

Principal Component Analysis of SDSS Stellar Spectra

Rosalie C. McGurk, Amy E. Kimball, Željko Ivezić

Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195

ABSTRACT

We apply Principal Component Analysis (PCA) to $\sim 100,000$ stellar spectra obtained by the Sloan Digital Sky Survey (SDSS). In order to avoid strong non-linear variation of spectra with effective temperature, the sample is binned into 0.02 mag wide intervals of the $g - r$ color ($-0.20 < g - r < 0.90$, roughly corresponding to MK spectral types A3 to K3), and PCA is applied independently for each bin. In each color bin, the first four eigenspectra are sufficient to describe the observed spectra within the measurement noise. We discuss correlations of eigencoefficients with metallicity and gravity estimated by the Sloan Extension for Galactic Understanding and Exploration (SEGUE) Stellar Parameters Pipeline. The resulting high signal-to-noise mean spectra and the other three eigenspectra are made publicly available. These data can be used to generate high quality spectra for an arbitrary combination of effective temperature, metallicity, and gravity within the parameter space probed by the SDSS. The SDSS stellar spectroscopic database and the PCA results presented here offer a convenient method to classify new spectra, to search for unusual spectra, to train various spectral classification methods, and to synthesize accurate colors in arbitrary optical bandpasses.

Subject headings: stars: abundances – stars: statistics – methods: data analysis – stars: fundamental parameters

1. Introduction

A large number of homogeneously-obtained stellar spectra have recently become available. For example, the Sloan Digital Sky Survey (SDSS) (York et al. 2000) has made publicly available¹ over 460,000 stellar spectra as a part of its Data Release 7 (Abazajian & Sloan Digital Sky Survey

¹See <http://www.sdss.org/dr7>

2008), and Radial Velocity Experiments² (RAVE) may provide up to a million spectra over the next few years. This rapid progress in the availability of stellar spectra re-opens the old question of optimal stellar parameter extraction. For example, the SDSS estimates effective temperature, gravity, and metallicity using a variety of standard methods implemented in an automated pipeline (SEGUE³ Stellar Parameters Pipeline, hereafter SSPP; Beers et al. 2006). A detailed discussion of these methods and their performance can be found in Allende Prieto et al. (2006, 2007) and Lee et al. (2007a,b). The results of different methods implemented in the SSPP are *averaged* to obtain the final adopted values in the SDSS Spectral Parameter Pipeline table (*sppParams*). Although a detailed analysis by Lee et al. (2007a,b) demonstrates that systematic metallicity differences between the methods used in averaging do not exceed ~ 0.1 dex (with random errors in the range 0.1–0.3 dex), it is fair to ask whether a single method could be used to obtain the same level of systematic and random errors, instead of combining different methods with varying error properties.

Principal Component Analysis (PCA) has been demonstrated as a viable tool in solving this classification problem (Connolly et al. 1995; Connolly & Szalay 1998; Bailer-Jones et al. 1998; and references therein). Yip et al. (2004) have developed a PCA-based analysis code specialized to SDSS spectra. Here we use the same code to investigate whether the PCA eigencoefficients are correlated with the metallicity and gravity obtained by the SSPP. Byproducts of this analysis are high signal-to-noise eigenspectra that can be used to generate spectra for any combination of basic stellar parameters (effective temperature, metallicity, and gravity) within the parameter space probed by SDSS. Hence, given an arbitrary spectrum, one can attempt a low-dimensional fit using our library of eigenspectra. Among numerous drivers for such a library, we single out a photometric calibration scheme for the Large Synoptic Survey Telescope (LSST)⁴. LSST plans to use an auxiliary spectroscopic telescope to obtain spectra of standard stars at the same time as the main imaging survey is performed (see Ivezić et al. 2008b). The atmospheric transmission properties, required to photometrically calibrate the imaging survey, will be obtained by simultaneously fitting the stellar spectrum and a sophisticated atmospheric model with six free parameters for each observation. The ability to describe the expected stellar spectra in a low-dimensional continuous space by using a small number of eigencomponents, with eigencoefficients that are not defined on a fixed grid, might increase the fidelity of the fitted model.

In Section 2 we describe our sample selection and the application of PCA to SDSS stellar

²See <http://www.rave-survey.aip.de/rave>

³Sloan Extension for Galactic Understanding and Exploration

⁴See <http://www.lsst.org/>

spectra. We discuss our results in Section 3, and end with a summary in Section 4.

2. Principal Component Decomposition of SDSS Stellar Spectra

2.1. The properties of SDSS spectra

In addition to massive amounts of optical photometry of unprecedented quality, the SDSS has also produced a large spectroscopic database. A compendium of technical details about SDSS can be found on the SDSS web site⁵, which also provides an interface for public data access. Targets for the spectroscopic survey are chosen from the SDSS imaging data based on their colors and morphological properties (Strauss et al. 2002; Eisenstein et al. 2001; Richards et al. 2002). In the spectroscopic survey, stars are targeted either as calibrators or for scientific reasons in specific parts of the four-dimensional SDSS color space (Yanny et al. 2009).

A pair of multi-object fiber-fed spectrographs mounted onto the SDSS 2.5m telescope (Gunn et al. 2006) are used to take 640 simultaneous spectra within a radius of 1.49 degrees, each with wavelength coverage 3800–9200 Å and spectral resolution of ~ 2000 , and with a signal-to-noise ratio of >4 per pixel at $g=20.2$. Spectro-photometric calibration of these spectra is exquisite; for example, the imaging magnitudes and the stellar magnitudes synthesized from SDSS spectra agree with an rms of only ~ 0.05 mag (see Smolčić et al. 2004).

2.2. Sample selection

We begin by selecting bright stars in SDSS Data Release 6 that have colors consistent with the main stellar locus (Lenz et al. 1998; Fan 1999; Finlator et al. 2000), or are found in the regions populated by RR Lyrae stars (Ivezić et al. 2005) and blue horizontal branch stars (Sirko et al. 2004). Stars that are probable white dwarf — red dwarf pairs (Smolčić et al. 2004) or single hot white dwarfs (Eisenstein et al. 2006) are not selected. We only use stars from the sky regions with modest interstellar dust extinction, determined using the interstellar dust maps of Schlegel et al. (1998).

The specific criteria applied to 130,620 entries from the SDSS DR6 version of *sppParams*

⁵See <http://www.sdss.org/>