第15卷 第3期 电子科学学刊 Vol. 15, No. 3
1993 年 5 月 JOURNAL OF ELECTRONICS May, 1993

一般二维数字系统的低噪声优化实现*

肖承山 冯振明
(清华大学电子工程系，北京 100084)

摘要 本文给出了一般二维数字系统的两种高效率状态空间结构；导出了这两种结构的能控性和能观性格拉姆矩阵的有关性质；建立了二维数字系统定点运算的全局噪声模型；据此获得了一般二维数字系统低噪声、高效率的优化实现. 文中举例说明了本文的两种结构的噪声性能和计算效率各不相同.

关键词 二维数字系统；状态空间实现；舍入噪声；优化

一、引 言

数字系统的舍入噪声是衡量该系统性能的重要技术指标. 由于采用不同的状态空间结构实现同一传输函数的数字系统，其噪声性能有很大的差异，有的结构甚至不能正常工作. 因此，如何寻找低噪声的状态空间结构，引起了许多学者的关注.

1985 年，Mertzios[1] 率先采用 Roesser 模型研究二维数字系统的舍入噪声，但其结论有严重错误[2]，于是 Lin 等人[3]和 Lu 等人[4]分别对二维系统的 Roesser 模型作了进一步的研究，建立了只考虑乘积舍入噪声的噪声模型，定义了能控性和能观性格拉姆矩阵(又叫协方差矩阵和单位噪声矩阵)，讨论了低噪声的优化算法. 但他们的研究有 3 个局限：(1)噪声模型没有考虑系数舍入噪声和输入信号舍入噪声；(2)对协方差矩阵和单位噪声矩阵的性质缺乏研究；(3)低噪声优化算法只适用于具有最小状态空间结构的二维系统，而且优化结构的工作效率最低.

在二维数字系统的优化实现中，还存在一个基本问题，即对于给定的传输函数，如何用最小状态空间实现. 目前对于一些特殊情形的二维系统，已找出了最小状态空间实现[5,6]. 对于一般的二维系统 Kung 等人[7]认为，在实数域内，不存在最小状态空间实现. 于是寻找性能好效率高的状态空间实现是十分有意义的. 张宏科等人[8]曾用梅森规则讨论二维系统的实现问题，但他们所得的状态空间只是一种较小的状态空间，而不是最小的状态空间. 顺便指出，该文公式有近十处印刷错误.

本文将讨论一般二维数字系统的高效率状态空间实现及其对应的两种格拉姆矩阵的有关性质，建立包括乘积舍入噪声、系数舍入噪声和输入信号舍入噪声在内的全局的噪声

1992.03.05 收到 1992.08.20 定稿.
* 国家自然科学青年基金和清华大学科研基金资助课题.
肖承山 男，1965 年生，讲师，现主要从事多维系统理论和多维数字滤波器的设计、状态空间实现和优化的研究工作.
冯振明 男，1946 年生，副教授，现主要从事模式识别和雷达杂波抑制等研究工作.

模型，并探讨一般二维系统的低噪声高效率优化实现的方法。

二、高效率的状态空间结构

设二维 (m,n) 阶系统的传输函数为[7]：

$$H(z_1,\ z_2)=\frac{\sum_{i=0}^{m}\sum_{j=0}^{n}b_{ij}z_1^{-i}z_2^{-j}}{\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}z_1^{-i}z_2^{-j}} \tag{1}$$

其中 $a_{ij}\in R$，$b_{ij}\in R$，$a_{00}=1$。

Kung 等人[7]给出了较小的状态空间结构，对该结构应用系统理论中的对偶原理，可得更高效率的新结构，其信号流图如图 1 所示。

根据图 1，可建立 Roesser 模型的状态方程和输出方程：

$$\begin{bmatrix}\mathbf{X}_{v_1}(i,\ j+1)\\ \mathbf{X}_{v_2}(i,\ j+1)\\ \mathbf{X}_h(i+1,\ j)\end{bmatrix}=A\begin{bmatrix}\mathbf{X}_{v_1}(i,\ j)\\ \mathbf{X}_{v_2}(i,\ j)\\ \mathbf{X}_h(i,\ j)\end{bmatrix}+Bu(i,j) \tag{2}$$

$$y(i,j)=C\begin{bmatrix}\mathbf{X}_{v_1}(i,\ j)\\ \mathbf{X}_{v_2}(i,\ j)\\ \mathbf{X}_h(i,\ j)\end{bmatrix}+Du(i,j) \tag{3}$$

式中 $\mathbf{X}_{v_1}\in R^m$，$\mathbf{X}_{v_2}\in R^n$，$\mathbf{X}_h\in R^m$

$$A=\begin{bmatrix}
\begin{matrix}0&&0\\1&\ddots&\\&\ddots&0\\0&&\end{matrix} & 0 & 0\\
b_{01}\cdots\cdots b_{0n} & -a_{01}\cdots\cdots -a_{0n} & \begin{matrix}1\\0\\\vdots\\0\end{matrix}\quad \begin{matrix}0\cdots\cdots 0\\0\end{matrix}\\
\begin{matrix}0\\\end{matrix} & \begin{matrix}1&&0\\&\ddots&\\0&&1\ \ 0\end{matrix} & \\
\tilde{b}_{11}\cdots\cdots\tilde{b}_{1n} & -\tilde{a}_{11}\cdots\cdots-\tilde{a}_{1n} & -a_{10}\quad \begin{matrix}1&&0\\&\ddots&\\0&&1\\0&&\end{matrix}\\
\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \\
\tilde{b}_{m1}\cdots\cdots\tilde{b}_{mn} & -\tilde{a}_{m1}\cdots\cdots-\tilde{a}_{mn} & -a_{m0}
\end{bmatrix} \tag{4}$$

$$B=[\,1\ \ 0\cdots 0\ \vdots\ b_{00}\ \ 0\cdots 0\ \vdots\ \tilde{b}_{10}\cdots\cdots\tilde{b}_{m0}\,] \tag{5}$$

$$C=[\,b_{01}\cdots\cdots b_{0n}\ \vdots\ -a_{01}\cdots\cdots -a_{0n}\ \vdots\ 1\ \ 0\cdots 0\,] \tag{6}$$

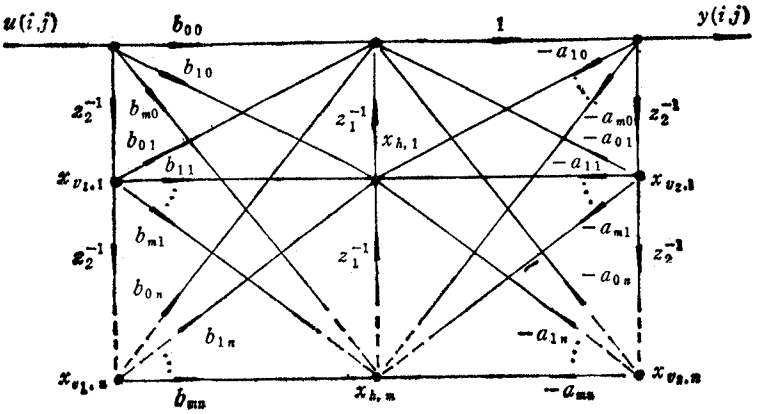


图 1　$H(z_1, z_2)$ 的一种信号流图

$$D = b_{00}$$

其中 $\tilde{b}_{ij} = b_{ij} - a_{i0}b_{0j}$，$\tilde{a}_{ij} = a_{ij} - a_{i0}b_{0j}$。

由于二维数字系统的传输函数可改写为

$$H(z_1, z_2) = \frac{\sum_{j=0}^{n}\sum_{i=0}^{m} b_{ji} z_2^{-j} z_1^{-i}}{\sum_{j=0}^{n}\sum_{i=0}^{m} a_{ji} z_2^{-j} z_1^{-i}} \tag{7}$$

所以该系统还有另一高效率状态空间结构为

$$\begin{bmatrix} \mathbf{X}_{h_1}(i+1, j) \\ \mathbf{X}_{h_2}(i+1, j) \\ \mathbf{X}_v(i, j+1) \end{bmatrix} = A \begin{bmatrix} \mathbf{X}_{h_1}(i, j) \\ \mathbf{X}_{h_2}(i, j) \\ \mathbf{X}_v(i, j) \end{bmatrix} + Bu(i,j) \tag{8}$$

$$y(i, j) = C \begin{bmatrix} \mathbf{X}_{h_1}(i, j) \\ \mathbf{X}_{h_2}(i, j) \\ \mathbf{X}_v(i, j) \end{bmatrix} + Du(i,j) \tag{9}$$

式中

$$A = \begin{bmatrix} \begin{matrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & 1 & 0 \end{matrix} & 0 & 0 \\ \begin{matrix} b_{10} & \cdots & b_{m0} \\ & 0 & \end{matrix} & \begin{matrix} -a_{10} & \cdots & -a_{m0} \\ 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} & \begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{matrix} \\ \begin{matrix} \tilde{b}_{11} & \cdots & \tilde{b}_{m1} \\ \cdots & & \cdots \\ \tilde{b}_{1n} & \cdots & \tilde{b}_{mn} \end{matrix} & \begin{matrix} -\tilde{a}_{11} & \cdots & -\tilde{a}_{m1} \\ \cdots & & \cdots \\ -\tilde{a}_{1n} & \cdots & -\tilde{a}_{mn} \end{matrix} & \begin{matrix} -a_{01} \\ \vdots \\ -a_{0n} \end{matrix} & \begin{matrix} 1 & \ddots & \\ & \ddots & \\ 0 & & 1 \\ & & 0 \end{matrix} \end{bmatrix} \tag{10}$$

$$B = [1 \quad 0\cdots0 \mid b_{00} \quad 0\cdots0 \mid \tilde{b}_{01}\cdots\cdots\tilde{b}_{0n}] \tag{11}$$

$$C = [b_{10}\cdots\cdots b_{m0} \mid -a_{10}\cdots\cdots-a_{m0} \mid 1 \quad 0\cdots0] \tag{12}$$

$$D = b_{00}$$

其中 $\tilde{b}_{ij} = b_{ij} - b_{i0}a_{0j}$，$\tilde{a}_{ij} = a_{ij} - a_{i0}a_{0j}$。

由上述可见，$H(z_1,z_2)$ 可由两种不同的高效率的结构实现，而且这两种结构的性能通常有较大的差别。以下称前者为结构 1 后者为结构 2。

如果 $H(z_1, z_2)$ 的分子或分母是部分可分解的二维系统，则可用小于 $2m+n$（或 $2n+m$）的状态变量完全描述，甚至达到最小状态空间实现[9]。

三、高效率结构的格拉姆矩阵

对于二维系统的 Roesser 模型，文献 [3] 给出了该系统的能控性格拉姆矩阵 K 和能观性格拉姆矩阵 W 的时域和复频域表达式，在此我们不赘述，但值得指出的一点是，文献[3]和文献[4]的 W 是有区别的，其原因详见文献[10—12]。

对于本文中的结构 1 和结构 2，我们将它们的 K 和 W 与其对应的 A 进行分块，即

$$K = \begin{bmatrix} K_1^{(1)} & K_1^{(2)} & K_2^{(1)} \\ K_1^{(3)} & K_1^{(4)} & K_2^{(2)} \\ K_3^{(1)} & K_3^{(2)} & K_4 \end{bmatrix}, \quad W = \begin{bmatrix} W_1^{(1)} & W_1^{(2)} & W_2^{(1)} \\ W_1^{(3)} & W_1^{(4)} & W_2^{(2)} \\ W_3^{(1)} & W_3^{(2)} & W_4 \end{bmatrix} \tag{13}$$

把结构 1 和结构 2 的系数 A,B,C 分别代入文献[3]的(20)和(23)式，应用矩阵分块求逆公式和留数定理，可以证明(步骤从略)结构 1 和结构 2 的 K 和 W 具有以下性质：

$K_1^{(1)}$ 为单位阵，$K_1^{(4)}$ 和 W_4 均为对称的 Toeplitz 矩阵，$K_1^{(2)}$ 为下三角 Toeplitz 矩阵，$K_2^{(1)} = 0$，$K^T = K$，$W^T = W$。

通过这些性质，不仅可以检验 K 和 W 计算结果的精度，而且它将指导一般二维数字系统的低噪声优化实现。

四、噪声模型和优化

数字系统定点运算舍入量化的噪声有 3 种来源，输入信号的舍入噪声，系数的舍入噪声和乘积的舍入噪声。在文献[1,3,4]中，只讨论了乘积舍入噪声，具有较大的局限性，本节将建立这 3 种噪声源共同作用的噪声模型。

经有限字长量化后，令 $A_q = A + \Delta A$，$B_q = B + \Delta B$，$C_q = C + \Delta C$，$D_q = D + \Delta D$，$u_q(i,j) = u(i,j) + \Delta u(i,j)$，$\mathbf{X}_q(i,j) = \mathbf{X}(i,j) + \Delta\mathbf{X}(i,j)$，$y_q(i,j) = y(i,j) + \Delta y(i,j)$，3 种噪声共同作用的二维系统为

$$\overset{*}{\mathbf{X}}_q(i,j) = A_q\mathbf{X}_q(i,j) + B_q u_q(i,j) + \alpha(i,j) \tag{14}$$

$$y_q(i,j) = C_q\mathbf{X}_q(i,j) + D_q u_q(i,j) + \beta(i,j) \tag{15}$$

式中

$$\mathbf{X}_q(i,j) = \begin{bmatrix} \mathbf{X}_h(i,j) \\ \mathbf{X}_v(i,j) \end{bmatrix}_q, \quad \overset{*}{\mathbf{X}}_q(i,j) = \begin{bmatrix} \mathbf{X}_h(i+1,j) \\ \mathbf{X}_v(i,j+1) \end{bmatrix}_q$$

$\alpha(i,j)$ 为 $A_q\mathbf{X}_q(i,j)$ 和 $B_qu_q(i,j)$ 的乘积舍入噪声，$\beta(i,j)$ 为 $C_q\mathbf{X}_q(i,j)$ 和 $D_qu_q(i,j)$ 的乘积舍入噪声。

将(14)和(15)式减去对应的理想情形的状态方程和输出方程，略去高阶小量得

$$\Delta\overset{*}{\mathbf{X}}(i,j)=A_q\Delta\mathbf{X}(i,j)+B_q\Delta u(i,j)+\Delta A\mathbf{X}(i,j)+\Delta Bu(i,j)+\alpha(i,j) \tag{16}$$

$$\Delta y(i,j)=C_q\Delta\mathbf{X}(i,j)+D_q\Delta u(i,j)+\Delta C\mathbf{X}(i,j)+\Delta Du(i,j)+\beta(i,j) \tag{17}$$

当各初始状态均为零时，求得

$$\Delta y(i,j)=\Delta y_1(i,j)+\Delta y_2(i,j)+\Delta y_3(i,j) \tag{18}$$

其中

$$\Delta y_1(i,j)=C_q\sum_{(0,0)\leqslant(k,l)<(i,j)}\sum\bar{A}_q^{i-k,j-l}B_q\Delta u(k,l)+D_q\Delta u(i,j) \tag{19}$$

$$\Delta y_2(i,j)=C_q\sum_{(0,0)\leqslant(k,l)<(i,j)}\sum\bar{A}_q^{i-k,j-l}[\Delta A\mathbf{X}(k,l)+\Delta Bu(k,l)]+\Delta C\mathbf{X}(i,j)+\Delta Du(i,j) \tag{20}$$

$$\Delta y_3(i,j)=C_q\sum_{(0,0)\leqslant(k,l)<(i,j)}\sum\bar{A}_q^{i-k,j-l}\alpha(k,l)+\beta(i,j) \tag{21}$$

$$\mathbf{X}(i,j)=\sum_{(0,0)\leqslant(k,l)<(i,j)}\sum\bar{A}^{i-k,j-l}Bu(k,l) \tag{22}$$

式中 $\bar{A}^{i,j}$ 的含义和求解公式详见文献[3]。

由上可见，在略去高阶小量后，$\Delta y_1(i,j)$ 为输入信号舍入噪声响应，$\Delta y_2(i,j)$ 为系数舍入噪声响应，$\Delta y_3(i,j)$ 为乘积舍入噪声响应。

众所周知，对于舍入量化，$\Delta u(i,j)$，$\alpha(i,j)$，$\beta(i,j)$ 都可视为均值为零的白噪声，且有 $E\{|\Delta u(i,j)|^2\}=\sigma^2$，$E\{\alpha(i,j)\alpha^T(i,j)\}=Q\sigma^2$，$E\{\beta^2(i,j)\}=\gamma\sigma^2$，$\sigma^2=2^{-2L}/12$，$L$ 为字长位数，Q 为对角阵，其对角元素 Q_i 是 A 和 B 的第 i 行中非 0 非 1 元素的个数，γ 是 C 和 D 中非 0 非 1 元素的个数。

当输入信号 $u(i,j)$ 是均值为零、方差为 1 的白噪声，则理想输出 $y(i,j)$ 的均值为零，方差为

$$P_y=CKC^T+D^2=B^TWB+D^2 \tag{23}$$

误差输出 $\Delta y(i,j)$ 的均值也为零，方差为

$$P_{\Delta y}=P_{\Delta y_1}+P_{\Delta y_2}+P_{\Delta y_3} \tag{24}$$

其中

$$P_{\Delta y_1}=\sigma^2[C_qK_qC_q^T+D_q^2]$$

$$P_{\Delta y_2}=\mathrm{tr}\{W_q\Delta AK\Delta A^T+W_q\Delta B\Delta B^T]+\Delta CK\Delta C^T+\Delta D^2\text{[注]} \tag{25}$$

$$P_{\Delta y_3}=\sigma^2[\mathrm{tr}\{QW_q\}+\gamma] \tag{26}$$

在 $P_{\Delta y}$ 的 3 项噪声中，(1)对于 $P_{\Delta y_1}$，当字长足够长时，$C_qK_qC_q^T+D_q^2$ 与 P_y 相近，因此，字长位数 L 对 $P_{\Delta y_1}$ 起决定作用；(2) $P_{\Delta y_2}$ 取决于系数量化误差，并将随系数灵敏度[13]的降低而减小；(3) $P_{\Delta y_3}$ 是乘积舍入噪声功率。根据文献[13]可知，在 L_2 范数约束条件下，即在内部状态无溢出条件下，乘积舍入的单位噪声增益 $\mathrm{tr}W$ 最小时，系数灵敏度也将最小。因此，在状态 $\mathbf{X}(i,j)$ 无溢出的条件下，使 $P_{\Delta y_3}$ 极小化，$P_{\Delta y_2}$ 也将大幅度降

[注] 对于一个确定的结构，其系数误差是确定的。

低.

文献[3,4]给出了在最小状态空间实现的内部状态无溢出的条件下，$\mathrm{tr}W$ 最小化的算法. 由于他们所得的 (m,n) 阶数字系统的优化结构，共有 $(m+n+1)^2$ 个非 0 非 1 系数，因此，他们的优化结构的工作效率是最低的. 而且由于 $Q=(m+n+1)I$，所以，$\mathrm{tr}W$ 最小，并不能保证 $P_{\Delta y_3}$ 最小.

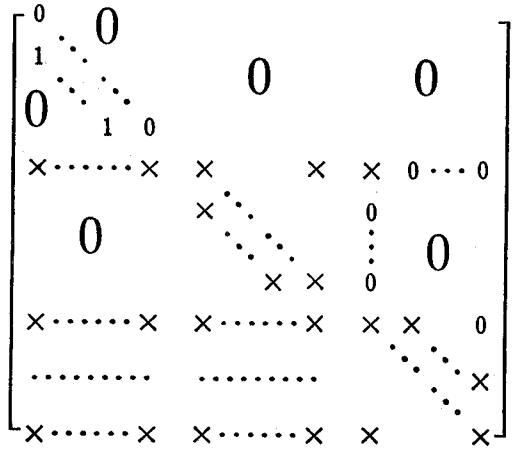
对于不能用最小状态空间实现的二维数字系统，若用文献[3,4]的算法，那所得结构的非 0 非 1 系数将会接近或达到 $(2m+n+1)^2$（或 $(2n+m+1)^2$）个，而且 $P_{\Delta y_3}$ 也会较大.

为使不存在最小状态空间实现的一般二维数字系统，得到较合理的优化，即既要噪声低，又要兼顾工作效率，我们对前面提出的两种状态空间结构作一定性分析.

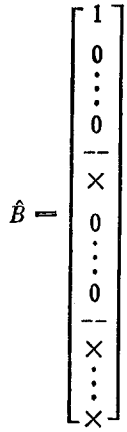
令 T_1 为上三角阵并与 $A_1^{(q)}$ 同维，T_2 为下三角阵并与 A_4 同维，$T=I\oplus T_1\oplus T_2$，对 $\mathbf{X}(i,j)$ 进行非奇异变换 $\hat{\mathbf{X}}(i,j)=T^{-1}\mathbf{X}(i,j)$，则得新结构 $\hat{A}=T^{-1}AT$，$\hat{B}=T^{-1}B$，$\hat{C}=CT$，$\hat{D}=D$，$\hat{K}=T^{-1}KT^{-T}$，$\hat{W}=T^TWT$，而且新结构与原结构有相同的传输函数 $H(z_1,z_2)$. 所得新结构有以下性质:

(1) $K_1^{(1)}$ 仍为单位阵，故 $\hat{\mathbf{X}}_{v_1}(i,j)$（或 $\hat{\mathbf{X}}_{h_1}(i,j)$）无溢出;

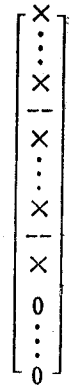
(2) $\hat{A}=$



$\hat{B}=$



$\hat{C}=$



可见新结构有较多的 0 系数和部分 1 系数，故效率较高，而且若令 $\hat{Q}=\mathrm{diag}\{\hat{Q}_0,\hat{Q}_1,\hat{Q}_2\}$，则有

对于结构 1:

$\hat{Q}_0=\mathbf{0}$，$\hat{Q}_1=\mathrm{diag}\{2n+2,n,n-1,\cdots,2\}$

$\hat{Q}_2=\mathrm{diag}\{2n+3,\cdots,2n+m,\ 2n+m+1,\ 2n+m+1\}$，$\hat{r}=2n+2$

对于结构 2:

$\hat{Q}_0=\mathbf{0}$，$\hat{Q}_1=\mathrm{diag}\{2m+2,m,m-1,\ \cdots,2\}$

$\hat{Q}_2=\mathrm{diag}\{2m+3,\cdots,\ 2m+n,\ 2m+n+1,\ 2m+n+1\}$，$\hat{r}=2m+2$

根据以上分析可知: (1)无溢出约束条件可化为 $\hat{\mathbf{X}}_{v_2}(i,j)$ 和 $\hat{\mathbf{X}}_{h_2}(i,j)$ [或 $\hat{\mathbf{X}}_{h_2}(i,j)$ 和 $\hat{\mathbf{X}}_{v_2}(i,j)$] 无溢出的条件;(2)求解 $T=I\oplus T_1\oplus T_2$ 使 $\hat{P}_{\Delta y_3}$ 最小（此时 $\hat{P}_{\Delta y_1}$ 也随之有较大的降低）可化为求解 T_1 使 $\mathrm{tr}\{\hat{Q}_1T_1^TW_1^{(q)}T_1\}$ 最小和求解 T_2 使 $\mathrm{tr}\{\hat{Q}_2T_2^TW_4T_2\}$ 最小两步进行. 因为

$$\hat{P}_{\Delta y_3} = \mathrm{tr}\{\hat{Q}\hat{W}\} + \hat{r} = \mathrm{tr}\{\hat{Q}T^TWT\} + \hat{r}$$
$$= \mathrm{tr}\{\hat{Q}_1T_1^TW_1^{(0)}T_1\} + \mathrm{tr}\{\hat{Q}_2T_2^TW_4T_2\} + \hat{r}$$

至于求解 T_1 和 T_2 的算法，作者在文献[14,15]中做了详细的讨论，在此不赘述。

在结束本节之前，我们必须指出，本文的优化方法是在保证效率较高、内部状态无溢出的前提下，进行低噪声优化的，而不是单方面追求某一指标最优。

五、举　例

给定(2,3)阶二维数字系统的传输函数为

$$H(z_1,z_2)=\frac{[1\ z_1^{-1}\ z_1^{-2}]\begin{bmatrix}0.00895000 & 0.01233749 & -0.03458808 & 0.02684839\\ -0.01451251 & -0.01993604 & 0.05629398 & -0.04353164\\ 0.00894946 & 0.01233784 & -0.03456347 & 0.02690451\end{bmatrix}\begin{bmatrix}1\\ z_2^{-1}\\ z_2^{-2}\\ z_2^{-3}\end{bmatrix}}{[1\ z_1^{-1}\ z_1^{-2}]\begin{bmatrix}1.000000 & -0.8281300 & -0.873048 & 0.7961280\\ -1.788130 & 1.489795 & 1.585434 & -1.433002\\ 0.829300 & -0.6965820 & -0.7347716 & 0.6703008\end{bmatrix}\begin{bmatrix}1\\ z_2^{-1}\\ z_2^{-2}\\ z_2^{-3}\end{bmatrix}}$$

由(4)—(6)式可得实现 $H(z_1, z_2)$ 的结构 1，根据文献[3]的(17)和(22)式，取(i,j)的上限为(200,200)，在 AT286 微机上用双精度运算 80min 求得结构 1 的 K 和 W 为：

$$K=\begin{bmatrix}1 & 0 & 0 & 0.0089500 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0.0197492 & 0.0089500 & 0 & 0 & 0\\ 0 & 0 & 1 & -0.0102918 & 0.0197492 & 0.0089500 & 0 & 0\\ 0.0089500 & 0.0197492 & -0.0102918 & 0.0694147 & -0.0064059 & 0.0637646 & -0.0065926 & 0.0053857\\ 0 & 0.0089500 & 0.0197492 & -0.0064059 & 0.0694147 & -0.0064059 & 0.0021569 & -0.0018123\\ 0 & 0 & 0.0089500 & 0.0637646 & -0.0064059 & 0.0694147 & 0.0022131 & -0.0018463\\ 0 & 0 & 0 & -0.0065926 & 0.0021569 & 0.0022131 & 0.0174690 & -0.0142134\\ 0 & 0 & 0 & 0.0053857 & -0.0018123 & -0.0018463 & -0.0142134 & 0.0122225\end{bmatrix}$$

$$W=\begin{bmatrix}0.0541191 & -0.0420449 & 0.0369561 & -0.0053868 & 0.1149342 & -0.1125615 & 0.1219424 & 0.1007760\\ -0.0420449 & 0.0486577 & -0.0503984 & -0.1288905 & -0.1365623 & 0.1833786 & -2.368836 & -2.398576\\ 0.0369561 & -0.0503984 & 0.1340506 & 0.5222185 & 0.1003747 & -0.4719583 & 11.43300 & 11.43824\\ -0.0053868 & -0.1288905 & 0.5222185 & 19.04625 & 1.058889 & -14.08105 & 4.322692 & 5.233853\\ 0.1149342 & -0.1365623 & 0.1003747 & 1.058889 & 3.526450 & -2.944859 & 0.3461197 & 0.1216215\\ -0.1125615 & 0.1833786 & -0.4719583 & -14.08105 & -2.944859 & 12.73105 & -3.651518 & -4.940275\\ 0.1219424 & -2.368836 & 11.43300 & 4.322692 & 0.3461197 & -3.651518 & 1425.905 & 1396.177\\ 0.1007760 & -2.398576 & 11.43824 & 5.233853 & -0.1216215 & -4.940275 & 1396.177 & 1425.905\end{bmatrix}$$

由 K 和 W 可求得对应结构 1 的变换阵 T_1 和 T_2 分别为

$$T_1=\begin{bmatrix}0.107198598 & -0.013359390 & 0.214222453\\ 0 & 0.278603042 & 0.069473698\\ 0 & 0 & 0.263466717\end{bmatrix},\quad T_2=\begin{bmatrix}0.132170461 & 0\\ -0.118476696 & 0.278861284\end{bmatrix}$$

于是结构 1 的优化结构——结构 3 的系数为

$$A=\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0.1150900 & -0.3226542 & 0.2504547 & 0.8760813 & 0.0508355 & -0.1660422 & 1.232950 & 0\\
0 & 0 & 0 & 0.3847718 & -0.3116420 & 0.7031614 & 0 & 0\\
0 & 0 & 0 & 0 & 1.057451 & 0.2636906 & 0 & 0\\
0.0160777 & -0.0420215 & 0.0338712 & -0.0072924 & 0.0034010 & 0.0048288 & 0.8917368 & 0.2109861\\
0.1438420 & -0.3893744 & 0.3102649 & 0.0067435 & -0.0009447 & -0.0019191 & -0.1419676 & 0.8963932
\end{bmatrix}$$

$$B=[1\ \ 0\ \ 0\ \ 0.0834899\ \ 0\ \ 0\ \ 0.0112828\ \ 0.1027024]^T$$

$$C=[0.0123375\ \ -0.0345881\ \ 0.0268484\ \ 0.0887744\ \ 0.2361425\ \ 0.0292952\ \ 0.1321705\ \ 0]$$

$$D=0.00895$$

由(10)—(12)式得结构 2,取(i,j)的上限为(200,200),用时 58min,求得结构 2 的K和W(限于篇幅,恕不列出),对应结构 2 的变换阵 T_1 和 T_2 分别为

$$T_1=\begin{bmatrix}0.116255118 & 0.226688019\\ 0 & 0.263466718\end{bmatrix}\qquad T_2=\begin{bmatrix}0.179434779 & 0 & 0\\ -0.119944428 & 0.182589788 & 0\\ -0.039042342 & -0.159049738 & 0.046777711\end{bmatrix}$$

结构 2 的优化结构——结构 4 的系数为

$$A=\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0\\
-0.1248333 & 0.0769812 & 0.9277252 & -0.0704378 & 1.543457 & 0 & 0\\
0 & 0 & 0.4412516 & 0.8604048 & 0 & 0 & 0\\
-0.1780830 & 0.1100632 & -0.0058253 & 0.0030509 & 0.1596730 & 1.017583 & 0\\
0.1208002 & -0.0735041 & -0.0030739 & 0.0019277 & 0.7630377 & -0.2026199 & 0.2561902\\
-0.4215116 & 0.2647821 & 0.0081007 & -0.0019711 & -0.3261831 & 0.1603795 & 0.8710769
\end{bmatrix}$$

$$B=[1\ \ 0\ \ 0.0769859\ \ 0\ \ 0.1100637\ \ -0.0736360\ \ 0.2631254]^T$$

$$C=[-0.0145125\ \ 0.0089495\ \ 0.2078793\ \ 0.1868547\ \ 0.1794348\ \ 0\ \ 0]$$

$$D=0.00895$$

如果用文献[3]或文献[4]的方法(这两种方法的结果相同[12])对结构 1 和结构 2 进行优化,则相应得结构 5 和结构 6,限于篇幅,此处不给出它们的系数,仅将这 6 种结构的主要参数列于表 1,表中 $P_{\Delta y_2}/\sigma^2$ 项为 $L=11$ 的结果。

表 1

结 构	1	2	3	4	5	6
$P_{\Delta y_1}/\sigma^2$	0.0671	0.0662	0.0693	0.0694	0.0693	0.0694
$P_{\Delta y_2}/\sigma^2$	12947.8	46222.4	47.42	297.06	71.04	415.57
$P_{\Delta y_3}/\sigma^2$	22697.5	25740.1	34.96	177.84	35.74	182.30
单位噪声增益	2887.35	4399.85	3.141	23.61	2.971	21.79
乘法器	30	28	39	37	81	64
加法器	29	27	33	32	72	56

我们用 64 位二进制浮点运算,求该二维系统的单位冲激响应 $h(i,j)$,用 12 位(扣除符号位,$L=11$)二进制定点运算,作计算机模拟,求以上 6 种结构的单位冲激响应 $\hat{h}(i,j)$,现将模拟结果列于表 2.

表 2

结　构	1	2	3	4	5	6
$\frac{\sum_{i=0}^{150}\sum_{j=0}^{150}[\hat{h}(i,j)-h(i,j)]^2}{\sum_{i=0}^{150}\sum_{j=0}^{150}[h(i,j)]^2}$	1.12×10^{-2}	1.88×10^{-2}	8.51×10^{-5}	3.42×10^{-4}	9.73×10^{-5}	4.71×10^{-4}

由表 1 可见，尽管本文给出的优化结构(结构 3 和4)比文献[3]和文献[4]方法所对应的优化结构（结构 5 和 6）的单位噪声增益略高，但结构 3 和 4 比对应的结构 5 和 6 有工作效率更高全局噪声更低的优点。表 2 的模拟结果从另一方面也验证了结构 3 和 4 的噪声低于对应的结构 5 和 6 的噪声。

此外，结构 1 和结构 2 以及它们对应的结构(结构 3 和结构 4)的工作效率和噪声性能也各有差异，设计者可以从中选择合适的结构予以实现 $H(z_1,z_2)$。

六、结　语

本文讨论了一般二维数字系统的状态空间实现问题；找出了两种效率高的状态空间结构；导出了这两种结构对应的能控性和能观性格拉姆矩阵的有关性质。根据数字系统有限字长定点运算法则，建立了二维数字系统的全局噪声模型；针对以上两种结构，讨论了二维数字系统低噪声高效率的优化问题。文中举例及其模拟结果，验证了本文的有关理论，说明了两种结构及其优化结构的噪声性能和工作效率各不相同。举例结果还充分地说明了本文的优化结构比文献[3]和文献[4]方法对应的优化结构，具有更低的全局噪声和更高的工作效率。

本文部分内容曾在 CCSP'92 (成都，1992 年 4 月) 会议上宣读。作者感谢张有正教授对本文的热情支持。

参　考　文　献

[1] B. G. Mertzios, *IEEE Trans. on CAS*, CAS-32(1985)2, 201—204.
[2] W. S. Lu et al., *IEEE Trans. on CAS*, CAS-32(1985)10, 1080—1081.
[3] T. Lin et al., *IEEE Trans. on CAS*, CAS-33(1986)7, 724—730.
[4] W. S. Lu et al., *IEEE Trans. on CAS*, CAS-33(1986)10, 965—973.
[5] S. J. Varoufakis et al., *IEEE Trans. on CAS*, CAS-34(1987)3, 289—292.
[6] G. E. Antoniou et al., *IEEE Trans. on CAS*, CAS-35(1988)8, 1055—1058.
[7] S. Y. Kung et al., *Proc. IEEE*, 65(1977)6, 945—961.
[8] 张宏科等，电子学报，18 (1990) 6, 107—109.
[9] A. Kawakami, *IEEE Trans. on CAS*, CAS-37(1990)3, 425—432.
[10] T. Hinamoto et al., *IEEE Trans. on CAS*, CAS-35(1988)5, 609—610.
[11] T. Lin et al., *IEEE Trans. on CAS*, CAS-35(1988)5, 610—611.
[12] W. S. Lu et al., *IEEE Trans. on CAS*, CAS-35(1988)5, 611—612.
[13] M. Kawamata et al., Minimization of sensitivity of 2-D state-space digital filters and its realization to 2-D balanced realizations, Proc. 1987 IEEE on ISCAS, Philadelphia, PA, May 1987, pp. 710—713.

[14] 肖承山等,二阶数字滤波器的近最佳实现,全国 NN-SP 会议论文集,南京,1991 年,第 287—290 页。
[15] 肖承山,状态空间数字滤波器高效率低噪声的优化实现,信号处理,将发表在 1993 年第 3 期上.

OPTIMAL STATE-SPACE REALIZATIONS OF GENERAL 2-D DIGITAL SYSTEMS WITH LOW ROUNDOFF NOISE

Xiao Chengshan Feng Zhenming

(Tsinghua University, Beijing 100084)

Abstract Two computationally efficient state-space structures for general 2-D digital systems described by Roesser's local state-space model are presented. Some properties of controllability and observability Gramians of the above two structures are derived. The overall roundoff noise model of 2-D digital systems implemented with fixed-point arithmetic is given. On the basis of the above theory, the low roundoff noise optimal state-space realizations of general 2-D digital systems are obtained. Finally, an example is given to demonstrate the differences of roundoff noise and computational efficiency between the above two structures.

Key words 2-D digital systems; State-space relazations; Roundoff noise; Optimization