

# Algorithms for Stellar Populations in 3D Spectroscopy

Luis Alvarez-Ochoa<sup>1,2</sup> and Leticia Flores-Pulido<sup>3</sup>, Roberto Terlevich<sup>2</sup>, Oleg Starostenko<sup>3</sup>

<sup>1</sup> Universidad Politécnica de Tlaxcala , Tlaxcala, C.P. 90180, México

<sup>2</sup> Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, C.P. 72840, México

<sup>3</sup> Universidad de las Américas Puebla, Puebla, C.P. 72820, México  
lochoa@inaoep.mx, leticia.florespo@udlap.mx, rjt@inaoep.mx,  
oleg.starostenko@udlap.mx

**Abstract.** The storage capacity of digital databases overcomes our ability to analyze data, so the need to narrow this gap as much as possible is imperative. In astronomy, telescopes can perform automatic surveys of the sky gathering a huge amount of data that are waiting to be analyzed in more detail. The deeper is the analysis the more elaborated and slower are the procedures. Computer science is facing the challenge of creating efficient algorithms . Machine learning (ML), a branch of computer science, is an option for many problems where there is no a simple definition of the relation input-output. ML techniques have been successively used in several astronomical and storage knowledge problems. We focus the research on various branches of machine learning by solving a practical problem: estimation of stellar population parameters (SPP) and kinematics of galactic spectra. The SPP have been tackled in multiple ways, but as the complexity and number of spectra grows, the techniques of ML gain importance. We define the stellar populations and we survey some of the most common methods for their estimating. We concentrate on a special type of spectra, called 3D spectra.

**Keywords:** Algorithms, Galactic spectra, Stellar population, Simulated annealing, Evolution strategies.

## 1 Introduction

Nowadays, technologies of storage and instrumentation are producing huge databases in several sciences, astronomy is a good example of this situation. Modern telescopes and instruments daily record spectra and images, these data have a lot of valuable implicit information which needs to be analyzed by efficient algorithms. Sophisticated computational algorithms have already been used for classifying astronomical objects from images and for estimating physical parameters from spectra. The complexity of the techniques changes according to the

size and intrinsic difficulty of the data. However, the astronomical observational techniques evolve and now they are producing more complex data each time. A relatively new technique for obtaining spectra is the Integral Field Spectroscopy (IFS) that can produce thousands of spectra of only one object. The collected data by this kind of instruments require efficient methods of analysis. Examples of the type of information that can be retrieved from an IFS spectrum of one galaxy are the distribution of ages of the component stars in each region of the galaxy, the kinematics of stars and gases. In principle, with these data, one can form the history of the galaxy, a very important problem in Astrophysics. Surely the study of the problems set by IFS spectroscopy will require to develop new algorithms that will contribute to computer science and to other disciplines. The main methods for estimating stellar populations are introduced section 2, two new strategies for improving efficiency are presented in section 3, the experimental results are discussed in section 4.

## 2 Stellar Populations

The stellar populations in galaxies are characterized for an age distribution, the chemical composition (metallicity) and the amount of gas that obscures star light (extinction). These properties can be directly inferred from the characterization of the absorption and emission lines of their spectra [1], this is an appropriate technique in case that just a few spectra and restricted properties have to be analyzed. Other common and more practical approach for estimating physical parameters is to synthesize a galactic spectrum from a spectral base set of simple stellar populations with known physical features, in such a way that this synthetic spectrum resembles the observed spectrum [2]. Basically the method try to find a synthetic spectrum that best fits with the observed one, see figure 3.

3D spectra of galaxies allow to know the stellar populations in greater spatial detail. 3D spectra are formed by a grid of spaxles, where each spaxel contains a spectrum that samples a region of the galaxy providing the necessary spatial resolution to synthesize stellar populations kinematics maps [3],[4],[5].

## 3 Methods

We believe that a machine learning approach will be successful in addressing our problem, this claim is based on both the state of the art review and in our own experience. The problem could be posed as an optimization problem, therefore we must review the available optimization algorithms, mainly those in the machine learning area. We also explore the preprocessing of data as a medium to improve the system performance. The research will lead us to design algorithms that hopefully would be of wide application.

Our problem of stellar population parameters is non-linear with constraints, the different objectives functions that model galactic spectra contain several independent variables, so we have a high dimensional problem, for instance,

the objective function 2 based on the model 1. On the other hand, the data have certain degree of noise. Therefore, we have selected stochastic and heuristic methods to start our experiments, in particular Simulated Annealing (SA) and Evolution Strategies (ES).

$$g(\lambda) = \sum_{i=1}^3 c_i s(a_i, \lambda) (1 - e^{r_i \lambda}) \quad (1)$$

$$f(P) = \sum_{\lambda} |o - g|^2 \quad (2)$$

In any machine learning system the preprocessing stage has particular relevance on the final performance, even more in cases where the data are of high dimension [6], [7]. Each spectrum of interest in this proposal has several thousands of variables, thus some research on this theme is mandatory before we focus on the optimization algorithms.

One way of accelerating the convergence of an optimization algorithm is start with an initial guess solution that near us to an acceptable solution. The guess can be produced by another algorithm. We test this strategy by designing two architectures. The first architecture sort the 3D spectra by means of a similarity metric. This ordering will allow us to estimate the parameters of a given spectrum and then use them as an initial solution for a similar spectra, then the optimization algorithm will refine this initial solution, see Figure 1. The second one, a neural network whose training set are the ordered pairs (spectral features, stellar population parameters) gives the initial guess, see Figure 2.

We start our work by taking both simulated annealing and evolution strategies [8], [9] (a type of EA) as an initial basis, these were chosen by their ability to optimize complex functions. We will use active learning, feature selection and prior knowledge to address our problem.

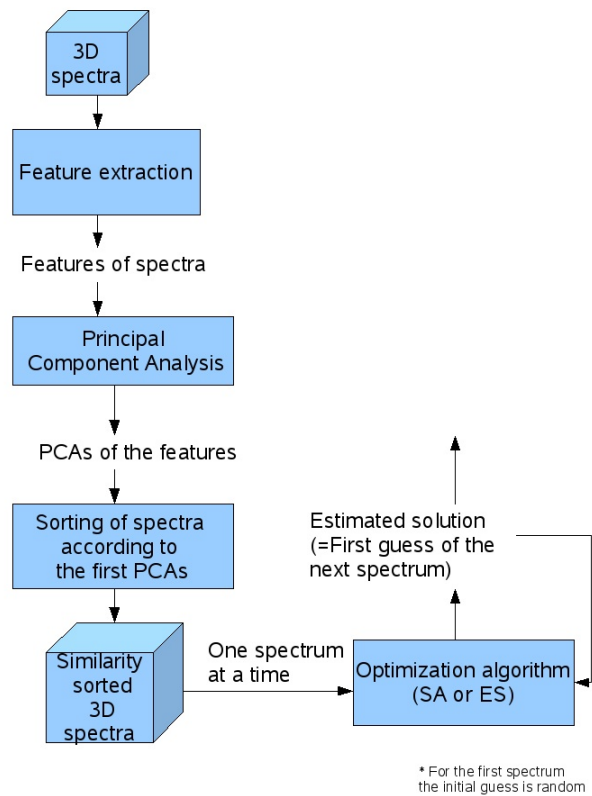
## 4 Preliminary Results

The first experiments are focused on stellar population, the issue of kinematics will be treated later.

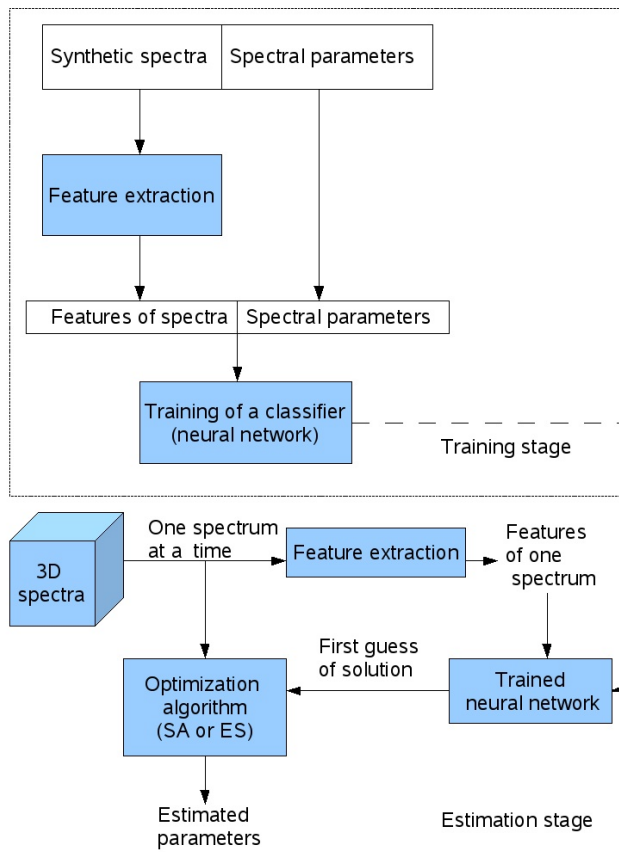
### 4.1 The Graphical User Interface

We have designed a simple graphical user interface (GUI), because given the amount of data graphical results can help us in debugging the approach, besides of integrating the system parts in an ordered way. Figure 3 shows the GUI, it has a menu bar, a central window that shows a test spectrum and its corresponding estimation, the upper right group *galaxy components* displays the selected options for data set and model, finally, the lower group *model parameters* shows the parameter values for both test and estimated spectrum, the number and type of parameters will depend on the chosen model.

The GUI contains already menus for the following tasks:

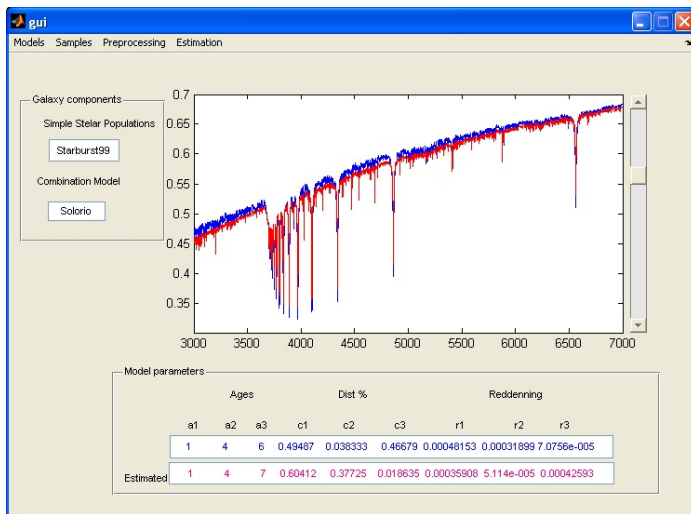


**Fig. 1.** Estimation architecture based on an ordering of spectra according to a similarity metric.



**Fig. 2.** Estimation architecture based on an classifier.

- Select data and models: menu **Model**. The data are simple stellar populations spectra and the model specifies how to combine these data in order to form a synthetic galaxy.
- Create training and test sets: menu **Samples**. The training set is used to train a neural network that can be used to make a first estimation of parameters.
- Calculate features and train a neural network: menu **Preprocessing**. The multiple metric option extracts from the training and test sets features as inputs of a backpropagation neural network.
- Select the optimization algorithm: menu **Estimation**. Here, we can choose between Simulated Annealing and Evolution Strategies as estimation algorithm.



**Fig. 3.** The GUI

## 4.2 The model

We can choose from several models of synthetic galaxy spectra, they basically make linear combinations of spectra of different ages, and then they include a non-linear factor for taking into account the extinction. We test the model reported in [10], [11], see equation 3.

$$g(\lambda) = \sum_{i=1}^3 c_i s(a_i, \lambda) (1 - e^{-r_i \lambda}) \quad (3)$$

where the value of the synthetic galaxy spectrum at wavelength  $\lambda$  is  $g(\lambda)$ ,  $a_1, a_2$  and  $a_3$  are the ages of the component spectra,  $c_1, c_2$  and  $c_3$  are the relative contributions to the galaxy and  $r_1, r_2$  and  $r_3$  are the extinction values.  $s(a_i, \lambda)$  refers to the value that a spectrum of age  $a_i$  at wavelength  $\lambda$  has.

A basis of seven high resolution spectra were calculated with a suitable small suitable basis for initial experiments. The ages of spectra range from 1 million to 1,000 million years. Each spectrum has 13,323 pixels that covers a spectral interval from 3,000 to 7,000 Å.

As a primer set of features we chose: euclidian distance, sum of squared differences, difference in total luminosity, maximal difference and difference between means in ten equal intervals through the spectra, all of these are defined below. The fourteen features are mapped to the nine parameters using a neural network, this is the first estimation that is used later by the optimization algorithm.

Let  $N$  be the number of spectra,  $f_\lambda^k$  the energy flux of the spectrum  $k$  at the point  $\lambda$ , where  $k = 1 \dots N$  and  $\lambda = 1 \dots m$ . The definition of features is the following:

We took an arbitrary spectrum and we named it  $f^0$ , all the features are metrics of similarity with respect to this reference spectrum.

$$f_{Tot}^k = \sum_{i=1}^m f_i^k \quad (4)$$

$$d_{lum}(f^0, f^k) = |f_{Tot}^0 - f_{Tot}^k| \quad (5)$$

$$d_{max}(f^0, f^k) = \max(\{|f_\lambda^0 - f_\lambda^k|\}) \quad (6)$$

$$d_{euc}(f^0, f^k) = \sqrt{\sum_{\lambda=1}^m (f_\lambda^0 - f_\lambda^k)^2} \quad (7)$$

$$d_{sqreuc} = \sum_{\lambda=1}^m (f_\lambda^0 - f_\lambda^k)^2 \quad (8)$$

We divide the spectrum into  $j$  intervals of the same length and we measure their means, the intervals are designated by  $[j_{low}, j_{up}]$

$$mean_j^k = \frac{1}{(j_{up} - j_{low} + 1)} \sum_{\lambda=j_{low}}^{j_{up}} f_\lambda^k \quad j = 1 \dots 10 \quad (9)$$

$$d_{mean_j}(f^0, f^k) = |mean_j^0 - mean_j^k| \quad (10)$$

### 4.3 Experimental Results

In order to assess the relevance of the selected features in the estimation process, we prepared a training set of 100 spectra, this set is the input to the neural network that will predict the population parameters. The prediction done by the network is the starting point of the optimization algorithms SA-NN and ES-NN.

Another way of generating a first guess is to order the spectra by a similarity metric, so in the estimation of a given spectrum we can use the solution of a neighboring spectrum as the starting point of the optimization algorithms SA-PCA and ES-PCA. The proximity in the feature space should reflect the proximity in the parameter space. We project the features on the principal components instead of calculating the distances over all the pairs of spectra, and we use first component as criterion of nearness.

The better results were obtained using a neural network (see figure 4). The PCAs did not work better than neural networks, a explanation of this result could be that the test set is randomly sparse and probably the mean distance between set of parameters is not suitable for this technique. In addition, in SA and ES algorithms, the rate of change of the actual solution is high at the beginning, so we must try to choose appropriately the parameters of the algorithms to control the rate when a starting guess is used.

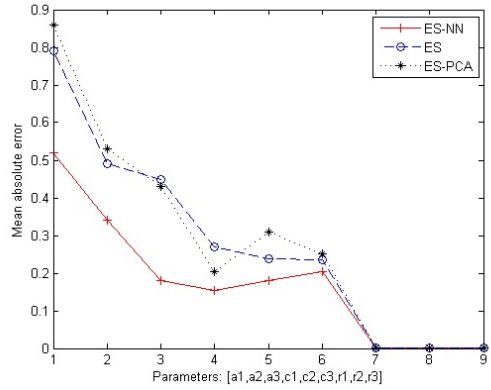
The real spectra of galaxies cover only certain regions of the parameter space, so we do not discard the use of PCA until dealing with real 3D spectra. The figure 5 shows that a set of parameters over a given direction, could be ordered by the PCA of the features. The distribution of the parameters over a direction was formed by sampling uniformly from minor to major each parameter range (coordinates). The 100 sets of parameters were integrated by combining the sampled values in this form: the first point is formed by the first elements of each coordinate, the second point by the second elements of each coordinate, and so on.

The stop criterion was the number of evaluations of the objective function: 2000. The training of the backpropagation network lasted three minutes with a training set of 100 spectra and the stop criterion was the prediction accuracy: sum of squared differences  $\leq$  to 0.01.

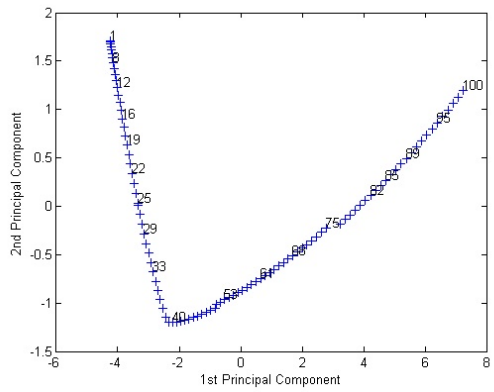
## 5 Conclusions

From these preliminary results, we conclude that the selected features used as inputs to a neural network can increase the accuracy of the optimization algorithms, at the same time the features can be useful in the definition of a measure of the nearness in the parameter space that can be exploited in optimization of multiple instances, under certain restrictions, i.e. the data should be clustered and no distributed over all parameter space. Furthermore, the parameters of the optimization algorithms must be initialized in a non standard way for taking advantage of the first guess.





**Fig. 4.** Mean absolute errors of 100 spectra using Evolution Strategies



**Fig. 5.** The first and second principal components of the features corresponding to spectra generated by 100 sets of parameters distributed along a hyper-line in the parameter space. The first component preserves the same order of the parameter space, this is indicated by the numbers.

## References

1. Terlevich, E. and Terlevich, R. and Torres-Papaqui, J.P. and Fernandes, Roberto Cid and Melnick, J. and Kunth, D. and Gu, Q. and Bressan, A., Characterizing the stellar population of the nuclei of active galaxies, Resolved Stellar Populations, D. Valls-Gabaud M. Chavez, TBA, ASP Conference Series, (2005).
2. Fernandes, Roberto Cid and Mateus, Abílio and Sodré, Laerte Jr. and Stasińska, Grazyna and Gomes, Jean M., Semi-empirical analysis of Sloan Digital Sky Survey galaxies- I. Spectral synthesis method, Monthly Notices of the Royal Astronomical Society, vol. 358, 363-368, month 04, (2005).
3. Rosado, Margarita; Some Astronomical Niches With 3D Spectroscopy; Revista Mexicana de Astronomia; Vol. 24, 92-101, (2005).
4. Kuntschner, Harald and Emsellem, Eric and Bacon, Roland and Bureau, Martin and Cappellari, Michele and Davies, Roger L. and de Zeeuw, Tim and Falcón-Barroso, Jesús and Krajnovic, Davor and McDermid, Richard M. and Peletier, Reynier F. and Sarzi, Marc, The stellar populations of E and S0 galaxies as seen with SAURON, arXiv, January, (2006).
5. Roth, Martin M. and Kelz, Andreas and Fechner, Thomas and Hahn, Thomas and Bauer, Svend-Marian and Becker, Thomas and Bohm, Petra and Christensen, Lise and Dionies, Frank and Paschke, Jens and Popow, Emil and Wolter, Dieter, PMAS: The Potsdam Multi-Aperture Spectrophotometer. I. Design, Manufacture and Performance, Publications of the Astronomical Society of the Pacific, no. 117, 620-642, June, (2005).
6. Bishop, Cristopher M., Neural Networks for Pattern Recognition, chapter 8, Oxford University Press, 2000.
7. Guyon, Isabelle and Elisseeff, André, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, 1157-1182, March (2003).
8. Back, Thomas, Evolutionary Algorithms in Theory and Practice, Oxford University Press, (1996).
9. Michalewicz, Zbigniew and Fogel, David B., How to Solve It: Modern Heuristics, Springer, (2002).
10. Solorio, Tamar and Fuentes, Olac and Terlevich, Roberto and Terlevich, Elena, An Active Instance-based Machine Learning method for Stellar Population Studies, Monthly Notices of the Royal Astronomical Society, (2005).
11. Alvarez, Luis and Fuentes, Olac and Terlevich, Roberto, Extracting stellar populations parameters of galaxies from photometric data using evolution strategies and locally weighted regression, Lecture Notes in Computer Science, 395, volume 3215 month 9, KES2004, Springer Verlag, (2004).