

可逆线性变换的整数实现*

郝鹏威 石青云

(北京大学信息科学中心, 视觉与听觉信息处理国家重点实验室, 北京 100871)

摘要 考虑一般线性变换的整数实现问题. 引入了整数因子和基本三角阵两个概念. 证明只要这种线性变换是有限维的和可逆的, 都存在相应的整数实现途径. 整数实现的线性变换一般都具有如下优点: (i) 把整数全部映射成整数; (ii) 仍然完全可逆; (iii) 原位计算, 无需申请任何新的空间. 给出了线性变换整数实现的构造性方法和一种优化的实现方法, 这些实现方法均可以自动生成. 在此基础上, 讨论了线性变换整数实现时所产生误差的估计. 最后, 给出了一个线性变换整数实现的例子.

关键词 线性变换 整数映射 三角分解 完全可逆

通过变换来对数字信号进行无损压缩编码或用硬件来实现信号变换时, 由于计算机的计算精度是有限的, 所用的变换要求必须是把整数映射为整数的变换, 而且必须仍然是可逆的. 在基于感兴趣区域的渐进性图像压缩等应用中, 有损和无损压缩是要求同时兼顾的. 因此, 整数变换对于数字信号的压缩有着重要的意义.

最初研究变换的整数实现方法, 只是针对一些简单变换, 通过发现它的某种近似的整数一一对应关系来建立, 如 Zandi 等人^[1]提到的 Blume 等人^[2]的 S 变换及他们自己提出的 TS 变换, Gormish 等人^[3]的颜色空间变换的近似整数实现. Bruekers 等人^[4]梯形结构的完全重构滤波网络的提出, 为这方面的研究开辟了一条新的途径. 但这一成果在变换的整数实现方面的应用并未得到足够的重视. 最早用类似途径系统地研究变换的整数实现方法的是在离散小波变换领域, 而在小波变换中最初是使用一种类似的网络来研究小波变换消失矩的可自动提升性和小波的构造^[5], 该实现方法被称为提升方法 (lifting scheme). Dewitte 等人^[6]似乎是最先应用提升方法实现小波整数变换的. 之后, Daubechies 等人^[7,8]及石青云^[9]将该方法推广和应用到一般小波变换情形, 并给出了小波变换自动转化为整数实现的步骤和方法. 这些成果均得到了广泛的认同和采纳. 后来闫宇松和石青云又将这一方法应用于离散 Fourier 变换 (FFT) 和离散余弦变换 (DCT) 的整数实现问题, 均得到了很好的结果.

本文则考虑了一般线性变换的整数实现问题. 通过研究可逆线性变换的基本结构, 指出梯形结构是整数实现最恰当的基本结构. 然后, 研究了可逆线性变换分解为一系列可整数实现的基本结构的机械化方法, 并给出了一种优化的分解策略. 毕竟, 整数实现相对于无限精度

的符号计算是有误差的,本文也讨论了其中误差的估计问题,并给出如何减少误差的一些原则性的建议. 最后,本文给出了一个例子来说明本文方法的有效性和可操作性.

1 可逆线性变换整数实现的基本结构

在以下的讨论中,线性变换为复数域内的变换,整型复数或整数是指复数的实部和虚部都是整数,复数的取整表示对复数的实部和虚部分别进行取整. 变换的完全可逆是指正变换和逆变换两个过程都是可逆的.

对于数 x 到数 y 的线性变换 $y = ax + b$, 如果 x 可取任意整数, 则使 y 成为整数的条件是 a 和 b 均为整数. 因为有可能在变换域执行某种处理, y 的值域(变换域)一般也认为是不加任何限制的整数集合, 所以, 该线性变换对应的逆变换把整数映射成整数的条件便是 $1/a$ 和 b/a 均为整数. 因此, 该线性变换整数完全可逆的条件便是 $a = \pm 1$ 或 $\pm i$ 且 b 为整数. 其中, a 和 $1/a$ 的取值范围为 $\{1, -1, i, -i\}$, $i = \sqrt{-1}$, 统称为整数因子. 为以下叙述方便, 我们用 j 来表示整数因子, 用 j' 来表示其倒数. 显然, j' 也是整数因子. 如果考虑数 x 和数 y 是 Z 变换的形式, 整数因子还可以是 js^k , 其中, k 为整数.

如果 b 不是整数, 其计算仍然可以是浮点的, 但只要用取整结果 $[b]$ 代替它进行变换运算, 仍可保证变换是整数的和可逆的. 取整方法可以是四舍五入取整、截尾取整、进位取整或其他取整形式.

因此, 可逆线性变换可整数实现的基本形式一般为: $y = jx + [b]$. 相应逆变换的基本形式为: $x = j'(y - [b])$. 其计算网络如图 1 所示.

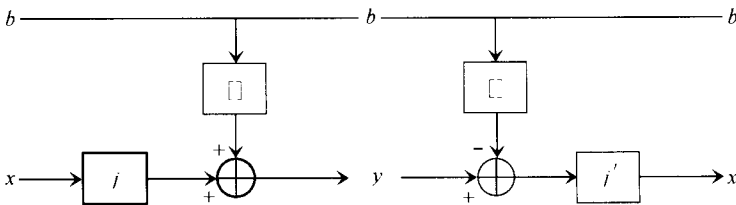


图 1 数的线性变换及其逆变换整数实现的计算网络结构

从图中可逆线性变换可整数实现的基本结构可以看出, 正逆两种变换的差异只有传输方向、数 b 取整结果的符号和整数因子(如果是 $\pm i$)的不同. 数 b 还必须满足两个条件:(i) 对于正变换, b 是不直接或间接地依赖于 x 的;(ii) 对于逆变换, b 是不直接或间接地依赖于 y 的. 无论变换是实数域的还是复数域的, 该计算网络结构是一种梯形结构. 在 $a = 1$ 时的计算网络与 Bruekers 等人^[4]的梯形结构和 Sweldens^[5]的提升方法的计算网络结构是一致的. 因此可以说, 梯形结构是可逆线性变换整数实现最恰当的基本结构.

对于一般 N 维可逆线性变换 $y = Ax$, 如果它是可以直接整数实现的和完全可逆的, 则其计算过程必满足上述数 b 的两个条件. 在适当安排计算次序的情况下, 一种整数实现且完全可逆的基本变换矩阵就是对角线上的元素都是整数因子的上三角矩阵或下三角矩阵, 我们统称之为基本三角阵. 如果对角线上的整数因子都是 1, 就是单位三角阵.

基本三角阵具有如下两个重要的性质:(i) 两个基本上三角阵的乘积仍然是基本上三角阵; 两个基本下三角阵的乘积仍然是基本下三角阵. (ii) 基本三角阵的行列式值是整数因

子.

如果 A 是基本上三角阵, 则 N 维可逆线性变换 $y = Ax$ 的计算次序为自上而下:

$$y_m = jx_m + \left[\sum_{n=m+1}^N a_{mn}x_n \right] = jx_m + [b], \quad m = 1, 2, \dots, N; \quad (1)$$

逆变换的计算次序刚好与正变换相反:

$$x_m = j' \left(y_m - \left[\sum_{n=m+1}^N a_{mn}x_n \right] \right) = j' (y_m - [b]), \quad m = N, N-1, \dots, 1. \quad (2)$$

如果 A 是基本下三角阵, 则变换 $y = Ax$ 及其逆变换的计算次序与基本上三角阵相反:

$$y_m = jx_m + \left[\sum_{n=1}^{m-1} a_{mn}x_n \right] = jx_m + [b], \quad m = N, N-1, \dots, 1, \quad (3)$$

$$x_m = j' \left(y_m - \left[\sum_{n=1}^{m-1} a_{mn}x_n \right] \right) = j' (y_m - [b]), \quad m = 1, 2, \dots, N. \quad (4)$$

2 可逆线性变换的分解

任意有限维的可逆线性变换是否可以用整数映射来实现, 在于该变换是否存在一系列可整数实现的基本结构的分解, 即变换矩阵是否存在一系列可整数实现的基本三角阵的分解. 对于有限维的线性变换 $y = Ax$, 问题便是 A 是否存在有限步的分解: $A = V_1 V_2 \cdots V_M$ 或 $A = PV_1 V_2 \cdots V_M$, 其中, P 为置换矩阵, 所有 V_i 均为基本三角阵.

对于任意 $N \times N$ 的可逆矩阵 A , 若 $|\det A| = \alpha \neq 1$, 则可以通过改造变换矩阵将其变为行列式值的模为 1 的矩阵. 如果不考虑变换结果所对应的物理意义, 变换的改造方法为对 A 中的 $m (\leq N)$ 行分别除以常数 $\alpha_1, \alpha_2, \dots, \alpha_m$, 这些常数满足

$$\prod_{k=1}^m |\alpha_k| = |\det A| = \alpha. \quad (5)$$

这种改造所具有的一个最大的优点是, 可以对变换各个分量的动态范围加以控制. 这对于仅仅用于无损压缩等不顾及变换的物理意义的应用, 将会非常有益.

考虑到变换结果的实际意义, 就不可以改变变换矩阵特征值的相对大小, 只能对变换进行比例改造, 即取 $\alpha_1 = \alpha_2 = \dots = \alpha_N = \sqrt[N]{\alpha}$. 因此, 以下讨论只针对 $|\det A| = 1$ 的情形. 如果 $|\det A| = 1$, 则存在如下定理:

定理 矩阵 A 存在有限步的基本三角阵分解 $A = PV_1 V_2 \cdots V_M D_R$ 的充分必要条件是 $|\det A| = 1$. 其中, M 有限, $V_k (k = 1, 2, \dots, M)$ 为基本三角阵, P 为置换矩阵, D_R 为一个复数的旋转变换矩阵.

证明 必要性是显而易见的. 下面给出该定理充分性的构造性证明.

由矩阵理论可知, 对非奇异矩阵 A , 必存在三角分解 $A = PLDU$, 其中 P, L, D, U 分别为置换矩阵、单位下三角阵、对角阵和单位上三角阵. $|\det D| = |\det PD| = |\det PLDU| = |\det A| = 1$. 假设 $D = \text{diag}(d_1, d_2, \dots, d_N)$, $\lambda_n = d_1 d_2 \cdots d_n (n = 1, 2, \dots, N)$, $D_R = \text{diag}(1, \dots, 1, \lambda_N)$, 并令

$$D_0 = \begin{cases} \text{diag}(\lambda_1, 1/\lambda_1, \lambda_3, 1/\lambda_3, \dots, \lambda_{N-1}, 1/\lambda_{N-1}), & N \text{ 为偶数,} \\ \text{diag}(\lambda_1, 1/\lambda_1, \lambda_3, 1/\lambda_3, \dots, \lambda_{N-2}, 1/\lambda_{N-2}, 1), & N \text{ 为奇数,} \end{cases}$$

$$D_E = \begin{cases} \text{diag}(1, \lambda_2, 1/\lambda_2, \lambda_4, 1/\lambda_4, \dots, \lambda_{N-2}, 1/\lambda_{N-2}, 1), & N \text{ 为偶数,} \\ \text{diag}(1, \lambda_2, 1/\lambda_2, \lambda_4, 1/\lambda_4, \dots, \lambda_{N-1}, 1/\lambda_{N-1}), & N \text{ 为奇数,} \end{cases}$$

则有 $D = D_O \cdot D_E \cdot D_R$ 或 $D = D_E \cdot D_O \cdot D_R$ 及 $\det D = \lambda_N = \det D_R = e^{i\theta}$. 其中, D_R 是一个对最末分量进行复数旋转变换的对角阵.

除了文献[8]所提供的分解形式外,两个对角元素互为倒数的二阶矩阵存在多种单位三角分解:

$$\begin{aligned} \begin{bmatrix} \alpha & 0 \\ 0 & 1/\alpha \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ (1/\alpha) & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \alpha & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/\alpha \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ -1/\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & \alpha - 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & (1/\alpha) - 1 \\ 0 & 1 \end{bmatrix} \\ &= \dots = V_1 V_2 V_3 V_4, \end{aligned} \tag{6}$$

因此,采用 2×2 的分块技巧, D_O 和 D_E 就可分别分解为 4 个单位三角矩阵的乘积: $D_O = V_{O1} V_{O2} V_{O3} V_{O4}$ 及 $D_E = V_{E1} V_{E2} V_{E3} V_{E4}$. 这样,整个分解便成为

$$A = PLDU = PLV_{O1} V_{O2} V_{O3} V_{O4} V_{E1} V_{E2} V_{E3} V_{E4} D_R U \tag{7}$$

或

$$A = PLV_{E1} V_{E2} V_{E3} V_{E4} V_{O1} V_{O2} V_{O3} V_{O4} D_R U, \tag{8}$$

其中对角阵 D_R 可以放到最末尾, $D_R U = U_R D_R$, 然后,最左边的两个单位下三角阵的乘积可合并为一个单位下三角阵 V_1 ,最右边的两个单位上三角阵的积可合并为一个单位上三角阵 V_8 ,再令中间的 6 个单位三角阵分别为 V_2, \dots, V_7 ,则整个变换的分解就是:

$$A = PV_1 V_2 V_3 V_4 V_5 V_6 V_7 V_8 D_R. \tag{9}$$

其结构与希望所证一致. 定理得证.

由这一分解可见,除了旋转变换和位置交换外,可以分解为最多 8 步基本整数变换操作. 如果 $e^{i\theta}$ 是整数因子,则可以将它直接与最末一个单位上三角阵的最右列元素合并,从而省去旋转变换操作. 从上述定理中去除旋转变换矩阵可以很容易得到以下两个推论:

推论 1 矩阵 A 存在有限步的基本三角阵分解 $A = PV_1 V_2 \dots V_M$ 的充分必要条件是 $\det A$ 为整数因子.

推论 2 矩阵 A 存在有限步的单位三角阵分解 $A = PV_1 V_2 \dots V_M$ 的充分必要条件是 $\det A = 1$.

如果 $e^{i\theta}$ 不是整数因子,则可用对复数的实部和虚部进行旋转变换来实现整数变换,而旋转变换只需 3 步基本整数变换操作也可以实现,因为旋转变换矩阵也存在多种单位三角分解^[8]:

$$\begin{aligned} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ (1 - \cos\theta)/\sin\theta & 1 \end{bmatrix} \begin{bmatrix} 1 & -\sin\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ (1 - \cos\theta)/\sin\theta & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & (\cos\theta - 1)/\sin\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sin\theta & 1 \end{bmatrix} \begin{bmatrix} 1 & (\cos\theta - 1)/\sin\theta \\ 0 & 1 \end{bmatrix} \\ &= \dots = V_{R1} V_{R2} V_{R3}. \end{aligned} \tag{10}$$

这样就证明任意有限维的可逆线性变换只要行列式绝对值为 1 就都存在整数实现的途径. 对

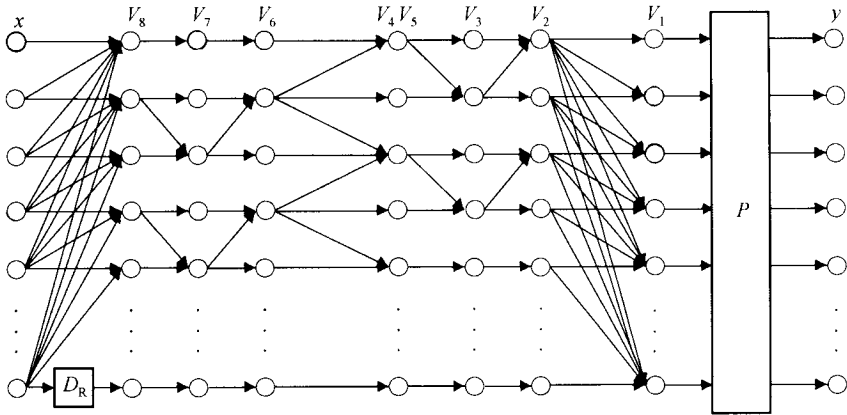


图 2 一般线性变换整数实现的计算流程图

现的. 如果用选主元的 Gauss 消去方法, 还可以使三角分解中的单位下三角矩阵对角线以外的元素的绝对值都不大于 1, 这对于线性变换整数实现的误差控制非常有益. 但是, 可以肯定地说, 优化的分解会有多种形式, 最优的分解形式还有待进一步研究.

4 整数实现的误差估计

由于在每一步变换中对累加数进行了取整, 这使得变换虽然整数化、可逆, 但相对于无限精度的原线性变换结果, 必然是有误差的. 如果采用四舍五入原则进行复数的取整, 则取整结果的实部和虚部的误差都在区间 \$[-0.5, 0.5]\$ 内, 而截尾取整则使之在 \$[0, 1)\$ 区间内. 这些区间实际上对应一个半开的复数矩形区域. 设变换矩阵 \$A\$ 存在有限步的基本三角阵分解 \$A = PV_1V_2 \cdots V_M\$, 用 \$u_m\$ 表示用 \$V_m\$ 进行第 \$m\$ 步基本变换时所产生的取整误差, 用 \$u\$ 表示整个变换所产生的误差, 并记 \$V_0 = P\$, 则整个变换的误差累进结果为

$$u = V_0 \cdot (u_1 + V_1 \cdot (u_2 + V_2 \cdots (u_{M-1} + V_{M-1} \cdot u_M) \cdots)) = \sum_{m=1}^M \left(\prod_{k=0}^{m-1} V_k \right) \cdot u_m. \quad (12)$$

用 \$\| \cdot \|_\infty\$ 表示矩阵或向量的无穷范数, 用 \$U\$ 表示取整误差的实部或虚部绝对值的上界, 由于 \$\| \operatorname{Re}(u_m) \|_\infty \leq U, \| \operatorname{Im}(u_m) \|_\infty \leq U, \| u_m \|_\infty \leq \sqrt{2} \cdot U\$, 则

$$\| u \|_\infty = \left\| \sum_{m=1}^M \left(\prod_{k=0}^{m-1} V_k \right) \cdot u_m \right\|_\infty \leq \sum_{m=1}^M \left\{ \left\| \prod_{k=0}^{m-1} V_k \right\|_\infty \cdot \| u_m \|_\infty \right\} \leq \sqrt{2} \cdot U \cdot \sum_{m=1}^M \left\| \prod_{k=0}^{m-1} V_k \right\|_\infty. \quad (13)$$

通过把复矩阵分为实矩阵和虚矩阵之和, 整个整数变换实部和虚部的误差界也可类似地分别给出.

显然, 如果 \$V_m\$ 的某一行均为整数, 则 \$u_m\$ 的相应分量为 0; 若 \$V_m\$ 的所有元素全是整数, 则 \$u_m = 0\$, 即该矩阵不产生新的取整误差, 但该矩阵仍然传递和累进已有误差. 由一般可逆线性变换的基本三角分解式((7)~(9)式)可知, \$V_1, \dots, V_8\$ 中至少可以有两个三角阵的所有元素全是整数而不产生任何新的取整误差. 因此, 整个变换的误差显然是不会达到上式所显示的误差上界的.

由这两个误差式可见:分解的三角阵越靠左就对最终误差的影响越大. 因此,优化分解应遵循的一些原则包括:(i) 大数靠近原则. 分解的三角阵中有绝对值大于1的数时,该数越大该阵应越靠后,以免放大累积误差. 最好是绝对值最大的数出现在最右边的三角阵中.(ii) 大数避开原则. 如果某个三角阵的某行已出现一个绝对值大的数,则在其左边的三角阵相应列上应尽量避免出现大数,最好是绝对值小于1的数.(iii) 整数优先原则. 能采用整数的地方,最好采用整数,以免取整出现误差.(iv) 主元分解原则. 在对变换矩阵作三角分解时,应使用选主元消去法,这样可使最左边单位下三角阵的元素绝对值都不大于1.

5 实例

为说明本文方法的有效性和可操作性,选择复数域 4×4 的离散 Fourier 变换作为本文的一个实例来构造和分析其整数实现的途径.

4×4 的离散 Fourier 变换矩阵及其逆变换矩阵为:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}, \quad A^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.$$

虽然 A 不加分解都是可以把整数映射为整数的,但从 A^{-1} 看,它不是直接整数完全可逆的. 由于 $\det A = 16i$, 所以其整数实现需要对变换进行改造.

如果不考虑变换结果的物理意义,可以改造变换使得整数实现步数减少到 2. 由于 $A = PLDU$, $D = \text{diag}(1, -1-i, 2-2i, -4i)$, $P = I$, 则对上述变换矩阵有:

$$PD^{-1}P^T A = \begin{bmatrix} 1 & & & \\ (-1+i)/2 & 1 & & \\ (1+i)/4 & (-1-i)/2 & 1 & \\ i/4 & (-1-i)/4 & (1-i)/2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1-i & -i \\ & & 1 & -i \\ & & & 1 \end{bmatrix}.$$

由于 $u_2 = 0$, 如此改造后的变换其整数实现的误差 $u = u_1$. 其中,第 1 分量的误差为 0.

若原向量各分量的动态范围都是 $[0, L]$, 则变换后各分量的动态范围分别为 $[0, 4L]$, $[-L(1+i), L(1+i)]$ (复数的矩形区域), $[-2L, 2L]$, $[-L(1+i), L(1+i)]$. 如果希望其变换后的第 1, 3 分量的动态范围分别为 $[0, L]$ 和 $[-L/2, L/2]$, 则变换的改造可以是乘一个如下对角阵: $D = \text{diag}(1/4, 1, 1/4, 1)$, 这也可使 $\det(DA) = \det D \cdot \det A = i$ 为整数因子,但其分解形式就会复杂一些:

$$DA = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1/4 & (1+i)/8 & 1 & \\ 1/4 & (1-i)/8 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -i & -1 & i \\ 2i & 0 & -2i & \\ & 1/2 & -1/2 & \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ 1 & & & \\ 1/4 & (1+i)/8 & & 1 \\ 1/4 & (1-i)/8 & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 2 & & \\ & & 1/2 & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -i & -1 & i \\ & i & 0 & -i \\ & & 1 & -1 \\ & & & 1 \end{bmatrix},$$

其中间的对角阵用对角元素互为倒数的二阶矩阵的单位三角分解公式就可以分解为 4 个单位三角矩阵的积：

$$\begin{bmatrix} 1 & & & \\ & 2 & & \\ & & 1/2 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & -1/2 & & 1 \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & -1/2 & \\ & & & 1 \\ & & & 1 \end{bmatrix}.$$

再把最左边的 2 个下三角矩阵合并，最右边的 2 个上三角矩阵合并，则有

$$DA = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ 1 & & & \\ 1/4 & (-3+i)/8 & & 1 \\ 1/4 & (-3-i)/8 & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & -1 & \\ & & & 1 \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -i & -1 & i \\ & i & -1/2 & 1/2 - i \\ & & 1 & -1 \\ & & & 1 \end{bmatrix}.$$

此时, $u_2 = u_3 = 0$, 相应误差贡献为 0. u_4 除第 2 分量实部的误差可能是 ± 0.5 外, 其他分量全为 0, 相应累积误差为 0 或 $[\pm(1-i)/8, 0, \pm(1+i)/8, \pm 1]^T$. u_1 的第 1, 2 分量为 0, 可能不为 0 的第 3, 4 分量只置换影响到最终结果的第 1, 3 分量. 因此, 整数变换总误差的各分量实部的绝对值上界按 (12) 和 (13) 式可以分别计算得到 $[\frac{1}{8} + U, 0, \frac{1}{8} + U, 1]^T$, 各分量虚部的绝对值上界可以分别是 $[\frac{1}{8} + U, 0, \frac{1}{8} + U, 0]^T$.

如果考虑到变换结果的意义, 就只能对变换进行比例改造:

$$\frac{1}{\sqrt[4]{16}}A = \frac{1}{2}A = \begin{bmatrix} 1 & & & \\ 1 & & & \\ 1 & 1-i & & 1 \\ 1 & -i & -i & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ & (-1-i)/2 & -1 & (-1+i)/2 \\ & & 1-i & -1-i \\ & & & -2i \end{bmatrix} \\ = \begin{bmatrix} 1 & & & \\ 1 & & & \\ 1 & 1-i & & 1 \\ 1 & -i & -i & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & & & \\ & 1/(1-i) & & \\ & & 1-i & \\ & & & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ & -1 & i-1 & i \\ & & 1 & -i \\ & & & -i \end{bmatrix},$$

其中间的对角阵用两个对角元素互为倒数的二阶矩阵的单位三角分解公式就可以分解为 4 个单位三角矩阵的积：

$$\begin{bmatrix} 1/2 & & & \\ & 1/(1-i) & & \\ & & 1-i & \\ & & & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & -i & 1 & \\ 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 1 \\ & 1 & 1 & 0 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} \\ \cdot \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & (i-1)/2 & 1 & \\ -1/2 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & -2 \\ & 1 & i-1 & 0 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}$$

本式最左边的下三角矩阵与前一式最左边的下三角矩阵合并为一个下三角矩阵,本式最右边的上三角矩阵与前一式最右边的上三角矩阵合并为一个上三角矩阵,则整个变换矩阵就分解成为 4 个基本三角矩阵的乘积:

$$\frac{1}{2}A = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 1-2i & 1 & \\ 2 & -1-i & -i & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 1 \\ & 1 & 1 & 0 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & (i-1)/2 & 1 & \\ -1/2 & 0 & 0 & 1 \end{bmatrix} \\ \cdot \begin{bmatrix} 1 & 1 & 1 & 1+2i \\ & -1 & 2i-2 & 1+2i \\ & & 1 & -i \\ & & & -i \end{bmatrix}$$

对于这一分解, $u_1 = u_2 = u_4 = 0$, 相应累进误差为 0. u_3 的第 1,2 分量为 0, 可能不为 0 的第 3 分量的实部和虚部和第 4 分量的实部和虚部的误差都可能是 ± 0.5 . 因此, 整数变换总误差的各分量的实部和虚部绝对值的上界分别是 $[0.5, 1, 2.5, 3]^T$.

6 结论

- (i) 线性变换只要是有限维的和可逆的, 都存在整数实现形式.
- (ii) 线性变换的整数实现是完全可逆的, 而且可以原位计算.
- (iii) 线性变换矩阵的行列式值为整数因子时, 整数实现的基本操作除位置交换外不超过 7 步.
- (iv) 线性变换分解的优化还有待进一步的研究.

参 考 文 献

- 1 Zandi A, Allen J, Schwartz E, et al. CREW: Compression with reversible embedded wavelets. In: James A S, Martin C, eds. Proceedings of IEEE Data Compression Conference. Snowbird: IEEE Computer Society Press, 1995. 212 ~ 221
- 2 Blume H, Fand A. Reversible and irreversible image data compression using the S-transform and Lempel-Ziv coding. Proceedings of SPIE, 1989, 1 091: 2 ~ 18
- 3 Gornish M J, Schwartz E L, Keith A F, et al. Lossless and nearly lossless compression for high quality images. Proceedings of SPIE, 1997, 3 025: 62 ~ 70
- 4 Bruekers F A M L, van den Enden A W M. New networks for perfect inversion and perfect reconstruction. IEEE J on Selected Areas

- in Communications, 1992, 10(1): 130 ~ 137
- 5 Sweldens W. The lifting scheme: A custom-design construction of biorthogonal wavelets. J of Applied and Computational Harmonic Analysis, 1996, 3(2): 186 ~ 200
 - 6 Dewitte S, Cornelis J. Lossless integer wavelet transform. IEEE Signal Processing Letters, 1997, 4(6): 158 ~ 160
 - 7 Calderbank A R, Daubechies I, Sweldens W, et al. Wavelet transforms that map integers to integers. J of Applied and Computational Harmonic Analysis, 1998, 5(3): 332 ~ 369
 - 8 Daubechies I, Sweldens W. Factoring wavelet transforms into lifting steps. J of Fourier Analysis and Application, 1998, 4(3): 247 ~ 269
 - 9 Shi Qingyun. Biorthogonal wavelet theory and techniques for image coding. Proceedings of SPIE, 1998, 3 545: 24 ~ 32