

基于动态线性回归的一阶导数分段法 在螺纹检测数据处理中的应用

周强 张慧慧

(北京工业大学机械工程与应用电子技术学院, 北京 100022)

摘要 提出了一种基于动态线性回归的一阶导数分段方法, 并成功应用于螺纹检测数据处理中. 实践证明这是一种有效的对测量数据进行分类的方法.

关键词 线性回归, 一阶导数, 管接箍内螺纹测量

分类号 TB115

测量数据按一定标准分类是数据处理的一个基本且重要的环节, 它直接关系到各项参数的计算结果和精度. 通常的方法有基于(快速傅立叶变换)FFT的抽取特征量的方法; 优化方法以及各种相关理论和谱分析算法. 这些算法都比较抽象、复杂, 并且算法的实时快速性不易控制. 基于动态线性回归的一阶导数分段法思路清楚、简洁、算法程序编写简单, 很适于对具有周期性并且具有线性规律的测量数据分类.

1 算法的基本原理和步骤

用数据的一阶导数寻找极值点, 对数据进行分段, 在实际运用中比较普遍, 是寻找数据极值点的基本方法. 此方法是先对数据 X 求取导函数 $\delta(X)$; 然后判断 $\delta(X)$ 的正负号, 当 $\delta(X)$ 由正变负时数据出现极大值; $\delta(X)$ 由负变正时数据出现极小值, 进而根据 $\delta(X)$ 的正负特性对数据进行分类. 在对数据分段方法的基础上再对每段数据 X_m (其中 n 为第 i 段数据个数) 按确定点数 N 依次拟合求取标准差

$$\delta(X_i)_j, j = 1, 2, \dots, n - N$$

选取标准差最小的那部分数据作为本段数据的测量计算用数据.

$$\delta(X_i)_m = \min(\delta(X_i)_j), j = 1, 2, \dots, n - N$$
$$\bar{X}_i = X_{i,m}, X_{i,m+1}, \dots, X_{i,m+N}$$

把这种方法定义为“基于动态线性回归的一阶导数分段法”. 图1为算法的程序框图.

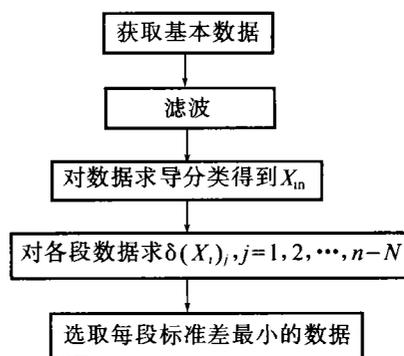


图1 算法的程序框图

2 算法在测量设备中的应用

以油管接箍的内螺纹检测为例. 油管接箍内螺纹检测方法有2种. 一种是使用各种量规现场检测; 另一种是实验室检测, 选取个别单件产品, 用笔式记录仪记录螺纹齿型, 然后手工测绘进行数据处理. 2种方法都存在劳动强度大, 测量结果受人为主观因素影响等缺点.

管接箍计算机辅助测量装置是测量油管接箍内螺纹的专用设备, 主要用于在线或离线(实验室)测试

各种符合美国石油学会(American Petroleum Institute)标准的油管接箍内螺纹^[1]。该设备主要特点是：①结构紧凑、体积小、重量轻；②在线测量、记录管接内箍螺纹的各项参数，实时、准确反映产品质量趋势，为现场质量控制提供重要的参数依据；③数据处理快速、准确、精度高。连接测量装置的数据处理过程中建立了符合实际要求的本质属性；④采用计算机辅助测量，测量效率高、准确；⑤设备通用性强，本套设备主要测量型号为UPTBG 2 7/8 EU的油管接箍内螺纹，测试其它型号只需调整标准数据库参数，不需对硬件进行改动。

2.1 工作原理

管接箍内计算机辅助测量装置的工作原理是以螺纹的某一截面齿型为主要研究对象，获取这一齿型截面的数据是该测量装置的核心任务。测量装置采用接触式测量策略，计算机控制步进电机带动测针沿管接箍轴向运动并记录这一运动轨迹 X ，同时测针带动位移传感器在内螺纹齿面滑动，并由计算机通过对位移传感器的采样来记录这一位移变化量 Y 。测量完一定长度内的螺纹齿型后测针缩回，步进电机带动测针回位。由数据 X, Y 不难和成内螺纹某一截面齿型，然后对数具进行合理性判断、滤波、分段，找到各齿面的直线段部分，进一步求出各直线段部分的交点并计算出管接箍内螺纹的基本参数，再计算导出参数，最后进误差补偿把测量数据存入数据库，一个测量周期结束。

2.2 数学模型的建立

通过对管接内螺纹(2 7/8EU)进行理论分析和数学推导，可以得出其数学表达式。

$$\begin{aligned} \text{当 } 0 \leq \delta x \leq L_4 & \quad y = y_2 - \sqrt{r_2^2 - (x - x_2)^2} \\ \text{当 } L_4 \leq \delta x \leq (L_4 + L_3) \text{ 时} & \quad y = (x - x_0) \times \cot \alpha_2 + y_0 \\ \text{当 } (L_4 + L_5) \leq \delta x \leq p_2 \text{ 时} & \quad y = y_4 + \sqrt{r_4^2 - (x - x_4)^2} \\ \text{当 } p_2 \leq \delta x \leq (p_2 + L_1) \text{ 时} & \quad y = y_3 + \sqrt{r_3^2 - (x - x_3)^2} \\ \text{当 } (p_2 + L_1) \leq \delta x \leq (p - L_3) \text{ 时} & \quad y = -(x - x_3) \times \cot \alpha_1 + y_5 \\ \text{当 } (p - L_3) \leq \delta x \leq p \text{ 时} & \quad y = y_1 - \sqrt{r_1^2 - (x - x_1)^2} \end{aligned}$$

其中： p 为螺距； β 为锥角； α_1 为左半角； α_2 为右半角； s_1 为齿顶截距； s_2 为齿根截距。

$$p_2 = p \times \sin \alpha_2 \times \cos (\alpha_1 - \beta / 2) / [\sin \alpha_2 \times \cos (\alpha_1 - \beta / 2) + \sin \alpha_1 \times \cos (\alpha_2 + \beta / 2)]$$

$$p_1 = p_2 \times \sin \alpha_1 \times \cos (\alpha_2 + \beta / 2) / [\sin \alpha_2 \times \cos (\alpha_1 - \beta / 2)]$$

$$r_1 = s_2 \times \sin \alpha_1 / (1 - \sin \alpha_1) \quad r_2 = s_2 \times \sin \alpha_2 / (1 - \sin \alpha_2)$$

$$r_3 = s_1 \times \sin \alpha_1 / (1 - \sin \alpha_1) \quad r_4 = s_1 \times \sin \alpha_2 / (1 - \sin \alpha_2)$$

$[x/p] = n []$ 表示取整

$$x - n \times p = \delta x$$

$$L_1 = s_1 \sin \alpha_1 \cos \alpha_1 / (1 - \sin \alpha_1) \quad L_3 = s_2 \sin \alpha_1 \cos \alpha_1 / (1 - \sin \alpha_1)$$

$$L_2 = p_1 - L_1 - L_3 \quad L_4 = s_2 \sin \alpha_2 \cos \alpha_2 / (1 - \sin \alpha_2)$$

$$L_5 = p_2 - L_4 - L_6 \quad L_6 = s_1 \sin \alpha_2 \cos \alpha_2 / (1 - \sin \alpha_2)$$

$$\begin{cases} x_0 = n \times p \\ y_0 = n \times p \times \tan \beta / 2 \end{cases} \quad \begin{cases} x_1 = (n + 1) \times p \\ y_1 = (n + 1) \times p \times \tan \beta / 2 + s_2 + r_1 \end{cases}$$

$$\begin{cases} x_2 = n \times p \\ y_2 = n \times p \times \tan \beta / 2 + s_2 + r_2 \end{cases} \quad \begin{cases} x_3 = n \times p \times p_2 \\ y_3 = n \times p \times \tan \beta / 2 + p_2 / \tan \alpha_2 - s_1 - r_3 \end{cases}$$

$$\begin{cases} x_4 = n \times p \times p_2 \\ y_4 = n \times p \times \tan \beta / 2 + p_2 / \tan \alpha_2 - s_1 - r_4 \end{cases} \quad \begin{cases} x_5 = (n + 1) \times p \\ y_5 = (n + 1) \times p \times \tan \beta / 2 \end{cases}$$

$$\begin{cases} x_6 = n \times p + p_2 \\ y_6 = n \times p \times \tan \beta / 2 + p_2 / \tan \alpha_2 \end{cases}$$

从上述公式可看出螺纹齿型函数 $f(x)$ 是 $p, \beta, \alpha_1, \alpha_2, s_1, s_2$ 等 6 个基本参数的函数。通过对螺纹曲线方程的分析可知, 螺纹曲线实际上是一个周期性的三角波函数(周期为 p), 沿水平轴旋转 $\beta / 2$, 然后再削顶、去底而得到的。所以螺纹曲线中的直线段是关键, 它相当于螺纹取线的基波成分。只要能够把直线段数据筛选出来, 也就是说, 通过适当的数字滤波方法找到直线段数据, 再找到各段直线的交点, 计算出螺纹、理论齿高和实际齿高以及齿顶截距和齿根截距, 然后利用线性回归可计算出 2 个半角, 从而得到基本参数, 再通过对标准样件的测定来修订这 6 个基本参数补偿误差, 最后计算出其他导出参数。

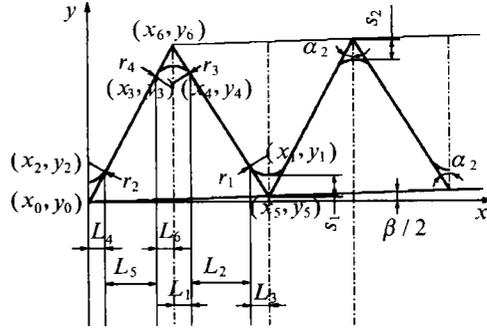


图2 螺纹曲线各项参数

2.3 基本数据的取得和处理

套管接箍内螺纹测量装置对管接箍内螺纹进行在线式接触测量, 得到螺纹某一截面上的齿型数据。采样后的数据写入采样数据文件, 计算程序调用采样文件进行计算得出 6 个基本参数, 然后再计算出导出参数完成整个计算过程^[2]。

采样后的数据必会存在误差, 所以在计算前首先要对数据进行低通滤波, 其方程为

$$y(n) = (1 - k) \times y(n - 1) + k \times x(n)$$

其中, $y(n)$ 及 $y(n - 1)$ 为前一采样是刻及采样时刻的数出, $x(n)$ 为本采样时刻的输入。

在调试过程中, 若由于测试过程中发生意外会产生非法测试数据, 这时就需要计算程序对数据进行合理性检验。采用 FFT 抽取特征量(螺距)对测试数据进行合理性检验。

2.4 基本参数的计算

分段回归分析的关键是采样数据进行分段。这里分段的概念是根据齿型方程以及螺纹齿型曲线, 把属于同一齿面的采样数据划分出来, 然后再分别根据这一段数据挑选出来线性度最好的给定点数拟合出一条曲线, 得到斜率与截距, 进而求出交点计算出 6 个基本参数。

在对螺纹数据进行分段时, 考虑了 2 种方案:

其一是用标准的 FFT 抽取基波频率 f , 最终计算出螺纹的螺距。这种方法在实际处理管接箍螺纹数据时存在一些问题: ① 在进行 FFT 变换时还无法精确确定螺纹数据的峰顶、峰谷的起点和终点, 这就造成由 FFT 的基波频率计算出的螺距 p 存在较大的误差; ② 如果要提高精度, 势必要在定距离内增加采样的点数, 而采样点数是以 $N = 2^m$ 的指数增长的。大量的采样点数和计算时间对于这套测量装置显然是不太现实的; ③ 进行完 FFT 变换后只能计算出螺距, 仍无法进行数据分段。

其二是采用基于动态线性回归的一阶到数分段法对数据进行分段。对原始数据导数研究不难发现, 在坐标轴零线附近导数多次穿越零线, 在理想极值点附近导数抖动很厉害。所以简单分段是不可能的, 必须对数据进行滤波。从滤波后的导数看出, 导数变化趋势比较理想, 可作为判断的依据。由于测试系统在起测时有一定的延时, 所以应去除起头的部分数据, 同样收尾部分也应去除一部分。这样找到的直线部分是比较粗略的, 不能把这些分段中的数据都作为直线部分处理, 而应根据齿型实际情况挑选线性度最好的那部分数据作为直线。

可以计算出管接箍内螺纹齿型曲线部分

$$L_2 = 0.745, \quad L_5 = 0.802$$

定义 L_2 和 L_5 占一个螺距长度的百分比分别为 Δ_2 和 Δ_5 :

$$\Delta_2 = L_2 / p = 0.745 / 3.175 = 0.2346 = 23.46\%$$

$$\Delta_5 = L_5 / p = 0.802 / 3.175 = 0.2526 = 25.26\%$$

所以直线部分占一个螺距的百分比为 $\Delta_2 + \Delta_3 = 48.82\%$, 这说明直线部分在一个螺距中并不是很多, 选择起来更加困难。

定义直线段数据点数标准 N_s , 采样步长 $L_s = 0.01$.

$$N_s = \min([L_2 / L_s], [L_5 / L_s]) = \min([0.745 / 0.01], [0.802 / 0.01]) = 73$$

实际测量中综合考虑其它因素 $N_s = 70$. 直线段数据点数标准 N_s 是一个非常重要的参数. 在进行动态线性回归寻找标准直线段时以 N_s 为单位进行线性回归, 选取方差最小的 N_s 个数据作为直线段数据.

综上所述, 不难得到选取螺纹数据直线段的步骤: ① 读取数据文件, 转换为螺纹采样数据; ② 利用 FFT 提取特征量, 判断数据是否为螺纹数据; ③ 对数据进行一阶低通滤波器; ④ 求取滤波后数据的导数; ⑤ 根据导数对数据进行粗略分段; ⑥ 对每段数据以 N_s 为单位进行线性回归, 比较每组回归的方法, 以方差最小的 N_s 个数点作为此段数据的直线段数据; ⑦ 根据要求决定是否剔除起头或结尾的数据.

分段后各类数据如图 3~图 6 所示.

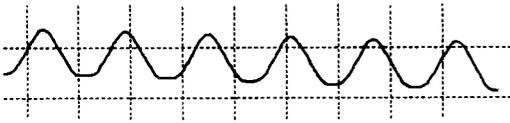


图3 采样滤波后的数据

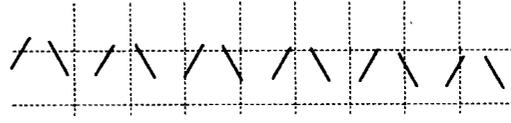


图4 分段后供拟合用的采样数据

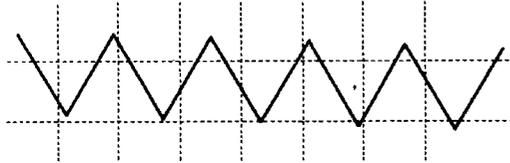


图5 拟合后的直线段

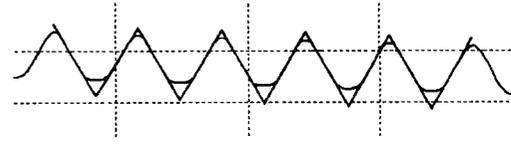


图6 拟合后的直线与原始采样数据

3 结束语

个据基于动态线性回归的一阶导数分段法对数据进行分段后计算出的各项基本参数和导出参数都满足测量的实际要求, 所以此方法是一种行之有效的对数据进行分段、分类的新方法.

参 考 文 献

- 1 石油工业部物资供应管理局组织翻译. API 专用管材手册. 北京: 石油工业出版社, 1980. 11~14
- 2 盛骤, 谢式千, 潘承毅. 概率论与数理统计. 北京: 高等教育出版社, 1989. 241~284

Application of First Differential Grouping Method Based on the Dynamic Linear Regression in Thread Testing

Zhou Qiang Zhang Huihui

(College of Mechanical Engineering and Applied Electronics Technology,
Beijing Polytechnic University, Beijing 100022)

Abstract A first differential grouping method based on dynamic linear regression is described. The method is successfully used in data processing of the thread testing. It has been verified that this method is efficient in grouping testing data.

Keywords linear regression, first differential, inner thread testing