

基于离散极大值法的太赫兹光谱特征吸收峰提取方法

曹雨齐², 康旭升^{1, 2*}, 陈飘云², 谢晨², 喻洁^{2*}, 黄平捷², 侯迪波², 张光新²

1. 浙大城市学院计算机与计算科学学院, 浙江 杭州 310015

2. 浙江大学控制科学与工程学院, 工业控制技术国家重点实验室, 浙江 杭州 310027

摘要 太赫兹波在电磁波谱中介于微波和红外辐射之间, 具有指纹特性、安全无损、强穿透性等特点, 因此太赫兹光谱技术在药品成分和组成检测领域具有广泛应用价值。针对高纯度物质识别研究中存在部分弱吸收峰不易识别, 以及混合物的太赫兹光谱中吸收峰强度降低而导致吸收峰位信息模糊化的问题, 提出了一种基于离散极大值法的光谱吸收峰位识别方法, 即伴随拐点法。伴随拐点法首先利用目标检测物太赫兹吸收系数谱图的一阶和二阶导数确定吸收峰位的伴随拐点和基线谱, 其次将原始吸收光谱与基线谱进行差分运算得到差谱, 最后根据离散极大值法确定吸收峰位, 从而实现特征吸收峰的识别。为验证伴随拐点法的有效性, 采用伴随拐点法对四种硝基呋喃类样品光谱进行吸收峰提取, 并将吸收峰位识别结果与仿真结果进行比较。实验结果证明, 伴随拐点法能有效识别目标检测物的强吸收峰和弱吸收峰。该方法不仅在含峰目标物的太赫兹特征吸收峰识别问题中具有广泛的应用前景, 还适用于其他光谱的谱峰峰位检测。

关键词 太赫兹时域光谱; 特征吸收峰识别; 光谱吸收峰; 光谱特征信息

中图分类号: O433 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3058-05

引言

太赫兹波是频率位于 0.1~10 THz (1 THz=10¹² Hz) 的电磁波, 其波长范围为 0.03~3 mm。由于太赫兹波具有低能性、指纹特性、强穿透性等特点, 因此太赫兹时域光谱技术被广泛应用于食品安全检测^[1-5]、药品质量检测^[6-10]、医学诊断^[11-16]等多个应用领域。

太赫兹光谱中的特征吸收峰在含峰目标物的鉴定研究中起着重要作用^[17-20]。浙江大学张宏建团队利用四种杀虫剂独特的特征吸收峰强度对四种杀虫剂进行定量检测^[17]。三聚氰胺混合物在 0.1~3.0 THz 频率范围内被观察到三个特征吸收峰, 分别位于 2.00, 2.26 和 2.60 THz 处。该研究根据混合物样品的太赫兹吸收系数谱图中的特征峰信息对三聚氰胺进行检测。结果表明, 太赫兹时域光谱技术在食品质量检测中具有潜在的应用价值^[18]。

上述研究中, 样品均具有明显特征吸收峰, 可被肉眼识别, 但部分弱峰可能被忽略, 从而影响含峰目标物含量的精准检测。因此研究一种由计算机自主有效地提取太赫兹光谱中的吸收峰位特征信息的方法, 是太赫兹波广泛应用于各领

域物质识别的必要性研究。

为实现药品的太赫兹弱吸收峰的精准识别, 本文提出了伴随拐点法, 并将其应用于食品添加剂硝基呋喃类化合物的太赫兹吸收峰识别中。研究结果表明伴随拐点法对弱吸收峰具有较好的识别作用。

1 实验部分

1.1 实验系统与样品制备

实验采用 Z3 太赫兹时域光谱系统 (Zomega Terahertz Corporation, USA), 系统的详细介绍可参考文献[21]。实验测试的各硝基呋喃类药物的纯物质均为结晶性粉末。为制备样品压片, 首先将样品粉末放置于真空干燥柜中干燥, 然后在玛瑙研钵中充分研磨。最后, 将样品粉末压制成厚度为 1 mm、直径为 13 mm 的样品压片。由于纯硝基呋喃粉末不易压制成型, 需与聚乙烯粉末按 1:1 的比例进行混合后压片。

1.2 离散极大值法

连续极大值法是指根据连续函数曲线 $f(x)$ 中 x 点的一阶和二阶导数 $f'(x)$ 和 $f''(x)$ 数值符号来判断该点是否为特征吸收峰位。如果 $f'(x)=0$ 且 $f''(x)<0$, 则将 x 点判定为 f

收稿日期: 2021-08-05, 修订日期: 2021-10-22

基金项目: 国家自然科学基金项目(61873234), 杭州市科委项目(20191203B43)资助

作者简介: 曹雨齐, 女, 1994年生, 浙江大学控制科学与工程学院特聘副研究员 e-mail: yuqicao1994@gmail.com

* 通讯作者 e-mail: kangxs@zucc.edu.cn; yu_jie@zju.edu.cn

(x) 的吸收峰位。

然而由于太赫兹时域光谱系统采样不连续，因此频域谱图为离散函数，则可能存在含峰物质的峰位未被采样，导致没有 x 满足 $f'(x)=0$ 。因此，针对离散数据，判别吸收峰位的条件为对点 x ， $f'(x)$ 的取值需经过从正值到负值的转换，同时 $f''(x)<0$ 。由于实验系统中 Δx 为固定的非零常数， $f'(x)$ 能够利用中心差分公式近似计算，如式(1)。同理计算得到 $f''(x)$ ，如式(2)

$$f'(x) \approx \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} \quad (1)$$

$$f''(x) \approx \frac{f(x+\Delta x) - 2f(x) + f(x-\Delta x)}{\Delta^2 x} \quad (2)$$

根据离散极大值法识别特征吸收峰的方法如下。设 x_i 为频段范围内任意频率点，分别利用式(1)和式(2)近似计算吸收系数谱图的一阶和二阶导数，若 $f'(x_{i-1})f'(x_i)<0$ 且 $f''(x_i)<0$ ，则判定 x_i 为特征谱的极大值点或者吸收峰位。

1.3 伴随拐点法

虽然离散极大值法对强吸收峰具有较强识别能力，但它对弱吸收峰的敏感性不强，这是因为离散极大值法仅根据该点或在邻域内确定峰值位置。若吸收峰位于采样间隔 Δx 内而没有被采样，则一阶导数的离散曲线可能并不形成过零点，导致弱吸收峰被忽略。因此，本文基于离散极大值法提出一种能够同时测定强吸收峰和弱吸收峰的吸收峰识别方法，即伴随拐点法。

伴随拐点法利用特征吸收峰的伴随拐点信息来识别对应的特征吸收峰，其示意图如图 1 所示，其中 a, b, c, d 为吸收峰， a_1 与 a_2, b_1 与 b_2, c_1 与 c_2, d_1 与 d_2 分别为 a, b, c, d 的伴随拐点对。伴随拐点法的具体步骤如下：(1) 获取样品的原始太赫兹吸收系数谱(Origin，如图 1 蓝色曲线所示，其中部分曲线被红色曲线重叠覆盖未显示)，确定拐点；(2) 确立拐点中的各伴随拐点对(如图 1 中的 a_1 与 a_2, b_1 与 b_2)；(3) 连接伴随拐点对，与原始谱共同确定基线谱(Baseline，如图 1 中红色曲线所示)；(4) 将原始谱与基线谱作差，确定原始谱与基线谱的差谱(Difference，如图 1 中褐色曲线所示)；(5) 利用离散极大值法计算差谱的吸收峰位，即为目标检测物的特征吸收峰。

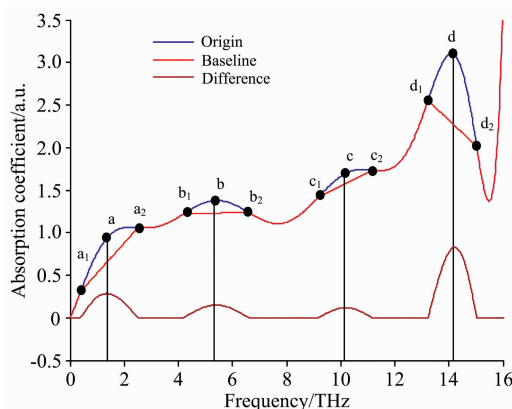


图 1 伴随拐点法示意图

Fig. 1 Schematic diagram of discrimination of absorption peaks using AIP method

伴随拐点法中伴随拐点对的确定方法如下。首先，设 x_i 为频段范围内任意频率点，若满足 $f''(x_{i-1})f''(x_i)<0$ ，则判定 x_i 为特征谱的拐点。然后，根据 $f''(x_{i-1})$ 和 $f''(x_i)$ 的符号将拐点分类，即起势拐点和落势拐点。定义若 $f''(x_{i-1})>0$ 且 $f''(x_i)<0$ ，则 x_i 为起势拐点；若 $f''(x_{i-1})<0$ 且 $f''(x_i)>0$ ，则 x_i 为落势拐点。最后，将连续出现的一对起势拐点和落势拐点作为一对伴随拐点。

为避免伴随拐点法过敏感，本研究提出差谱中对应峰值强度设定吸收峰的识别规则。具体为：如果利用“离散极大值法”未在差谱中的某伴随区间内识别出极大值点，则直接将对应候选吸收峰排除；否则如果在该区间内识别出极大值点，但在差谱中的峰值强度低于最大吸收峰峰值强度的 5% 或者在原始吸收谱中对应的吸收系数在 20% 分位数以下，则认为其吸收强度过低，也对应候选吸收峰排除；其余识别出的极大值点均可认定为原始吸收谱的吸收峰位。

2 结果与讨论

2.1 极大值法与伴随拐点法的比较分析

利用伴随拐点法对一种易被滥用的硝基咪唑类药物呋喃妥因进行了特征吸收峰的识别，结果如图 2 所示。为验证伴随拐点法的有效性，同样采用离散极大值法对该样品进行吸收峰位识别，并比较分析。黑色曲线为呋喃妥因在 THz 波段的原始吸收系数谱图，蓝色曲线为原始吸收系数谱图的一阶导数，红色曲线为对应二阶导数。

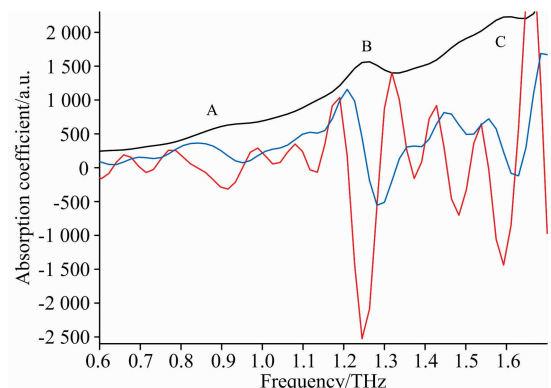


图 2 呋喃妥因的吸收系数曲线及一、二阶导数曲线
Fig. 2 Absorption coefficient spectrum of Furantoin and its corresponding first and second derivative function

从图中观察到，该物质至少存在三个吸收峰，峰位位于 0.90, 1.25 和 1.60 THz 附近，其中 1.25 THz 附近为强吸收峰(图 2 中 B 点)，而 0.90 和 1.60 THz 处的吸收峰较弱(图 2 中 A 和 C 点)。对于强吸收峰 B，在 1.25 THz 附近的每个一阶导数曲线均随频率的增大由正值转为负值，即 1.25 THz 附近恰为所有的一阶导数曲线的“过零点”。同时 1.25 THz 附近所有的二阶导数的取值均小于零，符合离散极大值法判别吸收峰的条件。同样，对于弱吸收峰 C，其导数特征也符合离散极大值法的评判指标，能够被离散极大值法识别到。但是，对于弱吸收峰 A，其峰位 0.90 THz 附近，虽然二

阶导数均小于零,但在 0.90 THz 附近的一阶导数均大于零,因此,离散极大值法无法识别到该吸收峰。

由此可见,离散极大值法用于判定强吸收峰十分有效,但用于弱吸收峰的判定时,有效性低,所以在使用此法寻找吸收峰位时有可能错过一些特征吸收峰,并不利于物质的识别和鉴定。而伴随拐点法恰好能满足这些需求。

呋喃妥因的吸收系数谱图及伴随拐点法的峰位识别结果如图 3 所示,其中蓝色圆点和红色圆点分别为用于构造伴随区间的起势拐点和落势拐点。图 3 中伴随区间的数量达到 7 个,对应的起势拐点和落势拐点的数量达到 14 个。但是根据差分光谱的谱峰识别,最终识别到 4 个吸收峰,分别为 A, B, C 和 D。值得注意的是,伴随拐点法不仅能识别弱吸收峰 A,而且能识别到另一个肉眼无法识别的弱吸收峰 D。伴随拐点法计算得到的差谱中有两个明显的峰和两个相对较弱的峰。因此,根据差分光谱,两个强吸收峰分别对应 B 和 C,两个相对较弱的吸收峰分别对应 A 和 D。差分光谱中的峰位代表了利用伴随拐点法识别到的呋喃妥因的吸收峰位。另外,差分光谱中 B 峰的强度最强,其次是吸收峰 C,这与离散极大值法得到的结果一致。因此,伴随拐点法不管是对强吸收峰还是对弱吸收峰的识别,均十分有效。

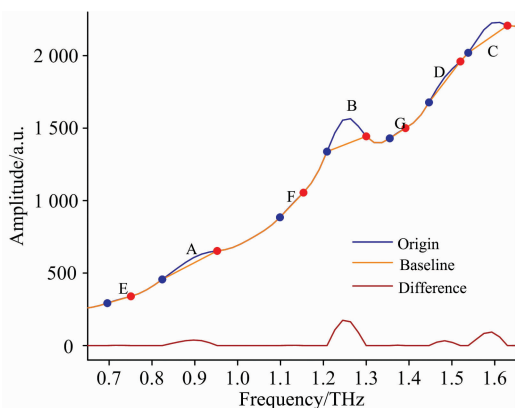


图 3 呋喃妥因的吸收系数谱图及伴随拐点法的峰位识别结果

Fig. 3 Result of absorption peak discrimination of nitrofurantoin using AIP

而对于候选弱吸收峰 E, F 和 G 而言,其强度在差谱中不清晰。根据峰位判别规则,利用离散极大值法未在差谱中识别出 E 和 F 点,则直接将其排除。同时,由于 G 峰值在差谱中强度低于吸收峰 B 的峰值强度的 5%,也排除吸收峰 G,其余识别出的极大值点(A, B, C, D)均可认定为原始吸收谱的吸收峰位。

2.2 基于密度泛函理论的呋喃妥因吸收峰的理论计算

为验证伴随拐点法对弱吸收峰的识别能力,将吸收峰识别结果与仿真结果进行对比分析。本研究利用密度泛函理论针对呋喃妥因的吸收峰振动模式进行了理论解析。输入呋喃妥因分子的三维结构文件,在 Chem3D 软件中采用半经验分子轨道 PM3 方法,获得呋喃妥因分子的初始最小能量结构,然后在 Gaussian 量子化学软件中采用密度泛函理论中的

B3LYP 方法,选取 6-311G 基组水平对此初始分子结构进行了几何优化和频率计算。计算结果没有虚频出现,说明几何优化计算得到了分子的最小能量结果。

图 4 为呋喃妥因在 0.2~1.8 THz 范围内的实验光谱与利用密度泛函理论计算的理论光谱对照图,蓝色曲线和红色曲线分别表示呋喃妥因的实验光谱和理论光谱。理论计算结果显示,呋喃妥因在 0.2~1.8 THz 范围内共有三个吸收峰,分别位于 0.89, 1.30 和 1.67 THz 处。由于理论光谱中的吸收系数与实验光谱不在一个数量级上,为便于观察,理论光谱中的各纵坐标采用了同时放大为 1 000 倍后的数据,其中放大结果不影响吸收峰位的指认。另外,理论光谱中 1.30 THz 处的吸收峰的峰值与其余两个吸收峰的峰值相比,相对较小,无法在图 4 中观察到,但我们仍在图 4 中做了标记,方便对比。

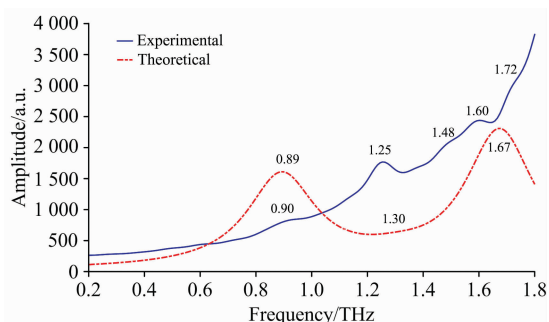


图 4 呋喃妥因的理论吸收谱与实验吸收谱对照图
Fig. 4 Comparison of theoretical and experimental absorption spectrum of furantoin

呋喃妥因的太赫兹吸收系数光谱中利用伴随拐点法识别出的在 1.25 THz 处的强吸收峰,和在 0.90 和 1.60 THz 处的两个弱吸收峰均有对应的理论计算结果,分别对应于理论光谱中 1.30, 0.89 和 1.67 THz 处的吸收峰。而其中实验光谱 0.90 THz 处的弱吸收峰是无法利用离散极大值法识别出来的,理论光谱中 0.89 THz 的吸收峰进一步验证了利用伴随拐点法识别实验光谱中吸收峰的合理性。不过实验光谱中利用伴随拐点法识别出的在 1.48 和 1.72 THz 处的弱吸收峰未在理论光谱中找到对应的吸收峰,这与理论计算模拟软件所采用分子模拟条件有关。

2.3 伴随拐点法对不同样品吸收峰的识别结果分析

为进一步验证伴随拐点法的有效性,利用该方法对呋喃妥因、呋喃西林、呋喃他酮和呋喃唑酮的吸收峰位也进行识别,结果如图 5 所示。每个样本吸收系数谱图均有多对伴随拐点,其中蓝点是起势拐点,红点表示落势拐点。以呋喃妥因为例,共识别出 10 对拐点对,但是最终识别出强度相对较大的 5 个吸收峰,分别为 1.25 THz 处的强吸收峰,0.90, 1.48, 1.60 和 1.72 THz 处的四个弱吸收峰。其中 0.90, 1.48 和 1.72 THz 处的弱吸收峰采用离散极大值法均无法识别。从结果中观察到,结合伴随拐点法的吸收峰位识别规则,能够有效避免过敏感现象,实现含峰目标物吸收峰位的精准识别。

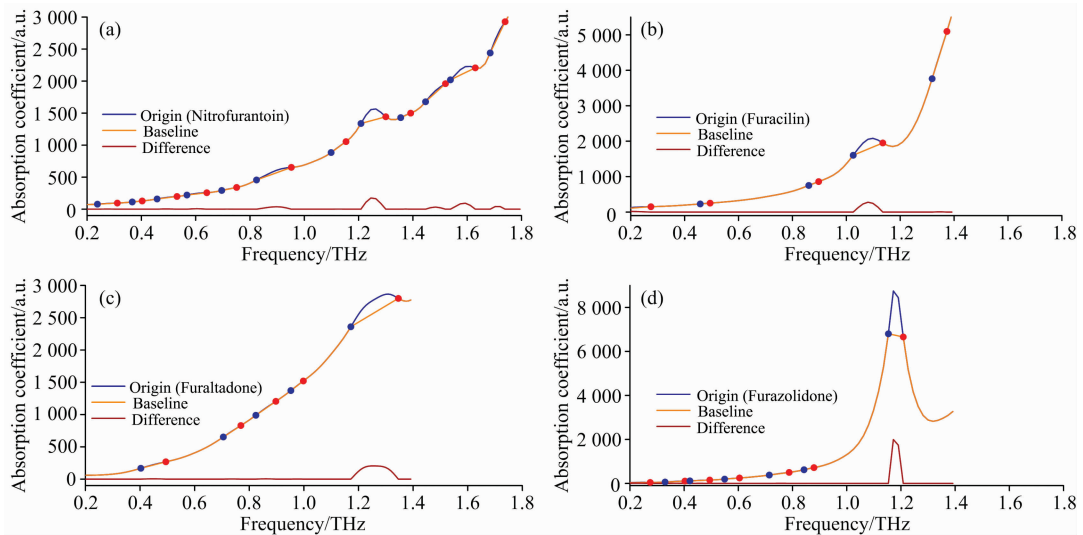


图 5 硝基呋喃类药物吸收系数的伴随拐点法结果

(a): 呋喃妥因; (b): 呋喃西林; (c): 呋喃他酮; (d): 呋喃唑酮

Fig. 5 Absorption peak discrimination results of four kinds of samples using AIP

(a): Nitrofurantoin; (b): Furacilin; (c): Furaltadone; (d): Furazolidone

3 结 论

提出了一种基于离散极大值法的含峰目标物的峰位检测方法伴随拐点法,并将其应用于识别和提取硝基呋喃类化合物的太赫兹吸收峰位。同时比较了离散极大值法和伴随拐点法提取的样品峰位信息,观察到伴随拐点法能够有效识别离

散极大值法无法识别的弱吸收峰,但同时具有较好的避免过敏感的能力。另外,将密度泛函理论计算得到的理论光谱与太赫兹谱图的识别结果进行比较,进一步验证了伴随拐点法对位于 0.89 THz 的弱吸收峰位识别的有效性。该结果表明,伴随拐点法在光谱谱图解析和物质检测等领域具有潜在的应用前景。

References

- [1] Afsah-Hejri L, Hajeb P, et al. *Comprehensive Reviews in Food Science and Food Safety*, 2019, 18(5): 1563.
- [2] Qin J Y, Xie L J, Ying Y B. *Analytical Chemistry*, 2014, 86(23): 11750.
- [3] Massaouti M, Daskalaki C, et al. *Applied Spectroscopy*, 2013, 67(11): 1264.
- [4] Yin M, Tang S F, Tong M M. *Analytical Methods*, 2016, 8(13): 2794.
- [5] Qin B, Li Z, et al. *IEEE Transactions on Terahertz Science and Technology*, 2018, 8(2): 149.
- [6] Shen Y C. *International Journal of Pharmaceutics*, 2011, 417(1-2): 48.
- [7] Takeuchi I, Tomoda K, et al. *Journal of Pharmaceutical Sciences*, 2012, 101(9): 3465.
- [8] Siegel P H. *IEEE Transactions on Microwave Theory and Techniques*, 2004, 52(10): 2438.
- [9] Qin J, Xie L, Ying Y. *Food Chemistry*, 2017, 224: 262.
- [10] Song M, Yang F, Liu L, et al. *Spectrochim Acta A Mol Biomol Spectrosc*, 2018, 191: 125.
- [11] Sy S, Huang S, et al. *Physics in Medicine and Biology*, 2010, 55(24): 7587.
- [12] Zaytsev K I, Kudrin K G, et al. *Applied Physics Letters*, 2015, 106(5): 053702.
- [13] Shi L Y, Shumyatsky P, et al. *Journal of Biomedical Optics*, 2016, 21(1): 015014.
- [14] Smolyanskaya O A, Chernomyrdin N V, et al. *Progress in Quantum Electronics*, 2018, 62: 1.
- [15] Kistenev Y V, Borisov A V, et al. *Journal of Biomedical Optics*, 2018, 23(4): 045001.
- [16] Sun C K, Chen H Y, et al. *Scientific Reports*, 2018, 8: 3948.
- [17] Hua Y F, Zhang H J. *IEEE Transactions on Microwave Theory and Techniques*, 2010, 58(7): 2064.
- [18] Baek S H, Lim H B, Chun H S. *Journal of Agricultural and Food Chemistry*, 2014, 62(24): 5403.
- [19] Cao B, Li H, Fan M, et al. *Analytical Methods*, 2018, 10(42): 5097.
- [20] Sun F, Fan M, Cao B, et al. *IEEE Transactions on Industrial Informatics*, Early Access.
- [21] Cao Y, Huang P, et al. *Biomedical Optics Express*, 2020, 11(2): 982.

Research on Discrimination Method of Absorption Peak in Terahertz Regime

CAO Yu-qi², KANG Xu-sheng^{1, 2*}, CHEN Piao-yun², XIE Chen², YU Jie^{2*}, HUANG Ping-jie², HOU Di-bo², ZHANG Guang-xin²

1. School of Computer and Computer Science, Zhejiang University City College, Hangzhou 310015, China

2. State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

Abstract Terahertz radiation bridges the gap between the microwave and optical regimes. It has unique properties such as fingerprint characteristics, non-destructive testing and transparency to various materials, which makes terahertz waves have significant scientific and technological potential in drug testing applications. Terahertz time-domain spectroscopy plays an important role in identifying target drugs containing absorption peaks. It can be used to discriminate specific molecules contained in drugs or the changes of components in samples, as many molecules have characteristic absorption peaks in the terahertz regime. Thus, to solve the problems of identifying weak absorption peaks of low content targets in the mixture, in this paper, the adjoint inflection point (AIP) method based on the discrete local maximum (DLM) method is proposed for identifying the characteristic absorption peaks in terahertz regime for the effective identification of the low content target. Firstly, the adjoint inflection points of potential absorption peaks are obtained using the first and second derivative of the terahertz absorption coefficient spectrum. Secondly, the difference spectrum is calculated by performing the operation between the original absorption spectrum and the baseline spectrum. At last, the absorption peak positions are determined by using the DLM method along the difference spectrum. Also, this AIP method is applied to the absorption peak extraction of four nitrofurans sample spectra. The result is compared with the peak positions determined by DLM, and the peak positions are also compared with the peaks calculated by the density functional theory. The better performance of the recognition capacity of the AIP method is observed and verified, especially for weak absorption peaks. This method suggests that it has profound application potential in spectroscopic analyses and even in determining curve peaks in various applications.

Keywords Terahertz absorption coefficient spectrum; Absorption peak recognition; Discrete local maximum method; Adjoint inflection point method

(Received Aug. 5, 2021; accepted Oct. 22, 2021)

* Corresponding authors