

LAMOST 二维光纤光谱抽谱方法的研究^{**}

崔 波 叶中付[†] 白仲瑞

(中国科学技术大学信号统计处理研究室 合肥 230027)

摘要 对 LAMOST 二维光纤光谱数据处理系统中的抽谱方法进行了阐述,分析了抽谱中关键参数采样点选取对结果的影响,并根据高斯分布函数的特性以及实验情况确定了采样点的选取范围,解决了抽谱结果出现负值的问题,进一步针对大噪声背景下抽谱存在误差较大的问题,提出了基于频域滤波思想的改进抽谱方法. 首先利用快速傅立叶变换和低通滤波器将影响光纤实际轮廓的尖锐噪声滤除,然后再进行正常的抽谱. 利用 LAMOST 二维光纤光谱数据处理系统提供的模拟数据进行了实验测试,结果表明改进抽谱方法的可行性与有效性.

关键词 望远镜; 技术; 频谱仪

中图分类号: P111; **文献标识码**: A

1 引言

LAMOST(Large Sky Area Multi-Object Fiber Spectroscopic Telescope)是我国在建的大天区多目标光纤光谱望远镜,它的建成将成为世界上口径最大和同时观测目标数最多的光纤光谱望远镜. 二维光纤光谱数据处理是该望远镜系统的一个重要组成部分,主要包括 CCD 改正、平场改正、宇宙线剔除、目标谱抽取(下文简称抽谱)、波长定标、减天光、流量定标等内容. 其中,抽谱的任务是从目标观测图像中获得各根光纤在各个波长上的流量值,它在光谱数据处理中占着相当重要的地位,抽谱精度直接影响光谱后续处理的质量.

抽谱方法主要有 3 种:(1) 是孔径抽取方法^[1,2],即对某波长空间方向上的光谱观测数据,在光纤轨迹中心两侧选取一定孔径,然后将此孔径内的数值直接相加即得到光纤在此波长上的流量;(2) 是最优化抽取方法^[2-4],它是在孔径抽取方法的基础上,通过对选定孔径内的流量进行权值评定,然后再进行加权平均得到最后的光纤流量;(3) 是轮廓拟合方法^[1,5,6,7],采用一定的准则,用空间采样点的流量值拟合光纤轮廓形式,得到表征一个光纤流量值的函数表达式,对其进行处理得到最后的光纤流量.

孔径抽取方法的缺点是很难选取合适的孔径,得到的光纤流量信噪比较低,误差较大;最优化抽取方法相比来说,抑制了一定的噪声干扰,改善了所得光纤流量的信噪比,

* 2007-06-21 收到原稿,2007-12-20 收到修改稿

* 国家九五科学工程项目

[†] yezf@ustc.edu.cn

但是孔径的选取仍是一个难题;轮廓拟合方法的效果最好,它不但能提高光纤流量的信噪比,还很好地排除了杂散光和部分宇宙线的影响,减少了交叉效应的作用,抽谱误差相对较小.

本文阐述了LAMOST二维光纤光谱数据处理中采用轮廓拟合方法进行抽谱的原理,确定了抽谱中关键参数采样点的选取范围,解决了抽谱结果出现负值和误差过大的问题. 同时在目标观测图像中可能存在的较大噪声会严重影响光纤实际轮廓,降低抽谱精度,对此作者提出了大噪声背景下基于频域滤波思想的改进抽谱方法.

2 LAMOST 光谱处理中的抽谱方法

2.1 光纤的轮廓形式

所谓光纤轮廓形式是指光纤光谱在 CCD 上成像后,光纤流量在给定波长的空间方向上的扩展形式. 在 LAMOST 数据处理中,假定每根光纤的轮廓都是高斯轮廓. 为了减小因轮廓标准差估算不准带来的误差,常采用如下的多高斯函数来逼近真实的光纤轮廓^[8]:

$$f(x) = \lambda_1 \eta_1(x) + \lambda_2 \eta_2(x), \tag{1}$$

其中 $\eta_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$ 是标准的高斯轮廓, x_0 是光纤轨迹中心, σ 是轮廓标准差; $\eta_2(x) = \frac{(x-x_0)^2}{\sigma^2}\eta_1(x)$ 是 $\eta_1(x)$ 的二阶矩函数,用于改正轮廓宽度. 逼近真实光纤轮廓的示意图见图 1:

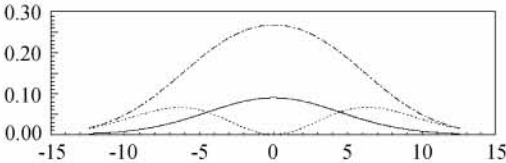


图 1 光纤轮廓示意图
Fig. 1 Spatial profile of one fiber

图 1 中的实线表示 $\eta_1(x)$, 点线表示 $\eta_2(x)$, 点划线表示按(1)式合成的光纤轮廓 $f(x)$, 也即逼近的真实光纤轮廓. $f(x)$ 扩散在 CCD 空间方向的每一像素点上,理论上在 $(-\infty, +\infty)$ 对其进行积分就可得到此根光纤在给定波长上的流量值.

由于

$$\int_{-\infty}^{+\infty} \eta_1(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\} dx = 1, \tag{2}$$

$$\int_{-\infty}^{+\infty} \eta_2(x) dx = \frac{1}{\sigma^2} \int_{-\infty}^{+\infty} (x-x_0)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\} dx = \frac{1}{\sigma^2} \sigma^2 = 1, \tag{3}$$

所以

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \lambda_1 \eta_1(x) dx + \int_{-\infty}^{+\infty} \lambda_2 \eta_2(x) dx = \lambda_1 + \lambda_2. \tag{4}$$

即此光纤在给定波长上的流量值为 $\lambda_1 + \lambda_2$, 这就是最后得到的每根光纤的抽谱结果.

2.2 抽谱原理

LAMOST 二维光纤光谱数据处理中的抽谱采用最小二乘法曲线拟合得到上述每根光纤的轮廓形式 $f(x)$, 并求得每根光纤流量. 最小二乘法是一种在科学计算中既重要又广泛使用的方法^[9,10], 它的一般提法是: 对给定的一组数据 $(x_i, y_i) (i = 1, 2, \dots, M)$, 要求在函数类 $\varphi = \{\varphi_1, \varphi_2, \dots, \varphi_N\}$ 中找一个函数 $y = \phi(x)$, 使误差平方和 χ^2 最小, 即

$$\min(\chi^2) = \min_{\phi(x) \in \varphi} \left(\sum_{i=1}^M [\phi(x_i) - y_i]^2 \right), \quad (5)$$

其中 $\phi(x) = \sum_{k=1}^N \alpha_k \varphi_k(x)$, $N < M$.

在抽谱中, $\phi(x)$ 是光纤轮廓 $f(x)$ 与部分杂散光的叠加, 通过选取每根光纤中心附近的采样点, 按照(5)式规则最终求得的 $f(x)$, 即与实际光纤轮廓“最贴近”的函数形式.

假定光纤总数为 P , 这样在某波长上, P 根光纤轮廓的叠加就构成了此波长空间方向上的目标观测图(见图 2), 此波长空间方向上每一像素点对应的流量值可用下式表示:

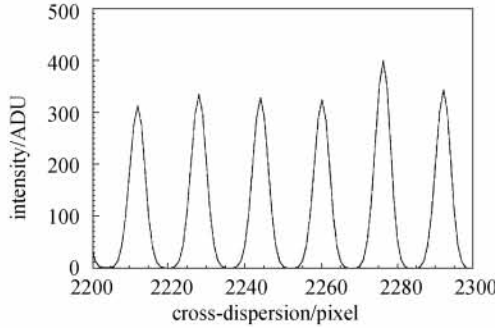


图 2 某波长上目标观测图的一部分

Fig. 2 Section of object spectra image along the cross-dispersion

$$y(x_i) = \sum_{n=1}^P f_n(x_i) = \sum_{n=1}^P [\lambda_{n1} \eta_{n1}(x_i) + \lambda_{n2} \eta_{n2}(x_i)]. \quad (6)$$

其中

$$\eta_{n1}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{(x_i - x_{n0})^2}{2\sigma_n^2}\right\} \quad n = 1, 2, \dots, P, \quad (7)$$

$$\eta_{n2}(x_i) = \frac{(x_i - x_{n0})^2}{\sigma_n^2} \eta_{n1}(x_i) \quad n = 1, 2, \dots, P. \quad (8)$$

x_i, x_{n0} 和 σ_n 分别表示像素点、第 n 根光纤的轨迹中心和第 n 根光纤的轮廓标准差.

同时还要考虑光谱中杂散光的影响. 可用 Q 阶契比雪夫多项式拟合此杂散光, 其形式如下:

$$S(x_i) = \sum_{m=0}^{Q-1} h_m T_m(x_i). \quad (9)$$

其中, $S(x_i)$ 是像素点 x_i 上的杂散光值, h_m 是契比雪夫多项式的系数, $T_m(x)$ 是对应的契比雪夫多项式函数.

最后得到：

$$y(x_i) = \sum_{n=1}^P f_n(x_i) + S(x_i) = \sum_{n=1}^P [\lambda_{n1} \eta_{n1}(x_i) + \lambda_{n2} \eta_{n2}(x_i)] + \sum_{m=0}^{Q-1} h_m T_m(x_i). \tag{10}$$

为讨论方便, 可将(10)式统一为下式：

$$y(x_i) = \sum_{k=1}^N \alpha_k \varphi_k(x_i). \tag{11}$$

这样求 λ_{n1} 和 λ_{n2} 的问题就转化为求 α_k 的问题.

由于噪声(如泊松噪声)或其他原因, 空间方向上的每一像素点值对光纤流量的贡献是有区别的, 因此在拟合计算时每个点的权重应该区别对待, 这样利用加权最小二乘法会得到更精确的抽谱结果. 以像素倒方差作为权重, 最后得到加权最小二乘法的目标函数为^[8]：

$$\chi^2 = \sum_{i=1}^M w(x_i) \left\{ y(x_i) - \sum_{k=1}^N \alpha_k \varphi_k(x_i) \right\}^2. \tag{12}$$

其中, $w(x_i)$ 即像素倒方差, M 是采样点的个数, $M > N$.

为使 χ^2 取得最小值, 求(12)式关于 $\alpha_l (l = 1, 2, \cdots, N)$ 的偏导并令其等于 0, 有：

$$\sum_{i=1}^M \left\{ w(x_i) \left[y(x_i) - \sum_{k=1}^N \alpha_k \varphi_k(x_i) \right] \right\} \varphi_l(x_i) = 0, l = 1, 2, \cdots, N. \tag{13}$$

进一步得

$$\sum_{i=1}^M w(x_i) y(x_i) \varphi_l(x_i) = \sum_{k=1}^N \left\{ \alpha_k \sum_{i=1}^M w(x_i) \varphi_k(x_i) \varphi_l(x_i) \right\}. \tag{14}$$

令

$$\sum_{i=1}^M w(x_i) \varphi_k(x_i) \varphi_l(x_i) = \text{cov } ar_{lk}, \sum_{i=1}^M w(x_i) y(x_i) \varphi_l(x_i) = \beta_l, \tag{15}$$

并代入(14)式, 得

$$\sum_{k=1}^N \alpha_k \text{cov } ar_{lk} = \beta_l, l = 1, 2, \cdots, N. \tag{16}$$

写成矩阵形式如下：

$$\mathbf{C} \boldsymbol{\alpha} = \boldsymbol{\beta}, \tag{17}$$

其中 $\mathbf{C} = \begin{bmatrix} \text{covar}_{11} & \cdots & \text{covar}_{1N} \\ \vdots & \ddots & \vdots \\ \text{covar}_{N1} & \cdots & \text{covar}_{NN} \end{bmatrix}, \boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T, \boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_N]^T$. 利用

Cholesky 分解原理即可求出方程组 $\mathbf{C} \boldsymbol{\alpha} = \boldsymbol{\beta}$ 的解^[11].

得到解 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T$ 后, 可以进一步获得第 n 根光纤在给定波长上的 λ_{n1} 和 λ_{n2} 参数以及流量 $\lambda_{n1} + \lambda_{n2}, n = 1, 2, \cdots, P$. 同理可求得每根光纤在其他波长上的流量. 至此, 抽谱的核心算法(即利用加权最小二乘法曲线拟合得到光纤的轮廓函数及其流量)已经实现.

2.3 采样点的选取

在抽谱过程中, 用于曲线拟合的采样点选取(即 x_i 的选取)至关重要, 它是抽谱方法中的关键参数. 若采样点选取较少则可能会丢失原目标谱中的很多信息, 抽谱质量很低;

若采样点选取过多则可能会增加不必要的运算时间,还可能会带来更多噪声,影响抽谱精度.在空间方向上,以前我们只选取每根光纤轨迹中心两侧各 2σ (σ 即是上述的轮廓标准差)范围内的点作为采样点,实际计算中发现这个范围比较小,造成了光纤流量出现很多负值和零值的情况.

根据高斯分布的“ 3σ 规则”^[12],一般认为采样点选取在轨迹中心两侧各 3σ 范围内可以满足要求.但是实际计算发现,因最小二乘法曲线拟合的抗噪性,采样点较多带来的误差并不大,而采样点较少会带来很大的误差;同时考虑由于 σ 估算不准确以及可能存在较严重的交叉效应,因此采样点选取范围应该稍大一些.从(10)式可见,像素点 x_i 上的值是所有光纤在此点的真实流量值 $f_n(x_i)$ ($n = 1, 2, \dots, P$) 和杂散光 $S(x_i)$ 的叠加,因此理论上最后求得的 $f_n(x)$ 是第 n 根光纤的真实流量,它不受光纤之间交叉污染的影响.这样,采样点的选取范围越宽,曲线拟合效果将越好.考虑到拟合精确和运算时间这两方面的需求,最后我们将采样点选取在轨迹中心两侧各 6σ 的范围,在实验中获得了较好的抽谱结果.

3 大噪声背景下抽谱方法的改进

从 2.2 节的抽谱原理可以看出,表征光纤轮廓的目标拟合函数 $\phi(x)$ 仅考虑了干净的光纤光谱和杂散光的影响,而没有考虑噪声因素.二维光纤光谱数据存在着来自 CCD 的噪声(如泊松噪声)或其他噪声^[13,14].尽管最小二乘法曲线拟合有一定的抗噪能力,但在大噪声环境下的结果仍然不够理想,因此大噪声背景下的抽谱方法是需要改进的.这里的“大噪声”明显影响了光纤的轮廓形式,表现为一种尖锐性.我们先采用信号处理中平滑滤波的方法去除这些噪声^[15],之后再进行正常的抽谱,主要思路如下:

在频域中,由于光谱信号的能量主要集中在低频段内,而尖锐噪声能量主要集中在高频段内,这样,对给定波长上的目标观测图像,先利用快速傅里叶变换(Fast Fourier Transformer, FFT)将其变换到频域,选择特定截止频率的理想低通滤波器对其进行滤波,然后再将滤波输出进行反变换,这样就得到了一个滤除高频成分后的目标图像.对此目标图像再进行抽谱就得到了此波长上的各目标光纤流量.

从上述思路中可以看出,算法的关键在于截止频率的选取.按照滤波理论,若截止频率取得当,就会最大程度上滤除尖锐噪声,同时不会引起有用光谱信号的失真.对于给定波长,所有光纤真实流量在空间方向每一像素点 x_i 上的叠加值可以用 2.2 节中的(6)式表示,空间方向每一像素点未被污染(不含杂散光、宇宙线和各种噪声)的值 $y(x_i)$ 可以用下式表示:

$$y(x_i) = \sum_{n=1}^P f_n(x_i) = \sum_{n=1}^P [\lambda_{n1} \eta_{n1}(x_i) + \lambda_{n2} \eta_{n2}(x_i)].$$

当 σ_n 是第 n 根光纤轮廓真实的标准差时,就无需再对轮廓宽度进行改正,这样 $y(x_i)$ 可简写为下式:

$$y(x_i) = \sum_{n=1}^P [\lambda_{n1} \eta_{n1}(x_i)]. \quad (18)$$

对该波长上的所有像素点值 $[y(x_1), y(x_2), \dots, y(x_M)]$ 进行 FFT 变换可得:

$$\begin{aligned}
 F(u) &= \frac{1}{M} \sum_{x_i=1}^M y(x_i) \exp[-j2\pi u(x_i - 1)/M] \\
 &= \frac{1}{M} \sum_{x_i=1}^M \left\{ \sum_{n=1}^P [\lambda_{n1} \eta_{n1}(x_i)] \right\} \exp[-j2\pi u(x_i - 1)/M] \\
 &= \sum_{n=1}^P \lambda_{n1} \left\{ \frac{1}{M} \sum_{x_i=1}^M [\eta_{n1}(x_i)] \exp[-j2\pi u(x_i - 1)/M] \right\} \\
 &= \sum_{n=1}^P \lambda_{n1} \{FFT[\eta_{n1}(x_i)]\}. \tag{19}
 \end{aligned}$$

根据(7)式知,此时 x_{n0} 与 σ_n 均已知,所以 $FFT[\eta_{n1}(x_i)]$ 的能量分布范围是确知的,这样由(19)式知 $F(u)$ 的能量分布范围也是确知的,即真实的光纤能量分布范围是确知的,而与光纤流量 λ_{n1} 无关.那么实际计算时就可以先假设每根光纤的流量值均为 1(即 $\lambda_{n1} = 1$),然后按照轮廓扩展形式合成该波长上假定的目标图像数据,并对其进行 FFT 变换,由频域内能量分布范围估计出上述截止频率.在确定频域上的能量分布范围时,需要定义一个比例关系,当依次叠加频域上的像素点值之和大于这个比例时,表明已尽可能地接近了总能量,从而估计出截止频率.通过实验,最后我们把这个比例定为 99.8%,它最多会引起 0.2% 的能量损失,但是对抽谱结果的影响已经很小,精度可以满足实际需求.另外由于此时的 σ_n 并非真实的轮廓标准差,它是在抽取平场谱时估算出来的,所以利用上述原理得到的截止频率并非真实值,我们将采用迭代算法得到一个更接近真实值的截止频率.

综上所述,具体改进算法如下:

(1) 利用已知的中心点 x_{n0} 和标准差 σ_n 构造出每根光纤的标准高斯轮廓 $\eta_{n1}(x)$,并由(18)式合成给定波长上新的目标观测图像,其中 $\lambda_{n1} = 1$.

(2) 对合成的目标图像 $y(x_i)$ 进行 FFT 变换并取模,得到对称的图像,从图像的两边同时向中间对像素点值进行叠加求和,找到叠加之和刚好等于或大于所有像素点值总和 99.8% 时对应的那两点 μ_1 和 μ_2 ,即是所要求的初始截止频率.

(3) 对此波长上的原目标图像,进行 FFT 变换后,分别将实部分量和虚部分量介于 μ_1 和 μ_2 之间的点值置零,然后再合并两分量,并反变换回去得到滤波后的目标图像,对此目标图像再进行 2.2 节的抽谱计算得到每根光纤新的流量 $\lambda_{n1} + \lambda_{n2}$ 以及新的标准差 σ_n .

(4) 由(3)中得到的 σ_n 再次重构每根光纤的标准高斯轮廓 $\eta_{n1}(x)$,并合成此波长上新的目标图像为 $y(x_i) = \sum_{n=1}^P (\lambda_{n1} + \lambda_{n2}) \eta_{n1}(x_i)$, λ_{n1} 和 λ_{n2} 也是上述(3)中得到的结果.重复步骤(2)得到新的截止频率 μ'_1 和 μ'_2 ,若 $|\mu'_1 - \mu_1| > \Delta_\mu$ (Δ_μ 是根据估计精度要求选取的门槛值),则重复步骤(3)、(4);否则退出,此时 λ_{n1} 与 λ_{n2} 之和即是给定波长上的抽谱结果.

4 实验

利用模拟程序(模拟程序是由作者所在实验室承担的 LAMOST 项目的一部分)产生

模拟数据(包括目标光纤光谱数据以及每根光纤的真实流量值等)对以上抽谱方法进行实验验证. 实验中的抽谱质量由抽谱结果与光纤真实流量值之间的均方根值(Root-Mean-Square, RMS)来衡量.

4.1 采样点选取对抽谱结果的影响

采样点分别选取在轨迹中心两侧各 2σ 和 6σ 范围内, 得到某根光纤的流量值与此光纤真实流量值对比见图 3.

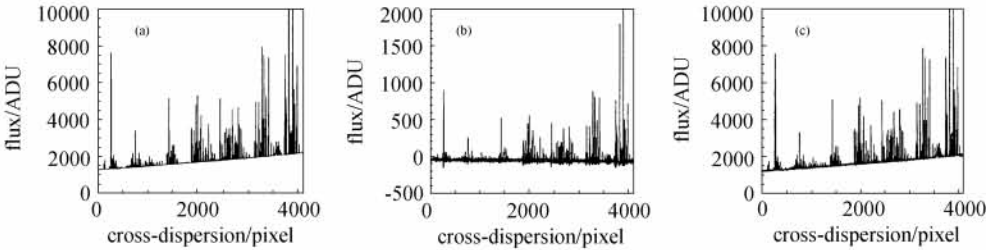


图 3 采样点选取范围不同得到的某根光纤的流量与光纤真实流量对比
Fig. 3 The flux comparison of one fiber between two different scopes of sampling point

其中, 图 3(a)是模拟程序产生的某光纤的真实流量, (b)是采样点选取在轨迹中心两侧 2σ 范围内得到的此光纤的流量, (c)是采样点选取在轨迹中心两侧 6σ 范围内得到的此光纤的流量. 与(a)图相比, 可以看到(b)图中的光纤流量有很多负值和零值, 某些发射线的位置也已经改变, 这是不正确的; 而(c)图中的光纤流量形状非常接近于(a)图. 得到的 250 根光纤流量对应的均方根值(抽谱质量)对比见图 4.

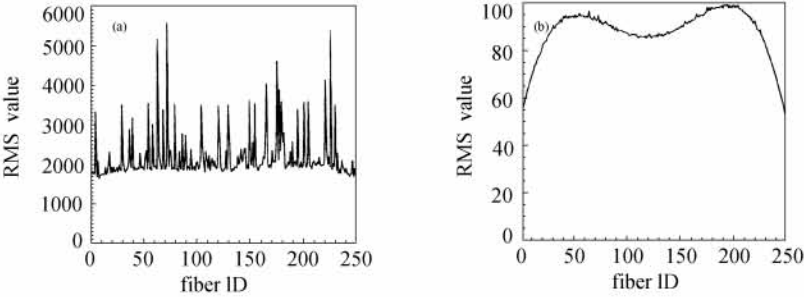


图 4 采样点选取范围不同所得到的光纤流量对应的均方根值对比
Fig. 4 The RMS value comparison of flux between two different scopes of sampling point

其中图 4(a)是采样点选取在轨迹中心两侧 2σ 范围内抽谱结果对应的均方根值, (b)是采样点选取在轨迹中心两侧 6σ 范围内抽谱结果对应的均方根值. 可见(b)的结果要比(a)好很多.

4.2 大噪声背景下改进抽谱方法的结果

我们定义目标观测图像每个波长上的信噪比为 $SNR = 10\lg \frac{\sum_i S_i^2}{\sum_i N_i^2}$, 其中 S_i 和 N_i 分别为此波长在空间方向上每一点的光谱有用信号值和噪声值. 那么整幅目标观测图像的

信噪比由所有波长上的平均值得到,用 $\text{avg}(\text{SNR})$ 表示.

由模拟程序产生 3 幅目标观测图像,它们的信噪比分别为 $\text{avg}_1(\text{SNR})=54.53$ 、 $\text{avg}_2(\text{SNR})=36.19$ 、 $\text{avg}_3(\text{SNR})=14.20$,在某一波长上的光纤轮廓形式见图 5.

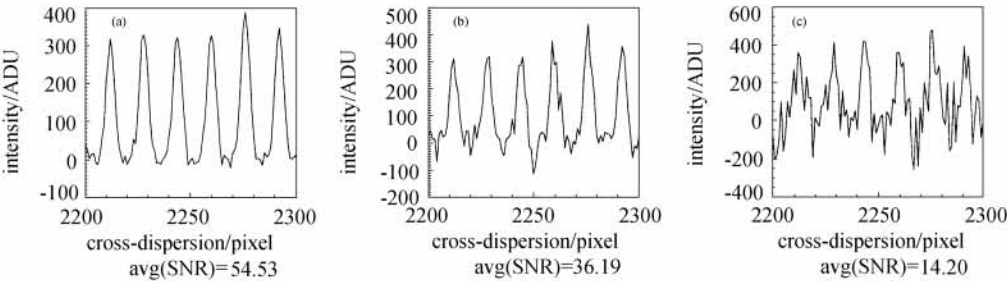


图 5 噪声背景下的目标光纤光谱在同一波长上的轮廓形式
Fig. 5 Three groups of object spectra image with noise along the cross-dispersion

利用第 3 节描述的算法对图 5 的 3 幅目标观测图像进行滤波并抽谱. 根据反复实验结果,算法中迭代条件的门限值取 $\Delta_{\mu}=40$. 以对 $\text{avg}_2(\text{SNR})=36.19$ 的目标观测图像处理为例说明截止频率的选取,见图 6.

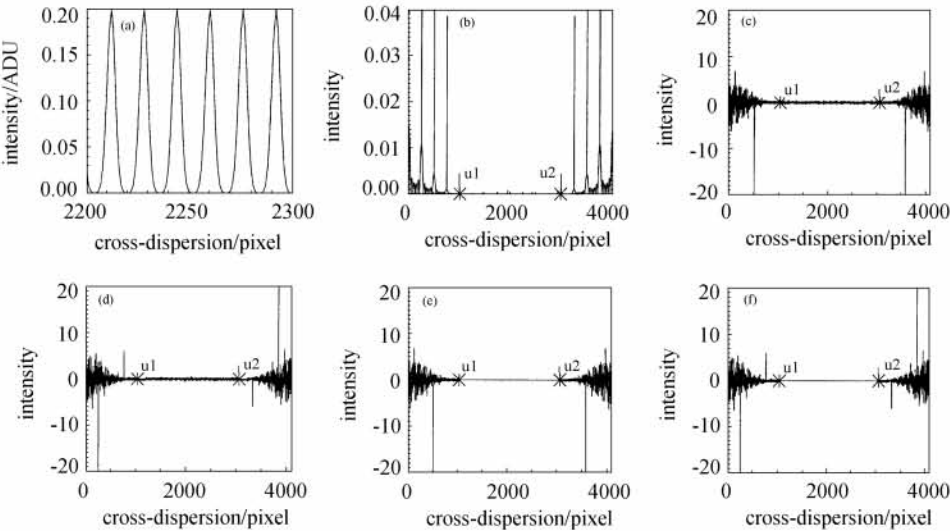


图 6 截止频率的选取
Fig. 6 The steps of selecting cutoff frequency

其中,图 6(a)是假设光纤流量值为 1 时合成的目标图像,(b)是对(a)图进行 FFT 变换后取模并找到截止频率 μ_1 、 μ_2 的幅频图像,(c)、(d)分别是对原始目标图进行 FFT 变换后的实部分量和虚部分量,(e)、(f)分别表示对(c)和(d)进行滤波(截止频率 μ_1 和 μ_2 之间的数据置零)后的结果. 最后得到对原始 3 幅目标观测图像滤波后同一波长上的光纤轮廓,如图 7 所示.

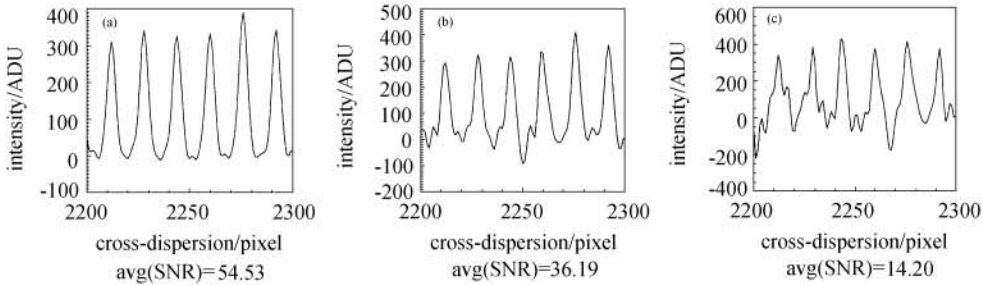


图 7 滤波后的目标光纤光谱在同波长上的轮廓形式

Fig. 7 Three groups of object spectra image along the cross-dispersion after filtering noise out

之后再对图 7 中的目标观测图像进行正常的抽谱即可。

比较图 5 与图 7 中的每幅目标观测图像,可见采用滤波方法仅仅去除了那些尖锐的噪声,并没有明显造成有用光谱信号的失真。

对图 5 所示的三幅目标观测图像用原抽谱方法和改进的抽谱方法进行抽谱,得到 250 根光纤流量对应的均方根值,分别见图 8。

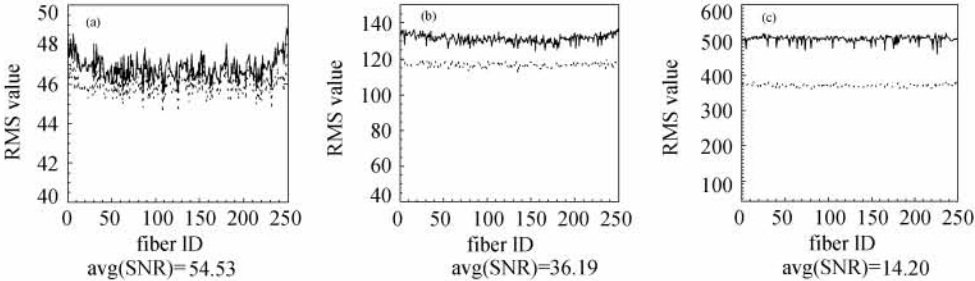


图 8 3 幅目标观测图像抽谱结果对应的均方根值比较

Fig. 8 The RMS value comparison of flux from the three spectrum images

其中,图 8(a)、(b)、(c)中上面的实线均表示未滤除噪声的抽谱方法得到的均方根值,下面的点线均表示滤除噪声的改进抽谱方法得到的均方根值.从图中可见,在各种信噪比条件下改进抽谱方法的性能均优于原抽谱方法,信噪比越低,改善效果越明显。

另外考察极端情况下上述滤波方法对原始目标观测图像的影响.由于 LAMOST 的 CCD 图像中最亮的源与最暗的源相差 5.6 个星等,我们作了相邻光纤相差 5.6 个星等的极端情况下,上述滤波方法的效果实验,原始目标观测图像及其滤波后的目标观测图像在某波长上的光纤轮廓分别见图 9 所示。

其中,图 9(a)是极端情况下含有 3 根最亮光线的原始目标观测图像在某波长上的光纤轮廓,其相邻光纤和其他光纤的流量相对太弱,在图中已经不能明显分辨出,将在图 (c)中另外给出,(b)是对(a)进行滤波后的光纤轮廓,(c)即是对(a)中流量相对较弱的光纤轮廓,(d)是对(c)进行滤波后的光纤轮廓.经计算,上述滤波前、后的光纤光谱同没有加入噪声的光纤光谱之间的均方根误差分别为 262.419 和 198.776.从这个结果以及图 9 的比较来看,极端情况下上述滤波方法并没有明显改变原始光谱信号。

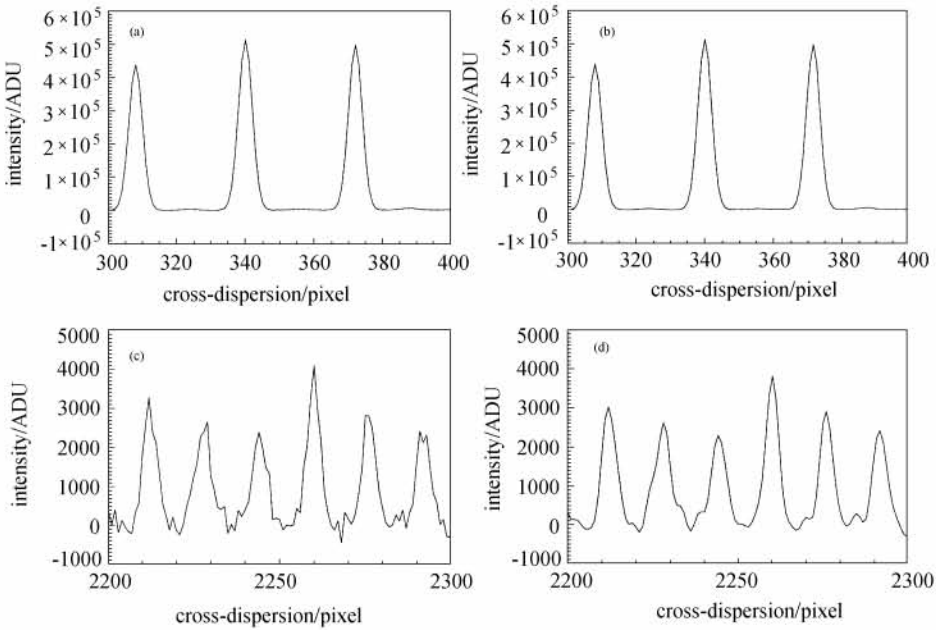


图 9 极端情况下光纤光谱在某波长上滤波前后的轮廓形式

Fig. 9 Object spectra image along the cross-dispersion before and after filtering noise out in extreme cases

5 结 论

本文阐述了LAMOST二维光纤光谱数据处理中的抽谱方法,其中采样点的选取是一个关键问题,综合考虑速度和精度的需求,我们选取了光纤轨迹中心两侧各 6σ 范围内的点作为采样点,解决了光纤流量出现负值的问题,有效地改善了抽谱质量。

对大噪声背景下的目标观测图像,我们提出了先滤波后抽谱的方法,算法核心是基于FFT的迭代算法。模拟实验表明,这种改进抽谱方法可有效地去除目标观测图像中的尖锐噪声,提高抽谱精度,并且没有歪曲有用光谱信号;同时在小噪声的情况下,其性能得到改善。另外在相邻光纤相差5.6个星等的极限情况下,上述滤波方法也没有明显改变原始光谱信号,在大噪声背景下改进抽谱方法是可取的。

参 考 文 献

1 Sanchez S F. *Astronomische Nachrichten*, 2006, 327: 850—861
2 Horne Keith. *PASP*, 1986, 98: 609—617
3 Marsh T R. *PASP*, 1989, 101: 1032—1037
4 Robertson J G. *PASP*, 1986, 98: 1220—1231
5 Roth M M, Kelz A, Fechner T, et al. *PASP*, 2005, 117: 620—642
6 Piskunov N E, Valenti J A. *A&A*, 2002, 385: 1095—1106
7 Blondin S, Walsh J R, Leibundgut B, et al. *A&A*, 2005, 431: 757—771
8 Zhang Hao-tong, Li Shao-ming. LAMOST-TR-BAO-L8-007, 2005

- 9 奚梅成. 数值分析方法(2003 年修订). 合肥: 中国科学技术大学出版社, 2003
- 10 张韵华, 奚梅成, 陈效群. 数值计算方法和算法. 北京: 科学出版社, 2000
- 11 张贤达. 矩阵分析与应用. 北京: 清华大学出版社, 2004
- 12 盛骤, 谢式千, 潘承毅. 概率论与数理统计(第二版). 北京: 高等教育出版社, 2002
- 13 Xiang Zhen. Opto-Electronic Engineering, 2001, 28(6): 66—68
- 14 Zhou Hong-chao, Zhu Ju-bo, Wang Zheng-ming. Chinese Space Science And Technology, 2005, 25(2): 1—5
- 15 徐守时. 信号与系统(修订版). 合肥: 中国科学技术大学出版社, 2003

The Extraction Algorithm for LAMOST Two-dimensional Spectrum Image

CUI Bo YE Zhong-fu BAI Zhong-rui

(Institute of Statistical Signal Processing, University of Science and Technology of China, Hefei 230027)

ABSTRACT Spectra extraction is an important procedure in reducing raw data of fiber-fed spectrographs. The flux of fiber on each wavelength will be obtained by extracting spectra from a two-dimensional image of fiber-fed spectrographs. Spectra extraction affects the later spectrum data processing procedures directly, and its accuracy guarantees the scientific value of spectra. The extraction algorithm is described for extracting one-dimensional object spectra from a two-dimensional spectrum image of LAMOST in this paper. The principle of spectra extraction based on the weighted least square method and polynomial fitting is presented in detail as well as fiber spatial profile. The raw data of fiber-fed spectrographs consist of a collection of spectra distributed along a certain axis of a two-dimensional frame. For each wavelength, spectrum expands along the cross-dispersion axis and covers some width, which is called spatial profile. The profile is considered Gaussian function in the paper. According to Gaussian spatial profile character of the fiber and experiments, the scope of sampling point is well selected in this paper. The good scope of sampling point solves the problem that the fibers' flux extracted is negative. The least square method is often used for scientific computing. To fit Gaussian spatial profile of fiber the weighted least square method is used for spectra extraction of LAMOST, and then the flux of fiber on each wavelength is obtained further.

The existence of noise is usually a problem in data processing. In reducing raw data of fiber-fed spectrographs, strong noise will distort fiber spatial profile seriously. It will lead the bad extraction results from the spectrum image and can't meet the scientific demand. In frequency domain, the spectra energy lies in the low frequency, but the noise energy lies in the high frequency on the contrary. So the improved extraction algorithm based on spectrum analysis in frequency domain is proposed against the affection of strong noise in this paper. In the first step, it filters the sharp noise out by Fast Fourier

Transformer and low-pass filter, and gets a more accurate spatial profile, which has less affection of noise. Then the spectra are extracted with existing algorithm in the later step. The cut-off frequency in improved spectrum extraction is a fatal parameter, and the iteration algorithm for selecting cut-off frequency is described in detail. Finally the improvement is tested with the simulated data provided by LAMOST 2d-pipeline system, the results demonstrate the improved spectrum extraction can remove sharp noise effectively and does not twist the original spatial profile. More accurate extraction results are also achieved. Furthermore, the improvement in extreme cases of 5, 6 magnitudes difference between neighborhood fibers is tested, and the results also demonstrate the feasibility and effectiveness of the proposed method.

Key words Telescopes; Techniques; spectroscopic