石油地球物理勘探

符号位快相关速度谱

物探局计算中心站

摘要 符号位技术是压缩数据的方法之一。它可以扩大计算机的处理能力，而且算法简单，计算速度快，这是目前值得重视的一个发展方向。

地震勘探数据处理要求从野外资料中提取更多的有用信息，因而需要更大的机器内存和更高的运算速度。例如地震勘探从研究构造形态发展到研究岩石物理性质，其主要标志之一，就是高精度地确定层速度，以及在二维、三维空间综合层速度剖面。这就要求在每一个共深度点至少在相当密的点作速度分析，这样，工作量相当可观。在目前条件下，实现起来是很困难的。因此发展一种数据压缩方法，既节省内存，又加快速度，显得非常必要而迫切。

数据压缩的问题带有普遍性，通信工程，电视传输等许多其它学科早已有所研究和应用。近年来，地震处理开始引用这些成果，在理论和应用方面取得了新的进展。这里介绍的符号位快相关速度谱，就是其中之一。

数据压缩研究的每一进展，将为提高地震处理质量，缩短处理周期，实现高度自动化——自动速度分析、自动迭加、自动斜率分析、自动静校正，以及自动滤波参数选择等处理手段，提供有利条件。

我们用八位二进制数，记波峰、波谷的幅值，用 1 毫秒的采样间隔，把每道地震记录压缩成如图 1 的形式。

每一个二进制位和一个采样时间相对应，0 代表该记录时间振幅不是极值。当取得极值后，在对应时间那一位写 1，后面八位记振幅值。这八位仍连续代表采样时间，这

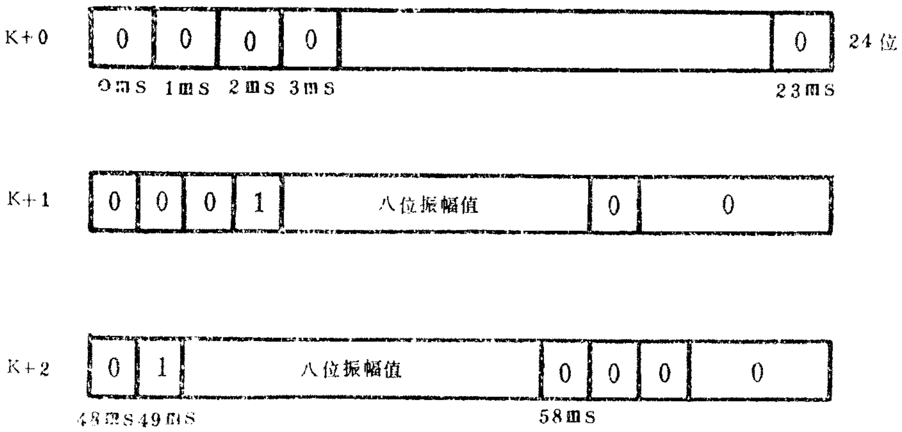

图 1

种表示法的优点是，当记录最大时间不变时，所占单元是等长度的，用这个方案作了原记录和压缩—恢复记录的对比，还作了速度谱，水平迭加等剖面的对比。结果证明，面貌极其相似，这一方案的好处是压缩了内存，记录可以保真，但没有能引起计算上的改革，从而不能大幅度提高运算速度。符号位技术，则不仅压缩比最大，而且引起了计算上的完全革新，它可以用逻辑运算来代替浮点数运算，从一个一个数的运算变为整体运算，大大加快了计算速度，但它的应用范围是有限的，所以并不能代替其它方法。

所谓符号位，国外文献是这样定义的：令 A_{ij} 代表第 i 道上观测时窗中的第 j 个采样值，其符号位的定义是：

$$\operatorname{sgn} A_{ij} \equiv \begin{cases} \dfrac{A_{ij}}{|A_{ij}|} & A_{ij} \neq 0 \\ 0 & A_{ij} = 0 \end{cases}$$

我们这里略加改变，为了区别起见，用 $\bar{A}_{ij}$ 代表 A_{ij} 的符号位，其定义是：

$$\bar{A}_{ij} \equiv \begin{cases} 1 & A_{ij} \geqslant 0 \\ -1 & A_{ij} < 0 \end{cases}$$

前者三个状态：+1、-1、0。后者两个状态，因此更简化了。这样存储一个采样值，只需二进数的一个位。为了存取方便，我们程序中是这样取的，对第 i 道记录，压缩形式如图 2 示。

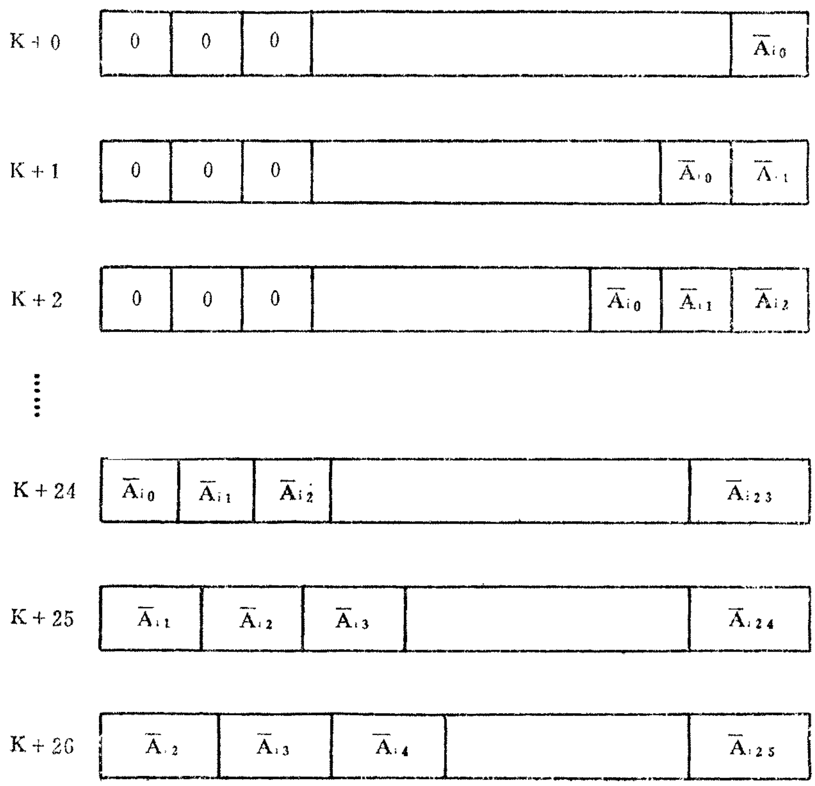

图 2

每个单元都有一部分采样值是重复存放的，不难看出，这样的好处是，每次取一个单元就把这道的观测时窗中的所有值都取来了。

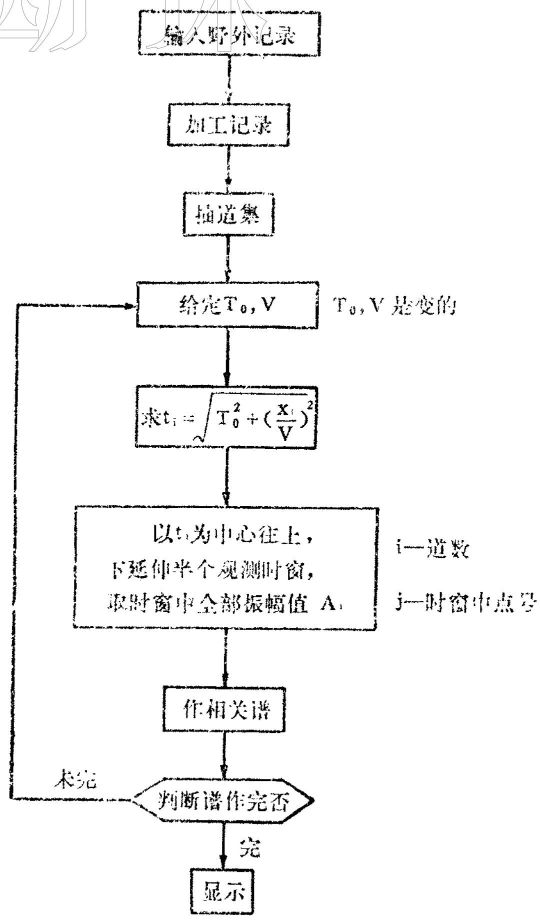

图 3

图 3 是作一个共深度点速度谱的程序粗框图。符号位速度谱和普遍速度谱有什么不同呢？从物理学角度来说，它把共深度点道集改造成为等幅矩形波（图 4）。分析一下速度谱的实质就可以知道，它只用到记录的时间和相位特征，无需严格地知道振幅、频率及波形特征。符号位恰恰保持了记录的时间和相位特征，因此作这么一番改造，并不影响分析速度。事实证明，符号位速度谱速度曲线容易连续追踪，弱层突出，丢层和假层的现象不严重，和以前资料对比，速度分析结果无差异。从程序上看，符号位速度谱和普通速度谱总框图是大同小异的。但从存数方式和计算方法上则根本不同。

我们知道相关公式可以有两种形式：

$$r(T_0, v)=\sum_i \sum_k \sum_j A_{ij} A_{kj} \qquad 其中\quad i \neq k \tag{1}$$

$$i=1\cdots\cdots n,\quad k=1\cdots\cdots n,\quad j=1\cdots\cdots m$$

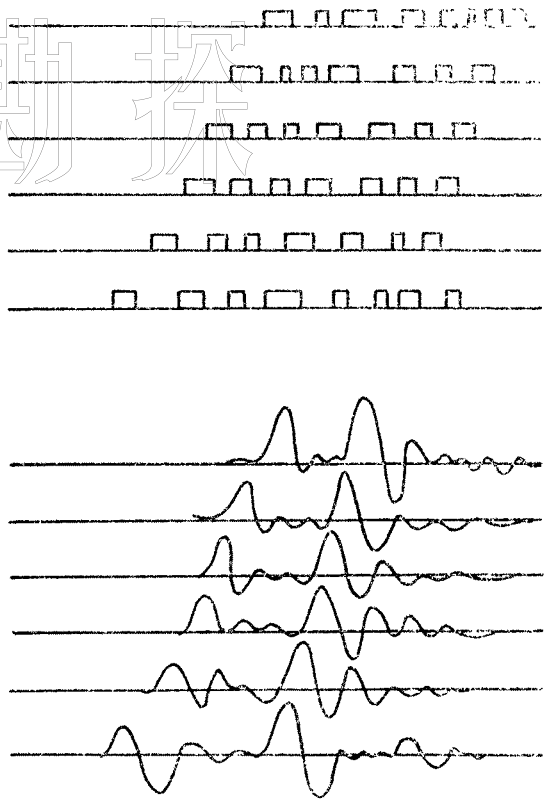

图 4

m 是观测时窗的长度（点数），n 是道集道数，等于复盖次数另一种 形式：

$$r(T_0, v) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} A_{ij} \right)^2 - \sum_{j}^{m} \sum_{i}^{n} A_{ij}^2 \tag{2}$$

这两个式子是完全等价的。前者要作道与道之间互相关，它还有 其它 用途，例 如自 动静校正，求道间随机时移。现在我们分别探讨上述两个式子用符号位技术怎样实现快速算法。用符号位定义，（1）式 改写 为：

$$r(T_0, v) = \sum_{i}^{n} \sum_{k}^{i \neq k} \sum_{j}^{m} \bar{A}_{ij} \bar{A}_{kj} \tag{3}$$

这个式子的直观几何意义就是用时距曲线公式

$$t_i = \sqrt{T_0^2 + \left(\frac{X_i}{V} \right)^2}$$

校正后的记录，如图 5 所示。对任意两道作互相关计算，公式

$$\sum_{j}^{m} \bar{A}_{ij} \bar{A}_{kj}$$

表示两个道的互相关。公式（3）前面两个和号就是要对任意不相重的所有可能的道间作互相关。根据 A_{ij} 的定义，只有 +1 和 −1，因此 $A_{ij}\cdot A_{kj}$ 有四种可能，如表所示。

表

$\overline{A}_{ij}$	$\overline{A}_{kj}$	$\overline{A}_{ij}\ \overline{A}_{kj}$	
1	1	1	同　相
1	−1	−1	不同相
−1	1	−1	不同相
−1	−1	1	同　相

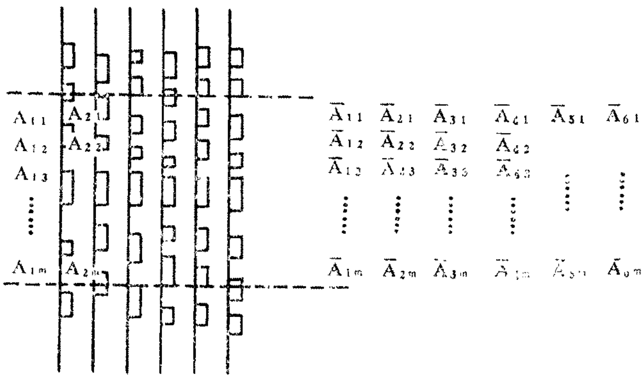

图　5

如果把这张表中的 +1 全换成 0，马上可以看出这张表就是150机指令中的按位加的真值表。

K+0	$\overline{A}_{i1}$	$\overline{A}_{i2}$	$\overline{A}_{i3}$	……	$\overline{A}_{im}$	0	…	0
K+1	$\overline{A}_{k1}$	$\overline{A}_{k2}$	$\overline{A}_{k3}$	……	$\overline{A}_{km}$	0	…	0

当我们从压存区取数，一下子就将时窗的全部采样值放好在一个单元中，进行一次按位加运算，m 个对位乘的结果已在累加器中了。即在 L 中存放的结果是

$\overline{A}_{i1}\ \overline{A}_{k1}$	$\overline{A}_{i2}$	$\overline{A}_{k2}$		$\overline{A}_{im}\ \overline{A}_{km}$	0	…	0

从上述真值表可以看出，$\overline{A}_{ij}\cdot\overline{A}_{kj}$ 非 0 即 1，0代表记录同相得数是 +1，1 代表不同相，得数是 −1，因此求 $\Sigma_j\overline{A}_{ij}\ \overline{A}_{kj}$ 可分为两部分。即

$$\sum_{j}^{m}\overline{A}_{ij}\cdot\overline{A}_{kj}=c_0-c_1=(m-c_1)-c_1=m-2c_1 \tag{4}$$

C_0 代表所有 +1 的和，即 O 的个数，

C_1 代表所有 −1 的和，即1的个数，它也是一个正整数，从（4）式可以看出求和的

问题就变成了数1的个数问题。总之求 $\sum_{j}^{m}\bar{A}_{ij}A_{k,j}$ 只需两步：按位加，数1的个数。这样计算就快多了。同样办法分析②式：

$$r(T_0,v)=\sum_{j=1}^{m}\left(\sum_{j=1}^{n}\bar{A}_{ij}\right)^2-\sum_{j}^{m}\sum_{i}^{n}\bar{A}_{ij}^2$$

结论是：①

$$\sum_{j}^{m}\sum_{i}^{n}\bar{A}_{ij}^2\equiv m\cdot n$$

②求 $\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\bar{A}_{ij}\right)^2$ 可以归结为数“1”个数和查表两步。

结论是：

需要特别一提的是取符号位如何清初至，因为初至区清0后取符号位永远是+1，如果0代表+1，则全部取到0，这样作相关，初至区的相关值均达到最大，为此给出一个判断，当按 t_i 去取时窗中所有值时，再往时窗前多取若干值，例如多取21个点，如图6示。

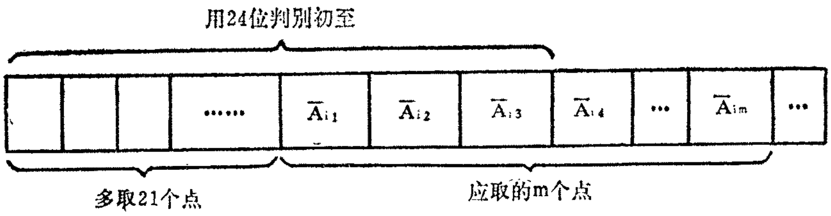

图 6

用24位判别是否全0，是，则表明落在初至区，该道不参加相关，其余道相关后，应按下式加数

$$\text{加权后相关值}=\frac{\text{最大相关次数}}{\text{实际相关次数}}\times\text{实际相关值}$$

以6次复盖只作5道相关为例，最大相关次数为 $\frac{6\times5}{2}=15$，实际相关次数 $\frac{5\times4}{2}=10$，这样处理以后，浅层资料有所改善。

符号位技术是压缩数据的方法之一，其特点是压缩比最大，算法简单，实现相关速度快，但应用范围有限。

数据压缩是扩大机器空间的有效手段，是提高计算速度，实现过去难以实现的处理方法和处理过程自动化的重要前提条件，是值得重视的发展方向。如果这种发展再和硬件的改造紧密配合起来，将有更好的效果，例如上述方法都要求数“1”的个数，现在用150机实现还要用移位循环来实现，如能有一条硬指令实现，此方法的速度还可大大加快。