Generalized Direct-Form II Realization of 2-D Separable-Denominator Digital Filters

Takao Hinamoto
Hiroshima University
Higashi-Hiroshima 739-8527, Japan
Email: hinamoto@ieee.org

Akimitsu Doi
Hiroshima Institute of Technology
Hiroshima 731-5193, Japan
Email: doi@cc.it-hiroshima.ac.jp

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

***Abstract*—In this paper, we explore generalized direct-form II realization of two-dimensional separable-denominator digital filters of order (m, n) based on the concept of polynomial operators, where we deal with a model with $3(m+n)+mn+1$ fixed parameters plus $m+n$ free parameters. An l_2-scaling method is introduced by utilizing different coupling coefficients at different branch nodes to avoid overflow. An expression for the roundoff noise gain in the realization is also examined. It is shown that the roundoff noise gain can be minimized with respect to the $m+n$ free parameters by means of exhaustive search in a finite element space.**

I. Introduction

Delta operator is well known for satisfactory finite-word-length (FWL) performance in systems with high sampling rate and has been widely used in the realization of digital filters [1]-[5]. Potential advantages of using delta operator state-space realizations rather than shift operator realizations of transfer functions have been examined in terms of minimizing the roundoff noise [1], and the implementation of IIR digital filters using direct form delta structures has been studied with the analysis and comparison of different second-order direct-form sections in terms of the roundoff noise [2]. A modified delta transposed direct-form II second-order section has been proposed in which parameter Δ of the delta operator $\delta \equiv (z-1)/\Delta$ and filter coefficients at different branches are separately scaled to minimize the roundoff noise [3]. An arbitrary order delta operator-based direct-form IIR filter with minimum roundoff noise and sensitivity has been derived [4]. A new structure based on the concept of polynomial operators has been explored for digital filter implementation [5]. The new structure is viewded as a generalization of the shift operator z-based direct-form II transposed structure. On the other hand, two-dimensional (2-D) separable-denominator digital filters form a very popular class of multidimensional digital signal processors because of their desirable properties and trivial stability verification, and have been a subject of study in the past [6]-[12].

In this paper, motivated by [4]-[5] and based on the concept of polynomial operators, two sets of operators are introduced to express 2-D separable-denominator digital filters by a generalized SIMO direct-form II and a generalized MISO transposed direct-form II connected in cascade. This structure is then realized by a local state-space model with very sparse parameters, and l_2 scaling is performed by utilizing different coupling coefficients at different branch nodes to avoid overflow. An expression for the roundoff noise gain in the realization is investigated. It is shown that the roundoff noise gain depends on the choice of several free parameters in the realization. Moreover, in the scenario where the free parameters are subject to FWL format for actual implementation, it is shown that the roundoff noise gain can be minimized with respect to the free parameters subject to l_2-scaling constraints by an exhaustive search of a set with finite elements.

II. Realization of 2-D Digital Filters

Consider a 2-D stable separable-denominator digital filter described by

$$H(z_1,z_2)=\frac{\displaystyle\sum_{k=0}^{m}\sum_{l=0}^{n}c_{kl}z_1^{-k}z_2^{-l}}{\displaystyle\Big(1+\sum_{k=1}^{m}a_kz_1^{-k}\Big)\Big(1+\sum_{l=1}^{n}b_lz_2^{-l}\Big)} \tag{1}$$

where the denominator and numerator are assumed to be co-prime. Let $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ be $(m+1)\times(m+1)$ and $(n+1)\times(n+1)$ nonsingular matrices, respectively, and

$$\boldsymbol{q}_1(z_1)=\boldsymbol{P}_1\boldsymbol{z}_1,\qquad \boldsymbol{q}_2(z_2)=\boldsymbol{P}_2\boldsymbol{z}_2 \tag{2}$$

where

$$\begin{aligned}\boldsymbol{q}_1(z_1)&=\big[q_0^h(z_1),q_1^h(z_1),\cdots,q_m^h(z_1)\big]^T\\ \boldsymbol{q}_2(z_2)&=\big[q_0^v(z_2),q_1^v(z_2),\cdots,q_n^v(z_2)\big]^T\\ \boldsymbol{z}_1&=[z_1^m,\cdots,z_1,1]^T,\qquad \boldsymbol{z}_2=[z_2^n,\cdots,z_2,1]^T\end{aligned}$$

If scalars $\{\alpha_k|\,k=1,2,\cdots,m\}$, $\{\beta_l|\,l=1,2,\cdots,n\}$ and $\{r_{kl}|\,k=0,1,\cdots,m;\,l=0,1,\cdots,n\}$ are defined so that

$$\begin{gathered}\kappa_1\,[1\ \alpha_1\ \alpha_2\ \cdots\ \alpha_m]\,\boldsymbol{P}_1=[1\ a_1\ a_2\ \cdots\ a_m]\\ \kappa_2\,[1\ \beta_1\ \beta_2\ \cdots\ \beta_n]\,\boldsymbol{P}_2=[1\ b_1\ b_2\ \cdots\ b_n]\\ \kappa_1\kappa_2\boldsymbol{P}_1^T\begin{bmatrix}r_{00}&\cdots&r_{0n}\\ \vdots&\ddots&\vdots\\ r_{m0}&\cdots&r_{mn}\end{bmatrix}\boldsymbol{P}_2=\begin{bmatrix}c_{00}&\cdots&c_{0n}\\ \vdots&\ddots&\vdots\\ c_{m0}&\cdots&c_{mn}\end{bmatrix}\end{gathered} \tag{3}$$

the transfer function in (1) can be written as

$$H(z_1,z_2)=\frac{\displaystyle\sum_{k=0}^{m}\sum_{l=0}^{n}r_{kl}q_k^h(z_1)q_l^v(z_2)}{\displaystyle\sum_{k=0}^{m}\alpha_kq_k^h(z_1)\sum_{l=0}^{n}\beta_lq_l^v(z_2)},\quad \alpha_0=\beta_0=1 \tag{4}$$

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

where κ_1 (κ_2) is a scaling factor such that $\alpha_0 = 1$ ($\beta_0 = 1$). From (4), it is obvious that the filter is characterized by scalars $\{\alpha_k | \, k = 1, 2, \cdots, m\}$, $\{\beta_l | \, l = 1, 2, \cdots, n\}$ and $\{r_{kl} | \, k = 0, 1, \cdots, n; \; l = 0, 1, \cdots, n\}$ under the polynomial operators $\{q_k^h(z_1) | \, k = 0, 1, \cdots, m\}$ and $\{q_l^v(z_2) | \, l = 0, 1, \cdots, n\}$. We now consider two sets of polynomial operators which, as we shall see shortly, lead to an interesting structure for filter implementation. Define

$$\begin{aligned} \rho_k^h(z_1) &= \frac{z_1 - \overline{\gamma}_k}{\overline{\Delta}_k} \quad \text{for} \quad k = 1, 2, \cdots, m \\ \rho_l^v(z_2) &= \frac{z_2 - \hat{\gamma}_l}{\hat{\Delta}_l} \quad \text{for} \quad l = 1, 2, \cdots, n \end{aligned} \tag{5}$$

where $\{\overline{\gamma}_k\}$, $\{\hat{\gamma}_l\}$, $\{\overline{\Delta}_k > 0\}$ and $\{\hat{\Delta}_l > 0\}$ are four sets of constants to be discussed later. Let the polynomial operators be chosen as

$$\begin{aligned} q_k^h(z_1) &= \rho_k^h(z_1)\rho_{k+1}^h(z_1) \cdots \rho_m^h(z_1), \quad k = 0, 1, \cdots, m \\ q_l^v(z_2) &= \rho_l^v(z_2)\rho_{l+1}^v(z_2) \cdots \rho_n^v(z_2), \quad l = 0, 1, \cdots, n \\ q_m^h(z_1) &= q_n^v(z_2) = 1, \qquad \rho_0^h(z_1) = \rho_0^v(z_2) = 1 \end{aligned} \tag{6}$$

With the choice of $\{\rho_k^h(z_1) | \, k = 1, 2, \cdots, m\}$ and $\{\rho_l^v(z_2) | \, l = 1, 2, \cdots, n\}$ in (5), we can specify the corresponding transformation matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and scalars $\kappa_1 = \overline{\Delta}_1 \overline{\Delta}_2 \cdots \overline{\Delta}_m$, $\kappa_2 = \hat{\Delta}_1 \hat{\Delta}_2 \cdots \hat{\Delta}_n$. From (6), it follows that for $k, l \geq 1$,

$$\begin{aligned} \frac{q_k^h(z_1)}{q_0^h(z_1)} &= \rho_1^h(z_1)^{-1} \rho_2^h(z_1)^{-1} \cdots \rho_k^h(z_1)^{-1} \\ \frac{q_l^v(z_2)}{q_0^v(z_2)} &= \rho_1^v(z_2)^{-1} \rho_2^v(z_2)^{-1} \cdots \rho_l^v(z_2)^{-1} \end{aligned} \tag{7}$$

Hence the transfer function in (4) can be written as

$$H(z_1, z_2) = \frac{\displaystyle\sum_{k=0}^{m} \sum_{l=0}^{n} r_{kl} \prod_{p=0}^{k} \rho_p^h(z_1)^{-1} \prod_{q=0}^{l} \rho_q^v(z_2)^{-1}}{\displaystyle\Big(\sum_{k=0}^{m} \alpha_k \prod_{p=0}^{k} \rho_p^h(z_1)^{-1}\Big)\Big(\sum_{l=0}^{n} \beta_l \prod_{q=0}^{l} \rho_q^v(z_2)^{-1}\Big)} \tag{8}$$

The implementations of $\rho_k^h(z_1)^{-1}$ and $\rho_l^v(z_2)^{-1}$ are depicted in Fig. 1. As an illustrative example, referring to shift-operator structure of a 2-D separable-denominator filter with $(m, n) = (3, 2)$ shown in [13] (p. 956), the generalized filter structure in (8) with $(m, n) = (3, 3)$ is illustrated in Fig. 2 where $u(i, j)$ is a scalar input and $y(i, j)$ is a scalar output.

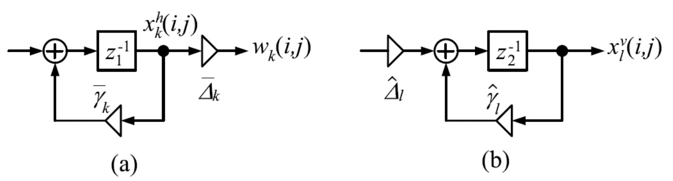

Fig. 1. (a) Implementation of $\rho_k^h(z_1)^{-1}$ and (b) that of $\rho_l^v(z_2)^{-1}$.

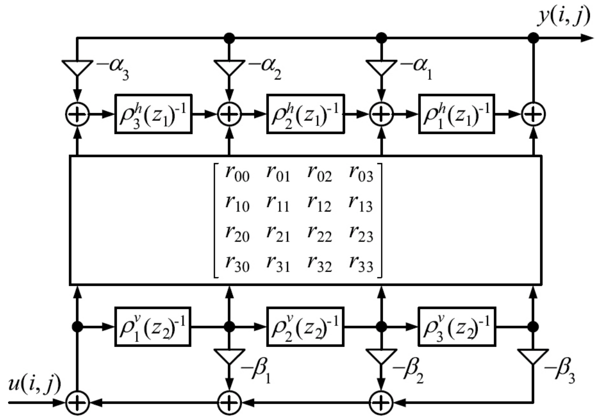

Fig. 2. The 2-D filter structure of (8) with $(m, n) = (3, 3)$.

From Figures 1 and 2, we obtain

$$\begin{aligned} y(i,j) &= w_1(i,j) + r_{00}\Big[u(i,j) - \sum_{l=1}^{n} \beta_l x_l^v(i,j)\Big] \\ &\quad + \sum_{l=1}^{n} r_{0l} x_l^v(i,j) \\ w_k(i,j) &= \rho_k^h(z_1)^{-1}\Big\{w_{k+1}(i,j) + \sum_{l=1}^{n} r_{kl} x_l^v(i,j) \\ &\quad + r_{k0}\Big[u(i,j) - \sum_{l=1}^{n} \beta_l x_l^v(i,j)\Big] - \alpha_k y(i,j)\Big\} \end{aligned} \tag{9}$$

for $k = 1, 2, \cdots, m$ where $w_{m+1}(i, j) = 0$. In addtion,

$$\begin{aligned} x_1^v(i, j+1) &= \hat{\gamma}_1 x_1^v(i,j) + \hat{\Delta}_1\Big[u(i,j) - \sum_{l=1}^{n} \beta_l x_l^v(i,j)\Big] \\ x_l^v(i, j+1) &= \hat{\Delta}_l x_{l-1}^v(i,j) + \hat{\gamma}_l x_l^v(i,j) \end{aligned} \tag{10}$$

for $l = 2, 3, \cdots, n$.

From (5), the second equation in (9) can be written as

$$\begin{aligned} x_k^h(i+1, j) &= -\alpha_k \overline{\Delta}_1 x_1^h(i,j) + \overline{\Delta}_{k+1} x_{k+1}^h(i,j) + \overline{\gamma}_k x_k^h(i,j) \\ &\quad + \sum_{l=1}^{n} \big(r_{kl} - \alpha_k r_{0l} - r_{k0}\beta_l + r_{00}\alpha_k\beta_l\big) x_l^v(i,j) \\ &\quad + (r_{k0} - r_{00}\alpha_k) u(i,j) \end{aligned} \tag{11}$$

for $k = 1, 2, \cdots, m$ where $x_{m+1}^h(i, j) = 0$.

From (9)-(11), it follows that the filter $H(z_1, z_2)$ in (8) can be realized by the Roesser local state-space model as

$$\begin{aligned} \begin{bmatrix} \boldsymbol{x}^h(i+1, j) \\ \boldsymbol{x}^v(i, j+1) \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{0} & \boldsymbol{A}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^h(i,j) \\ \boldsymbol{x}^v(i,j) \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} u(i,j) \\ y(i,j) &= [\; \boldsymbol{c}_1 \quad \boldsymbol{c}_2 \;] \begin{bmatrix} \boldsymbol{x}^h(i,j) \\ \boldsymbol{x}^v(i,j) \end{bmatrix} + d\, u(i,j) \end{aligned} \tag{12}$$

where

$$\begin{aligned} \boldsymbol{x}^h(i,j) &= [\; x_1^h(i,j), \quad x_2^h(i,j), \quad \cdots, \quad x_m^h(i,j) \;]^T \\ \boldsymbol{x}^v(i,j) &= [\; x_1^v(i,j), \quad x_2^v(i,j), \quad \cdots, \quad x_n^v(i,j) \;]^T \end{aligned}$$

$$
\boldsymbol{A}_1 = \begin{bmatrix} -\alpha_1\overline{\Delta}_1 & \overline{\Delta}_2 & \cdots & 0 \\ -\alpha_2\overline{\Delta}_1 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \overline{\Delta}_m \\ -\alpha_m\overline{\Delta}_1 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} \overline{\gamma}_1 & 0 & \cdots & 0 \\ 0 & \overline{\gamma}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \overline{\gamma}_m \end{bmatrix}
$$

$$
\boldsymbol{A}_2 = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} - \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} [r_{01} \;\; r_{02} \;\; \cdots \;\; r_{0n}]
$$

$$
- \begin{bmatrix} r_{10} \\ r_{20} \\ \vdots \\ r_{m0} \end{bmatrix} [\beta_1 \; \beta_2 \cdots \beta_n] + r_{00} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} [\beta_1 \; \beta_2 \cdots \beta_n]
$$

$$
\boldsymbol{A}_4 = \begin{bmatrix} -\beta_1\hat{\Delta}_1 & -\beta_2\hat{\Delta}_1 & \cdots & -\beta_n\hat{\Delta}_1 \\ \hat{\Delta}_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \hat{\Delta}_n & 0 \end{bmatrix} + \begin{bmatrix} \hat{\gamma}_1 & 0 & \cdots & 0 \\ 0 & \hat{\gamma}_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \hat{\gamma}_n \end{bmatrix}
$$

$$
\boldsymbol{b}_1 = \begin{bmatrix} r_{10} \\ r_{20} \\ \vdots \\ r_{m0} \end{bmatrix} - r_{00} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}, \qquad \boldsymbol{b}_2 = \begin{bmatrix} \hat{\Delta}_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

$$
\boldsymbol{c}_1 = [\; \overline{\Delta}_1 \;\; 0 \;\; \cdots \;\; 0 \;]
$$

$$
\boldsymbol{c}_2 = [\, r_{01} \;\; r_{02} \;\; \cdots \;\; r_{0n} \,] - r_{00}\, [\, \beta_1 \; \beta_2 \cdots \beta_n \,], \qquad d = r_{00}
$$

We now consider an equivalent state-space realization with $\overline{\Delta}_k = 1$ for $k = 1, 2, \cdots, m$ and $\hat{\Delta}_l = 1$ for $l = 1, 2, \cdots, n$ in (5). Then the other realization is obtained as

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

$$
y(i, j) = [\; \tilde{\boldsymbol{c}}_1 \;\; \tilde{\boldsymbol{c}}_2 \;] \begin{bmatrix} \tilde{\boldsymbol{x}}^h(i, j) \\ \tilde{\boldsymbol{x}}^v(i, j) \end{bmatrix} + \tilde{d}u(i, j) \tag{13}
$$

where

$$
\tilde{\boldsymbol{A}}_1 = \begin{bmatrix} -\tilde{\alpha}_1 & & & \\ -\tilde{\alpha}_2 & & \boldsymbol{I}_{m-1} & \\ \vdots & & & \\ -\tilde{\alpha}_m & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} \overline{\gamma}_1 & 0 & \cdots & 0 \\ 0 & \overline{\gamma}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \overline{\gamma}_m \end{bmatrix}
$$

$$
\tilde{\boldsymbol{A}}_4 = \begin{bmatrix} -\tilde{\beta}_1 & -\tilde{\beta}_2 & \cdots & -\tilde{\beta}_n \\ & & & 0 \\ & \boldsymbol{I}_{n-1} & & \vdots \\ & & & 0 \end{bmatrix} + \begin{bmatrix} \hat{\gamma}_1 & 0 & \cdots & 0 \\ 0 & \hat{\gamma}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\gamma}_n \end{bmatrix}
$$

$$
\tilde{\boldsymbol{c}}_1 = [\; 1 \;\; 0 \;\; \cdots \;\; 0 \;], \qquad \tilde{\boldsymbol{b}}_2 = [\; 1 \;\; 0 \;\; \cdots \;\; 0 \;]^T
$$

and coefficient matrices $\tilde{\boldsymbol{A}}_2$, $\tilde{\boldsymbol{b}}_1$, $\tilde{\boldsymbol{c}}_2$ and $\tilde{d}$ can be derived from $\{\tilde{r}_{kl} | \, k = 0, 1, \cdots, m; \, l = 0, 1, \cdots, n\}$, $\{\tilde{\alpha}_k | \, k = 1, 2, \cdots, m\}$ and $\{\tilde{\beta}_l | \, l = 1, 2, \cdots, n\}$ by the same manner as in (12).

If the local state-space model in (12) is related to the model in (13) by

$$
\begin{bmatrix} \tilde{\boldsymbol{x}}^h(i, j) \\ \tilde{\boldsymbol{x}}^v(i, j) \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_4 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}^h(i, j) \\ \boldsymbol{x}^v(i, j) \end{bmatrix} \tag{14}
$$

with

$$
\boldsymbol{T}_1 = \mathrm{diag}\{\bar{t}_1 \; \bar{t}_2 \; \cdots \; \bar{t}_m\}, \qquad \boldsymbol{T}_4 = \mathrm{diag}\{\hat{t}_1 \; \hat{t}_2 \; \cdots \; \hat{t}_n\}
$$

we obtain

$$
\begin{aligned}
\tilde{\boldsymbol{A}}_1 &= \boldsymbol{T}_1^{-1}\boldsymbol{A}_1\boldsymbol{T}_1, & \tilde{\boldsymbol{b}}_1 &= \boldsymbol{T}_1^{-1}\boldsymbol{b}_1, & \tilde{\boldsymbol{c}}_1 &= \boldsymbol{c}_1\boldsymbol{T}_1 \\
\tilde{\boldsymbol{A}}_4 &= \boldsymbol{T}_4^{-1}\boldsymbol{A}_4\boldsymbol{T}_4, & \tilde{\boldsymbol{b}}_2 &= \boldsymbol{T}_4^{-1}\boldsymbol{b}_2, & \tilde{\boldsymbol{c}}_2 &= \boldsymbol{c}_2\boldsymbol{T}_4 \\
\tilde{\boldsymbol{A}}_2 &= \boldsymbol{T}_1^{-1}\boldsymbol{A}_2\boldsymbol{T}_4, & \tilde{d} &= d & &
\end{aligned} \tag{15}
$$

By relations $\tilde{\boldsymbol{c}}_1 = \boldsymbol{c}_1\boldsymbol{T}_1$, $\boldsymbol{T}_1\tilde{\boldsymbol{A}}_1 = \boldsymbol{A}_1\boldsymbol{T}_1$, $\tilde{\boldsymbol{A}}_4\boldsymbol{T}_4^{-1} = \boldsymbol{T}_4^{-1}\boldsymbol{A}_4$ and $\tilde{\boldsymbol{b}}_2 = \boldsymbol{T}_4^{-1}\boldsymbol{b}_2$ it follows that

$$
\begin{aligned}
\boldsymbol{T}_1^{-1} &= \mathrm{diag}\big\{\overline{\Delta}_1, \; \overline{\Delta}_1\overline{\Delta}_2, \; \cdots, \; \overline{\Delta}_1\overline{\Delta}_2\cdots\overline{\Delta}_m\big\} \\
\boldsymbol{T}_4 &= \mathrm{diag}\big\{\hat{\Delta}_1, \; \hat{\Delta}_1\hat{\Delta}_2, \; \cdots, \; \hat{\Delta}_1\hat{\Delta}_2\cdots\hat{\Delta}_n\big\}
\end{aligned}
$$

$$
\begin{bmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_m \end{bmatrix} = \boldsymbol{T}_1^{-1} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}, \qquad \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_n \end{bmatrix} = \boldsymbol{T}_4^T \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}
$$

$$
\begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1n} \\ \tilde{r}_{21} & \tilde{r}_{22} & \cdots & \tilde{r}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{r}_{m1} & \tilde{r}_{m2} & \cdots & \tilde{r}_{mn} \end{bmatrix} = \boldsymbol{T}_1^{-1} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \boldsymbol{T}_4
$$

$$
\begin{aligned}
&[\, \tilde{r}_{01} \;\; \tilde{r}_{02} \;\; \cdots \;\; \tilde{r}_{0n} \,] = [\, r_{01} \;\; r_{02} \;\; \cdots \;\; r_{0n} \,]\, \boldsymbol{T}_4 \\
&[\, \tilde{r}_{10} \;\; \tilde{r}_{20} \;\; \cdots \;\; \tilde{r}_{m0} \,]^T = \boldsymbol{T}_1^{-1} [\, r_{10} \;\; r_{20} \;\; \cdots \;\; r_{m0} \,]^T
\end{aligned} \tag{16}
$$

because of $\tilde{\boldsymbol{A}}_2 = \boldsymbol{T}_1^{-1}\boldsymbol{A}_2\boldsymbol{T}_4$.

III. MINIMIZATION OF ROUNDOFF NOISE

A. l_2-Scaling Constraints

The horizontal and vertical controllability Grammians $\tilde{\boldsymbol{K}}^h$ and $\tilde{\boldsymbol{K}}^v$ of the local state-space model in (13) play an important role in the dynamic-range scaling of the local state vectors $\boldsymbol{x}^h(i, j)$ and $\boldsymbol{x}^v(i, j)$ in (12), and matrices $\tilde{\boldsymbol{K}}^h$ and $\tilde{\boldsymbol{K}}^v$ can be obtained by solving the Lyapunov equations [9]

$$
\begin{aligned}
\tilde{\boldsymbol{K}}^h &= \tilde{\boldsymbol{A}}_1\tilde{\boldsymbol{K}}^h\tilde{\boldsymbol{A}}_1^T + \tilde{\boldsymbol{A}}_2\tilde{\boldsymbol{K}}^v\tilde{\boldsymbol{A}}_2^T + \tilde{\boldsymbol{b}}_1\tilde{\boldsymbol{b}}_1^T \\
\tilde{\boldsymbol{K}}^v &= \tilde{\boldsymbol{A}}_4\tilde{\boldsymbol{K}}^v\tilde{\boldsymbol{A}}_4^T + \tilde{\boldsymbol{b}}_2\tilde{\boldsymbol{b}}_2^T
\end{aligned} \tag{17}
$$

respectively. With an equivalent realization specified in (15), the horizontal and vertical controllability Grammians $\boldsymbol{K}^h$ and $\boldsymbol{K}^v$ of the local state-space model in (12) assumes the form

$$
\begin{aligned}
\boldsymbol{K}^h &= \boldsymbol{T}_1\tilde{\boldsymbol{K}}^h\boldsymbol{T}_1^T, & \boldsymbol{T}_1 &= \mathrm{diag}\{\bar{t}_1, \; \bar{t}_2, \; \cdots, \; \bar{t}_m\} \\
\boldsymbol{K}^v &= \boldsymbol{T}_4\tilde{\boldsymbol{K}}^v\boldsymbol{T}_4^T, & \boldsymbol{T}_4 &= \mathrm{diag}\{\hat{t}_1, \; \hat{t}_2, \; \cdots, \; \hat{t}_n\}
\end{aligned} \tag{18}
$$

If l_2-scaling constraints are imposed on the horizontal and vertical state vectors $\boldsymbol{x}^h(i, j)$ and $\boldsymbol{x}^v(i, j)$ in (12), the horizontal

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

and vertical controllability Grammians $\boldsymbol{K}^h$ and $\boldsymbol{K}^v$ of the local state-space model in (12) are subject to the constraints

$$\begin{aligned}\overline{\boldsymbol{e}}_k^T \boldsymbol{K}^h \overline{\boldsymbol{e}}_k &= \overline{\boldsymbol{e}}_k^T \boldsymbol{T}_1 \tilde{\boldsymbol{K}}^h \boldsymbol{T}_1^T \overline{\boldsymbol{e}}_k = \bar{t}_k^2 \overline{\boldsymbol{e}}_k^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_k = 1 \\ &\text{for} \quad k = 1, 2, \cdots, m \\ \hat{\boldsymbol{e}}_l^T \boldsymbol{K}^v \hat{\boldsymbol{e}}_l &= \hat{\boldsymbol{e}}_l^T \boldsymbol{T}_4 \tilde{\boldsymbol{K}}^v \boldsymbol{T}_4^T \hat{\boldsymbol{e}}_l = \hat{t}_l^2 \hat{\boldsymbol{e}}_l^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_l = 1 \\ &\text{for} \quad l = 1, 2, \cdots, n\end{aligned} \tag{19}$$

where $\overline{\boldsymbol{e}}_k$ ($\hat{\boldsymbol{e}}_l$) indicates the kth (lth) column of $m \times m$ ($n \times n$) identity matrix $\boldsymbol{I}_m$ ($\boldsymbol{I}_n$).

Consequently, by noting that $\bar{t}_k = \overline{\Delta}_1^{-1} \overline{\Delta}_2^{-1} \cdots \overline{\Delta}_k^{-1}$ for $k = 1, 2, \cdots, m$ and $\hat{t}_l = \hat{\Delta}_1 \hat{\Delta}_2 \cdots \hat{\Delta}_l$ for $l = 1, 2, \cdots, n$, we obtain

$$\begin{aligned}&\overline{\Delta}_1 = \sqrt{\overline{\boldsymbol{e}}_1^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_1}, \quad \overline{\Delta}_2 = \sqrt{\frac{\overline{\boldsymbol{e}}_2^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_2}{\overline{\boldsymbol{e}}_1^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_1}}, \quad \cdots, \\ &\overline{\Delta}_m = \sqrt{\frac{\overline{\boldsymbol{e}}_m^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_m}{\overline{\boldsymbol{e}}_{m-1}^T \tilde{\boldsymbol{K}}^h \overline{\boldsymbol{e}}_{m-1}}}, \qquad \hat{\Delta}_1 = \frac{1}{\sqrt{\hat{\boldsymbol{e}}_1^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_1}} \\ &\hat{\Delta}_2 = \sqrt{\frac{\hat{\boldsymbol{e}}_1^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_1}{\hat{\boldsymbol{e}}_2^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_2}}, \quad \cdots, \quad \hat{\Delta}_n = \sqrt{\frac{\hat{\boldsymbol{e}}_{n-1}^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_{n-1}}{\hat{\boldsymbol{e}}_n^T \tilde{\boldsymbol{K}}^v \hat{\boldsymbol{e}}_n}}\end{aligned} \tag{20}$$

The parameters $\{\overline{\Delta}_k\}$ and $\{\hat{\Delta}_l\}$ are called the *coupling coefficients* [4]. It is clear from the above analysis that we can avoid overflow oscillations if these coupling coefficients are used to scale the state variables.

B. Roundoff Noise Analysis

It is known that the roundoff noise gain due to product quantization associated with the coefficient matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_4$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ and d in (12) can be expressed as [9]

$$J_{S\rho} = \text{tr}[\boldsymbol{Q}^h \boldsymbol{W}^h] + \text{tr}[\boldsymbol{Q}^v \boldsymbol{W}^v] + q \tag{21a}$$

where $\boldsymbol{W}^h$ and $\boldsymbol{W}^v$ are the horizontal and vertical observability Grammians of the local state-space model in (12), which can be obtained by solving the Lyapunov equations [9]

$$\begin{aligned}\boldsymbol{W}^h &= \boldsymbol{A}_1^T \boldsymbol{W}^h \boldsymbol{A}_1 + \boldsymbol{c}_1^T \boldsymbol{c}_1 \\ \boldsymbol{W}^v &= \boldsymbol{A}_4^T \boldsymbol{W}^v \boldsymbol{A}_4 + \boldsymbol{A}_2^T \boldsymbol{W}^h \boldsymbol{A}_2 + \boldsymbol{c}_2^T \boldsymbol{c}_2\end{aligned} \tag{21b}$$

respectively, $\boldsymbol{Q}^h$ ($\boldsymbol{Q}^v$) is a diagonal matrix whose kth (lth) diagonal element q_k^h (q_l^v) is the number of coefficients in the kth (lth) row of $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{b}_1$ ($\boldsymbol{A}_4$ and $\boldsymbol{b}_2$) that are neither 0 nor ± 1, and q is the number of neither 0 nor ± 1 constant in $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ and d.

If (21a) is applied to the state-space realization in (12), the corresponding roundoff noise gain $J_{S\rho}$ is characterized by

$$\begin{aligned}&q_1^h = n + 3, \quad q_k^h = n + 3 + \psi(\overline{\gamma}_k) \text{ for } k = 2, 3, \cdots, m-1 \\ &q_m^h = n + 2 + \psi(\overline{\gamma}_m), \qquad\qquad q = n + 2 \\ &q_1^v = n + 1, \quad q_l^v = 1 + \psi(\hat{\gamma}_l) \quad \text{for} \quad l = 2, 3, \cdots, n\end{aligned} \tag{22}$$

where

$$\psi(\gamma) = \begin{cases} 1 & \text{for} \quad \gamma \neq 0, \pm 1 \\ 0 & \text{for} \quad \gamma = 0, \pm 1 \end{cases}$$

Substituting (22) into (21a) yields

$$\begin{aligned}J_{S\rho} = &(n+3)\,\text{tr}[\boldsymbol{W}^h] - \overline{\boldsymbol{e}}_m^T \boldsymbol{W}^h \overline{\boldsymbol{e}}_m \\ &+ \sum_{k=2}^{m} \psi(\overline{\gamma}_k) \overline{\boldsymbol{e}}_k^T \boldsymbol{W}^h \overline{\boldsymbol{e}}_k + n\, \hat{\boldsymbol{e}}_1^T \boldsymbol{W}^v \hat{\boldsymbol{e}}_1 \\ &+ \text{tr}[\boldsymbol{W}^v] + \sum_{l=2}^{n} \psi(\hat{\gamma}_l) \hat{\boldsymbol{e}}_l^T \boldsymbol{W}^v \hat{\boldsymbol{e}}_l + n + 2\end{aligned} \tag{23}$$

C. Choosing Parameters $\{\overline{\gamma}_i\}$ and $\{\hat{\gamma}_l\}$

We now assume that $|\overline{\gamma}_k| \leq 1$ for $k = 1, 2, \cdots, m$ and $|\hat{\gamma}_l| \leq 1$ for $l = 1, 2, \cdots, n$. For a fixed-point implementation of B_c bits, every $\overline{\gamma}_i$ ($\hat{\gamma}_l$) must be truncated or rounded into a B_γ-bit number ($B_\gamma \leq B_c$) of the form

$$v = \pm \sum_{p=1}^{B_\gamma} b_p 2^{-p}, \quad b_p = 0, 1 \ \ \forall\ p \tag{24}$$

unless $v = \pm 1$. Hence $\{\overline{\gamma}_i\}$ and $\{\hat{\gamma}_l\}$ should take values within a *discrete space* defined by

$$S_\gamma = \{-1, 1\} \cup \left\{ \pm \sum_{p=1}^{B_\gamma} b_p 2^{-p}, \quad b_p = 0, 1 \ \ \forall\ p \right\} \tag{25}$$

which contains $(2^{B_\gamma + 1} + 1)$ elements.

D. Search of Optimal Vector $\boldsymbol{\gamma} = [\overline{\gamma}_1, \cdots, \overline{\gamma}_m, \hat{\gamma}_1, \cdots, \hat{\gamma}_n]^T$

Let $\tilde{S}_{\boldsymbol{\gamma}} \in \boldsymbol{R}^{(m+n)\times 1}$ be the space in which $\boldsymbol{\gamma}$ resides, i.e.,

$$\tilde{S}_{\boldsymbol{\gamma}} = \{\boldsymbol{\gamma} \,|\, \gamma_k \in S_\gamma \ \ \forall\ k\} \tag{26}$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_{m+n}]^T$. It is obvious that $J_{S\rho}$ in (23) is a function of vector $\boldsymbol{\gamma}$ because $\boldsymbol{W}^h$ and $\boldsymbol{W}^v$ depend on $\boldsymbol{\gamma}$. Consequently, the problem of minimizing $J_{S\rho}$ with respect to vector $\boldsymbol{\gamma}$ subject to l_2-scaling constraints can be considered as

$$\boldsymbol{\gamma}\big(J_{S\rho}^{opt}\big) = \arg \min_{\boldsymbol{\gamma} \in \tilde{S}_{\boldsymbol{\gamma}}} J_{S\rho} \quad \text{subject to} \quad (19) \tag{27}$$

The roundoff noise gain $J_{S\rho}$ in (23) can be minimized with respect to the $m + n$ free parameters $\{\overline{\gamma}_k\}$ and $\{\hat{\gamma}_l\}$ subject to l_2-scaling constraints in (19) through exhaustive search in a finite element space by applying the similar manner as in [5].

IV. Conclusion

Based on the concept of polynomial operators, generalized direct-form II realization for 2-D separable-denominator digital filters has been proposed. We have expressed 2-D separable in denominator digital filters by a generalized SIMO direct-form II and a generalized MISO transposed direct-form II connected in cascade. An l_2-scaling method has been introduced by using different coupling coefficients at different branch nodes to avoid overflow. An expression for the roundoff noise gain in the realization has been investigated. It has been shown that the roundoff noise gain can be minimized with respect to the free parameters subject to l_2-scaling constraints through exhaustive search in a finite element space. Numerical experiments have been performed satisfactorily. For the resulting realization, the analysis of l_2-sensitivity and its minimization with respect to the free parameters subject to l_2-scaling constraints will appear elsewhere.

REFERENCES

[1] G. Li and M. Gevers, "Roundoff Noise minimization using delta-operator realizations," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 629-637, Feb. 1993.

[2] J. Kauraniemi, T. I. Laakso, I. Hartimo and S. J. Ovaska, "Delta operator realizations of direct-form IIR filters," *IEEE Trans. Circuits Syst.-II*, vol. 45, no. 1, pp. 41-52, Jan. 1998.

[3] N. Wong and T.-S. Ng, "Roundoff noise minimization in a modified direct-form delta operator IIR structure," *IEEE Trans. Circuits Syst.-II*, vol. 47, no. 12, pp. 1533-1536, Dec. 2000.

[4] N. Wong and T.-S. Ng, "A generalized direct-form delta operator-based IIR filter with minimum noise gain and sensitivity," *IEEE Trans. Circuits Syst.-II*, vol. 48, no. 4, pp. 425-431, Apr. 2001.

[5] G. Li and Z. Zhao, "On the generalized DFIIt structure and its state-space realization in digital filter implementation," *IEEE Trans. Circuits Syst.-I*, vol. 51, no. 4, pp. 769-778, Apr. 2004.

[6] T. Hinamoto and F. W. Fairman, "Separable-denominator state-space realization of two-dimensional filters using a canonic form," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 4, pp. 846-853, Aug. 1981.

[7] P. K. Rajan and M. N. S. Swamy, "Design of separable denominator 2-dimensional digital filters possessing real circularly symmetric frequency responses," *Proc. Inst. Elect. Eng. G*, vol. 129, pp. 235-240, Oct. 1982.

[8] T. Hinamoto, "Design of 2-D separable-denominator recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-31, no. 11, pp. 925-932, Nov. 1984.

[9] M. Kawamata and T. Higuchi, "Synthesis of 2-D separable denominator digital filters with minimum roundoff noise and no overflow oscillations," *IEEE Trans. Circuits Syst.*, vol. CAS-33, no. 4, pp. 365-372, Apr. 1986.

[10] H. K. Kwan and C. L. Chan, "Design of linear phase circularly symmetric two-dimensional recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. 36, no. 7, pp. 1023-1029, July 1989.

[11] W.-S. Lu, H.-P. Wang and A. Antoniou, "Design and two-dimensional digital filters using singular-value decomposition and balanced approximation method," *IEEE Trans. Signal Process.*, vol. 93, no. 10, pp. 2253-2262, Oct. 1991.

[12] C. Xiao and P. Agathoklis, "Design and implementation of approximately linear phase two-dimensional IIR filters," *IEEE Trans. Circuits Syst.-II*, vol. 45, no. 9, pp. 1279-1288, Sep. 1998.

[13] S.-Y. Kung, B. C. Levy, M. Morf and T. Kailath, "New results in 2-D systems theory, Part II: 2-D state-space models—realization and the notions of controllability, observability, and minimality," *Proc. IEEE*, vol. 65, no. 6, pp. 945-961, June 1977.