

天体光谱信号的连续谱归一化新方法

赵瑞珍¹, 罗阿理²

1. 北京交通大学计算机与信息技术学院, 北京 100044

2. 中国科学院国家天文台, 北京 100012

摘 要 采用事先扣除强谱线的策略, 提出小波变换与样条拟合相结合的方法进行连续谱拟合, 并通过对实际光谱的试验, 与较常用的连续谱拟合方法进行比较, 结果表明文章提出的方法有明显的优越性。另外, 针对多项式拟合等方法得到的连续谱可能会出现负值的情况, 推导出有效的连续谱校正公式。

主题词 天体光谱; 小波变换; 连续谱归一化

中图分类号: TN911.7 **文献标识码**: A **文章编号**: 1000-0593(2006)03-0587-04

引 言

大天区面积多目标光纤光谱望远镜(LAMOST)是“九五”国家重大科学工程, 该项目的突出难点是红移参数的确定和证认问题同天体光谱的分类问题相耦合, 这给光谱的前期处理提出了更高的要求。本文主要针对其中的连续谱归一化问题进行研究。

每条观测得到的光谱都包含连续谱、谱线和噪声。尽管连续谱的信息在很多情况下很有用, 但对于LAMOST来说, 其连续谱是没有标定的, 也就是说它没有经过流量定标的处理, 不反映真实的强度分布信息。因而只有谱线才是有用的特征信息, 必须将连续谱进行归一化处理。

要去除连续谱, 首先要对它进行拟合。连续谱的拟合在交互式天文软件(如 MIDAS, FIGARO, IRAF 等)中都是标准的半自动处理过程。它们首先需要通过人工目测找到一些没有强谱线和吸收带的连续谱“窗口”, 然后再由程序对那些窗口外的点自动插值, 最后由软件自动进行低阶拟合。显然对于LAMOST这样大量的巡天数据, 必须摆脱对人工的依赖, 进行完全的自动处理^[1, 2]。

1 经典的连续谱拟合方法

关于连续谱的提取, 用多项式逼近连续谱是常用的方法之一; 同时还有许多别的方法, 比如形态滤波器, 中值滤波器, 小波变换等。连续谱逼近的结果直接影响到特征峰的提取, 另外在恒星分类中有重要的作用, 因为恒星的连续谱形

状是重要的分类特征。我们分别简要介绍一下较常用的连续谱拟合方法及特点。

1.1 多项式逼近

多项式逼近实际上是多项式拟合, 即寻找一个给定阶数的多项式, 使其最佳逼近连续谱, 采用的是最小二乘法技术。将光谱数据代入多项式, 利用最小二乘法计算待定系数, 即可完成多项式拟合。

采用最小二乘法进行预测或判断, 其关键是基向量的选择, 具体到多项式逼近问题, 确定基向量即成为阶数的选择问题了, 这要根据不同光谱而定。基向量选择的好就能更好地逼近数据点, 反之, 则有可能得到错误的结论, 使得获得的连续谱面目全非。

1.2 形态滤波器

形态滤波器是非线性滤波器中最具代表性和很有发展前途的一种滤波器, 因为它是以数学形态学为理论基础, 具有并行快速实现的特点, 一直受到国内外学者的普遍关注进行了广泛研究。形态滤波器是从数学形态学中发展出来的一类新型非线性滤波器, 它基于信号(图像)的几何结构特性, 利用预先定义的结构元素(相当于滤波窗)对信号进行匹配或局部修正, 以达到提取信号、抑制噪声的目的。

但由于非线性滤波器理论和算法的复杂性和多样性, 形态滤波器至今尚未形成系统的设计方法, 不适合变化剧烈的谱线。

1.3 中值滤波

中值滤波也是一种典型的非线性滤波技术。中值滤波器是在1971年由Jukey首先提出并应用在一维信号处理技术中, 后来被二维图像信号处理技术所引用。中值滤波器的方

收稿日期: 2005-01-06, 修订日期: 2005-04-26

基金项目: 国家自然科学基金(60202013; 60402041)和北京交通大学科技基金(2005SM011)资助

作者简介: 赵瑞珍, 1975年生, 北京交通大学计算机与信息技术学院副研究员

法简单易行,而且效果基本不错,是一种比较实用的方法。

中值滤波对脉冲噪声的去除很有效。脉冲噪声是指大而短暂的正值或负值的干扰。只要所选择的窗口大小至少是脉冲宽度的两倍,移动中值就很适合去除那些充分分散的噪声脉冲而保留彼此靠近的脉冲。由于去除了类似脉冲噪声的谱线影响,得到的移动中值基本上能保留连续谱的形状,但往往还不够平滑,就需要进行多次不同窗口大小的中值滤波。

中值滤波较适合于对发射线较强的活动星系或活动星系核光谱的连续谱提取,因为它们的发射谱线较类似脉冲噪声,但对于吸收型光谱如正常恒星和正常星系,则会由于吸收带的影响使得中值的位置与真实连续谱不一致,造成效果不好。另外,中值滤波中,其窗口大小是固定的,这就无法满足所有种类的光谱要求。如何采用动态的自适应窗口是一个难题。

2 小波滤波方法与其他方法的结合

理论上讲,只要选择一适当的带通滤波器,即可消除天体光谱信号中噪声的干扰,从而较好地拟合出连续谱。然而,由于光谱信号形态各异,很难找到这样一个统一的滤波器。近年来兴起的小波理论为处理上述问题提供了新的方法,由于小波变换具有多分辨特性,且可以同时实现低通和带通滤波,因此可以根据需要对原始信号进行若干次滤波,直到去除所有无用信号^[3-5]。

有关小波变换的原理与方法,有许多文献讨论过,我们不再仔细论述。在此,仅给出小波变换快速算法流程中信号在各尺度所占频带示意图,如图 1 所示。原始信号频带为 $0 \sim f_i$ kHz, 设 h_j 和 g_j 分别为理想的半带低通和半带高通滤波器, h_j 和 g_j 将信号 $S_{2^j}^d f$ 分解为低半带 $S_{2^{j+1}}^d f$ 和高半带 $W_{2^{j+1}}^d f$ 。比如,原始信号 $S_{2^0}^d f$ 占据频带为 $0 \sim f_i$ kHz, 经 h, g 滤波之后, $S_{2^1}^d f$ 占据频带 $0 \sim f_i/2$ kHz, $W_{2^1}^d f$ 则为 $f_i/2 \sim f_i$ kHz。

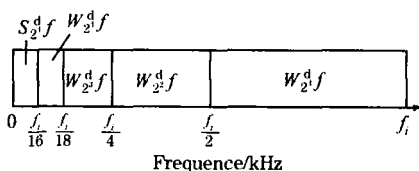


Fig 1 Frequency bands of a signal in wavelet domain

假设 LAMOST 观测并经过初步处理得到的光谱信号为

$$f(\lambda) = s(\lambda) + n(\lambda) \quad (1)$$

其中, $s(\lambda)$ 为真实的光谱信号, $n(\lambda)$ 为干扰噪声, 我们欲从 $f(\lambda)$ 中把 $n(\lambda)$ 分离出去, 并进一步拟合其连续谱。

对 $f(\lambda)$ 连续做小波分解之后, 其各尺度上的小波系数集中了不同频率成分的信号能量。很显然, 连续谱属于低频信息, 它应集中在图 1 中的 $S_{2^{j+1}}^d f$ 频带; 而谱线和噪声属于高频信息, 它们包含在 $W_{2^{j+1}}^d f$ 中的各个频带。这就达到了低通和带通的滤波效果。比如我们要进行连续谱拟合, 则丢掉 $W_{2^{j+1}}^d f$ 各频带中的信息, 只利用 $S_{2^{j+1}}^d f$ 中的信息进行小波重构, 即可得到较为满意的连续谱。

所谓连续谱归一化, 是指将原始光谱除以拟合得到的连续谱这一过程, 归一化的结果是得到谱线光谱。需要指出的是, 该谱线光谱中依然是含有噪声的。要想有效地提取谱线, 必须先对谱线光谱进行去噪处理。

将光谱信号做小波变换之后, 可将低频信息和高频信息分离开来, 从而将体现连续谱的低频信息单独提取出来。然而如此得到的连续谱由于受到强谱线的影响, 不够准确, 如果不加以处理则与真实的连续谱相差较大。因此, 我们在做小波变换的过程中, 尽量将强谱线扣除掉, 并且对所得到的连续谱进行多项式插值或样条拟合, 使其尽量逼近真实的连续谱。即基本步骤为, 1) 先扣除强线或坏线(如天光线或宇宙线等); 2) 小波滤波; 3) 多项式或样条拟合。

将光谱做一层小波变换, 首先在特定的波长位置判断是否有天光线或宇宙线等, 若有, 则将其去除, 并在相同位置处对原光谱进行修正。将修改之后的光谱重新做一层小波变换, 检测在小波域内有无强谱线, 若有, 则采用同样的方法将其去除, 并对原始光谱做相应修改。这一过程需要反复进行, 直到不再发现有强谱线为止。显然, 此时的光谱中只包含有连续谱和部分噪声信息。将其做小波变换, 取出低频系数, 则其中主要是连续谱信息了。为了得到较为理想的连续谱, 我们需要对低频系数进行数据拟合, 具体可采用多项式拟合或样条拟合。

以下给出本文算法的主要步骤:

1. 记 $f_0(\lambda) = f(\lambda)$, 令 $i = 1$;
2. 将 $f_{i-1}(\lambda)$ 做一层小波变换得到 $s_i(\lambda)$ 和 $w_i(\lambda)$, 计算 $w_i(\lambda)$ 的方差 σ , 记录 $w_i(\lambda)$ 中绝对值大于 3σ 的波长位置, 然后在相应的位置以 $s_i(\lambda)$ 的值替换 $f_{i-1}(\lambda)$ 的值, 得到 $f_i(\lambda)$ 。若 $f_i(\lambda) = f_{i-1}(\lambda)$, 转 3; 否则, 令 $i = i + 1$, 转 2;
3. 将 $f_i(\lambda)$ 作为原始信号进行 4~5 层小波变换, 然后将全部高频系数置零后再进行小波反变换, 得到 $\tilde{f}(\lambda)$;
4. 将 $\tilde{f}(\lambda)$ 进行样条拟合或多项式拟合后即可得较为满意的连续谱 $c(\lambda)$ 。

该算法稳定性较好, 适合于低信噪比的信号。从计算量的角度考虑, 虽然算法中第二步需要反复迭代, 但一般迭代次数不会很多即可满足要求。

多项式拟合实际上就是最小二乘法, 关键是要确定多项式的阶数, 不同光谱对阶数的要求不同。我们可以通过光谱信号的小波系数及其导数的能量大小来选择多项式的阶数。若选好阶数, 则多项式的形式就可写出来, 系数为待定系数。将光谱数据代入多项式, 然后利用最小二乘法可确定出多项式的待定系数, 从而完成多项式拟合。

样条拟合中的关键是如何确定型值点的问题, 尤其是对于恒星的不同子类, 其选择方案应该不同。

3 连续谱校正

在实际中, 不管是采取什么样的拟合方法, 都有可能会出现拟合出来的连续谱中出现负值的情况, 这在物理上是没法解释的。为了避免这种情况发生, 我们需要在技术上做一定处理, 称为连续谱校正。具体地讲, 如果拟合出的连续谱

$C(\lambda)$ 中出现负值, 即 $\min\{C(\lambda)\} < 0$, 则保持 $C(\lambda)$ 的最大值不变, 将其最小值校正为正值。记 $C_{\max} = \max\{C(\lambda)\}$, $C_{\min} = \min\{C(\lambda)\}$, 当 $C_{\min} < 0$ 时, 我们推导出连续谱校正公式

$$\mathcal{C}(\lambda) = C(\lambda) - (C_{\min} - \delta \cdot C_{\max}) \cdot \left[1 - \frac{C(\lambda) - C_{\min}}{C_{\max} - C_{\min}} \right] \quad (2)$$

其中参数 δ 为一较小的正数, 如可以取为 0.01, 它的引入避免了校正之后出现零值的情况, 即当某一 $C(\lambda_0) = C_{\min} < 0$ 时, 校正之后有 $\mathcal{C}(\lambda_0) = \delta \cdot C_{\max} > 0$ 。

连续谱校正有着重要的意义, 如果出现负值而不进行校正, 则在连续谱归一化的过程中, 当光谱与连续谱相除时, 得到的谱线光谱中会引入很多假线, 给后续的谱线提取与特征计算带来极大的不便。连续谱校正属于物理校正, 不改变连续谱的大致形状和性质。

4 实验结果

我们对本文所述的各种连续谱拟合方法做了实验和比

较, 对同样的光谱分别采用多项式拟合法、中值滤波法、小波变换法以及小波与样条拟合相结合的方法做了处理, 实验结果示于图 2。图中 (a) 为原始光谱; (b)~(e) 依次为采用多项式拟合法、中值滤波法、小波变换法以及小波与样条拟合相结合的方法提取出来的连续谱; (f)~(i) 依次表示利用相应方法进行连续谱归一化 (即原光谱除以连续谱) 之后的结果。

从实验结果可以看出, 采用多项式拟合法提取出来的连续谱过于光滑, 没有很好地体现光谱中较为细节的部分, 从而, 连续谱归一化之后的谱线光谱不够平坦; 中值滤波法得到的连续谱不够光滑, 象是分段函数, 有阶梯效应; 采用小波变换法提取的连续谱较好, 但在强谱线附近依然不够准确; 而小波与样条拟和相结合的方法比较好, 得到的连续谱是较为理想的。

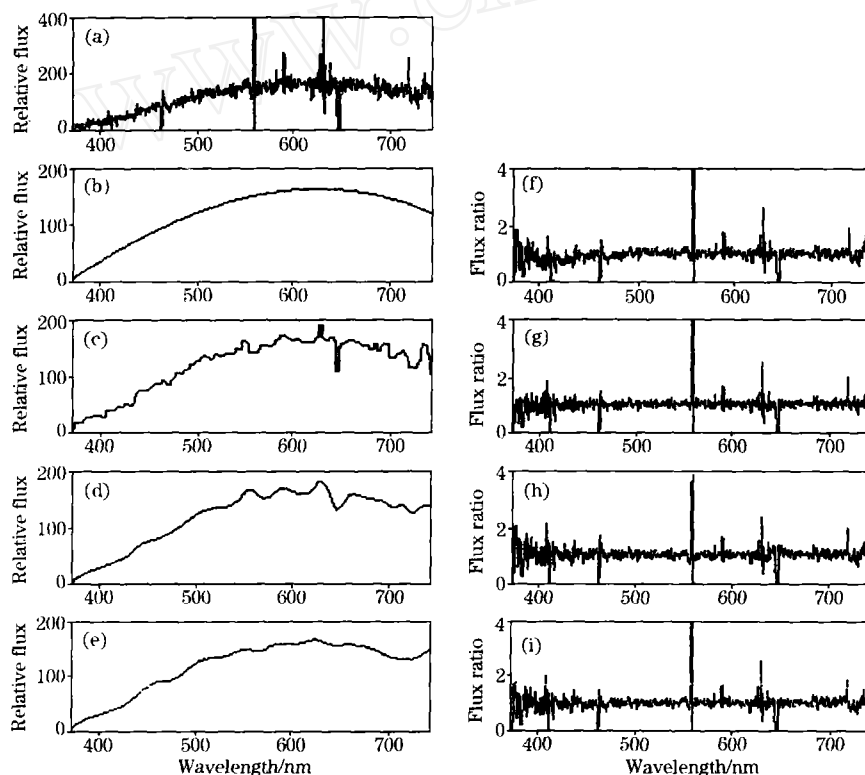


Fig 2 Comparison results of methods for continuum fitting where (a) is original spectrum; (b)-(e) are continua extracted by the methods of polynomial fitting, median filtering, wavelet transforming and the combined method of wavelet and spline fitting respectively; (f)-(i) are the corresponding results of continuum normalization

参 考 文 献

- [1] LUO A-li, ZHAO Yong-heng(罗阿理, 赵永恒). Acta Astrophysica Sinica(天体物理学报), 2000, 20(4): 427.
- [2] ZHAO Rui-zhen, HU Zhan-yi, ZHAO Yong-heng(赵瑞珍, 胡占义, 赵永恒). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(1): 153.
- [3] Faghhih F, Smith M. IEEE Trans. on IP, 2002, 11(9): 1062.
- [4] Daubechies I. Ten Lectures on Wavelets. Philadelphia: SIAM, 1992.
- [5] Mallat S, Hwang W L. IEEE Trans. on IT, 1992, 38(2): 617.

A Novel Method for Continuum Normalization of Astronomical Spectrum Signals

ZHAO Rui-zhen¹, LUO A-li²

- 1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
- 2. National Astronomical Observatory, Chinese Academy of Sciences, Beijing 100012, China

Abstract A combined method of wavelet transform and spline fitting is presented for continuum fitting with the strong spectral lines taken out beforehand. Some comparisons between normal methods and the method presented are made by the experiments on actual spectra. And the experimental results show that our method is superior to others. Besides, an effective continuum revision formula is derived to avoid the appearance of negative continua obtained by some methods such as polynomial fitting and so on.

Keywords Astronomical spectra; Wavelet transform; Continuum normalization

(Received Jan. 6, 2005; accepted Apr. 26, 2005)