# Weighted Pole and Zero Sensitivity Minimization for State-Space Digital Filters

Takao Hinamoto, Akimitsu Doi
Hiroshima Institute of Technology
Hiroshima 731-5193, Japan
Emails: hinamoto@ieee.org, doi@cc.it-hiroshima.ac.jp

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

***Abstract*— This paper investigates the problem of minimizing weighted pole and zero sensitivity for state-space digital filters. A new measure for pole and zero sensitivity is proposed and an efficient iterative technique for minimizing this measure is developed by employing a quasi-Newton algorithm. A novel and simple method is also explored for obtaining the optimal coordinate transformation matrix which minimizes a zero sensitivity measure subject to minimal pole sensitivity. A numerical example is presented to demonstrate the validity and effectiveness of the proposed technique.**

## 1. INTRODUCTION

When designing a transfer function with infinite accuracy coefficients to meet the filter specification requirements and then realizing it by a state-space model, the state-space parameters must be truncated or rounded to implement the filter in a finite binary representation that fits the finite-word-length (FWL) constraints. This coefficient quantization usually alters the characteristics of the filter. For instance, it may change a stable filter to unstable one. This motivates the study of minimizing coefficients sensitivity. As is well known, there are several ways to define sensitivity of a filter with respect to its realization coefficients. Two of them are based on a mixed $l_1/l_2$ norm or a pure $l_2$ norm that measures changes of a certain transfer function, while the other is defined in terms of the poles and zeros of a filter. Several techniques for minimizing the $l_1/l_2$-sensitivity measure [1]-[5] and the $l_2$-sensitivity measure [6]-[10] have been proposed. Alternatively, the pole and zero sensitivity of a filter with respect to state-space parameters has been analyzed, and its reduction or minimization have been considered [11], [12]. It has been mentioned in [11] that the sensitivities of the poles and zeros of a transfer function with respect to the variations in state-space parameters (due to FWL) are closely related to the sensitivity characteristics of its frequency transfer function. Moreover, minimal pole sensitivity is achieved when matrix $\boldsymbol{A}$ is normal, and minimal zero sensitivity is acheived when a certain matrix is normal. However, minimal pole sensitivity does not guarantee minimal zero sensitivity and vice versa. In [12], a method for minimizing a zero sensitivity measure subject to minimal ploe sensitivity for state-space digital filters has been explored, since it is generally impossible to minimize pole sensitivity and zero sensitivity simultaneously.

In this paper, the problem of minimizing weighted pole and zeo sensitivity for state-space digital filters is investigated. An efficient iterative technique for minimizing a weighted pole and zeo sensitivity measure is developed by employing a quasi-Newton algorithm. Moreover, a novel and simple method for minimizing a zero sensitivity measure subject to minimal pole sensitivity is explored. A numerical example is included to demonstrate the validity and effectiveness of the proposed technique.

## 2. PROBLEM FORMULATION

Consider a stable, controllable and observable state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_n$ of order $n$ described by

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{b}u(k) \\ y(k) &= \boldsymbol{c}\boldsymbol{x}(k) + du(k) \end{aligned} \tag{1}$$

where $\boldsymbol{x}(k)$ is an $n \times 1$ state-variable vector, $u(k)$ is a scalar input, $y(k)$ is a scalar output, and $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$ and $d$ are real constant matrices of appropriate dimensions. The transfer function of the digital filter in (1) can be expressed as

$$H(z) = \boldsymbol{c}(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}\boldsymbol{b} + d. \tag{2}$$

The poles $\{\lambda_l\}$ of $H(z)$ are denoted by $\{\lambda_l\} = \lambda(\boldsymbol{A})$, while the zeros $\{v_l\}$ of $H(z)$ are indecated by $\{v_l\} = \lambda(\boldsymbol{Z})$ with

$$\boldsymbol{Z} = \boldsymbol{A} - d^{-1}\boldsymbol{b}\boldsymbol{c} \tag{3}$$

provided that $d \neq 0$ [13]. Hereafter, it is assumed that a direct path from the input to the output always exists in (1), i.e., $d \neq 0$.

*Lemma 1* [12]: Let $\boldsymbol{x}_p(l)$ be a right eigenvector corresponding to $\lambda_l$ and let $\boldsymbol{y}_p(l)$ be the reciprocal left eigenvector that corresponds to $\boldsymbol{x}_p(l)$. Then, under the assumption that $\boldsymbol{A}$ has a full set of linearly independent eigenvectors, it follows that

$$\left(\frac{\partial \lambda_l}{\partial \boldsymbol{A}}\right)^T = \boldsymbol{x}_p(l)\boldsymbol{y}_p^H(l) \quad \text{for} \quad l = 1, 2, \cdots, n. \tag{4}$$

*Proof*: By virtue of $\lambda_l = \boldsymbol{y}_p^H(l)\boldsymbol{A}\boldsymbol{x}_p(l)$, $\boldsymbol{A}\boldsymbol{x}_p(l) = \lambda_l\boldsymbol{x}_p(l)$, $\boldsymbol{y}_p^H(l)\boldsymbol{A} = \lambda_l\boldsymbol{y}_p^H(l)$ and $\boldsymbol{y}_p^H(l)\boldsymbol{x}_p(l) = 1$, we can derive that

$$\frac{\partial \lambda_l}{\partial a_{ij}} = \boldsymbol{y}_p^H(l)\frac{\partial \boldsymbol{A}}{\partial a_{ij}}\boldsymbol{x}_p(l) = \boldsymbol{e}_j^T\boldsymbol{x}_p(l)\boldsymbol{y}_p^H(l)\boldsymbol{e}_i \tag{5}$$

which is equivalent to (4) where $a_{ij}$ is the $(i,j)$th entry of $\boldsymbol{A}$.

*Lemma 2* [12]: Let $\boldsymbol{x}_z(l)$ be a right eigenvector corresponding to $v_l$ and let $\boldsymbol{y}_z(l)$ be the reciprocal left eigenvector that corresponds to $\boldsymbol{x}_z(l)$. Then, under the assumption that $\boldsymbol{Z}$ has a linearly independent eigenvector set, we obtain

$$\left(\frac{\partial v_l}{\partial \boldsymbol{Z}}\right)^T = \boldsymbol{x}_z(l)\boldsymbol{y}_z^H(l) \quad \text{for} \quad l = 1, 2, \cdots, n. \tag{6}$$

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

*Lemma 3* [8]: If (6) is valid then it is derived that

$$\begin{aligned} \frac{\partial v_l}{\partial \boldsymbol{A}} &= \frac{\partial v_l}{\partial \boldsymbol{Z}}, & \frac{\partial v_l}{\partial \boldsymbol{b}} &= -d^{-1}\frac{\partial v_l}{\partial \boldsymbol{Z}}\boldsymbol{c}^T \\ \frac{\partial v_l}{\partial \boldsymbol{c}} &= -d^{-1}\boldsymbol{b}^T\frac{\partial v_l}{\partial \boldsymbol{Z}}, & \frac{\partial v_l}{\partial d} &= d^{-2}\boldsymbol{b}^T\frac{\partial v_l}{\partial \boldsymbol{Z}}\boldsymbol{c}^T. \end{aligned} \tag{7}$$

*Proof*: By making use of (3), (6) and $v_l = \boldsymbol{y}_z^H(l)\boldsymbol{Z}\boldsymbol{x}_z(l)$, it can easily be shown that (7) is valid.

Using the Frobenius norm for an $m \times n$ complex matrix $\boldsymbol{M}$ defined by

$$||\boldsymbol{M}||_F = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|(\boldsymbol{M})_{ij}|^2\Big)^{\frac{1}{2}} \tag{8}$$

we define the pole sensitivity measure $J_p$ and the zero sensitivity measure $J_z$ for the digital filter in (1) as

$$\begin{aligned} J_p &= \sum_{l=1}^{n}\left\|\frac{\partial \lambda_l}{\partial \boldsymbol{A}}\right\|_F^2 = \sum_{l=1}^{n}\left\|\Big(\frac{\partial \lambda_l}{\partial \boldsymbol{A}}\Big)^T\right\|_F^2 \\ J_z &= \sum_{l=1}^{n}\Big\{\left\|\frac{\partial v_l}{\partial \boldsymbol{A}}\right\|_F^2 + \left\|\frac{\partial v_l}{\partial \boldsymbol{b}}\right\|_F^2 + \left\|\frac{\partial v_l}{\partial \boldsymbol{c}}\right\|_F^2 + \left\|\frac{\partial v_l}{\partial d}\right\|_F^2\Big\} \end{aligned} \tag{9}$$

respectively. Noting that

$$||\boldsymbol{M}||_F^2 = \mathrm{tr}[\boldsymbol{M}\boldsymbol{M}^H] = \mathrm{tr}[\boldsymbol{M}^H\boldsymbol{M}] = ||\boldsymbol{M}^H||_F^2 \tag{10}$$

and by virtue of (4), (6) and (7), we can write (9) as

$$J_p = \sum_{l=1}^{n}\Big(\boldsymbol{x}_p^H(l)\boldsymbol{x}_p(l)\Big)\Big(\boldsymbol{y}_p^H(l)\boldsymbol{y}_p(l)\Big) \tag{11a}$$

$$J_z = \sum_{l=1}^{n}\Big(\boldsymbol{x}_z^H(l)\boldsymbol{x}_z(l) + \alpha_l^2\Big)\Big(\boldsymbol{y}_z^H(l)\boldsymbol{y}_z(l) + \beta_l^2\Big) \tag{11b}$$

where

$$\alpha_l = |d^{-1}\boldsymbol{c}\boldsymbol{x}_z(l)|, \qquad \beta_l = |d^{-1}\boldsymbol{y}_z^H(l)\boldsymbol{b}|. \tag{11c}$$

If a coordinate transformation defined by

$$\overline{\boldsymbol{x}}(k) = \boldsymbol{T}^{-1}\boldsymbol{x}(k) \tag{12}$$

is applied to the digital filter in (1), then the new realization $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_n$ can be characterized by

$$\overline{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}, \quad \overline{\boldsymbol{b}} = \boldsymbol{T}^{-1}\boldsymbol{b}, \quad \overline{\boldsymbol{c}} = \boldsymbol{c}\boldsymbol{T}. \tag{13}$$

From (2) and (13), it is obvious that the transfer function $H(z)$ is invariant under the coordinate transformation in (12). For the new realization $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_n$, the right eigenvector $\boldsymbol{x}_p(l)$ corresponding to $\lambda_l$ and the reciprocal left eigenvector $\boldsymbol{y}_p(l)$ corresponding to $\boldsymbol{x}_p(l)$ are changed to

$$\overline{\boldsymbol{x}}_p(l) = \boldsymbol{T}^{-1}\boldsymbol{x}_p(l), \qquad \overline{\boldsymbol{y}}_p(l) = \boldsymbol{T}^T\boldsymbol{y}_p(l), \tag{14}$$

respectively, for $l = 1, 2, \cdots, n$. Similarly, the right eigenvector $\boldsymbol{x}_z(l)$ corresponding to $v_l$ and the reciprocal left eigenvector $\boldsymbol{y}_z(l)$ corresponding to $\boldsymbol{x}_z(l)$ are changed to

$$\overline{\boldsymbol{x}}_z(l) = \boldsymbol{T}^{-1}\boldsymbol{x}_z(l), \qquad \overline{\boldsymbol{y}}_z(l) = \boldsymbol{T}^T\boldsymbol{y}_z(l), \tag{15}$$

respectively, for $l = 1, 2, \cdots, n$. Therefore, for the new realizatiom $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_n$ specified in (13), the pole and zero sensitivity measures in (11a) and (11b) can be expressed as

$$\begin{aligned} J_p(\boldsymbol{T}) &= \sum_{l=1}^{n}\Big(\boldsymbol{x}_p^H(l)\boldsymbol{T}^{-T}\boldsymbol{T}^{-1}\boldsymbol{x}_p(l)\Big)\Big(\boldsymbol{y}_p^H(l)\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{y}_p(l)\Big) \\ J_z(\boldsymbol{T}) &= \sum_{l=1}^{n}\Big(\boldsymbol{x}_z^H(l)\boldsymbol{T}^{-T}\boldsymbol{T}^{-1}\boldsymbol{x}_z(l) + \alpha_l^2\Big) \\ &\qquad \cdot\Big(\boldsymbol{y}_z^H(l)\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{y}_z(l) + \beta_l^2\Big) \end{aligned} \tag{16}$$

respectively. Note that for any $n \times n$ nonsingular matrix $\boldsymbol{T}$

$$J_p(\boldsymbol{T}) \geq n, \qquad J_z(\boldsymbol{T}) \geq \sum_{l=1}^{n}(1 + |\alpha_l\beta_l|)^2 \tag{17}$$

are always hold true [11], [12]. In this paper, we consider a weighted pole and zero sensitivity measure $J_\gamma(\boldsymbol{T})$ defined by

$$J_\gamma(\boldsymbol{T}) = \gamma J_p(\boldsymbol{T}) + (1-\gamma)J_z(\boldsymbol{T}) \tag{18}$$

where $0 \leq \gamma \leq 1$ is a weighting factor. Evidently, a $J_\gamma(\boldsymbol{T})$ with a greater $\gamma$ represents a measure that places more emphasis on pole sensitivity while a $J_\gamma(\boldsymbol{T})$ with a smaller $\gamma$ serves as a measure that weights more heavily on zero sensitivity. In addition, by setting $\gamma$ to one or zero $J_\gamma(\boldsymbol{T})$ becomes $J_p(\boldsymbol{T})$ or $J_z(\boldsymbol{T})$, respectively.

The problem of minimizing the weighted pole and zero sensitivity measure is now formulated as follows: *For given $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $d$, obtain an $n \times n$ nonsingular matrix $\boldsymbol{T}$ which minimizes $J_\gamma(\boldsymbol{T})$ in (18) with $\gamma$ specified by the designer.*

## 3. POLE AND ZERO SENSITIVITY MINIMIZATION

### A. Minimization of Weighted Pole and Zero Sensitivity Measure $J_\gamma(\boldsymbol{T})$

In order to minimize (18) over an $n \times n$ nonsingular matrix $\boldsymbol{T}$, we define

$$\boldsymbol{T}^{-1} = \big[\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n\big], \quad \boldsymbol{x} = \big[\boldsymbol{t}_1^T, \boldsymbol{t}_2^T, \cdots, \boldsymbol{t}_n^T\big]^T \tag{19}$$

and denote $J_\gamma(\boldsymbol{T})$ in (18) as function $J_\gamma(\boldsymbol{x})$ of the length-$n^2$ vector $\boldsymbol{x}$. In the $k$th iteration, a quasi-Newton algorithm updates the current point $\boldsymbol{x}_k$ to point $\boldsymbol{x}_{k+1}$ as [14]

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k \tag{20}$$

where

$$\begin{aligned} \boldsymbol{d}_k &= -\boldsymbol{S}_k\nabla J_\gamma(\boldsymbol{x}_k), \quad \alpha_k = arg\Big[\min_{\alpha}\ J_\gamma(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k)\Big] \\ \boldsymbol{S}_{k+1} &= \boldsymbol{S}_k + \Big(1 + \frac{\boldsymbol{\varphi}_k^T\boldsymbol{S}_k\boldsymbol{\varphi}_k}{\boldsymbol{\varphi}_k^T\boldsymbol{\delta}_k}\Big)\frac{\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\varphi}_k^T\boldsymbol{\delta}_k} - \frac{\boldsymbol{\delta}_k\boldsymbol{\varphi}_k^T\boldsymbol{S}_k + \boldsymbol{S}_k\boldsymbol{\varphi}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\varphi}_k^T\boldsymbol{\delta}_k} \\ \boldsymbol{S}_0 &= \boldsymbol{I}, \ \ \boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \ \ \boldsymbol{\varphi}_k = \nabla J_\gamma(\boldsymbol{x}_{k+1}) - \nabla J_\gamma(\boldsymbol{x}_k). \end{aligned}$$

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

$$|J_\gamma(\boldsymbol{x}_{k+1}) - J_\gamma(\boldsymbol{x}_k)| < \varepsilon \tag{21}$$

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

The gradient of $J_\gamma(\boldsymbol{x})$ can be evaluated using closed-form expressions as

$$\begin{aligned}\frac{\partial J_\gamma(\boldsymbol{x})}{\partial t_{ij}} &= \lim_{\Delta\to 0}\frac{J_\gamma(\boldsymbol{T}_{ij})-J_\gamma(\boldsymbol{T})}{\Delta}\\ &= \gamma\frac{\partial J_p(\boldsymbol{x})}{\partial t_{ij}}+(1-\gamma)\frac{\partial J_z(\boldsymbol{x})}{\partial t_{ij}}\end{aligned} \tag{22}$$

where $\boldsymbol{T}_{ij}$ is the matrix obtained from $\boldsymbol{T}$ with a perturbed $(i,j)$th component. Then it follows that [15, p. 655]

$$\begin{aligned}\boldsymbol{T}_{ij} &= \boldsymbol{T}-\frac{\Delta\boldsymbol{T}\boldsymbol{e}_i\boldsymbol{e}_j^T\boldsymbol{T}}{1+\Delta\boldsymbol{e}_j^T\boldsymbol{T}\boldsymbol{e}_i}\simeq\boldsymbol{T}-\Delta\boldsymbol{T}\boldsymbol{e}_i\boldsymbol{e}_j^T\boldsymbol{T}\\ \boldsymbol{T}_{ij}^{-1} &= \boldsymbol{T}^{-1}+\Delta\frac{\partial\boldsymbol{t}_j}{\partial t_{ij}}\boldsymbol{e}_j^T=\boldsymbol{T}^{-1}+\Delta\boldsymbol{e}_i\boldsymbol{e}_j^T\end{aligned} \tag{23}$$

and

$$\begin{aligned}\frac{\partial J_p(\boldsymbol{x})}{\partial t_{ij}} &= \sum_{l=1}^{n}\Big\{\Big(\boldsymbol{x}_p^H(l)\boldsymbol{M}(\boldsymbol{T})\boldsymbol{x}_p(l)\Big)\Big(\boldsymbol{y}_p^H(l)\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{y}_p(l)\Big)\\ &\quad-\Big(\boldsymbol{x}_p^H(l)\boldsymbol{T}^{-T}\boldsymbol{T}^{-1}\boldsymbol{x}_p(l)\Big)\Big(\boldsymbol{y}_p^H(l)\boldsymbol{N}(\boldsymbol{T})\boldsymbol{y}_p(l)\Big)\Big\}\\ \frac{\partial J_z(\boldsymbol{x})}{\partial t_{ij}} &= \sum_{l=1}^{n}\Big\{\boldsymbol{x}_z^H(l)\boldsymbol{M}(\boldsymbol{T})\boldsymbol{x}_z(l)\Big(\boldsymbol{y}_z^H(l)\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{y}_z(l)+\beta_l^2\Big)\\ &\quad-\Big(\boldsymbol{x}_z^H(l)\boldsymbol{T}^{-T}\boldsymbol{T}^{-1}\boldsymbol{x}_z(l)+\alpha_l^2\Big)\boldsymbol{y}_z^H(l)\boldsymbol{N}(\boldsymbol{T})\boldsymbol{y}_z(l)\Big\}\end{aligned} \tag{24}$$

where

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{T}) &= \boldsymbol{e}_j\boldsymbol{e}_i^T\boldsymbol{T}^{-1}+\boldsymbol{T}^{-T}\boldsymbol{e}_i\boldsymbol{e}_j^T\\ \boldsymbol{N}(\boldsymbol{T}) &= \boldsymbol{T}\big[\boldsymbol{e}_i\boldsymbol{e}_j^T\boldsymbol{T}+\boldsymbol{T}^T\boldsymbol{e}_j\boldsymbol{e}_i^T\big]\boldsymbol{T}^T.\end{aligned}$$

### *B. Minimization of Zero Sensitivity Subject to a Constraint on Pole Sensitivity*

Suppose that the coordinate transformation matrix $\boldsymbol{T}_o$ yields $J_p(\boldsymbol{T}_o)=n$. Then it can be verified that $J_p(\pm\sqrt{\zeta}\,\boldsymbol{T}_o)=n$ is always satisfied for any scalar $\pm\sqrt{\zeta}$. Therefore, one can reduce the zero sensitivity as much as possible while keep the pole sensitivity unaltered by adequately tuning the value of $\zeta$. To this end, we compute

$$\begin{aligned}\frac{\partial J_z(\pm\sqrt{\zeta}\,\boldsymbol{T}_o)}{\partial\zeta} &= \zeta^{-2}\sum_{l=1}^{n}\Big\{\zeta^2\alpha_l^2\boldsymbol{y}_z^H(l)\boldsymbol{T}_o\boldsymbol{T}_o^T\boldsymbol{y}_z(l)\\ &\quad-\beta_l^2\boldsymbol{x}_z^H(l)\boldsymbol{T}_o^{-T}\boldsymbol{T}_o^{-1}\boldsymbol{x}_z(l)\Big\}.\end{aligned} \tag{25}$$

By solving $\partial J_z(\pm\sqrt{\zeta}\,\boldsymbol{T}_o)/\partial\zeta=0$, the optimal value of $\zeta$ is found to be

$$\zeta=\sqrt{\frac{\displaystyle\sum_{l=1}^{n}\beta_l^2\boldsymbol{x}_z^H(l)\boldsymbol{T}_o^{-T}\boldsymbol{T}_o^{-1}\boldsymbol{x}_z(l)}{\displaystyle\sum_{l=1}^{n}\alpha_l^2\boldsymbol{y}_z^H(l)\boldsymbol{T}_o\boldsymbol{T}_o^T\boldsymbol{y}_z(l)}}. \tag{26}$$

This implies that the optimal coordinate transformation matrix $\boldsymbol{T}^{opt}$ that minimizes $J_z(\pm\sqrt{\zeta}\,\boldsymbol{T}_o)$ with respect to $\zeta$ subject to $J_p(\boldsymbol{T}_o)=n$ can be obtained by

$$\boldsymbol{T}^{opt}=\pm\sqrt{\zeta}\,\boldsymbol{T}_o \tag{27}$$

where $\zeta$ is given by (26). Notice that by applying the arithmetic-geometric mean inequality, the following relation holds:

$$\begin{aligned}&J_z(\boldsymbol{T}_o)-J_z(\boldsymbol{T}^{opt})\\ &=\sum_{l=1}^{n}\alpha_l^2\boldsymbol{y}_z^H(l)\boldsymbol{T}_o\boldsymbol{T}_o^T\boldsymbol{y}_z(l)+\sum_{l=1}^{n}\beta_l^2\boldsymbol{x}_z^H(l)\boldsymbol{T}_o^{-T}\boldsymbol{T}_o^{-1}\boldsymbol{x}_z(l)\\ &\quad-2\sqrt{\sum_{l=1}^{n}\alpha_l^2\boldsymbol{y}_z^H(l)\boldsymbol{T}_o\boldsymbol{T}_o^T\boldsymbol{y}_z(l)\sum_{l=1}^{n}\beta_l^2\boldsymbol{x}_z^H(l)\boldsymbol{T}_o^{-T}\boldsymbol{T}_o^{-1}\boldsymbol{x}_z(l)}\\ &\geq 0.\end{aligned} \tag{28}$$

Therefore, as expected, by using the optimal $\boldsymbol{T}^{opt}$ instead of matrix $\boldsymbol{T}_o$ the zero sensitivity always gets reduced.

## 4. A NUMERICAL EXAMPLE

Consider a 4th-order state-space digital filter $(\boldsymbol{A},\boldsymbol{b},\boldsymbol{c},d)_4$ in (1) studied in [12]

$$\boldsymbol{A}=\begin{bmatrix}3.7183 & 1 & 0 & 0\\ -5.2153 & 0 & 1 & 0\\ 3.2689 & 0 & 0 & 1\\ -0.7724 & 0 & 0 & 0\end{bmatrix},\quad \boldsymbol{b}=10^{-3}\begin{bmatrix}0.1755\\ 0.0178\\ 0.1652\\ 0.0052\end{bmatrix}$$
$$\boldsymbol{c}=\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix},\qquad d=0.0227.$$

The eigenvalues of matrices $\boldsymbol{A}$ and $\boldsymbol{Z}$ were found to be

$$\begin{aligned}\lambda &= 0.963556\pm j0.156437,\quad 0.895594\pm j0.092081\\ v &= 1.147681\pm j0.329541,\quad 0.707604\pm j0.202980\end{aligned}$$

respectively. By using (11c), we obtained

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{bmatrix}=\begin{bmatrix}12.215615\\ 12.215615\\ 9.687213\\ 9.687213\end{bmatrix},\quad \begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{bmatrix}=\begin{bmatrix}0.386463\\ 0.386463\\ 0.310390\\ 0.310390\end{bmatrix}.$$

The original pole and zero sensitivity measures in (11a) and (11b) were computed as

$$J_p=7.209829\times 10^6,\qquad J_z=1.135788\times 10^6.$$

By choosing $\gamma=1$ in (18) and $\boldsymbol{T}^{-1}=[\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{t}_3,\boldsymbol{t}_4]=\boldsymbol{I}_4$ in (19) as an initial estimate in the iteration vector $\boldsymbol{x}_k$ of (20), it took the proposed algorithm 26 iterations to converge to

$$\boldsymbol{T}^{-1}=\begin{bmatrix}0.715371 & 0.705875 & 0.669384 & 0.604262\\ 0.711906 & 0.761567 & 0.806011 & 0.843410\\ 0.399603 & 0.506828 & 0.618995 & 0.736068\\ -0.147317 & -0.016446 & 0.120132 & 0.258514\end{bmatrix}$$

with

$$J_p(\boldsymbol{T})=4.000000,\quad J_z(\boldsymbol{T})=1.888985\times 10^8.$$

In this case, from (13) it was observed that $\overline{\boldsymbol{A}}\;\overline{\boldsymbol{A}}^T=\overline{\boldsymbol{A}}^T\overline{\boldsymbol{A}}$ holds. This implies that matrix $\overline{\boldsymbol{A}}$ is normal and the lower bound $n=4$ of $J_p(\boldsymbol{T})$ is achieved. The pole sensitivity performance of 26 iterations is shown in Fig. 1 where $\gamma=1$ and $J_\gamma(\boldsymbol{T})=J_p(\boldsymbol{T})$, from which it is seen that the iterative algorithm converges with 26 iterations.

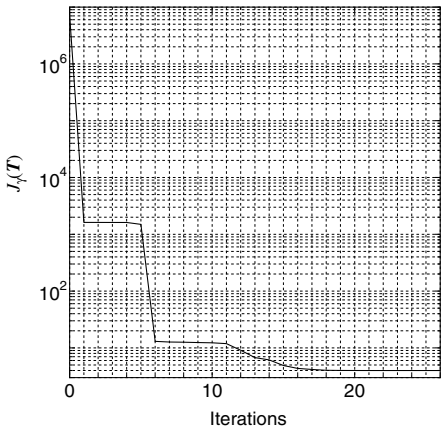


Fig. 1. Profile of $J_\gamma(\boldsymbol{T})$ during the first 26 iterations.

Since $J_p(\boldsymbol{T}) = 4$ is satisfied above, we set $\boldsymbol{T} = \boldsymbol{T}_o$. Then, from (26) and (27) it was derived that

$$\boldsymbol{T}^{opt} = \begin{bmatrix} -0.130599 & 0.337514 & -0.321396 & 0.119229 \\ 0.333112 & -0.928941 & 0.921529 & -0.371809 \\ -0.279701 & 0.858912 & -0.888673 & 0.381883 \\ 0.076747 & -0.265901 & 0.288445 & -0.129859 \end{bmatrix}$$

$$J_p(\boldsymbol{T}^{opt}) = 4.000000, \qquad J_z(\boldsymbol{T}^{opt}) = 663.119100$$

where $\zeta = 7.336475 \times 10^{-7}$. It is noted that $\boldsymbol{T}^{opt}$ obtained above corresponds to the optimal coordinate transformation matrix that minimizes a zero sensitivity measure $J_z(\pm\sqrt{\zeta}\,\boldsymbol{T}_o)$ with respect to $\zeta$ subject to minimal pole sensitivity satisfying $J_p(\boldsymbol{T}_o) = 4$.

The whole numerical results of the pole and zero sensitivity measures are summarized in Table I where

$$J_z(\boldsymbol{T}) \geq \sum_{l=1}^{4}(1 + |\alpha_l\beta_l|)^2 = 97.566165$$

always holds. From Table I, it is observed that the lower bound of $J_z(\boldsymbol{T})$ was achieved in case $\gamma = 0$.

TABLE I
PERFORMANCE COMPARISON

| $\gamma$ | $J_\gamma(\boldsymbol{T})$ | $J_p(\boldsymbol{T})$ | $J_z(\boldsymbol{T})$ | $J_p(\boldsymbol{T})+J_z(\boldsymbol{T})$ |
|---|---|---|---|---|
| 1.0 | 4.000000 | 4.000000 | $1.88898\times10^8$ | $1.888985\times10^8$ |
| 0.9 | 26.147775 | 9.786615 | 173.398216 | 183.184831 |
| 0.8 | 40.878190 | 13.926809 | 148.683716 | 162.610524 |
| 0.7 | 53.474454 | 18.020834 | 136.199568 | 154.220402 |
| 0.6 | 64.780550 | 22.716019 | 127.877347 | 150.593365 |
| 0.5 | 74.608206 | 27.598917 | 121.617495 | 149.216412 |
| 0.4 | 83.435795 | 34.001559 | 116.391953 | 150.393512 |
| 0.3 | 91.030285 | 42.698201 | 111.744036 | 154.442237 |
| 0.2 | 97.096840 | 56.211254 | 107.318237 | 163.529491 |
| 0.1 | 100.813973 | 83.685516 | 102.717135 | 186.402651 |
| 0.0 | 97.566165 | 283.698405 | 97.566165 | 381.264570 |

In addition, the performances of the proposed algorithm in terms of $J_p(\boldsymbol{T})$ and $J_z(\boldsymbol{T})$ are summarized in Table II to compare with these achieved by the method in [12] for three implementation settings. This table shows that the proposed method consistently outperforms the method reported in [12].

TABLE II
PERFORMANCE COMPARISON BETWEEN TWO METHODS

| Method | Measure | Min $J_p(\boldsymbol{T})$ w. r. t. $\boldsymbol{T}$ | Min $J_z(\boldsymbol{T})$ w. r. t. $\boldsymbol{T}$ | $\boldsymbol{T}^{opt}$ |
|---|---|---|---|---|
| Proposed | $J_p$ | 4 | 283.7 | 4 |
| | $J_z$ | $1.8890 \times 10^8$ | 97.566 | 663.12 |
| In [12] | $J_p$ | 4 | 415.8 | 4 |
| | $J_z$ | $3.7724 \times 10^8$ | 98.232 | 732.23 |

## 5. CONCLUSION

The problem of minimizing pole and zero sensitivity for state-space filters has been investigated. An efficient iterative technique for minimizing a weighted pole and zero sensitivity measure has been developed by employing a quasi-Newton algorithm. Moreover, a novel and simple method has been explored to obtain the optimal coordinate transformation matrix which minimizes a zero sensitivity measure subject to minimal pole sensitivity. Computer simulation results have demonstrated the validity and 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 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 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. ICASSP 1992*, San Francisco, CA, August 24-28, 2009, 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 systems," *IEEE Trans. Signal Processing*, vol.45, pp.837-840, April 1997.

[10] 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. II*, vol.52, pp.641-645, Oct. 2005.

[11] D. Williamson, "Roundoff noise minimization and pole-zero sensitivity in fixed-point digital filters using residue feedback," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-34, pp.1210-1220, Oct. 1986.

[12] G. Li, "On pole and zero sensitivity of linear systems," *IEEE Trans. Circuits Syst. I*, vol. 44, pp.583-590, Jul. 1997.

[13] R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Addison-Wesley Publishing Company, Inc., 1987.

[14] R. Fletcher, *Practical Methods of optimization*, 2nd ed. New York: Wiley, 1987.

[15] T. Kailath, *Linear System*. Englewood Cliffs, NJ: Prentice-Hall, 1980.