left\|\boldsymbol{g}(z_1,z_2)^T\boldsymbol{f}(z_2,z_3)^T\right\|_2^2+\left\|\boldsymbol{g}(z_1,z_2)^T\frac{\boldsymbol{Z}_3^T}{D_3(z_3)}\right\|_2^2\\ &\quad+\left\|\frac{\boldsymbol{Z}_1}{D_1(z_1)}\boldsymbol{f}(z_2,z_3)^T\right\|_2^2\end{aligned} \tag{7}$$

where

$$\begin{aligned}\boldsymbol{f}(z_2,z_3) &= (z_2\boldsymbol{I}_p-\boldsymbol{A}_2)^{-1}\boldsymbol{B}_2\frac{\boldsymbol{Z}_3}{D_3(z_3)}\\ \boldsymbol{g}(z_1,z_2) &= \frac{\boldsymbol{Z}_1^T}{D_1(z_1)}\boldsymbol{C}_2(z_2\boldsymbol{I}_p-\mathbf{A}_2)^{-1}\\ \frac{\boldsymbol{Z}_1^T}{D_1(z_1)} &= \boldsymbol{c}_1(z_1\boldsymbol{I}_{N_1}-\boldsymbol{A}_1)^{-1}\boldsymbol{B}_1+\boldsymbol{d}_1\\ \frac{\boldsymbol{Z}_3}{D_3(z_3)} &= \boldsymbol{C}_3(z_3\boldsymbol{I}_{N_3}-\boldsymbol{A}_3)^{-1}\boldsymbol{b}_3+\boldsymbol{d}_3.\end{aligned}$$

The l_2-sensitivity measure in (7) can be written as

$$S=\mathrm{tr}[\boldsymbol{M}_A(\boldsymbol{I}_p)]+\mathrm{tr}[\boldsymbol{W}_B]+\mathrm{tr}[\boldsymbol{K}_C] \tag{8}$$

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

$$\begin{aligned}\boldsymbol{X} &= \frac{1}{(2\pi j)^3}\oint_{|z_1|=1}\oint_{|z_2|=1}\oint_{|z_3|=1}\boldsymbol{Y}(z_1,z_2,z_3)\\ &\qquad\cdot\boldsymbol{Y}^*(z_1,z_2,z_3)\frac{dz_1}{z_1}\frac{dz_2}{z_2}\frac{dz_3}{z_3}\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{Y}(z_1,z_2,z_3) &= \boldsymbol{g}(z_1,z_2)^T\boldsymbol{f}(z_2,z_3)^T \quad \text{for} \quad \boldsymbol{X}=\boldsymbol{M}_A(\boldsymbol{I}_p)\\ \boldsymbol{Y}(z_1,z_2,z_3) &= \boldsymbol{g}(z_1,z_2)^*\frac{\boldsymbol{Z}_3^*}{D_3(\overline{z}_3)} \quad \text{for} \quad \boldsymbol{X}=\boldsymbol{W}_B\\ \boldsymbol{Y}(z_1,z_2,z_3) &= \frac{\boldsymbol{Z}_1}{D_1(z_1)}\boldsymbol{f}(z_2,z_3)^T \quad \text{for} \quad \boldsymbol{X}=\boldsymbol{K}_C.\end{aligned}$$

The Gramians $\boldsymbol{M}_A(\boldsymbol{P})$, $\boldsymbol{W}_B$, and $\boldsymbol{K}_C$ can be computed using

$$\begin{aligned}\boldsymbol{M}_A(\boldsymbol{P}) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}[\mathbf{0}\quad \boldsymbol{I}_p]\begin{bmatrix}\boldsymbol{A}_2^T & \mathbf{0}\\ \boldsymbol{C}_2^T\boldsymbol{R}_{ij}^T\boldsymbol{B}_2^T & \boldsymbol{A}_2^T\end{bmatrix}^k\\ &\quad\cdot\begin{bmatrix}\boldsymbol{P}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_2 & \boldsymbol{B}_2\boldsymbol{R}_{ij}\boldsymbol{C}_2\\ \mathbf{0} & \boldsymbol{A}_2\end{bmatrix}^k\begin{bmatrix}\mathbf{0}\\ \boldsymbol{I}_p\end{bmatrix}\\ \boldsymbol{W}_B &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(\boldsymbol{A}_2^T)^k\boldsymbol{C}_2^T\boldsymbol{R}_{ij}^T\boldsymbol{R}_{ij}\boldsymbol{C}_2\boldsymbol{A}_2^k\\ \boldsymbol{K}_C &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\boldsymbol{A}_2^k\boldsymbol{B}_2\boldsymbol{R}_{ij}\boldsymbol{R}_{ij}^T\boldsymbol{B}_2^T(\boldsymbol{A}_2^T)^k\end{aligned} \tag{9}$$

where

$$\frac{\boldsymbol{Z}_3\boldsymbol{Z}_1^T}{D_3(z_3)D_1(z_1)} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{R}_{ij}z_1^{-i}z_3^{-j}$$
$$\frac{\boldsymbol{Z}_1^T}{D_1(z_1)} = \sum_{i=1}^{\infty}\boldsymbol{c}_1\boldsymbol{A}_1^{i-1}\boldsymbol{B}_1 z_1^{-i} + \boldsymbol{d}_1$$
$$\frac{\boldsymbol{Z}_3}{D_3(z_3)} = \sum_{j=1}^{\infty}\boldsymbol{C}_3\boldsymbol{A}_3^{i-1}\boldsymbol{b}_3 z_3^{-j} + \boldsymbol{d}_3.$$

Applying a coordinate transformation defined by $\overline{\boldsymbol{x}}(k) = \boldsymbol{T}^{-1}\boldsymbol{x}(k)$ for the 1-D system $(\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{C}_2, \boldsymbol{\Delta}_0)_p$ in (3), we obtain a new realization $(\overline{\boldsymbol{A}}_2, \overline{\boldsymbol{B}}_2, \overline{\boldsymbol{C}}_2, \boldsymbol{\Delta}_0)_p$ characterized by

$$\overline{\boldsymbol{A}}_2 = \boldsymbol{T}^{-1}\boldsymbol{A}_2\boldsymbol{T}, \quad \overline{\boldsymbol{B}}_2 = \boldsymbol{T}^{-1}\boldsymbol{B}_2, \quad \overline{\boldsymbol{C}}_2 = \boldsymbol{C}_2\boldsymbol{T}. \tag{10}$$

For the new realization, the l_2-sensitivity measure in (8) is changed to

$$\begin{aligned} S(\boldsymbol{T}) = \ & \mathrm{tr}[\boldsymbol{T}^T\boldsymbol{M}_A(\boldsymbol{T}\boldsymbol{T}^T)\boldsymbol{T}] + \mathrm{tr}[\boldsymbol{T}^T\boldsymbol{W}_B\boldsymbol{T}] \\ & + \mathrm{tr}[\boldsymbol{T}^{-1}\boldsymbol{K}_C\boldsymbol{T}^{-T}]. \end{aligned} \tag{11}$$

Noting that $\boldsymbol{f}(z_2, z_3)$ is the transfer function from the filter input to the state-variable vector $\boldsymbol{x}(k)$, a controllability Gramian $\boldsymbol{K}$ can be derived from

$$\begin{aligned} \boldsymbol{K} &= \frac{1}{(2\pi j)^2}\oint_{|z_2|=1}\oint_{|z_3|=1}\boldsymbol{f}(z_2,z_3)\boldsymbol{f}^*(z_2,z_3)\frac{dz_2}{z_2}\frac{dz_3}{z_3} \\ &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{A}_2^k\boldsymbol{B}_2\boldsymbol{r}_j\boldsymbol{r}_j^T\boldsymbol{B}_2^T(\boldsymbol{A}_2^T)^k \end{aligned} \tag{12}$$

where

$$\frac{\boldsymbol{Z}_3}{D_3(z_3)} = \sum_{j=0}^{\infty}\boldsymbol{r}_j z_3^{-j}.$$

In this case, l_2-scaling constraints are given by

$$(\overline{\boldsymbol{K}})_{ii} = (\boldsymbol{T}^{-1}\boldsymbol{K}\boldsymbol{T}^{-T})_{ii} = 1 \quad \text{for} \quad i = 1, 2, \cdots, p. \tag{13}$$

The problem at hand can now be formulated as to obtain the coordinate transformation matrix $\boldsymbol{T}$ that minimizes $S(\boldsymbol{T})$ in (11) subject to the l_2-scaling constraints in (13).

3. L_2-SENSITIVITY MINIMIZATION

Defining

$$\hat{\boldsymbol{T}} = \boldsymbol{T}^T\boldsymbol{K}^{-\frac{1}{2}} \tag{14}$$

the l_2-scaling constraints in (13) can be written as

$$(\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{T}}^{-1})_{ii} = 1 \quad \text{for} \quad i = 1, 2, \cdots, p. \tag{15}$$

It is obvious that the conditions in (15) are always satisfied by choosing $\hat{\boldsymbol{T}}^{-1}$ as

$$\hat{\boldsymbol{T}}^{-1} = \left[\frac{\boldsymbol{t}_1}{||\boldsymbol{t}_1||}, \frac{\boldsymbol{t}_2}{||\boldsymbol{t}_2||}, \cdots, \frac{\boldsymbol{t}_p}{||\boldsymbol{t}_p||}\right]. \tag{16}$$

Substituting matrix $\boldsymbol{T}$ which satisfies (14) into $S(\boldsymbol{T})$ in (11), the l_2-sensitivity measure can be expressed as

$$\begin{aligned} J_o(\boldsymbol{x}) = \ & \mathrm{tr}[\hat{\boldsymbol{T}}\hat{\boldsymbol{M}}_A(\hat{\boldsymbol{T}})\hat{\boldsymbol{T}}^T] + \mathrm{tr}[\hat{\boldsymbol{T}}\hat{\boldsymbol{W}}_B\hat{\boldsymbol{T}}^T] \\ & + \mathrm{tr}[\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{K}}_C\hat{\boldsymbol{T}}^{-1}] \end{aligned} \tag{17}$$

where

$$\begin{aligned} \boldsymbol{x} &= (\boldsymbol{t}_1^T, \boldsymbol{t}_2^T, \cdots, \boldsymbol{t}_p^T)^T \\ \hat{\boldsymbol{M}}_A(\hat{\boldsymbol{T}}) &= \sum_{i=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\begin{bmatrix}\boldsymbol{0} & (\boldsymbol{K}^{\frac{1}{2}})^T\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_2^T & \boldsymbol{0} \\ \boldsymbol{C}_2^T\boldsymbol{R}_{ij}^T\boldsymbol{B}_2^T & \boldsymbol{A}_2^T\end{bmatrix}^k \\ & \cdot \begin{bmatrix}(\boldsymbol{K}^{-\frac{1}{2}})^T\hat{\boldsymbol{T}}^{-1}\hat{\boldsymbol{T}}^{-T}\boldsymbol{K}^{-\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \\ & \cdot \begin{bmatrix}\boldsymbol{A}_2 & \boldsymbol{B}_2\boldsymbol{R}_{ij}\boldsymbol{C}_2 \\ \boldsymbol{0} & \boldsymbol{A}_2\end{bmatrix}^k\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{K}^{\frac{1}{2}}\end{bmatrix} \\ \hat{\boldsymbol{W}}_B &= \boldsymbol{K}^{\frac{1}{2}}\boldsymbol{W}_B\boldsymbol{K}^{\frac{1}{2}} \\ \hat{\boldsymbol{K}}_C &= \boldsymbol{K}^{-\frac{1}{2}}\boldsymbol{K}_C\boldsymbol{K}^{-\frac{1}{2}}. \end{aligned}$$

This shows that the original constrained optimization problem can be converted into an unconstrained optimization problem of obtaining a $p^2 \times 1$ vector $\boldsymbol{x}$ which minimizes $J_o(\boldsymbol{x})$ in (17).

Applying a quasi-Newton algorithm to minimize $J_o(\boldsymbol{x})$ in (17), in the kth iteration the most recent point $\boldsymbol{x}_k$ is updated to point $\boldsymbol{x}_{k+1}$ as

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

where

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

In the above, $\nabla J_o(\boldsymbol{x})$ 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 $\hat{\boldsymbol{T}} = \boldsymbol{I}_p$, and this iteration process continues until

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

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

The implementation of (18) requires the gradient of $J_o(\boldsymbol{x})$, which can be efficiently evaluated using closed-form expressions as follows.

$$\begin{aligned} \frac{\partial J_o(\hat{\boldsymbol{T}})}{\partial t_{ij}} &= \lim_{\Delta\to 0}\frac{J_o(\hat{\boldsymbol{T}}_{ij}) - J_o(\hat{\boldsymbol{T}})}{\Delta} \\ &= 2\beta_1 - 2\beta_2 + 2\beta_3 - 2\beta_4 \end{aligned} \tag{20}$$

where

$$\hat{\boldsymbol{T}}_{ij} = \hat{\boldsymbol{T}} + \frac{\Delta\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}\boldsymbol{e}_j^T\hat{\boldsymbol{T}}}{1-\Delta\boldsymbol{e}_j^T\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}},\ \hat{\boldsymbol{T}}_{ij}^{-1} = \hat{\boldsymbol{T}}^{-1} - \Delta\boldsymbol{g}_{ij}\boldsymbol{e}_j^T$$

$$\boldsymbol{g}_{ij} = \partial\left\{\frac{\boldsymbol{t}_j}{||\boldsymbol{t}_j||}\right\}/\partial t_{ij} = \frac{1}{||\boldsymbol{t}_j||^3}(t_{ij}\boldsymbol{t}_j - ||\boldsymbol{t}_j||^2\boldsymbol{e}_i)$$

$$\beta_1 = \boldsymbol{e}_j^T\hat{\boldsymbol{T}}\hat{\boldsymbol{M}}_A(\hat{\boldsymbol{T}})\hat{\boldsymbol{T}}^T\hat{\boldsymbol{T}}\boldsymbol{g}_{ij},\ \beta_3 = \boldsymbol{e}_j^T\hat{\boldsymbol{T}}\hat{\boldsymbol{W}}_B\hat{\boldsymbol{T}}^T\hat{\boldsymbol{T}}\boldsymbol{g}_{ij}$$

$$\beta_2 = \boldsymbol{e}_j^T\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{N}}_A(\hat{\boldsymbol{T}})\boldsymbol{g}_{ij},\quad \beta_4 = \boldsymbol{e}_j^T\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{K}}_C\boldsymbol{g}_{ij}$$

$$\hat{\boldsymbol{N}}_A(\hat{\boldsymbol{T}}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\begin{bmatrix}(\boldsymbol{K}^{-\frac{1}{2}})^T & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_2 & \boldsymbol{C}_2\boldsymbol{R}_{ij}\boldsymbol{B}_2\\ \boldsymbol{0} & \boldsymbol{A}_2\end{bmatrix}^k$$

$$\cdot\begin{bmatrix}(\boldsymbol{K}^{\frac{1}{2}})\hat{\boldsymbol{T}}^T\hat{\boldsymbol{T}}(\boldsymbol{K}^{\frac{1}{2}})^T & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$$

$$\cdot\begin{bmatrix}\boldsymbol{A}_2^T & \boldsymbol{0}\\ \boldsymbol{B}_2^T\boldsymbol{R}_{ij}^T\boldsymbol{C}_2^T & \boldsymbol{A}_2^T\end{bmatrix}^k\begin{bmatrix}(\boldsymbol{K}^{-\frac{1}{2}})^T\\ \boldsymbol{0}\end{bmatrix}.$$

4. A NUMERICAL EXAMPLE

Consider a stable 3-D SD digital filter specified by

$$\boldsymbol{\Delta}_0 = 10^{-2}\begin{bmatrix}0.00730 & 0.34297 & -0.09594 & 0.20541\\ 3.33408 & -5.73707 & 3.94939 & -1.61598\\ -1.46081 & 2.66051 & -1.68094 & 0.68022\\ 1.12651 & -1.62192 & 1.24735 & -0.55781\end{bmatrix}$$

$$\boldsymbol{\Delta}_1 = 10^{-2}\begin{bmatrix}2.81318 & -5.00467 & 3.46926 & -0.84798\\ -5.29980 & 9.24831 & -6.29206 & 2.80791\\ 4.95232 & -8.39641 & 5.73329 & -1.62170\\ 0.72029 & -1.34272 & 0.95941 & 0.54827\end{bmatrix}$$

$$\boldsymbol{\Delta}_2 = 10^{-2}\begin{bmatrix}-0.69409 & 1.54874 & -0.94779 & 0.39116\\ 3.93785 & -6.79910 & 4.66564 & -1.96344\\ -2.37995 & 4.20737 & -2.75482 & 0.95329\\ 0.70545 & -0.90615 & 0.73168 & -0.55633\end{bmatrix}$$

$$\boldsymbol{\Delta}_3 = 10^{-2}\begin{bmatrix}1.67681 & -2.69078 & 1.98218 & -0.33567\\ -0.59937 & 1.11289 & -0.71981 & 0.43504\\ 1.82472 & -2.93685 & 2.11591 & -0.43417\\ 1.28875 & -2.01749 & 1.51782 & -0.09016\end{bmatrix}$$

$$\begin{aligned}[b_{11}\quad b_{12}\quad b_{13}] &= [b_{31}\quad b_{32}\quad b_{33}]\\ &= [-1.81600\quad 1.23756\quad -0.31382]\\ [b_{21}\quad b_{22}\quad b_{23}] &= [-1.81611\quad 1.23775\quad -0.31391].\end{aligned}$$

The above data can be realized by a minimal state-space model in (3) as

$$\boldsymbol{A}_2 = \begin{bmatrix}0.00000 & -0.19089 & 0.29060\\ 0.74393 & -86.40470 & 133.71075\\ -0.27211 & -57.01643 & 88.22081\end{bmatrix}$$

$$\boldsymbol{B}_2 = 10^3\begin{bmatrix}0.00602 & -0.00921 & 0.00699 & -0.00095\\ -1.10247 & 1.68622 & -1.27902 & 0.17267\\ -0.71455 & 1.09291 & -0.82977 & 0.11192\end{bmatrix}$$

$$\boldsymbol{C}_2 = \begin{bmatrix}0.07236 & 0.06711 & -0.10298\\ 0.01930 & 0.01789 & -0.02745\\ 0.05887 & 0.05460 & -0.08378\\ 0.07079 & 0.06565 & -0.10073\end{bmatrix}.$$

Equations (9) and (12) were used with truncation $(0,0,0) \le (i,j,k) \le (100,100,100)$ to evaluate the Gramians $\boldsymbol{M}_A$, $\boldsymbol{W}_B$, $\boldsymbol{K}_C$ and $\boldsymbol{K}$, and the coordinate transformation matrix $\boldsymbol{T}$ was chosen as

$$\boldsymbol{T} = 10^3\mathrm{diag}\{0.01077, 2.58588, 1.68384\}$$

which yields

$$\boldsymbol{K} = \begin{bmatrix}1.00000 & -0.84067 & -0.83915\\ -0.84067 & 1.00000 & 0.99999\\ -0.83915 & 0.99999 & 1.00000\end{bmatrix}$$

$$\boldsymbol{K}_C = 10\begin{bmatrix}5.70413 & -4.79529 & -4.78664\\ -4.79529 & 5.70413 & 5.70410\\ -4.78664 & 5.70410 & 5.70413\end{bmatrix}$$

$$\boldsymbol{W}_B = 10^8\begin{bmatrix}0.00014 & 0.02404 & -0.02399\\ 0.02404 & 4.36159 & -4.35410\\ -0.02399 & -4.35410 & 4.34663\end{bmatrix}$$

$$\boldsymbol{M}_A = 10^7\begin{bmatrix}0.00021 & 0.03478 & -0.03471\\ 0.03478 & 5.83540 & -5.82400\\ -0.03479 & -5.82400 & 5.81261\end{bmatrix}.$$

In this case, the l_2-sensitivity measure was computed as

$$S = 9.87318749 \times 10^8.$$

By choosing $\hat{\boldsymbol{T}} = \boldsymbol{I}_p$ as an initial estimate in (18) and a tolerance $\epsilon = 10^{-8}$ in (19), the quasi-Newton algorithm took 39 iterations to converge to the solution

$$\hat{\boldsymbol{T}}^{opt} = \begin{bmatrix}0.423192 & -2.435713 & 0.101976\\ 0.969196 & -0.865289 & 0.939063\\ -0.921737 & -1.670782 & 0.229632\end{bmatrix}$$

or equivalently,

$$\boldsymbol{T}^{opt} = \begin{bmatrix}1.274934 & 0.777636 & -0.194201\\ -1.683744 & -0.336138 & -0.571737\\ -1.680959 & -0.331102 & -0.572584\end{bmatrix}.$$

The l_2-sensitivity measure in (17) was then computed as

$$J_o(\hat{\boldsymbol{T}}^{opt}) = 3.243563304 \times 10^3$$

and the optimal state-space model in (10) was constructed as

$$\overline{\boldsymbol{A}}_2 = \begin{bmatrix}0.578361 & -0.146041 & -0.025781\\ 0.028916 & 0.626452 & 0.438734\\ -0.198149 & -0.319445 & 0.611297\end{bmatrix}$$

$$\overline{\boldsymbol{B}}_2 = \begin{bmatrix}0.283548 & -0.434959 & 0.405584 & -0.045694\\ 0.202161 & -0.306822 & 0.076466 & -0.029747\\ -0.208199 & 0.320789 & -0.374275 & 0.035264\end{bmatrix}$$

$$\overline{\boldsymbol{C}}_2 = \begin{bmatrix}0.279941 & -0.313437 & -0.082577\\ 0.068751 & -0.084653 & -0.024085\\ 0.217945 & -0.256949 & -0.070517\\ 0.246633 & -0.311955 & -0.090080\end{bmatrix}.$$

The profile of the l_2-sensitivity measure $J_o(\boldsymbol{x})$ during the first 39 iterations is shown in Fig. 1, from which it is observed that with a tolerance $\epsilon = 10^{-8}$ the algorithm converges within 39 iterations.

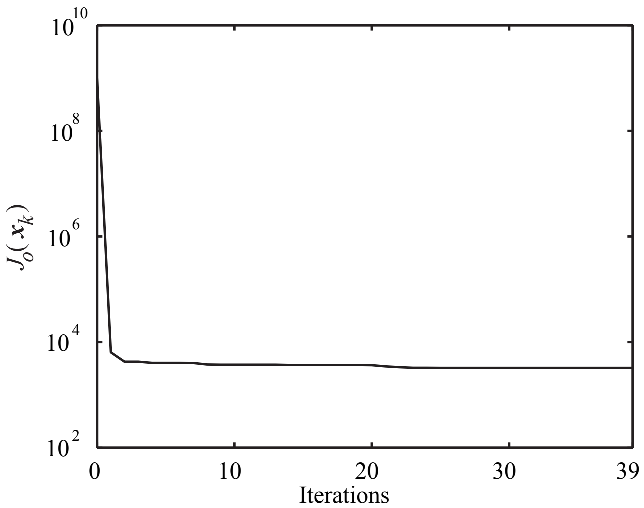

Figure 1: l_2-Sensitivity Performance.

5. CONCLUSION

We have investigated the problem of constructing a 3-D SD digital filter which reduces the l_2-sensitivity under l_2-scaling constraints. To this end, the constrained optimization problem has been converted into an unconstrained optimization formulation by using linear-algebraic techniques. An efficient quasi-Newton algorithm is then applied to solve the unconstrained optimization problem iteratively. Computer simulation results have demonstrated the validity and effectiveness of the proposed technique.

REFERENCES

[1] A. Zilouchian and R. L. Carroll, "A coefficient sensitivity bound in 2-D state-space digital filtering," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 665-667, June 1986.

[2] 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," *Proc. 1987 IEEE Int. Symp. Circuits Syst.*, Philadelphia, PA, USA, May 4-7, 1987, pp. 710-713.

[3] 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.

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

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

[6] 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.

[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] T. Hinamoto, Y. Sugie, A. Doi and M. Muneyasu, "Synthesis of 3-D separable-denominator state-space digital filters with minimum L_2-sensitivity," *Multidimensional Systems and Signal Processing*, vol. 15, pp. 147-167, Apr. 2004.

[9] 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.

[10] T. Hinamoto, T. Oumi, O. I. Omoifo and W.-S. Lu, "Minimization of frequency-weighted l_2-sensitivity subject to l_2-scaling constraints for two-dimensional state-space digital filters," *IEEE Trans. on Signal Processing*, vol. 56, pp. 5157-5168, Oct. 2008.

[11] 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.

[12] K. Hirano, M. Sakane and M. Z. Mulk, "Design of three-dimensional recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 550-561, June 1984.

[13] H. Mutluay and M. M. Fahmy, "Frequency-domain design of N-D digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 1226-1233, Dec. 1985.

[14] M. E. Zervakis and A. N. Venetsanopoulos, "Design of three-dimensional digital filters using two-dimensional rotated filters," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 1452-1469, Dec. 1987.

[15] H. K. Kwan and C. L. Chan, "Design of multidimensional spherically symmetric and constant group delay recursive digital filters with sum of power-of-two coefficients," *IEEE Trans. Circuits Syst.*, vol. CAS-37, pp. 1027-1035, Aug. 1990.

[16] M. Muneyasu and T. Hinamoto, "Analytical design of multidimensional recursive digital filters with cubic symmetry and constant group delays," *J. Franklin Inst.*, vol. 330, pp. 1087-1100, 1993.