

# 基于遗传算法和模拟退火算法的振动光谱成分分析算法

黄凡, 张旭坤, 孙陆\*, 刘伟伟

南开大学现代光学研究所, 天津市微尺度光学信息技术科学重点实验室, 天津 300071

**摘要** 在大气污染物中有害气体成分众多, 运用太赫兹、红外等振动光谱检测所得不同成分的吸收峰较难辨别。设计了一种根据所测光谱中可能包含的物质的理论光谱分析所测光谱中不同物质所占比重的算法。利用实验峰值与不同理论峰值的对应关系构建超定方程组, 分别基于遗传算法和模拟退火算法进行求解, 得到实验光谱中各物质理论光谱的贡献。为了验证所提算法的准确性, 利用该算法对已知水团簇组成的水汽系统的实验光谱进行了分析, 结果表明: 所提算法能够得到不同成分的比重, 具有较强的实用性。

**关键词** 光谱学; 振动光谱; 成分分析; 超定方程组; 遗传算法; 模拟退火算法

中图分类号 O433.4

文献标志码 A

doi: 10.3788/LOP57.093001

## Vibration Spectral Component Analysis Based on Genetic Algorithm and Simulated Annealing Algorithm

Huang Fan, Zhang Xukun, Sun Lu\*, Liu Weiwei

Tianjin Key Laboratory of Micro-Scale Optical Information Science and Technology, Institute of Modern Optics, Nankai University, Tianjin 300071, China

**Abstract** In the atmospheric pollutants, there are many harmful gas components that is difficult to distinguish the absorption peaks of different components obtained by infrared and terahertz spectral detection. This paper designs an algorithm for analyzing the proportion of different substances in the measured spectrum based on the theoretical spectrum of the substances that may be included in the measured spectrum. The overdetermined equations constructed by using the corresponding relationship between experimental peaks and different theoretical peaks are solved based on genetic algorithm and simulated annealing algorithm, so the contribution of theoretical spectra of various materials in the experimental spectrum is obtained. In order to verify the accuracy of the proposed algorithm, we use this algorithm to analyze the experimental spectra of the water vapor system composed of known water clusters. The results show that the proposed algorithm can get the specific gravity of different components better and has stronger practicality.

**Key words** spectroscopy; vibrational spectroscopy; component analysis; overdetermined equations; genetic algorithm; simulated annealing algorithm

**OCIS codes** 300.6390; 300.6495; 070.4790

## 1 引言

近年来,随着工业的发展,各种工业废气的排放使得大气污染愈加严重,大气污染对人体的危害是多方面的,主要表现为呼吸道疾病与生理机能障碍,

以及眼鼻等粘膜组织受到刺激而患病,是造成老年哮喘的慢性因素。因此,对大气污染成分进行检测和治理,已经成为人们日益关切的健康与安全问题。振动光谱作为一种高效且有特征性的检测手段成为越来越常用的大气污染检测手段,振动光谱位于红

收稿日期: 2019-08-29; 修回日期: 2019-09-03; 录用日期: 2019-09-10

基金项目: 国家重点研发计划(2018YFB0504400)、天津市自然科学基金(16JCQNJC05100)

\* E-mail: lusun@nankai.edu.cn

外区,可用于表征不同气体分子的特征振动,振动光谱的特征吸收峰数量较多且在对大气污染气体进行振动光谱分析时,气体自身复杂的成分以及各成分之间的质量差异使得表现在振动光谱中不同物质的吸收峰在位置与强度上存在重叠、交错等情况,难以鉴别不同气体成分及其占比。因此对大气污染气体不同成分的分析已经成为大气污染检测的重要环节。目前,崔方晓等<sup>[1]</sup>提出了基于 Lasso 方法的污染气体自适应算法,通过仿真不同大气状态下的背景光谱特征集,利用 Lasso 算法进行特征筛选和重构,实现了对目标光谱的提取。由于 Lasso 算法需要覆盖尽可能多的测量条件,算法数据集具有很大的先验性。Harig 等<sup>[2]</sup>对大气光谱进行模拟,并采用最小二乘法扣除背景,提取出目标光谱特征。由于模拟是针对已设定的仿真环境参数,其结果存在一定的偏差。这些方法都只针对提取单一气体成分的光谱,而无法从整体上对污染气体成分进行定性和定量的分析。

本文基于遗传算法和模拟退火算法设计了一种能够分析气体中不同成分占比的算法。遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种通过模拟自然进化过程搜索最优解的方法,具有算法独立于求解域,能够求出优化问题的全局最优解等优点<sup>[3]</sup>。模拟退火算法来源于固体退火原理,是一种基于概率的算法,具有局部搜索能力强,运行时间短的特点<sup>[4]</sup>。本文将理论中可能存在的不同成分的峰与实验气体光谱的峰值一一对应,从而将实验图改写为几幅理论图的线性叠加。经过峰的匹配过程后,与实验峰有关的理论峰数据转换入一个  $m$  行  $n$  列的矩阵与之对应,实验峰强度也储存在一个  $m$  行向量中,进而得到超定方程组<sup>[5]</sup>。分别利用遗传算法和模拟退火算法搜索最优解,得到不同气体成分的占比。利用已知水团簇组成的水汽系统的太赫兹光谱对算法的可行性进行验证,并比较遗传算法和模拟退火算法所得结果的差别。结果发现,两种算法所得结果都能够与已知的成分组成较好地符合,其中遗传算法相比模拟退火算法耗时更长,但输出结果更为稳定;而模拟退火算法虽然运行较快,但容易出现陷入局部最优解的情况,需要多次运行取一次最好结果。本文所设计的算法可通过气体中不同成分的理论光谱对气体实验光谱中各种目标成分的光谱进行提取,具有算法时间复杂度低、实用性强等特点,在振动光谱检测大气污染成分方面有着潜在的应用前景。

## 2 基本原理

### 2.1 遗传算法的基本原理

遗传算法模拟了自然界生物进化的交叉、变异、选择等遗传过程,是一种新型的全局优化搜索算法。遗传算法的操作对象是一个可行解的集合。该集合中的元素经遗传和进化操作,按照适应度的不同,有不同概率保留至下一代<sup>[6]</sup>。每代集合中有一个最优的个体,它对应的表现型为这一代中的最优解;随着数代的演化,每代的最优个体也在进化,最终产生无限接近问题最优解的解。在遗传算法中,需要对不同个体进行“编码”,即生成它们的“基因型”,再对各“基因型”进行操作。算法流程如下:

1) 初始化。产生初始种群,而后进行的一切遗传操作都针对这一种群。

2) 个体评价。根据个体表现型的不同决定其对环境的适应度,以此作为接下来选择的依据。

3) 选择运算。以适应度为依据制定合理的选择策略,适应度更好的个体应有更大几率保留。

4) 交叉运算。产生子代的主要手段,一个子代个体由亲代两两交叉产生。

5) 变异运算。为防止过早收敛或收敛于局部最优解,增加基因多样性,每个个体有一定的概率进行随机的变异。

### 2.2 模拟退火算法基本原理:

模拟退火算法是一种常用的概率演算法,根据固体退火原理,将固体加温至足够高,再使其自然冷却<sup>[7]</sup>。在这一过程中,固体内能降低,原先处于活跃的、无序状态的固体分子,慢慢趋于有序,最后达到平衡。整个过程足够缓慢,可认为每一温度下固体都处于平衡态<sup>[8]</sup>。由 Metropolis 准则,粒子处于温度  $T$  的概率为

$$P = \exp\left(\frac{-\Delta E}{kT}\right), \quad (1)$$

式中: $\Delta E$  为温度  $T$  发生微小改变时内能的增量; $k$  为玻尔兹曼常量。算法的流程如下:

1) 初始化。从可行解空间中选取一个解为初始解,虽然模拟退火法对初始解无要求,但为了简化过程、节约计算时间,一般选取已知可行解中的最优解作为初始解。

2) 产生新解。对初始解进行一个随机的扰动,作为新解。

3) 评估新解。计算新解与旧解的目标函数之差,若差值小于零,则接受新解,以新解替代旧解进

入下一循环;若大于零,则根据 Metropolis 准则决定是否保留新解。

### 3 算法的设计

成分复杂的物质的实验光谱图中吸收峰的判定需要同时考虑峰高和相邻吸收峰的间隔,设定峰高大于  $(Q-q)b$  的峰计入吸收峰(其中  $Q$  为光谱图中的峰高的最大值,  $q$  为最小值,  $b$  介于 0 与 1 之间),避免出现过小的、局部的峰。设定一个正整数  $c$ ,表示两峰间至少隔有  $c-1$  个点,可避免出现峰过密的情况。当实验光谱中存在十分密集的高度高于  $(M-m)b$  的峰时,设定  $h$  值,若两相邻峰之间所有数据起伏不超过  $h$ ,则只计单个峰;若超过,则两峰都保留。通过判定过程得到吸收峰的数量为  $m$ ,对各实验峰强度进行计算,每个峰可近似为一个三角形,峰的强度表示为半峰全宽乘以峰高。最后将得到的峰的强度存储在一个  $m \times 1$  的矩阵中  $[b_1 \cdots b_n]^T$ ,每个元素为依次得到的实验峰的强度。

峰的匹配是算法设计的关键部分,通过匹配生成一个直接联系实验与理论的矩阵,每个理论图是指在成分复杂物质中每个可能存在物质的理论上的光谱图。通过程序在实验峰的每一个位置处设定一个范围,在这个范围内寻找几幅理论图像在该处的峰,若第  $i$  个实验峰匹配到了与它对应的第  $j$  幅理论图的峰,则矩阵的  $i$  行  $j$  列元素为该理论峰的强度,未匹配到的位置为 0。经过峰的匹配生成一个

$$m \times n \text{ 的矩阵 } \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \text{ 其中 } m \text{ 为找到的实}$$

验峰的数目,  $n$  为给出的理论图的数目,一般有  $m > n$ 。在匹配过程中,峰的范围的设定需根据实验光谱的情况来制定,既不能过大,否则匹配到的峰无意义,也不能过小,否则难以找到对应。理论峰的强度计算方法和实验峰强度计算方法相同。

通过上述过程,得到所要求解的超定线性方程组<sup>[9]</sup>,超定方程组的表达式为

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad (2)$$

式中:  $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$  为系数矩阵;  $[x_1 \cdots x_n]^T$  为解向量,其中每个元素代表每个可能存在的物质的理论光谱在实验光谱中所占的比例,这也是每个物质

成分在成分复杂物质中所占的比例;  $[b_1 \cdots b_n]^T$  中的每个元素为对应的实验峰的强度。之后设计了遗传算法、模拟退火算法求出超定方程组的正归一化最小二乘解<sup>[10]</sup>。遗传算法的设计如下:

1) 初始化和编码。已得的系数矩阵为

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \text{ 理论上可从中得到 } C_m^n \text{ 个方程组,}$$

$$\text{每个方程组系数矩阵为 } n \times n \text{ 方阵 } \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}.$$

再对它们进行秩的判断,解出有解的方程组的解,选取各分量大于 0 的解作为初始种群。编码的目的在于产生个体的基因型,便于其后对种群进行各种遗传操作。在这里由于问题的特殊性,所得解向量本身就可作为基因型。

2) 选择算子设计。将解向量  $[x_1 \cdots x_n]^T$  本身确定为基因型,要进行选择过程首要确定个体表现型,而后根据表现型确定其适应度,最后以适应度为标准决定个体的存活概率。由于程序目的是得到原方程组的正归一化最小二乘解,则定义表现型为解向量符合期望的程度,其表达式为

$$D = \frac{|\mathbf{AX} - \mathbf{B}|}{\sum_i X_i}, \quad (3)$$

式中:  $\mathbf{A}$  为原系数矩阵;  $\mathbf{X}$  为解向量(个体基因型);  $\mathbf{B}$  为原方程组等号右边的列向量。  $D$  越小即方程组的解向量越符合期望,适应度  $F$  的表达式为

$$F = \max(D) - D, \quad (4)$$

式中:  $\max(D)$  为表现型中最不符合期望的个体的  $D$  值,越优良的个体适应度  $F$  越高。对已选定的适应度进行赌盘选择,决定个体的去留,完成选择过程。

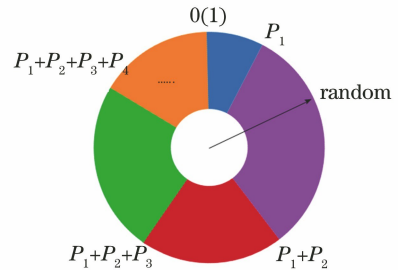


图 1 赌盘选择示意图

Fig. 1 Schematic diagram of the market selection

图 1 为赌盘选择示意图,其中  $P_n$  满足下式

$$P_n = \frac{F_n}{\sum_i F_i}. \quad (5)$$

每次进行选择时,由 0 到 1 间产生一个随机数,观察其落入图 1 的区域,落入哪个区域则选择哪个个体存活,此即为赌盘选择。适应度越高的个体,  $P_n$  值越高,在赌盘中所占的面积越大,被选中的概率越高。

3) 交叉算子的设计。由于亲代个体的向量分量和不同,交叉时不能用亲本 1 与亲本 2 之和的一半进行,否则可能会造成子代明显偏向某个亲本,造成基因多样性的丧失,而应该考虑两个亲本的权重,使它们在交叉过程中尽量均匀。于是采用了以分量和加权平均得子代的算法,设已知亲代解向量分别为  $[x_{10} \cdots x_{n0}]^T$ 、 $[x_{11} \cdots x_{n1}]^T$ ,子代向量为  $[x_{12} \cdots x_{n2}]^T$ ,子代向量中每一个元素  $x_{i2}$  ( $i=1, \dots, n$ ) 的表达式为

$$x_{i2} = \left[ \left( \sum_{j=1}^n x_{j1} \right) \cdot x_{i0} + \left( \sum_{k=1}^n x_{k0} \right) \cdot x_{i1} \right] / \left( \sum_{j=1}^n x_{j1} + \sum_{k=1}^n x_{k0} \right), \quad (6)$$

式中:  $\sum_{j=1}^n x_{j1}$ 、 $\sum_{k=1}^n x_{k0}$  为亲代向量所有元素之和,每个子代解向量的元素都为对应亲代解向量元素的加权平均。

4) 变异算子。在遗传算法中,为防止过早收敛或收敛于局部最优解,增加基因多样性,变异操作是必要的,在本程序选择算子的设计中,赌盘选择已通过生成随机数的形式对不同个体的不同位置的基因进行了随机的改变。

5) 精英策略。如前所述的选择过程和变异过程有其随机性,固然适应度大的个体更有可能保留,但总有优良个体消亡的可能,有效的防治方法是采用精英策略,即经过交叉、变异、选择后的个体和上一代中表现最为优良的个体共同组成子代,即将上一代适应度  $F$  值较高的解向量与经过交叉,变异,选择后得到的解向量共同组成解向量集,增大优良基因的比例。经过数代繁衍,最后得到最优解。

基于模拟退火算法进行求解的算法设计如下:

1) 初始化。模拟退火算法与遗传算法不同,它的操作对象是一个个体,对初始值的要求不高,但为了减少计算量,缩短计算用时,一般采用能得到的最优解作为初始值进行迭代。在遗传算法的计算过程中,已经得到了初始化的种群,选取种群中的最优个体作为模拟退火法的初始值。

2) 退火策略。首先选取较高的初始温度  $T$  和极低的末温度  $T_0$ ,再制定退火策略  $T(k+1) = \alpha T(k)$ ,

这些参数直接决定选择过程和迭代次数(其中  $\alpha \approx 0.9$ ,三个参数的具体值应根据运行结果选取)。

3) 新解的产生。对解向量  $\mathbf{X}$ ,新的解向量与旧的解向量之间的关系式为

$$\mathbf{X}_{\text{new}} = \mathbf{X}_{\text{old}} + \Delta \mathbf{X}, \quad (7)$$

式中:向量  $\Delta \mathbf{X}$  为随机产生的一个介于  $-\mathbf{X}/2$  与  $\mathbf{X}/2$  之间的随机向量; $\mathbf{X}_{\text{new}}$  为新的解向量; $\mathbf{X}_{\text{old}}$  为旧解。

4) 选择过程。对解向量  $\mathbf{X}$ ,根据(3)式,计算  $D$ ,即解向量符合期望的程度,若  $D_{\text{new}}$  小于  $D_{\text{old}}$ ,则保留解向量  $\mathbf{X}_{\text{new}}$ ;否则根据 Metropolis 准则,在  $[0, 1]$  之间生成一个随机数  $d$ ,概率  $P$  表达式为

$$P = \exp \left[ - \frac{D_{\text{new}} - D_{\text{old}}}{kT} \right]. \quad (8)$$

若概率  $P$  大于随机数  $d$ ,则接受新的解向量  $\mathbf{X}_{\text{new}}$ ,否则保留原解  $\mathbf{X}_{\text{old}}$ 。经过数次迭代后得到最优解。

## 4 基于水汽太赫兹光谱的算法验证

水团簇是一种由分立的水分子氢键构成的化学组装体或水分子簇,存在于冰和自由液态水中,最简单的构型是  $(\text{H}_2\text{O})_2$ 。对这些团簇物的研究有助于对水的一些异常特征的理解,如其反常的密度随温度的变化曲线、理论与实际熔沸点的差距。水团簇也涉及到了超分子的稳定性问题,在化学中被普遍认为是一个悬而未决的问题。尽管承认了水单分子的几何结构,但由于分立的水分子之间氢键的多样的组合形式,即使是极小的水蒸气(少于 10 个水分子)的构型也极为复杂。氢键可以在皮秒内断裂和形成。太赫兹光谱技术以探测寿命在皮秒范围内的振动为特征,因此能够有效的测量和分析一定湿度的水汽中的不同水团簇成分及其振动模式<sup>[11]</sup>。目前,Dai 等<sup>[12]</sup>通过太赫兹时域光谱技术结合量子力学计算分析了在不同湿度下水汽吸收强度未按比例增长现象中不同水团簇组成的影响。

现有已知水团簇组成比例的水汽系统  $S$  及其实际的太赫兹光谱,在该系统中水团簇摩尔分数为 10% 的  $(\text{H}_2\text{O})_4$ 、10% 的  $(\text{H}_2\text{O})_5$ 、10% 的  $(\text{H}_2\text{O})_6$ 、20% 的  $(\text{H}_2\text{O})_7$ 、30% 的  $(\text{H}_2\text{O})_8$ 、20% 的  $(\text{H}_2\text{O})_9$ 。已知 6 种不同水团簇的理论光谱如图 2 所示,图 2 (a) ~ (f) 分别为  $(\text{H}_2\text{O})_4$ 、 $(\text{H}_2\text{O})_5$ 、 $(\text{H}_2\text{O})_6$ 、 $(\text{H}_2\text{O})_7$ 、 $(\text{H}_2\text{O})_8$ 、 $(\text{H}_2\text{O})_9$  的理论太赫兹光谱,红点表示曲线的峰,分别通过遗传算法和模拟退火算法对水汽系统  $S$  的组成成分进行模拟和估算

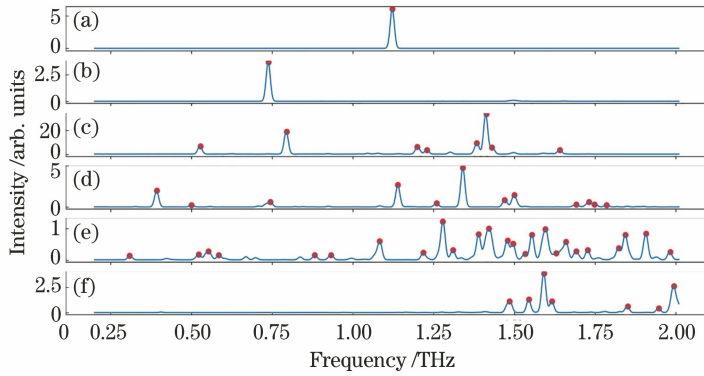


图 2 6 种水团簇的理论太赫兹光谱。(a)  $(\text{H}_2\text{O})_4$  理论太赫兹光谱;(b)  $(\text{H}_2\text{O})_5$  理论太赫兹光谱;(c)  $(\text{H}_2\text{O})_6$  理论太赫兹光谱;(d)  $(\text{H}_2\text{O})_7$  理论太赫兹光谱;(e)  $(\text{H}_2\text{O})_8$  理论太赫兹光谱;(f)  $(\text{H}_2\text{O})_9$  理论太赫兹光谱

Fig. 2 Simulated terahertz spectroscopy of six water clusters. (a)  $(\text{H}_2\text{O})_4$  theoretical terahertz spectroscopy; (b)  $(\text{H}_2\text{O})_5$  theoretical terahertz spectroscopy; (c)  $(\text{H}_2\text{O})_6$  theoretical terahertz spectroscopy; (d)  $(\text{H}_2\text{O})_7$  theoretical terahertz spectroscopy; (e)  $(\text{H}_2\text{O})_8$  theoretical terahertz spectroscopy; (f)  $(\text{H}_2\text{O})_9$  theoretical terahertz spectroscopy

并得到模拟光谱, 并比较它们与实际光谱图的误差, 以及遗传算法与模拟退火算法所得模拟光谱的差别, 从而检验程序的处理结果和比较遗传算法与模拟退火算法的优缺点。

图 3(a) 为遗传算法所得模拟光谱与实际光谱的对比图, 可以看出, 程序对峰的位置拟合较为准确, 但在峰值强度和比例组成上存在误差, 其中在强吸收峰 0.79 THz 和 1.41 THz 处, 模拟峰值强度较实际强度相差一半, 实际光谱中 1.60 THz 处的吸收峰在模拟光谱中并未出现, 模拟光谱在 1.64 THz 处相比实际光谱出现了吸收峰的分裂。图 3(b) 为遗传算法中  $D$  值的变化曲线, 由(2)式可知,  $D$  值反映了超定方程组的正的归一化最小二乘解符合期望的程度,  $D$  值越小越符合期望。结果表明, 种群中最优解的  $D$  值在前 10 代中急剧下降, 10 代到 30 代左右缓慢下降, 直到 40 代后趋于稳定, 此时  $D$  值稳定在 1.58, 最优解经过了 50 代演化, 该算法最终耗时约 285 s。

图 4(a) 为模拟退火算法所得模拟光谱与实际光谱的对比图, 与遗传算法结果相似, 但误差较遗传算法更大, 其中, 实际光谱 0.74 THz 和 0.39 THz 处的吸收峰在模拟光谱中没有出现, 模拟光谱 1.12 THz 处吸收峰的半峰全宽远小于实际光谱, 强吸收峰 0.79 THz 和 1.41 THz 处, 模拟峰值强度较实际强度相差一半, 在实际光谱 0.56 THz 吸收峰处, 模拟光谱出现吸收峰的分裂, 1.60 THz 和 1.64 THz 处模拟光谱与实际光谱的吸收峰交错。图 4(b) 为模拟退火算法的  $D$  值变化曲线, 由(2)式可知,  $D$  值反映了超定方程组的正的归一化最小二

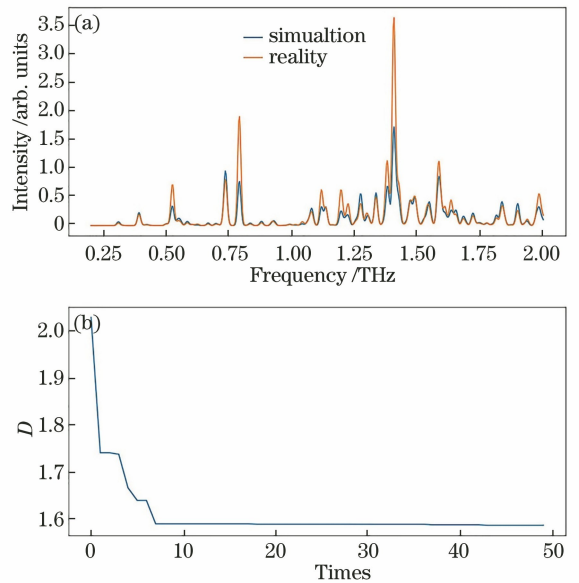


图 3 遗传算法运行结果。(a) 模拟-实际光谱强度对比图 (遗传算法); (b) 运行时遗传算法中最优解  $D$  值变化曲线

Fig. 3 Genetic algorithm operation results. (a) Simulation-actual spectral intensity comparison chart (genetic algorithm); (b) optimal solution  $D$  value curve in runtime genetic algorithm

乘解符合期望的程度,  $D$  值越小越符合期望, 设置初温  $T_0 = 109300100$ , 末温  $T = 0.01$ , 退火系数  $\alpha = 0.95$ , 经检测迭代了 449 次,  $D$  值达到 3.67, 耗时约 17 s。利用该算法时并不是每次计算过程中  $D$  都如图 4(b) 一样单调下降, 而是有一定概率会出现上升的情况, 因此程序中有一变量负责记录到目前循环为止出现过的最小值。

在测试过程中发现, 遗传算法相比模拟退火算

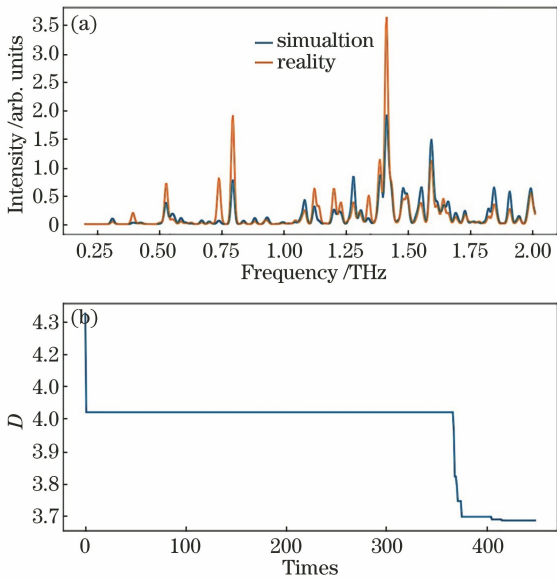


图4 模拟退火算法运行结果。(a) 模拟-实际光谱强度对比图(模拟退火算法); (b) 运行时模拟退火算法中最优解  $D$  值变化曲线

Fig. 4 Simulated annealing algorithm operation results. (a) Simulation-actual spectral intensity comparison chart (simulated annealing algorithm); (b) optimal solution  $D$  value variation curve in runtime simulated annealing algorithm

法耗时更长,但输出结果更为稳定,且得到的  $D$  值较模拟退火算法更低,超定方程组的正的归一化最小二乘解更符合期望。而模拟退火算法虽然快,但容易出现陷入局部最优解的情况,需要运行多次,取一次最好结果,且得到的最好结果的  $D$  值较遗传算法更高。由于两种算法自带的随机性,应对其中的参数进行多次调试,对变异算子、退火策略进行不同设计,对同一组数据进行多次运行后取其最优解。

## 5 结 论

针对大气污染成分的振动光谱进行了分析,提出了一种能够根据不同成分的理论光谱分析实际光谱中不同成分占比的算法。通过理论峰与实验峰的匹配,生成矩阵,得到超定方程组,利用遗传算法和模拟退火算法搜索最优解,得到不同成分对实验光谱的贡献,具有高效性。利用已知不同水团簇成分的水汽系统的太赫兹光谱对算法进行验证,所提算法对峰的位置拟合准确,可实现对气体振动光谱中不同成分的分析。比较遗传算法和模拟退火算法的优劣可知,遗传算法耗时更长,但输出结果稳定;模拟退火算法速度更快,但容易陷入局部最优解。对两种算法的各参数选择、变异、迭代方式进行更大范

围的试探,不断提高其程序精度及运行效率并引入神经网络算法,将是下一步的研究内容。

## 参 考 文 献

- [1] Cui F X, Li D C, Wu J, et al. Adaptive feature extraction algorithm based on lasso method for detecting polluted gas[J]. Acta Optica Sinica, 2019, 39(5): 406-414.  
崔方晓, 李大成, 吴军, 等. 基于 Lasso 方法的污染气体自适应探测算法[J]. 光学学报, 2019, 39(5): 406-414.
- [2] Harig R, Keens A, Rusch P, et al. Hyperspectral sensor for analysis of gases in the atmosphere (HYGAS) [C]//SPIE Defense, Security, and Sensing. Proc SPIE 7695, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI, Orlando, Florida, USA, 2010, 7695: 76950B.
- [3] Ge J K, Qiu Y H, Wu C M, et al. Summary of genetic algorithms research[J]. Application Research of Computers, 2008, 25(10): 2911-2916.  
葛继科, 邱玉辉, 吴春明, 等. 遗传算法研究综述[J]. 计算机应用研究, 2008, 25(10): 2911-2916.
- [4] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing [J]. Science, 1983, 220(4598): 671-680.
- [5] Duff I S, Reid J K. A comparison of some methods for the solution of sparse overdetermined systems of linear equations [J]. IMA Journal of Applied Mathematics, 1976, 17(3): 267-280.
- [6] Xu Z B, Gao Y. Characteristic Analysis and Prevention of Premature Convergence of Genetic Algorithms[J]. Science in China (Series E), 1996 (4): 364-375.  
徐宗本, 高勇. 遗传算法过早收敛现象的特征分析及其预防[J]. 中国科学(E辑), 1996(4): 364-375.
- [7] Liu F P. Simulated annealing algorithm based on seed strategy for cluster structure optimization [J]. Modern Computer, 2018(25): 23-26.  
刘凤萍. 改进的模拟退火算法用于团簇结构优化[J]. 现代计算机, 2018(25): 23-26.
- [8] Jia Z X, Duan H B, Shi Y H. Hybrid brain storm optimisation and simulated annealing algorithm for continuous optimisation problems [J]. International Journal of Bio-Inspired Computation, 2016, 8(2): 109-121.
- [9] Madsen K. An algorithm for minimax solution of overdetermined systems of non-linear equations [J].

- IMA Journal of Applied Mathematics, 1975, 16(3): 321-328.
- [10] Cadzow J A. Spectral estimation: an overdetermined rational model equation approach[J]. Proceedings of the IEEE, 1982, 70(9): 907-939.
- [11] Jepsen P U, Cooke D G, Koch M. Terahertz spectroscopy and imaging - Modern techniques and applications[J]. Laser & Photonics Reviews, 2011, 5(1): 124-166.
- [12] Dai Z J, Su Q, Lu D, et al. A combined experimental and theoretical study on the terahertz vibrations of water vapors[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 214: 277-284.