论 文

基于输入平衡实现的一类高效 IIR 数字滤波器结构

黄朝耿^{①*}, 陈琰宏^①, 徐红^②, 何熊熊^②, 周志光^①^① 浙江财经大学信息学院, 杭州 310018^② 浙江工业大学信息工程学院, 杭州 310023

* 通信作者. E-mail: cghuang@zufe.edu.cn

收稿日期: 2015-05-04; 接受日期: 2015-07-02; 网络出版日期: 2015-07-16

国家自然科学基金 (批准号: 61273195, 61303133, 61473262)、浙江省自然科学基金 (批准号: LQ14F030008) 和浙江省科技厅公益性技术应用研究计划 (批准号: 2014C31057) 资助项目

摘要 基于一种输入平衡实现, 本文导出了一种简洁有效的 IIR 数字滤波器结构, 并导出了传函灵敏度函数和舍入噪声增益的数学表达式. 对于一个 N 阶的数字滤波器, 该结构只需 $5N-1$ 个乘法器, 比归一化格型结构少了两个乘法器. 对于全通系统, 新结构的乘法器个数和舍入噪声增益都为 $4N$, 极大地降低了最佳状态空间实现具有的 $(N+1)^2$. 通过仿真实例验证了所提结构对有限字长效益的高鲁棒性和实现的简洁性.

关键词 状态空间实现 数字滤波器结构 有限字长 输入平衡实现 全通滤波器

1 引言

众所周知, 一个设计好的数字滤波器 $H(z)$ 必须由数字器件来实现, 而数字器件的有限字长限制必将带来所谓的有限字长 (finite word length, FWL) 效应, 导致实际的数字滤波器性能大大降低. 同时, 数字滤波器的性能与其实现结构¹⁾紧密相连, 即一个好的结构具有非常强的抵抗有限字长效应的能力. 因此, 从设计的角度上来讲, 研究者们往往针对某一给定的 FWL 标准设计高性能的结构^[1,2]. FWL 标准一般分为两类: 传函灵敏度函数 (sensitivity measure)^[3~8] 和舍入噪声增益 (roundoff noise gain)^[9~13]. 近些年, 一种叫做误差反馈的技术被广泛应用于基于状态空间实现的最优舍入噪声设计^[14~16]. 事实上, 如何设计高性能低复杂度的数字滤波器结构一直是数字信号处理领域的研究热点之一^[17~22].

最佳状态空间实现对 FWL 效益有着非常强的鲁棒性, 但是它属于满参的结构, 即对于一个 N 阶的 IIR 数字滤波器, 实现需要 $(N+1)^2$ 个乘法器和 $N(N+1)$ 个加法器. 近期, Li 等^[21] 综合了注入式和抽取式两类格型结构的特点, 提出了一类基于格型思想的滤波器结构. 该结构不仅具有标准的乘法器个数, 而且具有非常低的参数灵敏度和平稳的状态信号能量. 在文献^[21] 的基础之上, Huang 进一步推广并提出了一种广义的格型滤波器结构^[23]. 非常著名的归一化格型结构, 记为 S_{nl} , 同样具有非常好

1) 在本文, 结构只指数字滤波器由给定输入计算输出的一种方式或方法.

的数字特性, 且只含有 $5N + 1$ 个乘法器, 其实质上是一种特殊的正交结构 (orthogonal structure) [24]. 从状态空间的角度分析, 正交结构也叫输入平衡实现 (input-balanced realization), 表现为其可控 Gram 矩阵为单位阵. 文献 [12] 基于一类特殊的输入平衡实现, 导出了一种稀疏正交结构, 对于每一次输出该结构只需做 $4N - 1$ 次乘法和 $4N - 1$ 次加法.

需要特别指出, 输入平衡实现并不是唯一存在的. 归一化格型结构和文献 [12] 提出的结构都属于输入平衡实现. 本文的主要目的是深入研究输入平衡实现的特性, 导出高性能的简洁数字滤波器结构. 本文的主要内容如下所述.

- 基于一类 Hessenberg 的输入平衡实现, 本文提出了一种高鲁棒性的正交结构, 为方便论述, 新结构命名为 S_{new} . 与 S_{nl} 相比, 新结构较前者少两个乘法器, 其个数为 $5N - 1$.

- 对所提新结构进行了性能分析, 并导出了传输函数的灵敏度函数和舍入噪声增益表达式.

- 对于全通系统, 本文导出了 S_{new} 结构对应的简化形式. 简化后的结构只含有 $4N$ 个乘法器且对应的舍入噪声增益也为 $4N$, 其性能远远优越于具有 $(N + 1)^2$ 个乘法器和舍入噪声增益为 $(N + 1)^2$ 的最佳舍入噪声增益结构.

本文的概述如下, 在第 2 部分, 给出了状态空间实现结构的一些准备知识和两类 FWL 衡量标准, 即灵敏度函数和舍入噪声增益. 这部分也给出了一些经典的数字滤波器结构. 本文的主要贡献体现在第 3 ~ 5 部分. 第 3 部分从一类特殊的输入平衡实现入手, 导出了新结构. 在第 4 部分给出了新结构的传函灵敏度函数和舍入噪声增益表达式. 第 5 部分则致力于推导新结构对于全通系统的简化形式, 简化后的结构只需 $4N$ 个乘法器和 $2N$ 个加法器, 并且具有非常低的舍入噪声增益, 其值为 $4N$. 在第 6 部分, 本文给出了两个仿真实例, 验证了新结构的良好性能. 本文的结论由第 7 部分给出.

2 预备知识

对于一个 N 阶的 IIR 数字滤波器, 其传输函数 $H(z)$ 可表示为

$$H(z) = \frac{\sum_{k=0}^N b_k z^{-k}}{1 + \sum_{k=1}^N a_k z^{-k}} \triangleq \frac{N(z)}{D(z)}. \quad (1)$$

显然, $H(z)$ 在无限精度下存在无穷种等价的实现结构. 状态空间实现 (state-space realization) 是一类特殊的方式, 其表示如下:

$$\begin{cases} x(n+1) = Ax(n) + Bu(n), \\ y(n) = Cx(n) + du(n), \end{cases} \quad (2)$$

其中, $u(n)$ 和 $y(n)$ 分别表示滤波器的输入和输出, $A \in \mathbb{R}^{N \times N}$, $B \in \mathbb{R}^{N \times 1}$, $C \in \mathbb{R}^{1 \times N}$ 和 $d \in \mathbb{R}$ 是状态空间实现的结构参数且满足

$$H(z) = d + C(zI - A)^{-1}B, \quad (3)$$

这里, I 表示维数为 N 的单位矩阵.

定义 $R \triangleq (A, B, C, d)$ 为任意状态空间实现, 则其可控和可观 Gram 矩阵 W_c, W_o 可用 Lyapunov 方程求得

$$\begin{cases} W_c = AW_cA^T + BB^T, \\ W_o = A^TW_oA + C^TC, \end{cases} \quad (4)$$

其中, T 表示转置操作.

状态空间实现存在这样一个特性: 若初始实现为 $R_0 \triangleq (A_0, B_0, C_0, d)$, 则 (A, B, C, d) 可通过相似变换由 R_0 变换得到, 即

$$A = T^{-1}A_0T, \quad B = T^{-1}B_0, \quad C = C_0T, \quad (5)$$

其中, $T \in \mathbb{R}^{N \times N}$ 为非奇异矩阵, 又称相似矩阵. 此时,

$$W_c = T^{-1}W_c^{(0)}T^{-T}, \quad W_o = T^TW_o^{(0)}T, \quad (6)$$

这里, $W_c^{(0)}$ 和 $W_o^{(0)}$ 分别为 R_0 的可控和可观 Gram 矩阵.

2.1 灵敏度

对于一个给定的结构, 其结构参数的集合定义为 $P_\tau \triangleq \{\tau_k\}$. 如前所述, 由于硬件的 FWL 约束, 理想的参数只能被有限字长表示. 因此, 实际参数 τ_k^* 与理想参数 τ_k 之间必然存在量化误差 $\Delta\tau_k$. 即满足 $\tau_k^* = \tau_k + \Delta\tau_k$, 此时, 实际的传输函数 $H^*(z)$ 也会偏离理想的传输函数 $H(z)$. 根据 Taylor 展开并利用其一阶近似可得

$$\Delta H(z) = H(z) - H^*(z) \approx \sum_{\tau_k \in P_\tau, \forall k} \frac{\partial H(z)}{\partial \tau_k} \Delta\tau_k, \quad (7)$$

其中, $S_{\tau_k}(z) \triangleq \frac{\partial H(z)}{\partial \tau_k}$ 定义为传输函数对参数 τ_k 的灵敏度函数. 基于 l_2 -norm 可计算得到整个频率范围内的偏差总和 [6], 即

$$M \triangleq \sum_{\tau_k \in P_\tau, \forall k} \|S_{\tau_k}(z)\|_2^2. \quad (8)$$

这里, l_2 -norm 的定义为

$$\|S_{\tau_k}(z)\|_2^2 \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} |S_{\tau_k}(e^{j\omega})|^2 d\omega. \quad (9)$$

如果 $S_{\tau_k}(z)$ 用状态空间的形式表示, 即 (A_k, B_k, C_k, D_k) , 则 $\|S_{\tau_k}(z)\|_2^2$ 可由留数定理直接得到

$$\|S_{\tau_k}(z)\|_2^2 = \text{tr}[D_k^T D_k + B_k^T W_o^{(k)} B_k], \quad (10)$$

其中, $\text{tr}[\cdot]$ 表示矩阵的迹运算, $W_o^{(k)}$ 为 (A_k, B_k, C_k, D_k) 实现的可观 Gram 矩阵.

2.2 舍入噪声增益

当参数与信号相乘时会使小数位数加倍, 由于字长的限定会产生 FWL 量化误差, 这种误差通常称作舍入噪声²⁾. 在实际应用中, 通常假设信号的截断发生在相乘之后, 其模型建立如下: 假设 $s(n)$ 为信号与某一参数 τ 乘积的结果, 则其必将被量化器 $q[\cdot]$ 量化. 定义舍入噪声为 $\epsilon_s(n) \triangleq q[s(n)] - s(n)$. 根据文献 [2] 可知, 舍入噪声可以建模为平稳高斯白噪声, 且其方差为常数, 满足 $E[\epsilon_s^2(n)] = \sigma_0^2$, 其中 $E[\cdot]$ 表示信号的统计平均. 存在的 $\epsilon_s(n)$ 必然对滤波器的输出产生影响, 假设 $\Delta y(n)$ 是由于存在 $\epsilon_s(n)$ 而导致实际输出与理想输出的偏差量, 则对于参数 τ 的舍入噪声增益定义为

$$G_\tau \triangleq \frac{E[(\Delta y(n))^2]}{E[\epsilon_s^2(n)]}.$$

2) 特别地, 当系统参数为 0, ± 1 时, 系统并不产生舍入噪声.

图 1 归一化格型结构 S_{nl} 的结构框图

Figure 1 The normalized lattice filter structure

若 $H_\tau(z) = D_\tau + C_\tau(zI - A_\tau)^{-1}B_\tau$ 为 $\epsilon_s(n)$ 与 $\Delta y(n)$ 之间的传输函数, 则 G_τ 的计算可参考式 (10).

因此, 整个滤波器结构的舍入噪声增益为

$$G = \sum_{\tau} G_{\tau}. \quad (11)$$

2.3 经典滤波器结构

(1) 最佳状态空间实现. 为了降低状态变量的溢出概率, 往往采用 l_2 -norm 对状态变量进行约束. 对于状态空间实现 (A, B, C, d) , 只需约束其可控 Gram 矩阵 W_c , 即³⁾

$$W_c(k, k) = 1, \quad \forall k. \quad (12)$$

对于全参的状态空间实现, 相应的舍入噪声增益为

$$G = [\text{tr}(W_o) + 1](N + 1). \quad (13)$$

而最佳状态空间实现就是以式 (12) 为约束条件, 对式 (13) 进行优化设计得到的. Mullis 等^[9] 和 Hwang^[10] 分别对这个命题进行了优化求解, 为方便起见, 优化得到的最佳结构记为 R_{mrh} , 相应的最优舍入噪声增益为

$$G_{mrh} = \left[1 + \frac{1}{N} \left(\sum_{l=1}^N \sigma_l \right)^2 \right] (N + 1). \quad (14)$$

其中, $\{\sigma_l\}$ 是滤波器 Hankel 矩阵的奇异值, 且 σ_l 表示矩阵 $W_c W_o$ 的第 l 个特征值.

(2) 归一化格型结构. 非常著名的归一化格型结构, 其结构框图见图 1. 该结构具有非常好的 FWL 特性, 详细见文献 [21].

(3) 输入平衡实现. 如果一个实现的可控 Gram 矩阵为单位阵, 则该实现叫做输入平衡实现, 记为 R_{ib} . 对于一个满参的输入平衡实现, 其相应的舍入噪声增益 G_{ib} 为

$$G_{ib} = [\text{tr}(W_o) + 1](N + 1) = \left(\sum_{k=1}^N \sigma_k^2 + 1 \right) (N + 1). \quad (15)$$

这里, σ_k 的定义如前所述, 为滤波器 $W_c W_o$ 的奇异值. 显然, $G_{ib} \geq G_{mrh}$, 当且仅当 $\sigma_i = \sigma_j$ 时取等号.

需要特别指出的是, 状态空间的输入平衡实现并不唯一. 事实上, 从状态空间的角度分析 S_{nl} 结构, 其可控 Gram 矩阵也是单位阵. 文献 [12] 针对一种特殊的输入平衡实现, 提出了更加有效的输入平衡实现. 继续深入研究探索输入平衡实现的特性是本文的主要目的.

3) 本文采用 matlab 中的表示方法, $M(i, j)$, $M(i, :)$, $M(:, j)$ 分别表示矩阵 M 的第 (i, j) 个元素, 第 i 行, 第 j 列.

3 一种新的输入平衡实现

在推导新结构前, 需给出以下的引理^[2].

引理 1 若数字滤波器的传输函数 $H(z)$ 对应的状态空间实现为 $\hat{R} \triangleq (\tilde{A}, \tilde{B}, \tilde{C}, d)$, 则通过相似变换可以使得 $\hat{R}$ 变换到一个具有 Hessenberg 形式的状态空间实现 (A, B, C, d) , 其结构为

$$A = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times & \times \\ 0 & \times & \cdots & \times & \times & \times & \times \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 0 & \times & \times & \times \\ 0 & 0 & \cdots & 0 & 0 & \times & \times \end{bmatrix}, B = \begin{bmatrix} \times \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad (16)$$

C 没有特定结构且满参.

基于引理 1, 这里将给出定理 1.

定理 1 任意一个 N 阶稳定的数字滤波器 $H(z)$, 存在一类特殊的输入平衡实现 (A, B, C, d) , 其结构满足

$$\begin{cases} A = \begin{bmatrix} \cos \phi_1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} Q \triangleq Q_1 Q, \\ B = \begin{bmatrix} -\sin \phi_1 & 0 & \cdots & 0 \end{bmatrix}^T, \\ C = \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_N \end{bmatrix}, d = \xi_0, \end{cases} \quad (17)$$

其中, $|\phi_k| < \frac{\pi}{2}$, $Q \triangleq \prod_{k=2}^N Q_k = Q_2 Q_3 \cdots Q_N$, $Q_k, k = 2, 3, \dots, N$ 是旋转矩阵 (givens matrix)⁴⁾,

$$Q_k \triangleq \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \cos \phi_k & \sin \phi_k & & \\ & & & -\sin \phi_k & \cos \phi_k & & \\ & & & & & 1 & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{bmatrix}. \quad (18)$$

证明 假设 $(\hat{A}, \hat{B}, \hat{C}, \hat{d})$ 为任一输入平衡实现, 根据引理 1, 通过相似变换可以得到一类特殊的输入平衡实现 (A, B, C, d) , 并具有 Hessenberg 形式.

4) 矩阵 Q_k 的第 $(k-1, k)$ 个元素为 $\sin \phi_k$.

输入平衡实现的可控 Gram 矩阵为单位阵, 即 $AA^T + BB^T = I$, 由于向量 B 只有第一个元素为结构参数, 则上式变为 $A(1, :)(A(1, :))^T + B^2(1) = 1$. 进一步地, 若设定 $A(1, :) = \cos \phi_1 Q(1, :)$ 和 $B(1) = -\sin \phi_1$ 即可满足上式等式. 这里, Q 为正交矩阵且满足 $A = Q_1 Q$.

任意正交矩阵 U 满足 $UU^T = U^T U = I$, 则矩阵 U 至少含有 $N(N+1)/2$ 个独立的参数. 然而, 由于 A 具有 Hessenberg 形式, Q 可以被分解为 $N-1$ 个连乘的旋转矩阵.

如前所述, A 具有 Hessenberg 形式且 Q_1 是对角矩阵, 则 Q 同样具有 Hessenberg 形式. 因为

$$QQ^T = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times & \times \\ 0 & \times & \cdots & \times & \times & \times & \times \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 0 & \times & \times & \times \\ 0 & 0 & \cdots & 0 & 0 & \times & \times \end{bmatrix} \begin{bmatrix} \times & \times & \cdots & \times & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times & \times \\ 0 & \times & \cdots & \times & \times & \times & \times \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 0 & \times & \times & \times \\ 0 & 0 & \cdots & 0 & 0 & \times & \times \end{bmatrix}^T = I,$$

则 $Q^2(N, N-1) + Q^2(N, N) = 1$, 因此, 存在一个 ϕ_N 使得 $Q(N, N-1) = -\sin \phi_N$, $Q(N, N) = \cos \phi_N$. 根据式 (18) 定义的 Q_N , 存在

$$Q^{(1)} \triangleq QQ_N^T = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times & \times \\ 0 & \times & \cdots & \times & \times & \times & \times \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 0 & \times & \times & \times \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

等价地, $Q = Q^{(1)}Q_N$.

由于 $(Q^{(1)})^T Q^{(1)} = I$, 即 $(Q^{(1)}(:, N))^T Q^{(1)}(:, N) = 1$, 则 $Q^{(1)}$ 可以被简化为

$$Q^{(1)} = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times & 0 \\ \times & \times & \cdots & \times & \times & \times & 0 \\ 0 & \times & \cdots & \times & \times & \times & 0 \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 0 & \times & \times & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

需要指出的是, $Q^{(1)}(1:N-1, 1:N-1)$ 是具有 $N-1$ 维的 Hessenberg 矩阵.

同理, 按照以上的思想方法可以得到

$$Q = Q^{(2)}Q_{N-1}Q_N = Q^{(3)}Q_{N-2}Q_{N-1}Q_N = \cdots = Q^{(N-2)}Q_3 \cdots Q_{N-1}Q_N = Q^{(N-1)}Q_2Q_3 \cdots Q_N,$$

这里, $Q^{(N-1)} = I$. 证毕.

图 2 新结构的结构框图

Figure 2 The proposed new structure

基于定理 1, 一个 N 阶的数字滤波器 $H(z)$ 可用如下结构实现, 并命名为 S_{new} ,

$$\begin{cases} x^{(1)}(n) \triangleq x(n), \\ x^{(m)}(n) = Q_{N-m+2}x^{(m-1)}(n), \quad 2 \leq m \leq N, \\ w(n) = x^{(N)}(n), \\ x(n+1) = Q_1w(n) + Bu(n), \\ y(n) = Cx(n) + du(n). \end{cases} \quad (19)$$

对应的结构框图见图 2. 图 2 所示的新结构与归一化格型结构 (见图 1) 非常相似, 但从实现复杂度上比较, 新结构只含有 $5N-1$ 个乘法器, 比 S_{nl} 少两个. 由于新结构是由输入平衡实现等价导出, 其对量化误差具有非常好的鲁棒性, 这在下一节将做详细论述.

4 性能分析

根据第 3 部分的论述可知, 新结构的参数包含 $\{\alpha_k, \beta_k, \xi_k\}$, 其中, $\alpha_k \triangleq \cos \phi_k, \beta_k \triangleq \sin \phi_k, \forall k$. 本节将分析这些结构参数的灵敏度及其带来的舍入噪声增益总和, 并分别给出二者的计算表达式.

4.1 灵敏度分析

假定 τ 为一系统结构参数. 根据灵敏度的定义, 这个参数关于传输函数 $H(z) = d + C(zI - A)^{-1}B$ 的灵敏度函数为

$$\frac{\partial H(z)}{\partial \tau} = \frac{\partial d}{\partial \tau} + \frac{\partial C}{\partial \tau}(zI - A)^{-1}B + C(zI - A)^{-1}\frac{\partial A}{\partial \tau}(zI - A)^{-1}B + C(zI - A)^{-1}\frac{\partial B}{\partial \tau}. \quad (20)$$

因为 α_1 和 β_1 分别存在于 Q_1 和 B 中, 根据式 (17) 和 (20) 可以直接得到二者的灵敏度表达式

$$\begin{cases} \frac{\partial H}{\partial \alpha_1} = C(zI - A)^{-1}v_1v_1^TQ(zI - A)^{-1}B, \\ \frac{\partial H}{\partial \beta_1} = -C(zI - A)^{-1}v_1, \end{cases} \quad (21)$$

其中, v_k 是维数为 N 的单位阵的第 k 列, 即 $v_k = I(:, k), \forall k$.

同理, 对于 $\alpha_k, \beta_k, k > 1$, 其仅存在于 Q_k , 故

$$\begin{cases} \frac{\partial H}{\partial \alpha_k} = C(zI - A)^{-1}P_k[v_{k-1}v_{k-1}^T + v_kv_k^T]R_k(zI - A)^{-1}B, \\ \frac{\partial H}{\partial \beta_k} = C(zI - A)^{-1}P_k[v_{k-1}v_k^T - v_kv_{k-1}^T]R_k(zI - A)^{-1}B, \end{cases} \quad (22)$$

其中,

$$\begin{cases} P_k \triangleq \prod_{m=1}^{k-1} Q_m, & k = 2, \dots, N, \\ R_k \triangleq \begin{cases} I, & k = N, \\ \prod_{m=k+1}^N Q_m, & 1 < k < N, \end{cases} \end{cases} \quad (23)$$

关于 C, d 的参数灵敏度, 可以直接给出二者的灵敏度计算表达式

$$\begin{cases} \frac{\partial H}{\partial \xi_k} = v_k^T(zI - A)^{-1}B, \quad \forall k, \\ \frac{\partial H}{\partial d} = 1, \end{cases} \quad (24)$$

利用式 (10), β_1, ξ_k, d 的灵敏度可以直接计算得到

$$\begin{cases} \left\| \frac{\partial H}{\partial \beta_1} \right\|_2^2 = \text{tr}(v_1^T W_o v_1) = W_o(1, 1), \\ \left\| \frac{\partial H}{\partial \xi_k} \right\|_2^2 = \text{tr}(v_k^T W_c v_k) = W_c(k, k), \quad \forall k, \\ \left\| \frac{\partial H}{\partial d} \right\|_2^2 = 1, \end{cases} \quad (25)$$

其中, W_c 和 W_o 分别为新结构 S_{new} 的可控和可观 Gram 矩阵.

然而, 对于 α_k, β_k 的灵敏度计算, 由于其表达式是两个传输函数的级联, 可用如下方法求解. 假定 $S(z) = S_2(z)S_1(z)$, 其中 $S_m(z) = D_m + C_m(zI_m - A_m)^{-1}B_m, m = 1, 2, I_m$ 为单位阵且其维数与 A_m 相同. 以 $\frac{\partial H}{\partial \alpha_k}$ 为例, 记 $S_1(z) = C(zI - A)^{-1}P_k, S_2(z) = [v_{k-1}v_{k-1}^T + v_kv_k^T]R_k(zI - A)^{-1}B$.

若定义 $S(z) \triangleq \bar{D} + \bar{C}(z\bar{I} - \bar{A})^{-1}\bar{B}, \bar{I} \triangleq \text{diag}(I_1, I_2)$, 根据 $S_1(z)$ 和 $S_2(z)$ 可直接求得 $S(z)$:

$$\begin{cases} \bar{A} = \begin{bmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} B_1 \\ B_2D_1 \end{bmatrix}, \\ \bar{C} = \begin{bmatrix} D_2C_1 & C_2 \end{bmatrix}, \quad \bar{D} = D_2D_1. \end{cases}$$

此时, 利用式 (10) 可直接求得

$$\|S(z)\|_2^2 = \text{tr}[\bar{D}\bar{D}^T + \bar{C}\bar{W}_c\bar{C}^T], \quad (26)$$

其中, $\bar{W}_c$ 是实现 $(\bar{A}, \bar{B}, \bar{C}, \bar{D})$ 的可控 Gram 矩阵.

因此, 所提新结构的灵敏度可根据下式计算得到

$$M_{\text{new}} = \sum_{k=1}^N \left(\left\| \frac{\partial H}{\partial \alpha_k} \right\|_2^2 + \left\| \frac{\partial H}{\partial \beta_k} \right\|_2^2 + \left\| \frac{\partial H}{\partial \xi_k} \right\|_2^2 \right) + 1. \quad (27)$$

4.2 舍入噪声增益

舍入噪声增益的计算也是与系统参数相关, 新结构的参数存在于 Q_k 中的 $\{\alpha_k, \beta_k\}$, C 中的 $\{\xi_k\}$ 和 d . 接下来将分析各个参数与信号相乘导致的舍入噪声, 并导出相应的舍入噪声增益表达式.

首先, 对于 Q_1 中的 α_1 , 由于量化的影响, 式 (19) 变为

$$\begin{cases} \tilde{x}^{(1)}(n) = \tilde{x}(n), \\ \tilde{x}^{(m)}(n) = Q_{N-m+2}\tilde{x}^{(m-1)}(n), \quad 2 \leq m \leq N, \\ \tilde{w}(n) = \tilde{x}^{(N)}(n), \\ \tilde{x}(n+1) = q[Q_1\tilde{w}(n)] + Bu(n), \\ \tilde{y}(n) = C\tilde{x}(n) + du(n). \end{cases}$$

定义 $e_1(n) \triangleq q[Q_1\tilde{w}(n)] - Q_1\tilde{w}(n)$, 则 $e_1(n) = v_1\epsilon_{Q_1}(n)$, 其中 $\epsilon_{Q_1}(n)$ 为 Q_1 产生的舍入噪声. 假定 $\Delta x(n) \triangleq \tilde{x}(n) - x(n)$, $\Delta x^{(m)}(n) \triangleq \tilde{x}^{(m)}(n) - x^{(m)}(n)$, $\Delta w(n) \triangleq \tilde{w}(n) - w(n)$ 和 $\Delta y(n) \triangleq \tilde{y}(n) - y(n)$. 分别代入上式可得

$$\begin{cases} \Delta x^{(1)}(n) = \Delta x(n), \\ \Delta x^{(m)}(n) = Q_{N-m+2}\Delta x^{(m-1)}(n), \quad 2 \leq m \leq N, \\ \Delta w(n) = \Delta x^{(N)}(n), \\ \Delta x(n+1) = Q_1\Delta w(n) + v_1\epsilon_{Q_1}(n), \\ \Delta y(n) = C\Delta x(n). \end{cases} \quad (28)$$

不难发现, 式 (28) 对应的实现为

$$\begin{cases} \Delta x(n+1) = A\Delta x(n) + v_1\epsilon_{Q_1}(n), \\ \Delta y(n) = C\Delta x(n). \end{cases}$$

因此, Q_1 带来的舍入噪声增益为

$$G_{Q_1} = \text{tr}(v_1^T W_o v_1) = W_o(1, 1).$$

类似地, 定义 $e_k(n) \triangleq q[Q_{N-k+2}x^{(k-1)}(n)] - Q_{N-k+2}x^{(k-1)}(n)$, $k = 2, \dots, N$. 由式 (18), $e_k(n)$ 为

$$e_k(n) = [\epsilon_{11}(n) + \epsilon_{12}(n)]v_{p-1} + [\epsilon_{21}(n) + \epsilon_{22}(n)]v_p,$$

其中, $p = N - k + 2$, 4 个舍入噪声源 $\epsilon_{ij}(n)$ 是各自独立的噪声源.

采用 $e_1(n)$ 相同的定义, 由 $e_k(n)$ 产生的噪声对应的传输函数为

$$\begin{cases} \Delta x^{(1)}(n) = \Delta x(n), \\ \Delta x^{(m)}(n) = Q_{N-m+2}\Delta x^{(m-1)}(n), \quad \forall m \neq k, \\ \Delta x^{(k)}(n) = Q_{N-k+2}\Delta x^{(k-1)}(n) + e_k(n), \\ \Delta w(n) = \Delta x^{(N)}(n), \\ \Delta x(n+1) = Q_1\Delta w(n), \\ \Delta y(n) = C\Delta x(n), \end{cases}$$

其等价于

$$\begin{cases} \Delta x(n+1) = A\Delta x(n) + P_{N-k+2}e_k(n), \\ \Delta y(n) = C\Delta x(n), \end{cases}$$

这里, $P_m, 2 \leq m \leq N$ 的定义与式 (23) 一样.

根据式 (10), 由于 $Q_k, k = 2, 3, \dots, N$ 所产生的舍入噪声增益为

$$G_{Q_k} = \text{tr}(P_{N-k+1}^T W_o P_{N-k+1} V_k), \quad k = 2, \dots, N, \quad (29)$$

其中, $p = N - k + 2$ 和

$$V_k \triangleq E[e_k(n)e_k^T(n)] = 2(v_{p-1}v_{p-1}^T + v_p v_p^T). \quad (30)$$

而关于新结构 S_{new} 中, B 的第一个元素, C 和 d , 其舍入噪声增益都可以非常简便的计算得到, 其值分别为 $W_o(1, 1)$, N 和 1.

因此, 整个结构的舍入噪声增益为

$$G_{\text{new}} = \sum_{k=1}^N G_{Q_k} + W_o(1, 1) + N + 1. \quad (31)$$

5 全通系统的新结构实现

全通滤波器被广泛应用于数字处理的各个领域, 包括单通道滤波器、均衡器、变压器等等. 在本节, 将给出全通系统的新结构实现.

对于一个任意的 N 阶全通系统, 其传输函数可以描述为 $H_{\text{ap}}(z) = \frac{z^{-N}D(z^{-1})}{D(z)}$, 这里, 假设 $z^{-N}D(z)$ 的根都在单位圆内且不在圆点 $z = 0$.

5.1 全通系统的一些性质

首先将给出定理 2.

定理 2 若任意一个 N 阶的全通滤波器 H_{ap} 对应的状态空间实现为 $R \triangleq (A, B, C, d)$. 进一步地, 若其可控和可观 Gram 矩阵分别用 W_c 和 W_o 表示, 则存在

$$\begin{cases} d^2 + CW_c C^T = 1, \\ dB + AW_c C^T = \mathbf{0}, \end{cases} \quad (32)$$

其中, $\mathbf{0}$ 是全零的向量, 且

$$W_c = W_o = I. \quad (33)$$

证明 参照文献 [2], 由于 $H_{\text{ap}}(z) = d + C(zI - A)^{-1}B \triangleq \sum_{k=0}^{+\infty} q_k z^{-k}$, 则有 $H_{\text{ap}}(z)H_{\text{ap}}(z^{-1}) = p_0 + \sum_{m=1}^{+\infty} p_m(z^{-m} + z^m)$, 这里,

$$p_m = q_0 q_m + \sum_{k=1}^{+\infty} q_{k+m} q_k, \quad \forall m.$$

因为 $q_k = CA^{k-1}B$, $\forall k > 0$ 和 $q_0 = d$, 通过计算可以得到

$$\begin{cases} p_0 = q_0^2 + C \sum_{k=1}^{+\infty} A^{k-1} B B^T (A^{k-1})^T C^T = q_0^2 + C W_c C^T, \\ p_m = q_0 C A^{m-1} B + A^m W_c C^T = C A^{m-1} (q_0 B + A W_c C^T), \quad \forall m \neq 0. \end{cases} \quad (34)$$

$H_{ap}(z)$ 是全通滤波器, 因此 $p_0 = 1$ 和 $p_m = 0, \forall m \neq 0$, 这使得式 (32) 的第一条成立. 因为 $W_o = \sum_{k=1}^{+\infty} (CA^{k-1})^T CA^{k-1}$ 是非奇异的, 则可由 $p_m = 0, \forall m \neq 0$ 直接得出 $q_0 B + A W_c C^T = 0$.

为了论证式 (33), 定义输入平衡实现 $\hat{R} \triangleq (\hat{A}, \hat{B}, \hat{C}, d)$ 由 R 经相似变换 $\hat{T}$ 得到, 则自然满足 $\hat{A}\hat{A}^T + \hat{B}\hat{B}^T = I$. 定义

$$P \triangleq \begin{bmatrix} \hat{A} & \hat{B} \\ v & \varrho \end{bmatrix},$$

这里, $v \in \mathbb{R}^{1 \times N}$, $\varrho \in \mathbb{R}$ 并满足 $PP^T = I_{N+1}$, I_{N+1} 是维数为 $N+1$ 的单位矩阵. 显然

$$\begin{cases} \hat{A}v^T + \varrho\hat{B} = \mathbf{0}, \\ vv^T + \varrho^2 = 1, \end{cases}$$

化简上式可得到

$$\begin{cases} \varrho = \pm[\hat{B}^T(\hat{A}^T\hat{A})^{-1}\hat{B} + 1]^{-1}, \\ v^T = \pm\hat{A}^{-1}\hat{B}. \end{cases}$$

又因为 $PP^T = P^TP = I_{N+1}$, 则

$$\begin{cases} \hat{A}\hat{A}^T + \hat{B}\hat{B}^T = I, & \hat{A}^T\hat{A} + v^Tv = I, \\ \hat{A}v^T + \varrho\hat{B} = \mathbf{0}, & \hat{A}^T\hat{B} + \varrho v^T = 0, \end{cases} \quad (35)$$

显然, 式 (35) 中的前两个式子对于输入平衡系统 $G(z) \triangleq \varrho + v(zI - \hat{A})^{-1}\hat{B}$ 恒成立, 此时 $G(z)$ 的可控和可观 Gram 矩阵满足 $\hat{W}_c = \hat{W}_o = I$.

同时, $G(z)$ 和 $H_{ap}(z)$ 都是全通滤波器且具有相同的极点 (具有相同的状态矩阵), 则存在 $G(z) = \pm H_{ap}(z)$. 不难发现, 只要选择 $C = \varrho, d = v$, 就可使 $G(z) = H_{ap}(z)$. 因此, R 的可控和可观 Gram 矩阵与 $\hat{R}$ 的一样, 分别为单位矩阵. 式 (33) 成立. 证毕.

5.2 新结构的简化形式

根据定理 2, 可以得到定理 3.

定理 3 假设 $R_h \triangleq (A, B, C, d)$ 为 N 阶全通系统对于新结构 S_{new} 的状态空间实现, 则

$$C = \beta_1 v_1^T Q, \quad (36)$$

其中, v_1 就是前文定义的 N 维单位阵的第一列. 此时, 式 (19) 中新结构的输出可简化为

$$y(n) = \beta_1 w_1(n) + du(n), \quad (37)$$

图 3 简化结构 $S_{\text{new}}^{\text{ap}}$ 的结构框图Figure 3 The simplified structure $S_{\text{new}}^{\text{ap}}$ for allpass filter

图 4 数字滤波器的级联实现

Figure 4 The cascade realization of IIR filter

这里, $w_1(n)$ 为向量 $w(n)$ 的第一个元素. 简化后的新结构记为 $S_{\text{new}}^{\text{ap}}$, 其结构框图见图 3. 进一步地, 该结构的舍入噪声增益为

$$G_{\text{new}}^{\text{ap}} = 4N. \quad (38)$$

证明 由于新结构为输入平衡实现, 即满足 $W_c = I$, 由定理 2 可得, $d^2 + CC^T = I$, $dB + AC^T = \mathbf{0}$ 和 $W_o = I$, 也就是 $A^T A + C^T C = I$.

注意到 $dB + AC^T = \mathbf{0}$, 则其等价于 $CA^T A = -dB^T$. 把上面的两式代入即可得到

$$C(I - C^T C) = -dB^T A \Rightarrow C = -d^{-1} B^T A.$$

利用 $A = Q_1 Q$ 和 $d = a_N = \alpha_1$, 则 $C = -d^{-1} B^T A = \beta_1 v_1^T Q$.

不难看出, S_{new} 和 $S_{\text{new}}^{\text{ap}}$ 两个结构的区别在于输出方程. 因为 $S_{\text{new}}^{\text{ap}}$ 的 $W_o = I$, 则式 (29) 变为 $G_{Q_k} = 4, k = 2, 3, \dots, N, G_{Q_1} = W_o(1, 1) = 1$. β_1 产生的舍入噪声增益为 1, 比 S_{new} 结构的全参 C 产生的舍入噪声减少了 $N - 1$. 因此, $S_{\text{new}}^{\text{ap}}$ 结构的舍入噪声增益为 $4N$. 证毕.

由上述内容可见, 对于一个 N 阶的全通滤波器, $S_{\text{new}}^{\text{ap}}$ 和 S_{nl} 具有相同的舍入噪声增益, 同时二者具有相同的实现复杂度: 包含 $4N$ 个乘法器和 $2N$ 个加法器. 见文献 [22]. 与最佳状态空间实现 R_{mrh} 相比, $S_{\text{new}}^{\text{ap}}$ 和 S_{nl} 包含的 $4N$ 舍入噪声增益远优越于前者的 $(N + 1)^2$, 并且在实现复杂度上也较前者有较大改进. 由文献 [25] 可知, 有一大类阶数为奇数的低通数字滤波器, 如 Butterworth 滤波器、Chebyshev 滤波器、椭圆滤波器等, 可分解两个全通滤波器级联, 见图 4. 整个系统的传输函数可表示为 $H(z) = \frac{1}{2}[H_1(z) + H_2(z)]$, 其中 $H_k(z)$ 表示阶数为 $N_k, k = 1, 2$ 的全通子滤波器且 $N = N_1 + N_2$. 如图 4 所示的实现方式称之为 S_{VMN} . 当 $H(z)$ 用 S_{VMN} 实现时, 其总的舍入噪声增益为 $G_{\text{VMN}} = (\frac{1}{2})^2 G_1 + (\frac{1}{2})^2 G_2 + 1$, 这里, G_k 表示 $H_k(z)$ 的舍入噪声增益. 进一步地, 若 $H_1(z)$ 和 $H_2(z)$ 都采用 S_{nl} 和 $S_{\text{new}}^{\text{ap}}$ 实现, $H(z)$ 的舍入噪声增益为 $N + 1$. 此时, 该实现方式每计算一次输出只需做 $4N$ 次乘法, $2N + 1$ 次加法和 1 次移位操作. 因此, S_{VMN} 无论从性能上还是从实现复杂度上都优越于 $S_{\text{new}}^{\text{ap}}$ 和 S_{nl} , 但遗憾的是, S_{VMN} 只适用于一部分有限的数字滤波器.

表 1 新结构 S_{new} 的结构参数

Table 1 The coefficient values of the proposed structure

m	$\sin \phi_m$	$\cos \phi_m$	ξ_m
0	—	—	0.014427493014264
1	0.904965936056452	0.425483986872526	-0.037247062470238
2	0.556303524464708	0.830979174488647	0.083619241461273
3	0.400797025064193	0.916166876018430	-0.201734809075682
4	0.298581868505560	0.954384025327783	0.252418400184547
5	0.451414984388138	0.892314132954024	-0.230422477147784
6	0.086491222166119	0.996252612788254	0.167480285339615
7	0.576204027226514	0.817305890721261	0.056002192351766

表 2 两个全通子滤波器采用 S_{new} 实现时的结构参数

Table 2 The coefficients of the two allpass sub-filters for S_{new}

m	$H_1(z)$		$H_2(z)$	
	$\sin \phi_m$	$\cos \phi_m$	$\sin \phi_m$	$\cos \phi_m$
1	0.7451666	0.6668784	0.7700170	0.6380234
2	0.4801625	0.8771795	0.3541893	0.9351737
3	0.1426787	0.9897691	0.4957126	0.8684866
4	-0.5728920	0.8196309	—	—

表 3 实例 I: 6 种结构的性能比较

Table 3 Statistics for each of the six structures for Example I

Structure	S_{DFIIt}	R_{mrh}	R_{ib}	S_{nl}	S_{new}	$S_{\text{VMN}}^{\text{new}}$
G	8.2151×10^7	19.2149	26.0157	18.2024	14.7041	8
M	2.4212×10^8	93.5447	127.1465	177.4678	134.9451	—
N_{mul}	$3N + 1$	$(N + 1)^2$	$(N + 1)^2$	$5N + 1$	$5N - 1$	$4N$

6 仿真实例

本节将提供两个仿真实例来验证所提结构的性能, 并与经典结构比较, 包括最佳状态空间实现 R_{mrh} , 输入平衡实现 R_{ib} , 归一化格型结构 S_{nl} . 当 N 阶的 IIR 数字滤波器采用 $S_{\text{VMN}}^{\text{new}}$ 实现且子滤波器采用新结构实现时, 这样的结构称之为 $S_{\text{VMN}}^{\text{new}}$.

实例 I 第 1 个实例是一个 7 阶的低通椭圆滤波器, 由 Matlab 直接产生: $\text{ellip}(7, 0.25, 40, 0.2)$, 其通带波纹、阻带波纹、归一化通带带宽分别为 0.25, 40, 0.1.

新结构 S_{new} 的结构参数由表 1 给出.

根据 S_{VMN} 的实现结构, 该滤波器可以被分解为两个全通滤波器级联, 即 $H(z) = \frac{1}{2}[H_1(z) + H_2(z)]$, 这里, $H_1(z), H_2(z)$ 分别为 4 阶和 3 阶的全通滤波器, 且两个全通滤波器都采用 S_{new} 实现. 表 2 给出了相应的结构参数.

表 3 给出了 6 种不同实现结构的性能比较, 其中, G, M 和 N_{mul} 分别表示舍入噪声增益、灵敏度、乘法器个数.

图 5 实例 I 幅频响应比较 ($B_c=9$). 实线 – 理想, 虚线 – S_{new} , 破折线 – R_{ib} , 点划线 – R_{mrh}

Figure 5 Magnitude response comparison for Example I. Solid line for the ideal response; dotted line, dashed line and dashdotted line for the proposed, R_{ib} and R_{mrh} structure implemented with $B_c = 9$ bits, respectively

图 6 实例 I 幅频响应比较. 实线 – 理想, 虚线 – S_{new} ($B_c=9$), 破折线 – S_{nl} ($B_c=9$), 点划线 – S_{DFII_t} ($B_c=15$)

Figure 6 Magnitude response comparison for Example I. Solid line for the ideal response; dashdotted line for S_{DFII_t} structure with $B_c = 15$ bits; dotted line for S_{new} and dashed line for S_{nl} both with $B_c = 9$ bits

表 4 实例 II: 5 种结构的性能比较

Table 4 Statistics for each of the five structures for Example II

Structure	S_{DFII_t}	R_{mrh}	R_{ib}	S_{nl}	S_{new}
G	6.8406×10^3	23.5815	30.8365	20.3082	15.8906
M	1.8847×10^4	54.2544	70.9842	83.2267	46.3561
N_{mul}	$3N + 1$	$(N + 1)^2$	$(N + 1)^2$	$5N + 1$	$5N - 1$

由表 3 可以看出, 除 S_{VMN} 之外, 所提结构 S_{new} 具有最小的舍入噪声增益. 与 R_{mrh} 和 R_{ib} 结构相比, 所提结构的灵敏度虽然略大, 但其只需 $5N - 1$ 个乘法器, 极大地降低了实现复杂度. 同时, 所提结构所需的乘法器个数比 S_{nl} 少两个. 而与 S_{DFII_t} 相比, S_{new} 具有较好的抗 FWL 误差能力. 对于给定的 $H(z)$, 基于 S_{VMN} 形式实现的系统性能往往较直接实现具有更强的抗 FWL 误差能力. 从表 3 可知, $S_{\text{VMN}}^{\text{new}}$ 具有最小的实现复杂度和舍入噪声增益. 若两个子系统都采用 R_{mrh} 和 R_{ib} , 整个系统的舍入噪声增益为 11.25, 且需要 41 个乘法器, 比 $S_{\text{VMN}}^{\text{new}}$ 多了 13 个. 需要特别指出的是, S_{new} , R_{mrh} , R_{ib} 和 S_{nl} 具有同级别的参数灵敏度, 远远优越于 S_{DFII_t} 结构. 如期望, S_{new} , R_{ib} 和 R_{mrh} 三者具有类似的性能, 与理想幅频响应非常接近, 见图 5. 而与 S_{nl} 相比, 所提结构的通带性能较前者好, 见图 6. 这里, B_c 表示系统参数保留的小数位数.

实例 II 第 2 个实例是一个八阶的带通滤波器, 其产生指令为 `ellip(4, 0.25, 40, [0.3 0.5])`. 表 4 给出了 5 种结构的性能比较⁵⁾. 由表 4 可以看出, 与实例 I 类似, 新结构 S_{new} 具有最小的舍入噪声增益和参数灵敏度. 事实上, 以上仿真只局限于 $N = 4$, 对于不同的阶数具有相同的结论, 图 7 和 8 分别给出了 4 种结构的舍入噪声增益和灵敏度随阶数 N 的变化曲线.

5) 对于带通滤波器, S_{VMN} 并不适用.

图 7 舍入噪声增益随阶数 N 的变化曲线

Figure 7 Relationship between roundoff noise gain $\lg(G)$ and filter order N

图 8 灵敏度随阶数 N 的变化曲线

Figure 8 Relationship between sensitivity measure $\lg(M)$ and filter order N

7 结论

本文的主要贡献是导出了一种基于 Hessenberg 的输入平衡实现, 该结构非常稀疏, 只需 $5N - 1$ 个乘法器. 对于全通系统, 该结构可进一步简化, 可以把乘法器个数降低到 $4N$. 本文分析并导出了新结构的舍入噪声增益和传函灵敏度函数的数学表达式. 对于一个 N 阶的全通数字滤波器, 所提结构的舍入噪声增益为 $4N$. 两个仿真实例表明, 与传统经典的结构相比, 所提结构对 FWL 误差具有更强的鲁棒性. 在系统的实现复杂度上, 新结构具有显著优势, 适合应用于实时处理的系统.

致谢 感谢浙江省信号处理重点实验室对本课题的支持.

参考文献

- 1 Roberts R A, Mullis C T. Digital Signal Processing. New Jersey: Addison-Wesley Publishing Company, 1987
- 2 Gevers M, Li G. Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects(Communication and Control Engineering). London: Springer Verlag, 1993
- 3 Rao D V B. Analysis of coefficient quantization errors in statespace digital filters. IEEE Trans Acoust Speech Signal Process, 1986, ASSP-34: 131-139
- 4 何明一, 保铮. 神经网络与信号处理系统 —— 有限精度射击理论. 西安: 西北工业大学出版社, 1998
- 5 Thiele L. On the sensitivity of linear state-space systems. IEEE Trans Circuits Syst, 1986, CAS-33: 502-510
- 6 Xiao C. Improved L_2 -sensitivity for state-space digital system. IEEE Trans Signal Process, 1997, 45: 837-840
- 7 Hinamoto T, Iwata K, Lu W S. L_2 -sensitivity minimization of one- and two-dimensional state-space digital filters subject to L_2 -scaling constraints. IEEE Trans Signal Process, 2006, 54: 1804-1812
- 8 Huang C G, Li G, Xu Z X, et al. Design of optimal digital lattice filter structures based on genetic algorithm. Signal Process, 2012, 92: 989-998
- 9 Mullis C T, Roberts R A. Synthesis of minimum roundoff noise fixed-point digital filters. IEEE Trans Circuits Syst, 1976, CAS-23: 551-562
- 10 Hwang S Y. Minimum uncorrelated unit noise in state-space digital filtering. IEEE Trans Acoust Speech Signal Process, 1977, ASSP-25: 273-281
- 11 Wong N, Ng T S. A generalized direct-form delta operator-based IIR filter with minimum noise gain and sensitivity. IEEE Trans Circuits Syst II, 2001, 48: 425-431

- 12 Li G, Chu J, Wu J. A matrix factorization-based structure for digital filters. *IEEE Trans Signal Process*, 2007, 55: 5108–5122
- 13 Li G, Huang C G, Hua J, et al. An improved orthogonal digital filter structure. *Signal Process*, 2010, 90: 679–683
- 14 Higgins W E, Munson D C. Noise reduction strategies for digital filters: error spectrum shaping versus the optimal linear state-space formulation. *IEEE Trans Acoust Speech Signal Process*, 1982, ASSP-30: 963–973
- 15 Hinamoto T, Ohnishi H. Roundoff noise minimization of state-space digital filters using separate and joint error feedback/ coordinate transformation. *IEEE Trans Circuits Syst I*, 2003, 50: 23–33
- 16 Lu W S, Hinamoto T. Jointly optimized error-feedback and realization for roundoff noise minimization in state-space digital filters. *IEEE Trans Signal Process*, 2005, 53: 2135–2145
- 17 Yu Y J, Lim Y C. Design of linear phase FIR filters in subexpression space using mixed integer linear programming. *IEEE Trans Circuits Syst I*, 2007, 54: 2330–2338
- 18 Yan S, Ma Y. Optimal design and verification of temporal and spatial filters using second-order cone programming approach. *Sci China Ser F-Inf Sci*, 2006, 49: 235–253
- 19 Gao X, Zhong H. Parameterization of 3-channel non-separable 2-D wavelets and filters. *Sci China Ser F-Inf Sci*, 2004, 47: 362–371
- 20 Skaf J, Boyd S P. Filter design with low complexity coefficients. *IEEE Trans Signal Process*, 2008, 56: 3162–3169
- 21 Li G, Lim Y C, Huang C G. Very robust low complexity lattice filters. *IEEE Trans Signal Process*, 2010, 58: 6093–6104
- 22 Li G, Hua J Y, Bi G. An efficient and robust orthogonal structure for linear discrete-time systems. *J Circuits Syst Signal Process*, 2010, 29: 727–743
- 23 Huang C G, Lim Y C, Li G. A generalized lattice filter for finite wordlength implementation with reduced number of multipliers. *IEEE Trans Signal Process*, 2014, 62: 2080–2089
- 24 Gray Jr A, Markel J. A normalized filter structure. *IEEE Trans Acoust Speech Signal Process*, 1975, ASSP-23: 268–277
- 25 Vaidyanathan P P, Mitra S K, Neuvo Y. A new approach to the realization of low-sensitivity IIR digital filters. *IEEE Trans Acoust Speech Signal Process*, 1986, ASSP-34: 350–361

An input-balanced realization-based structure for efficient implementation of digital IIR filters

HUANG ChaoGeng^{1*}, CHEN YanHong¹, XU Hong², HE XiongXiong² & ZHOU ZhiGuang¹

¹ School of Information, Zhejiang University of Finance & Economics, Hangzhou 310018, China;

² College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

*E-mail: cghuang@zufe.edu.cn

Abstract Based on an input-balanced state-space realization, this paper derives a novel structure for efficient implementation of digital IIR filters. Such a structure is very similar to a normalized lattice and, for an N th order IIR filter, it possesses $5N - 1$ non-trivial multipliers: two multipliers less than in a normalized lattice. Expressions for transfer function sensitivity measures and roundoff noise gain for the proposed structure are derived. Interestingly, it is shown that, when the filter is all-pass, the proposed structure requires $4N$ multiplications in order to compute one output sample. Moreover, the corresponding roundoff noise gain is $4N$, much smaller than that of $(N + 1)^2$ given by the classical minimum roundoff state-space realizations. Results of simulations performed show that the proposed structure outperforms many classical ones in terms of implementation efficiency and robustness against finite-word-length effects.

Keywords state-space realization, digital filter structures, finite word length, input-balanced realization, all-pass filters

HUANG ChaoGeng received a B.E. degree in electronic and information engineering and a Ph.D. degree in control theory and control engineering, both from Zhejiang University of Technology, Zhejiang, China, in 2007 and 2013, respectively. He has been with the School of Information, Zhejiang University of Finance & Economics, since 2013. His research interests are in the areas of digital signal processing and control theory.

CHEN YanHong received an M.S. degree in technology of computer applications from Northeast Dianli University, Jilin, China, in 2005. He has been with the School of Information, Zhejiang University of Finance & Economics, since 2005. His research interests are in the areas of design and analysis of algorithms.

XU Hong received an M.S. degree in information and communication engineering from Zhejiang University, Zhejiang, China, in 2004. She has been with the College of Information Engineering, Zhejiang University of Technology, since 2004. Her research interests are in the areas of design and implementation of digital filters.

HE XiongXiong received his M.S. degree from the Institute of Automation, Qufu Normal University, China, in 1994, and his Ph.D. degree from the Institute of Industrial Control Technology, Zhejiang University, China, in 1997. From 1998 to 2000, he was a post-doctoral fellow at the Harbin Institute of Technology, China. Since 2000, he has been with the College of Information Engineering, Zhejiang University of Technology, where he is currently Professor and Director of the Institute of Intelligence Systems and Robotics. His research interests include robotics, intelligence systems, and signal processing.