Identification of Different Dairy Products Using Raman Spectroscopy Combined with Fused Lasso Distributionally Robust Logistic Regression

Xiang Xu (✉ xuxiang2_2009@foxmail.com)  
Nanjing University of Finance and Economics

Wentao Xiao  
Nanjing University of Finance and Economics

Yiyun Cao  
Nanjing University of Finance and Economics

Zhengyong Zhang  
Nanjing University of Finance and Economics

Research Article

Keywords: Fused Lasso, Distributionally Robust Optimization, Raman Spectroscopy

Posted Date: September 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2013460/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Identification of Different Dairy Products Using Raman Spectroscopy Combined with Fused Lasso Distributionally Robust Logistic Regression

Xiang Xu*, Wentao Xiao, Yiyun Cao

School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing, PR China

Abstract: Raman spectroscopy has been more and more widely used in the quality detection of dairy products. Due to the characteristics of high dimension and small sample size, the rapid, intelligent and accurate identification method of Raman spectroscopy is becoming one of the hot issues for researchers. In order to improve the robustness and accuracy of logistic regression identification method, a new Raman spectroscopy identification method is proposed by combining distributionally robust optimization technique and fused lasso technique with logistic regression. Raman spectroscopy of two kinds of dairy products were collected respectively for anti-jamming identification testing to verify the effectiveness of the new method. Experimental results show that the proposed method is more robust and has higher recognition accuracy than the traditional logistic regression.

Key words: Fused Lasso, Distributionally Robust Optimization, Raman Spectroscopy

1. Introduction

As one of the important food sources for human beings, dairy products are consumed in huge quantities. In recent years, some outlaws sell seconds at best quality prices and wantonly counterfeit dairy products driven by economic interests, which has seriously endangered the life and health of consumers. How to efficiently identify and detect adulterated substances in milk powder has become a research hotspot in the field of dairy product safety, which is of vital significance under China's current national conditions. Among the various detection technologies, Raman spectroscopic imaging technology is one of the few technologies that can simultaneously make the rapid detection of micro-region analysis and large-area scanning possible, and has a great application potential in the analysis of adulterated substances of milk powder with its unique advantages of high resolution and high throughput. Raman spectroscopy is a scattering spectrum. Raman spectral analysis technique is based on the Raman scattering effect discovered by Indian scientist C.V. Raman. It analyzes the scattering spectrum different from the incident light frequency to obtain information on molecular

* Corresponding author
E-mail address: xuxiang2_2009@foxmail.com (X. Xu)
vibration and rotation, and it is applied to the study of molecular structure.

Figure 1 shows the Raman spectroscopy of two different substances, where the X-axis coordinate shows the Raman reflection frequency of two dairy products. Z-coordinate shows the Raman light intensity at different frequency bands. We can see that the differences in the Raman spectroscopy of F and I are mainly reflected in the stokes peaks, and the algorithm proposed in this paper is mainly based on the extraction of peak segment characteristics as a means of identifying different material components.

![Raman spectroscopy of two dairy products](image)

The Raman spectroscopy has many raw data features, but small sample size, and several distinct peaks, but a large amount of noise interference. Therefore, it is particularly important to pre-treat the spectrum, remove noise and preserve stokes peaks. In the analysis and application of Raman spectroscopy, due to the influence of noise, small sample size and sequence correlation of features, the existing methods face many challenges such as noise interference, low efficiency and difficulty in identification. In order to solve these problems, this paper proposed a fused lasso logistic regression model based on the distributionally robust optimization, which uses the robust optimization idea to weaken the influence of noise. The fused lasso method extracts the key feature terms and ensures the sequencing between the feature to highlight spectral peaks. Meanwhile, we designed an algorithm for the model, and the iterative process of the algorithm has a simple closed form, which is convenient to be applied in practice.

Dairy products have long faced challenges such as adulteration and counterfeiting, drug residues, contaminants, excessive use of food additives and illegal additives. Compared with the high cost and long cycle of laboratory testing methods, Raman spectroscopy detecting technique is non-destructive, fast and pollution-free. Raman spectroscopy was used to perform real-time detection of the quality of different types of cooking oil under frying conditions (Angeyo and Gari, 2022). Campmajó et al. (2022) classified and certified 70 servings of Spanish peppers by Raman spectroscopy. Rios-
Reina et al. (2017) distinguished Spanish Protected Designation of Origin wine vinegar for categories with Raman spectroscopy. The study by Ardila et al. (2017) used Raman spectroscopy to classify and identify Spanish gasoline. Ge et al. (2019) analyzed the exhaust soot of diesel engines and gasoline engines under different laser powers, and used Raman spectroscopy to classify and identify them. Haiyan Wang et al. (2017) identified five fat samples with the Raman spectroscopy by using the kernel principal components analysis nearest neighbor model.

Because the signal is subject to experimental testing conditions and baseline interference, there are interference factors such as noise, baseline drift, overlapping peak and cosmic ray in the analysis of Raman spectroscopy data. For these problems, some researchers have made a lot of efforts and research in recent years. Ehrentreich and Sümmchen (2001) made use of the correlation coefficients of wavelet of different levels to identify the noise position and completed the denoising of Raman spectroscopy. Angeyo and Gari (2022) conducted correlation analysis between the spectrum of seaweed samples and the traditional inductively coupled plasma spectrum. The fully cross-verified partial least squares regression method and nonlinear iterative partial least squares algorithm were used in their research. Lin et al. (2016) used principal component analysis combined with linear discriminant analysis for multivariate statistical analysis to distinguish Raman spectroscopy of different blood groups. However, most of Raman spectroscopy applications in dairy products are still in qualitative detection. There are still some interference factors in the actual quantitative analysis process, such as poor reproducibility, signal susceptibility to experimental conditions. Coitinho et al. (2017) combined infrared spectroscopy with stoichiometry to detect the quality parameters in milk powder. Fourier Transform Raman spectroscopy was explored as a fast and reliable screening method for assessing milk powder quality and identifying doped whey (Almeida et al., 2011, Mabood et al., 2017). Based on Swiss cheese maturation process, Vásquez et al. (2018) utilized the partial least squares regression and artificial neural networks to model the relationship between spectral profiles and hardness values.

With the development of machine learning technology, Chen et al. (2019) established a quantitative analysis method for polycyclic aromatic hydrocarbon surface enhanced Raman spectroscopy with principal component analysis dimensionality reduction and support vector machine algorithm. Li et al. (2017) proposed an adaptive genetic algorithm for point-by-point selecting mixed terahertz absorption spectral wavelengths, dynamically adjusting the crossover probability and mutation probability. Sevetlidis and Pavlidis (2019) used the extremely randomized trees model to accurately match the entire spectral range to their respective minerals. Liu et al. (2019) developed a dynamic spectrum matching method based on convolutional conjoined neural
networks. Sha et al. (2019) studied the uniformity of rice flours of four different particle sizes using relative standard deviation analysis of Raman spectroscopy and hierarchical cluster analysis. Barman et al. (2010) designed a robust correction model for blood glucose spectral monitoring with support vector machine. Most studies have shown that Raman spectroscopy has broad application prospects for rapid and non-destructive quality testing of dairy products. However, the advanced analysis methods of Raman spectroscopy in dairy products analysis applications still need to be further explored.

2. Spectrum Recognition Method Based on DRO

In standard statistical learning, it is assumed that all training and testing samples \((x, y) \in \Xi = \mathbb{X} \times \{0, 1\}\) follow the distribution \(\mathbb{P}\) independently, where \(x \in \mathbb{X} \subseteq \mathbb{R}^{1 \times m}\) represents feature vector, \(y \in \{0, 1\}\) is classification label. If the distribution \(\mathbb{P}\) is known, the undetermined parameter \(b \in \mathbb{R}^{m}\) of the logistic regression model can be obtained by solving the following random optimization problem.

\[
\inf_b \left\{ \mathbb{E}\left[ l_b(x,y) \right] = \int_{\mathbb{X} \times \{0,1\}} l_b(x,y) \mathbb{P}(d(x,y)) \right\},
\]

where the loss function \(l_b(x,y) = -yx + \log(1 + e^{xb})\).

However, distribution \(\mathbb{P}\) cannot be directly observed by training samples in practical application. Therefore, the distribution \(\mathbb{P}\) is uncertain, and we can solve the problem (1) by using the distributionally robust optimization (DRO) theory, that is, to solve the following problem.

\[
\inf_b \sup_{\mathbb{Q} \in \mathbb{P}} \mathbb{E}\left[ l_b(x,y) \right],
\]

where \(\mathbb{P}\) is a fuzzy set of distributions supported by training samples and contains the unknown distribution \(\mathbb{P}\). In probability space we use the Wasserstein metric to characterize a spherical fuzzy set \(\mathbb{P}\). And then, we need to introduce the following definition of Wasserstein Distance.

**Definition 1** (Wasserstein Distance). Let \(M(\Xi^2)\) denote the set of probability distributions on \(\Xi \times \Xi\). The Wasserstein distance between two distributions \(\mathbb{P}\) and \(\mathbb{Q}\) supported on \(\Xi\) is defined as

\[
W(\mathbb{Q}, \mathbb{P}) = \inf_{\mathbb{H} \in M(\Xi^2)} \left\{ \int_{\Xi^2} d(\xi, \xi') \mathbb{H}(d\xi, d\xi') : \int_{\Xi} \mathbb{H}(d\xi) = \mathbb{Q}(\xi'), \int_{\Xi} \mathbb{H}(\xi, d\xi') = \mathbb{P}(\xi) \right\},
\]

where \(\xi = (x,y)\), and \(d(\xi, \xi')\) is a metric on \(\Xi\).

In this paper we use Wasserstein Balls \(\mathbb{B}_\delta(\mathbb{P}) = \{ \mathbb{Q} : W(\mathbb{Q}, \mathbb{P}) \leq \delta \}\) with radius \(\delta\) and center \(\mathbb{P}\) as fuzzy sets. Given the training samples \(\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n\), we can use the empirical distribution \(\hat{\mathbb{P}}_n = \frac{1}{n}\) to approximate the unknown distribution \(\mathbb{P}\). Therefore, problem (2) in this study can be written as the following form.
(DRO) \[ \inf_b \sup_{Q \in \mathbb{B}(P_n)} \mathbb{E}^Q[l_b(x,y)]. \] (3)

Problem (DRO) cannot be solved directly, and Theorem 1 shows that it is equivalent to a tractable semi-infinite programming problem.

**Theorem 1** (Tractable Reformulation). The problem (DRO) is equivalent to the following semi-infinite programming problem.

\[
\begin{align*}
\min_{b,u,v} & \quad \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u \\
\text{s. t.} & \quad \{ l_b(\xi) - v_i - ud(\xi, \xi_i) \leq 0, \forall \xi \in \Xi, \forall i \in [n], \} \\
& \quad u \geq 0.
\end{align*}
\] (4)

Where \( \xi = (x,y) \), and \( d(\xi, \xi_i) \) is a metric on \( \Xi = \mathbb{R} \times \{0,1\} \).

**Proof** By definition of the Wasserstein ball and the empirical distribution we have

\[
\mathbb{B}(\mathbb{P}_n) = \left\{ Q : \exists \Pi \text{ s. t. } \sum_{i=1}^{n} \Pi(i, \xi_i) = Q(\xi), \forall \xi \in \Xi, \int_{\Xi} \Pi(d(\xi, \xi_i)) = \frac{1}{n}, \forall i \in [n], \sum_{i=1}^{n} d(\xi, \xi_i) \Pi(d(\xi, \xi_i)) \leq \delta \right\}.
\] (5)

Then, the inner problem (DRO) is formulated as the following conic linear program.

\[
\begin{align*}
\sup_{Q \in \mathbb{B}(\mathbb{P}_n)} & \quad \mathbb{E}^Q[l_b(x,y)] = \sup_{Q \in \mathbb{B}(\mathbb{P}_n)} \int_{\Xi} l_b(\xi) Q(d\xi) \\
\text{s. t.} & \quad \sup_{\Pi \in \mathbb{M}(\Xi)} \int_{\Xi} \sum_{i=1}^{n} l_b(\xi) \Pi(d(\xi, \xi_i)) \\
& \quad \int_{\Xi} \sum_{i=1}^{n} d(\xi, \xi_i) \Pi(d(\xi, \xi_i)) = \frac{1}{n}, \forall i \in [n], \] \\
& \quad \int_{\Xi} \sum_{i=1}^{n} d(\xi, \xi_i) \Pi(d(\xi, \xi_i)) \leq \delta, \\
& \quad \Pi \geq 0.
\end{align*}
\] (6)

Thus, we obtain the following sample-based dual problem of (7) (Shapiro, 2001).

\[
\begin{align*}
\min_{u,v} & \quad \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u \\
\text{s. t.} & \quad \{ l_b(\xi) - v_i - ud(\xi, \xi_i) \leq 0, \forall \xi \in \Xi, \forall i \in [n], \} \\
& \quad u \geq 0.
\end{align*}
\] (7)

Where \( \xi = (x,y) \), and \( d(\xi, \xi_i) \) is a metric on \( \Xi = \mathbb{R} \times \{0,1\} \). Therefore, the result of Theorem 1 follows by combing the problem (8) with the outer minimization problem (DRO).

The Raman spectroscopy data is not only high in dimension, but also has a close
correlation between adjacent features, which is known as Raman spectral peaks. Therefore, the fused lasso term \( R(b) \) is added to the model to preserve the sparsity of the difference between neighboring features to highlight the importance of spectral peaks. The fused lasso model can automatically and continuously select features while performing parameter estimation, which makes model simplification explanatory.

\[
R(b) = \lambda_1 \|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1, \lambda_1 \geq 0, \lambda_2 \geq 0. \tag{9}
\]

Thus, we propose the following fused lasso distributionally robust optimization (FLDRO) model.

\[
\text{min} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1 \tag{11a}
\]

\[
\text{s.t.} \begin{cases} 
- \|x - \hat{x}_i\|^2 + \|y - \hat{y}_i\|^2 + v_i - u & \leq 0, \forall x \in X, \forall i \in [n], \\
\log(1 + e^{x \hat{x}_i}) - v_i - \delta d((x, y), (\hat{x}_i, \hat{y}_i)) & \leq 0, \forall x \in X, \forall i \in [n], \\
u \geq 0.
\end{cases} \tag{11b-11d}
\]

In the next section, we will introduce an algorithm to solve the problem (11).

3. Algorithm for the Tractable Reformulation FLDRO

In the previous section, we gave the tractable form of the problem (11), which is a semi-infinite programming problem that can be solved by using the cutting-surface algorithm. The idea of the cutting-surface method is to solve a relaxation problem of the semi-infinite program at each iteration, where the relaxation problem has a finite number of constraints. Constraints are added to the current set of constraints at the solution of the relaxation problem.

Let \( \chi = [b, u, v] \) be the decision variables of problem (11), and define the following functions.

\[
f(\chi) = \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1, \tag{12}
\]

\[
g_i(\chi, \xi) = l_b(\xi) - v_i - ud(\xi, \xi_i), \forall \xi \in \Xi, \forall i \in [n]. \tag{13}
\]

Thus, the problem (11) can be rewritten as

\[
\min_{\chi} f(\chi) \tag{14a}
\]
\begin{equation}
\begin{array}{l}
\text{s.t. } \{g_i(\chi, \xi) \leq 0, \forall \xi \in \Xi, \forall i \in [n]\}, \\
\chi \in X.
\end{array}
\end{equation}

Problem (14) is a semi-infinite program. We give the following cutting-surface algorithm to solve this problem.

**Algorithm 1** A cutting-surface algorithm to solve (14).

**Input:** training data \{\((\hat{x}_i, \hat{y}_i)\}\), parameters \(\delta, \lambda_1, \lambda_2\).

**Output:** \(\chi^*\)

1: (Initialization) Set \(\Xi^0 \leftarrow \emptyset\), \(k \leftarrow 0\).

2: Determine an optimal solution \(\chi^k\) of the problem

3: \(\min_{\chi \in X} f(\chi) \text{ s.t. } g_i(\chi, \xi) \leq 0, \forall \xi \in \Xi^k, \forall i \in [n]\).

4: for \(i = 1, 2, \ldots, n\) do

5: Determine an optimal solution \(\xi^{k+1}\) of \(i\)-th separation problem

6: \(\max_{\xi \in \Xi} g_i(\chi^k, \xi)\)

7: if \(g_i(\chi^k, \xi^{k+1}) > 0\)

8: \(\Xi^{k} \leftarrow \Xi^{k} \cup \{\xi^{k+1}\}\)

9: end if

10: end for

11: if no cut is added in the current iteration, then

12: STOP.

13: end if

14: Update \(\Xi^{k+1} \leftarrow \Xi^{k}\), \(k \leftarrow k + 1\) and go to Step 2.

The key to Algorithm 1 is to solve the following two optimization problems.

\begin{equation}
\min_{\chi \in X} f(\chi) \text{ s.t. } g_i(\chi, \xi) \leq 0, \forall \xi \in \Xi^k, \forall i \in [n].
\end{equation}

\begin{equation}
\max_{\xi \in \Xi} g_i(\chi^k, \xi), \forall i \in [n].
\end{equation}

Since separation problem (16) has the same structure for all \(i \in [n]\), according to the problem (11) we drop the index \(i\) in \(g_i(\chi, \xi)\) and use the following notations to simplify our presentation.

\[X = \begin{pmatrix}
\chi_1 \\
\vdots \\
\chi_{2n}
\end{pmatrix},\]  

\[(17)\]
Without loss of generality, given \( |X| \), where \( X \) is Wasserstein Balls radius, \( \hat{x}_i \in \mathbb{R}^{1 \times m} \) is the feature of \( i \)-th training sample, and \( \hat{y}_i \in \{0,1\} \) is the corresponding label. Through analysis, the objective function of problem (21) is a non-smooth, the constraints (21b) and (21c) are nonlinear, we can use the VAPP (Variant Auxiliary Problem Principle) method proposed by Zhao and Zhu (2022) to solve it. For the sake of brevity, we write problem (21) as follows:

\[
\begin{align*}
\min_{b,u,v,\eta} & \quad \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \|\eta\|_1 \\
\text{s.t.} & \quad -x b + \log(1 + e^{x \cdot b}) - v_i - u(\|x - \hat{x}_i\|_2 + \|1 - \hat{y}_i\|_2) \leq 0, \forall x \in \mathbb{X}, \forall i \in [n], \\
& \quad \log(1 + e^{x \cdot b}) - v_i - u(\|x - \hat{x}_i\|_2 + \|\hat{y}_i\|_2) \leq 0, \forall x \in \mathbb{X}, \forall i \in [n], \\
& \quad Ab = \eta, \\
& \quad u \geq 0.
\end{align*}
\]

Where \( A = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(m-1) \times m} \), \( b \in \mathbb{R}^m \) is the regression coefficient, \( \lambda_1 \geq 0 \) and \( \lambda_2 \geq 0 \) are the penalty parameters, \( \delta \geq 0 \) is Wasserstein Balls radius, \( \hat{x}_i \in \mathbb{R}^{1 \times m} \) is the feature of \( i \)-th training sample, and \( \hat{y}_i \in \{0,1\} \) is the corresponding label. Through analysis, the objective function of problem (21) is a non-smooth, the constraints (21b) and (21c) are nonlinear, we can use the VAPP (Variant Auxiliary Problem Principle) method proposed by Zhao and Zhu (2022) to solve it. For the sake of brevity, we write problem (21) as follows:

\[
\begin{align*}
\min_{b,u,v,\eta} & \quad \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \|\eta\|_1 \\
\text{s.t.} & \quad -x b + \log(1 + e^{x \cdot b}) - v_i - u(\|x - \hat{x}_i\|_2 + \|1 - \hat{y}_i\|_2) \leq 0, \forall x \in \mathbb{X}, \forall i \in [n], \\
& \quad \log(1 + e^{x \cdot b}) - v_i - u(\|x - \hat{x}_i\|_2 + \|\hat{y}_i\|_2) \leq 0, \forall x \in \mathbb{X}, \forall i \in [n], \\
& \quad Ab = \eta, \\
& \quad u \geq 0.
\end{align*}
\]

Thus, problem (16) can be solved by the projected gradient descent algorithm, and its iterative process is as follows:

\[
X^{k+1} = \Pi_X\left(X^k + \varepsilon \nabla_X g(x^k, X^k)\right),
\]

where \( \Pi_X(\cdot) \) is the projection function onto \( \mathbb{X} \), \( \varepsilon > 0 \) which is the learning rate.
\[
\begin{align*}
\min_{b \in \mathbb{R}^m, u, v, \eta} & \quad f(b, u, v, \eta) \\
\text{s.t.} & \quad \Theta(b, u, v, \eta) \in \{x \in \mathbb{R}^{2n_l} | x \leq 0 \} \times \{0\}^m.
\end{align*}
\] (22a)

Where

\[
f(b, u, v, \eta) = \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \|\eta\|_1,
\] (23)

\[
\Theta(b, u, v, \eta) = \Omega(b) - \begin{pmatrix} \text{vec}(V) \\ \text{vec}(V) \\ \eta - Ab \end{pmatrix} + \Phi(u),
\] (24a)

\[
\Omega(b) = \begin{pmatrix} \Omega_1(b) \\ \Omega_2(b) \\ \mathbb{0}_{m-1} \end{pmatrix} \in \mathbb{R}^{2n_l+m-1},
\] (24b)

\[
V = \begin{pmatrix} v_1 & \cdots & v_1 \\ \vdots & \ddots & \vdots \\ v_n & \cdots & v_n \end{pmatrix} \in \mathbb{R}^{n \times l},
\] (24c)

\[
\Phi(u) = \begin{pmatrix} \Phi_1(u) \\ \Phi_2(u) \\ \mathbb{0}_{m-1} \end{pmatrix} \in \mathbb{R}^{2n_l+m-1},
\] (24d)

\[
\text{mat}(\Omega_1(b))_{ij} = -x_j b + \log(1 + e^{x_j b}), \forall i \in [n], \forall j \in [l],
\] (24e)

\[
\text{mat}(\Omega_2(b))_{ij} = \log(1 + e^{x_j b}), \forall i \in [n], \forall j \in [l],
\] (24f)

\[
\text{mat}(\Phi_1(u))_{ij} = -u \left( \|x_j - \tilde{x}_j\|^2 + \|1 - \tilde{y}_i\|^2 \right), \forall i \in [n], \forall j \in [l],
\] (24g)

\[
\text{mat}(\Phi_2(u))_{ij} = -u \left( \|x_j - \tilde{x}_j\|^2 + \|\tilde{y}_i\|^2 \right), \forall i \in [n], \forall j \in [l],
\] (24h)

The main idea of VAPP is to linearize the non-separable term \(\Omega(b)\) to achieve the purpose of parallel computing. Then the iterative process of solving problem (22) by VAPP is obtained.

\[
(b^{k+1}, u^{k+1}, v^{k+1}, \eta^{k+1}) = \arg\min_{b \in \mathbb{R}^m, u, v, \eta \in \mathbb{R}^+} f(b, u, v, \eta)
\]

\[
+ \langle \Pi(p^k + \gamma \Theta(b^k, u^k, v^k, \eta^k)), \nabla \Omega(b^k) b - \begin{pmatrix} \text{vec}(V) \\ \text{vec}(V) \\ \eta - Ab \end{pmatrix} + \Phi(u) \rangle
\]

\[
+ \frac{1}{2\varepsilon} \left( \|b - b^k\|^2 + \|u - u^k\|^2 + \|v - v^k\|^2 + \|\eta - \eta^k\|^2 \right),
\] (25)

where \(\gamma > 0, \varepsilon > 0\).
Since the above optimization problems are separable, we can write it as follows:

\[
\begin{align*}
\n_{k+1} &= \arg\min_{b \in \mathbb{R}^m} \lambda_1 \|b\|_1 + \langle \Pi_1^k, \nabla \Omega_1(b^k) b \rangle + \langle \Pi_2^k, \nabla \Omega_1(b^k) b \rangle + \langle \Pi_3^k, A b \rangle \\
&\quad + \frac{1}{2\varepsilon} \|b - b^k\|^2,
\end{align*}
\]

(27a)

\[
\begin{align*}
\n_{k+1} &= \arg\min_{\delta \geq 0} \delta u + \langle \Pi_1^k, \Phi_1(u) \rangle + \langle \Pi_2^k, \Phi_2(u) \rangle + \frac{1}{2\varepsilon} \|u - u^k\|^2,
\end{align*}
\]

(27b)

\[
\begin{align*}
\n_{k+1} &= \arg\min_{\nu \geq 0} \frac{1}{n} \langle \mathbb{1}_n, \nu \rangle - \langle \Pi_1^k, \text{vec}(\mathbb{V}) \rangle - \langle \Pi_2^k, \text{vec}(\mathbb{V}) \rangle + \frac{1}{2\varepsilon} \|\nu - \nu^k\|^2,
\end{align*}
\]

(27c)

\[
\begin{align*}
\n_{k+1} &= \arg\min_{\eta} \lambda_2 \|\eta\|_1 - \langle \Pi_3^k, \eta \rangle + \frac{1}{2\varepsilon} \|\eta - \eta^k\|^2,
\end{align*}
\]

(27d)

where \( \Pi(p^k + \gamma \theta(b^k, u^k, \nu^k, \eta^k)) = \left( \begin{array}{c} \Pi_1^k \\ \Pi_2^k \\ \Pi_3^k \end{array} \right), \Pi_1^k = \left( \begin{array}{c} p_1^k + \gamma (\Omega_1(b^k) - \text{vec}(\mathbb{V}^k) + \Phi_1(u^k)) \\ p_2^k + \gamma (\Omega_1(b^k) - \text{vec}(\mathbb{V}^k) + \Phi_2(u^k)) \end{array} \right) \in \mathbb{R}^{n^l}, \Pi_2^k = \left( \begin{array}{c} p_2^k + \gamma (\Omega_1(b^k) - \text{vec}(\mathbb{V}^k) + \Phi_2(u^k)) \end{array} \right) \in \mathbb{R}^{n^l}, \Pi_3^k = p_3^k + \gamma (A b^k - \eta^k) \in \mathbb{R}^{m-1}. \)

And the dual variables iteration process is

\[
\begin{align*}
p^{k+1} &= \Pi(p^k + \gamma \theta(b^{k+1}, u^{k+1}, \nu^{k+1}, \eta^{k+1})),
\end{align*}
\]

(28)

where \( p^k = \left( \begin{array}{c} p_1^k \\ p_2^k \\ p_3^k \end{array} \right), \) then

\[
\begin{align*}
p_1^{k+1} &= \left( p_1^k + \gamma (\Omega_1(b^{k+1}) - \text{vec}(\mathbb{V}^{k+1}) + \Phi_1(u^{k+1})) \right)_+ \in \mathbb{R}^{n^l},
\end{align*}
\]

(29a)

\[
\begin{align*}
p_2^{k+1} &= \left( p_2^k + \gamma (\Omega_2(b^{k+1}) - \text{vec}(\mathbb{V}^{k+1}) + \Phi_2(u^{k+1})) \right)_+ \in \mathbb{R}^{n^l},
\end{align*}
\]

(29b)

\[
\begin{align*}
p_3^{k+1} &= p_3^k + \gamma (A b^{k+1} - \eta^{k+1}) \in \mathbb{R}^{m-1}.
\end{align*}
\]

(29c)

In order to get the closed form for the above iterative process, we need to introduce
the soft-thresholding operator

\[
\min_{b \in \mathbb{R}^m} \|b\|_1 + \frac{\beta}{2} \|b - r\|^2,
\]

(30)

where \( \beta > 0, r \in \mathbb{R}^m \). The solution to the problem can be expressed by using a soft threshold operator.

\[
b^* = \text{shrink}(r, 1/\beta) = \text{sign}(r) \cdot \max\{0, |r| - 1/\beta\},
\]

(31)

where \( \text{sign}(\cdot) \) is the sign function. And then, we can write the analytical solution of \((b, u, v, \eta)\) as follows:

\[
b^{k+1} = \text{sign}(r_1^k) \cdot \max\{0, |r_1^k| - \lambda_1 \epsilon\},
\]

(32a)

\[
r_1^k = b^k - \epsilon (\nabla \Omega_1 (b^k)^T \Pi_1^k + \nabla \Omega_2 (b^k)^T \Pi_2^k + A^T \Pi_3^k),
\]

(32b)

\[
u^{k+1} = \left( u^k - \epsilon \left( \delta - \sum_{i=1}^n \sum_{j=1}^l \text{mat}(\Pi_1^k)_{ij} \left( \|x_j - \hat{x}_i\|^2 + \|1 - \hat{y}_i\|^2 \right) \right) \right. \\
\left. - \sum_{i=1}^n \sum_{j=1}^l \text{mat}(\Pi_2^k)_{ij} \left( \|x_j - \hat{x}_i\|^2 + \|\hat{y}_i\|^2 \right) \right)_{+},
\]

(32c)

\[
v^{k+1} = \left( v^k - \epsilon \left( \frac{1}{n} \mathbb{1}_n - \sum_{j=1}^l \text{mat}(\Pi_1^k)_{.j} - \sum_{j=1}^l \text{mat}(\Pi_2^k)_{.j} \right) \right)_{+},
\]

(32d)

\[
\eta^{k+1} = \text{sign}(r_2^k) \cdot \max\{0, |r_2^k| - \lambda_2 \epsilon\},
\]

(32e)

\[
r_2^k = \eta^k + \epsilon \Pi_3^k.
\]

(32f)

4. Experimental Results and Discussions

In our experiments, we selected two liquid milky yellow cheese products that are indistinguishable to the naked eyes, and their Raman spectroscopy is shown in figure 2 (denoted as MK and SL respectively). The integration time of Raman spectrometer was 70 seconds, and 25 samples were collected from each dairy product. A total of \( n = 50 \) original Raman spectroscopy samples were collected, and the number of features of every sample was \( m = 2090 \). 80 percent of the samples were randomly selected as training samples and the rest were testing samples.
In addition, standard logistic regression (LR), L1 regularized logistic regression (LR-L1) and L2 regularized logistic regression (LR-L2) were taken as the control group. Set FLDRO parameters $\delta = 2, \lambda_1 = 2, \lambda_2 = 2, \mathbf{X} = [\bar{x}_{\text{min}}, \bar{x}_{\text{max}}]$ in equation (20). The parameter $C$ is equal to 1 in LR-L1 and LR-L2.

\[
\begin{align*}
\text{(LR)} & \quad \min_{\mathbf{b}} l_{\mathbf{b}}(x, y) \quad (33) \\
\text{(LR - L1)} & \quad \min_{\mathbf{b}} \|\mathbf{b}\|_1 + Cl_{\mathbf{b}}(x, y) \quad (34) \\
\text{(LR - L2)} & \quad \min_{\mathbf{b}} \frac{1}{2} \|\mathbf{b}\|^2 + Cl_{\mathbf{b}}(x, y) \quad (35)
\end{align*}
\]

Figure 3 shows the sparsity of regression coefficient $\mathbf{b}$ obtained by the four methods and corresponding testing accuracy. It can be seen from the figure that the recognition accuracy by FLDRO, LR and LR-L2 is significantly better than that of LR-L1. And in terms of sparsity, the FLDRO and LR-L1 are higher than the other two methods. In order to visually demonstrate the feature selection results by every method, without loss of generality, we plot the absolute normalized value of $\mathbf{b}$ in one figure with the testing data (as shown in figure 4). As can be seen from figure 4, LR and LR-L2 not only take all spectral peaks as key features, but also include a large number of noise features, which will reduce the generalization ability of the model. Although LR-L1 has no noise feature, only one spectral peak is selected as the key feature, which will lead to underfitting problem. FLDRO takes the spectral peaks marked by the box in the figure as the key feature to distinguish two kinds of dairy products. The experimental results also show that FLDRO has obvious advantages both in regression coefficient
sparsity and recognition accuracy.

Figure 3 Sparsity of regression coefficient $b$ and corresponding testing accuracy

Figure 4 (a) Testing samples and regression coefficients obtained by four methods. (b, c) local magnification of selected spectral peaks by FLDRO

Next, we test the robustness of the four methods by adding noise to the original samples. As we can see from figure 5, although the raw samples retain obvious features after adding noise, a large number of interference factors appear. To test the anti-noise performance of each model, we added Gaussian noise with mean 0.01 and variance 0.01, and multiplicative noise with variance 0.04 to the original test set.
The boxplot of MAE (Mean Absolute Error) by the four methods is shown in figure 6.

\[
MAE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| \log \left( 1 + e^{(x_{test}^i - b)} \right) - y_{test}^i \right|
\]  

(36)

where the \( x_{test}^i \) represents the \( i \)-th test Raman spectroscopy, \( y_{test}^i \) is the corresponding label and \( n_{test} \) represents the number of testing samples. The results of robustness test in figure 6 show that the MAE box size of FLDRO is the smallest in both noise scenarios, and that the median of the box is also significantly lower than other models, which means that the robustness and accuracy of distributionally robust optimization are significantly better than other models. Among them, the robustness of L1 regularized logistic regression is similar to FLDRO. This is mainly caused by the sparsity of these two methods.

Furthermore, we test the recognition accuracy of the four methods under different Gaussian noise intensities. As shown in figure 7, when the mean value is fixed (increasing), and the variance is increasing (fixed), the recognition accuracy of all
methods is decreasing. However, compared with other methods, FLDRO has the highest accuracy and the smallest influence of variance on accuracy variation. In addition, the accuracy of the three logistic regressions in the figure is close to 50% under the condition of large noise intensity. This is because the sparsity of $b$ by LR and LR-L2 is zero, in other words, all features are selected, which results in serious distortion of its recognition results. Besides, only one spectral peak is selected by LR-L1, which makes it difficult to make a correct identification.

![Figure 7](image)

**Figure 7** Trend of recognition accuracy by different methods with Gaussian noise. (a) Trend of accuracy with variance varying and fixed mean value 0.01. (b) Trend of accuracy with mean varying and fixed variance 0.01.

**5. Conclusions**

In order to verify the advantages of FLDRO in recognition accuracy and robustness of Raman spectral identification, we adopt FLDRO, LR, LR-L1, and LR-L2 methods to perform recognition experiments on spectral data of two kinds of dairy products. Firstly, based on the regression coefficients obtained from training samples by different methods, the recognition accuracy of FLDRO is 90%, and the obtained $b$ sparsity is 0.9737, which is the best among all methods. Secondly, it can be seen from figure 4 that FLDRO provides a significant regression coefficient value at the spectral peak of Raman spectroscopy, which indicates that FLDRO has a strong ability to identify spectral peaks. Moreover, figure 6 shows that MAE boxes size of FLDRO are significantly smaller than those of the other three models under the interference of two different noises, indicating that FLDRO model has a strong anti-interference ability. Finally, in the sensitivity test experiments of Gaussian noise, FLDRO outperforms the other three methods in terms of both recognition ability and robustness.

**Declarations**

**Conflict of Interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Ethics Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability https://github.com/xuxiang74/Raman-Spectroscopy.git.
References


