

# 基于整数小波变换的地震数据压缩<sup>\*</sup>

王喜珍 滕云田 高孟潭 姜 慧

(中国北京 100081 中国地震局地球物理研究所)

**摘要** 由于地震数据的特殊性,在某些情况下需要进行无损压缩.本文通过应用提升方法构造的整数小波变换无损压缩地震数据,与传统的无损压缩算法相比,可以较好地提高压缩率. CDF(2,  $n$ ) 系列双正交小波可以较好地消除数据间冗余,集中数据特征,比其它 CDF 系列、SWE 和 CRF 双正交小波有更好的压缩效果. CDF(2,  $n$ ) 系列双正交小波更适合作为整数小波变换无损压缩地震数据的小波函数.

**关键词** 无损压缩 整数小波变换 提升方法 双正交小波

**中图分类号**: P315.61

**文献标识码**: A

## 引言

随着地震观测技术的发展和应用领域的深入,特别是地震台网的数字化,数据精度的提高,观测数据呈几何级数增长,给存储和传输带来了一定的困难.地震数据与医学图象数据一样,在许多情况下要求无损压缩,以保留全部信息.总体上讲,传统的压缩方法主要采用高效编码数据和消除数据间冗余两种方法.高效编码方法如 miniSEED 地震数据格式采用的 Steim-1、Steim-2 压缩算法(陈运泰等, 2000)、李沙白等(1994)提出的差分编码实时压缩算法、哈弗曼编码、算术编码和基于字典编码,以及通过这些方法引伸出来的其它方法(袁玫, 袁文, 1995)都是减少数据的有效编码长度来实现压缩.但它们的压缩率一般都不高,不能满足台网间海量数据共享的要求.另一种有效的压缩方法就是消除数据间的冗余.地震观测数据往往不是独立的,数据在空间和时间上具有一定的相关性,因此,可以通过预测或变换提取特征数据,以达到压缩数据的目的.预测编码是通过设计一定的预测器进行编码的方法,如自适应线性预测编码(Mandyam *et al*, 1996)、预测树方案编码(Memon *et al*, 1994).变换方法可以在变化域上使数据间的相关性最小,如 DCT 变换(Wang, Wu, 2000)、Walsh 变换(Wood, 1974)、小波变换(Villasenor *et al*, 1996).预测过程往往不是完全可逆的,因此不能完全重构原始信号.变换方法在压缩之前先将数据变换到变换域,再对系数进行量化和编码.由于变换方法往往涉及到浮点运算,在一般情况下是一种不完全可逆的过程,因此在信号重构过程中会引起信号畸变.小波变换可以在小波域上使很少的数据集中了全部数据的大部分能量,因此可以有效地减少编码数据量,提高压缩比,是普遍采用的变换方法.近年来,随着小波理论和技术的深入研究,人们找到了一种从整数到整数的小波变换方法,就是整数小波变换.整数小波变换由于是一种完全

<sup>\*</sup> 中国地震局地球物理研究所论著 04AC1013.

2003-06-26 收到初稿, 2003-10-15 收到修改稿, 2004-02-02 决定采用.

可逆的小波变换，可以用于无损压缩地震数据。

本研究提出一种基于整数小波变换的压缩地震数据的方法。首先，基于提升方法思想，采用 CDF(Cohen-Daubechies-Feauveau)、SWE 和 CRF 双正交小波对地震数据进行整数小波变换和压缩编码。其结果与常用的基于 LZW、HUFFMAN、ARITH 的压缩编码作比较。其次，对不同双正交小波间作压缩效果比较。实验结果表明，基于整数小波变换的压缩比其它几种压缩编码有更高的压缩比。同时，CDF(2,  $n$ ) 系列双正交小波比其它几种小波有较好的压缩效果。

1 提升方法和整数小波变换

为了提高计算离散小波的效率和获取离散小波的特殊性能，Sweldens(1995)提出了提升方法(lifting scheme)。其思想是由简单的多分辨分析逐渐向具有某种特定性能的多分辨分析集中或提升。提升方法不依赖于 Fourier 变换，很大程度上提高了离散小波变换的计算性能，后来也成为构造第二代小波的有效方法。提升方法的详细介绍可参考文献(Sweldens, 1996; Daubechies, Sweldens, 1998)。

Daubechies 和 Sweldens (1998)证明任何具有有限长度滤波器的小波变换的滤波器多项式矩阵都可分解成单位上三角  $2 \times 2$  矩阵、单位下三角  $2 \times 2$  矩阵和对角归一化矩阵：

$$P(z) = \underbrace{\begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}}_{\text{归一化因子}} \prod_{i=m}^1 \overbrace{\left\{ \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \right\}}^{\text{主提升}} \quad (1)$$

式中， $s_i(z)$  和  $t_i(z)$  为 Laurent 多项式， $K_1$  和  $K_2$  为归一化因子。因此，对任何由小波高通和低通滤波器系数构成的  $P(z)$  多项式矩阵，总可以通过分解得到  $s_i(z)$  和  $t_i(z)$ 。基于欧几里德算法，Daubechies 和 Sweldens(1998) 指出，任何小波变换都可以通过提升方法实现，并给出了实现的具体步骤。对原始信号  $S$ ，在  $j+1$  水平的整数小波变换的正变换可以表示为：

- 1) 分解(lazy wavelet transform). 将数据序列分解成偶序列  $\lambda_j$  和奇序列  $\gamma_j$ 。

2) 对偶提升(dual lifting). 由  $\lambda_j$  预测  $\gamma_j$  的值，并由预测误差代替  $\gamma_j$ ，形成分解的高通部分。

$$\gamma_j \leftarrow \gamma_j - \{p(\lambda_j)\}$$

其中， $p(\lambda_j)$  由  $t_i(z)$  求得并作规整。

- 3) 主提升(primal lifting). 由  $\gamma_j$  更新  $\lambda_j$  并由更新值代替  $\lambda_j$ ，形成分解的低通部分。

$$\lambda_j \leftarrow \lambda_j + \{u(\gamma_j)\}$$

其中， $u(\gamma_j)$  由  $s_i(z)$  求得并作规整。

因为小波变换涉及到浮点运算，因此通常变换是不可逆的。即使输入信号为整数也因滤波过程是浮点的而得不到整数的输出。提升方法可以很好地执行从整数到整数的变换，而且允许完美重构。提升步骤一般涉及到除法运算，得到的预测值和更新值通常不是整数，因此需要作规整操作，以形成新的整数  $\{p(\lambda_j)\}$ ， $\{u(\gamma_j)\}$ 。当输入数据序列是整数时，经过规整操作，输出数据序列也是整数，这就实现了从整数集到整数集的变换。对于式  $a_{\text{new}} \leftarrow a_{\text{old}} + \{b\}$ ，提升操作并不改变  $b$  值(Calderbank *et al*, 1998)，因此， $a_{\text{old}}$  总能由  $a_{\text{new}}$  在

没有任何畸变的情况下重构, 即  $a_{old} \leftarrow a_{new} - \{b\}$ . 因此, 上述整数小波变换是完全可逆的, 逆变换可以表示为

- 1) 反主提升(inverse primal lifting):  $\lambda_j \leftarrow \lambda_j - \{u(\gamma_j)\}$ .
- 2) 反对偶提升(inverse dual lifting):  $\gamma_j \leftarrow \gamma_j + \{p(\lambda_j)\}$ .
- 3) 合并(inverse lazy wavelet transform):  $\lambda_{j+1} \leftarrow \lambda_j \cup \gamma_j$ .

2 双正交小波和数据压缩

双正交小波变换滤波器系数是对称的, 因此其传递函数具有线性相位特征, 这是变换过程所渴望的. 此外, 它还具有其它一些很有用的特征, 如是紧支撑的、尺度函数  $\varphi(x)$  总是对称的、小波函数总是对称或反对称的. 因此, 双正交小波函数更有利于数据特征的集中, 有效减少数据间冗余. 它的滤波系数是二分的, 如它们具有形式  $z/2^n$ , 这就可以减少执行费用. 我们采用 CDF(Cohen-Daubechies-Feauveau) 双正交小波系列、JEPG2000 采用的 SWE(13, 7) 和 CRF(13, 7) 双正交小波作为我们实验的小波函数(Cohen *et al*, 1992). 其滤波器系数可参考 Uytterhoeven(1999) 文章. 由于小波系数对称性原因(Uytterhoeven, 1999), 我们只选择 CDF(2,  $n$ )、CDF(4,  $n$ ),  $n=2, 4, 6$  的小波系列作研究.

数据的无损压缩过程可以分为 3 步: 第一步, 对小波滤波器系数进行分解, 分解可以得到归一化因子  $K_1$  和  $K_2$ 、Laurent 多项式  $s_i(z)$  和  $t_i(z)$ . 通过分解,  $s_i(z)$  和  $t_i(z)$  总能表示为  $a \times \{b_j, j=0, 1, 2, \cdots\}$ ,  $a$  和  $b_j$  是整数; 第二步, 对数据序列作 3 水平分解, 去除数据冗余. 在提升方法的对偶提升和主提升步骤中我们作规整操作

$$x_i \leftarrow x_i - \left\{ \frac{\sum_j b_j y_j}{a} \right\}$$

(2)

第三步, 对分解系数作熵编码. 我们以压缩比作各种编码方法压缩效果的评判标准. 压缩率可以表示为

$$r = \frac{s_o - s_n}{s_o} \times 100\%$$

式中,  $s_o$  为原始文件大小,  $s_n$  为压缩后文件大小. 我们选取基于字典压缩算法(LZW)、自适应哈弗曼压缩算法(HUFFMAN)、算术压缩算法(ARITH)作为比较对象. 这主要是因为根据熵理论, 上述几种方法压缩熵可以达到最大.

3 实验结果

实验数据取自中国国家数字地震台网中心. 取 5 个台站不同时期地震数据资料, 每段数据在 P 波前 20 分钟至 P 波后 60 分钟之间. 表 1 给出了地震数据资料信息.

表 1 数据资料信息

编号	台站	日 期 年-月-日	发震时间 时:分:秒	$\varphi_N/(^\circ)$	$\lambda_E/(^\circ)$	震级	数据量/字节
1	BNX	2001-11-14	14:22:10	36.04	90.22	8.1	360 000
2	SSE	2002-11-16	08:52:24	29.67	90.59	4.5	432 000
3	BJI	2003-02-24	02:03:44	39.58	77.10	6.5	720 000
4	ENS	2001-02-23	00:09:22	29.55	101.14	6.0	840 000
5	BJI	2003-01-16	11:36:52	29.70	88.21	5.7	1008 000
6	TIA	2002-03-31	06:52:50	24.22	122.22	7.4	1980 000

首先,我们分别采用基于字典压缩算法(LZW)、自适应哈弗曼压缩算法(HUFFMAN)、算术压缩算法(ARITH)和整数小波变换的压缩算法(IWT1,IWT2),对上述数据进行压缩.其中小波函数采用CDF(2,2)双正交小波,对6组数据进行阶次为3的整数小波变换,数据流采用两种不同的编码方法进行编码,实验结果如表2.通过计算每种压缩算法的平均压缩率,我们可以得到:基于整数小波变换的压缩算法,比基于LZW算法提高了68.6%,比基于ARITH算法提高了172.6%,比基于HUFFMAN算法提高了296.9%.而基于整数小波变换的压缩算法中,第一种算法又比第二种算法提高了23.8%.图1给出了各种压缩算法的平均压缩率.

其次,采用CDF双正交小波函数家族、SWE(13,7)双正交小波函数和CRF(13,7)双正交小波函数,对实验数据进行阶次为3的整数小波变换,数据流采用第一种方法编码.CDF双正交小波函数家族取CDF(2,2)、CDF(2,4)、CDF(2,6)、CDF(4,2)、CDF(4,4)和CDF(4,6).表3给出了实验结果.

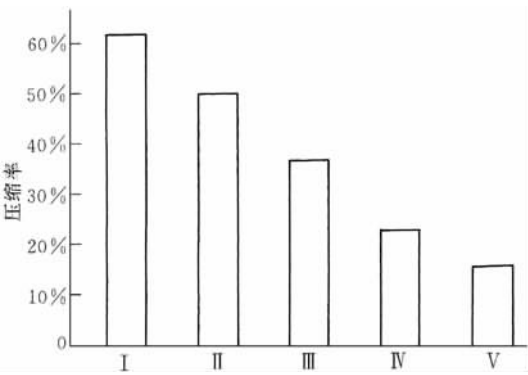


图1 不同压缩编码方法的压缩率  
I. 整数小波变换压缩方法一; II. 整数小波变换压缩方法二; III. 基于字典压缩算法; IV. 算术压缩算法; V. 哈弗曼压缩算法

表2 不同压缩编码方法的压缩结果 字节

编号	编 码				
	IWT1	IWT2	LZW	ARITH	HUFFMAN
1	67 349	102 400	146 767	216 758	274 422
2	197 773	252 922	304 349	388 735	415 225
3	233 153	311 002	371 938	484 990	579 267
4	449 551	565 241	681 385	825 496	867 236
5	358 623	472 416	579 911	484 990	855 966
6	856 649	1110 933	1556 698	1980 933	2064 814

表3 不同双正交小波函数的压缩结果 字节

编号	小 波						SWE	CRF
	CDF(2,2)	CDF(2,4)	CDF(2,6)	CDF(4,2)	CDF(4,4)	CDF(4,6)		
1	67 349	67 358	67 373	70 109	70 161	70 244	62 210	62 203
2	197 773	197 856	197 505	211 241	212 090	212 876	221 809	222 364
3	233 153	233 722	234 224	250 153	249 870	252 562	228 399	228 594
4	449 551	454 819	454 139	471 255	484 288	479 074	513 809	514 162
5	358 623	359 300	359 690	373 924	373 937	376 118	363 860	364 494
6	856 649	852 916	850 761	873 610	869 220	877 065	1155 815	1157 812

图2给出了基于各种小波函数压缩算法的平均压缩率.从图2我们可以看到,基于

CDF(2,  $n$ )系列小波函数压缩算法有更好的压缩效果. 我们将 CDF(2,  $n$ )、CDF(4,  $n$ )的压缩率分别作平均, 与 SWE(13, 7)、CRF(13, 7)的压缩率作比较, 可以得到, 基于 CDF(2,  $n$ )压缩率比基于 CDF(4,  $n$ )、SWE(13, 7)、CRF(13, 7)分别增加了 3.36%、7.68%和 7.79%. 图 3 给出了不同双正交小波函数系列的平均压缩率.

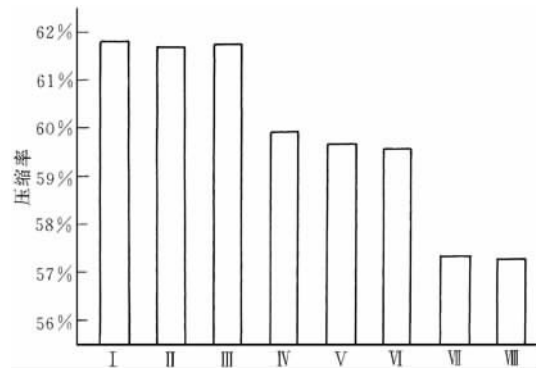


图 2 不同双正交小波函数的压缩率

I. CDF(2, 2); II. CDF(2, 4); III. CDF(2, 6);  
IV. CDF(4, 2); V. CDF(4, 4); VI. CDF(4, 6);  
VII. SWE(13, 7); VIII. CRF(13, 7)

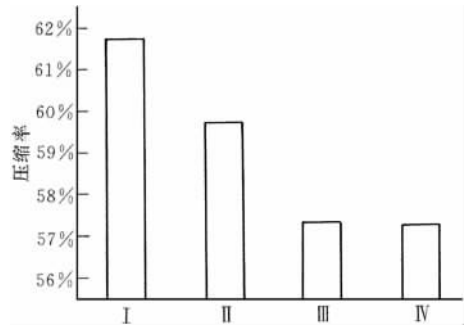


图 3 不同双正交小波函数系列的平均压缩率

I. CDF(2,  $n$ ); II. CDF(4,  $n$ );  
III. SWE(13, 7); IV. CRF(13, 7)

4 结论

本文对地震数据无损压缩方法进行了研究, 提出了应用整数小波变换无损压缩地震数据的方法. 双正交小波变换可以较好地消除数据间的冗余, 且具有线性分解的特点. 作为整数小波变换的小波函数, 与传统的无损压缩算法相比, 可以较好地提高无损压缩比. CDF 双正交小波系列、SWE 双正交小波和 CRF 双正交小波具有不同的特征, CDF 小波系列用于地震数据压缩, 比其它两种双正交小波有更好的压缩效果, 而 SWE 双正交小波和 CRF 双正交小波具有几乎相同的压缩比. 这可能与二者具有相似的特征有关. 对 CDF 双正交小波系列而言, CDF(2,  $n$ )系列又比其它几种双正交小波系列的压缩效果好. 由于双正交小波的滤波系数的二分性, 运算复杂度为  $O(n \lg n)$ , 而小波变换的运算复杂度为  $O(n^3)$ , 这就可以减少执行费用, 提高运算效率.

参 考 文 献

陈运泰, 吴忠良, 王培德, 等. 2000. 数字地震学[M]. 北京: 地震出版社, 35~38

李沙白, 刘启元, 沈立人. 1994. 宽频带大动态地震压缩数据的一种实时算法[J]. 地震学报, 16(1): 111~117

袁玫, 袁文编著. 1995. 数据压缩技术及其应用[M]. 北京: 电子工业出版社, 9~115

Calderbank A R, Daubechies I, Sweldens W, et al. 1998. Wavelet transforms that map integers to integers[J]. Appl Comput Harmon Anal, 5(3): 332~369

Cohen A, Daubechies I, Feauveau J C. 1992. Biorthogonal bases of compactly supported wavelets[J]. Comm Pure Appl Math, 45: 485~560

Daubechies I, Sweldens W. 1998. Factoring wavelet transforms into lifting steps[J]. J Fourier Anal Appl, 4(3): 245~267

Mandyam G, Magotra N, McCoy W. 1996. Lossless seismic data compression using adaptive linear prediction[J]. IEEE

- Int Geosci Remote Sensing Symposium*, **2**: 1 029~1 031
- Memon N D, Sayood K, Nagliras S S. 1994. Lossless compression of multispectral image data[J]. *IEEE Trans Geosci Remote Sensing*, **32**(2): 282~289
- Sweldens W. 1995. The lifting scheme: A new philosophy in biorthogonal wavelet constructions[J]. *Proc SPIE*, **2 569**: 68~79
- Sweldens W. 1996. The lifting scheme: A custom-design construction of biorthogonal wavelets[J]. *Appl Comput Harmon Anal*, **3**(2): 186~200
- Uytterhoeven G. 1999. *Wavelets: Software and Applications*[D]: [Dissertation]. K U Leuven, 38~98
- Villasenor J P, Ergas R A, Donoho P L. 1996. Seismic data compression using high-dimensional wavelet transforms[A]. *Data Compression Conf*[C]. Snowbird: UT IEEE Computer Society Press, 396~405
- Wood L C. 1974. Seismic data compression methods[J]. *Geophysics*, **39**(8): 499~525
- Wang Y Z, Wu R S. 2000. Seismic data compression by an adaptive local cosine/sine transform and its effects on migration[J]. *Geophys Prosp*, **48**: 1 009~1 031

## SEISMIC DATA COMPRESSION BASED ON INTEGER WAVELET TRANSFORM

Wang Xizhen Teng Yuntian Gao Mengtan Jing Hui

(Institute of Geophysics, China Earthquake Administration, Beijing 100081, China)

**Abstract:** Due to the particularity of the seismic data, they must be treated by lossless compression algorithm in some cases. In the paper, based on the integer wavelet transform, the lossless compression algorithm is studied. Comparing with the traditional algorithm, it can better improve the compression rate. CDF (2,  $n$ ) biorthogonal wavelet family can lead to better compression ratio than other CDF family, SWE and CRF, which is owe to its capability in canceling data redundancies and focusing data characteristics. CDF (2,  $n$ ) family is suitable as the wavelet function of the lossless compression seismic data.

**Key words:** lossless compression; integer wavelet transform; lifting scheme; biorthogonal wavelet