L_2-Sensitivity Minimization for a Class of 2-D Digital Filters Using a Quasi-Newton Method

Takao Hinamoto, Ken-ichi Iwata
Graduate School of Engineering
Hiroshima University
Higashi-Hiroshima 739-8527, Japan
Email: {hinamoto, iwata}@hiroshima-u.ac.jp

Wu-Sheng Lu
Dept. of Elec. and Comp. Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
Email: wslu@ece.uvic.ca

***Abstract*— The minimization problem of an L_2-sensitivity measure subject to L_2-norm dynamic-range scaling constraints is formulated for a class of two-dimensional (2-D) state-space digital filters. First, the problem is converted into an unconstrained optimization problem by using linear-algebraic techniques. Next, the unconstrained optimization problem is solved by applying an efficient quasi-Newton algorithm with closed-form formula for gradient evaluation. The coordinate transformation matrix obtained is then used to synthesize the optimal 2-D state-space filter structure that minimizes the L_2-sensitivity measure subject to the L_2-scaling constraints. Finally, a numerical example is presented to illustrate the utility of the proposed technique.**

I. INTRODUCTION

This paper is concerned with the optimal realization of a fixed-point state-space digital filter with finite word length (FWL). The efficiency and performance of the filter are directly influenced by selecting its state-space filter structure. When designing a transfer function with infinite accuracy coefficients so as to meet the filter specification requirements, and implementing it by a state-space model with a finite binary representation, the coefficients in the state-space model must be truncated or rounded to fit the FWL constraints. This coefficient quantization usually alters the characteristics of the filter and may change a stable filter to an unstable one. This motivates the study of the coefficient sensitivity minimization problem. In [1]-[10], two main classes of techniques have been proposed for constructing state-space digital filters that minimize the coefficient sensitivity, that is, L_1/L_2-sensitivity minimization [1]-[5] and L_2-sensitivity minimization [6]-[10]. It has been argued that the sensitivity measure based on the L_2 norm is more natural and reasonable relative to that based on the L_1/L_2-sensitivity minimization [6]-[10]. For 2-D state-space digital filters, the L_1/L_2-mixed sensitivity minimization problem [11]-[15] and L_2-sensitivity minimization problem [10],[16]-[19] have also been investigated. However, to our best knowledge, little has been done for the minimization of L_2-sensitivity subject to the L_2-norm dynamic-range scaling constraints for state-space digital filters [20], although it has been known that the use of scaling constraints can be beneficial for suppressing overflow oscillations [21],[22].

This paper investigates the problem of minimizing an L_2-sensitivity measure subject to L_2-norm dynamic-range scaling constraints for a class of 2-D state-space digital filters [23]. To this end, we introduce an expression for evaluating the L_2-sensitivity and formulate the L_2-sensitivity minimization problem subject to L_2-norm dynamic-range scaling constraints. Next, the constrained optimization problem is converted into an unconstrained optimization problem by using linear-algebraic techniques. The unconstrained optimization problem is then solved using an efficient quasi-Newton algorithm [24]. A numerical example is presented to demonstrate that the proposed algorithm offers much reduced L_2-sensitivity.

II. L_2-SENSITIVITY ANALYSIS

Consider a local state-space model $(\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{b}, \boldsymbol{c}_1, \boldsymbol{c}_2, d)_n$ for a class of 2-D recursive digital filters which is stable, locally controllable and locally observable [23]

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

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 real constant matrices of appropriate dimensions. The transfer function of (1) is given by

$$H(z_1,z_2) = (z_1^{-1}\boldsymbol{c}_1 + z_2^{-1}\boldsymbol{c}_2) \cdot \left(\boldsymbol{I}_n - z_1^{-1}\boldsymbol{A}_1 - z_2^{-1}\boldsymbol{A}_2\right)^{-1}\boldsymbol{b} + d. \tag{2}$$

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 then 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 $\boldsymbol{X}$.

From (2) and *Definition 1*, it can easily be shown that

$$\begin{aligned} \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{A}_k} &= z_k^{-1}[\boldsymbol{F}(z_1,z_2)\boldsymbol{G}(z_1,z_2)]^T \\ \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{b}} &= \boldsymbol{G}^T(z_1,z_2) \\ \frac{\partial H(z_1,z_2)}{\partial \boldsymbol{c}_k^T} &= z_k^{-1}\boldsymbol{F}(z_1,z_2), \qquad k = 1,2 \end{aligned} \tag{4}$$

where

$$\boldsymbol{F}(z_1,z_2)=\left(\boldsymbol{I}_n-z_1^{-1}\boldsymbol{A}_1-z_2^{-1}\boldsymbol{A}_2\right)^{-1}\boldsymbol{b}$$

$$\boldsymbol{G}(z_1,z_2)=(z_1^{-1}\boldsymbol{c}_1+z_2^{-1}\boldsymbol{c}_2)\left(\boldsymbol{I}_n-z_1^{-1}\boldsymbol{A}_1-z_2^{-1}\boldsymbol{A}_2\right)^{-1}.$$

The term d in (2) and its sensitivity are independent on the State-Space coordinate and therefore they are neglected here.

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_2 norm of $\boldsymbol{X}(z_1,z_2)$ is then defined as

$$\begin{aligned}&||\boldsymbol{X}(z_1,z_2)||_2\\&=\left(\mathrm{tr}\left[\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{X}(z_1,z_2)\boldsymbol{X}^*(z_1,z_2)\frac{dz_1dz_2}{z_1z_2}\right]\right)^{\frac{1}{2}}\end{aligned}\tag{5}$$

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

From (4) and *Definition 2*, the overall L_2-sensitivity measure for the local state-space (LSS) model in (1) is evaluated by

$$\begin{aligned}S=&\sum_{k=1}^{2}\left\|\frac{\partial H(z_1,z_2)}{\partial \boldsymbol{A}_k}\right\|_2^2+\left\|\frac{\partial H(z_1,z_2)}{\partial \boldsymbol{b}}\right\|_2^2\\&+\sum_{k=1}^{2}\left\|\frac{\partial H(z_1,z_2)}{\partial \boldsymbol{c}_k^T}\right\|_2^2\\=&\,2\left\|[\boldsymbol{F}(z_1,z_2)\boldsymbol{G}(z_1,z_2)]^T\right\|_2^2+\left\|\boldsymbol{G}^T(z_1,z_2)\right\|_2^2\\&+2\left\|\boldsymbol{F}(z_1,z_2)\right\|_2^2.\end{aligned}\tag{6}$$

The L_2-sensitivity measure in (6) can be written as

$$S=2\,\mathrm{tr}[\boldsymbol{M}]+\mathrm{tr}[\boldsymbol{W}_o]+2\,\mathrm{tr}[\boldsymbol{K}_c]\tag{7}$$

where

$$\begin{aligned}\boldsymbol{M}=&\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}[\boldsymbol{F}(z_1,z_2)\boldsymbol{G}(z_1,z_2)]^T\\&\cdot\boldsymbol{F}(z_1^{-1},z_2^{-1})\boldsymbol{G}(z_1^{-1},z_2^{-1})\frac{dz_1dz_2}{z_1z_2}\\\boldsymbol{K}_c=&\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{F}(z_1,z_2)\boldsymbol{F}^T(z_1^{-1},z_2^{-1})\frac{dz_1dz_2}{z_1z_2}\\\boldsymbol{W}_o=&\frac{1}{(2\pi j)^2}\oint_{\Gamma_1}\oint_{\Gamma_2}\boldsymbol{G}^T(z_1,z_2)\boldsymbol{G}(z_1^{-1},z_2^{-1})\frac{dz_1dz_2}{z_1z_2}.\end{aligned}$$

III. L_2-SENSITIVITY MINIMIZATION

If a coordinate transformation defined by

$$\overline{\boldsymbol{x}}(i,j)=\boldsymbol{T}^{-1}\boldsymbol{x}(i,j)\tag{8}$$

is applied to the LSS model in (1), we obtain a new realization $(\overline{\boldsymbol{A}}_1,\overline{\boldsymbol{A}}_2,\overline{\boldsymbol{b}},\overline{\boldsymbol{c}}_1,\overline{\boldsymbol{c}}_2,d)_n$ characterized by

$$\begin{aligned}\overline{\boldsymbol{A}}_1=\boldsymbol{T}^{-1}\boldsymbol{A}_1\boldsymbol{T},\quad&\overline{\boldsymbol{A}}_2=\boldsymbol{T}^{-1}\boldsymbol{A}_2\boldsymbol{T}\\\overline{\boldsymbol{b}}=\boldsymbol{T}^{-1}\boldsymbol{b},\quad&\overline{\boldsymbol{c}}_1=\boldsymbol{c}_1\boldsymbol{T},\quad\overline{\boldsymbol{c}}_2=\boldsymbol{c}_2\boldsymbol{T}\\\overline{\boldsymbol{K}}_c=\boldsymbol{T}^{-1}\boldsymbol{K}_c\boldsymbol{T}^{-T},\quad&\overline{\boldsymbol{W}}_o=\boldsymbol{T}^T\boldsymbol{W}_o\boldsymbol{T}.\end{aligned}\tag{9}$$

The coordinate transformation in (8) transforms (7) into

$$S(\boldsymbol{T})=2\,\mathrm{tr}[\boldsymbol{M}(\boldsymbol{T})]+\mathrm{tr}[\overline{\boldsymbol{W}}_o]+2\,\mathrm{tr}[\overline{\boldsymbol{K}}_c]\tag{10}$$

where

$$\boldsymbol{M}(\boldsymbol{T})=\boldsymbol{T}^T\left[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{H}^T(i,j)\boldsymbol{T}^{-T}\boldsymbol{T}^{-1}\boldsymbol{H}(i,j)\right]\boldsymbol{T}$$

Moreover, if the L_2-norm dynamic-range scaling constraints are imposed on the local state vector $\overline{\boldsymbol{x}}(i,j)$, then

$$(\overline{\boldsymbol{K}}_c)_{ii}=(\boldsymbol{T}^{-1}\boldsymbol{K}_c\boldsymbol{T}^{-T})_{ii}=1\tag{11}$$

is required for $i=1,2,\cdots,n$.

The problem considered here is as follows: *Given* $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{b}$, $\boldsymbol{c}_1$ *and* $\boldsymbol{c}_2$, *obtain an* $n\times n$ *nonsingular matrix* $\boldsymbol{T}$ *which minimizes (10) subject to the scaling constraints in (11).*

When the LSS model in (1) is assumed to be stable and locally controllable, the local controllability Gramian $\boldsymbol{K}_c$ is symmetric and positive-definite [15]. This implies that $\boldsymbol{K}_c^{1/2}$ satisfying $\boldsymbol{K}_c=\boldsymbol{K}_c^{1/2}\boldsymbol{K}_c^{1/2}$ is also symmetric and positive-definite. Defining

$$\hat{\boldsymbol{T}}=\boldsymbol{T}^T\boldsymbol{K}_c^{-\frac{1}{2}},\tag{12}$$

the scaling constraints in (11) can be expressed as

$$(\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{T}}^{-1})_{ii}=1,\qquad i=1,2,\cdots,n.\tag{13}$$

The constraints in (13) simply state that each column in $\hat{\boldsymbol{T}}^{-1}$ must be a unity vector. If matrix $\hat{\boldsymbol{T}}^{-1}$ is assumed to have the form

$$\hat{\boldsymbol{T}}^{-1}=\left[\frac{\boldsymbol{t}_1}{||\boldsymbol{t}_1||},\frac{\boldsymbol{t}_2}{||\boldsymbol{t}_2||},\cdots,\frac{\boldsymbol{t}_n}{||\boldsymbol{t}_n||}\right],\tag{14}$$

then (13) is always satisfied. From (12), it follows that (10) is changed to

$$\begin{aligned}J_o(\hat{\boldsymbol{T}})=&\,2\,\mathrm{tr}[\hat{\boldsymbol{T}}\left[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\hat{\boldsymbol{H}}^T(i,j)\hat{\boldsymbol{T}}^{-1}\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{H}}(i,j)\right]\hat{\boldsymbol{T}}^T]\\&+\mathrm{tr}[\hat{\boldsymbol{T}}\hat{\boldsymbol{W}}_o\hat{\boldsymbol{T}}^T]+2\,\mathrm{tr}[\hat{\boldsymbol{T}}^{-T}\hat{\boldsymbol{T}}^{-1}]\end{aligned}\tag{15}$$

where

$$\hat{\boldsymbol{H}}(i,j)=\boldsymbol{K}_c^{-\frac{1}{2}}\boldsymbol{H}(i,j)\boldsymbol{K}_c^{\frac{1}{2}},\qquad\hat{\boldsymbol{W}}_o=\boldsymbol{K}_c^{\frac{1}{2}}\boldsymbol{W}_o\boldsymbol{K}_c^{\frac{1}{2}}.$$

From the foregoing arguments, the problem of obtaining an $n\times n$ nonsingular matrix $\boldsymbol{T}$ which minimizes (10) subject to the scaling constraints in (11) can be converted into an unconstrained optimization problem of obtaining an $n\times n$ nonsingular matrix $\hat{\boldsymbol{T}}$ which minimizes (15).

Now we apply a quasi-Newton algorithm [24] to minimize (15) with respect to matrix $\hat{\boldsymbol{T}}$ given by (14). Let $\boldsymbol{x}$ be the column vector that collects the variables in matrix $\hat{\boldsymbol{T}}$. Then $J_o(\hat{\boldsymbol{T}})$ is a function of $\boldsymbol{x}$, which we denote by $J(\boldsymbol{x})$. The algorithm starts with a trivial initial point $\boldsymbol{x}_0$ obtained from an initial assignment $\hat{\boldsymbol{T}}=\boldsymbol{I}_n$. Then, in the kth iteration a quasi-Newton algorithm updates the most recent point $\boldsymbol{x}_k$ to point $\boldsymbol{x}_{k+1}$ as

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k\tag{16}$$

where

$$\boldsymbol{d}_k = -\boldsymbol{S}_k \nabla J(\boldsymbol{x}_k)$$

$$\alpha_k = arg \min_{\alpha} J(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$$

$$\boldsymbol{S}_{k+1} = \boldsymbol{S}_k + \left(1 + \frac{\boldsymbol{\gamma}_k^T \boldsymbol{S}_k \boldsymbol{\gamma}_k}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k}\right) \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k} - \frac{\boldsymbol{\delta}_k \boldsymbol{\gamma}_k^T \boldsymbol{S}_k + \boldsymbol{S}_k \boldsymbol{\gamma}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k}$$

$$\boldsymbol{S}_0 = \boldsymbol{I}, \;\; \boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \;\; \boldsymbol{\gamma}_k = \nabla J(\boldsymbol{x}_{k+1}) - \nabla J(\boldsymbol{x}_k).$$

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 $J(\boldsymbol{x})$. This iteration process continues until

$$|J(\boldsymbol{x}_{k+1}) - J(\boldsymbol{x}_k)| < \varepsilon \tag{17}$$

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

The implementation of (16) requires the gradient of $J(\boldsymbol{x})$. Closed-form expressions for $\nabla J(\boldsymbol{x})$ are given below.

$$\begin{aligned} \frac{\partial J(\hat{\boldsymbol{T}})}{\partial t_{pq}} &= \lim_{\Delta \to 0} \frac{J(\hat{\boldsymbol{T}}_{pq}) - J(\hat{\boldsymbol{T}})}{\Delta} \\ &= 2\beta_1 - \beta_2 + 2\beta_3 \end{aligned} \tag{18}$$

where $\hat{\boldsymbol{T}}_{pq}$ is the matrix obtained from $\hat{\boldsymbol{T}}$ with its (p,q)th component perturbed by Δ:

$$\hat{\boldsymbol{T}}_{pq} = \hat{\boldsymbol{T}} + \frac{\Delta \hat{\boldsymbol{T}} \boldsymbol{g}_{pq} \boldsymbol{e}_q^T \hat{\boldsymbol{T}}}{1 - \Delta \boldsymbol{e}_q^T \hat{\boldsymbol{T}} \boldsymbol{g}_{pq}}$$

$$\beta_1 = \boldsymbol{e}_q^T \hat{\boldsymbol{T}} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \hat{\boldsymbol{H}}^T(i,j) \hat{\boldsymbol{T}}^{-1} \hat{\boldsymbol{T}}^{-T} \hat{\boldsymbol{H}}(i,j) \right] \hat{\boldsymbol{T}}^T \hat{\boldsymbol{T}} \boldsymbol{g}_{pq}$$

$$\beta_2 = \boldsymbol{e}_q^T \hat{\boldsymbol{T}}^{-T} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \hat{\boldsymbol{H}}(i,j) \hat{\boldsymbol{T}}^T \hat{\boldsymbol{T}} \hat{\boldsymbol{H}}^T(i,j) \right] \boldsymbol{g}_{pq}$$

$$\beta_3 = \boldsymbol{e}_q^T \hat{\boldsymbol{T}} \hat{\boldsymbol{W}}_o \hat{\boldsymbol{T}}^T \hat{\boldsymbol{T}} \boldsymbol{g}_{pq}$$

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

where $\boldsymbol{e}_p$ is an $n \times 1$ unit vector whose pth entry equals unity.

IV. NUMERICAL EXAMPLE

Let a class of 2-D digital filters $(\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{b}, \boldsymbol{c}_1, \boldsymbol{c}_2, d)_n$ in (1) be specified by

$$\boldsymbol{A}_1 = \begin{bmatrix} 0 & 0.481228 & 0 & 0 \\ 0 & 0 & 0.510378 & 0 \\ 0 & 0 & 0 & 0.525287 \\ -0.031857 & 0.298663 & -0.808282 & 1.044600 \end{bmatrix}$$

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

$$\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 & 0.198473 \end{bmatrix}^T$$

$$\boldsymbol{c}_1 = \begin{bmatrix} -0.567054 & 0.231913 & 0.197016 & 0.239932 \end{bmatrix}$$

$$\boldsymbol{c}_2 = \begin{bmatrix} 0.464344 & 0.441837 & -0.061100 & 0.105505 \end{bmatrix}$$

$$d = 0.00943.$$

In this case, it is follows from (7) that the Grammians $\boldsymbol{K}_c$, $\boldsymbol{W}_o$, and $\boldsymbol{M}$ are calculated as

$$\boldsymbol{K}_c = \begin{bmatrix} 1.000000 & 0.987279 & 0.940868 & 0.844274 \\ 0.987279 & 1.000000 & 0.976755 & 0.888478 \\ 0.940868 & 0.976755 & 1.000000 & 0.952963 \\ 0.844274 & 0.888478 & 0.952963 & 1.000000 \end{bmatrix}$$

$$\boldsymbol{W}_o = 10 \cdot \begin{bmatrix} 1.337108 & -1.304050 & 0.189462 & -0.556646 \\ -1.304050 & 1.637345 & -0.429399 & 0.576183 \\ 0.189462 & -0.429399 & 2.122604 & -2.191942 \\ -0.556646 & 0.576183 & -2.191942 & 2.672484 \end{bmatrix}$$

$$\boldsymbol{M} = 10^3 \cdot \begin{bmatrix} 1.001461 & -1.050382 & 0.582275 & -0.913062 \\ -1.050382 & 1.182943 & -0.755388 & 1.062465 \\ 0.582275 & -0.755388 & 2.170753 & -2.398972 \\ -0.913062 & 1.062465 & -2.398972 & 2.814168 \end{bmatrix}.$$

The L_2-sensitivity measure S_2 in (7) is then computed as

$$S = 1.442435 \times 10^4.$$

Applying the quasi-Newton algorithm in (16) to the minimization of (15), it took 30 iterations to converge to

$$\hat{\boldsymbol{T}} = \begin{bmatrix} 0.501598 & -0.103085 & -0.225600 & 0.300033 \\ 0.181452 & 0.691561 & -0.214926 & 0.233838 \\ 0.395655 & 0.681436 & 0.410789 & 0.036465 \\ -0.747617 & -0.216243 & 0.856834 & 0.924105 \end{bmatrix}$$

which leads to

$$\boldsymbol{T} = \begin{bmatrix} 0.693066 & 0.209465 & -0.093284 & 1.044811 \\ 0.556893 & 0.363710 & -0.027678 & 0.956211 \\ 0.527631 & 0.371601 & 0.240955 & 0.815707 \\ 0.354314 & 0.260633 & 0.335061 & 0.822833 \end{bmatrix}.$$

In this case, the new realization $(\overline{\boldsymbol{A}}_1, \overline{\boldsymbol{A}}_2, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}_1, \overline{\boldsymbol{c}}_2, d)_n$ in (9) is constructed as

$$\overline{\boldsymbol{A}}_1 = \begin{bmatrix} 0.205926 & 0.094730 & -0.183488 & 0.105715 \\ 0.326754 & 0.273061 & 0.651563 & -0.077094 \\ -0.281487 & -0.176174 & 0.154131 & 0.288202 \\ 0.029260 & 0.034210 & -0.007899 & 0.411483 \end{bmatrix}$$

$$\overline{\boldsymbol{A}}_2 = \begin{bmatrix} 0.350051 & 0.129690 & -0.166229 & 0.174840 \\ -0.285268 & 0.510377 & -0.289206 & -0.041400 \\ -0.271866 & 0.174523 & 0.238315 & -0.005608 \\ 0.076961 & 0.060882 & 0.194526 & 0.395246 \end{bmatrix}$$

$$\overline{\boldsymbol{b}} = \begin{bmatrix} -0.729312 & -0.408558 & 0.239624 & 0.587085 \end{bmatrix}^T$$

$$\overline{\boldsymbol{c}}_1 = \begin{bmatrix} -0.074892 & 0.101317 & 0.174342 & -0.012574 \end{bmatrix}$$

$$\overline{\boldsymbol{c}}_2 = \begin{bmatrix} 0.573022 & 0.262758 & -0.034917 & 0.944616 \end{bmatrix}$$

$$d = 0.00943$$

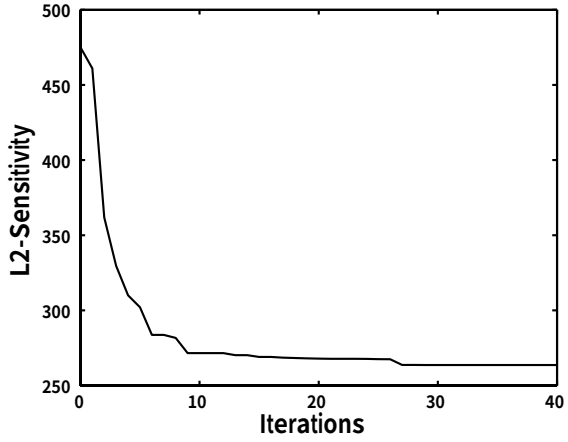

Fig. 1. L_2-Sensitivity Performance

which yields

$$\overline{\boldsymbol{K}}_c = \begin{bmatrix} 1.000000 & 0.505058 & -0.630213 & -0.483523 \\ 0.505058 & 1.000000 & -0.030736 & -0.044198 \\ -0.630213 & -0.030736 & 1.000000 & 0.688839 \\ -0.483523 & -0.044198 & 0.688839 & 1.000000 \end{bmatrix}$$

$$\overline{\boldsymbol{W}}_o = \begin{bmatrix} 0.905086 & 0.323428 & -0.444821 & 0.602792 \\ 0.323428 & 0.885081 & 0.323665 & 0.801004 \\ -0.444821 & 0.323665 & 0.968093 & 0.640803 \\ 0.602792 & 0.801004 & 0.640803 & 2.330628 \end{bmatrix}$$

$$\boldsymbol{M}(\boldsymbol{T}) = 10 \cdot \begin{bmatrix} 2.131163 & 0.659968 & -1.689791 & -0.292979 \\ 0.659968 & 2.779735 & 1.223639 & 1.911711 \\ -1.689791 & 1.223639 & 3.345663 & 2.646575 \\ -0.292979 & 1.911711 & 2.646575 & 4.337863 \end{bmatrix}.$$

The L_2-sensitivity measure in (10) is then minimized subject to the scaling constraints in (11) to

$$S(\boldsymbol{T}) = 2.649774 \times 10^2.$$

The L_2-sensitivity performance of 50 iterations in (15) is shown in Fig. 1, from which it is observed that the iterative algorithm converges with 30 iterations.

II. CONCLUSION

We have investigated the problem of minimizing the L_2-sensitivity measure subject to L_2-norm dynamic-range scaling constraints for a class of 2-D state-space digital filters. We have shown that the L_2-sensitivity minimization problem subject to L_2-scaling constraints can be converted into an unconstrained optimization problem by using linear algebraic techniques. An efficient quasi-Newton algorithm has then been applied to solve the unconstrained optimization problem. The coordinate transformation matrix obtained has allowed us to construct the optimal 2-D state-space filter structure. Computer simulation results have demonstrated the effectiveness of the proposed technique.

REFERENCES

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

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

[3] 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.

[4] G. Li and M. Gevers, "Optimal finite precision implementation of a state-estimate feedback controller," *IEEE Trans. Circuits Syst*., vol.37, pp.1487-1498, Dec. 1990.

[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] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, London, 1994.

[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] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] T. Hinamoto, H. Ohnishi and W.-S. Lu, "Minimization of L_2-sensitivity For state-space digital filters subject to L_2-scaling constraints," in *Proc. 2004 IEEE Int. Symp. Circuits Syst*., vol.III, pp.137-140.

[21] 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.

[22] 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.

[23] 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.

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