Adjustment of Observational Data to Specific Functional Forms Using Particle Swarm Algorithm and Differential Evolution: Rotational Curves of Spiral Galaxy as Case Study

Miguel Cárdenas-Montes, Mercedes Mollá, Miguel A. Vega-Rodríguez, Juan José Rodríguez-Vázquez and Antonio Gómez-Iglesias

Abstract The fitting of experimental or observational data to specific functional forms requires high computational capacities in order to tackle with the complexity of the calculations. This complexity makes compulsory the usage of efficient search procedures, such as evolutionary algorithms. Evolutionary algorithms have proved their capability to find sub-optimal high-quality solutions to problems with large search space. In this context, Particle Swarm Algorithm and Differential Evolution are used to fit a data set to a serial expansion of Legendre polynomial. Concerning the data set, 56 rotation curves of spiral galaxies are used to build up a serial expansion —physical meaningless— retaining the essential information of the curves. The final goal of this work is two-fold: firstly, to provide a theoretical functional form representing the features of the rotational curves of spiral galaxies in order to be coupled to other computational models; and secondly, to demonstrate the applicability of the evolutionary algorithms to the matching between astronomical data sets and theoretical models.

Miguel Cárdenas-Montes
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, Department of Fundamental Research, Avda. Complutense 22, Madrid, Spain, e-mail: miguel.cardenas@ciemat.es

Mercedes Mollá
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, Department of Fundamental Research, Avda. Complutense 22, Madrid, Spain, e-mail: mercedes.molla@ciemat.es

Miguel A. Vega-Rodríguez
University of Extremadura, ARCO Research Group, Dept. Technologies of Computers and Communications, Cáceres, Spain, e-mail: mavega@unex.es

Juan José Rodríguez-Vázquez
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, Department of Fundamental Research, Avda. Complutense 22, Madrid, Spain, e-mail: jj.rodriguez@ciemat.es

Antonio Gómez-Iglesias
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, National Fusion Laboratory, Avda. Complutense 22, Madrid, Spain, e-mail: antonio.gomez@ciemat.es

1 Introduction

This article focus on the construction of a model for the rotational curves of spiral galaxies. For this, the observational data are normalized and merged, and next, fitted to physical meaningless functional forms. Due to the large search space, Evolutionary Algorithms (EAs) are used to find sub-optimal high-quality solutions.

EAs, like Particle Swarm Algorithm (PSO) and Differential Evolution (DE) are powerful methods for solving many tough optimization problems. In science, the EAs have been profusely used to solve complex problems. In this work, PSO and DE are implemented to adjust a large observational data set —56 rotational curves of spiral galaxies— to functional forms. The huge volume of data under treatment forces the use of this kind of techniques.

PSO and DE are well known EA, widely adopted and suitable for the first approximation to any optimization problem. Regarding the functional form, Legendre polynomial and normal polynomial are considered to reproduce the essential information of the rotational curves.

This paper is organised as follows: Section 2 summarises the Related Work and previous efforts done. In Section 3, the Evolutionary Algorithms used in this article are briefly described. In Section 4, the details of the implementation and the Production Setup are shown. The Results and the Analysis are displayed in Section 5. And finally, the Conclusions and the Future Work are presented in Section 6.

2 Related Work

In the bibliographic search, few related studies have been found. It exists an old work which has inspired partially this survey. In this work, the author used a genetic algorithm to adjust the observational data of the spiral galaxy NGC 6946 [1]. Instead of using a physical meaningless, the author uses an equation with physical meaning describing the four mass contributions to the rotation curve —bulge, disk, interstellar gas and halo— (Eq. 1).

$$v^2(r) \;=\; v_D^2(r) \,+\, v_B^2(r) \,+\, v_H^2(r) \,+\, v_G^2(r) \tag{1}$$

Except for the halo, the other three contributions are merged in a variable, whereas the halo contribution is modelled by the Eq. 2. Therefore, the number of parameters to adjust is only three.

$$v_H^2(r) \;=\; 2\cdot\sigma^2\cdot(1-(\frac{r}{\alpha})\,\cdot\, tang^{-1}(\frac{\alpha}{r})) \tag{2}$$

In spite of the similarities —the application of a Genetic Algorithm to fit a data set—, the target of this work is very different. Whereas in [1] the focus is clearly on the physical behaviour of the rotation; in the present work, our study focuses on the extraction of the essential information of the curves involved in order to produce a universal curve.

3 Evolutionary Algorithms

EAs are stochastic search methods which maintain a population of tentative solutions that are manipulated competitively by applying some variation operators to find satisfactory solutions. The skeleton of a standard EA is as follows: EA proceeds in an iterative manner by generating new populations *P(t)* of individuals from the former population, every individual in the population is the encoded version of a tentative solution, an evaluation function associates a fitness value to every individual indicating its suitability to the problem, the canonical algorithm applies stochastic operators in order to compute a whole generation of new individuals. In a general formulation, variation operators to create a temporary population *P'(t)* are applied. Next, the resulting individuals are evaluated. Finally, a new population *P(t+1)* is obtained by using individuals from *P'(t)* or *P(t)*.

In all the EAs used in this work, the population structure is panmictic. Thus, the intrinsic operations to each EA take place globally over the whole population. Furthermore, in all cases the EAs follow a generational model, in which an entire new population of individuals *P'(t)* replaces the old one *P(t)* [2].

3.1 Particle Swarm Algorithm

In PSO initially, a set of particles are randomly created. During the process of particles movement, each particle keeps track of its coordinates in the problem space that are associated with the best solution it has achieved so far. Not only the best historical position of each particle is kept, also the associated fitness is stored. This value is called *localbest*.

Another "best" value that is tracked and stored by the global version of the particle swarm optimiser is the overall best value, and its location, obtained so far by any particle in the population. This location is called *globalbest*.

The PSO [3], [4], [5] concept consists of, at each time step, changing the velocity (accelerating) of each particle toward its *localbest* and the *globalbest* locations (in the global version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *localbest* and *globalbest* locations.

The process for implementing the global version of PSO is as follows:

1. Creation of a random initial population of particles. Each particle has a position vector and a velocity vector on N dimensions in the problem space.
2. Evaluation of the desired (benchmark function) fitness in N variables for each particle.

3. Comparison of each particle fitness function with its *localbest*. If the current value is better than the recorded *localbest*, it is replaced. Additionally, if replacement occurs, the current position is recorded as *localbest position*.
4. For each particle, comparison of the present fitness with the global best fitness, *globalbest*. If the current fitness improves the *globalbest* fitness, it is replaced, and the current position is recorded as *globalbest position*.
5. Updating the velocity and the position[1] of the particle according to Eqs. 3 and 4:

$$v_{id}(t+\delta t) \leftarrow v_{id}(t) + c_1 \cdot Rand() \cdot (x_{id}^{localbest} - x_{id}) + c_2 \cdot Rand() \cdot (x_{id}^{globalbest} - x_{id}) \tag{3}$$

$$x_{id}(t+\delta t) \leftarrow x_{id}(t) + v_{id} \tag{4}$$

6. If an end execution criterion —fitness threshold or number of generations— is not met, back to the step number 2.

In the implementation of PSO algorithm, the c_1 and c_2 constants were established as $c_1 = c_2 = 1$ and the maximum velocity of particles $V_{max} = 2$ [4], being these values the most typical ones. The rest of the configuration used is: 100 particles as population size and 5,000 cycles.

3.2 Differential Evolution

DE was proposed by Storn and Price [6], [7] in 1997. It is a non-deterministic technique based on the evolution of a population of individuals representing candidate solutions. The generation of new individuals is carried out with two operators: mutation and recombination.

Mutation adds the proportional difference between two randomly-selected individuals to a third individual (also randomly-selected). With these three randomly-selected and different individuals: v_1, v_2 and v_3, a new individual w_i —termed *mutant vector*— is generated using Eq. 5.

$$w_i \; = \; v_1 + \mu \cdot (v_2 - v_3) \tag{5}$$

where μ is the *mutation rate*.

After the *mutation operator*, a second operator —termed *recombination operator*— is executed. A recombination on each individual v_i (target individual) to generate a trial individual u_i is performed. The trial vector, u_i, is constructed mixing w_i and v_i individuals, Eq. 6, under a predefined *recombination rate*, $C_r \in [0,1]$; or if the equality $j = j_r$ is met —being j an integer random number $j \in [1,D]$.

[1] Apparently, in Eq. 3, a velocity is added to a position. However, this addition occurs over a single time increment (iteration), so the equation keeps its coherency.

$$u_i(j) = \begin{cases} w_i(j) \text{ if } rand \leq C_r \text{ or } j = j_r; \\ v_i(j) \text{ otherwise.} \end{cases} \quad (6)$$

Finally the *selector operator* decides, based on the improvement of the fitness, whether the trial individual is accepted, and then replaces the target vector, or the trial individual is rejected, and then the target vector remains in the next generation.

In the implementation of DE algorithm, the mutation rate was established as $\mu = 0.5$ and the recombination rate as $C_r = 0.5$, being these values the most typical ones in the literature [6]. The rest of the configuration used is: 100 particles as population size and 5,000 cycles.

4 Production Setup

Diverse serial expansions were tested to fit the experimental data to the theoretical physical-meaningless curve. In spite of the equal a priori capacity, the Legendre polynomial —50 degrees in all serial expansions— serial expansion showed a major sensitiveness to reproduce the data behaviour and produced the lowest values of the fitness function.

According to the usual practice in adjustment of experimental data to theoretical curve, the chi-square test —χ^2—- has been chosen in this work [8] as fitness function. The lower the χ^2 value is, the closer the solution is to the objective —the fitter experimental data is to the theoretical curve—. Thus, the aim is to minimise χ^2.

Considering a standard fitting problem, where one is given a discrete set of N data points with associated measured errors σ, and is asked to construct the best possible fit to these data using a specific functional form for the fitting function, the most appropriated fitness function is the merit function χ^2, Eq. 7 [9]. Therefore, independently of the specific functional form chosen, the fitness function used in this work is χ^2, Eq. 7.

$$\chi^2 = \sum_{for\ all\ points} \left(\frac{y_{simulated} - y_{observed}}{\sigma_{observed}}\right)^2 \quad (7)$$

For each case —each EA and type of polynomial— a total of 25 tests were executed in order to reach the desired statistical relevance.

As pseudorandom number generator, a subroutine based on Mersenne Twister has been used [10].

In order to fairly compare the curves of the galaxies, a double normalization has been applied. First of all, the size of the galaxies has to be homogenized. For this normalization, the radius where the maximum velocity is reached is settled —in arbitrary units— at 0.1 units. Consequently, all the radii measured for the galaxy under modification are conveniently scaled.

Second of all, the maximum velocity of each galaxy is settled at 1 —in arbitrary units—-. As consequence, the rest of measured velocities are also appropriately scaled. Finally, resulting of the scaling in velocities, the velocity error must be rescaled proportionately to the velocity associated.

As result of this double normalization, all the curves have a common coordinate at (0.1, 1). Once the normalization process has proceeded, the extraction of a pattern representing all the curves can be executed. In Fig. 1 two figures are presented: in the left figure the complete observational data set, and in the right the data without the error bars. Particularly, the galaxy rotation curves used in this work were extracted from a large astronomical data set [11], covering approximately 60 galaxies, being involved a total of 5051 points.

5 Results and Analysis

It is well known in Evolutionary Computing that it is not possible to know a priori which EA will perform the best for a particular problem. For this reason, optimization problems are treated with a variety of techniques, retaining the best ones for further improvements.

In Fig. 2 —left— the comparative box plots of the best results for the algorithms PSO and DE are presented. As can be appreciated the PSO algorithm performs better than DE, in both: the absolute best result obtained after the 25 test, as well as the median of the samples. Therefore, the use of DE will be rejected for this problem.

The application of the Wilcoxon signed-rank test [12] to the data shown in the left panel of Fig. 2 indicates that the differences are significant from the statistical point of view for $\alpha = 0.05$.

In Fig. 2 —right— the evolution of the best result for each case studied is presented. In this figure, the evolution of PSO with Legendre polynomial can distinguished from the other cases by the rapid evolution during the first half of generations. However, for the second half the fitness evolution stagnates. The other two cases show a lower ability to evolve along the generations.

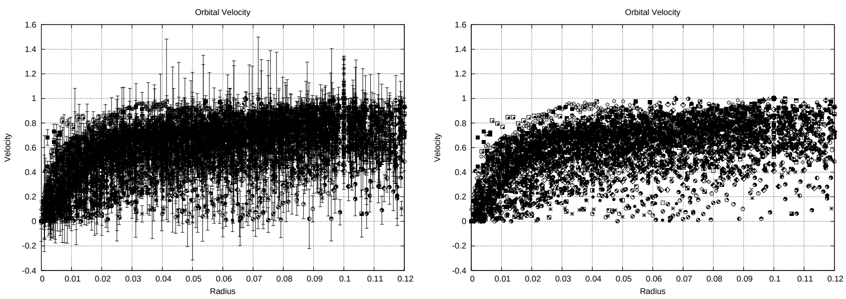

Fig. 1 All rotation curves doubly normalized. Panel (a, left) shows data with errorbars, while panel (b, right) without errorbars

In Fig. 3, two views of the absolute best result —the fittest solution to the observational data— are presented. As can be appreciated in the general view —left in Fig. 3—, the adjustment is far from the optimum for values of normalized radii higher than 0.2. This value corresponds to the double of the radius where the velocity reaches its maximum value. In this range [0.2, 1.0] few observational data exist, therefore, it is more difficult to fit accurately the functional form to the data. Probably the own nature of the Polynomial of Legendre, producing oscillation for these values, deteriorates the final result impeding finer adjustment.

On the contrary, in the inner segment [0, 0.2], the main part of the observational data are concentrated, and thus a better adjustment is expected. The observation of the area where most of the data are concentrated [0,0.2] —right in Fig. 3— shows an excellent adjustment to the observational data. As is appreciated, the functional form chosen accurately adjusts the observational data.

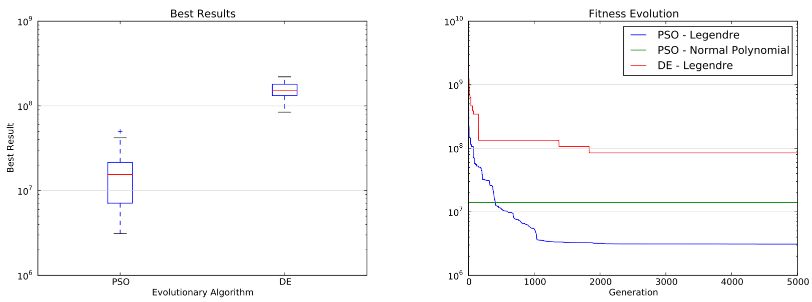

Fig. 2 Panel (a, left) shows the comparative box plots for the best results obtained for PSO and DE algorithms, while panel (b, right) shows the fitness evolution for the best result of each case studied

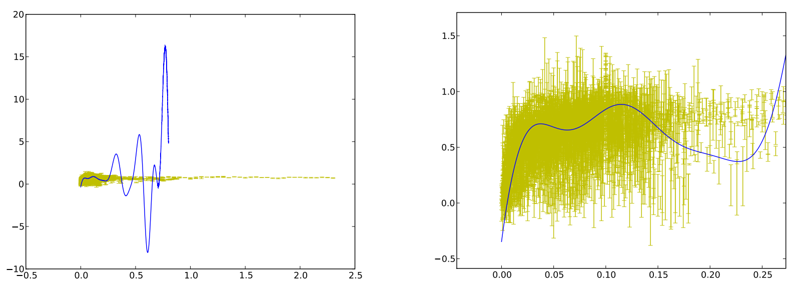

Fig. 3 Absolute best result —the fittest adjustment to observational data— obtained. Configuration used PSO with configuration of 100 particles and 5,000 cycles, and a series of Legendre Polynomials of 50 degrees. Panel (a, left) shows a general view, while panel (b, right) shows a closer view of the smaller radii

It can be concluded that a high-quality adjustment is produced in the area corresponding to the smaller radii [0, 0.2], being the adjustment less optimum for the external segment [0.2, 1.0].

6 Conclusions and Future Work

This paper deals with the use of Evolutionary Algorithms to adjust observational data —rotational curves of spiral galaxies— to specific functional forms. The numerical experiments performed show that PSO algorithm obtains more accurate results than DE algorithm. In general, the results obtained demonstrate the effectiveness of the application of Evolutionary Algorithms to cope with the extraction of essential information from huge volume of astronomical and astrophysical observational data.

The natural forthcoming step is to implement the population-diversification mechanisms necessary to avoid the stagnation of the fitness evaluation. Besides, the method to generate the initial population will be revisited, replacing the random generator by low-discrepancy numbers sequences generator. Finally, the checking of other EAs, as well as other functional forms to generate a fitter adjustment will be taken into account.

Acknowledgements This work has been partially supported by DGICYT grant AYA2010–21887–C04–02. Also, by the Comunidad de Madrid under grant CAM S2009/ESP-1496 (AstroMadrid) and by the Spanish MICINN under the Consolider-Ingenio 2010 Program grant CSD2006-00070: First Science with the GTC (http://www.iac.es/consolider-ingenio-gtc) which are acknowledged.

References

1. P. Charbonneau, The Astrophysical Journal Supplement Series **101**, 309 (1995)
2. E. Alba, M. Tomassini, IEEE Trans. Evolutionary Computation **6**(5), 443 (2002)
3. J. Kennedy, R.C. Eberhart, Proceedings of the IEEE International Conference on Neural Networks **IV**, 1942 (1995)
4. R.C. Eberhart, Y. Shi, J. Kennedy, *Swarm Intelligence (The Morgan Kaufmann Series in Artificial Intelligence)*, 1st edn. (Morgan Kaufmann, 2001)
5. R.C. Eberhart, J. Kennedy, (1995), pp. 39–43. DOI 10.1109/MHS.1995.494215
6. K.V. Price, R. Storn, J. Lampinen, *Differential Evolution: A practical Approach to Global Optimization* (Springer-Verlag, Berlin, Germany, 2005)
7. R. Storn, K. Price, J. of Global Optimization **11**(4), 341 (1997). DOI http://dx.doi.org/10.1023/A:1008202821328
8. D. Montgomery, G. Runger, *Applied Statistics and Probability for Engineers* (John Wiley and Sons Ltd, New York, USA, 2002)
9. W. Press, B. Flannery, S. Teukolsky, W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, 1992)
10. M. Matsumoto, T. Nishimura, ACM Trans. Model. Comput. Simul. **8**(1), 3 (1998)
11. I. Marquez, et al., Astron. Astrophys. **393**, 389 (2002). DOI 10.1051/0004-6361:20021036
12. S. García, A. Fernández, J. Luengo, F. Herrera, Soft Comput. **13**(10), 959 (2009)