Astron. Astrophys. 337, 681–684 (1998)

ASTRONOMY AND ASTROPHYSICS

Estimation of multiple time delays in complex gravitational lens systems

J. Pelt[1,6], **J. Hjorth**[2,6], **S. Refsdal**[3,6], **R. Schild**[4,6], **and R. Stabell**[5,6]

[1] Tartu Observatory, EE-2444 Tõravere, Estonia
[2] NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
[3] Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg-Bergedorf, Germany
[4] Harvard-Smithsonian Center for Astrophysics, MS-19 60 Garden Street, Cambridge, MA 02138, USA
[5] Institute of Theoretical Astrophysics, P.O. Box 1029, Blindern, N-0315 Oslo, Norway
[6] Centre for Advanced Study, Drammensveien 78, N-0271, Oslo, Norway

Received 27 March 1998 / Accepted 26 June 1998

Abstract. We generalize the nonparametric dispersion minimization method of time delay estimation for the case when more than two quasar images are available. This allows us to analyse time delays in complex gravitational lens systems in a model independent manner. We apply the new method to the recent observations of the quadruply imaged quasar PG 1115+080 and show that our nonparametric estimates roughly coincide with the values obtained with optimal filtering and smoothing techniques. Because our method uses a minimal set of assumptions and is computationally simpler it may be a method of choice, especially for the early stages of an investigation when the available data for a new lens system are scarce.

Key words: methods: statistical – quasars: individual: PG 1115+080 – gravitational lensing

1. Introduction

The dispersion minimization method developed in Pelt et al. (1994, Paper I) and Pelt et al. (1996, Paper II) was successfully used to estimate the time delay for the gravitational lens system QSO 0957+561 A,B. The method was also used by other research groups to analyse different data sets (Haarsma et al. 1997; Kundić et al. 1997; Oscoz et al. 1997; Goicoechea et al. 1998). It is important to check the applicability of this simple algorithm in more demanding situations.

In this short note we show that a fairly staightforward generalization of the method allows us to estimate time delays in systems with more than two images. For illustration purposes we use photometric data for the quadruple gravitational lens PG 1115+080. There is a difference between the two sets of time delay estimates published so far: first by Schechter et al. (1997, below S97) and then by Barkana (1997, below B97). We will show that the dispersion minimization approach tends to give higher weight to the S97 set. However, the data available are insufficient to yield conclusive results.

Send offprint requests to: R. Stabell

2. Method

Let A and $B^{(r)}, r = 1, \ldots, R$ be magnitude measurements of the $R+1$ different components of a gravitational lens system. We single out one arbitrary image (A) for a simplification of the description. The data model for the reference image A is then:

$$A_i = q(t_i) + \epsilon_A(t_i),\ i = 1, \ldots, N^{(A)}$$

and for additional (shifted in time by $\tau^{(r)}$) images:

$$\begin{aligned} B_j^{(r)} \ = \ & q(t_j - \tau^{(r)}) + l^{(r)}(t_j) + \epsilon^{(r)}(t_j), \\ & j = 1, \ldots, N^{(r)}, \\ & r = 1, \ldots, R, \end{aligned}$$

where $q(t)$ is the intrinsic variability of the quasar, the $l^{(r)}(t)$ are constant differences in magnitudes (macrolensing and/or absorption) or smooth nuisance components (due to possible microlensing), and $\epsilon^{(r)}(t)$ are observational errors. For simplicity we will in the following discussion assume that microlensing is absent or is constant during the observing season ($l^{(r)}(t) = l_0^{(r)}$).

For every time delay and magnitude difference set

$$\tau^{(r)}, l_0^{(r)}, r = 1, \ldots, R,$$

we can generate a combined light curve by appropriately shifting the original light curves in time and in magnitudes:

$$C_k(t_k) = \begin{cases} A_i, & \text{if } t_k = t_i, \\ B_j^{(r)} - l_0^{(r)}, & \text{if } t_k = t_j^{(r)} + \tau^{(r)}, \end{cases}$$

where $k = 1, \ldots, N$ and $N = N^{(A)} + N^{(1)} + \ldots + N^{(R)}$. The dispersion spectra

$$\begin{aligned} D^2(\tau^{(1)}, \ldots, \tau^{(R)}) \ = & \\ & \min_{l_0^{(1)}, \ldots, l_0^{(R)}} \ D^2(\tau^{(1)}, \ldots, \tau^{(R)}, l_0^{(1)}, \ldots, l_0^{(R)}), \end{aligned}$$

are then computed for every set of $\tau^{(r)}, r = 1, \ldots, R$ from a search space. For the experiments described in this report we chose dispersion estimates from Paper I in the form

$$D_1^2(\tau^{(1)}, \ldots, \tau^{(R)}) = \min_{l_0^{(1)}, \ldots, l_0^{(R)}} \frac{\sum_{k=1}^{N-1} W_{k,k+1} G_{k,k+1} (C_{k+1} - C_k)^2}{2 \sum_{k=1}^{N-1} W_{k,k+1} G_{k,k+1}},$$

and from Paper II

$$D_2^2(\tau^{(1)}, \ldots, \tau^{(R)}) = \min_{l_0^{(1)}, \ldots, l_0^{(R)}} \frac{\sum_{n=1}^{N-1} \sum_{m=n+1}^{N} W_{n,m} S_{n,m} G_{n,m} (C_n - C_m)^2}{2 \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} W_{n,m} S_{n,m} G_{n,m}},$$

where the $S_{n,m}$ are used to downweight or exclude pairs with long time difference

$$S_{n,m} = \begin{cases} 1 - \frac{|t_n - t_m|}{\Delta t}, & \text{if } |t_n - t_m| \le \Delta t, \\ 0, & \text{if } |t_n - t_m| > \Delta t \end{cases}.$$

$W_{i,j}$ are combined statistical weights computed from the standard errors σ_i, σ_j of the original observations

$$W_{i,j} = \frac{W_i W_j}{W_i + W_j} = \frac{1}{\sigma_i^2 + \sigma_j^2},$$

and $G_{k,k+1}$ are weights which switch off pairs of C_k, C_{k+1} which are coming from one and the same curve (A or $B^{(r)}$). The computational scheme is quite similar to the standard dispersion minimization algorithms from Papers I and II. However, now the search space for a solution is multidimensional and the magnitude shifts $l_0^{(r)}, r = 1, \ldots, R$ must be determined from a least squares solution of an $R * N$ dimensional matrix for every particular combination of the time shifts $\tau^{(1)}, \ldots, \tau^{(R)}$. To speed up the computations, it is reasonable to evaluate the dispersion spectra first on a rough grid of trial shifts and then proceed with a refined search in the surroundings of the provisional minima.

It is important to note that the choice of downweighting function $S_{n,m}$ and bandwidth parameter Δt in the proposed scheme is far from being unique. When working with actual data it is reasonable to vary the bandwith parameter and may be also the downweighting function (as done in Paper II) to get a feeling for the robustness of the results obtained. Our experience shows that for the right time delay system there is a wide range of experimental setups which all give very similar dispersion spectra (see also Sect. 4).

For both dispersion estimates D_1^2 and D_2^2 we use similar normalization, so that their actual value estimates the true variance around the hypothetical light curve.

3. Refinements

The proposed scheme allows several straightforward modifications. We can use low degree polynomials as nuisance parameters $l^{(r)}$ to model time dependent microlensing. In this case the computational scheme remains practically the same, but there will be some extra columns in the least sqares estimation matrix. Slightly more complicated is the introduction of the additional multiplicative factors (as used for instance in B97). Instead of the original curves $B^{(r)}$ we use curves with multiplicative factors $a^{(r)} B^{(r)}$, with corresponding reformulation of the dispersion estimates. Because multiplicative amplifications $a^{(r)}$ will have an effect also on the weights, the scheme becomes iterative. First we estimate free parameters using the original weights, then we plug in corrected weights

$$W_{i,j}^* = \frac{W_i W_j}{(a^{(r)})^2 W_i + W_j},$$

where $a^{(r)}$ is estimated (by least squares minimization) in the previous iteration as a multiplicative factor for an image r, and repeat all this cyclically until convergence is obtained. We will demonstrate the use of these techniques in a subsequent paper.

4. Application to the gravitational lens system PG 1115+080

To check the proposed method we used the data set from S97 which was also used in B97. This allows us to compare the results from three different approaches. Because our method does not use interpolation we were able to retain the last observation in the data set, which was excluded by other researchers. However, it was only marginally usable in our search space of delays from $[-40, 40] \times [-40, 40]$ days because of the gap of 37 days in between the last and penultimate nights. As in S97 and B97 we used two images (B and C) and the combined image $A = A1 + A2$ of the badly resolved images $A1$ and $A2$. Quite arbitrarily we chose A as a reference image in our scheme. We used exactly the same corrected error estimates as S97.

The application of the totally parameter free dispersion estimate D_1^2 gave us the strongest minimum at $D_1^2 = 0.0001631$ for $t_{BA} = 10.0, t_{AC} = 15.9, t_{BC} = 25.9$ (here and below we use notations for time delays $t_{BA} = -t_{AB} = -\tau^{(r)}$ for a particular r). There were at maximum 71 pairs and at minimum 52 pairs which were included in the summation. In Fig. 1 a contrast enhanced plot

$$\rho(t_{BA}, t_{CA}) = \frac{1}{\frac{D^2 - D^2_{min}}{D^2_{max} - D^2_{min}} + 0.02},$$

of the peaks in the spectrum $D_1^2(t_{BA}, t_{CA})$ is depicted. The maxima in $\rho(t_{BA}, t_{CA})$ indicate minima of the dispersion spectra. Because of the enhanced contrast (emphasizing peaks above 98% of the total variability level) only the strongest dispersion minima can be seen in the plot. It is seen that there are several peaks with practically the same strength. Consequently the totally parameter free version of the dispersion estimate does not help conclusively to single out one of the peaks in the search space. There are simply too few data pairs to be averaged over to get a stable spectrum. However when we include more pairs in the averaging, using dispersion estimate D_2^2, the picture becomes clearer. In Table 1 the results for D_2^2 runs with different

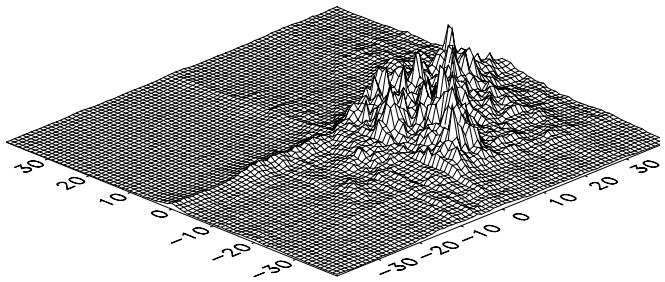

Fig. 1. Contrast enhanced plot $\rho(t_{BA}, t_{CA})$ of the dispersion spectrum D_1^2

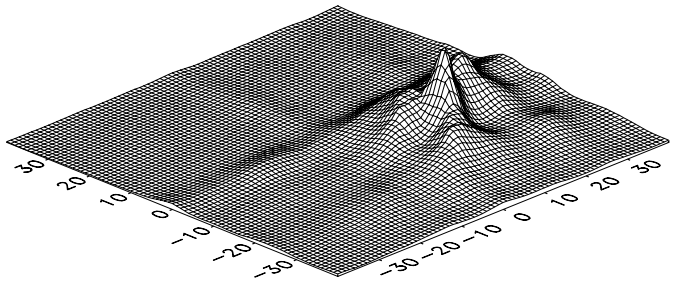

Fig. 2. Contrast enhanced plot $\rho(t_{BA}, t_{CA})$ of the dispersion spectrum D_2^2, $\Delta t = 5.5$

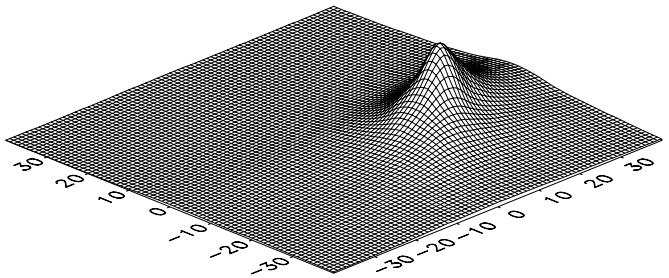

Fig. 3. Contrast enhanced plot $\rho(t_{BA}, t_{CA})$ of the dispersion spectrum D_2^2, $\Delta t = 15.5$

downweighting parameter Δt on the rough grid of delays (with one day step) are depicted. We can see that for a wide range

Table 1. Rough estimates for different values of Δt

Δt	t_{BA}	t_{AC}	t_{BC}	D_2^2	N_{min}	N_{max}
3.5	14	12	26	0.0001422	87	345
4.5	14	12	26	0.0001521	128	371
5.5	14	12	26	0.0001640	160	402
6.5	14	12	26	0.0001757	189	441
7.5	14	12	26	0.0001868	226	471
8.5	14	11	25	0.0001953	257	510
9.5	14	10	24	0.0002015	291	547
10.5	14	10	24	0.0002066	316	573
11.5	14	11	25	0.0002116	357	616
12.5	15	11	26	0.0002161	395	663
13.5	15	10	25	0.0002208	447	683
14.5	15	10	25	0.0002255	485	708
15.5	16	10	26	0.0002296	520	753
16.5	16	10	26	0.0002351	555	783
17.5	16	11	27	0.0002412	587	815
18.5	17	10	27	0.0002474	614	856
19.5	17	10	27	0.0002547	645	892
20.5	16	11	27	0.0002627	673	937
21.5	15	11	26	0.0002707	715	977
22.5	15	11	26	0.0002789	740	1009
23.5	15	12	27	0.0002867	762	1039
24.5	15	12	27	0.0002954	793	1080
25.5	16	11	27	0.0003041	817	1113
26.5	16	11	27	0.0003127	850	1140
27.5	17	11	28	0.0003211	886	1187
28.5	18	11	29	0.0003300	925	1219
29.5	19	10	29	0.0003396	961	1256

(3.5-27.5 days) of the Δt values the delays at absolute minimum of D_2^2 are practically the same and belong to the subset $[14, 17] \times [-12, -10]$. In Figs. 2 and 3 we see how the contrast enhanced plot $\rho(t_{BA}, t_{CA})$ changes with the value of Δt. For $\Delta t = 5.5$ the plot is multipeaked and somewhat similar to the case of D_1^2 whereas for $\Delta t = 15.5$ only one single peak can be clearly observed. (There are of course more peaks with lower amplitude, but they are damped out by the contrast enhancing transform.) The refined search in a square $[10, 20] \times [-15, -5]$ with $\Delta t = 15.5$ gave us a set of delays $t_{BA} = 15.5 \pm 1.8, t_{AC} = 10.3 \pm 1.9, t_{BC} = 25.8 \pm 2.4, r_{ABC} = 0.66 \pm 0.15$, where $r_{ABC} = t_{AC}/t_{BA}$. We estimated errors for the time delays using the bootstrap method described in Papers I and II. These error brackets are quite formal, and somewhat optimistic, because of their dependence on the downweighting parameter Δt. The estimates with bigger Δt tend to be more stable (the errors computed using bootstrap are smaller), but they can also contain larger bias due to the over-smoothing. The same is true for the method used in B97. Different parameterizations will give different error brackets. We consider the ± 2 day stability of our time delay estimates for a wide range of only one single input parameter, Δt, a more relevant precision indicator than formal Monte-Carlo type calculations.

When we compare our results with those of S97 and B97, we find a rather good agreement with the S97 delay set: $t_{BA} = 14.3, t_{AC} = 9.4 \pm 3.4, t_{BC} = 23.7 \pm 3.4, r_{ABC} = 0.7 \pm 0.3$ (method of Press et al. 1992). The agreement with B97 estimates $t_{BA} = 11.7^{+1.5}_{-1.6}, t_{AC} = 13.3^{+0.9}_{-1.0}, t_{BC} = 25.0^{+1.5}_{-1.7}, r_{ABC} = 1.13^{+0.18}_{-0.17}$ (smoothing parameter $L = 20$, bigger value corresponds to less smoothing) is not so good. If we adopt another value for the smoothing parameter in the B97 method, $L = 10$, then the agreement between the two methods is better ($t_{BA} = 14.0^{+1.6}_{-1.6}, t_{AC} = 12.5^{+0.9}_{-0.9}, t_{BC} = 26.5^{+1.7}_{-1.7}, r_{ABC} = 0.89^{+0.12}_{-0.12}$). The strongest minima for the D_1^2 statistic occured near the B97 solution with smoothing parameter value $L = 30$. Because of the apparent instability of the D_1^2 spectrum, we do not want to stress this coincidence too much. If we for the moment put aside the additional assumptions in the B97 analysis, then we can say that our nonparametric method (with D_2^2) tends to agree with the B97 method with stronger smoothing. Part of the differences between the two methods can be ascribed to

the fact that the B97 method takes into account probable correlated errors resulting from inclusion of the comparison star *B magnitudes into the estimates of A1+A2, B and C brightnesses. Barkana also included an additional magnification parameter into his model to describe conjectured "static" microlensing. We will present a more complete analysis of the intricate photometry of PG 1115+080 which takes into account these refinements in a subsequent paper.

5. Discussion

There are several advantages to the dispersion minimization scheme over the more complicated time delay estimation methods.

1. The method is essentially parameter free. Even when particular runs with the D_2^2 statistic are done with a fixed value of the downweighting parameter Δt the overall conclusions are always drawn from the full series of computations with different values of Δt.
2. The weights $G_{n,m}$ allow one to use only the information in the observed data relevant to the time delay estimation. In the methods by Press et al.(1992) and Barkana (1997) part of the fit quality is determined by the placement of the observed points of one or other single brightness curve.
3. The method provides a simple way to take into account heterogenity of the data sets and include different models for nuisance parameters (see Papers I and II).
4. The method avoids any interpolation into the gaps in the original data sets. As has been shown in Paper I this can be extremely dangerous in some cases.
5. The method is also free from bias introduced by variable amount of overlaps of the multiple lightcurves for different time delays.

There are of course typical pitfalls for the nonparametric methods. In general the computed estimates will have larger error bars when compared with estimation schemes with fully specified (and correct) models. For some sampling schemes the method can unexpectedly break down, because of odd sampling. This is a well known problem with other schemes too (compare Press et al. 1992 and Kundić et al. 1997). There is always a certain trade-off between dispersion of the estimates and the bias.

Nevertheless, we think that time delay estimates computed using the proposed scheme can be a good and safe starting point before a more detailed analysis.

In conlusion, we have demonstrated that the method of dispersion minimization can be easily generalized for the case, where light curves for more than two images of a gravitationally lensed quasar are available. The application of the new method gave consistent results in the case of the gravitational lens system PG 1115+080.

Acknowledgements. We would like to thank Paul L. Schechter for providing the data and his comments on them, and Nordic Scholarship Scheme for the Baltic Countries and Northwest Russia who through the Research Council of Norway supported financially one of the authors (J.P.).

References

Barkana R., 1997, ApJ 489, 21 (B97)
Goicoechea L.J., Oscoz A., Mediavilla E., Buitrago J., Serra-Ricart M., 1998, ApJ 492, 74
Haarsma D.B., Hewitt J.N., Lehár J., Burke B.F., 1997, ApJ 479, 102
Kundić T., Turner E.L., Colley W.N., et al., 1997, ApJ 482, 75
Oscoz A., Mediavilla E., Goicoechea L.-J., Serra-Ricart M., Buitrago J., 1997, ApJ 479, L89
Pelt J., Hoff W., Kayser R., Refsdal S., Schramm T., 1994, A&A 256, 775 (Paper I)
Pelt J., Kayser R., Refsdal S., Schramm T., 1996, A&A 305, 97 (Paper II)
Press W.H., Rybicki G.B., Hewitt J.N., 1992, ApJ 385, 404
Schechter P.L., Bailyn C.D., Barr R., et al., 1997, ApJ 475, L85 (S97)