Multi-Objective Optimisation in Time Series: Time Delay Agreement

Juan Carlos Cuevas-Tello[1] **and Hector G. Perez-Gonzalez**[1]
[1]Autonomous University of San Luis Potosi, Engineering Faculty, San Luis Potosi, SLP, Mexico
{cuevas,hectorgerardo}@uaslp.mx

Abstract— *Several time delay estimates have been reported for the quasar Q0957+561. They come from distinct data sets and published separately. This paper presents a methodology to estimate a single time delay given several data sets by using multi-objective optimisation. We use General Regression Neural Networks (GRNN) to estimate the time delay, which is one of the most accurate time delay estimators – and faster. For the time delay agreement, we use hill-climbing search. We found that the best agreement for the time delay on Q0957+561 is* $\Delta = 420$ *days.*

Keywords: Neural networks and applications, Heuristic searching methods, Applications: time series in astronomy

1. Introduction

Since it was predicted that the Hubble parameter can be estimated through time delays on gravitational lenses [1], many observation campaigns have been lunch since then [2], [3], [4], [5], [6], and new projects for ambitious surveys like Large Synoptic Survey Telescope (LSST) and the SuperNova Acceleration Probe (SNAP) devoted to study dark matter are in development. Moreover, current surveys like The Sloan Digital Sky Survey (SDSS) and Sloan Lens ACS (SLACS) are generating a tremendous amount of large monitoring data sets. The above surveys are not useful only to estimate the Hubble's parameter, because they are also important to study lensed supernovae (SNe) [7]. Therefore, time delay estimations become a big issue to study dark matter and microlensing.

So far methods to estimate time delays have been used along a monitoring campaign devoted to a single quasar [2], [3]. However, it is important to have a methodology that allows to estimate a time delay if one has several data sets coming from different surveys. In particular, it is well know that given a data set you can estimate a time delay which may be different from another data set, even if they come from the same source. It is the case for the most studied quasar Q0957+561 [2], [3], where a controversy started in the 90's and apparently stopped in 1997 with the work of Kundic et al. [2]. However, the definite time delay is based on a single data set (g-band). After that, more time delay estimates have been published for this quasar, e.g. see [4], [5], [8], [9].

Here we present a methodology to estimate time delay using several data sets from the same source. In particular we study data sets from the quasar Q0957+561. As we said above, each data set may give you a distinct time delay estimation, so we do multi-objective optimisation with hill climbing search. This allows us to have a time delay agreement given several data sets. The idea of multi-objective optimisation is that there is not only a single solution, then there is a set of solutions [10]. We found that the best time delay is $\Delta = 420$ days for the Q0957+561.

According to the best of our knowledge, this is the first approach to deal with this problem in time series via multi-objective optimisation. In practice, astronomers estimate the time delay separately for each pair of time series and by hand they study the time delay agreement for the same quasar.

The reminder of the paper is orginised as follows: the next section describes the data sets used in this research. In §3, we describe the proposed methodology. It follows the experiments and results section, and finally it comes the conclusions and future work.

2. Data sets

We do study six different data sets from the same quasar Q0957+561. The details are in Table 1 and the plots in Fig. 1. For DS1, The whole series was provided by R. Schild [11], private communication. The column labelled n corresponds to the amount of observations per data set. Through the third column, we specify the kind of data, optical o radio, and the Type means the filter and the frequency used to obtain such data sets. The Q0957+561 is a two images quasar, so there is either an offset or a ratio between the two components: image-A and image-B, which correspond to optical and radio data respectively.

3. Methodology

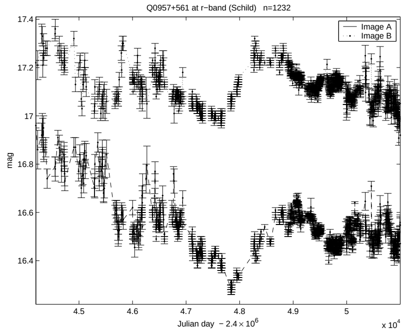
Assume that you have several data sets for the same quasar, which are obtained through a monitoring campaign manually or automatically (e.g. SNAP, LSST, SDSS), we denote them as:

$$d_i \quad i = 1, 2, ..., N \tag{1}$$

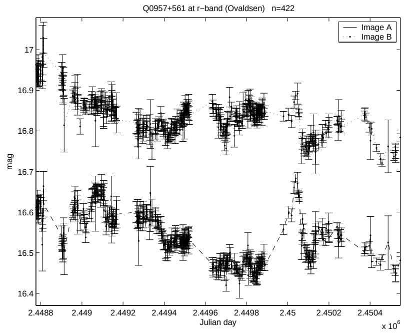
where N is the number of data sets, and D contains all data sets d_i. For each data set d_i, we have a time delay estimation

Table 1: Data sets: Q0957+561

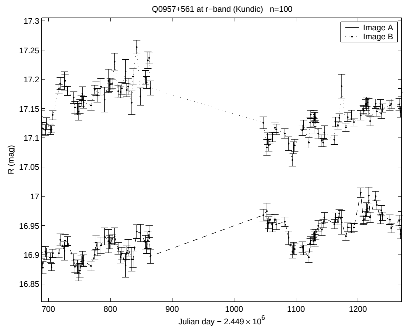
Id	n	Data	Type	Ratio/Offset	Monitoring Range	Ref
DS1	1232	optical	r-band	0.05	16/11/1979 – 4/7/1998	[11]
DS2	422	optical	r-band	0.076	2/6/1992 – 8/4/1997	[5]
DS3	100	optical	r-band	0.21	3/12/1994 – 6/7/1996	[2]
DS4	97	optical	g-band	0.117	3/12/1994 – 6/7/1996	[2]
DS5	143	radio	6cm	1/1.43	23/6/1979 – 6-Oct-1997	[3]
DS6	58	radio	4cm	1/1.44	4/10/1990 – 22/9/1997	[3]



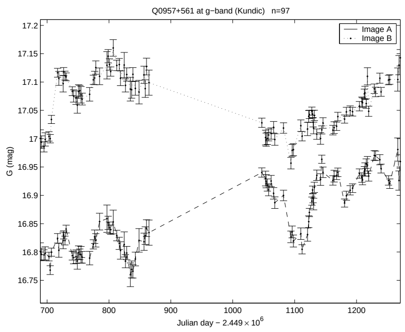
(a) DS1



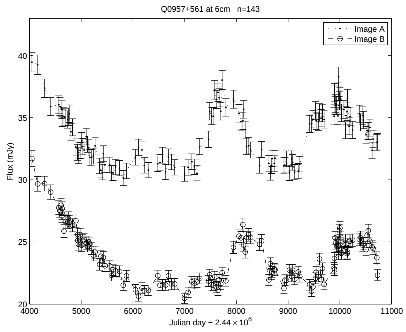
(b) DS2



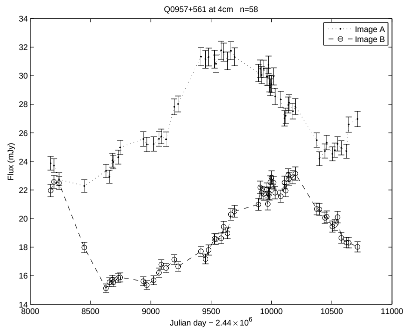
(c) DS3



(d) DS4



(e) DS5



(f) DS6

Fig. 1: Data sets: Q0957+561. Image-A on DS1 is shifted up 0.6mag for clarity; image-A of DS2 was shifted up 0.25mag; for DS4, the shift up is 0.05mag. For more details on these data sets see Table 1.

Δ_i and also a set of parameters P_i. The size of the set of parameters P_i depend on the method used to estimate the time delay. The best fitting error associated to each Δ_i is denoted by e_i.

The first step is the initialisation of all time delays Δ_i^0 and parameters involved P_i^0, see Algorithm 1.

The next step is obtaining the set of solutions, see Algorithm 2.

What really the Algorithm 2 does, it is to find a set of solutions rather than a single solution. In other words, we are performing multi-objective optimisation.

On one hand, the time delay estimation methods are looking for the best fitting given a time delay Δ and two time series (image-A and image-B for Q0957+561). This works for a single data set, and this is one objective In Algorithm

Algorithm 1: Initialization (D)

```
for each d_i ∈ D
{
        Read file for d_i ;
        Estimate Δ_i^0, e_i^0 and P_i^0 ;
}
```

Algorithm 2: Find set of solutions (D, Δ_i^0, e_i^0, P_i^0)

```
/* MAXG is the # of iterations to optimise E[Δ̄ − Δ_i]                    */
Δ̄ ← E[Δ_i^0] ;
ē ← E[e_i^0] ;
Δ_i ← Δ_i^0 ;
P_i ← P_i^0 ;
g ← 1 ;
plot Δ̄, ē ;
while E[Δ̄ − Δ_i] > 0 and ē > 0 and g <MAXG
{
        for each d_i ∈ D
        {
                Perturb P_i to obtain P̃_i ;
                Get Δ̃_i with P̃_i ;
                /* Hill Climbing search, optimise E[Δ̄ − Δ_i]              */
                if E[Δ̄ − Δ_i] > E[Δ̄ − Δ̃_i]
                {
                        P_i ← P̃_i ;
                        Δ_i ← Δ̃_i ;
                        Δ̄ ← E[Δ_i] ;
                }
        }
        plot E[Δ̄ − Δ_i], ē ;
        g ← g + 1 ;
}
```

2, we refer to the best fitting as e_i, where $\overline{e} = E[e_i] = 1/N \sum_i^N e_i$. So we assume that e_i is calculated when the best time delay Δ_i is estimated. Most of the time delay methods start with a set of suggested time delays, then the best time delay is found when the fitting error e_i is minimal.

On the other hand, here we want that all time delays, for each data set d_i, converge to a single time delay. Remember that the theory predicts that the time delay of the same gravitational lens must be the same, regardless the observation methodology (e.g. optical or radio telescopes). Then, the other objective is to get the minimal $E[\overline{\Delta} - \Delta_i]$, where $E[\overline{\Delta} - \Delta_i] = 1/N \sum_i^N (\overline{\Delta} - \Delta_i)$ and $\overline{\Delta} = E[\Delta_i] = 1/N \sum_i^N \Delta_i$.

Finally, the Algorithm 2 ends when either $\overline{e}$ is zero or $E[\overline{\Delta} - \Delta_i]$ is zero. Otherwise, Algorithm 2 stops when the maximum number of iterations (MAXG) has been reach.

4. Experiments and Results

We have two accurate methods to estimate time delays: i) EA-M-CV method is an evolutionary algorithm with mixed representation (integer and real numbers), and a objective function based on kernel formulation and cross-validation [12]. This method models a single underlying function that generates the two images plus the delay Δ between them. ii) GRNN method, which is based on radial basis functions (RBF) from neural networks theory [13]. As many methods in the literature, GRNN method mixes the two images into a single one given a trial time delay Δ, and when the best fitting e is reached the best time delay comes up (see [14]). The only parameter to estimate with GRNN method is the spread, which is the width of the basis functions ω, Gaussian functions. Therefore, $P_i = \omega_i$.

In practice, EA-M-CV with a population of 300 individuals and 150 generations was used to estimate a time delay on DS4 (see §2. Since EA-M-CV is stochastic, it is necessary to run several realisations, and 10 realisations takes about 3 hours (Matlab program). GRNN takes only 20 minutes (Matlab program), and GRNN with a C program takes only one second. All these experiments with the same data set (DS4) and running on the same machine (MacBook Pro, 2.4Ghz, Intel Core 2 Duo, 4Gb RAM, MacOS-X ver. 10.5.6).

Consequently, all experiments to test the methodology in §3 are performed with GRNN (C program). The results after performing the Algorithm 1 on the data sets in Table 1 are shown in Table 2. We test ω in the range of 5 to 8 with increments of 0.1 units, and trial time delays in the range of 400 to 449 with unitary increments. For perturbing P_i to obtain $\widetilde{P}_i$, in Algorithm 2, we generate random numbers with a Gaussian distribution, zero mean and standard deviation set to 5.

The results after 100 iterations (MAXG=100) are in Table 3. The total elapsed time was 4421 seconds (1 hour, 13

Table 2: Time Delays after initialisation

DS	Δ^0	e	$P_i = \omega_i$
1	428	8.510292e-04	7.3
2	427	2.491830e-04	7.45
3	426	2.112661e-04	7.7
4	420	3.081128e-04	7.1
5	449	9.356649e-02	6.5
6	402	8.736157e-02	8.0
	$\overline{\Delta} = 425.33$	$\overline{e} = 0.03$	$E[\overline{\Delta} - \Delta_i] = 9.0$

Table 3: Final Time Delays after 100 iterations

DS	Δ^0	e	$P_i = \omega_i$
1	422	2.666853e-04	0.84
2	424	1.514531e-04	1.91
3	420	1.052118e-04	2.80
4	420	3.081128e-04	7.10
5	419	4.397565e-03	0.25
6	416	1.792195e-03	0.66
	$\overline{\Delta} = 420.17$	$\overline{e} = 1.17E-3$	$E[\overline{\Delta} - \Delta_i] = 1.5$

minutes). The best time delay by using this methodology is $\Delta = 420$ days.

The set of solutions are depicted in Fig. 2. The first solution, during initialisation stage, is when $E[\overline{\Delta} - \Delta_i] = 9.0$, on the right. The best solution is when $E[\overline{\Delta} - \Delta_i] = 1.5$, the best time delay agreement. Note that in Fig. 2, following the initialisation stage, at the right, the following solutions are getting closer to the best solution in terms of $E[\overline{\Delta} - \Delta_i]$, but the fitting error is increasing. This is the tradeoff in multi-objective optimisation.

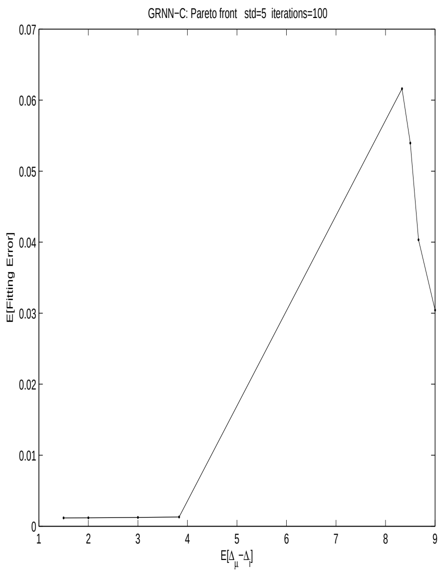


Fig. 2: Set of solutions, starting from the right-hand side to the left-hand side.

5. Conclusions and Future Work

The GRNN method is an accurate and fast method for time delay estimation. The EA-M-CV is robust time delay estimation but it cannot manage to analyse DS1 because the data set is large (more than one thousand observations).

Most of the data sets converge or are close to the time delay of 420 days. The only data set that underestimate this time delay is DS6. In Fig. 3, we can see that in the curves showing where is the best time delay. In Fig 3a, when ω is low, several time delay are feasible: 400, 416 and 420 days because the curve is sharp. However, at the initialisation stage, Δ is 402 with $\omega = 8$ (see Fig. 3c). Now some value in between, that is, when $\omega = 6.5$ we can see that there is some feature at $\Delta = 420$, that is, the agreed time delay for all other data sets (see Fig. 3b).

The set of solutions in Fig. 2 may change since the perturbing method used in the experiments is stochastic (see Algorithm 2). We did several simulations and most of them converge to the solution shown in Fig. 2.

As part of the future work, it is desirable to compare the performance of this methodology with a evolutionary-based methodology, that is, evolutionary computation (EC). In theory, EC is a global optimisation algorithm that may find a better set of solutions.

Another idea is to fix the initial Δ_i^0 into a single time delay Δ^0 (say the mean) and then plug it back into the variable representing the time delay estimation into the model, regardless the time delay estimation method. Then, for each data set obtain the log likelihood. The best time delay it may come with the sum of all likelihoods.

6. Acknowledgements

This paper was possible thanks to Peter Tino and Somak Raychaudhury our collaborators in this topic at University of Birmingham, UK. This work has been partially supported by grants C09-FAI-03-35.35, PROMEP/103.5/08/1696 and PROMEP/103.5/09/4280.

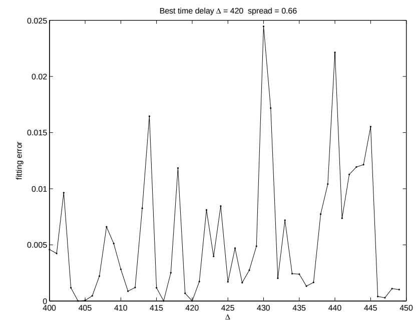
References

[1] S. Refsdal, "On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect," *Monthly Notices of The Royal Astronomical Society*, vol. 128, pp. 307–+, 1964.

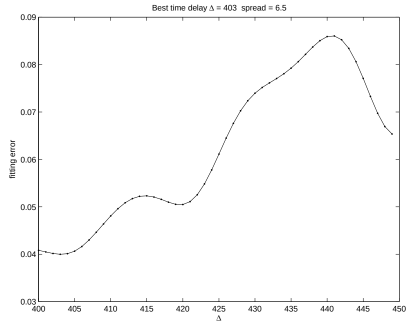
[2] T. Kundic, E. Turner, W. Colley, J. Gott-III, J. Rhoads, Y. Wang, L. bergeron, K. Gloria, D. Long, S. Malhorta, and J. Wambsganss, "A robust determination of the time delay in 0957+561A,B and a measurement of the global value of Hubble's constant," *Astrophysical Journal*, vol. 482, no. 1, pp. 75–82, 1997.

[3] D. Haarsma, J. Hewitt, J. Lehar, and B. Burke, "The radio wavelength time delay of gravitational lens 0957+561," *Astrophysical Journal*, vol. 510, no. 1, pp. 64–70, 1999.

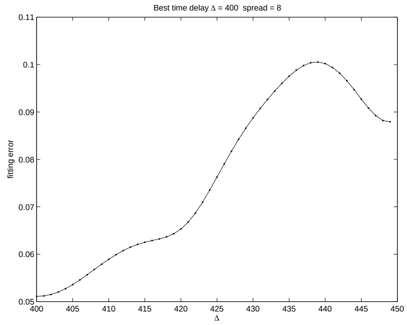
[4] W. Colley, R. Schild, C. Abajas, D. Alcalde, Z. Aslan, I. Bikmaev, V. Chavushyan, L. Chinarro, J. Cournoyer, R. Crowe, V. Dudinov, A. Evans, Y. Jeon, L. Goicoechea, O. Golbasi, I. Khamitov, K. Kjernsmo, H. Lee, J. Lee, K. Lee, M. Lee, O. Lopez-Cruz, E. Mediavilla, A. Moffat, R.Mujica, A. Ullan, J. Munoz, A. Oscoz, M. Park, N. Purves, O. Saanum, N. Sakhibullin, M. Serra-Ricart, I. Sinelnikov, R. Stabell, A. Stockton, J. Teuber, R. Thompson, H. Woo, and A. Zheleznyak, "Around-the-clock observations of the



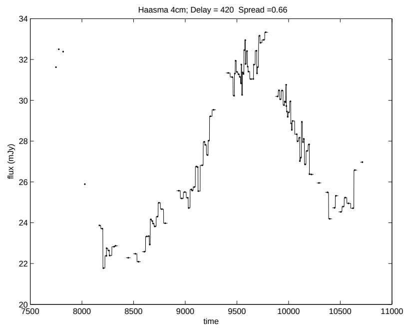
(a) DS6, $\omega = 0.66$



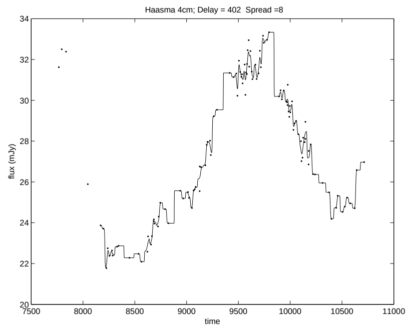
(b) DS6, $\omega = 6.5$



(c) DS6, $\omega = 8.0$



(d) DS6, $\Delta = 420$, $\omega = 0.66$



(e) DS6, $\Delta = 402$, $\omega = 8$

Fig. 3: DS6: Q0957+561. a)-c) show the curves showing the best time delay. d)-e) depicts the reconstructions of the combined curve given a time delay Δ and $P_i = \omega$.

Q0957+561 A,B gravitationally lensed quasar. II. results for the second observing season," *Astronomy and Astrophysics*, vol. 587, no. 1, pp. 71–79, 2003.

[5] J. Ovaldsen, J. Teuber, R. Schild, and R. Stabell, "New aperture photometry of QSO 0957+561; application to time delay and microlensing," *Astronomy and Astrophysics*, vol. 402, no. 3, pp. 891–904, 2003.

[6] A. Eigenbrod, F. Courbin, C. Vuissoz, G. Meylan, P. Saha, and S. Dye, "COSMOGRAIL: The COSmological MOnitoring of GRAvItational Lenses. I. How to sample the light curves of gravitationally lensed quasars to measure accurate time delays," *Astronomy and Astrophysics*, vol. 436, pp. 25–35, June 2005.

[7] G. Dobler and C. R. Keeton, "Microlensing of Lensed Supernovae," *Astrophysical Journal*, vol. 653, pp. 1391–1399, Dec. 2006.

[8] J. Cuevas-Tello, P. Tiňo, and S. Raychaudhury, "How accurate are the time delay estimates in gravitational lensing?" *Astronomy and Astrophysics*, vol. 454, pp. 695–706, August 2006.

[9] ——, "A kernel-based approach to estimating phase shifts between irregularly sampled time series: an application to gravitational lenses," in *Machine Learning: ECML 2006*, ser. Lecture Notes in Artificial Intelligence (LNAI 4212). Springer-Verlag, Sept. 2006, pp. 614–621.

[10] *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed. Springer, 2007.

[11] R. Schild and D. Thomson, "The Q0957+561 time delay from optical data," *Astronomical Journal*, vol. 113, no. 1, pp. 130–135, 1997.

[12] J. Cuevas-Tello, "Estimating time delays between irregularly sampled time series," Ph.D. dissertation, School of Computer Science, University of Birmingham, 2007, http://etheses.bham.ac.uk/88/.

[13] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[14] R. Gonzalez-Grimaldo and J. Cuevas-Tello, "Analysis of time series with neural networks," in *Proceedings of 7th Mexican International Conference on Artificial Intelligence (MICAI)*. IEEE Computer Society, Nov. 2008, pp. 131–137.