Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 206 (2018) 147–153

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Nonlinear least squares with local polynomial interpolation for quantitative analysis of IR spectra

Yifan Wu[a,b], Silong Peng[a,b,*], Qiong Xie[a], Pengcheng Xu[c]

[a] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] University of Chinese Academy of Sciences, Beijing 100190, China
[c] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China

ARTICLE INFO

Article history:
Received 8 April 2018
Received in revised form 23 July 2018
Accepted 3 August 2018
Available online 4 August 2018

Keywords:
Spectroscopy
Quantitative analysis
Nonlinear least squares
Local polynomial interpolation

ABSTRACT

When using spectroscopic instrumentation for quantitative analysis of mixture, spectral intensity nonlinearity and peak shift make it challenging for building calibration model. In this study, we investigated the performance of a nonlinear model, namely nonlinear least squares with local polynomial interpolation (NLSLPI). In NLSLPI, the parameters to be optimized are the concentrations of the components. Levenberg-Marquardt (L-M) method is used to solve the nonlinear-least-squares optimization problem and local polynomial interpolation is used to generate the nonlinear function for each component. We tested the robustness of NLSLPI on a computer-simulation dataset. We also compared NLSLPI, in terms of RMSEP, to partial least squares (PLS), classical least squares (CLS) and piecewise classical least squares (PCLS) on a real-world dataset. Experimental results demonstrate the effectiveness of the proposed method.

1. Introduction

Infrared (IR) spectroscopic analysis, whose advantages are no requirement for sample preparation, fast response and high accuracy, has been used for the quantitative analysis of mixture in wide applications [1-3]. In order to estimate the concentration of an interesting component from a mixture spectrum which usually contains hundreds of measurement values, chemometric algorithms are developed.

Usually, two situations will be met when dealing with IR spectroscopy. The first one is that only one or more components among the mixture are interested and most information (such as the reference spectrum of the interested component) are not obtainable. To deal with this case, a large amount of representative samples of mixture together with the concentrations of the interested component are needed. Inverse modeling methods such as principle component regression (PCR) [4] and partial least squares (PLS) [5] are usually applied to predict the concentrations of the future samples. The second case is that the mixture contains several major components and the reference spectra of these components are known. In this case, classical least squares (CLS) [6,7] is usually applied to simulate the spectrum of the mixture. Assuming that the mixture spectrum is a linear combination of reference spectra, CLS works by finding the concentrations that minimizes difference between the measured spectrum and the simulated spectrum. When the linear assumption failed, the piecewise classical least squares (PCLS) was used [8,9]. In PCLS, there are multiple reference spectra for each target component and then multiple CLS sub-models are generated. The two sub-models that predict concentrations in the measured spectrum nearest to those of specific reference spectra included in the set of sub-models are selected. The prediction value of PCLS is a weighted sum of the values of the selected sub-models. Even though CLS is widely used in commercial chemometric software, the investigation of its nonlinear extension is rare.

In this study, we focus on the second case with nonlinear assumption. The nonlinearity is presented in two aspects [10,11]. The first is a nonlinear relationship between spectral intensity and concentration. The second is peak shift with increasing concentration. In our point of view, CLS can still be used after some mathematical transformation when only the first nonlinearity presents. When the second nonlinearity or a combination of them present, nonlinear least squares with local polynomial interpolation (NLSLPI) is introduced to deal with quantitative analysis. Nonlinear least squares (NLS) has been proved to be a powerful tool for handling nonlinear systems [12,13]. In this study, NLS was employed to estimate the concentrations of the interested components. The Levenberg-Marquardt (L-M) method was used to solve the NLS problem. When using L-M, a difficulty is to estimate the function value together with

* Corresponding author.
E-mail address: silong.peng@ia.ac.cn (S. Peng).

https://doi.org/10.1016/j.saa.2018.08.002
1386-1425/

its derivative, which are unknown usually. In this study, local polynomial interpolation is used to generate the nonlinear function for each component.

In the following, we will discuss the nonlinear system and introduce NLSLPI in Section 2. Experimental data and results are presented in Sections 3 and 4 respectively. Finally, we make conclusions of NLSLPI in Section 5.

2. Methodology

2.1. Nonlinear Model of Quantitative Analysis

Given that the mixture is composed of J components, whose spectral responses are linear to the concentrations, CLS can be used to predict the concentration of each component from the spectrum of the mixture:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{argmin} \left\| \boldsymbol{s} - \sum_{j=1}^{J} x_j \quad \boldsymbol{f}_j \right\|_2^2 \tag{1}$$

where the mixture spectrum $\boldsymbol{s}$ is row vector of length p. x_j and $\boldsymbol{f}_j$ represent the concentration and the reference spectrum for component j. $\boldsymbol{x} = [x_1, x_2, \ldots, x_j]^\intercal$ is the concentration vector.

If the linearity condition can not be guaranteed, for example, the concentration falls outside the linear range, Eq. (1) should be modified:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{argmin} \left\| \boldsymbol{s} - \sum_{j=1}^{J} \quad \tilde{\boldsymbol{f}}_j(x_j) \right\|_2^2 \tag{2}$$

where the spectral response function $\tilde{\boldsymbol{f}}_j(x_j)$ can be any function of the concentration, not necessarily linear one. When spectral response function exhibits the first nonlinearity, we only need to use a nonlinear function $g_j(x_j)$ to represent the nonlinear relationship:

$$\tilde{\boldsymbol{f}}_j(x_j) = g_j(x_j) \quad \boldsymbol{f}_j \tag{3}$$

In practice, $g_j(x_j)$ can be generated by interpolating to the reference spectra with various concentrations of component j. Therefore, instead of estimating true concentration directly, we estimate pseudo-concentration first:

$$\hat{\boldsymbol{g}} = \underset{\boldsymbol{g}}{argmin} \left\| \boldsymbol{s} - \sum_{j=1}^{J} g_j \quad \boldsymbol{f}_j \right\|_2^2 \tag{4}$$

where $\boldsymbol{g} = [g_1, g_2, \ldots, g_J]^\intercal$ is the pseudo-concentration vector. Then the true concentration can be obtained by:

$$\hat{x}_j = g_j^{-1}\left(\hat{g}_j\right), j = 1, 2, \ldots, J \tag{5}$$

where g_j^{-1} is the inverse function of g_j.

When the mixture exhibits the combination of two nonlinearity, the above linear squares method fails. Then the nonlinear least squares method must be developed to predict the components of the mixture. In this study, Levenberg-Marquardt (L-M) method is employed.

2.2. Levenberg-Marquardt Method

L-M method is a classical method for solving nonlinear least squares, which is based on the trust-region framework [14]. In this study, L-M method is used for minimizing the following objective function:

$$l(\boldsymbol{x}) = \left\| \boldsymbol{s} - \boldsymbol{F}(\boldsymbol{x}) \right\|_2^2 \tag{6}$$

where $\boldsymbol{F}(\boldsymbol{x}) = \sum_{j=1}^{J} \tilde{\boldsymbol{f}}_j(x_j)$ is the sum of the pure component spectra.

L-M method performs optimization in a stepwise manner. Suppose that in iteration k, the estimation is $\boldsymbol{x}^k$, according to L-M method the solution for updating $\boldsymbol{x}^k$ can be calculated as following:

$$\boldsymbol{\delta} = \left(\boldsymbol{J}^{k\intercal}\boldsymbol{J}^k + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{J}^{k\intercal}\left(\boldsymbol{s} - \boldsymbol{F}^k\right) \tag{7}$$

where $\boldsymbol{\delta}$ is the solution for updating $\boldsymbol{x}^k$ and $\boldsymbol{F}^k = \boldsymbol{F}(\boldsymbol{x}^k)$. $\boldsymbol{J}(\boldsymbol{x}) = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}$ is first derivative matrix and $\boldsymbol{J}^k = \boldsymbol{J}(\boldsymbol{x}^k)$. λ is the regularization parameter and it is critical to the convergence speed of the optimization. When λ is large enough, $\left(\boldsymbol{J}^{k\intercal}\boldsymbol{J}^k + \lambda \boldsymbol{I}\right)^{-1} \approx 1/\lambda$ and $\boldsymbol{\delta}$ can be regarded yielded by Steepest Decent method. In this case, $\boldsymbol{\delta}$ guarantees the reduction of the objective function but with slow convergence speed. When $\lambda = 0$, $\boldsymbol{\delta}$ can be regarded yielded by Quasi Newton method. In this case, $\boldsymbol{J}^{k\intercal}\boldsymbol{J}^k$ is not necessarily invertible and does not guarantees the reduction of the objective function. However, when the reduction condition is fulfilled, Quasi Newton method converges to the minimum much faster than Steepest Decent method. In practice, λ usually decreases from a relatively large value and we check the reduction condition after each decrease. The smallest value which fulfills the reduction condition is selected as the final solution of λ.

After obtaining $\boldsymbol{\delta}$, we update $\boldsymbol{x}^k$ by:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{\delta} \tag{8}$$

2.3. Local Polynomial Interpolation

In order to implement L-M method in our study, $\boldsymbol{F}(\boldsymbol{x})$ and $\boldsymbol{J}(\boldsymbol{x})$ are obtained by interpolating to reference spectra with various concentrations. Polynomial interpolation is a common method for interpolation [15]. Suppose that $(n+1)$ reference spectra are available, then polynomial interpolation with degree of n can be obtained:

$$\tilde{\boldsymbol{f}}_j(x_j) = \boldsymbol{a}_{j0} + x_j\boldsymbol{a}_{j1} + x_j^2\boldsymbol{a}_{j2} \ldots + x_j^n\boldsymbol{a}_{jn}, \quad j = 1, 2, \ldots, J \tag{9}$$

where $\boldsymbol{a}_{j0}, \boldsymbol{a}_{j1}, \ldots, \boldsymbol{a}_{jn}$ are the polynomial coefficients for component j. The derivative of Eq. (9) is:

$$\frac{\partial \tilde{\boldsymbol{f}}_j}{\partial} x_j = \boldsymbol{a}_{j1} + 2x_j\boldsymbol{a}_{j2} + \ldots + nx_j^{n-1}\boldsymbol{a}_n, \quad j = 1, 2, \ldots, J \tag{10}$$

Therefore, we can obtain:

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{x}) &= \sum_{j=1}^{J} \quad \tilde{\boldsymbol{f}}_j\left(x_j\right) \\ \boldsymbol{J}(\boldsymbol{x}) &= \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}} \\ &= \left[\frac{\partial \tilde{\boldsymbol{f}}_1}{\partial x_1}, \frac{\partial \tilde{\boldsymbol{f}}_2}{\partial x_2}, \ldots, \frac{\partial \tilde{\boldsymbol{f}}_J}{\partial x_J}\right] \end{aligned} \tag{11}$$

Local polynomial interpolation (LPI) aims to find the closest data points to perform interpolation with low degree polynomial. For example, when three is set to be the polynomial degree, four pure component samples, $\left(x_j^{(1)}, \tilde{\boldsymbol{f}}_j^{(1)}\right), \left(x_j^{(2)}, \tilde{\boldsymbol{f}}_j^{(2)}\right), \left(x_j^{(3)}, \tilde{\boldsymbol{f}}_j^{(3)}\right), \left(x_j^{(4)}, \tilde{\boldsymbol{f}}_j^{(4)}\right)$, which satisfy the condition of $x_j^{(1)} \le x_j^{(2)} \le x_j \le x_j^{(3)} \le x_j^{(4)}$ are located, and then Eqs. (9) and (10) are employed. If x_j is close to the endpoints,

linear extension is used to generate virtual data points. The advantage of LPI is that it can model the local feature of the data well and avoid Runge's phenomenon [15] at the same time.

2.4. NLSLPI Algorithm

The framework of NLSLPI algorithm is described in the following. Given the spectra of pure components with various concentrations and the measured spectrum of the mixture $\boldsymbol{s}$, the goal is to estimate the concentrations of each component $\boldsymbol{x} = [x_1, x_2, \ldots, x_J]^\intercal$. Set the initial value of the regularization parameter λ_0 and the stopping tolerance of L-M iteration e. Step 1: Initiate the value of the solution $\boldsymbol{x}^0$; Step 2: In the k-th iteration, use local polynomial interpolation to calculate $\boldsymbol{F}^k$ and $\boldsymbol{J}^k$; Step 3: Calculate $\boldsymbol{\delta}$ according to according to Eq. (7). When λ decreases from λ_0 to 0, the smallest value which fulfills the condition of reduction of objective function is selected. Step 4: Update the estimated concentrations by $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{\delta}$; Step 5: Set $k = k + 1$ and return to Step 2 until $\|\boldsymbol{\delta}\| < e$ satisfies.

3. Experimental

3.1. Dataset

In this study, two datasets, one is the simulation data and the other is gas mixture data, are applied to validate the proposed method. The simulation data is applied to demonstrate the robustness of the proposed method while the gas mixture data is applied to demonstrate the advantage of the proposed method for a real dataset.

3.1.1. Simulation Data

There are four components in the simulation study, whose spectra are generated by the following equation:

$$\boldsymbol{f}(x_j) = a_j(x_j)e^{-\frac{(\boldsymbol{t}-\mu_j(x_j))^2}{\sigma_j^2}} \tag{12}$$

where x_j is the concentration of component j, $\boldsymbol{t}$ represents the range of spectra. $a_j(x_j)$ is a nonlinear function of x_j, which leads to nonlinear spectral intensity. $\mu_j(x_j)$ is a linear function of x_j, which leads to peak shift. σ_j represents the width of spectral peak. For the parameters setting, we set $\boldsymbol{t} = 1:200$; x_j increases from 0 to 1 with the interval of 0.1 and thus there are eleven reference spectra for each component. The details are shown in Table 1.

The mixture spectra are composed of all combinations of the pure spectra of concentrations of 0.25, 0.55 and 0.85 of each component. Therefore, there are 81 mixture spectra in the test set. We add Gaussian noise with different level to the mixture spectra. The influence of spectral overlap and spectral intensity to the robustness of NLSLPI is studied.

3.1.2. Gas Mixture Data

In this study, the gas samples were prepared and the spectra were measured in the laboratory. There were seven gases involved in this experiment, which were methane, ethane, propane, iso-butane, normal butane, iso-pentane and normal pentane. Pure gases with standard concentrations (100%, 10%, 5%, 1%) were blended with nitrogen to generate the designed concentrations. The process of gas mixing was controlled by a multichannel automated gas divider (GAINWAY GW-5000).

The calibration set was composed of pure gases at different concentrations. To ensure that the calibration samples were representative, the concentration of each gas varied from zero to an empirical value determined by practical situation. For methane, we observed that the nonlinearity in the low concentration range was obvious. The spectral areas of three different wavelength intervals were calculated and the scatter plots of spectral area-concentration were drawn (Supporting Information, Fig. S1). Therefore, the density of samples in the low concentration range was higher than the density in the high concentration range. The median of the calibration concentrations of methane favored low concentration. For other component gases, the responses of the spectra were linear. Therefore, the calibration concentrations were set to even-spacing. The concentrations of the pure gases are listed in Table 2. The test set was composed of nine samples of gas mixture and the concentrations of the component gases are listed in Table 3. The concentrations were carefully chosen so that the concentration ranges were inside of the calibration concentrations.

The mid-infrared (MIR) spectra of the gas samples, collected with a fourier transform infrared spectrometer (Bruker Alpha), were measured from 4000 to 1000 cm^{-1} with 64 scans at the resolution of 4 cm^{-1}. White cell with the pathlength of 2.4 m was used as measuring accessory. A temperature controller was used to keep the measuring temperature constant. Fig. 1 shows the process of gas mixing and FITR measurements.

3.2. Calculations

In our experiment, four quantitative models, PLS, CLS, PCLS and NLSLPI, were used for the determination of the gas mixture data. To generate the calibration set for PLS, additional gas mixture samples were prepared, from which the calibration spectra were collected. The simplest model (fewest number of latent variables) such that the predicted residual error sum of squares (PRESS) for this model was not significantly greater than the minimum PRESS was adopted. Therefore, a criterion [16] based on F-test was used to select the number of latent variables. In this work, the significance level of the F-test was set to 0.25 and PRESS was estimated for up to 10 factors by 5-fold cross-validation. For PCLS, since methane showed strong nonlinearity on the spectral response, eleven pure spectra were selected as reference spectra. For the other component gases, one pure spectrum was selected as reference spectrum for each component since they are linear systems. The concentrations of the reference spectra for PCLS are listed in Table 4. CLS was used as benchmark model. The reference spectra of CLS were the same as PCLS except that only one reference spectrum was used for methane, whose concentration was 0.497.

Root mean square error of prediction (RMSEP) was used as index of the quantitative performance for each model. RMSEP is calculated by:

$$RMSEP_j = \sqrt{\sum_{i=1}^{N}(x_{ij} - \hat{x}_{ij})^2}, j = 1, 2, \ldots, J \tag{13}$$

where x_{ij} and $\hat{x}_{ij}$ are the reference concentration and the predicted concentration of component j in sample i respectively. N is the number of mixture samples.

All computations were performed in Matlab 2015a (Mathworks, Inc., Natick, MA, USA)and run on a personal computer with 2.20-GHz Intel Core 2 processor, 4 GB RAM, and Windows 7 operating system. The programs were written in-house using Matlab language.

Table 1
The generation parameters of the reference spectra of different components in simulation study.

	a	x	μ	σ
Comp. 1	$log_{10}(9x + 1)$	0:0.1:1	$45 + 5x$	30
Comp. 2	$log_{10}(9x + 1)$	0:0.1:1	$60 + 5x$	30
Comp. 3	$log_{10}(9x + 1)$	0:0.1:1	$105 + 5x$	30
Comp. 4	$0.4log_{10}(9x + 1)$	0:0.1:1	$155 + 5x$	30

Table 2
The concentrations of the pure gases samples (unit: %).

	Methane	Ethane	Propane	Iso-butane	Normal butane	Iso-pentane	Normal pentane
Number of samples	16	12	21	21	20	16	16
Minimum	0	0	0	0	0	0	0
Median	0.849	0.3007	0.2374	0.2373	0.104	0.1716	0.1646
Maximum	4.3362	0.5739	0.4841	0.5025	0.1946	0.3842	0.3838

Table 3
The component concentrations of gas mixture samples (unit: %).

Sample number	Methane	Ethane	Propane	Iso-butane	Normal butane	Iso-pentane	Normal pentane
1	0	0.4058	0	0.02	0.0189	0.0157	0.0115
2	0	0	0	0.0124	0.0217	0.0117	0.0191
3	0	0.2281	0.019	0	0.0087	0.018	0.02
4	0.2874	0.3017	0.0172	0	0.0095	0.0087	0.0127
5	0.4616	0	0.0334	0.0063	0	0.0064	0.007
6	0.5794	0	0.0575	0.0119	0.0085	0.0082	0.0091
7	0.8364	0	0	0.0274	0.0193	0.0175	0.0292
8	1.2594	0.2694	0	0.0145	0	0.013	0.0141
9	1.5668	0.2652	0	0.0146	0.0145	0.0132	0.0142

4. Results and Discussion

4.1. Simulation Data

In the simulation settings, the pure spectra of each component contain both intensity nonlinearity and peak shift. The reference spectra dataset for NLSLPI are shown in Fig. 2. Among the four components, components 1 and 2 suffer from stronger spectral overlap than components 3 and 4. The spectral intensity of component 4 is lower than those of the other components. The distribution of the predictive errors of NLSLPI was analyzed under different noise levels. The polynomial degree of NLSLPI was set to three for the simulated data. Gaussian noise was added to the mixture spectra and the noise level was controlled by its standard deviation which denoted as σ_{noise}. The value of σ_{noise} increased from 0 to 0.030 with an interval of 0.005.

Firstly, RMSEP was used to present an overall description of the predictive errors (Fig. 3). It could be seen that component (component 3) suffered from weak spectral overlap was more robust to noise than components (components 1 and 2) suffered from strong spectral overlap. Moreover, component (component 3) with high spectral intensity was more robust than component (component 4) with low spectral intensity. Secondly, we analyzed the predictive errors by assigning them to different intervals. Under each noise level, we counted the numbers of samples whose predictive errors fell into the intervals of ±0.01, ±0.02, ±0.03, ±0.05 and ±0.07 respectively and calculated the ratios of the interval-samples to all samples. These intervals were called confidence interval and the ratios were called confidence level. Tables 5–8 show the calculation results of the confidence levels. We considered that the confidence interval was reliable when its confidence level was larger than 0.95. In comparison, the confidence intervals of component 3 were the most anti-noise. For instance, to achieve a reliable interval of ±0.01, the largest noise level of component 3 was 0.015, while it was smaller than 0.005 for component 1, and it was 0.005 for both components 2 and 4.

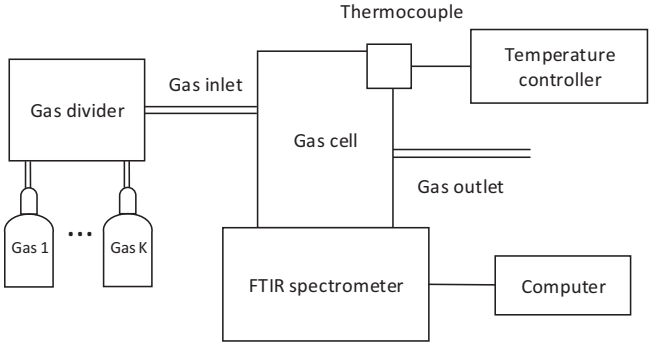

Fig. 1. The process of gas mixing and FTIR measurements.

4.2. Gas Mixture Data

Methane exhibited complex nonlinearity on the MIR spectra. In this study, the range of 3300–2400 cm^{-1} was selected for analyzing since all component gases had the strongest absorbance in this range. Fig. 4 (a) shows the spectra of pure methane in the selected range. Fig. 4 (b) shows that peak shift presents in the spectra of methane. Fig. 4 (c) shows the absorbance vs concentration at 3014 cm^{-1}, where intensity nonlinearity presents. It was well known that each wavenumber corresponded to a specific vibrational phenomena and did not depend on concentration. In this case, the phenomenon of peak shift might be due to the overlap of very similar energetic transitions. Specially in gas phase, we expected observable "peak shift" since the change in concentration might dramatically affect intermolecular distances and promote overlapping signals.

NLSLPI performs spectrum decomposition because it determinates a group of component spectra which minimize the difference between the sum of the estimated spectra and the spectrum of the mixture. The component spectra are obtained by interpolating the reference spectra with local polynomial. The polynomial degree

Table 4
The concentrations of the reference spectra for PCLS in gas mixture data.

Methane	Ethane	Propane	Iso-butane	Normal butane	Iso-pentane	Normal pentane
0.0935∼ 2.1504	0.329	0.057	0.0526	0.025	0.0266	0.0288

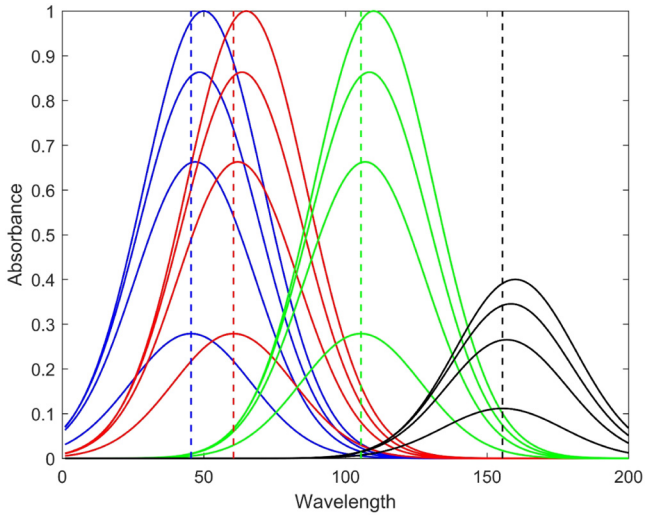

Fig. 2. Representative reference spectra in the simulation study. Blue, red, green and black curves represent the spectra of component 1, 2, 3 and 4 respectively. The dash line denotes the center of the peak at concentration 0.1 for each component. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
The estimated confidence levels of different confidence intervals (CI) under different noise levels (NL) for component 1.

NL	CI				
	±0.01	±0.02	±0.03	±0.05	±0.07
0.005	0.93	**1**	**1**	**1**	**1**
0.010	0.84	**0.95**	**0.99**	**1**	**1**
0.015	0.69	0.88	0.93	**0.98**	**0.99**
0.020	0.64	0.86	0.94	**0.98**	**0.99**
0.025	0.59	0.81	0.89	**0.96**	**0.96**
0.030	0.53	0.78	0.90	**0.98**	**0.99**

The bold values indicate that the confidence level is larger than 0.95.

Table 6
The estimated confidence levels of different confidence intervals (CI) under different noise levels (NL) for component 2.

NL	CI				
	±0.01	±0.02	±0.03	±0.05	±0.07
0.005	**1**	**1**	**1**	**1**	**1**
0.010	0.89	**1**	**1**	**1**	**1**
0.015	0.67	0.94	**1**	**1**	**1**
0.020	0.67	0.93	**0.98**	**1**	**1**
0.025	0.63	0.85	0.90	**0.99**	**1**
0.030	0.59	0.77	0.93	**0.99**	**1**

The bold values indicate that the confidence level is larger than 0.95.

of NLSLPI was set to one for the gas mixture data. Fig. 5 shows the result of spectrum decomposition for sample 4 of gas mixture. It can be seen that NLSLPI gave small residual of spectral fitting. The peak at 3016 cm^{-1} of the mixture spectrum was due to the spectrum of methane. Ethane made the largest contribution to the mixture spectrum while iso-butane and normal butane made little contribution.

Finally, we investigated the predictive results of NLSLPI of the gas mixture data. The results of CLS, PLS and PCLS were used for comparison. In this study, two combination rules were used for PCLS. One is the simple average rule and the other is the weighted average rule used in Ref. [8]. The RMSEP values for each component gas are listed in Table 9. For methane, the RMSEP of CLS was the highest, which was due to the highly nonlinearity of methane spectra and yet CLS used only one reference spectrum. PLS was a linear method but many researches found that it could deal with nonlinearity in some degree by using more latent variables. It could be seen that the RMSEP of PLS was much lower than CLS. The number of latent variables of methane was 9, which was larger than the number of components, 7. Details of the PLS could be referred to Supporting Information, Table S1. Both PCLS and NLSLPI used multiple reference spectra for modeling nonlinearity and they yielded considerable drop in RMSEP values when compared with CLS. Moreover, it could be seen that different combination rules lead to different predictive performance for PCLS. The advantage of NLSLPI over PCLS was that NLSLPI employed a single model and thus it did not need to consider the problem of sub-model combination. For the rest six component gases, the results of NLSLPI were better than or comparable to those of CLS, PLS and PCLS.

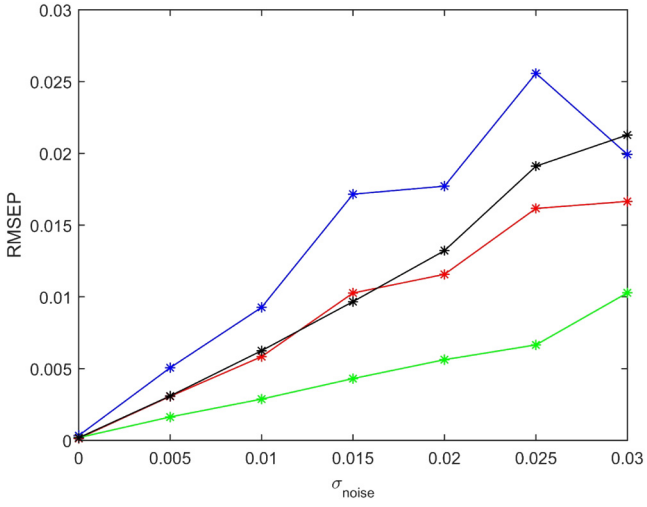

Fig. 3. The RMSEP of each component under different intensities of Gaussian noise. Blue, red, green and black lines represent components 1, 2, 3 and 4 respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
The estimated confidence levels of different confidence intervals (CI) under different noise levels (NL) for component 3.

NL	CI				
	±0.01	±0.02	±0.03	±0.05	±0.07
0.005	**1**	**1**	**1**	**1**	**1**
0.010	**1**	**1**	**1**	**1**	**1**
0.015	**0.95**	**1**	**1**	**1**	**1**
0.020	0.89	**1**	**1**	**1**	**1**
0.025	0.86	**1**	**1**	**1**	**1**
0.030	0.70	**0.95**	**0.99**	**1**	**1**

The bold values indicate that the confidence level is larger than 0.95.

Table 8
The estimated confidence levels of different confidence intervals (CI) under different noise levels (NL) for component 4.

NL	CI				
	±0.01	±0.02	±0.03	±0.05	±0.07
0.005	**1**	**1**	**1**	**1**	**1**
0.010	0.89	**0.99**	**1**	**1**	**1**
0.015	0.72	**0.95**	**1**	**1**	**1**
0.020	0.64	0.88	**0.95**	**1**	**1**
0.025	0.53	0.78	0.89	**0.98**	**0.99**
0.030	0.46	0.70	0.79	**0.98**	**1**

The bold values indicate that the confidence level is larger than 0.95.

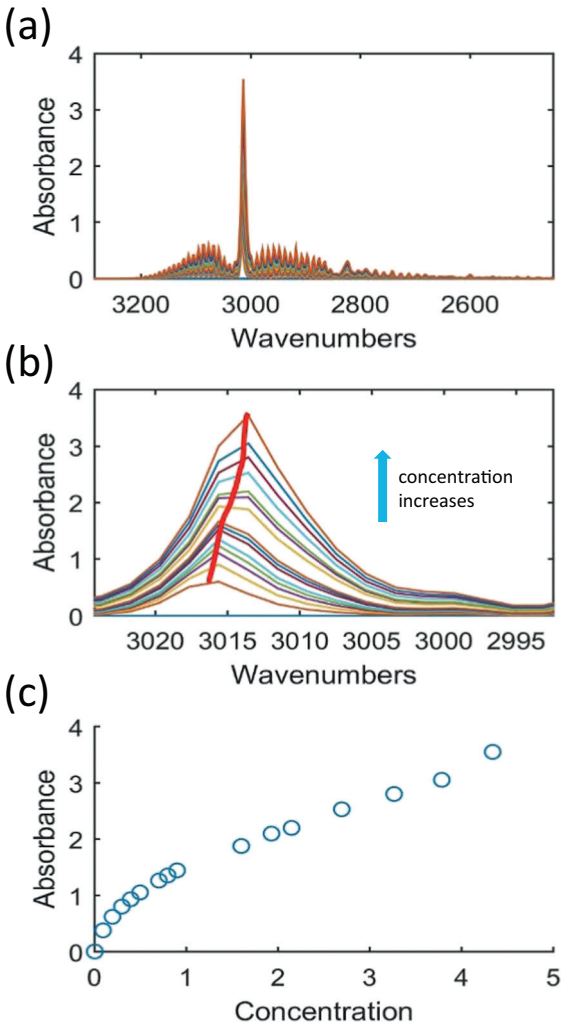

Fig. 4. Nonlinearity of methane spectra. (a) The spectra of pure methane in the region of 3300–2400 cm^{-1}; (b) the detail of the spectral peak, the red bold line represents the peak position with the increase of concentration; (c) absorbance vs concentration at 3014 cm^{-1}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

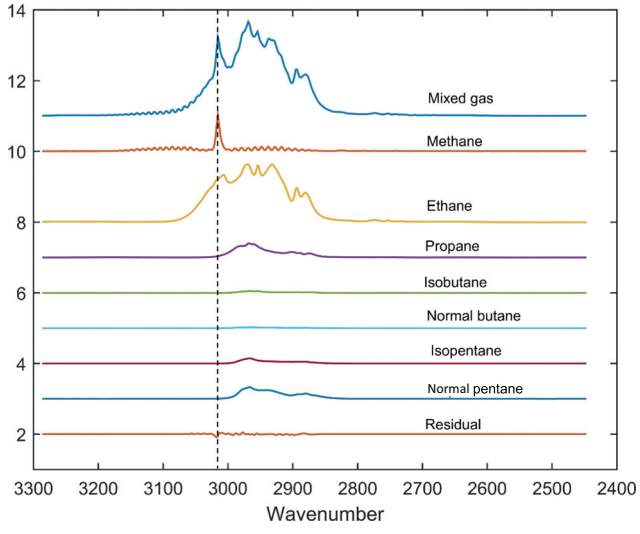

Fig. 5. Spectrum decomposition result of NLSLPI for sample 4 of gas mixture. Dash line denotes the position of 3016 cm^{-1}.

Table 9
The RMSEP of different quantitative methods for each component gas.

Method	Methane	Ethane	Propane	Iso-butane	Normal butane	Iso-pentane	Normal pentane
CLS	0.3115	0.0414	0.0095	0.0041	0.0134	0.0018	0.0051
PLS	0.0845	**0.0311**	0.0067	**0.0033**	0.0077	0.0034	0.0060
PCLS_sa[a]	0.0825	0.0357	0.0092	0.0038	0.0138	0.0018	0.0062
PCLS_wa[b]	0.1295	0.0335	0.0097	0.0039	0.0132	**0.0016**	0.0069
NLSLPI	**0.0416**	0.0354	**0.0053**	0.0037	**0.0070**	0.0020	**0.0048**

Bold indicates the smallest value.
[a] PCLS with the simple average combination rule.
[b] PCLS with the weighted average combination rule.

5. Conclusion

In this study, we introduced a nonlinear-least-squares model for quantitative analysis of the mixture. The goal of NLSLPI is to tackle with the problem that the component spectra show complex nonlinearity including intensity nonlinearity and peak shift. We proposed to use local polynomial interpolation to approximate such nonlinearity. The simulation study showed that the performance of NLSLPI was related to spectral overlap and spectral intensity. The gas mixture study showed that NLSLPI could give better performance than PLS, CLS and PCLS in practice. Therefore, NLSLPI can be considered as an alternative for modeling nonlinear system. Although NLSLPI was evaluated on MIR spectra, it can be applied to analyze other types of spectra data such as NIR spectra and Raman spectra. Some factors shall be considered to improve the performance of NLSLPI. For example, the components may react with each other when they are mixed; Baseline may exist when measuring spectra. Our future study will focus on extending NLSLPI to tackle with these factors.

Acknowledgments

Thanks are due to the Editor and reviewers for their valuable comments; to those anonymous scholars who offered advice on the improvement of the paper. This work is financially supported by the National Natural Science Foundation of China (Grant Nos.61571438 and 61601104).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2018.08.002.

References

[1] C.A.T. dos Santos, R.N.M.J. Pascoa, J.A. Lopes, A review on the application of vibrational spectroscopy in the wine industry: from soil to bottle, TrAC Trends Anal. Chem. 88 (2017) 100–118.
[2] J. Kiefer, Recent advances in the characterization of gaseous and liquid fuels by vibrational spectroscopy, Energies 8 (4) (2015) 3165–3197.
[3] L. Shao, P.R. Griffiths, A.B. Leytem, Advances in data processing for open-path Fourier transform infrared spectrometry of greenhouse gases, Anal. Chem. 82 (19) (2010) 8027–8033.
[4] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1-3) (1987) 37–52.
[5] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2) (2001) 109–130.
[6] M.K. Antoon, J.H. Koenig, J.L. Koenig, Least-squares curve-fitting of Fourier transform infrared spectra with applications to polymer systems, Appl. Spectrosc. 31 (6) (1977) 518–524.
[7] P.R. Griffiths, J.A. De Haseth, Fourier Transform Infrared Spectrometry, vol. 171. John Wiley and Sons. 2007.
[8] J.W. Childers, W.J. Phillips, E.L. Thompson, D.B. Harris, D.A. Kirchgessner, D.F. Natschke, M. Clayton, Comparison of an innovative nonlinear algorithm to classical least-squares for analyzing open-path Fourier transform infrared spectra collected at a concentrated swine production facility, Appl. Spectrosc. 56 (3) (2002) 325–336.

[9] N.N. Mahamuni, Y.G. Adewuyi, Fourier transform infrared spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions, petrodiesel? Biodiesel blends, and blend adulteration with soy oil, Energy Fuels 23 (7) (2009) 3773–3782.
[10] R.M. Balabin, E.I. Lomakina, Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data, Analyst 136 (8) (2011) 1703–1712.
[11] P. Beumers, D. Engel, T. Brands, H.J. Ko?, A. Bardow, Robust analysis of spectra with strong background signals by First-Derivative Indirect Hard Modeling (FD-IHM), Chemom. Intell. Lab. Syst. 172 (2018) 1–9.
[12] u. J. Witek, A.M. Turek, A novel algorithm for resolution of three-component mixtures of fluorophores by fluorescence quenching, Chemom. Intell. Lab. Syst. 160 (2017) 77–90.
[13] M.A. Maris, C.W. Brown, D.S. Lavery, Nonlinear multicomponent analysis by infrared spectrophotometry, Anal. Chem. 55 (1983) 1694–1703.
[14] J. Nocedal, S. Wright, Numerical Optimization, Springer Science and Business Media. 2006.
[15] K.E. Atkinson, An Introduction to Numerical Analysis, John Wiley and Sons. 2008.
[16] D.M. Haaland, E.V. Thomas, A. Chem, Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of quantitative information, Anal. Chem. 60 (1988) 1193–1202.