

文章编号 1000-5013(2001) 03-0312-05

# 伏安法实验数据预处理方法

谢维波<sup>①</sup> 徐金瑞<sup>②</sup> 林劲松<sup>③</sup>

( ① 华侨大学信息科学与工程学院; ② 华侨大学材料科学与工程学院, 泉州 362011;

③ 厦门大学自动化系, 厦门 361005)

**摘要** 研究伏安法实验数据的预处理方法. 采用基于奇异值分解原理的 TLS 算法, 对经过预处理后的实验数据进行再处理, 取得良好的效果. 文中着重讨论各种预处理方法的合理性和有效性, 并对 TLS 算法的适用性进行探讨.  
**关键词** 基线提取, 特征识别, 峰点偏移, 子波形成, TLS 算法  
**中图分类号** TP 391 : O 657.129 **文献标识码** A

文献 [1] 针对计算机模拟的伏安法实验数据, 采用基于奇异值分解原理的 TLS 算法<sup>[2]</sup>, 有效地滤除了模拟数据中的白噪声. 为了把伏安法的实验数据改造成适于 TLS 算法处理, 实验数据的预处理采取基线提取和峰点对准两种方法. (1) 针对伏安法实验数据, 又可作出如下两点假设. (a) 底液的伏安曲线为二次曲线, 没有明显的峰形. (b) 一定浓度待测物质的伏安曲线为正向峰形曲线, 远离峰形的数据接近于零值. 所谓数据的趋势项, 是由全体数据拟合的二次曲线<sup>[3]</sup>. 上述的基线(底液的伏安曲线)不是数据的趋势项, 而是由非峰形(不包含待测物质的浓度信息)的数据点拟合而成的二次曲线. 这些数据点包括负值数据点, 以及远离正向峰形, 接近于零值的一段数据点. (2) 基线提取就是对伏安法实验数据的特征提取, 与上述关于伏安法实验数据的两点假设相吻合. 通过伏安法实验数据预处理, 获得了比较满意的处理结果.

## 1 数据预处理算法的构造与实现

### 1.1 基线提取

1.1.1 基线提取方法 1 如图 1 所示,  $Y_1$  为两组分溶液实验测得的 1 000 点数据波形. 其中两组分浓度分别为  $CS_1(1) = 5.0$ ,  $CS_1(2) = 5.0$ ,  $YY_1$  为基线提取后的波形. 基线提取的算法有 7 个步骤.

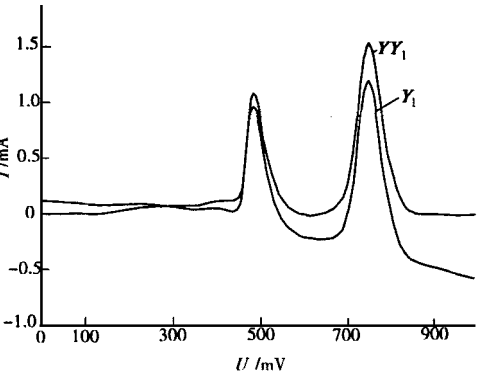


图 1 基线的提取

(1) 求取  $Y_1$  与横轴的交点  $k_0, k_1, k_2 (k_0 < k_1 < k_2)$ . (2) 求取  $Y_1$  在  $[k_0, k_1]$  区间内最小值对应的横坐标  $k_{\min}$ . (3) 在  $[1, 350]$  区间内, 用  $Y_1$  在  $[1, 350]$  区间内的数据拟合 1 条二次曲线. (4) 在  $[351, k_{\min} - 25]$  区间内, 用  $Y_1$  在  $[350, 360]$  和  $[k_{\min} - 50, k_{\min} - 25]$  区间内的数据拟合 1 条二次曲线. (5) 在  $[k_{\min} - 24, 1\ 000]$  区间内, 用  $Y_1$  在  $[k_{\min} - 24, k_{\min}]$  和  $[900, 1\ 000]$  区间内的数据拟合 1 条二次曲线. (6) 上述 3 条拟合曲线的拼接, 即为提取的基线. (7) 从  $Y_1$  中扣除提取的基线, 即为  $YY_1$  波形.

1.1.2 基线提取方法 2 如图 1 所示, 基线提取算法有 3 个步骤. (1) 在  $[1, 1\ 000]$  的区间内, 用  $Y_1$  在  $[1, 350]$  的区间内的数据拟合 1 条二次曲线. 从  $Y_1$  中减去这条曲线, 得到  $YY_1$ . (2) 在  $[1, 1\ 000]$  区间内, 用  $YY_1$  所有小于等于零的数据点拟合 1 条二次曲线. 从  $YY_1$  中减去这条曲线, 得到更新的  $YY_1$ . (3) 重复步骤 (2), 直至控制量满足给定精度 ( $1.0 \times 10^{-5}$ ). 控制量可为  $YY_1$  负向波形的最大幅度或拟合曲线的最大幅度.  $YY_1$  与图 1 中  $YY_1$  相比, 在  $[1, 400]$  非峰形区间内略有差别, 其余类似. 两者的特征都是正向峰形曲线, 其远离峰形的数据接近于零值. 符合上述关于物质浓度伏安曲线的客观假设. 图 1 中  $Y_1$  与  $YY_1$  峰点的偏移情况基本一致.  $Y_1$  与  $YY_1$  峰点的偏移, 反映了底液 (包括杂质和溶剂) 对峰点偏移的影响, 但可以看出校正的效果是很弱的. 基线提取方法 1 有着直观的合理性, 但过分依赖于主观的判断或选择, 缺乏科学的依据. 基线提取方法 2 仅依据浓度伏安曲线的特征识别, 所构造的叠代又能收敛于和方法 1 类似的结果. 它反映了浓度伏安曲线中基线的客观存在, 体现了方法的科学性、有效性及可行性.

## 1.2 峰点对准

子波曲线和浓度伏安曲线峰点偏移的校正, 可以有两种方法. (1) 截断移位法. 有限长度的波形平移实现峰点的对准, 需要对波形的截断. 由于波形峰点偏移的情况各不相同, 会造成各波形长度的不同. (2) 周期移位法. 将有限长度的波形视为周期拓展的周期波形, 周期平移实现峰点的对准, 则各波形长度相同. 这给数学处理带来方便, 也便于今后频域分析研究. 在数值计算中, 采用哪种移位方法, 最终结果几乎没有区别. 本研究只采用周期移位法, 即保持浓度伏安曲线不动, 只移位子波曲线. 同时, 将子波曲线  $E_{21}$  周期移位后的结果, 记为  $E_{21}\text{SHIFT}$ .

## 1.3 子波形成

伏安法中的标准加入法, 是一种近似的线性模型. 一个重要的原因是, 尽管采用了标准加入, 所获得的子波曲线并不“标准”. 因此, 子波的形成是本问题的关键. 图 2 为  $Y_1, Y_2$  和  $Y_3$  等 3 个波形. 其中  $Y_2$  是在  $Y_1$  的基础上, 标准加入为  $\Delta C(1) = 5.0$  后得到的实验数据,  $Y_3$  是在  $Y_2$  的基础上, 标准加入为  $\Delta C(2) = 2.5$  后得到的实验数据. 子波形成有 3 种方法. (1) 在基线提取之前形成子波. 记为  $EE_{21} = Y_2 - Y_1, EE_{32} = Y_3 - Y_2$ . (2) 在基线提取之后形成子波. 记为  $E_{21} = YY_2 - YY_1, E_{32} = YY_3 - YY_2$ . (3) 在实现两条浓度伏安曲线峰点的某种对准后, 形成子波. 比如保持  $YY_1$  浓度伏安曲线不动, 周期移位  $YY_2$ , 实现两者在  $[400, 600]$  (或  $[600, 900]$ ) 区间内峰点的对准. 记为  $E_{211} = YY_2\text{SHIFT} - YY_1$  (或  $E_{212} = YY_2\text{SHIFT} - YY_1$ ). 图 3 为  $EE_{21}$  和  $E_{21}$  的波形. 两者差别很小, 但在数

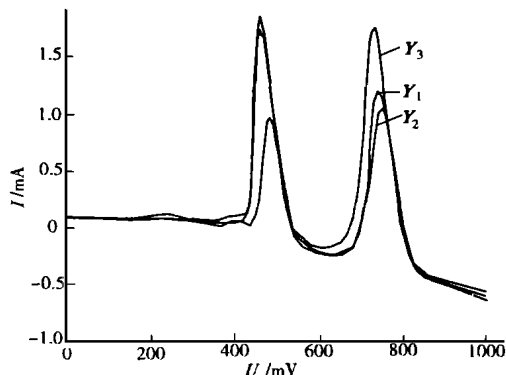


图 2 实验数据

值计算中, 最终结果却有区别. 图 4 为  $E_{211}$  和  $E_{212}$  的波形.  $E_{211}$  失真严重, 在数值计算中使用  $E_{212}$  最终结果也很差.

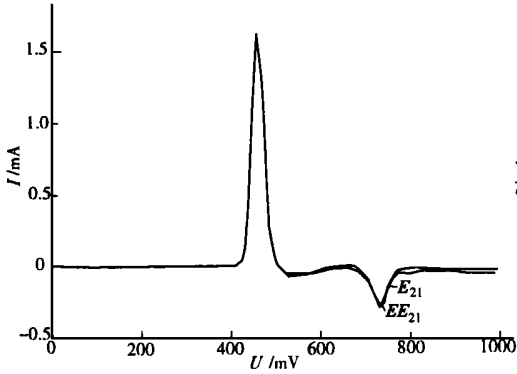


图 3 基线提取前后形成的子波

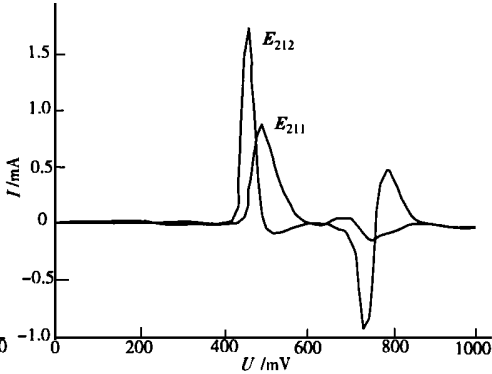


图 4 峰点对准后形成的子波

2 计算方案

综上所述, 可以组合出多种计算方案. 本节中的计算结果, 都是以标准加入浓度为单位 ( $\Delta C(1) = 5.0, \Delta C(2) = 2.5$ ) 而进行的讨论.

2.1 计算方案 1

用预处理前的数据构成线性方程, 进行 TLS 求解. (1)  $GS_1 = [-Y_1, EE_{21}, EE_{32}]$ ,  $GS_2 = [-Y_2, EE_{21}, EE_{32}]$ ,  $GS_3 = [-Y_3, EE_{21}, EE_{32}]$ .  $GS_1, GS_2, GS_3$  皆为  $1\ 000 \times 3$  的矩阵. (2)  $[U_1, S_1, V_1] = SVD(GS_1)$ ,  $[U_2, S_2, V_2] = SVD(GS_2)$ ,  $[U_3, S_3, V_3] = SVD(GS_3)$ . SVD 为奇异值分解运算.  $V_1, V_2, V_3$  皆为  $3 \times 3$  的矩阵. (3)  $CS_1 = [V_1(2, 3), V_1(3, 3)] / V_1(1, 3)$ ,  $CS_2 = [V_2(2, 3), V_2(3, 3)] / V_2(1, 3)$ ,  $CS_3 = [V_3(2, 3), V_3(3, 3)] / V_3(1, 3)$ .  $CS_1, CS_2, CS_3$  皆为二元行向量, 分别表示  $Y_1, Y_2, Y_3$  中两组分浓度的浓度向量, 如表 1 所示.

表 1  $YY_1, YY_2$  和  $YY_3$  两组分浓度的计算结果

方 案	$CS_1(1)$	$CS_1(2)$	$CS_2(1)$	$CS_2(2)$	$CS_3(1)$	$CS_3(2)$
计算方案 1	1.271 3	2.814 4	2.208.5	1.692 3	1.851 0	2.615 9
计算方案 2	0.996 6	1.898 7	1.678 7	1.929 5	1.723 5	2.751 6
计算方案 3	0.877 0	1.785 2	1.552 0	1.852 1	1.548 0	2.593 7
准确值	1.0	2.0	2.0	2.0	2.0	3.0

2.2 计算方案 2

用提取方法 2 和周期移位法实现数据的预处理(在基线提取之后形成子波), 构成线性方程, 进行 TLS 求解. 即  $GS_1 = [-YY_1, E_{21} \text{ SHIFT}, E_{32} \text{ SHIFT}]$ ,  $GS_2 = [-YY_2, E_{21} \text{ SHIFT}, E_{32} \text{ SHIFT}]$ ,  $GS_3 = [-YY_3, E_{21} \text{ SHIFT}, E_{32} \text{ SHIFT}]$ (表 1).

2.3 计算方案 3

用提取方法 2 和周期移位法实现数据的预处理(在基线提取之前形成子波), 构成线性方程, 进行 TLS 求解. 即  $GS_1 = [-YY_1, EE_{21} \text{ SHIFT}, EE_{32} \text{ SHIFT}]$ ,  $GS_2 = [-YY_2, EE_{21} \text{ SHIFT}, EE_{32} \text{ SHIFT}]$ ,  $GS_3 = [-YY_3, EE_{21} \text{ SHIFT}, EE_{32} \text{ SHIFT}]$ (表 1).

### 3 计算结果的修正

图5为子波 $E_{21}$ 和 $E_{32}$ 的波形. 观察 $E_{21}$ 和 $E_{32}$ (或 $EE_{21}$ 和 $EE_{32}$ )波形的关系, 可以看出两者波形失真的内在联系. 这在文献[1]中有提到.  $E_{21} + E_{32}$ 的波形失真一定较小. 因此, 由方程 $YY_3 - YY_1 = CS(1)E_{21}SHIFT + CS(2)E_{32}SHIFT$ 求出的 $CS(1)$ ,  $CS(2)$ 值, 一定都很接近于1. 计算结果确实如此. 据此可定出修正方案. 下面举例说明.

(1) 计算方案2的结果分别除以 $CS(1)$ ,  $CS(2)$ , 可得 $CS_1(1) = 0.9978$ ,  $CS_2(1) = 1.6806$ ,  $CS_3(1) = 1.7255$ ;  $CS_1(2) = 1.9076$ ,  $CS_2(2) = 1.9385$ ,  $CS_3(2) = 2.7645$ .

(2) 计算方案3的结果分别除以 $CS(1)$ ,  $CS(2)$ , 可得 $CS_1(1) = 0.9518$ ,  $CS_2(1) = 1.6205$ ,  $CS_3(1) = 1.6163$ ;  $CS_1(2) = 1.8718$ ,  $CS_2(2) = 1.9419$ ,  $CS_3(2) = 2.7195$ .

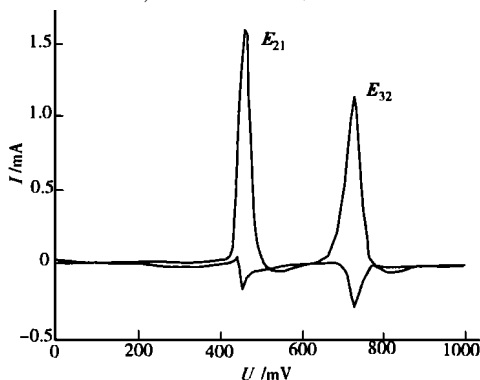


图5 两子波的畸变及相关性

所有数据的精度均获得一致性的改善. 另外, 理论上有如下恒等式, 即

$$CS_2(1) - CS_1(1) = 1\Delta C(1), \quad CS_3(1) - CS_2(1) = 0\Delta C(1),$$

$$CS_2(2) - CS_1(2) = 0\Delta C(2), \quad CS_3(2) - CS_2(2) = 1\Delta C(2),$$

以及 $CS_1(1)$ ,  $CS_1(2)$ 总是相对地给出较高精度. 因此在上述修正的基础上, 可定出关于 $CS_2$ ,  $CS_3$ 的二次修正方案, 以计算方案3的结果为例. (1)  $GGS = [-YY_2, (1.0 + CS_1(1))E_{21}SHIFT, CS_1(2)E_{32}SHIFT]$ ,  $[UU, SS, VV] = SVD(GGS)$ ,  $CCS = [VV(2, 3), VV(3, 3)]/VV(1, 3)$ . (2)  $CS_2 = [CS_2(1)/CCS(1), CS_2(2)/CCS(2)]$ . 所得为 $CS_2(1) = 2.0768$ ,  $CS_2(2) = 2.0495$ . (3)  $GGS = [-YY_3, (1.0 + CS_1(1))E_{21}SHIFT, (1.0 + CS_1(2))E_{32}SHIFT]$ ,  $[UU, SS, VV] = SVD(GGS)$ ,  $CCS = [VV(2, 3), VV(3, 3)]/VV(1, 3)$ . (4)  $CS_3 = [CS_3(1)/CCS(1), CS_3(2)/CCS(2)]$ . 所得为 $CS_3(1) = 2.0144$ ,  $CS_3(2) = 3.1758$ . 其它计算方案经二次修正后, 所有数据的精度也均能获得一致性的改善. 可见计算结果修正的有效性.

### 4 讨论

(1) 子波的引用至关重要. 上述计算都是以标准加入浓度的数据构成线性方程, 而不是以单位浓度的数据构成线性方程. 对于线性问题, 两者不应有区别. 以计算方案1为例, 其计算结果尚有参考价值. 而使用单位浓度数据 $ee_{21} = EE_{21}/\Delta C(1)$ 和 $ee_{32} = EE_{32}/\Delta C(2)$ 求解时, 计算结果严重失真(括号内为准确值). 即

$$CS_1(1) = 24.3301(5.0), \quad CS_2(1) = 13.1233(10.0), \quad CS_3(1) = 12.0024(10.0),$$

$$CS_1(2) = 5.9579(5.0), \quad CS_2(2) = 3.0135(5.0), \quad CS_3(2) = 5.7185(7.5).$$

这说明了本问题的非线性. 从文献[2]中可以知道, 基于奇异值分解原理的TLS算法适用于线性模型, 叠加在数据上的干扰应为强度均匀的白噪声. 当数据上的干扰不均匀时, 应对数据进行加权处理. 本问题的求解, 应采用某种加权的TLS算法. 以上述较为准确的方程 $YY_3 - YY_1 =$

$CS(1)E_{21}SHIFT + CS(2)E_{32}SHIFT$  为例, 矩阵 $[-(YY_3 - YY_1), E_{21}SHIFT, E_{32}SHIFT]$  各列数据的干扰强度较为均匀, 而矩阵 $[-(YY_3 - YY_1), E_{21}SHIFT/\Delta C(1), E_{32}SHIFT/\Delta C(2)]$  各列数据干扰强度的均匀性就被破坏. 其它计算方案也都能有相同的结论, 即使用标准加入浓度的数据优于使用单位浓度的数据.

(2)  $E_{211}$  期望的峰形与浓度数据的峰形吻合得较好, 但其上叠加的干扰太强, 以致完全丧失了子波曲线具有单峰的特征. 或许可以构造出  $E_{211}$  和  $E_{321}$  的联合相关滤波, 则可望滤除这种干扰. 实际上,  $E_{21}$  或  $E_{32}$  的峰形与浓度伏安曲线的峰形差异太大. 尽管我们较为准确地计算出它们之间的数值关系( $CS(1)$ ,  $CS(2)$ ), 却没有解决子波峰形的辨识问题.

(3) 在基线提取之后, 峰点偏移是非线性的主要根源, 也是对子波峰形辨识的难点所在. 文献[4]采用卡尔曼滤波, 在实验数据有明显的峰点偏移时, 浓度的求解精度大大下降. 峰点偏移的复杂性, 在于不仅峰形的宽度发生了变化, 而且双峰偏移的程度并不一致. 这应属于时变滤波的问题. 对子波峰形尝试使用某种类型的函数进行拟合, 均告失败.

(4) 在各种计算方案结果中,  $CS_1$  最准,  $CS_2$  和  $CS_3$  较差. 按理,  $Y_2, Y_3$  中待求的浓度值较高, 信息量较大,  $CS_2$  和  $CS_3$  的计算值应该较准才对. 究其原因, 形成的子波与  $Y_2, Y_3$  的线性程度较  $Y_1$  差. 这正说明了实现  $Y_1, Y_2$  和  $Y_3$  联合的相关滤波, 是本问题求解的关键.

(5) 计算结果的修正实际上是一种加权处理. 这种加权处理是对整个子波乘上一个系数, 无助以子波峰形的辨识.

## 参 考 文 献

- 1 谢维波, 徐金瑞. 新极谱法实验数据分析的数值研究[J]. 学会, 2000, (10): 182 ~ 184
- 2 张贤达. 现代信号处理[M]. 北京: 清华大学出版社, 1995. 68 ~ 72
- 3 程乾生. 信号数字处理的数学原理[M]. 北京: 石油工业出版社, 1979. 504 ~ 529
- 4 许荣榕. 新极谱法实验数据的卡尔曼滤波[D]. [学位论文]. 福州: 福州大学化学系, 1989. 22 ~ 23

## Method for Preprocessing Experimental Data of Voltammetry

Xie Weibo<sup>①</sup> Xu Jinrui<sup>②</sup> Lin Jinsong<sup>③</sup>

(<sup>①</sup> College Info. Sci. & Eng., <sup>②</sup> College of Mater. Sci. & Eng., Huaqiao Univ., 362011, Quanzhou;

<sup>③</sup> Dept. of Automotion, Xiamen Univ., 361005, Xiamen))

**Abstract** A study is made on the method for preprocessing the data from voltammetric experiment. Good effect can be obtained for the experimental data passed through preprocessing and then processed by adopting total least square (TLS) algorithm which is based on the principle of singular value decomposition. The authors discuss the reasonableness and effectiveness of various methods of preprocessing, and inquire into the compatibility of TLS algorithm.

**Keywords** datum-line drawing, feature recognition, peak point shifting, wavelet formation