AN OBJECTIVE APPROACH TO SPECTRAL CLASSIFICATION

A.J. CONNOLLY AND A.S. SZALAY
Johns Hopkins University

1. Introduction

The next generation of spectroscopic surveys, both Galactic and extragalactic (*e.g.*, SDSS, 2dF), present the challenge of classifying spectra in an efficient and objective manner. The standard approach to this problem has been to visually classify spectra based on a number of spectral features (*e.g.*, the equivalent widths of emission lines). The size of new spectral surveys ($> 10^6$ galaxies) and the desire to compare the luminosity and environments of galaxies with their spectral properties make these techniques infeasible. We describe here an automated classification scheme that is being developed for the SDSS.

2. The Karhunen-Loève Transform

The application of Principal Component Analysis to multivariate astrophysical data has been described in detail in a number of publications (Efstathiou and Fall 1984, Connolly *et al.* 1995). For details of the Karhunen-Loève transform (KL; Karhunen 1947, Loève 1948) and its use in the analysis of spectroscopic data we refer the reader to Connolly *et al.* (1995). Here, we describe an extension to these standard techniques that applies to censored data (*i.e.*, data that contain spectral regions where the flux is unknown).

We can describe an observed spectrum, $f_i^{'}(\lambda)$, in terms the true spectrum, $f_i(\lambda)$, and a mask, $m_i(\lambda)$. The mask is defined to be zero where the data are unknown and one in wavelength regions where the data are secure. If we decompose this spectrum onto a previously defined eigenbasis,

B.J. McLean et al. (eds.), New Horizons from Multi-Wavelength Sky Surveys, 376–378.

https://doi.org/10.1017/S0074180900129079 Published online by Cambridge University Press

$e_j(\lambda)$, then, $f_i'(\lambda) = m_i(\lambda)\sum_{i=1}^{n} y_{ij}e_j(\lambda)$, where, y_{ij}, are the decomposition coefficients.

Clearly the observed spectrum and the eigensystem are not orthogonal over the masked spectral range. Therefore, projecting the observed spectrum onto this eigensystem will lead to biased coefficients. We, therefore, determine an error, E, that describes the difference between $f_i'(\lambda)$ and $\sum_{j=1}^{n} y_{ij}e_j(\lambda)$ over the wavelength region where the data are good (Everson and Sirovich 1995), $E_i = \int d\lambda \sum_i [f_i'(\lambda) - \sum_{j=1}^{n} y_{ij}e_j(\lambda)]^2$. By minimizing E with respect to y_{ij} we can define, $y_i = M_{ij}^{-1}g_j$, where $M_{ij} = \int_{\lambda\forall m(\lambda)=1} d\lambda e_i(\lambda)e_j(\lambda)$ and $g_j = \int_{\lambda\forall m(\lambda)=1} d\lambda f_j'(\lambda)e_j(\lambda)$.

In Figure 1 we show that these corrected coefficients can be used to optimally interpolate across the masked spectral regions. The typical rms deviation between the true and interpolated regions is $\sim$ 10%). By iteratively replacing the masked regions with the interpolated values the above analysis can be extended to the case where we wish to derive the eigensystem from a sample of galaxy spectra that cover different spectral ranges (Everson and Sirovich, 1995).

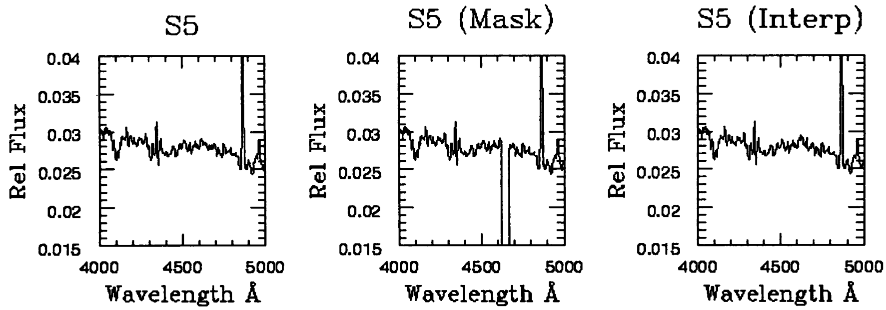

Figure 1. Spectra with missing data, *e.g.*, due to sky lines, can be optimally interpolated across using the derived eigensystem. In this figure (a) shows the original spectrum, (b) the spectrum with a 40 Å region masked out and (c) the interpolated spectrum.

3. A Spectral Classification Scheme

From the decomposition coefficients described above, we can classify galaxy spectra into different spectral types. Figure 6 in Connolly *et al.* (1995) shows this for the Kinney *et al.* (1996) spectral energy distributions. By plotting the mixing angles between the first three eigenspectra we find that galaxies from elliptical to starburst form a linear sequence *i.e.*, to first order, galaxies separate out into a simple monotonic spectral sequence.

To extend this analysis to grouping galaxies into subclasses along this spectral sequence (analogous to the sub groups in stellar classification) requires that we consider individual features within each spectrum (*e.g.*, lines or line complexes). This naturally results in additional dimensions and increased complexity to any classification scheme. While this is straightforward to undertake statistically the question remains how do you describe the classification of a galaxy spectrum in a simple and physical manner.

For the SDSS we choose to approach this problem by building on the work of Morgan and Mayall (1957). In their Yerkes Y system (and revisions thereof) they describe the spectral types of galaxies in terms of stellar classifications (*e.g.*, classes of galaxies go through A, AF, F, G, and K stellar types). In defining the subclasses they consider only a limited spectral range for each galaxy (3850–4100 Å). From the Kinney *et al.* (1996) eigenspectra we have shown that the statistical spectra derived from the KL analysis separate naturally into different stellar types (G, O and A). To subdivide further we apply the KL analysis to smaller spectral regions (*e.g.*, the 4000 Å break). In such a way a hierarchical classification of galaxy spectra can be built up.

4. Conclusions

(1) The Karhunen-Loève transform provides an elegant method for the derivation of eigenfunctions that describe multidimensional datasets. These techniques have been modified to account for data with imperfect spectral coverage, *e.g.*, due to sky lines or different rest-frame wavelength coverage. From this, an optimal interpolation scheme can be derived to account for the masked spectral regions.
(2) Decomposing spectra into whose eigenspectra their coefficients can be used as an objective spectral classification scheme. This natural classification can be described in terms of a one parameter problem and, in the mean, is correlated with independent morphological classifications.
(3) The resolution of any classification can be optimized to a particular range of spectral types by adopting an iterative scheme where we vary the normalizations of the input SEDs.

References

Connolly, A.J., Szalay, A.S., Bershady, M.A., Kinney, A.L. and Calzetti, D. 1995. A.J. 110, 1071.
Efstathiou, G. and Fall, M.S., 1984. Mon. Not. R. astron. Soc., 206 453
Everson, M. and Sirovich, L., 1995. J. Opt. Soc. Am. A, 12, 1657
Karhunen, H., 1947. Ann. Acad. Science Fenn, Ser. A.I. 37
Kinney, A.L., Calzetti, D., Bohlin, R.C., McQuade, K., and Storchi-Bergmann, T., 1994. Astrophys. J., 467, 38
Loève, M., 1948. Processus Stochastiques et Mouvement Brownien, Hermann, Paris, France