Extracting Stellar Population Parameters of Galaxies from Photometric Data Using Evolution Strategies and Locally Weighted Linear Regression *

Luis Alvarez[1], Olac Fuentes[1], and Roberto Terlevich[12]

[1] Instituto Nacional de Astrofísica Óptica y Electrónica,
Luis Enrique Erro # 1 Santa María Tonantzintla, Puebla, 72840, México
lochoa@ccc.inaoep.mx, fuentes@inaoep.mx, rjt@inaoep.mx
[2] Institute of Astronomy, University of Cambridge
Madingley Road, Cambridge CB3 0HA, UK

Abstract. There is now a huge amount of high quality photometric data available in the literature whose analysis is bound to play a fundamental role in studies of the formation and evolution of structure in the Universe. One important problem that this large amount of data generates is the definition of the best procedure or strategy to achieve the best result with the minimum of computational time.

Here we focus on the optimization of methods to obtain stellar population parameters (ages, proportions, redshift and reddening) from photometric data using evolutionary synthesis models. We pose the problem as an optimization problem and we solve it with Evolution Strategies (ES). We also test a hybrid algorithm combining Evolution Strategies and Locally Weighted Linear Regression (LWLR). The experiments show that the hybrid algorithm achieves greater accuracy, and faster convergence than evolution strategies. On the other hand the performance of ES and ES-LWLR is similar when noise is added to the input data.

1 Introduction

The main aim of this work is to explore automatic techniques for obtaining stellar population parameters (***spp***) from photometric data of galaxies (***pd***). Given the huge amount of information available in the form of photometric data, for example Sloan Digital Sky Survey [1], it is necessary to test faster data analysis methods.

Using evolutionary algorithms in different problems of astronomy was suggested in [5]. In [10] genetic algorithms were used to predict parameters of interacting galaxies. The analysis of stellar spectra with evolution strategies is

* This work was partially supported by CONACyT (the Mexican Research Council) under grants 177932 and J31877A.

presented in [8]. The extraction of ***spp*** has had several approaches. Some methods for determining age and reddening of stellar clusters are reviewed in [9]. Neural networks have also been used for calculating photometric redshifts [6].

Here we consider the problem as an optimization one, where finding a solution is difficult due to the number of variables involved and the level of noise in the data. Evolution Strategies (ES) seem to perform well in these particular conditions but, for problems with many dimensions ES can be slow. To speed up the convergence we have combined ES with Locally Weighted Linear Regression (LWLR). LWLR creates local linear models of a function around a query point. We used the candidate solutions found in each iteration of the ES for predicting another solution, possibly better, that will be included in the solution set.

This paper is structured as follows: Section 2 describes the procedure for creating data. Section 3 briefly summarizes the methods. Section 4 presents the results of experiments and Section 5 shows the conclusions.

2 Data

We have utilized a set of theoretical synthetic spectra, **F** of simple stellar populations along isochrones of given age and metallicity. They were computed for solar metallicity $Z = 0.02$ with logarithmic ages of: 6, 8, 8.3, 8.6, 9, 9.6, 9.78, 10, 10.2 yr. The resolution is 20Å and all were computed for a Salpeter Initial Mass Function.

Using this set we then form a set of **synthetic galactic** spectra [4]. First we separated the nine spectra into three age groups: the first contains one spectrum of a young stellar population, the second contains four intermediate-age spectra and the third has four old spectra. In the first step of the procedure we only have to specify F_2 and F_3. Since the proportions must sum 1, only p_1 and p_2 need to be specified.

We normalized the spectra, each component of a normalized spectrum F_i^{norm} being given by $F_{i,\lambda}^{norm} = F_{i,\lambda} / \sum_\lambda F_{i,\lambda}$, where $\lambda \in [890 - 23010\text{Å}]$. For clarity we now change F_i^{norm} to F_i.

The procedure for forming the galactic spectra and the photometric data is:

1. Randomly select populations F_i and proportions p_i.
2. Combine three spectra, F_i, of different ages at given proportions p_i.

$$F_{\mathrm{combined}} = F_1 p_1 + F_2 p_2 + F_3 p_3 \ . \tag{1}$$

3. Apply a simplified model of reddening, R, to F_{combined}

$$F_{\lambda,\mathrm{reddened}} = F_{\lambda,\mathrm{combined}} - F_{\lambda,\mathrm{combined}} \left(\frac{kR}{\lambda} \right) \ . \tag{2}$$

4. Apply redshift, Z, to F_{reddened} according to the formula

$$\lambda = \lambda_0 (Z + 1) \ . \tag{3}$$

We obtain a new spectrum, $F_{\mathrm{redshifted}}$.

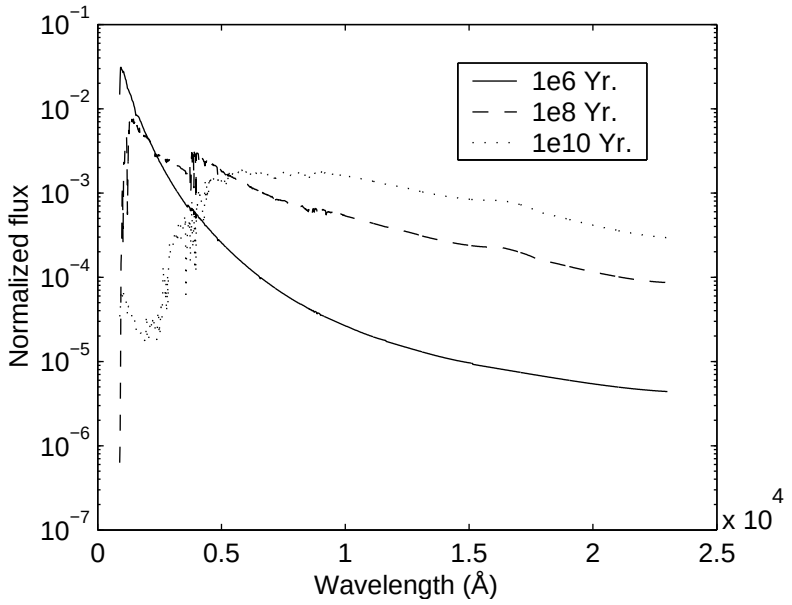

Fig. 1. Three spectra from which we construct a galactic spectrum. These are chosen from a grid of nine spectra

5. Finally we divide, $F_{\text{redshifted}}$, into fifteen intervals of the same width and we average their fluxes in each interval, simulating wide-band filters.

To summarize, the six ***spp*** we wish to extract are $[F_2, F_3, p_1, p_2, R, Z]$. Figure 1 shows three selected spectra, and Figure 2 depicts their combination following the procedure just described.

3 Method

In this section we describe ES, LWLR and the ES-LWLR hybrid in the context of the ***spp*** problem. Our main aim is to test the combination of ES with LWLR for speeding up the convergence into a suitable solution. This hybrid approach is based on the idea that it is possible to use the solutions generated in each iteration of the ES to form a local linear model of the objective function. For this purpose, we chose LWLR [2]. We then use the current model to predict another, hopefully better, solution. The predicted solution is added to the current population. A similar technique, but in conjunction with Active Learning is reported in [7].

3.1 Evolution Strategies

Among the several varieties of ES [3] we selected the $(\mu + \lambda)$-ES because it adapts better to the requirements of the hybrid algorithm. In this variant of

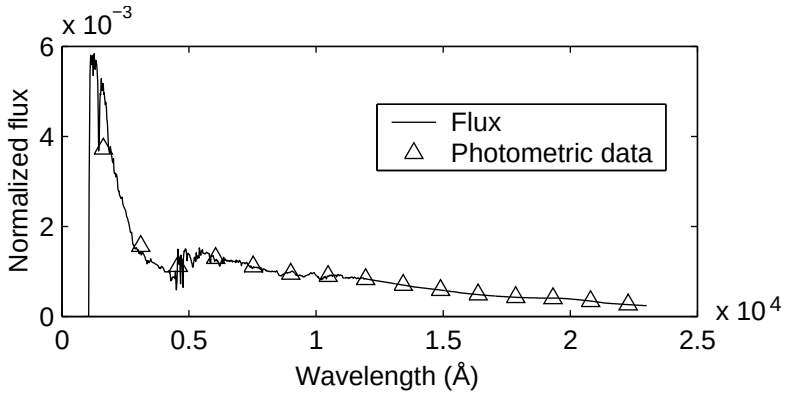

Fig. 2. A spectrum formed from the spectra of Figure 1, according to the procedure described in Section 2. The parameters used in this case are: $F_2 = 1\,(10^8 yr)$, $F_3 = 3\,(10^{10} yr)$, $p_1 = 0.23$, $p_2 = 0.37$, $R = 0.3$ and $Z = 0.2$. The photometric data indicated by triangles result from dividing the spectrum into fifteen intervals of equal width and averaging their corresponding fluxes

ES, the candidate solutions are chosen from a population formed by μ parents and λ offspring. These results in the population concentrating around one solution allowing the fit of finer models using LWLR. Other variants like (μ, λ)-ES that only choose their candidate solutions from λ offspring, are slower and their solutions are more sparse, although they are generally more noise tolerant.

How ES implements the principles of evolution in optimization is explained next: Given the problem of finding the $\mathbf{x}$ that minimizes the function $f(\mathbf{x})$ in a domain, M, where $\mathbf{x}$ satisfies the constraints $g_i(\mathbf{x}) > 0$, $i \in \{1, \ldots, k\}$, ES perform the following process in order to find a solution.

1. Generate a set of μ random solutions called initial population, $\mathbf{P}^0$.
2. Recombine the population, applying operator r, for creating λ offspring.
3. Mutate the offspring, applying operator m.
4. Select from the total population of parents and offspring the best μ solutions (the new population), and eliminate the rest.
5. Add a new individual applying the LWLR algorithm to the current population (hybrid part).
6. Go back to the second step if the termination criterion, t, is not satisfied.

The hybrid step 5 will be explained in subsection 3.2. All components of the ES are summarized in the tuple $(\mathbf{P}^0, \mu, \lambda; r, m, s;\ \Delta\sigma;\ f, g, t)$.

The recombination and mutation operators (also called genetic operators) work on extended vectors, $\boldsymbol{a} = (\boldsymbol{x}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ has the same length as $\boldsymbol{x}$, and contains the standard deviations for carrying out the mutation.

The recombination operator, r, produces one individual, $\boldsymbol{a}^{'}$ by selecting randomly with uniform probability two elements, $\boldsymbol{a}_a$ and $\boldsymbol{a}_b$, from the current population $\mathbf{P}$, and mixing them. We use discrete recombination in the $\boldsymbol{x}$'s, which consists of randomly selecting with uniform probability the components of the offspring from the two parents, (see Equation (5)). For the $\boldsymbol{\sigma}$'s, average recombination is used, (see Equation (6)). The recombination operator resembles the sexual reproduction in biological organisms.

$$r(\mathbf{P}^t) = \boldsymbol{a}^{'} = (\boldsymbol{x}^{'}, \boldsymbol{\sigma}^{'}) \ . \tag{4}$$

$$x_i^{'} = x_{i,b} \ \ or \ \ x_{i,a} \ . \tag{5}$$

$$\sigma_i^{'} = \frac{1}{2}(\sigma_{i,a} + \sigma_{i,b}) \ . \tag{6}$$

The mutation operator, m, randomly changes the vector $\boldsymbol{a}^{'}$, as occurs in nature. This brings improvements in the population, i.e., better adapted individuals (solutions), those that produce smaller objective function values than past generations did.

$$m(\boldsymbol{a}') = \boldsymbol{a}^{''} = (\boldsymbol{x}^{''}, \boldsymbol{\sigma}^{''}) \ . \tag{7}$$

$$\sigma_i^{''} = \sigma_i^{'} \exp N_0(\Delta\sigma) \ . \tag{8}$$

$$x_i^{''} = x_i^{'} + N_0(\sigma_i^{''}) \ . \tag{9}$$

$N_0(\sigma)$ is a random number generator from a normal distribution with mean 0 and standard deviation σ. $\Delta\sigma$ is a meta-parameter that controls the rate of change of the $\sigma' s$.

The selection operator s evaluates the objective function, f, for the total population and chooses the best μ (those that evaluate f to the smallest absolute values),that will in turn form the new generation of solutions. In our problem the $\mathbf{x}$ part of the extended $\mathbf{a}$ vector corresponds to the ***spp***. The objective function, f, is the sum of the quadratic differences among the photometric data given as query, and the photometric data produced by one solution of the population. The stopping criterion, t, finalizes the algorithm at the $50th$ iteration (generation), although it could be also a minimum error or an elapsed time.

3.2 Locally Weighted Linear Regression

As mentioned at the beginning of Section 3, the solutions generated in each generation of the ES could be used to form a linear model for predicting another solution. The information available for creating the model are the ordered pairs $\langle \boldsymbol{spp}, \boldsymbol{pd} \rangle$. The ***spp*** are generated randomly by the ES and the photometric data ***pd*** are calculated by means of the procedure outlined in Section 2. In this particular case, we need to predict the ***spp***, so we reverse the pairs to obtain

$\langle \boldsymbol{pd}, \boldsymbol{spp} \rangle$. The linear model is represented by Equation (10), with unknown coefficients, $\boldsymbol{\beta}$.

$$\mathbf{PD}\,\boldsymbol{\beta} = \mathbf{SPP} \ . \tag{10}$$

$\mathbf{PD}$ and $\mathbf{SPP}$ are matrices that contain all the pairs of a generation. The unknown vector, $\boldsymbol{\beta}$, is found by minimizing the least squares criterion

$$C = \sum_{for\,each\,row\,i} (\mathbf{PD}_i^T \boldsymbol{\beta} - \mathbf{SPP}_i)^2 \ . \tag{11}$$

The solution is also called normal equations

$$\boldsymbol{\beta} = (\mathbf{PD}^T\mathbf{PD})^{-1}\mathbf{PD}^T\mathbf{SPP} \ . \tag{12}$$

Vector $\boldsymbol{\beta}$ is used to predict any $\boldsymbol{spp}$ given a new set of $\boldsymbol{pd}$ (here named $\boldsymbol{pd}_{\text{query}}$), at the accuracy allowed by a global linear model. The accuracy is improved if we make local models, around the $\boldsymbol{pd}_{\text{query}}$, instead of global models. We have achieved this using proportionally weighting the data $\langle \mathbf{PD}, \mathbf{SPP} \rangle$ according to the Euclidean distance between each row $\mathbf{PD}_i$ and the $\boldsymbol{pd}_{\text{query}}$. In this way, the $\mathbf{PD}_i$ near $\boldsymbol{pd}_{\text{query}}$, will have greater influence in the model than farther $\mathbf{PD}_i$.

$$\mathbf{Z} = \mathbf{W} \cdot \mathbf{PD} \ . \tag{13}$$

$$\mathbf{V} = \mathbf{W} \cdot \mathbf{SPP} \ . \tag{14}$$

Equations (13) and (14) weight the data by multiplying them by the diagonal matrix $\mathbf{W}$, containing the inverse Euclidean distances of the $\boldsymbol{pd}_{\text{query}}$ to each row vector in $\mathbf{PD}$. Replacing these weighted variables in the normal equations, $\boldsymbol{\beta}$ can be determined. An $\boldsymbol{spp}$ is estimated given a $\boldsymbol{pd}_{\text{query}}$

$$\boldsymbol{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{V} \ . \tag{15}$$

$$\hat{\boldsymbol{spp}}(\boldsymbol{pd}_{\text{query}}) = \boldsymbol{pd}_{\text{query}}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{V} \ . \tag{16}$$

Equation (16) is included in step 5 of the ES algorithm, in this way we construct the hybrid algorithm ES-LWLR.

4 Results

We have run a series of tests in order to critically assess the performance of ES and ES-LWLR. We adopted as the measure of quality the mean absolute errors (MAE) of the $\boldsymbol{spp}$ and the execution time of the algorithms. The parameters for both algorithms are the same: population $\mu = 50$, number of offspring $\lambda = 100$, $\Delta\sigma = 0.25$, stopping criterion $t = 50$ generations for each $\boldsymbol{spp}$.

The different tests arise from variations of the procedure described in Section 2. The first test set consisted of 100 random sets of $\boldsymbol{spp}$ generated according to the mentioned procedure (see Section 2). We then reduced the resolution to

just 100 pixels by uniformly sampling the original 1107 pixels spectra. Two additional sets come from the addition of Gaussian noise to both low and high resolution spectra after we have applied the redshift to them. The aggregated noise has mean zero and standard deviation equal to 0.1 times the maximum flux of the spectrum. Thus we have a total of 400 ***spp*** for testing. We test both algorithms over these 400 ***spp***, giving a total of eight test.

As can be seen in Table 1 the hybrid has better performance. Figure 3 shows the differences between some given spectra and their predicted spectra. On low resolution spectra, again, the hybrid achieves better accuracy (see Table 1). After adding noise the accuracy reached by both algorithms is almost the same (see Table 1).

Because the domains of each stellar population parameter are different: F_2 and $F_3 \in [0, 4]$, p_1, p_2 and $R \in [0, 1]$ and $z \in [0, 2]$. Then the meaning of the MAE's of each parameter (see Table 1) depends on their respective domain.

In order to evaluate the efficiency of the proposed method we compare the rate of convergence of the objective function, f (see Section 3.1). This function is implemented in the algorithms. The hybrid algorithm converge in less generations than the ES (see Figure 4). Table 2 shows how the convergence time is reduced for the hybrid when the stopping criterion is a given error.

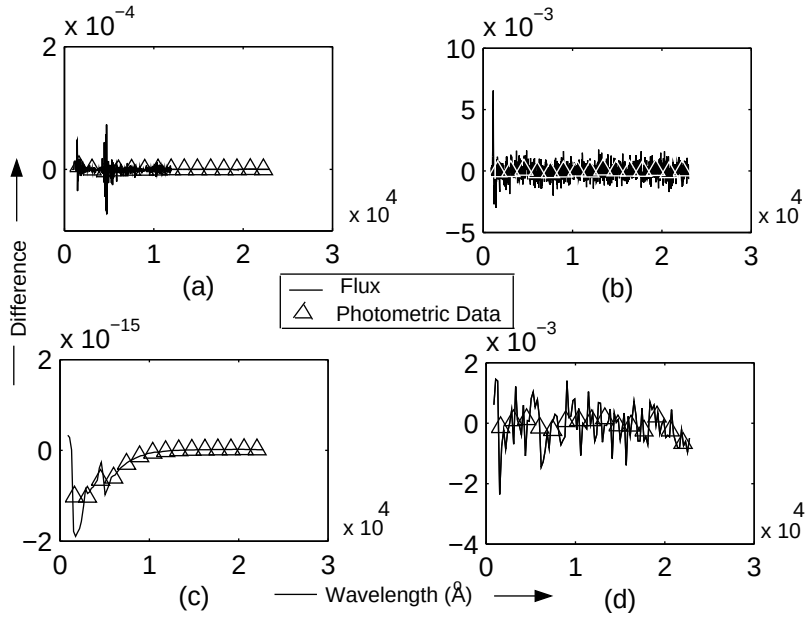

Fig. 3. Differences between spectra, formed from the ***spp*** of the Figure 2, and their predicted spectra. (a) and (b) correspond to high resolution spectra with and without noise respectively. (b) and (c) belong to low resolution spectra with and without noise respectively

Table 1. Mean absolute error of ***spp*** over 100 spectra (h=high, l=low, n=no, y=yes)

Algorithm	Resolution/Noise	F_2	F_3	p_1	p_2	p_3	R	Z
ES	h/n	0.67	1.03	0.0342	0.048	0.0544	0.0959	0.0254
ES-LWLR	h/n	0.12	0.28	0.0004	0.0017	0.0017	0.0006	0.0003
ES	l/n	0.58	1.14	0.039	0.0479	0.0524	0.0758	0.027
ES-LWLR	l/n	0.09	0.24	0.0005	0.0007	0.001	0.0005	0.003
ES	h/y	0.62	1.19	0.0344	0.0529	0.587	0.1001	0.0373
ES-LWLR	h/y	0.89	0.85	0.0411	0.0481	0.0541	0.0847	0.389
ES	l/y	0.94	1.08	0.0696	0.0936	0.0833	0.1641	0.0731
ES-LWLR	l/y	0.8	1.06	0.574	0.0888	0.0949	0.1461	0.0654

Table 2. Mean absolute error of ***spp*** over 100 high resolution spectra. Stopping criterion error=1e-10

Algorithm	F_2	F_3	p_1	p_2	p_3	R	Z	t (seg)
ES	0.87	1.07	0.0366	0.0450	0.0560	0.0922	0.0289	2017
ES-LWLR	0.1	0.45	0.0005	0.003	0.0029	0.0017	0.0006	843

5 Conclusions

We approached the problem of extracting stellar population parameters from photometric data as an optimization problem and we demonstrated that it can be solved with Evolution Strategies. We have also shown that the combination of ES with Locally Weighted Regression speeds up the convergence. This is due to the use of the solutions generated in each iteration of ES to form a linear model; this linear model is then used to predict a better solution. Our experiments show that in noiseless data, the hybrid algorithm considerably reduces the time of computing with respect to ES when the stopping criterion is a given error (see Table 2 and Figure 4). The calculated MAE's of each stellar population parameter show that the hybrid algorithm has better accuracy than the ES, in noiseless spectra, as was expected. On the other hand the performance of the hybrid algorithm after adding 10% Gaussian noise to the spectra is comparable to that of ES alone.

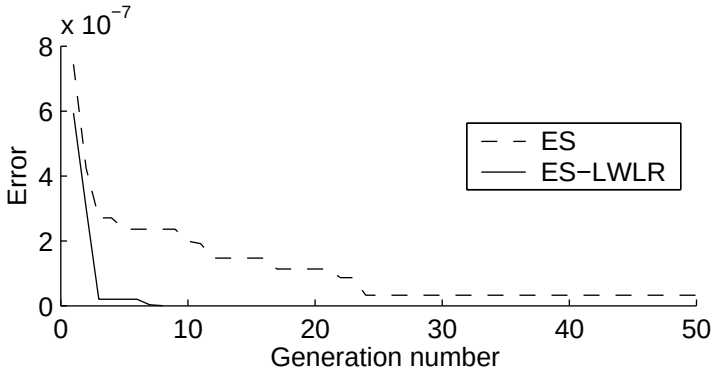

Fig. 4. Error of the objective function

References

1. K. Abazajian, et al.: The first data release of the Sloan Digital Sky Survey. The Astronomical Journal. **126** (October 2003) 2081–2086
2. C.G. Atkenson, A.W. Moore, and S. Schaal.: Locally weighted learning. Artificial Intelligence Review. **11** (1997) 11–73
3. T. Bäck, F. Hoffmeister, H.P. Schwefel.: A survey of evolution strategies. Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA. (1991) 2–9
4. A. Bressan, C. Chiosi, F. Fagotto.: Spectrophotometric evolution of elliptical galaxies. 1: Ultraviolet excess and color-magnitude-redshift relations. The Astrophysical Journal. **94** (1994) 63–115
5. P. Charbonneau. 1995.: Genetic algorithms in astronomy and astrophysics. Astrophysical Journal, Supplement Series. **101** (December 1995) 309–334
6. A.E. Firth, O. Lahav, R.S. Somerville.: Estimating photometric redshifts with artificial neural networks. Monthly Notices of the Royal Astronomical Society. **339** (March 2003) 1195-1202
7. O. Fuentes, T. Solorio.: Interferogram analysis using active instance-based learning. IASTED International Conference Artificial Intelligence and Applications, Benalmádena, Málaga, Spain. (2003) 386–390
8. J.F. Ramírez, O. Fuentes.: Spectral analysis using evolution strategies. IASTED, International Conference on Artificial Intelligence and Soft Computing, Banff, Alberta, Canada. (July 2002) 208–213
9. J.F.C. Santos, E. Bica.: Reddening and age for 11 Galactic open clusters from integrated spectra. Monthly Notices of the Royal Astronomical Society. **206** (1993) 915–924
10. M. Wahde.: Determination of orbital parameters of interacting galaxies using a genetic algorithm. Astronomy and Astrophysics Supplement Series. **132** (November 1998) 417–429