

最优化算法在 EDXRF 谱线拟合中的应用

刘明博^{1,2,3}, 廖学亮³, 胡学强³, 韩冰³, 沈学静^{1,2,3}, 陈吉文⁴, 王国华⁵, 王海舟^{1,2,3*}

1. 钢铁研究总院, 北京 100081
2. 金属材料表征北京重点实验室, 北京 100081
3. 钢研纳克检测技术股份有限公司, 北京 100081
4. 北方工业大学电气与控制工程学院, 北京 100144
5. 北京大学附属中学, 北京 100080

摘要 在 X 射线荧光光谱(EDXRF)测试中, 由于探测器分辨率的影响, 谱线会有不同程度的展宽, 多元素存在时谱线还会有一定的重叠。获得纯净的谱线强度, 并保证准确测试结果的过程, 通常被称为 EDXRF 解谱。传统的解谱方法包括感兴趣区加和、纯元素谱线剥离以及干扰系数法等, 但这些方法在解决 EDXRF 解谱问题时都有不同程度的局限性。最优化算法本质上是一种多自变量求极值的方法, 首先列出具有多个自变量参数并能够描述真实物理学过程的目标方程, 然后设定参数初值和边界条件, 通过数学(如共轭梯度法)的运算, 得到最优化的目标方程解。将最优化计算的思想应用于 EDXRF 解谱过程中, 假设 EDXRF 谱图是一系列理想高斯峰的叠加, 可以列出其与原始谱图残差的目标方程, 根据物理学现象对这些高斯峰的高度、位置和宽度三个参数分别进行估计, 此时目标方程的值通常较大, 通过设定所有参数各自的边界条件, 用共轭梯度法不断调整, 做最优化计算, 直至该值达到极小, 此时多峰叠加的结果与实测谱最为贴合, 解谱精度大大提高。利用这种方法研究开发了能用于商业化 EDXRF 仪器的软件程序。对最优化计算用于 EDXRF 解谱的方法进行了介绍, 并以 Pr 和 Nd 混合液的 L 系 EDXRF 谱图的解析举例, 这段谱线由 11 个谱峰组成, 通过设定 33 个高斯峰参数的初始值, 用共轭梯度法执行 33 个自变量的最优化计算, 运行于普通计算机, 经过 580 ms 的计算, 拟合谱与实测谱的残差从 37.645 减小为 1.699 4, 二者在对比谱图上也极为吻合, 从而说明这是一种比较有效的谱线拟合方法。通过对结果的分析还发现, 部分谱线的宽度发生了变化, 真实地反映了该条谱线是由多条相邻谱线构成的事实。研究创新性在于, 将数学中的最优化计算原理应用于 EDXRF 谱图解析过程中, 获得了较好的效果, 并以较为复杂的 Pr 和 Nd 双稀土元素的 L 系谱线解析为例进行了说明。

关键词 X 射线荧光光谱; 能谱; 解谱; 最优化计算

中图分类号: O582 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)04-0997-04

引言

在能量色散 X 射线荧光光谱(EDXRF)分析中, 当多种元素共存时, 由于探测器分辨率的影响, 谱线重叠不可避免, 要得到准确的谱峰强度(或峰面积), 即解谱就变得十分困难, 而要在后续的标定或测试时得到准确的结果, 却又必须有待测谱线的准确强度。因此解谱就成了 EDXRF 的关键技术。普遍采用的解谱方法包括感兴趣区积分(ROI)法、纯

元素谱扣除法、干扰系数法^[1-2]等, 以图 1 所示的 Pr 和 Nd 混合液样品谱图的解析为例分别介绍三种方法。

感兴趣区积分法是划定谱线左右边界直接对区域内的信号进行累加, 这种方法仅适用于谱线比较简单, 当有一定的重叠时就会有很大偏差, 如图 1 中 Pr: La 和 Nd: La 由于重叠而被互相抬高, 划定区域积分会导致 2 个峰的强度都偏大。

纯元素谱扣除方法需要先在相同的测试条件下采集所有待测元素对应纯物质的谱图, 在测试混合样品时依一定的次

收稿日期: 2019-03-14, 修订日期: 2019-07-22

基金项目: 国家重点研发计划(2017YFD0801201, 2017YFF0108900), 科技部国家重大科学仪器设备开发专项(2012YQ050076)资助

作者简介: 刘明博, 1981 年生, 钢铁研究总院博士研究生 e-mail: liumingbo@ncschina.com

* 通讯联系人 e-mail: wanghaizhou@ncschina.com

序将纯物质谱剥离,但当谱线有重叠时,先扣除的谱强度偏大,后扣除的则偏小,另外由于吸收增强效应的存在,在混合样品中单元素的谱峰分布会与纯物质中的不同,也会引起解谱误差。如图 1 中,用纯元素谱扣除方法进行解谱时,无论用哪个峰对齐进行扣除都会导致偏差的产生,此外由于高浓度的 Pr 和 Nd 共存,两种物质的谱分量形状跟纯物质有很大不同,更会引起解谱误差。如果结合基本参数法和最优化方法的计算对纯元素扣除方法进行矫正,是一种比较精确的方法。

干扰系数法则是根据谱峰位置,计算谱峰重叠干扰程度,确定相邻峰在叠加峰中的比例,但这种方法仅适用于两个峰的情形,当多峰共存时,由于干扰系数过于复杂,误差将无法预料,图 1 中 11 个峰共存,每个峰都会受到相邻峰的干扰,用这种方法的解谱误差会很大。

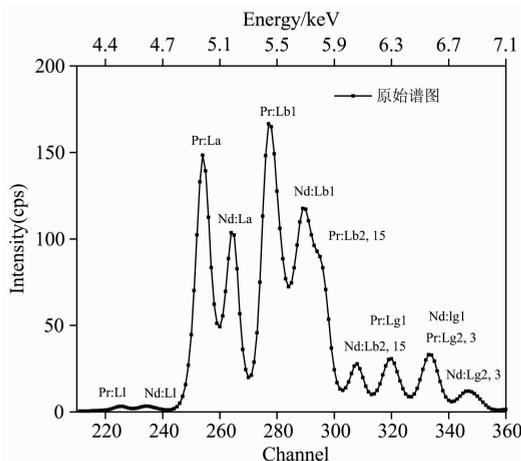


图 1 某 Pr 和 Nd 混合液样品的 EDXRF 谱图局部
Fig. 1 A part of Pr/Nd solution EDXRF spectrum

简单的感兴趣区积分法、纯元素谱扣除法和干扰系数法都无法很好地对图 1 的谱图完成精确的解析。本文介绍了一种基于最优化算法(optimize method)的多高斯函数拟合算法,用于 EDXRF 谱线的解析。

1 最优化算法

最优化算法本质上是一种多元函数求极值的数值计算方法,通过对一系列自变量设定初值,代入约束函数并计算与目标值的偏差,以一定的规则改变自变量分布,迭代计算偏差,直到计算的偏差结果达到极小值,即认为得到了最优的自变量分布。最优化计算方法在工程设计、理化分析甚至人文和社会科学等领域都有广泛应用。但在 EDXRF 的解谱方面,却少有应用的先例。

最优化计算过程通常是寻找多自变量非线性方程的极值,往往需要大量的数值计算来完成,在过去计算能力不足时,往往通过减少参量、精确设定初值来减少运算量,但随着现代计算机技术的发展,CPU 采用多核主频达数 GHz,计算机就可以在一秒时间内完成数十自变量的最优化计算,因此更精确的优化计算变得非常容易。在此条件下,研究中

将复杂的最优化计算引入看似简单实则极其复杂的 EDXRF 解谱过程中。

2 最优化算法的谱线拟合

EDXRF 谱图可以看成一系列符合高斯分布的谱峰叠加在连续背景上的结果,在执行了有效的背景扣除算法之后,谱图就可以看成是一系列不同高度、位置和宽度的高斯函数的加和。

一个独立的高斯函数由高度 A 、位置 c 、代表宽度的 ω 三个参数确定

$$y = A \exp \left[\frac{-(x-c)^2}{2\omega^2} \right] \quad (1)$$

由 n 个高斯函数组成的 XRF 谱图可以如式(2)表示

$$y_i = \sum_{j=1}^n A_j \exp \left[\frac{-(i-c_j)^2}{2\omega_j^2} \right] \quad (2)$$

($i = 1, 2, 3, \dots, 2048$ 为谱图道址)

其与原始谱图 Y_i 的残差

$$\sigma = \sum_{i=1}^{2048} (Y_i - y_i)^2 = \sum_{i=1}^{2048} \left\{ Y_i - \sum_{j=1}^n A_j \exp \left[\frac{-(i-c_j)^2}{2\omega_j^2} \right] \right\}^2 \quad (3)$$

式(3)中,道址 i 和原始谱 Y_i 为已知量; A_j , c_j , ω_j ($j=1, 2, 3, \dots, n$) 为自变量,分别表示 n 个高斯函数的高度、位置和宽度参数。

上述自变量还需要满足下列边界条件:

- (1) $A_j > 0$, 高斯函数高度不能为负;
- (2) $|c_j - c_{j0}| < 3$, 探测器在开机经过峰位微调后,峰位 c_j 与其假设的初始值偏差不会过大;

(3) $0.8 < |\omega_j / \omega_{j0}| < 1.25$, 谱线宽度与其初始假设值不能偏差过大,即谱线分辨率不会与理论值有较大偏差,但考虑到部分谱线可能由多条谱线叠加而成,因此宽度范围适度放大。

利用最优化算法对有 $3n$ 个自变量的式(3)求极值,即可得到一系列自变量分布,此时,在数值上 σ 最小,表现为拟合得到的 n 个高斯峰叠加结果与原始谱图最为接近。

3 实际应用

以图 1 所示 Pr 和 Nd 混合溶液的 EDXRF 谱图解析为例,在 $4.7 \sim 7.0$ keV 能量范围内,执行了背景扣除和峰位识别后,可以认为该谱图由 Pr 和 Nd 的各 6 条较为明显的 L 系谱线高斯峰构成,根据谱峰能量确定高斯峰的位置初始值 c_{j0} , 根据该位置处的谱峰高度确定高度初始值 A_{j0} , 根据探测器分辨率和能量的关系确定宽度初始值 ω_{j0} 。

将表 1 中 12 条谱线 11 个谱峰(Nd: Lg1 和 Pr: Lg2, 3 能量过于接近,合并为 1 个峰)所对应高斯函数的 33 个参数,代入式(3)中,此时 11 个高斯峰叠加结果与原始谱图的对比关系如图 2(a)所示,两者差异明显,残差高达 37.645。

以表 1 的 33 个参数为残差方程的初始值,用共轭梯度算法进行最优化计算之后,11 个高斯函数的参数如表 2 所

表 1 谱峰参数初始估计值

Table 1 Initial guess of gaussian peak parameters

<i>j</i>	元素	谱线	能量 /keV	c_{j0} /ch	A_{j0} /cps	ω_{j0} /chs
1	Pr	L1	4.453	224.648	2.859	2.578
2	Nd	L1	4.633	234.869	3.317	2.623
3	Pr	La	5.035	254.036	148.331	2.707
4	Nd	La	5.228	264.547	103.514	2.753
5	Pr	Lb1	5.492	277.176	166.525	2.810
6	Nd	Lb1	5.719	289.210	117.653	2.863
7	Pr	Lb2, 15	5.849	295.898	89.772	2.893
8	Nd	Lb2, 15	6.088	307.859	26.460	2.947
9	Pr	Lg1	6.325	319.690	30.206	3.001
10	Nd	Lg1	6.601	333.497	33.084	3.064
11	Pr	Lg2, 3	6.617	333.497	33.084	3.064
12	Nd	Lg2, 3	6.901	347.756	12.045	3.130
σ				37.645		

列, 叠加结果与原始谱图对比如图 2(b)所示, 吻合度较好, 残差也减小为 1.699 4。

表 1 和表 2 的对比可以发现, 宽度偏差较大, 这是由于部分峰(如 Lb1 等)实际是由一系列位置比较接近的谱峰叠加而成的, 这会导致实际峰型宽度变大, 初始估计时忽略这一现象, 但最优化计算会对其进行弥补; 高度偏差较大, 说明初始的粗略估计有偏差; 峰位差异不明显。

表 2 谱峰参数优化结果

Table 2 Parameters of Gaussian peaks optimized

<i>j</i>	元素	谱线	能量 /keV	c_{j0} /ch	A_{j0} /cps	ω_{j0} /chs
1	Pr	L1	4.453	224.800	2.454	2.805
2	Nd	L1	4.633	234.818	2.618	3.722
3	Pr	La	5.035	254.182	143.005	2.760
4	Nd	La	5.228	264.243	101.085	2.665
5	Pr	Lb1	5.492	277.645	163.533	3.002
6	Nd	Lb1	5.719	289.369	113.770	4.303
7	Pr	Lb2, 15	5.849	296.473	47.692	2.439
8	Nd	Lb2, 15	6.088	307.568	25.922	3.107
9	Pr	Lg1	6.325	319.755	29.239	3.306
10	Nd	Lg1	6.601	333.379	32.098	3.176
11	Pr	Lg2, 3	6.617	333.379	32.098	3.176
12	Nd	Lg2, 3	6.901	346.739	11.246	3.778
σ				1.6994		

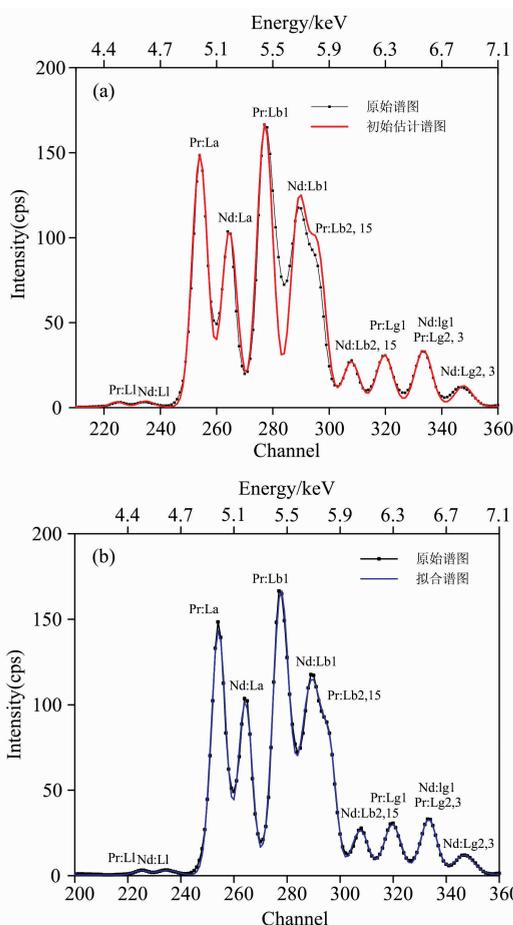


图 2 (a)初始估计参数构成的 11 个高斯峰叠加结果与原始谱图的对比; (b)最优化计算后的高斯峰叠加与原始谱图的对比

Fig. 2 (a) Composition of 11 Gaussian peaks with initial guess parameters and the raw spectrum; (b) Composition of 11 Gaussian peaks with parameters optimized and the raw spectrum

用 EDXRF 仪器测试, 最优化算法解谱, 经过标定之后, 对某一稀土湿法生产线上 Pr/Nd 混合液样品中的 Pr 和 Nd 含量配分进行多次连续测试, 并与 ICP 方法测得结果进行对比, 结果如表 3。该方法稳定性较好, $RSD < 0.2\%$, 准确性与 ICP 标准方法相差 0.08, 相对较小。

表 3 实际测试及与 ICP 方法对比

Table 3 Comparison of EDXRF and ICP test

	Pr/%	Nd/%
1	64.49	35.51
2	64.47	35.53
3	64.53	35.47
4	64.49	35.51
5	64.38	35.62
6	64.57	35.43
7	64.37	35.63
相对标准偏差 RSD/%	0.106	0.193
XRF 测试平均值	64.47	35.53
ICP 测试值	64.55	35.45
偏差(XRF-ICP)	-0.08	0.08

4 结 论

借助于现代计算机和软件技术, 在普通 PC 上大计算量的数值模拟计算变得可能, 本研究将最优化计算方法应用于看似简单的 EDXRF 谱图解析过程中, 通过对 Pr 和 Nd 混合

液 EDXRF 谱图的解析举例可见, 拟合的谱图与实测谱图的残差较少, 吻合也较好, 可以大幅度提高解谱精度, 且计算

的理论结果与实测结果自动匹配, 需要的人工干预较小, 该方法可以在 EDXRF 解谱过程中推广应用。

References

- [1] JI Ang, TAO Guang-yi, ZHUO Shang-jun, et al(吉 昂, 陶光仪, 卓尚军, 等). X-Ray Fluorescence Spectrum Analysis(X 射线荧光光谱分析). Beijing: Science Press(北京: 科学出版社), 2003. 99
- [2] GUO Cheng, LAI Wan-chang, YI Xin, et al(郭 成, 赖万昌, 易 欣, 等). Science and Technology of West China(中国西部科技), 2013, (4): 46.

Application of Optimize Method in EDXRF Spectrum Unfolding

LIU Ming-bo^{1, 2, 3}, LIAO Xue-liang³, HU Xue-qiang³, HAN Bing³, SHEN Xue-jing^{1, 2, 3}, CHEN Ji-wen⁴, WANG Guo-hua⁵, WANG Hai-zhou^{1, 2, 3*}

1. Central Iron and Steel Research Institute, Beijing 100081, China

2. Beijing Key Laboratory of Metal Materials Characterization, Beijing 100081, China

3. NCS Testing Technology Co., LTD., Beijing 100081, China

4. School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China

5. The Affiliated High School of Peking University, Beijing 100080, China

Abstract With the restriction of detector resolution in EDXRF, the peaks will be widened, and the overlapping of several peaks cannot be neglected especially in multi elements sample. The process of achieving the intensity of pure peaks, named spectrum unfolding is difficult. There are three unfolding methods traditionally, ROI integration, pure element peeling, and the coefficient method. Both of the methods cannot unfold the complex EDXRF spectrum perfect. In this work, we introduce an optimized algorithm to EDXRF spectrum unfolding process. Pure peaks, processed by applying the smooth and debackground operation to raw spectrum, can be considered as the sum of several Gaussian peaks, which can be formulated by Gaussian functions with 3 parameters each, height, width and position. By alternating the values of the 3N parameters within certain boundary conditions, the error between theory function and raw spectrum can be smaller and smaller. In mathematics, this alternating process can be finished by Optimize algorithm. When we developed the algorithm by 3N parameters Conjugate Gradient Method and coding the program, an EDXRF spectrum of Pr/Nd solution was studied. Sum of 11 Gaussian peaks with 33 parameters was very well closed to the raw spectrum composed of 11 L-series peaks of Pr and Nd, and the error calculated changed from 37.645 to 1.699 4. The innovative of this paper is the application of optimized method in complex EDXRF spectrum unfolding process.

Keywords Energy dispersive X-ray fluorescence; Energy spectroscopy; Spectrum unfolding; Optimization calculation

(Received Mar. 14, 2019; accepted Jul. 22, 2019)

* Corresponding author