KL 变换在位场数据处理中的应用

夏克文

(西安石油勘探仪器总厂, 710061)

潘作枢

(西安地质学院,710054)

摘 要 应用 KL 变换方法在位场数据处理中消除信号频段上的干扰噪声,确定出信号和随机噪声所对应的特征范围,提取信号特征并重构数据,使随机噪声的影响降到最低限度,从而提高信噪比。
 主题词 重力勘探,磁法勘探/KL变换,位场数据处理,随机噪声

分类号 P631.1; P631.2

0 前 言

KL(Karhuen-Loeve)变换也即主分量分析,对信号波形数据进行波形特征提取并 重构波形,因此它又称特征向量变换或霍特林(Hotelling)变换,其是一种正交变换。用 KL变换对数据进行分析,已在图象分析处理、数字通讯、医用及地震资料处理中得到广泛 应用。如在地震资料处理中它被用于改善信噪比(C.H.埃蒙等,1978;付才芳,1986), 从波形中移去不相关噪声(Jones, 1987),从反射中分离绕射(Yedlin etal, 1987);在 垂直地震剖面(VSP)法中,用KL变换从上行波场中分离下行波场;在声波全波列测井 中,用KL变换压制随机噪声和多次反射一转换波,提取组分波信息等^[1]。本文论述这种 方法在位场数据处理中剔除噪声提高信噪比的理论和方法,并给出对人工数据和实际数据进 行处理的实例。

1 方法原理

1.1 KL 变换数学原理

KL变换是与 Fourier 变换、Walsh 变 换等相似的变 换, 用 矩 阵 表 示为: Y = TX, 式中 X 是要变换的矢量, T 是变换矩阵, Y 是变换后的矢量。一般 T与X 无关, 如Fourier 变换中, T 的元素是指数型的, Walsh 变换中 T 的元素是 + 1 或 - 1, 并且元 素是按已知 的递推法则来确定的。但 KL 变换中, 矩阵 T 是要由变换的矢量 X 推算出来的, 其元 素由 X 的协方差矩阵的特征矢量构成。KL 变换有两个最大特点, 一是唯一产生不相关系数的变

第一作者简介: 夏克文, 男, 1965年出生, 硕士, 工程师, 应用地球物理专业。

¹⁹⁹²年11月11日收稿。

换,二是它的均方误差最小。因此这是一种最佳变换,故其有广泛的应用价值。

设 X 是一个 N 维实随机向量,记 X = $(x_1, x_2, ..., x_N)^T$,其 均 值 向 量 为 mx = E(x) = $(E(x_1), E(x_2), ..., E(x_N))^T$,X的协方差矩阵记为 CX = E((x - mx))(x - mx)^T], CX 为实对称矩阵,将其特征值按其递减顺序排列为 $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_N \ge 0$, 与特征值对应的特征向量 a_i 满足关系:

 $CXa_i = \lambda_i a_i (i = 1, 2, ..., N)$ (1) 由于协方差矩阵的实对称性,故存在完备特征向量组 $\{a_i\}$ (*i* = 1, 2, ..., N)满足正交归 一化条件: $a_i a_i = \{ \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$ (*i*, *j* = 1, 2, ..., N),构成N维空间的一个完备正交系。因 而对 N 维空间中的任一向量 X 均可展开为: $X = \sum_{i=1}^{N} Y_i a_i$,其中展开系数 Y_i 可求得为:

$$Y_{i} = a_{i}^{T} X (i = 1, 2, ..., N)$$
(2)

用特征向量组 $\{a_i\}$ 组成正交变换矩阵: $A = [a_1, a_2, ..., a_N]$,则(1)、(2)两式可 分别记为:

$$CXA = AR \tag{3}$$

$$Y = A^{\mathsf{T}}X \tag{4}$$

其中 $R = diag [\lambda_1, \lambda_2, ..., \lambda_N]$ 为一按特征值递减顺序排列的对角矩阵, Y为由(2)式 求得的展开系数矢量:

$$Y = (Y_1, Y_2, \dots, Y_N)^{\mathrm{T}}$$

$$(5)$$

称 ^Y 为 X 的 KL 变换,或 Y, (*i* = 1, 2, …, N) 为 X 的 主分 量,由于 A 为正交归一化 矩阵,由(4)式可求得 X 的准确表达式:

$$X = AY \tag{6}$$

Y的协方差矩阵为:

 $CY = E\{ (Y - E(Y)) (Y - E(Y))^{\mathsf{T}} \} = E\{ A^{\mathsf{T}} (X - E(x)) (X - E(X))^{\mathsf{T}} \} = A^{\mathsf{T}}CXA$

由(4)式得:

 $CY = R = diag(\lambda_1, \lambda_2, ..., \lambda_N)$,这说明各主分量 Y_i 间两两正交、互不相关,且 第 i 个主分量的方差就等于 A 的第 i 个特征值,即 $\sigma^2 Y_i = E\{ \{ Y_i \sim E(Y_i) \} \} = \lambda_i (i = 1, 2, ..., N)$ 。

考虑均值 mx = E(x) = 0的随 机矢 量 X,此时 CX 的迹 $Tr(CX) = \sum_{i=1}^{N} X^{2}$ 代 表向 量 X 的总能量,由于 $E(Y) = A^{T}E(X) = 0$,所以方差 $\sigma^{2}Y_{i}E[Y_{i}^{2}] = \lambda_{i}$ 代表了第 i 个 主分量包含的能量。根据正交变换性质有:

$$T_{i}(CX) = T_{r}(CY) = \sum_{i=1}^{N} X_{i} = \sum_{i=1}^{N} \lambda_{i}$$
 (7)

这说明经 KL 变换后总能量不变,且等于 CX 的特征值之和。若用前 M 个特征值对应的特征向量组成的变换矩阵 A 来代替 A, 即令:

$$A = [a_1, a_2, \cdots, a_M] \qquad M < N$$
(8)

则有: $\widetilde{Y} = \{Y_1, Y_2, ..., Y_M\}^T$, \widetilde{Y} 可由 X 表示为:

$$\widetilde{Y} = A^{\mathsf{T}} \widetilde{X} \tag{9}$$

由(9)式来重构 X, 可得到:

$$\widetilde{X} = \sum_{i=1}^{M} Y_i a_i = \widetilde{A} \widetilde{Y}$$
(10)

该值与原来的值有 X 差异,其均方误差为:

$$\varepsilon(M) = E\{ \| X - \widetilde{X} \|^2 \} = E\{ \| \sum_{i=M+1}^{N} Y_i a_i \|^2 \} = E\{ \sum_{i=M+1}^{N} X_i^2 \} = \sum_{i=M+1}^{N} \lambda_i$$
(11)

由于 λ_i 是单调递减的,而所选用的 M 个特 征向量与 前 M 个最 大 特 征 值 相 对 应,则由 (10)式求得的重构向 量 \tilde{X} 与实际向量 \tilde{X} 的误差最小。故从最小均方误差意义上讲,KL 变换是最佳变换。

1.2 位场数据处理方法

在位场测量中所观测到的数据一般说来是信号和噪声两部分的叠加,而噪声是由各种干 扰和误差所造成的。为了将噪声压低到最低限度,可以在整个测区范围采用KL 变换方法进 行数据处理提高信噪比。

假定实测位场数据可以构成一个 $N \times P$ 阶矩阵 X,例如,数据由 P 条剖面,每 条剖面 有 N 个测点组成;或者是由 N 个测点,每个测点有 P 个频率或时间取样道的数据组成。若 将每道液形数据看成为一个离散随机信号,则 X 可表示为一个 N 维随机向量: $X = (X_1, X_2, ..., X_N)^T$,其中 $X_i = (X_{i1}, X_{i2}, ..., X_i)$,若将 X 分为主要 信号 X_i 和非主要 信号 X_i ,即令 $X = X_i + X_i$,利用 KL 变换,可将 X_i 与 X_i 分离开。其出发点是基于如下 原理:主要信号通常具有较大的能量,从信号的某个特征出发,它们在各道之间具有较强的 相干性。KL 变换后,这些相干性能量主要集中在与 X 的协方差矩 阵 CX 的前 n 个最 大特 征值相对应的主分量上,而那些非主要信号具有的能量较小,各道间的相干性也很弱,它们 与那些具有较小特征值的主分量相对应。由于各主分量之间互不相关,故可通过与 CX 较大 特征值相对应的主分量来构成主要信号,而用与较小特征 值 对应 的主分量来构成非主要信 号,从而达到信嗓分离的目的。若主要 信号 X_i 由前 M 个主分量构成,则由(4)式可得 经 KL 变换后的主分量提取为:

 $\begin{array}{c}
 a_{11}, a_{21}, \dots, a_{N1} \\
 a_{12}, a_{22}, \dots, a_{N2} \\
 \dots \dots \dots \dots \dots \\
 a_{1N}, a_{2N}, \dots, a_{NN}
\end{array}
\begin{pmatrix}
 X_{11}, X_{12}, \dots, X_{1P} \\
 X_{21}, X_{22}, \dots, X_{2P} \\
 \dots \dots & X_{N1}, X_{N2}, \dots, X_{NP}
\end{array}
=
\begin{pmatrix}
 Y_{11}, Y_{12}, \dots, Y_{1P} \\
 Y_{21}, Y_{22}, \dots, Y_{2P} \\
 Y_{N1}, Y_{N2}, \dots, Y_{NP}
\end{pmatrix}
\begin{pmatrix}
 \textbf{第二主分量} \\
 \textbf{第N主分量} \\
 \textbf{第N主分量}
\end{array}$ $\begin{array}{c}
 \textbf{12} \\
 \textbf{#取为}
\end{array}$

$$X_{s} = \begin{array}{c} a_{11}, a_{12}, \cdots, a_{1M} \\ a_{21}, a_{22}, \cdots, a_{2M} \\ \cdots & \cdots & \cdots \\ a_{N1}, a_{N2}, \cdots, a_{NM} \end{array} \begin{pmatrix} Y_{11}, Y_{12}, \cdots, Y_{1P} \\ Y_{21}, Y_{22}, \cdots, Y_{2P} \\ \cdots & \cdots & \cdots \\ Y_{M1}, Y_{M2}, \cdots, Y_{MP} \end{array}$$
(13)

这样,重构的信号即去掉了 X₁。

对零均值的波形信号,重构能量占原能量的百分数 $\eta(M)$ 由(7)式得:

$$\eta(M) = \left(\sum_{i=1}^{M} \lambda_{i} / \sum_{j=1}^{N} \lambda_{j}\right) \times 100\%$$
(14)

在实际应用中, *M* 的具体选取原则有二: 一是在理论上, 当信号的百分含量比 $\eta(M)$ 最接近于去掉噪声时百分含量比 $1 - \eta_a(\eta_a 为噪声百分 数)$, 即 $\eta(M-i) < 1 - \eta_a(i = 1, 2, ..., M-1)$ 且 $\eta(M) \ge 1 - \eta_0$ 时的 *M* 即为所求; 二是根据 $\eta(M)$ 的曲线形态判定。从下述人工资料处理图 3 也可知道, 无噪声曲线随 *M* 递增迅速使 $\eta(M)$ 等于 1; 有噪声时曲线随 *M* 递增不是迅速使 $\eta(M)$ 为 1 而是缓缓地趋于 1, 曲线接近于一条水平延伸直线。此水平直线是为随机噪声所致,故选取 $\eta(M)$ 曲线 变为直线时起始位置 *M* 即为所求。

2 算 例

2.1 人工资料处理

这里以重力数据处理为例进行讨论。首先用若干理论模型通过计算叠加成面积上的理论 重力数据,并在这个理论数据中加入满足高斯分布等的随机噪声,构成带有噪声的数据,然 后对这样的数据进行 KL 变换,再提取特征和重构数据。

图 1 是由若干个球体在地面产生的重力异常等值线图,其中剖面测线为东西方向,每条 剖面上取采样点 N = 40, 剖面测线数 P = 40, 异常的极大值为150 mgal。在重力测量中,干 扰因素较多,尽管在接收系统中已将噪声压到最小限度,但是在测量记录上随机噪声仍明显 可见。为模拟实际噪声特征,本文用 Monte-Carlo 法在计算机上计 算出服从高斯分布的随 机序列,然后由此构造出具有实际意义的离 散噪声序列,将 由 此得到 $N \times P$ 个均值近似为 零,均方差为 9 mgal 的随机噪声(相当于异常极大值的 6 %,即 $\eta_n = 6 %$)叠 加在理论数 据上。

图 2 是用加入噪声后的数据绘成的平面图,从图上可见等值线变得比较零乱。对图 2 带 有噪声的数据进行 KL 变换,求出主分量所对应的特征值,再由(14)式绘出 $\eta(M)$ 曲线 于图 3 上。由图 3 可知,不带噪声的纯信号 $\eta(M)$ 曲线在M = 4时就迅速趋于 1;而带噪 的 $\eta(M)$ 曲线,第一分量特征约占70%的总能量,第一第二特征共占86%,而前三个特征 所占能量为95%,即 $\eta(1) = 70\%$, $\eta(2) = 86\%$, $\eta(3) = 95\% \ge 1 - \eta_a$;此外,根 据带噪曲线的形状可知,M从4开始变成一条水平延伸的直线。因此由M的选取原则,提 取前三个特征,应用公式(11)重构数据,得出剔除噪声到最低限度的平面等值线(图 4),对 比图 1 可知,异常形态复原程度较高。

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



图1.由若干个球体所构成的重力异常图 (等值线间隔为8 magl)

Fig.1 The gravity anomaly of some spheroids



图 3 η(M)—M 关系曲线图 Fig.3 η(M)—M curve 1-育噪声曲线; 2-无噪声曲线



Fig.2 Noise data map



图4 重构数据的平面图 Fig.4 Reconstitution data map

2.2 实际资料处理

选取某地区1:20万航磁测量结果经过频率域化极后的平面等值线①(图5)。由图可 看出其异常形态比较复杂,高频干扰比较严重,且呈南北向拉长形态,其主要原因是化极场 误差的影响,包括处理时人为取数所带来的误差和测量结果预处理时给位场值产生的一定误 差等随机噪声。为了剔除这些噪声,采用 KL 变换方法,将图5以50×40的矩形网格划分, 即取南北向剖面数40条,每条剖面上取50个测点数据,构成一个50×40数据矩阵。按 KL 变 换求出数据协方差矩阵的特征值,绘制 η(*M*)曲线,然后根据*M*的选取原则二提取第一第 二特征即可,再重构数据则得到剔除约为12%的噪声后的平面等值线(图6)。图6的等值 线形态简单,对称性较好,结果令人满意。

①刘建军, 磁测数据处理---化极问题的研究, 西安地质学院硕士论文, 1992



图 5 某地区航磁资料经化极所得的平面图 (据浏建军, 1992, 等值线间隔为4nf) Fig.5 Aeromagnetic surveying data map in some place



图 6 别除随机噪声后得到的平面等值线图 Fig.6 Contour diagram after eliminating the random noise

3 结 束 语

(1)本文提出了 KL 变换方法用于位场数据处理。KL 变换方法在数据处理中的基本 步骤是对数据进行 KL 变换,然后提取信号特征和重构数据。

(2)划分信号和噪声,提取信号特征的两个原则,能够使重构数据将随机噪声剔除到 最低限度。

(3)通过人工资料和实际资料的处理结果,表明应用 KL 变换方法能成功地从位场数据中剔除噪声,增进信噪比。

(4)在测区范围大、要求处理的剖面条数很多,且每条剖面上的测点数目又大的情况下,构成的数据矩阵很大,考虑到计算机的容量问题,可进行分块处理达到剔除噪声的目的。

参考文献

- 1 K Hsu. Wave separation and feature extraction of acoustic well-logging waveforms using Karhunen-Loeve transformation. Geophysics, 1990, vol.55, p176-184
- 2 郑治真. 数字信号处理基础. 北京: 地震出版社, 1988
- 3 程民德等.图象识别导论.上海:上海科技出版社, 1983
- 4 张宏基. 数字通信数学基础。北京: 人民邮电出版社, 1982

198

APPLICATION OF KL TRANSFORMATION TO POTENTIAL FIELD DATA PROCESSING

Xia Kewen

(Logging Institute, Xi'an Petroleum Exploration Instrument Complex) Pan Zhoushu (Xi'an College of Geology)

Abstract

This paper expounds the mathematical theory of KL transformation method, and presents a new approach to eliminate the random noise which has the same frequency spectrum range as the signal in potential field data processing by using the KL transformation. This approach is based on the KL transformation of potential field data into characteristic vectors. The distribution of characterise indicates which characterise are related to the signal and which to the random noise. By extracted the characterise associated with the signal, the data may be reconstructed and the random noise eliminated. The test of the method on gravity and magnetic data shows that the approach can reduce the random noise to the inferior limit, thus the signal-noise radio fully increased.

Key words gravity prospecting; magnetic prospecting/KL transformation, potential field data processing; random noise

			-	ሻ	期	要	8			
1.	城市地质3	环境的工程	地质分析	折与图	系编绘			• • • • • • • • • • • •	吴	恒
2.	柳州市区。	人工地基的	测试与记	泙价…	•••••	· • · • · • • • • • •	••••••		关	键超
3.	沉管灌注	脏施工对邻	近建筑	物的影	响预测	•••••	••••••		周	东等
4.	三轴压缩	下大理岩应	力一应到	变全过	程的力	学特征	••••••		xß	小亮
5.	柳州市红料	粘土的变形	和强度	時征…	• • • • • • • • • • •			• • • • • • • • • • • •	陈	先华
6.	分形几何有	在工程地质	中的应用	刊		••••••	••••••	• • • • • • • • • • • • •	袁	宝远
7.	中国城市	地质环境及	环境地	贡问题	į	••••••	••••••	• • • • • • • • • • • • •	杨	忠耀
8.	临淄地区	同位素水文	地质研究	究	• • • • • • • • • • •	••••••		• • • • • • • • • • • • • • • • • • • •	陈	余道
9.	用直线高速	差法测定水	准仪望远	远镜调	焦透镜	运行差		• • • • • • • • • • • •	陈	继红
10.	一种双点	后方交会的	坐标及,	点位中	误差计	算方法	••••••	• • • • • • • • • • •	汤	善忠
						300630063006	30063006300	130003000000000	a a63c a63c a63c a63	