

## 基于 Laplacian-Markov 先验数据的加权光谱反卷积模型

李 姣, 张天序

(华中科技大学 图像识别与人工智能研究所, 湖北 武汉 430074)

**摘要:** 针对由光谱仪器测得的拉曼光谱数据经常会受到随机噪声和仪器误差等的影响而导致低分辨率的问题, 文中提出了一种能在恢复光谱结构的同时又能抑制光谱噪声的方法, 即基于 Laplacian-Markov 约束的数据加权光谱反卷积模型。该模型将退化光谱中恢复真实光谱的问题转化为最大后验概率的求解问题, 推导出了拉曼光谱恢复的变分模型。模型中利用 Laplacian-Markov 作为光谱数据的光滑性先验, 提出加权光谱反卷积来恢复退化的光谱, 并使用分裂 Bregman 迭代法求解。文中对该算法利用实验数据进行了验证, 实验表明该方法既能恢复退化光谱细节又能抑制光谱噪声, 并且求解速度快, 有较强的实用价值。

**关键词:** Laplacian-Markov 先验; 光谱反卷积; 分裂 Bregman 迭代法; 拉曼光谱

**中图分类号:** O657.37   **文献标志码:** A   **文章编号:** 1007-2276(2013)12-3464-06

## Data-weighted spectral deconvolution with Laplacian-Markov priori

Li Jiao, Zhang Tianxu

(Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology,  
Wuhan 430074, China)

**Abstract:** Raman spectroscopic data often suffers from common problems of bands overlapping and random Gaussian noise. Spectral resolution can be improved by mathematically removing the effect of the instrument response function. In this paper, a novel method to deconvolute the degraded spectrum with the Laplacian-Markov priori was proposed, solving by split Bregman optimization scheme, which was fast, robust to noise and easy to implement. The Laplacian-Markov priori was proposed to save the shape peaks and suppress the noise. A data weighted operator was introduced to spectral deconvolution to find a balance between band narrowing and noise suppression. The method could estimate spectral structural details as well as suppress the noise effectively. Experimental results with real Raman spectra manifest that this algorithm can deconvolute the overlapping peaks as well as suppress the noise effectively. Owing to the fast of computing time, it is expected that the proposed method has considerable value in practice.

**Key words:** Laplacian-Markov priori; spectral deconvolution; split Bregman iteration;

Raman spectroscopy

收稿日期: 2013-04-09; 修订日期: 2013-05-13

基金项目: 国家自然科学基金科学仪器专项(61227007)

作者简介: 李姣(1987-), 女, 硕士生, 主要从事图像分析, 模式识别与智能系统, 多光谱信息处理技术方面的研究。

Email: lijiaotiantian365@163.com

导师简介: 张天序(1947-), 男, 教授, 主要从事图像分析与目标识别, 精确制导智能化信息处理医学图像处理, 实时并行处理方面的研究。Email: tianxuzhang@hust.edu.cn

## 0 引言

近年来,人们对拉曼光谱技术的研究相当活跃,使得拉曼光谱在食品质量监测<sup>[1]</sup>、化学品快速检测、材料分析<sup>[2]</sup>和生学医学等领域的得到了广泛的应用。然而,光谱数据在光谱仪测量过程中,经常会受到谱线的自然展宽、碰撞展宽、探测器和电路低通特性等<sup>[3]</sup>因素的影响,导致光谱数据强噪声和子带重叠现象。这些噪声干扰给光谱应用带来了困难,甚至会导致严重的鉴定和估计错误,这极大限制了光谱的应用。因此,对数据进行预处理、提高光谱的分辨率就显得不可或缺。

目前,恢复退化光谱的方法有很多,这些方法主要可分为两大类:曲线拟合<sup>[4]</sup>和光谱反卷积<sup>[5-7]</sup>。曲线拟合的方法认为光谱曲线是通过高斯函数和洛伦兹函数的组合而成,即实测光谱能通过这些函数来拟合得到。这种方法能够很好的解决子带重叠的问题,即能拟合出重叠的波峰,但是对噪声极其敏感。光谱反卷积的方法主要是通过不同的约束来恢复退化的光谱,这些方法有高阶统计<sup>[5]</sup>(HOS)、Tikhonov 正则化<sup>[6,8]</sup>、同态滤波<sup>[9]</sup>、等等。在这些方法中,约束较强的能够很好的抑制拉曼光谱噪声(Tikhonov 正则化<sup>[6,8]</sup>),但约束过强会导致退化严重的光谱细节不能恢复。约束较弱的能够将退化的细节较好的恢复出来(高阶统计<sup>[5]</sup>),但是抑制噪声能力会减弱。因此,这类方法在恢复光谱的分辨率和噪声抑制能力上很难达到平衡。

事实上,对于拉曼光谱数据曲线来讲,并不是所有的数据受到干扰后对本身数据的影响都是一样的,在强度变化平缓的波数受到噪声的影响较大,而在强度变化剧烈的波数,受到噪声的影响较小。如果根据强度变化的不同进行自适应处理,这样就可以达到恢复分辨率和噪声抑制能力的平衡。因此文中提出了一个带有数据加权的光谱反卷积的方法,数据加权即根据强度变化不同而采取不同的约束,在强度变化平缓的地方采取较强的约束,抑制好噪声;在强度变化剧烈的地方采取较弱的约束,保存好光谱的细节。另一方面,Laplacian-Markov 先验是一种弱的光滑性先验。文中采这种先验约束拉曼光谱的光滑性。实验结果也证明了该方法对恢复退化的光谱很有效。

## 1 模型提出

文中提出了一种新的恢复退化光谱数据的方法。大多数利用拉曼光谱仪测得的拉曼光谱数据可以描述为真实信号与测量仪器相应函数的卷积:

$$g(v)=Hf(v)+n(v) \quad (1)$$

式中: $f(v)$ 为真实信号; $g(v)$ 为测量仪器测得的观测信号; $H$ 为仪器响应函数,也称为核函数,主要有内部的线性函数和仪器加宽函数等组成。最大后验概率的方法是获得真实信号的一个通用方法。在观测光谱  $g(v)$  和仪器响应函  $H$  已知的情况下,求得最大化真实光谱  $f(v)$  的条件概率,即:

$$\hat{f}=\operatorname{argmax}_f P(f|g) \quad (2)$$

运用贝叶斯公式,然后取负对数,可以得到:

$$\hat{f}=\operatorname{argmin}_f \{-\log P(f|g)-\log P(f)\} \quad (3)$$

式中: $\log P(f|g)$ 为取对数的似然函数。对于每一个波数上的光谱噪声,假设它们是满足独立同分布的高斯分布,因此光谱数据上的噪声可以表示为:

$$P(n)=\prod_i \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{\|n_i\|^2}{2\sigma_i^2}\right)=\frac{1}{Z_1} \exp\left(-\frac{\|Hf-g\|^2}{2\sigma^2}\right) \quad (4)$$

式中: $\sigma$ 为噪声的标准差; $n_i$ 为每一个波数点的噪声强度; $N$ 为光谱的长度; $\|\cdot\|$ 表示欧几里德范数。由公式(1)可知, $P(g|f)=P(n)$ ,因此由公式(4)可以推出:

$$P(g|f)=\frac{1}{Z_1} \exp\left(-\frac{\|Hf-g\|^2}{2\sigma^2}\right)=\frac{1}{Z_1} \exp(-\|Hf-g\|^T Q^T Q(Hf-g)) \quad (5)$$

$$\frac{1}{Z_1} \exp(-\|Q(Hf-g)\|^2)$$

式中: $Q$ 为对角矩阵; $q_{ii}$ 为对角线上的元素  $\sigma_i$ ,其他的元素都为零。即:

$$Q=\begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \sigma_N \end{bmatrix}$$

在文中,将  $q_{ii}$  归一化得到:

$$q_{ii}=\ln\left[\frac{(e-1)(\text{std}-\min)}{\max-\min}+1\right] \quad (6)$$

式中:std 为当前光谱点  $f_i$  的邻域( $f_{i-3}, f_{i-2}, f_{i-1}, f_i, f_{i+1}, f_{i+2}, f_{i+3}$ ) 的标准差;max 和 min 分别为 std 的最大和最小值,光谱数据强度变化的越剧烈, std 的值越大。

公式(4)中的先验概率  $P(f)$  可由 Laplacian-Markov 随机场求得, 即:

$$P(f) = \frac{1}{Z_2} \exp\left(-\frac{1}{\lambda} \sum_{i \in C} |df_i|\right) \quad (7)$$

式中:  $df_i = (f_{i+1} - f_i)/2$ , 将公式(5)和公式(7)带入公式(3), 并消除参数  $Z_1$  和  $Z_2$ , 得到文中的能量泛函:

$$E(f) = \frac{1}{2} \|Q(Hf - g)\|_2^2 + \lambda \|Df\|_1 \quad (8)$$

式中: 矩阵  $D$  是光谱的差分矩阵, 等号右边公式的第一项为数据项, 第二项为正则项, 定义这种方法为 Laplacian-Markov 光谱反卷积, 即 LMSD (Laplacian-Markov Spectral Deconvolution)。

## 2 优化求解

对  $f$  求变分, 得到关于常规梯度下降方法  $f$  的 Euler-Lagrange 方程,

$$\nabla E(f) = H^T Q^T Q(Hf - g) - 2\lambda D^T Df \quad (9)$$

对于公式(9)求变分时在某些点会存在  $Df=0$  而导致不可微的问题, 为了克服常规方法的困难, 文中引进了分裂 Bregman 迭代法<sup>[10]</sup>。具体过程如下, 让  $d$  代替  $Df$ , 公式(8)转化成了一个受约束的优化问题:

$$\min_{f, d} \frac{1}{2} \|Q(Hf - g)\|_2^2 + \lambda \|d\|_1 + \frac{\alpha}{2} \|d - Df\|_2^2 \quad (10)$$

分裂 Bregman 迭代法将公式(8)分裂成  $L_1$  范数和  $L_2$  范数, 因此, 这个公式(10)可以分为两个子问题:

$$f^{k+1} = \operatorname{argmin}_f \frac{1}{2} \|Q(Hf - g)\|_2^2 + \frac{\beta}{2} \|d^k - Df - b^k\|_2^2 \quad (11)$$

$$d^{k+1} = \operatorname{argmin}_d \lambda \|d\|_1 + \frac{\beta}{2} \|d - (Df)^{k+1} - b^k\|_2^2 \quad (12)$$

其中, 公式(11)只有  $L_2$  范数, 线性的求解, 迭代式表示为:

$$f^{k+1} = f^k + \Delta t (\nabla E(f)) \quad (13)$$

式中:  $\nabla E(f) = H^T Q^T Q(Hf - g) + \beta(d^k - Df - b^k)$ ,  $H^T$  为矩阵  $H$  的转置。

对于公式(12)中只需优化  $L_1$  范数, 能够通过软阈值公式直接求解,

$$d^{k+1} = \max\left(|(Df)^{k+1} + b^k| - \frac{\lambda}{\beta}, 0\right) \frac{(Df)^{k+1} + b^k}{|(Df)^{k+1} + b^k|} \quad (14)$$

$$b^{k+1} = b^k + ((Df)^{k+1} - d^{k+1}) \quad (15)$$

软阈值公式(14)解决了不可微的问题。如果差值  $|(Df)^{k+1} + b^k|$  比阈值  $\lambda/\beta$  小, 则认为是噪声点, 则噪声点的强度用 0 代替; 如果比阈值  $\lambda/\beta$  大, 被认为

是光谱的细节点, 则用  $L_1$  范数对其光滑性约束。公式(15)更新了中间变量  $b^k$ , 它的物理含义是对过度去噪进行补偿。

根据上述分析, 模型的求解算法流程如图 1 所示, 其中  $\varepsilon_1$  是一个取值在  $10^{-9}$  和  $10^{-7}$  间小的正常数。

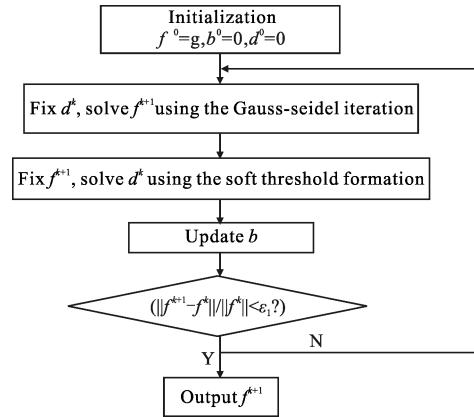


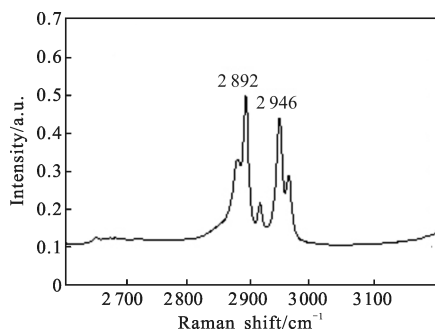
图 1 Laplacian-Markov 光谱反卷积算法(LMSD)

Fig.1 Laplacian-Markov spectral deconvolution

## 3 实验分析

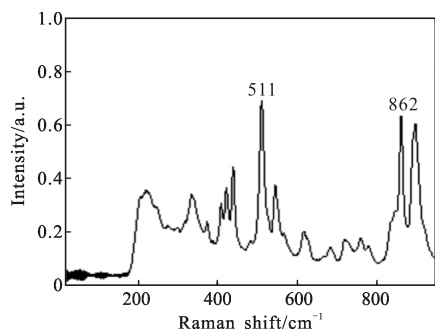
光谱反卷积时, 相应的核函数被选择来消除光谱仪器在测量中产生的退化。对于 LMSD 算法, 文中选择高斯形状函数作为核函数。在实验中, 最窄峰的宽度作为此高斯核的宽度。实验是在 Matlab 2012a 的环境下运行的。文中对光谱的光谱分辨率恢复能力和噪声抑制能力方面都进行了实验测试。文中主要进行四组实验, 笔者分别为 Tikhonov 正则化光谱反卷积<sup>[6]</sup>(TRSD)的方法、高阶统计<sup>[5]</sup>(HOS)的方法、无加权的 Laplacian-Markov 光谱反卷积 (LM-SD)的方法以及加权的 Laplacian-Markov 光谱反卷积(LM-QSD)的方法。TRSD、HOS、LM-QSD 实现方法间的对比, LM-SD、LM-QSD 实现方法内的对比。

在实验中才采用了两个实测光谱信号, 如图 2 所示, 图 2(a)和图 2(b)分别是葡萄糖( $C_6H_{12}O_6$ )和木精( $C_{15}H_{20}NO_7$ )的实测拉曼光谱曲线, 分辨率为  $1 \text{ cm}^{-1}$ , 曲线 2(a)在  $2850 \sim 2900 \text{ cm}^{-1}$  波数范围存在严重子带重叠现象, 该光谱可用于测试算法对光谱曲线的分辨率恢复能力。曲线 2(b)受到了严重的噪声影响, 该光谱可用于测试算法对噪声的抑制能力。公式(10)与(11)中的两个正则参数  $\alpha$  与  $\beta$  平衡了光谱窄化(第一项)与噪声抑制(第二项)。实验结果表明, 合适的取值为:  $\alpha=0.05$ ,  $\beta=0.1$ 。



(a) 葡萄糖

(a) D(+)-Glucopyranose



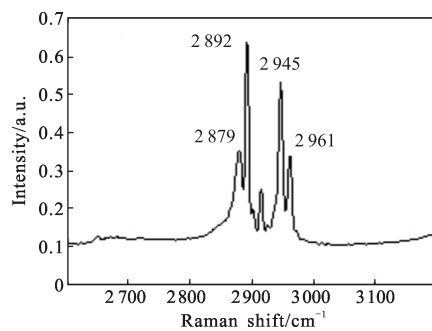
(b) 木精

(b) Methyl

图 2 实测拉曼光谱图

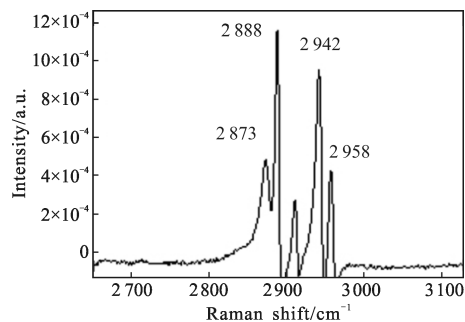
Fig.2 Real Raman spectra test set

如图 3 所示, 图 3 为不同的方法对退化葡萄糖拉曼光谱恢复的结果对比图, 图 3(a)为 TRSD 方法的结果<sup>[6]</sup>, 图 3(b)为 HOS 方法的结果<sup>[5]</sup>, 图 3(c)和图 3(d)分别 LM-SD 和加权 LM-QSD 方法的结果。四个实验中,  $\lambda$  都取值均为:  $\lambda=0.05$ 。由图 1(a)可知, 氨基葡萄糖的拉曼光谱在波数为  $2946\text{ cm}^{-1}$  时有较轻微的子带重叠现象, 对于这种波峰重叠现象, 这四种方法都能将重叠的波峰分开。而对于波峰重叠比较严重的波数 ( $2892\text{ cm}^{-1}$ ), 图 3(a)TRSD、图 3(b)HOS、图 3(c)LM-SD 方法只能将其分为两个峰, 分别为  $2879$  和  $2892\text{ cm}^{-1}$ (图 3(a))、 $2873$  和  $2888\text{ cm}^{-1}$ (图 3(b))、 $2879$  和  $2892\text{ cm}^{-1}$ (图 3(c)), 而图 3(d)LM-QSD 方法可以将其分为三个峰 ( $2879$ 、 $2892$ 、 $2903\text{ cm}^{-1}$ ), 对比 TRSD、HOS、LM-QSD 方法说明 LM-QSD 方法较其他方法提高光谱分辨率能力更强, 对比 LM-SD 和 LM-QSD 说明提高光谱分辨率能力增强的关键因素是加权  $Q$  矩阵。此外, 由实验结果可知, 对于这种受噪声影响较小的光谱, HOS、LM-SD、LM-QSD 方法都能抑制噪声, 而 HOS 方法却使强噪声有稍微的加强。



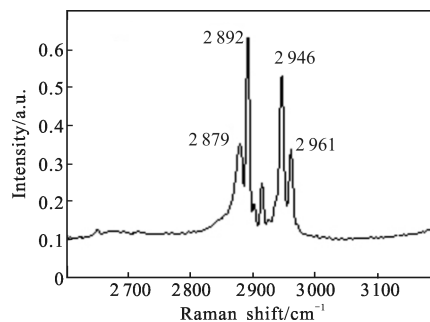
(a) TRSD 方法

(a) TRSD method



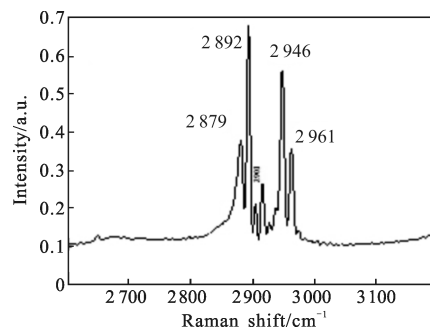
(b)HOS 方法

(b) HOS method



(c) LM-SD 方法

(c) LM-SD method



(d) LM-QSD 方法

(d) LM-QSD method

图 3 不同的方法对退化葡萄糖光谱恢复的结果对比  
Fig.3 Compared results for real Raman spectrum of  
D(+)-Glucopyranose(C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>)

图 4 所示为对图 1(b) 中木精(C<sub>15</sub>H<sub>29</sub>NO<sub>7</sub>)的拉

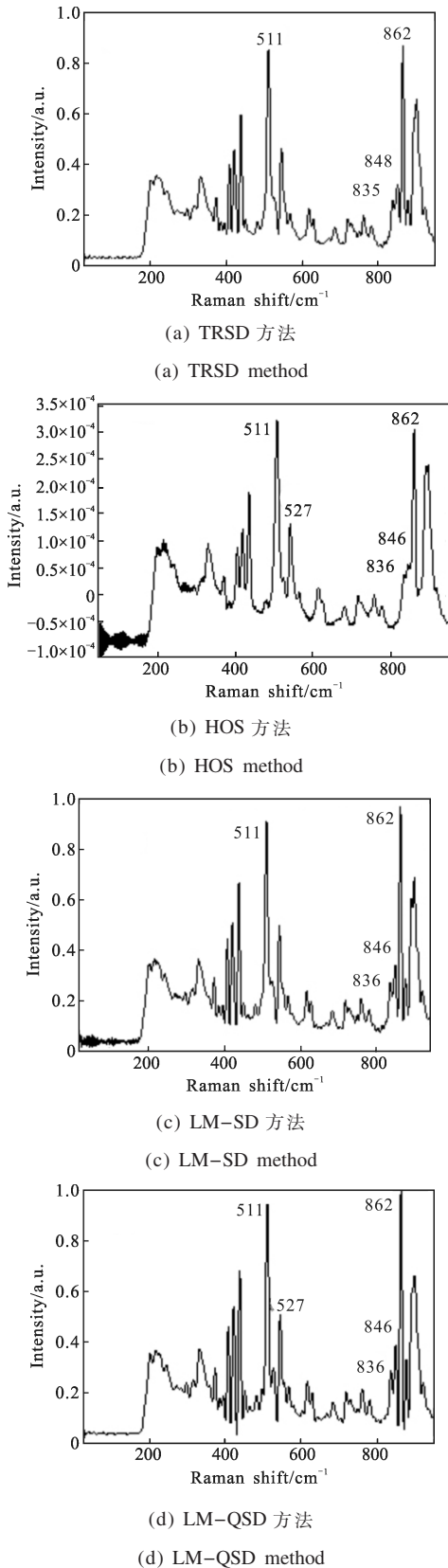


图 4 不同的方法对退化木精光谱恢复的结果对比

Fig.4 Compared results for real Raman spectrum of Methyl(C<sub>15</sub>H<sub>29</sub>NO<sub>7</sub>)

曼观测光谱的恢复处理的结果,图 4(a)~(d)分别为 TRSD 方法、HOS 方法、LM-SD 和 LM-QSD 方法的实验结果,在图 1(b)中可以看到光谱在波数范围20~200 cm<sup>-1</sup> 受到严重的高斯噪声影响,在图(3)实验结果中可以看到,只有图 3(a)TRSD 方法和图 3(d)的 LM-QSD 方法对噪声有较好的抑制效果,图 3(b) HOS 方法由于对噪声及其敏感反而放大了噪声,使得光谱曲线振荡的更剧烈,图 3(c) LM-SD 方法对噪声抑制能力较弱,不能达到去噪效果。因此比较 TRSD、HOS、LM-QSD 方法实验结果可知,TRSD、LM-QSD 方法有较好的噪声抑制能力;比较 LM-SD、LM-QSD 方法实验结果可知,LM-QSD 有较好噪声抑制能力主要取决于加权矩阵 *Q* 的贡献。虽然 TRSD、LM-QSD 都有较好的噪声抑制能力,但是比较图 3(a)和图 3(d)在波数分别为 936,846,862 cm<sup>-1</sup> 的恢复结果可以看出,图 3(d)LM-QSD 方法的峰分的更开,这说明 LM-QSD 方法既有较好的噪声抑制能力又有较强的恢复光谱分辨率能力。

值得注意的是,在给定高斯核的情况下,数据项是起到恢复重叠的波峰产生细节的作用,而正则项起到抑制噪声,使得光谱曲线更加平滑的作用,归一化参数越大,得到光谱曲线就会越光滑。数据项和正则项是两个相互竞争的关系。上述所有实验的参数设置都是一致的,即 LM-QSD 方法可很好的平衡这个竞争关系达到噪声抑制能力好和恢复光谱分辨率能力强的双重效果。

此外,文中提供三种算法总的 CPU 时间,如表 1 所示,纵向比较,可以看出,计算的时间跟光谱的长度成正比;横向比较,可以看出,LM-QSD 算法较 TRSD、HOS 两种算法快,这主要得益于 LM-QSD 算法中引进的分裂法。

表 1 三种算法的 CPU 耗时对比(单位:s)

Tab.1 Comparison of computing time between TRSD, HOS and proposed methods (Unit: s)

Spectra	TRSD <sup>[6]</sup>	HOS <sup>[5]</sup>	LM-QSD
D(+)-Glucopyranose	15.365	17.524	4.365
Methyl	25.625	20.152	5.574



## 4 结 论

针对光谱退化问题,文中提出了一种能同时恢复光谱结构又能抑制光谱噪声的方法。将退化光谱中恢复出真实光谱的问题转化成最大后验概率求解问题,推导出拉曼光谱恢复的变分模型。其中第一项似然项对噪声建模,充分利用了光谱在强度变化不同而受到噪声影响不同的特点,对不同强度变化施加不同的权重;第二项先验项,假设光谱的光滑性满足拉普拉斯马尔可夫先验。此模型采用分裂Bregman迭代法来求解,速度快,效率高,并且克服了常规算法中分母为零的问题。对比实验结果证明了该算法的有效性,具有很强的实用价值。

### 参 考 文 献:

- [1] Wang Yingwei, Chen Haobo. Preparation of ZnSe crystal of 0.5–22  $\mu\text{m}$  wide spectrum[J]. *Infrared and Laser Engineering*, 2006, 35(S2): 396–398. (in Chinese)  
王瑛伟,程灏波. 0.5~22  $\mu\text{m}$  宽光谱 ZnSe 单晶窗口制备[J]. 红外与激光工程, 2006, 35(S2): 396–398.
- [2] Zhang Meng, Zhao Huijie, Li Na. Analysis of the influence of hyperspectral spectral resolution on the mineral recognition [J]. *Infrared and Laser Engineering*, 2006, 35(S4): 493–498. (in Chinese)  
张萌,赵慧洁,李娜. 高光谱数据光谱分辨率对矿物识别的影响分析[J]. 红外与激光工程, 2006, 35(S4): 493–498.
- [3] Yang Huidong, Xu Li, Chen Kexin, et al. Blind deconvolution algorithm for spectrogram super-resolution restoration [J]. *Spectroscopy and Spectral Analysis*, 2005, 25 (9): 1520. (in Chinese)  
杨怀栋,徐立,陈科新,等. 盲目反卷积老谱图超分辨率复原算法[J]. 光谱学与光谱分析. 2005, 25(9): 1520.
- [4] Alsmeyer F, Marquardt W. Automatic generation of peak-shaped models[J]. *Appl Spectrosc*, 2004, 58: 986–994.
- [5] Yuan J, Hu Z. Gauss-Newton based kurtosis blind deconvolution of spectroscopic data [J]. *Chin Opt Lett OSA*, 2006: 490–492.
- [6] Ottaway J, Kalivas J H, Andries E. Spectral multivariate calibration with wavelength selection using variants of tikhonov regularization [J]. *Appl Spectrosc*, 2010, 64: 1388–1395.
- [7] Lórenz-Fonfría V A, Padrós E. Maximum entropy deconvolution of infrared spectra: use of a Novel Entropy Expression Without Sign Restriction [J]. *Appl Spectrosc*, 2005, 59: 474–486.
- [8] Meng L J, Ramsden D. An inter-comparison of three spectral-deconvolution algorithms for gamma-ray spectroscopy [J]. *Nuclear Science IEEE Transactions on*, 2000, 47: 1329–1336.
- [9] Senga Y, Minami K, Kawata S, Minami S. Estimation of spectral slit width and blind deconvolution of spectroscopic data by homomorphic filtering [J]. *Appl Opt*, 1984, 23: 1601–1608.
- [10] Goldstein T, Osher S. The split bregman method for L1 – regularized problems[J]. *SIAM J Imag Sci*, 2009, 2: 323–343.