Applications of Manifold Learning Techniques to Spectral Classification of Quasars

I. Jankov, D. Ilić and A. Kovačević

Department of Astronomy, Faculty of Mathematics, University of Belgrade

INTRODUCTION

Quasars are extremely luminous objects powered by accretion of matter onto a supermassive black hole located in their center. The last three decades of research related to spectral properties of type 1 quasars have revealed that they occupy a specific parameter space, the so-called Eigenvector 1 (E1) (Boroson & Green, 1992; Sulentic, et al., 2000), where they form an elbow-shaped main sequence (MS), analogous to the MS of stars in the H-R diagram. Sulentic et al. (2000) have shown that E1 parameter space could be used to distinguish between different spectral populations of type 1 quasars, namely populations xA, A and B, based on the Hβ line width (FWHM) and the strength of the iron line - the two of four spectral parameters describing E1 (Fig. 1). It is believed that the potential driving mechanism behind the quasar MS, and thus the physical reason behind the need for definition of populations, is the Eddington ratio convolved with the line-of-sight orientation of the source (Marziani, et al., 2001; Shen & Ho, 2014). We propose an improvement to the interpretation of quasar spectral diversity by applying manifold learning technique called locally linear embedding – LLE (Roweis & Saul, 2000) in the context of E1 parameters.

MOTIVATION

The E1 parameter space was first described using the Principal Component Analysis (PCA), which is considered to be a linear dimensionality reduction technique. It was repeatedly demonstrated that PCA is an extremely useful tool for finding meaningful linear relationships in high-dimensional data sets, but it has its shortcomings when applied to inherently non-linear data sets (e.g. galaxy spectra). Manifold learning techniques address this problem by taking into account the non-linear geometry of the data embedded in the original parameter space when calculating low-dimensional projection of that data. The main motivation behind this contribution is to take advantage of manifold learning methods in order to improve our understanding of the quasar main sequence and reveal the information that is potentially lost in the optical plane of the E1 parameter space.

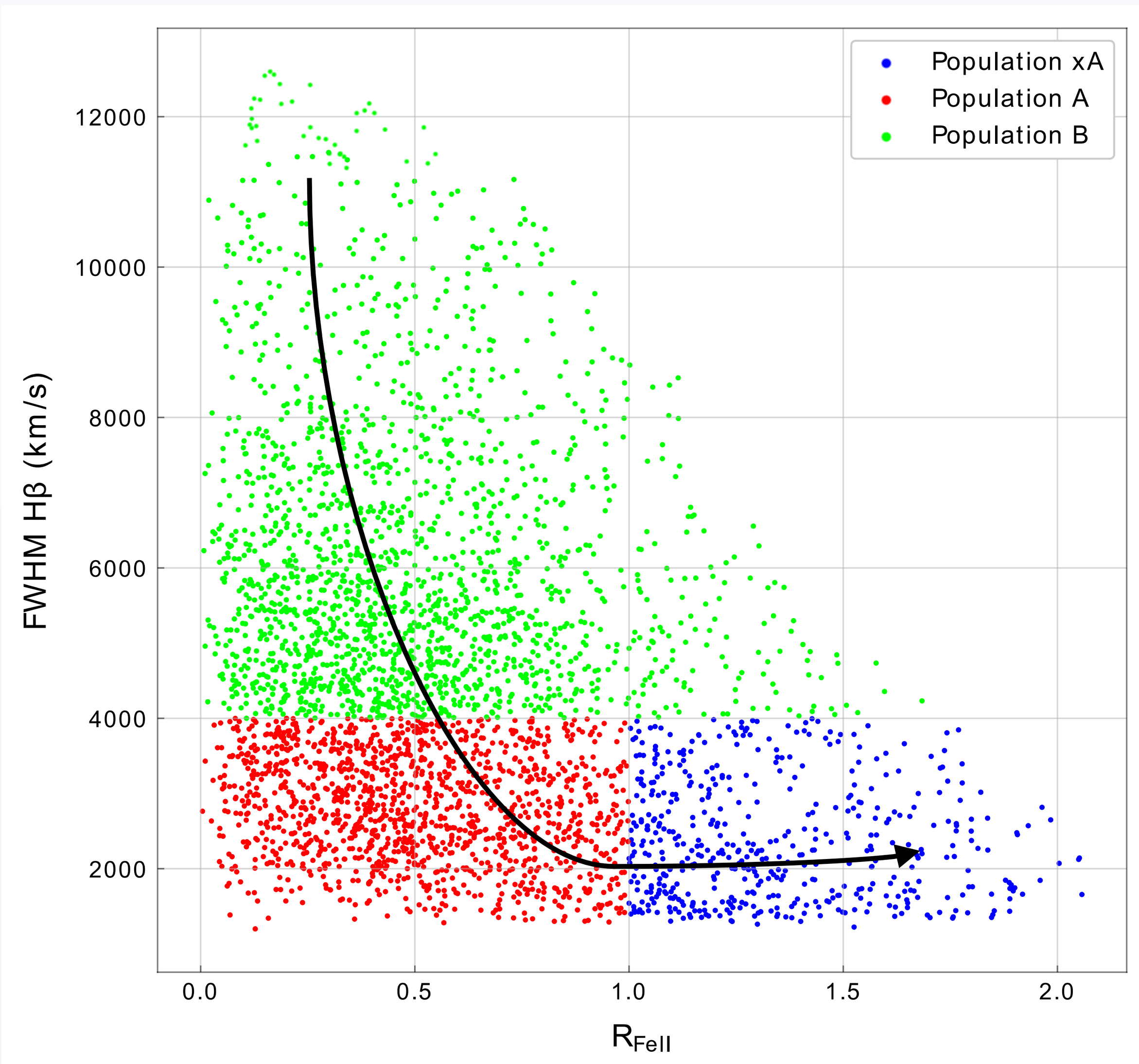

Figure 1. Optical plane of the E1 parameter space for our sample with outliers removed. Iron strength (R_{FeII}) is indicated on the horizontal axis and represents the equivalent width ratio of optical Fe II and broad Hβ emission lines. Vertical axis presents the width of the broad Hβ line. Different spectral populations of quasars are indicated with different colors. The main sequence direction is indicated with an arrow.

DATA ANALYSIS

Sample selection:

We selected a sample of low-redshift (z < 0.39) type-1 quasars from the Sloan Digital Sky Survey Data Release 7 quasar catalog (Shen, et al., 2011). The sample contained only objects with measured spectral properties possibly relevant to E1 parameter space (Hα, Hβ, [O III] λ5007 Å, optical iron and continuum luminosity). LLE algorithm can be very sensitive to outliers, so we removed those from E1 optical plane by eliminating low density regions which were identified using kernel density estimation (e.g. Silverman, 1986), leaving us with the final sample of 3720 objects.

Parameter selection for LLE:

The LLE algorithm requires only one free parameter – the number of nearest neighbors (k). The algorithm uses this parameter to learn the local geometry of the manifold. In order to select optimal value for k we calculated a matrix of pairwise geodesic distances of the original parameter space, as well as of the resulting low-dimensional parameter space. Next, we compared the matrices using the modified RV coefficient (Smilde, et al., 2008). This process was repeated for each value of k in the range $4 \leq k \leq 30$ and the optimal value was found to be $k_{opt} = 12$, the one with the highest value of the RV coefficient.

RESULTS & CONCLUSIONS

We applied LLE algorithm ($k_{opt} = 12$) to our sample of 3720 quasars described by ten spectral parameters (Table 1). Output dimension was chosen to be three ($n = 3$) in order to have a visual representation of the resulting space.

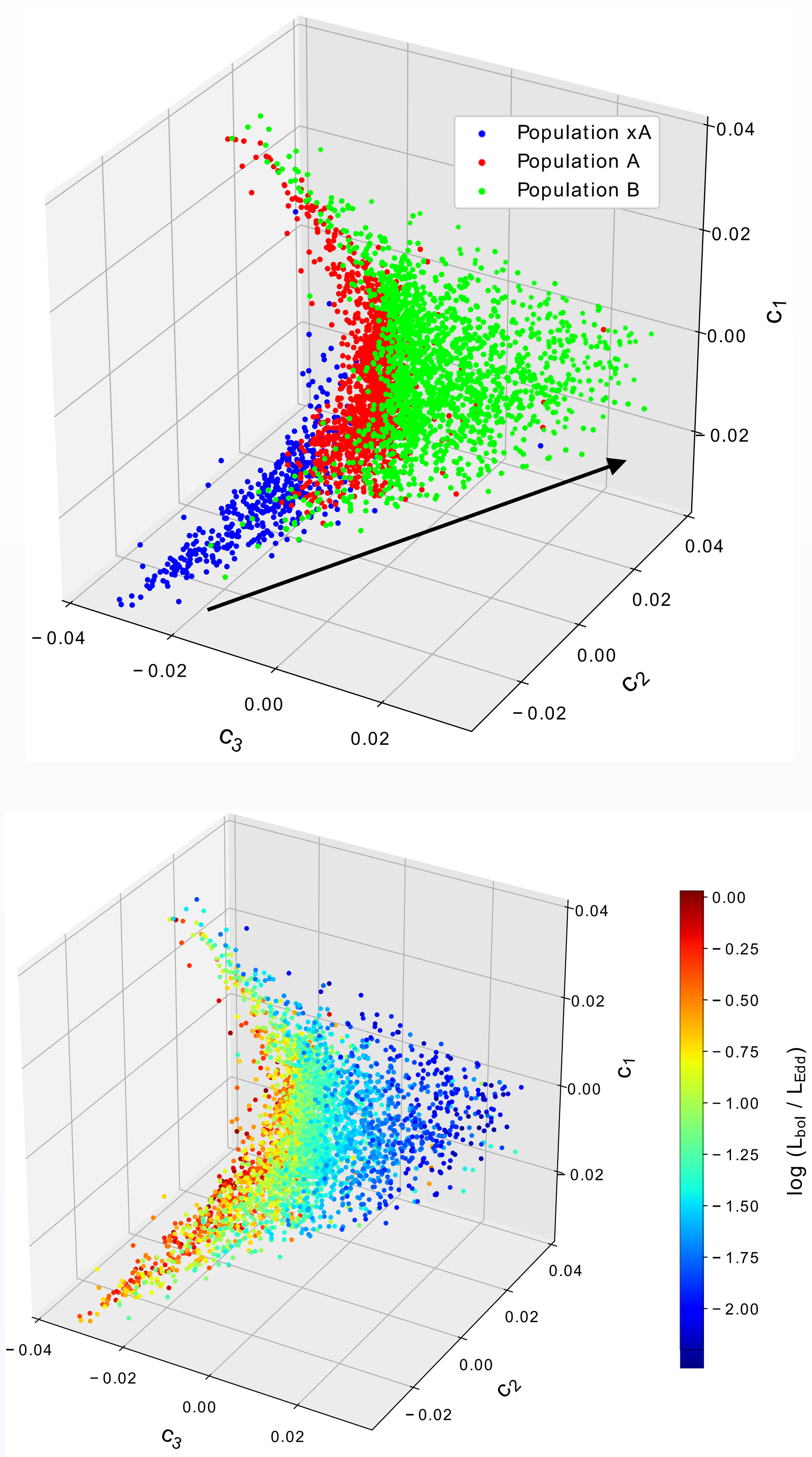

Figure 2. 3D projection of the original manifold embedded in ten-dimensional space. Axes are in arbitrary units and correspond to three components of LLE decomposition. Top – populations xA, A and B are marked with blue, red and green, respectively. The arrow points in the direction of the MS. Bottom - gradient of the Eddington ratio in the resulting projection.

Table 1. Input parameters for the LLE algorithm.

Input parameters
Broad Hα (FWHM, EW, L)
Broad Hβ (FWHM, EW, L)
[O III] λ5007 (EW, L)
log L_{5100}
$R_{FeII} = \frac{\text{EW FeII}}{\text{EW H}\beta}$

EW – equivalent width;
L – luminosity.

After applying the algorithm, we have obtained a 3D space where different quasar populations occupy distinct regions in accordance with their spectral differences (top panel of Fig. 2.). Please note that LLE provides us with a 3D projection of original ten-dimensional space where the relationships between the points are close to their relationships from the original space, which is not necessarily the case with linear dimensionality reduction methods (e.g. PCA). In this way, we tend to retain as much information as possible (including information about non-linear relationships) while lowering the number of dimensions. In addition, we can use these projections to follow variations of different physical parameters and identify which ones are correlated with the MS (bottom panel of Fig. 2).

Our preliminary findings are outlined below:

- LLE may be used as a tool in data exploration and identification of objects with distinct spectral properties, potentially aiding future type 1 quasar classification tasks in large spectroscopic surveys.
- It is possible to find three eigenvectors giving projection in 3D that contains information potentially lost in 2D projection.
- Presence of quasar MS driven by Eddington ratio was confirmed in a space with maximum preservation of the original manifold geometry.

REFERENCES

Boroson, T. A., Green, R. F. : 1992, *The Astrophysical Journal Supplement Series*, **80**, 109

Marziani, P. et al. : 2001, *The Astrophysical Journal*, **558**, 553-560

Roweis, S. T., Saul, L. K. : 2000, *Science*, **290**, 2323-2326

Shen, Y., Ho, L. C. : 2014, *Nature*, **513**, 210-213

Shen, Y. et al. : 2011, *The Astrophysical Journal Supplement Series*, **194**, 45

Smilde, A. K. et al. : 2008, *Bioinformatics*, **25**, 401-405

Silverman, B. W. : 1986, Density estimation for statistics and data analysis, Chapman and Hall, London

Sulentic, J. W. et al. : 2000, *Annual Review of Astronomy and Astrophysics*, **38**, 521-571

CONTACT

Isidora Jankov: isidora_jankov@matf.bg.ac.rs

Dr. Dragana Ilić: dilic@matf.bg.ac.rs

Dr. Anđelka Kovačević: andjelka@matf.bg.ac.rs