DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS
16

DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS
16

ANTI HIRV

Estimation of time delays from light curves of gravitationally lensed quasars

TARTU UNIVERSITY PRESS

This study was carried out at the Tartu Observatory, Estonia.

The Dissertation was admitted on June 17, 2011, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in physics (astrophysics), and allowed for defence by the Council of the Institute of Physics, University of Tartu.

Supervisor:	Ph.D. Jaan Pelt, Tartu Observatory, Estonia
Opponents:	Prof. Lutz Wisotzki, Leibniz-Institut für Astrophysik (AIP), Potsdam, Germany
	Prof. dr. Luitje Vincent Ewoud Koopmans, Kapteyn Astronomical Institute, Groningen, the Netherlands
Defence:	August 26, 2011, at the University of Tartu, Estonia

ISSN 1406-0302
ISBN 978-9949-19-758-3 (trükis)
ISBN 978-9949-19-759-0 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 438

Contents

List of original publications

This thesis is based on the following publications:

I Hirv A., Eenmäe T., Liimets T., Liivamägi L. J. and Pelt J., *"Estimation of time delays from unresolved photometry"*, 2007, A&A, 464, 471

II Hirv A., Eenmäe T., Liivamägi L. J. and Pelt J., *"Estimation of time delays from two blended light curves of gravitational lenses"*, 2007, Baltic Astronomy, 16, 241

III Hirv A., Olspert N. and Pelt J., *"Towards the automatic estimation of gravitational lenses' time delays"*, 2011, Baltic Astronomy, accepted for publication on 23 May 2011, e-print arXiv:1105.5991

The dissertant started his Doctor of Philosophy program in 2003 with observational studies of highly evolved stars. The publications related to this work are:

IV Kipper T., Klochkova V. G., Annuk K., Hirv A., Kolka I., Leedjärv L., Puss A., Skoda P. and Slechta M., *"The peculiar variable V838 Monocerotis"*, 2004, A&A, 416, 1107

V Hirv A., Annuk K., Eenmäe T., Liimets T., Pelt J., Puss A. and Tempel M., *"Orbital elements and mass-loss rate of V 444 Cyg"*, 2006, Baltic Astronomy, 15, 405

VI Nugis T., Annuk K. and Hirv A., *"Mass loss parameters of WNE stars: dependence on metallicity"*, 2007, Baltic Astronomy, 16, 227

VII Nugis T., Annuk K., Hirv A., Niedzielski A. and Czart K., *"Near infrared spectra of Galactic Wolf-Rayet stars"*, 2008, Baltic Astronomy, 17, 39

Author's contribution to the publications I–III

The author of the thesis has

- improved and tested the time delay estimation algorithms;
- composed computer code to perform the computations;
- carried out major part of computations;
- participated in writing the papers;
- communicated with the editors and referees.

Author's contribution to the publications IV–VII

The dissertant has contributed to the observations of Galactic Wolf-Rayet stars and V838 Monocerotis at the Tartu Observatory. In publication **V** the dissertant also carried out major part of data analysis and prepared the paper for publication. In publication **VII** the dissertant corrected the near infrared spectra against the telluric absorption.

Introduction

Already in 1704 Isaac Newton proposed that gravitational field of massive bodies could possibly bend light rays and the deflection could be strongest at the least distance. In 1804 Johann Georg von Soldner (see Soldner 1804) calculated using Newtonian mechanics the deflection angle for a light ray passing near the limb of the Sun. However, nobody took these results seriously as the wave description of light was widely acknowledged in 18th and 19th century.

In 1911, before completing his theory of general relativity, Albert Einstein investigated the influence of gravity on light (see Einstein 1911) and got a value for the deflection angle for the Sun that was very close to the result derived by Soldner. In his unpublished notes in 1912 Einstein (see Renn et al. 1997) has analysed the possibility of double image due to the gravitational deflection of light and the magnification of these images. In 1915, with the help of his theory of general relativity, Einstein was able to predict the angle of light deflection near the Sun correctly.

According to Einstein's theory of general relativity mass warps space-time. The light rays, propagating along the geodesics of curved space-time are bended as well. If there is a massive object – *gravitational lens* – near the line of sight between a distant source and an observer, the *gravitational lensing*, *i.e.* gravitational deforming of light rays, will result. Arthur Eddington and his collaborators (see Dyson et al. 1920) registered the gravitational displacement of the background star images near the limb of the Sun during the solar eclipse in 1919. They confirmed that gravity bends light rays, and the deflection angle predicted by Einstein is right within their observational accuracy.

Eddington (1920) was probably the first to publish the discussion on possible formation of multiple images of a background star by the gravitational lensing effect of a foreground one (see Refsdal & Surdej 1994; Renn et al. 1997). In 1924 Orest Chwolson (see Chwolson 1924) predicted again the possible forming of multiple images of a background star, when a foreground star acts as a lens. Furthermore, he suggested that, if two stars at different distances and an observer are perfectly aligned, the observer would see a ring-shaped image of the more distant star around the closer one. In 1936 Albert Einstein published a more detailed and famous article about this topic. As Einstein (1936) also considered only stars for the lens, he claimed that the gravitational lensing effects can not be observed at cosmological distances, as the angular separation of the lensed images of a background star would be too small (in order of milli-arcseconds).

In 1937 Fritz Zwicky (Zwicky 1937a,b) predicted that these effects can be detected in cosmological scales if we have a galaxy for the lens. He also predicted that distant galaxies can be used as telescopes to observe otherwise too faint objects behind them. However, the observations must be carried out with high accuracy, as the effect is still small. If a galaxy acts as a lens, the separation between lensed images is only a few arcseconds (see Refsdal 1964a; Refsdal & Surdej 1994).

As lensed images have different *flight paths*, the corresponding *flight times* are in general not equal (due to different path lengths and Shapiro effect). If the source signal is variable, we may first identify a feature in one light curve and after some *time delay* find it in the other. Why are the time delays important? The cosmological aspects of the time delay gravitational lensing were first discussed by Sjur Refsdal (see Refsdal 1964b,a). Using the red-shifts of the source and the lens, angular separation and intensities of the lensed images, the estimated or assumed angular mass distribution of the lens, and the measured time delay, one can fix the Hubble parameter H_0 independently of the "cosmic distance ladder". (For larger red-shifts we must also include the values of cosmological curvature and density parameters into the computation.)

Refsdal considered supernovae for the sources, as the gravitationally lensed quasars were not yet discovered (although, the first optical quasar 3C 273 was identified by Maarten Schmidt 1963, it took more than a decade to discover a lensed one). In 1979 the first extra-galactic gravitationally lensed object, the "double" quasar QSO 0957+561 was discovered by Dennis Walsh et al. (1979). By analysing the spectra and light curves of the "double" quasar, it has been established that the two images A and B are actually gravitational mirages of a single quasar. Today it is known that a massive foreground galaxy produces the two images.

This was a brief review of the history of the gravitational lensing. For more detailed treatments, the reader may have a look at Coles (2001); Refsdal & Surdej (1994); Schneider et al. (2006); Wambsganss (1998).

Although there are about 200 known galaxy scale lens systems (see Treu 2010), the number of lensed quasars which have the time delays estimated and can be used for determining H_0, is still close to $10\ldots20$. One problem is that the time delay estimation needs a much longer time series than is the delay value itself. Typical time delays for galaxy scale lens systems are in the range of weeks to months (with tails extending from hours to years). On the other hand, all the detected lenses are not suitable for establishing the Hubble parameter, as their observables may fall too close to the detection limits. As every observed lens system has its own problems with observational accuracy and established mass distribution of the lens, there are much higher number of individual time delay measurements needed to settle the H_0. We can overcome the uncertainty of H_0 arising from the degeneracies in building the mass distribution models, if we use a large number of different lens systems with measured time delays. As described by Coe & Moustakas (2009); Oguri & Marshall (2010), it is then possible to suppress the errors due to incorrect lens models. A recent step in this direction was taken by Paraficz & Hjorth (2010) who use 18 lenses with measured time delays and non-parametric modelling of the mass profiles to estimate H_0.

Although we will introduce the cosmological applications of the time delay gravitational lensing in the context of fixing the H_0, there are important parallel efforts going on. Knowing the H_0, we can use time delay gravitational lenses for measuring masses of galaxies and estimating the distribution of galaxy masses in the Universe. For an example of deriving the rotation curve of the lensing galaxy from the time delay measurements see Kochanek et al. (2006). Moreover we can estimate the time delays between (somehow selected) nearby images in massive photometry programs (for an example, see LSST Science Collaborations et al. 2009) and detect new lens systems this way. We can also shift the light curves of a given source by the estimated time delay to get a combined time series with a better sampling, which in turn, can be used for studying the source itself. Just to mention, correct time delay measurements are also needed in other astrophysical applications where the reverberation of variable radiation is registered and analysed (for example the reverberation mapping of the active galactic nuclei).

As observational time series are noisy and often gapped, we can not print two nice light curves on transparencies and find the time delay just by looking at the data. Although, there have been much work done already, present time delay estimation algorithms often work correctly only with the time series they have tested to work. Some methods can not deal with gapped data, other can not take the microlensing into account. Every method has at least some parameters, that are set subjectively by user.

From the introduction just presented, we formulate the following goals for the thesis.

- For a large number of lenses to be used in either fixing H_0 or measuring the masses of galaxies in the near future, there is need for an automatic time delay estimation method that does not involve the problems mentioned above. The same type of algorithm is also needed for discovering new gravitational lenses.

- As the approach proposed by Press et al. (1992a) seems to be very close to an automatic method, we are going to point out, why it may give wrong results, especially, with the Vanderriest et al. (1989) data.

- As longer time series can be often obtained with smaller telescopes in not very good seeing conditions, there is also need for time delay estimation methods from blended light curves. It may be also useful to apply algorithms for blended data to reveal possible unresolved images in the case of massive photometry (LSST).

As the result of the study presented in the current thesis[1], an automatic time delay estimation method was developed and evaluated that gives correct results with different sets of observational data with as few as possible "user set parameters" (see Hirv et al. 2011). The approach, to be proposed, connects the good ideas from previous well known methods avoiding their weaknesses. Two special algorithms for blended images were built and tested with artificial data (see Hirv et al. 2007a,b). Many problems were removed from methodology that could lead to wrong time delay estimations. As there was a large amount of time used for testing and improving the algorithms, the final results reflect only a part of the actual work that has been done.

The thesis is organised as follows. In Chapter 1 we present a brief overview of the theory of strong gravitational lensing. We show how multiple images and time delays between their light curves are formed. We describe the problem of estimating H_0 as one possible application of the time delay gravitational lensing. We provide a short discussion of the popular time delay estimation methods, and formulate the motivation of the present study by pointing out their major drawbacks. In Chapter 2 we

[1] The relevant computer programs can be requested from the author of the thesis.

give a more detailed treatment of the problems involved in the time delay estimation process. We form our combined algorithm for measuring time delays from light curves of resolved images by complementing the selected old good ideas with our improvements. We also present the two methods for blended data. We discuss the possibilities of testing the methods and estimating the precision of the measured time delays. In Chapter 3 we present the results of applying developed methods, as well as the explanation, why we may get wrong results with the original Press et al. (1992a) approach. We will also point out how to establish requirements for sufficient observational quality of the data to be processed. In Chapter 4 we summarise the results and introduce the problems left to be analysed in further studies.

Chapter 1

Overview

1.1 Three types of gravitational lensing

Gravitational lensing phenomena are classified into three types based on the strength of the effect. *Strong lensing* involves large image distortions like Einstein ring, arcs, multiple images. The (total) intensity of lensed images may be many times greater than the intensity of the source would be without lensing, as shown by Refsdal (1964b). As such, strong lenses may be used to investigate very faint sources that would be invisible otherwise. Strong lensing is the only type, where we can measure time delays between different images. *Weak lensing* (see Schneider et al. 2006) introduces only minor image distortions that can be detected statistically using a large number of observations of different background objects. We may study the foreground mass (and dark matter) distribution by analysing slight changes of the shape of many background objects. *Microlensing* effect (see Schneider et al. 2006) alters only the brightness of the source as the deflection angle is here too small to be resolved. Microlensing in lensing galaxies results from the fact, that the mass of the galaxies is not distributed smoothly, but some fraction of it is in stars. The moving stars in the lensing galaxy produce many microimages of the distant source. These microimages can not be resolved, as the typical deflection angles are a few micro-arcseconds. However, as the stars move and the lens, source and observer move as well, the net effect of many microlenses changes in time. Hence, microlensing adds additional variability to the macroimages produced by strong lensing. The microlensing effect should produce variability that is uncorrelated between different macroimages. The characteristic time-scales of microlensing are of order a decade or less.

1.2 Theory of strong gravitational lensing and an application to cosmology

As time delays can be measured between light curves of separated images caused by strong lensing, we present here a brief introduction into the relevant theory (for more detailed treatment see Refsdal & Surdej 1994; Courbin & Minniti 2002; Schneider et al. 2006). Although there are developments for strong field gravitational lensing and large deflection angles that describe strong lensing in close vicinity of black holes (see for instance Virbhadra & Ellis 2000; Virbhadra & Keeton 2008; Bozza 2010), the chances of observing involved *relativistic images* and, moreover, obtaining good enough light curves for quite a short time delay measurements, are poor at present time. Just to mention, the angular separation of relativistic images for the Galactic supermassive black hole should be a few microarcseconds and time delays related to not relativistic images should be in the order of minutes. We work always in the weak field limit and, consequently, with small deflection angles when time delays between distant quasar images are measured.

1.2.1 The deflection angle

The deflection angle in the weak field approximation of Einstein's theory of general relativity for a light ray passing near a spherically symmetric mass M at a distance ξ is

$$\hat{\alpha} = \frac{4GM}{c^2 \xi}, \tag{1.1}$$

where G is the constant of gravitation and c is the velocity of light. For the weak field approximation to hold, ξ must be much larger than the Schwarzschild radius of the mass, $\xi \gg R_{sc} \equiv 2GMc^{-2}$. According to this condition, the deflection angle must be small, $\hat{\alpha} \ll 1$.

For a not spherically symmetric mass distribution and weak field we may calculate the deflection angle as a vector sum of elementary deflections caused by an ensemble of mass points. Let a light ray move along a spatial trajectory $(\xi_1(\lambda), \xi_2(\lambda), r_3(\lambda))$ and pass a mass distribution with volume density $\rho(\boldsymbol{r})$. Let us choose the coordinates so that the light ray propagates along the base vector $\hat{\boldsymbol{r}}_{\mathbf{3}}$ before reaching the mass distribution. If the deflection angle is small and the lens has small size compared to the distances involved, we may use the *geometrically thin lens* approximation where the path of a light ray can be approximated as a straight line in the neighbourhood of the lens. If this condition is satisfied, and keeping in mind the chosen coordinate system, the impact vector of the light ray relative to an arbitrary optical axis drawn parallel to $\hat{\boldsymbol{r}}_{\mathbf{3}}$, $\boldsymbol{\xi} = (\xi_1, \xi_2)$, is independent of r_3 and affine parameter λ throughout all the lens. The impact vector relative to a mass element at $\boldsymbol{r'} = (\xi'_1, \xi'_2, r'_3)$, $\boldsymbol{\xi} - \boldsymbol{\xi'}$ is also independent of r'_3 as $\boldsymbol{\xi}$ is independent of r_3.

As the impact vectors relative to all mass elements of the lens are independent of r_3 in geometrically thin lens approximation, we may project the mass density of the lens onto a plane perpendicular to the incoming light ray and define the *surface mass density*

$$\Sigma(\boldsymbol{\xi}) \equiv \int \mathrm{d}r_3 \rho(\xi_1, \xi_2, r_3). \tag{1.2}$$

Finally we may write the deflection angle for a not spherically symmetric mass distribution as a double integral

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \Sigma(\boldsymbol{\xi'}) \frac{\boldsymbol{\xi} - \boldsymbol{\xi'}}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2}, \tag{1.3}$$

where $\mathrm{d}^2\xi' \equiv \mathrm{d}\xi'_1 \mathrm{d}\xi'_2$, $\mathrm{d}^2\xi'\Sigma(\boldsymbol{\xi'})$ is the mass element at $\boldsymbol{\xi'}$ and $\frac{\boldsymbol{\xi}-\boldsymbol{\xi'}}{|\boldsymbol{\xi}-\boldsymbol{\xi'}|}$ gives the direction of the elementary deflection caused by the mass element. $\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is a two-dimensional vector.

The geometrically thin lens model is valid in most astrophysical situations including quasar lensing by galaxies and clusters of galaxies. The geometrically thin lens approximation can not be applied if the lensing mass is distributed along all the way between the source and observer, which may be the case in weak lensing.

1.2.2 The lens equation

There is a typical gravitational lensing situation depicted in Fig. 1.1. An observer at point O sees a lensed image I of a distant source S. The lensing mass distribution is located at distance D_d from the observer. The source lies at distance D_s from the observer and at distance D_{ds} from the lens. The actual light ray curves smoothly due to the lensing mass, but as the D_d and D_{ds} are much larger than the dimensions of the lens in our practical situations, we may approximate the real lensed light ray with a straight line with a kink at the lens position. Moreover, we may well use the weak field and geometrically thin lens approximations. As the deflection angle $\hat{\boldsymbol{\alpha}}$ is very small, the exact definition of the *optical axis* does not matter. We define the optical axis as a straight line from the observer to the position

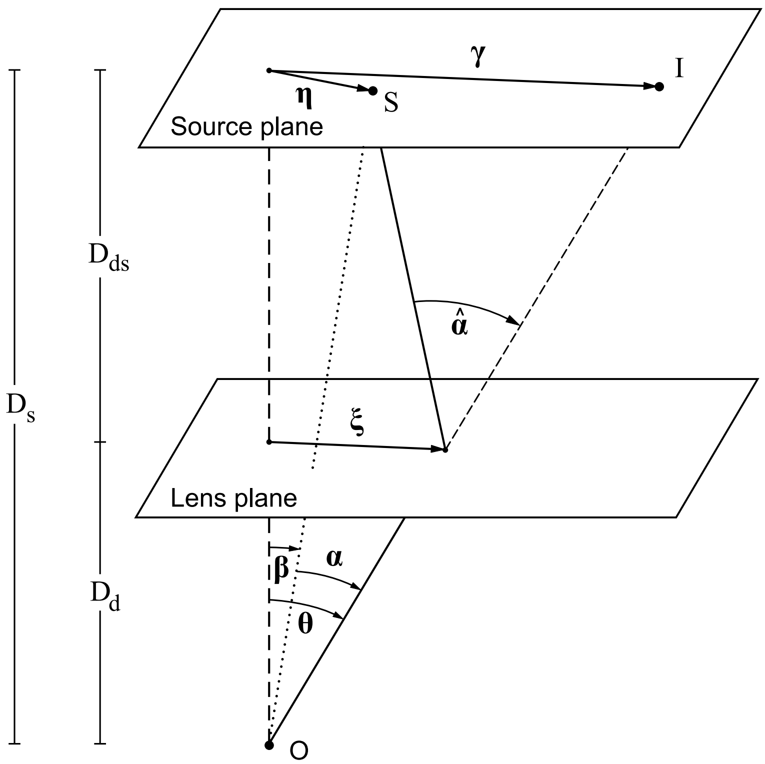

Figure 1.1: The basic geometry of a gravitational lens. (Figure with little modifications from Schneider 1995).

of the lens. The *source* and *lens planes* are perpendicular to the optical axis and are located at the distances of the source and lens respectively. We denote the position vector of the source in the source plane as $\boldsymbol{\eta}$, the position vector of the lensed image in the source plane as $\boldsymbol{\gamma}$ and the impact vector of the light ray in the lens plane as $\boldsymbol{\xi}$. The observer sees lensed image at an angular position $\boldsymbol{\theta}$ but he or she can not, in general case, see the unlensed source, as we can not remove the lens physically. In other words, the true position angle $\boldsymbol{\beta}$ of the source can not be observed directly.

Note, that $\boldsymbol{\eta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ as well as $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are two-dimensional vectors. $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are not parallel in general case. The position and deflection angles in Fig. 1.1 have been highly magnified for clarity. We need not to worry about the fact, that the light ray from the source does not hit the lens plane (where the components of $\boldsymbol{\xi}$ are measured) at right angle. The deviation from right angle is very small as the deflection angle is very small. We may well use Eq. 1.3 to calculate $\hat{\boldsymbol{\alpha}}$. The same considerations apply when we approximate the triangle with $\hat{\boldsymbol{\alpha}}$ as a right triangle.

Keeping in mind that for small angles $\tan x = x$, we can read from Fig. 1.1 that $\boldsymbol{\eta} = D_s\boldsymbol{\beta}$, $\boldsymbol{\gamma} = D_s\boldsymbol{\theta}$ and $\boldsymbol{\gamma} - \boldsymbol{\eta} = D_{ds}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi})$. Now we can write

$$D_s\boldsymbol{\theta} = D_s\boldsymbol{\beta} + D_{ds}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{1.4}$$

Reading from Fig. 1.1 that $\boldsymbol{\xi} = D_d\boldsymbol{\theta}$, we obtain the *lens equation* from Eq. 1.4

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{ds}}{D_s}\hat{\boldsymbol{\alpha}}(D_d\boldsymbol{\theta}), \tag{1.5}$$

which allows us to calculate the generally not observable $\boldsymbol{\beta}$. By defining the *scaled deflection angle* as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{ds}}{D_s}\hat{\boldsymbol{\alpha}}(D_d\boldsymbol{\theta}), \tag{1.6}$$

we can rewrite the lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1.7}$$

Let us define the *critical surface mass density*

$$\Sigma_{cr} = \frac{c^2}{4\pi G}\frac{D_s}{D_d D_{ds}} \tag{1.8}$$

and the *dimensionless surface mass density* or *convergence*

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_d\boldsymbol{\theta})}{\Sigma_{cr}}. \tag{1.9}$$

There may be more than one value of $\boldsymbol{\theta}$ that satisfies the lens equation for a given $\boldsymbol{\beta}$, i.e., the lens may produce multiple images of a single source. It can be shown that a mass distribution which has $\kappa \geq 1$ somewhere, produces multiple images for some source positions. Hence, strong lenses have $\Sigma \geq \Sigma_{cr}$.

Using Eqs. 1.3, 1.6, 1.8, 1.9 and keeping in mind that $\boldsymbol{\xi} = D_d\boldsymbol{\theta}$ and $\mathrm{d}^2\xi' = D_d^2\mathrm{d}^2\theta'$, we can write the scaled deflection angle in terms of the observable angular position $\boldsymbol{\theta}$

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi}\int \mathrm{d}^2\theta'\kappa(\boldsymbol{\theta'})\frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2}. \tag{1.10}$$

We used relations between angles and distances from Euclidean geometry to derive the lens equation and formulae for the scaled deflection angle. The derived equations hold also in an expanding universe if we interpret the distances as angular diameter distances.

1.2.3 Some cosmological notes

We will represent here some cosmological notations that are needed to understand the time delay and Hubble constant relation in gravitational lensing. For more detailed treatment see Schneider et al. (2006) and references therein.

The light, radiated from a comoving source at time t_2 and observed at time $t_1 > t_2$ by a comoving observer, is *redshifted* by a factor z_{12}. The redshift and the *cosmic scale factor* $a(t)$ are related as

$$1 + z_{12} = a(t_1)/a(t_2). \tag{1.11}$$

$a(t)$ is interpreted as the expansion history of the Universe and is normalised so that today, $a(t_0) = 1$.

The *expansion rate* $H(t)$ of the Universe is defined as

$$H(t) = \dot{a}a^{-1}, \tag{1.12}$$

and its current value H_0 is called *Hubble constant.*

If we set the cosmological constant $\Lambda = 0$ and curvature parameter $K = 0$, then the current density of the Universe can be expressed as

$$\rho_{cr} = \frac{3H_0^2}{8\pi G}. \tag{1.13}$$

The *critical density* ρ_{cr} is used to define the *density parameters*

$$\Omega_m = \frac{\rho_{m0}}{\rho_{cr}}; \qquad \Omega_r = \frac{\rho_{r0}}{\rho_{cr}}; \qquad \Omega_\Lambda = \frac{\rho_v}{\rho_{cr}}, \tag{1.14}$$

where ρ_{m0}, ρ_{r0} and ρ_v are the present densities of matter, radiation and vacuum respectively. (The vacuum density is assumed to be constant in time.)

Let $\omega(z_1)$ and $\omega(z_2)$ be the comoving radial coordinates of two sources at redshifts $z_1 < z_2$. The *comoving distance* $\omega(z_1, z_2) = \omega(z_2) - \omega(z_1)$ between these sources can be expressed as

$$\omega(z_1, z_2) = \frac{c}{H_0} \int_{a(z_2)}^{a(z_1)} [a\Omega_m + a^2(1 - \Omega_m - \Omega_\Lambda) + a^4\Omega_\Lambda]^{-1/2} \mathrm{d}a. \tag{1.15}$$

The *angular diameter distance* $D^{ang}(z_1, z_2)$ of a source at redshift z_2 seen by an observer at redshift $z_1 < z_2$ can be written as

$$D^{ang}(z_1, z_2) = a(z_2) f_K[\omega(z_1, z_2)], \tag{1.16}$$

where the *comoving angular diameter distance* $f_K(\omega)$ depends on the curvature parameter K as

$$f_K(\omega) = |K|^{-1/2} \sinh[|K|^{1/2}\omega]. \tag{1.17}$$

Note that $f_K(\omega) = \omega$ if $K = 0$ (the flat Universe).

1.2.4 The time delay and Hubble constant

Following Cooke & Kantowski (1975) the *time delay* between alternative images in a gravitational lens system must be sum of two components. First, as the path lengths of light rays of individual images are generally not equal, the propagation times of light must be uneven. Second, the propagation times differ also due to the Shapiro effect that retards light in gravitational field of the lens.

By defining the *deflection potential*

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}^2\theta' \kappa(\boldsymbol{\theta'}) \ln |\boldsymbol{\theta} - \boldsymbol{\theta'}|, \tag{1.18}$$

and the *Fermat potential*

$$\tau(\boldsymbol{\theta}; \boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}), \tag{1.19}$$

the time delay between two lensed images at positions $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ can be written as (see Schneider 1985; Schneider et al. 2006)

$$\Delta t = \frac{D_d^{ang} D_s^{ang}}{c D_{ds}^{ang}} (1 + z_d)[\tau(\boldsymbol{\theta}^{(1)}; \boldsymbol{\beta}) - \tau(\boldsymbol{\theta}^{(2)}; \boldsymbol{\beta})], \tag{1.20}$$

where z_d is the redshift of the lens and the distances are measured as angular diameter distances. The source position $\boldsymbol{\beta}$ enters into the Fermat potential as parameter. The deflection potential $\psi(\boldsymbol{\theta})$ describes the potential time delay and the geometric time delay is described by the $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2$ term.

Note, that there is an alternative wavefront method for deriving the expression of the time delay introduced by Refsdal (1964b,a, 1966) and Kayser & Refsdal (1983).

Now we have all the necessary relations to describe how the Hubble constant is connected to the time delay between the two gravitationally lensed images of a distant source. H_0 enters into Eq. 1.20 through the angular diameter distances. We can see from Eqs. 1.15, 1.17 and 1.16 that $D^{ang}(z_1, z_2) \sim cH_0^{-1}$. As the redshift and Fermat potential are both dimensionless, we can read from Eq. 1.20 that $\Delta t \sim H_0^{-1}$. In order to express the angular diameter distances in the units of cH_0^{-1} we need to take the following two steps. First, to calculate the values of the cosmic scale factor (Eq. 1.11), we have to estimate the redshifts of the source and the lens. Second, as $D^{ang}(z_1, z_2)$ depends also on the curvature parameter and density parameters, we need to fix these as well. To derive the values of the Fermat potential (Eq. 1.19), we have to measure the positional angles of the images and determine the positional angle of the source using the lens equation (Eq. 1.7) with Eqs. 1.10, 1.9 and 1.8. Additionally we need to fix the model of the surface mass density of the lens to compute the values of the scaled deflection angle (Eq. 1.10) and the deflection potential (Eq. 1.18). Finally, to estimate H_0 from Eq. 1.20 we need to measure the time delay, which is the topic of the present thesis.

1.2.5 Images and magnification

The gravitational lensing effect does not depend on wavelength of light. When we talk about light, the same applies for any kind of electromagnetic radiation. The gravitational lensing conserves the surface brightness of the source. We will not prove this here, but if it is true, we can expect that a sheet-like source with constant surface brightness is gravitationally lensed into a sheet-like image with constant surface brightness. This is exactly what happens with the cosmic microwave background radiation – we observe smoothly distributed radiation and no high amplifications due to the large scale structure (Courbin & Minniti 2002).

The relative positioning of the source, the lens and the observer, and also the surface mass distribution of the lens, as well as the shape of the source determine what the observer will see. For example, if a spatially unresolved source, an intervening axisymmetric massive object and an observer lie on a straight line, the observer will see an *Einstein ring*. If the lens is not so well alined, the observer will see arcs of light around the lensing object (see Refsdal & Surdej 1994). It is also possible to see the Einstein ring for a not perfectly axisymmetric lens due to proper combination of the mass distribution of the lens and the extent of the physical source (see Schneider et al. 2006).

The Fermat's principle says that real light rays take paths that correspond to the stationary points of the arrival-time surface. *I.e.* the images can be found in either minima, maxima or saddle points of the arrival-time surface. As has been shown by Schneider (1985), the Fermat potential $\tau(\boldsymbol{\theta}; \boldsymbol{\beta})$ is, up to an affine transformation, the light travel time in a gravitational lensing situation. Hence,

$$\nabla\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = \mathbf{0} \tag{1.21}$$

is the condition not only for the stationary points of the Fermat potential, but also for the arrival-time surface and gives us the location of the images. Eq. 1.21 is equivalent to the lens equation.

If a source is much smaller than the angular scale on which the lens properties change, we can express the inverse of *magnification* as a Jacobian matrix

$$M^{-1}(\boldsymbol{\theta}) = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}} = \nabla\nabla\tau(\boldsymbol{\theta};\boldsymbol{\beta}), \tag{1.22}$$

which tells us how much source-plane displacement is needed to produce a given small image displacement. Eq. 1.22 means that the curvature of the arrival-time surface is the inverse of magnification. Broad hills in the arrival-time surface correspond to highly magnified images while sharp peaks correspond to images demagnified into unobservability. The magnification M is a 2D tensor. The determinant $\det M$ defines a scalar magnification, or ratio of image area to source area for a 'small' source.

As the gravitational lensing may change the apparent shape of lensed objects and conserves the surface brightness, it may give significant flux amplification. (This can also be applied to the unresolved microlensing case, if we analyse every microlensing act separately and assume that physical images are unresolved only due to our poor telescopes.) If $\mathrm{d}\omega_i$ and $\mathrm{d}\omega_s$ are the solid angles covered by a lensed image and an unlensed source with constant surface brightness (which is the practical case in quasar lensing), the ratio

$$|\mu| = \frac{\mathrm{d}\omega_i}{\mathrm{d}\omega_s} \tag{1.23}$$

gives the *flux amplification* due to lensing. For a 'small' source $\mu = \det M$. μ can be either positive or negative, the observed fluxes of images are determined by the absolute value $|\mu|$. If the lens produces several images of a single source, the total flux amplification is given by the sum of all individual image amplifications.

In any lens there can be closed, smooth curves, on which the determinant $\det M^{-1}(\boldsymbol{\theta}) = 0$. These curves are called *critical curves*. The corresponding curves in the source plane, which are obtained from the critical curves using the lens equation are known as *caustics*. A gravitational lens may produce very high flux amplification, if the source lies on or close to a caustic ($\mu = \det M$ is formally infinite in this case). If we move a ('small') source across the caustic, a pair of images near the corresponding critical curve is either created or destroyed, depending on the direction of the movement.

1.2.6 Model degeneracies and making use of future optical imaging surveys

It occurs that the observed image positions, flux ratios and time delay ratios can be reproduced with different mass models of the lens. If $\kappa(\boldsymbol{\theta})$ is a mass distribution that provides a good fit to these observables, there is the whole family of mass models

$$\kappa_\lambda(\boldsymbol{\theta}) = (1 - \lambda) + \lambda\kappa(\boldsymbol{\theta}), \tag{1.24}$$

that provides an equally good fit to the data. λ may have all values that do not turn the surface mass density negative. As the first term corresponds to adding a homogeneous mass-sheet with constant surface mass density $\kappa_c = 1 - \lambda$ to the mass distribution, the problem is known as *mass-sheet degeneracy*. The second term corresponds to scaling of the 'original' surface mass density due to the mass-sheet. Adding a constant surface density mass-sheet corresponds to inserting a group or a cluster of galaxies into the model of the lens. As lensing galaxies belong always into some group, determining the group contribution is critical but difficult part of our modelling. If we insert a mass-sheet with the surface mass density κ_c into our 'original' model which gave time delay Δt, the resulting time delay will transform into $\lambda\Delta t$, while the observed image positions, flux ratios and time delay ratios remain unchanged. As determining the real κ_c with high enough precision is rather difficult, we may get wrong H_0 estimation from the given time delay lens. Neglecting the extra surface mass density coming from nearby objects leads to an overestimate of the Hubble constant. For most cases this error is probably $\lesssim 10\%$.

In addition, for many lens systems there is a degeneracy (within observational errors) in choosing the 'original' surface mass density model even without considering the possible mass-sheet.

It seems that there is little hope for finding a 'golden' lens for a precise H_0 estimation. In some cases the error of time delay measurement due to little quasar variability dominates, in other cases we have major problems with the model of the lens or even with the astrometry of the quasar images. On the other hand, according to Coe & Moustakas (2009); Oguri & Marshall (2010) we can hope to reduce the uncertainty of H_0 down to $\sim 1\%$ (1σ precision) if thousands of strong gravitational lens systems detected by the

Large Synoptic Survey Telescope (LSST) are to be analysed. When using large number of gravitational lens systems with measured time delays from future optical imaging surveys, we do not need to know the exact surface mass density profile of a lens. An average profile is used instead. The mass-sheet contributions to a single lens system also need not to be measured as we can make use of the knowledge about the distribution of the mass-sheet surface mass densities. In a recent study Paraficz & Hjorth (2010) take a step in this direction and use 18 lenses with measured time delays and non-parametric modelling of the mass profiles to estimate H_0.

The time delay estimation procedure for a large number of lenses is easier and more reliable, if we have an automatic scheme for that. Therefore, moving towards an automatic time delay estimation is the aim of the present thesis.

1.3 Time delays and basic problems

The light from the source quasar is deflected in a gravitational lens system into multiple images $f_1, \ldots, f_R$ due to the gravitational field of the intervening galaxy. The source variability $g(t)$ shows itself in the measured light curves. Because of different flight paths (and also different lags due to the Shapiro effect) the total flight times $\Phi_r, r = 1, \ldots, R$ differ. Consequently, we can only measure replicas of the source curve with different delays:

$$f_r(t) = F(g(t - \Phi_r)), r = 1, 2, \ldots, R. \tag{1.25}$$

Additional distortions (physical and instrumental) are depicted here using the function F (the exact form of it depends on particular experiment). For a fully resolved case we will have in total R continuous curves $f_r(t)$. What observer measures are values of $f_r(t)$ or their combinations (blends) at certain moments of time $t_i^*, i = 1, 2, \ldots, N$. As an example of a real time series to work with, we represent in Fig. 3.2 the 131 point and 2926 day long light curve of the QSO 0957+561 provided by Vanderriest et al. (1989).

We can measure the differential time delay $\Delta t_{o,p} = \Phi_p - \Phi_o$ between each pair of images f_o, f_p. The time delay $\Delta t_{o,p}$ is positive if the variability of the image o is preceding the variability of the p image. From $R(R-1)/2$ delays only $R - 1$ can be considered as independent.

Depending on a number of measured channels and their content we can now have different schemes to work with:

- Unresolved measurements. Sometimes we have only summed up combination of different channels. It occurs that when certain conditions are met, it is still possible to unscramble the single input curve and estimate separate time delays involved, see Geiger & Schneider (1996) for details.

- Two separate fully resolved channels. This is the most often treated case.

- Multiple ($R > 2$) of resolved channels. For a particular method see Pelt et al. (1998a).

- At least one unblended channel and some blends. This particular case is elaborated in Hirv et al. (2007a).

- Only blends. Very special cases can be strictly analysed, see Hirv et al. (2007b) for an example.

To expose the basic problems involved in the time delay analysis, we will center our attention to the classical scheme with two fully resolved observed curve. Very often the computational approaches for two curves are illustrated by using two graphical images of the input curves depicted on transparencies. By shifting plots against each other along time direction and along magnitudes we can seek for a combination where both curves form together a picture with less scatter. The basic problems of the time delay analysis are revealed already in this demonstration:

- For different time delays the area where curves can be compared is different. Computational methods must take this into account by proper normalisation of statistics involved.

- The final image of juxtaposed curves depends on mutual position of the two curves as well as on the inherent scatter of each single channel. The both aspects must be considered with care. The measurement errors assigned by observers should be taken into account properly.

- We must be extra careful to make difference between the cases when continuous pattern occurs due to the real similarity of the two curves or due to the occurring of data points of the one curve in gaps of the other data set.

- After finding the best juxtaposition it is still possible that the parts of the curves show significant differences. In the context of gravitational lensing it is often assumed that microlensing effects for one or both curves can be involved.

As we will see below there are different approaches to treat the listed problems.

1.4 Time delay estimation methods and our motivation

Here we will briefly discuss different methods for time delay estimation in astrophysical problems. As it will be shown below the severity of problems involved in the algorithms developed so far is strong motivation for working towards a better method, which depends less on user set parameters, can properly handle observational gaps and microlensing effect. The need for methods for blended light curves is emphasised as well.

To unify the treatment below we use the following notation. There are two sets of data points A and B (time, measured value and standard deviation for every point) with N_A points $t_i^*, a_i^*, \sigma_i^*, i = 1, 2, \ldots, N_A$ and N_B points $t_j^*, b_j^*, \sigma_j^*, j = 1, 2, \ldots, N_B$. It is not assumed that $N_A = N_B$. If the observer given estimates for standard deviations are missing then we can set $\sigma_i^* = 1$ and $\sigma_j^* = 1$ (in the sense of relative weights). When comparing two data sets we often need to *adjust* the data to take into account time delay, differences in magnification, baseline levels etc. For adjusted data sets we will use the simplest notation: if both sets are treated separately then we have triples t_i, a_i, σ_i and t_j, b_j, σ_j and if the sets are combined using certain trial time delay Δt^* we use triples $t_l, y_l, \sigma_l,\ l = 1, 2, \ldots, L$. In some cases $L = N_A + N_B$ but not always. The procedure of shifting in time and scaling in amplitude or shifting in baseline of two light curves for estimating the time delays is called *matching*.

In some formulae the notation for statistical weights $W = 1/\sigma^2$ (with proper indexes) is more appropriate.

1.4.1 Resolved images

The basics of time delay estimation between two sampled light curves can be formulated as follows. We shift one light curve by a trial differential time delay and correct it for different magnification (for the data in relative flux units) or baseline shift (for the data in magnitudes). Then we estimate the goodness of match by computing the value of a certain statistic or *merit* *function.* The maximum (or minimum) value of the merit function indicates the best combination of the trial parameters.

There are a number of different time delay estimation methods used by various research groups. For a short review of the popular methods see for instance Kundic et al. (1997). Some recent more peculiar approaches can be found in Hjorth et al. (1992), Pijpers (1997), Barkana (1997), Burud et al. (2001), Gil-Merino et al. (2002), and Cuevas-Tello et al. (2006). These methods can roughly be divided into three classes:

- cross-correlation based methods;
- methods based on interpolation (linear, polynomial, spline, etc.);
- methods which use dispersion spectra.

1.4.1.1 Methods based on cross-correlation

Two continuous curves $a(t)$ and $b(t)$ can be correlated for various delays Δt^* by computing

$$CF(\Delta t^*) = \frac{\mathbf{E}\{[a(t) - \bar{a}][b(t + \Delta t^*) - \bar{b}]\}}{\sigma_a \sigma_b}, \tag{1.26}$$

where application of $\mathbf{E}\{\ldots\}$ denotes taking statistical expectation, $\bar{a}$, $\bar{b}$ are the mean values (estimated or known) and σ_a and σ_b are the corresponding standard deviations. It is hoped that correct delay Δt_{AB} will reveal itself as a strongest or at least major maximum in correlation curve.

There are many ways to approximate notion of correlation function for discrete time series. For instance we can define certain fixed step (say $t_l = l\delta t + t_0, l = 0, 1, \ldots, L - 1$) grid in time, where t_0 is the starting time point of our time series; interpolate every observed point of the A curve to the nearest grid point; shift the B curve in time by Δt^* and interpolate every shifted point of the B curve to the nearest grid point. Finally we select pairs a_u, b_u with same time moments and, using standard definition for the discrete correlation function, we can compute an approximation

$$DCF_s(\Delta t^*) = \frac{1}{U - 1} \frac{\sum\limits_{u=1}^{U} (a_u - \bar{a})(b_u - \bar{b})}{\sigma_a \sigma_b}, \tag{1.27}$$

where U is the number of pairs. There is no need to say that for our sparse data sets the resulting correlation function estimate will probably have gaps and considerable scatter due to the fact that the large part of the pairs to be correlated is missing. Sometimes it is proposed that we can add to our data sets artificial points which are obtained by linear interpolation, see for instance Gaskell & Sparke (1986) or Gaskell & Peterson (1987). This approach, even when useful in some contexts, can not be used for a data with significant gaps.

More common and often used is an approach proposed in Edelson & Krolik (1988). First, for each pair of observations they define:

$$F_{ij} = \frac{(a_i - \bar{a})(b_j - \bar{b})}{\sqrt{(\sigma_a^2 - e_{a_i}^2)(\sigma_b^2 - e_{b_j}^2)}}, \tag{1.28}$$

where $e_{a_i}^2$ and $e_{b_j}^2$ are measurement errors for A and B set correspondingly. Then the moving window with width ν is used to compute the discrete correlation function:

$$DCF(\Delta t^*) = \frac{\sum\limits_{i,j} S_{ij} F_{ij}}{\sum\limits_{i,j} S_{ij}}, \tag{1.29}$$

with inclusion condition:

$$S_{ij} = \begin{cases} 1, & \text{when } |t_i - t_j - \Delta t^*| \leq \nu/2, \\ 0, & \text{otherwise.} \end{cases} \tag{1.30}$$

Depending on circumstances we can compute $DCF(\Delta t^*)$ for overlapping windows or just for a row of nonoverlapping but fully covering set of windows. The free parameter of the procedure - width of the window ν - is chosen as a compromise value to get enough resolution when trading it against statistical stability. Without certain objective method to fix it we can not consider DCF computation as an automatic procedure.

In the original formulation of the $DCF(\Delta t^*)$ the means and dispersions for computing F_{ij} values are global, they are computed for the full data sets A and B. We can also consider a form of the correlation function where these values are computed separately - for the each bin (see Lehar et al. 1992; Gil-Merino et al. 2002). This allows us to take into account possible nonstationarity of the underlying processes. However, the problem with freely chosen bin size remains.

All DCF based methods do not take explicitly into account the possibility of microlensing.

1.4.1.2 Dispersion spectra

The simplest dispersion spectrum can be computed from two time series quite similarly to DCF. We just define differences:

$$D_{ij} = \frac{(a_i - b_j)^2}{2}, \tag{1.31}$$

and form Δt^* dependent function

$$DS(\Delta t^*) = \frac{\sum\limits_{i,j} S_{ij} D_{ij}}{\sum\limits_{i,j} S_{ij}}, \tag{1.32}$$

where the inclusion condition S_{ij} can be computed as in Eq. 1.30. The full range of different implementations of this simple scheme is presented in Pelt et al. (1996). For instance we can use one of the input series in adjusted form $b_j = mb_j^* + h$ where m and h are unknown magnification and baseline shift correspondingly. The differences will now depend on unknown parameters:

$$D_{ij}(m,h) = \frac{(a_i - b_j)^2}{2}, \tag{1.33}$$

and so do also the dispersion spectra $DS(\Delta t^*, m, h)$. Resulting spectrum can be computed by performing minimisation:

$$DS(\Delta t^*) = \min_{m,h} DS(\Delta t^*, m, h). \tag{1.34}$$

If our data is in magnitudes, we include the baseline shift h as a free parameter in the computation of the dispersion spectrum and fix magnification $m \equiv 1$; otherwise we use magnification m as a free parameter and fix baseline shift $h \equiv 0$. Methods using dispersion spectra allow also easily to take into account unequal quality of different observations (by introducing weights into squared differences), and microlensing.

The averaging bin size ν is still a free parameter for DS-type methods and consequently we are not too much better off if to compare with DCF-style methods.

1.4.1.3 Interpolation based methods

To fill gaps in observed data series we can use different interpolation or approximation schemes. In principle it is possible to fit certain model curves (polynomials, splines etc) to the both curves and then compare continuous (or regularly sampled) model curves. However, most often the researchers use methods where a model for source curve is built and both observed sequences are then fitted (with proper delay or delays) to the model curve (see for instance Lehar et al. 1992; Barkana 1997; Burud et al. 2001; Cuevas-Tello et al. 2006). In these methods statistical weights can be easily used.

The resolution – statistical stability trade-off for interpolation methods is achieved by proper (but basically, not automatic) choice of the model form (polynomial degree, number of nodes for splines etc). The microlensing effects or other low-frequency disturbations can not be detected easily, because they will be hidden in the common model curve for both light curves. Consequently, in some contexts it would be useful to interpolate (or approximate) input curves separately. Then the misfit between them will indicate possible distortions.

Most important deficiency of the fitting to the common model curve is of course tendency to obtain results with good characteristics of the fit but still indicating wrong delays. This happens if the data contains nearly periodic gaps. It can fairly well happen (and it indeed does so often) that with certain delay the observed points of the A curve fit into time gaps of the B curve.

1.4.1.4 Using optimal prediction

Somewhat apart from other methods stands a method developed by Press et al. (1992a, below PRH), refined by Rybicki & Kleyna (1994) and lucidly presented by Haarsma et al. (1997). In principle, this method is nearest to that which can be called automatic, but it still has some drawbacks which we are going to analyse and resolve in Section 2.1.1. Because our discussion below heavily uses notions and ideas promoted in these papers we will describe here the original method in some detail.

We start from a model of the observed data

$$y(t) = s(t) + n(t), \tag{1.35}$$

where $s(t)$ is original (source) signal and $n(t)$ is observational noise. The observed sample vector is then

$$y_i \equiv y(t_i), \quad i = 1, 2, \ldots, N. \tag{1.36}$$

Here y_i denotes any time series, not explicitly combined ones. Our goal for a moment is to "predict" signal value for a particular time point so that predicted value is as close as possible to the real one or so that prediction error:

$$\mathbf{E}\{e^2(t)\} \equiv \mathbf{E}\{[\hat{s}(t) - s(t)]^2\} \tag{1.37}$$

is minimised. Estimate $\hat{s}(t)$ that is linear in the data points y_i:

$$\hat{s}(t) = \sum_{i=1}^{N} y_i q_i(t), \tag{1.38}$$

can be formed using N "trial" functions $q_i(t)$. By substituting estimate $\hat{s}(t)$ into expression to be minimised we can get formulae for "trial" functions and predicted value.

Using notations

$$\begin{aligned} &C_{ij} \equiv \mathbf{E}\{s(t_i)s(t_j)\} \\ &c_i(t) \equiv \mathbf{E}\{s(t_i)s(t)\} \\ &c_c(t) \equiv \mathbf{E}\{s(t)s(t)\} \\ &n_i^2 \equiv \mathbf{E}\{n(t_i)n(t_i)\} \\ &B_{ij} = C_{ij} + n_i^2\delta_{ij} \end{aligned} \tag{1.39}$$

(where $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ elsewhere) we introduce covariance matrix $\mathbf{B}$ with elements B_{ij} and two vectors $\mathbf{q}$ and $\mathbf{c}$ with elements $q_i(t)$ and $c_i(t)$.

It occurs (see PRH), that using these notations the "trial" functions in the form

$$\mathbf{q}(t) = \mathbf{B}^{-1}\mathbf{c}(t), \tag{1.40}$$

can be used to get actual "optimal" predictions

$$\hat{s}(t) = \mathbf{y}^T\mathbf{B}^{-1}\mathbf{c}(t), \tag{1.41}$$

with corresponding expected estimation errors

$$\mathbf{E}\{e^2(t)\} = c_c(t) - \mathbf{c}^T(t)\mathbf{B}^{-1}\mathbf{c}(t). \tag{1.42}$$

For practical computations PRH assume the stationarity of the source process. We will show in Section 2.1.1.5 how the values of C_{ij}, $c_i(t)$ (and $c_c(t)$) are estimated. From this point on, to use the optimal prediction for time delay measurements, one can proceed through three different paths.

In PRH authors show that optimal reconstruction of $s(t)$ is equivalent to a reconstruction that minimises the value of χ^2:

$$\chi^2 = \sum_{i,j} y_i A_{ij} y_j, \tag{1.43}$$

where the matrix $\mathbf{A}$ is the matrix inverse of the total covariance matrix $\mathbf{B}$

$$\mathbf{A} = \mathbf{B}^{-1} \equiv \left\{C_{km} + n_k^2 \delta_{km}\right\}^{-1}, \tag{1.44}$$

and y_i-s are adjusted observational data (shifted by time delay and magnitude difference B curve combined with A curve, mean $\bar{y}$ subtracted). Correspondingly they use the following scheme to estimate the time delays:

- Let us assume that the underlying source process $s(t)$ is stationary and its autocorrelation function can be described by simple analytical form (with small number of parameters).
- For a particular trial time delay Δt^* and other free parameters (general mean of the process $\bar{y}$, magnitude difference between two components Δy_{AB}) we can compute χ^2 and use it as a criterion to compare different parameterisations.
- The time delay and parameters which minimise criterion value is then used as a solution.

In actual computations authors use certain optimisations and approximations to obtain final solution. For instance they minimise χ^2 along the free parameter $\bar{y}$ analytically and estimate magnitude differences using "pointwise" fitting of the two curves (we come to that after a while). But these are minor aspects of the adopted procedure.

The second method is important modification for the PRH method suggested in Rybicki & Kleyna (1994) and applied by Haarsma et al. (1997). Instead of formal minimisation of the χ^2 they take off from Gaussian model of the process with probability distribution of the data vector:

$$P(\mathbf{y}) = \left[(2\pi)^N |\mathbf{B}|\right]^{-1/2} e^{-1/2\chi^2}. \tag{1.45}$$

Correspondingly, for correct solution of the time delay estimation problem they propose to minimise *log likelihood* Q

$$Q = \log(|\mathbf{B}|) + \sum_{i,j} y_i A_{ij} y_j, \tag{1.46}$$

where adjusted observations y_i as well as determinant $|\mathbf{B}|$ depend on free parameters.

As the problems hidden in the described above algorithm will be analysed in detail later, we only name here the most important ones. First, the algorithms which are based on the minimisation of the global χ^2 or Q

statistic of the combined curve do not allow to take the microlensing effect into account properly. Second, the problem with correct normalisation of the χ^2 or Q arises if we are unable to eliminate the microlensing effect. Third, using the global χ^2 or Q, it is possible that we fit data of one curve into the gaps of the other and get a wrong result for the time delay.

The third method which uses prediction ideology is somewhat naive use of the predicted process values itself. Let us call this method as a "pointwise minimisation". Let us fix a particular trial time delay Δt^*. For each component we can interpolate values at times corresponding to the observations of the *other* component and compute standard errors for predicted values (restricting the pairing to the overlapping area of the two curves). Using the obtained pairs we can now form a standard χ^2 measure of goodness-of-fit:

$$\chi^2 = \sum_k \frac{(\hat{a}_k - b_k)^2}{\delta_k^2} + \sum_l \frac{(a_l - \hat{b}_l)^2}{\delta_l^2}, \tag{1.47}$$

where δ_k^2 and δ_l^2 are combined variances (observer given variance plus variance of predicted value). And again, the best combination, to be adopted as a solution, is set of parameters which minimises the χ^2. This type of criterion function was used already by PRH, but only in the context of estimating magnitude difference between two light curves. We take Eq. 1.47 as a starting point for the criterion function for our combined time delay estimation method.

1.4.2 Unresolved images

As the time delays may be in order of hundreds of days for real gravitational lens systems, long-lasting monitoring of quasar images is often needed. Longtime use of telescopes, that can resolve all the subimages of a lensed quasar, is rather expensive. Time is money, especially for large telescopes! As we need to use smaller ones at not so good seeing conditions, photometric aperture may cover several images and we will get blended light curves. It occurs that it is possible to measure time delays between different images even if the data is blended. On the other hand, if the data is blended and we do not take this fact into account, we may get incorrect results for the time delay. When trying to find a new lens system by its time delays, we do not know the mass distribution of the lens – consequently, it is worth considering the possible blending as well.

From physical considerations we can predict that the components of the blends are certain weighted sums of the time-shifted source variability curves. Typically, the time delays between blended components are remarkably shorter than the delays between images, which are significantly separated. A singular case of blending, where we observe only the sum of multiple components, was analysed in Geiger & Schneider (1996). In this method a certain amount of extra information is used from additional interferometric observations.

Below we will introduce time delay estimation algorithms for the cases, where one image is composed of two blended subimages and the other may be a single component image (see Hirv et al. 2007a) or a blend of two subimages (see Hirv et al. 2007b) as well.

Chapter 2

Methodology

2.1 Estimating time delays from resolved images

We continue here with a more detailed discussion of the different statistical and computational problems involved in the time delay estimation process. Through the following treatment we will formulate the combined time delay estimation algorithm for two light curves of resolved images (see Hirv et al. 2011).

2.1.1 Problems and solutions in the time delay estimation

2.1.1.1 Normalisation

If we compare curve A with shifted in time by Δt^* curve B then there is only certain interval in time where both curves have observed values. For longer delays this overlap area is shorter and for shorter delays it is longer.

For methods with χ^2 calculation for the combined curve the correct approach is not self-evident. Original authors (as far as we understand from PRH) use for evaluation of the different delays the same number of data points $L = N_A + N_B$. It may fairly well be that for an ideal case, where actual data is a realisation of a stochastic process whose properties match these of hypothesised model, this approach can be considered as correct. But for real data this is certainly not so.

For instance, when input data contains correlated errors due to the microlensing, the scatter for combined curve in overlap area is certainly higher if to compare with scatter in the parts where only one curve is observed. The length of the high scatter area depends on delay and this dependency will show up in $\chi^2(\Delta t^*)$ or $Q(\Delta t^*)$ curves. One possible solution is a computation of criteria for only overlapping subsets and using proper normalisation using degrees of freedom involved. Unfortunately this approach can sometimes fail - due to the long and nearly periodic gaps in the data.

In methods where dispersion spectra or pointwise differences (Eq. 1.47) are calculated it is quite easy to take the microlensing and periodic gaps into account (see the discussion below).

2.1.1.2 Fitting data into gaps

As we saw above, the different time delay estimation schemes can be divided into two classes.

In the first class the input data sets A and B are merged using trial delay, and combined data set is used as it is – without taking into account the origin (A or B) of data points. Different delays are then compared by modelling of the combined data set using certain analytical (polynomials, splines) or statistical (χ^2,Q) models and criteria.

In the second class the original data points enter into estimation scheme only in pairs where one point is from A curve and the other one from B curve.

The principal difference of the two schemes reveals itself in the cases where input data has long and more or less periodic gaps in it (say, due to the skipping of certain seasons, when observations are not feasible). For some trial delays it can now happen that for a particular shift in time the observations from curve A happen to fit into gaps of the B curve. The first class methods can be quite happy with this, the general scatter around continuous model is at the level of the observational errors and only these parts of the combined curve, where data points are mixed, add extra scatter. As a result there is quite high probability to get spurious minima in the criterion curves. The susceptibility of the data to such distortions can be estimated by computing *data windows* either in the form proposed in PRH (Fig. 8) or as a pair count spectra as this is done in Pelt et al. (1994).

The DS and DCF type methods overcome the problem of nearly periodic gaps in an obvious way – we do not fit data onto a common model curve and regions where *one* curve has long gaps are not used in computing the merit function. The scatter due to data points from different light curves is analysed only.

The pointwise minimisation (Eq. 1.47) also measures the difference of the two light curves only, and, in regions where *one* curve has gaps, appropriate weights for interpolated points are assigned. Even the long and continuous stretches of the *one* input curve do not have any effect on final dispersion estimates. We think that this is one of the important properties of the pointwise minimisation.

2.1.1.3 Decorrelation length

The long gaps in input data sets, even if not periodic, are of grave concern from another point of view too – in methods where only data point pairs enter into valuations, the pairs whose time moments differ too much can obtain great influence. However, the probability that a_i point and corresponding b_j point are correlated is quite low if time difference $t_i - t_j$ is long enough. The term *decorrelation* is often used in this context. The inclusion parameter ν introduced in the DCF and DS methods is used to skip all the data point pairs from our statistics if the time difference exceeds this prescribed value.

Using the pointwise minimisation (Eq. 1.47), we do not need either inclusion condition S_{ij}, nor inclusion parameter ν. In the computation of the criterion function we can always use pairs of observed and interpolated points (with appropriate weights) from different curves at the same time moments. In this way we can get rid off from another free parameter. We begin to see that the pointwise minimisation is the best type of merit function.

2.1.1.4 Correlated errors

In the context of gravitational lens research the time delay estimation is often complicated by the feature called microlensing. From the mathematical point of view this means that the two curves to be compared are not exactly similar but one or both of them contain extra low frequency component and we can not hope to achieve perfect fit of the two components, even for a correct delay.

The possibility of extra nuisance components is generally ignored and only perfect matches are seeked for. In this case the extra components (as supposed) are included into schemes as a part of observational noise. But often the complications can be severe enough to spoil whole analysis. This is especially true when after matching the final difference curve will have long and systematic excursion away from a mean (zero) level. These kinds of correlated errors indicate that we need to work with models where possible long term low frequency components are included somehow into matching scheme. For instance in Kochanek et al. (2006), Courbin et al.

(2010) authors use separate polynomial models to describe intrinsic (source) variability and extrinsic variability (microlensing). It is also possible to take into account low frequency trends by preprocessing input data sets. In this case the low degree polynomials are fitted into both light curves before matching procedure (see Pelt et al. 1994).

As we do not know, which image is affected by microlensing or other distortions, immediate subtracting of smooth trend models from observed data is not the best solution. However, if the selected trial time delay Δt^* is correct, we may suppose that the variability of the *difference curve* of the appropriately shifted A and B sequences can contain smooth trend component.

Inserting this idea and proper normalisation by the sums of weights into Eq. 1.47 we get finally the following match criterion (or *combined dispersion spectrum*) for a trial time delay:

$$CDS(\Delta t^*) = \min_{p_1, p_2, \ldots, p_P} \frac{1}{2} \left[\frac{\sum\limits_k [\hat{a}_k - b_k - h(p_1, p_2, \ldots, p_P, t_k)]^2 W_k}{\sum\limits_k W_k} + \right.$$
$$\left. + \frac{\sum\limits_l [a_l - \hat{b}_l - h(p_1, p_2, \ldots, p_P, t_l)]^2 W_l}{\sum\limits_l W_l} \right], \tag{2.1}$$

where $h(p_1, p_2, \ldots, p_P, t)$ is a smooth time dependent trend model (polynomial or spline) with parameters $p_1, p_2, \ldots, p_P$. The combined weights W_k are calculated as

$$W_k = \frac{W_{\hat{a}_k} W_{b_k}}{W_{\hat{a}_k} + W_{b_k}}, \tag{2.2}$$

where $W_{\hat{a}_k}$ and W_{b_k} are weights of predicted and observed points from A and B curve respectively. (The combined weights W_l are calculated in the similar way. When writing out the Eq. 2.2 we keep in mind that in the standard formulation of the χ^2 statistic the weights are calculated for the differences.)

Using the combined dispersion spectrum for merit function allows us to take properly into account the microlensing effect; fitting data into gaps and corresponding drawbacks are avoided; correct weighting and normalisation of this statistic is quite obvious.

Note, that as we work with data in magnitudes, we assume the magnification $m \equiv 1.0$. Adjusting the data for the baseline shift is included in the procedure of subtracting the polynomial trend ($P = 0$ corresponds to the constant shift). We have divided the criterion by 2 to get it as the estimator of the dispersion. The trend parameters $p_1, p_2, \ldots, p_P$ can be estimated using standard least squares fit method. The number of trend parameters P is a free parameter of the procedure.

To include the extrinsic component elimination procedure into fully automatic algorithm we need a method to fix the parameter P. Our experience with actual computations shows that the best way to do this is to perform analysis with a full range of possible P values. In most cases the combined dispersion spectra do not change significantly if we change P value. Only seldom the presence of a strong nuisance component demands inclusion of a significant trend component. In our final computer code we used simple polynomials with the degrees $P = 0 \ldots 10$. The final result of the time delay analysis is then formulated as a particular delay value with estimated error bars and a range of the trend parameter P values for which alternative solutions remain inside the claimed interval.

The proposed method for subtraction of the trend component is usable if the number of time points in a time series is sufficient and we do not seek for too large Δt_{AB} for the time coverage of the given data set. Otherwise the polynomial may fit and reduce the variability of the quasar. The same happens, if we use too high polynomial degree for the given time series. The results should be taken with extra care if our method suddenly starts reporting large time delays near the varying limits of the trial parameter, when the degree of the polynomial is increased. The best way to get an idea what is going on then, is just to look at the final combined dispersion spectra (Eq. 2.1) for different polynomial degrees. For most light curves the trial time delay Δt^* can be safely varied from 0.0 to $\pm(timecoverage)/2.5$. Then we can get stable time delay for some subset of polynomials and may improve the result compared to matching schemes, where correlated errors are not included. Note, that in our combined time delay estimation method to be outlined in Section 2.1.2, Δt^* can be varied in even wider range – from 0.0 to $\pm(timecoverage)/2.0$, if we do not want to subtract the microlensing distortions.

Subtracting different trends from the difference curve may help us to identify new lens systems. In this case we do not know whether there are any time delays between nearby images in a field of a photometric survey. However, we may get a stable time delay, which is in the detection range of our method, with different polynomial degrees. This may be considered as motivation for further studies of the given pair of images.

2.1.1.5 Optimal prediction

To get the interpolated intensities for Eq. 2.1 we will use the optimal prediction presented by PRH. For the completeness and also to bring out problems encountered, we describe here some important details of the method.

Estimation of the autocorrelation function Prediction based methods involve assumption that underlying source process which is observed through different channels is stationary and consequently with simple correlation structure:

$$C_{ij} = \mathbf{E}\left\{s(t_i)s(t_j)\right\} \equiv C(t_i - t_j) \equiv C(T), \tag{2.3}$$

where the autocorrelation function $C(T)$ is to be estimated from data. For that purpose PRH introduce first-order structure function $V(T)$ of the source process:

$$V(T) = \frac{1}{2}\mathbf{E}\left\{[s(t+T) - s(t)]^2\right\}, \tag{2.4}$$

and then get

$$C(T) = \mathbf{E}\left\{s^2\right\} - V(T), \tag{2.5}$$

where $\mathbf{E}\left\{s^2\right\}$ is the estimated variance of the source process. It is important to notice that in prediction procedures we need $C(T)$ values for a continuous range of argument and therefore we should have a certain parametric model for it.

From the observed data we can compute point estimates for the structure function of the source process

$$v_{ij} = \frac{1}{2}\left[(y_i - y_j)^2 - n_i^2 - n_j^2\right], T_{ij} = |t_i - t_j|, \tag{2.6}$$

which can be binned and averaged. Finally a continuous parametrised model is fitted into the binned curve to get a continuous approximation of the $V(T)$. In PRH the power-law type model for a structure function is postulated and consequently the linear model in log-log coordinates is used. Authors of the original method claim that the overall procedure of time delay estimation is quite robust against small changes in the linear model parameters estimated from data.

Structure function Following PRH – to estimate $V(T)$, we compute time-lags T_{ij} and point estimates of the structure function v_{ij} for every independent pair of data points. We sort T_{ij} and v_{ij} pairs by the value of

T_{ij}, bin and compute the bin averages $\overline{T}_{ij}$ and $\overline{v}_{ij}$. We average P_{bin} points for every bin. Next we skip bins unsuitable for the model of $V(T)$ (see the discussion below) and compute $\overline{v}_l = log(\overline{v}_{ij}{}^{\frac{1}{2}})$ and $\overline{T}_l = log(\overline{T}_{ij})$ and find the linear model for the logarithmic structure function $\overline{v}_l = a\overline{T}_l + b$. The particular value for the bin size P_{bin} is in principle free parameter of the procedure. However the final results practically do not depend on it. The light curves we analysed were long enough to use $P_{bin} = 85$ as in PRH, but for shorter data sets, where there would be too few bins, P_{bin} should be reduced.

Next we will discuss the problems we found in the PRH treatment of the structure function. PRH use pointwise subtraction of observational noise to estimate $V(T)$, as shown in Eq. 2.6. The use of the linear model for the logarithmic structure function is certainly over-simplification. As discussed by Hovatta et al. (2007) and Hughes et al. (1992), an ideal structure function of observational data should contain a plateau at the variance level of observational noise, rising part, and finally, a plateau at the total variance level at long time-lags. The structure function $V(T)$ of the source process should begin from zero level at zero time-lag and have a long time-lag plateau at the level of $\mathbf{E}\left\{s^2\right\}$. Usually we do not have zero time-lags in our time series, even if observations from different sites are combined. However, the beginning of $V(T)$ estimated from observational data may also lie on negative level in real cases, as for nearby to the zero lag data points y_i and y_j point estimates in Eq. 2.6 tend to be negative (estimated observational dispersions can be quite large, if to compare with differences). The $V(T)$ values can not be negative by definition. To avoid negative bin averages we need then – either rather large averaging bins, or we can skip the bins with negative means all together. In the original paper (PRH) the first averaging bin for the A curve was skipped from computations but it was retained for the B curve. The skipping of data points or adjustment of the bin size – both methods involve manual nudging which is unacceptable for fully automatic methods.

The real structure function may sometimes have highly oscillating large time-lag end, that may even cause negative slope for the linear fit (in log-log scale). As discussed by Emmanoulopoulos et al. (2010), the position of the upper turning point of a structure function depends not only on the underlying process but also on the length of the time series. Consequently, the possible (oscillating) large time-lag plateau may not be connected to the real underlying quasar variability and should be excluded from the model.

There are even more critical remarks about the use of the structure function. Authors in Emmanoulopoulos et al. (2010) claim that estimating the slope of the linear part of the logarithmic structure function of the source process can be done with a rather large error. But, as it was stated already in PRH, the results of optimal prediction are not sensitive to the exact parameters (slope and intercept) of the logarithmic structure function. We implemented the simple linear fit model of the PRH, but added some modifications to the structure function building procedure.

First, to avoid problems around zero lag value we skipped from our fit all the bins that were smaller than the squared mean observational noise level $\overline{n}^2$. Bins with bigger values are not so sensitive to observational uncertainties and are supposed to provide information about the power law type behaviour of the structure function. This is fully automatic step and can be always performed.

Second, we also implemented the skipping of the possible (oscillating) high level plateau from the linear fit. However, during concrete time delay estimation computations it occurred that this procedure was redundant for our test data and the results did not depend on its use. The robustness of the linear model assumption was also stated in the original PRH paper.

And finally we postulated high level constant plateau value at the level of asymptotic variance $\mathbf{E}\left\{s^2\right\}$. In this way our structure functions can have two parts: rising part from the linear fit in log-log plane and horizontal plateau at the variance level.

Variance estimation The variance value has an important role in the algorithmic implementation of the PRH type methods. However the precise or well founded estimate for it is seldom available. For instance, to estimate $\mathbf{E}\left\{s^2\right\}$, authors in Rybicki & Press (1992) suggest to take it as $10\ldots100$ times the data sample variance. We found the result of optimal prediction to be quite insensitive to this arbitrary constant, even if it was taken as large as 10^4 times of the measured variance.

As we do not know more about the variance of the source process than it can be guessed from the observed part of the time series, we estimate $\mathbf{E}\left\{s^2\right\}$ as the value of the largest bin ($\overline{v}_{ij}$) of the structure function. For the large sample of concrete computations this occurred good enough and covariance matrices were invertible. However, in the final code we also allow iterative doubling of the variance value, until to the point where covariance matrices can be correctly inverted. This procedure is again fully automatic and does not need manual nudging. As an alternative, the singular value decomposition technique can be used when the covariance matrix is ill conditioned (see Cuevas-Tello et al. 2006).

Having now the value for the variance, we can fix final form for the structure function. First, we calculate the time-lag T_{max},

$$T_{max} = \left[\frac{\mathbf{E}\left\{s^2\right\}}{10^{2b}}\right]^{\frac{1}{2a}}, \tag{2.7}$$

where the linear model of $V(T)$ in log-log coordinates reaches the $log(\mathbf{E}\left\{s^2\right\}^{\frac{1}{2}})$ value and turn our model to plateau of $\mathbf{E}\left\{s^2\right\}$ for longer time-lags. We can do that, since we do not know anything about the structure function above the estimated $\mathbf{E}\left\{s^2\right\}$ level. The final model for the structure function of the underlying process $V(T)$ is now:

$$V(T) = \begin{cases} 10^{2b}T^{2a}, & \text{when } 0 \leq T \leq T_{max}, \\ \mathbf{E}\left\{s^2\right\}, & \text{otherwise.} \end{cases} \tag{2.8}$$

It is clear from the definition that our structure functions and also corresponding autocorrelation functions are always positive.

2.1.2 The outline of the combined method

Taking into account ideas discussed in Section 2.1.1, we formulate now the combined procedure for the time delay estimation.

- First we subtract mean values from both light curves and subtract mean time moment of one curve from the time points of both curves. This puts our data into general position so that irrelevant particularities of the time and amplitude measurements will be ignored.

- For each trial time delay Δt^*

  - We shift the A light curve by a trial time delay Δt^* and select points in time-shifted curves that overlap in time. Important point here is that for different delays, regions where match can be performed are of different length. This is taken into account by normalising in the criterion function $CDS(\Delta t^*)$.
  - Next we interpolate using optimal prediction technique values for A curve at the time points of the B curve and *vice versa*. We have now two curves that have the same number of data points and the same sampling structure. For every time point we will have one original value and one interpolated value.

- – We fit smooth polynomial trend into the difference curve to eliminate the possible low frequency distortions due to the microlensing and compute $CDS(\Delta t^*)$ value (see Eq. 2.1).

- The global minimum in the run of $CDS(\Delta t^*)$ is the statistic used to select the best candidate for the true delay value Δt_{AB}. This minimisation procedure can be repeated for different trend models.

- For the established best delay values the error bars can be computed using bootstrap or a simple procedure described in Section 2.5.

2.2 Estimating time delays from one resolved image and a blend of two subimages

The resolution of our observational equipment may be good enough to resolve the most separated images of a lensed quasar. We can estimate the time delay between these images using the combined method presented in Section 2.1. But we should make sure, if either one or both of our images are blends or not. Sometimes we know it from higher resolution observations, but we can also test the data ourselves. In the present section we describe a time delay estimation algorithm for a resolved image and a blend of two subimages. The algorithm was presented and tested by Hirv et al. (2007a).

As the work with blended cases was carried out before developing the combined method which uses optimal prediction and proper treatment of the microlensing effect, we will use the standard form of the dispersion spectrum for the merit function[1].

To give the reader basic guidelines for time delay estimation in this blended case, we will use here somewhat oversimplified model for the time delayed components of the lensed image:

$$f_r(t) = a_r g(t - \Phi_r), \quad r = 1, 2, 3, \tag{2.9}$$

where a_r enables us to take into account the possible different amplification of the images. We ignore the microlensing effect (variability of amplification coefficients a_r in time) and also other possible distortions.

We observe one pure image, say $f_1(t) = a_1 g(t - \Phi_1)$ and one blend $\Theta(t) = f_2(t) + f_3(t) = a_2 g(t - \Phi_2) + a_3 g(t - \Phi_3)$. These curves can not be matched directly to recover the differential time delays $\Delta t_{1,2} = \Phi_2 - \Phi_1$

[1]Note, it is also possible to modify the combined method presented above for this blended case. For a short discussion see Section 2.2.2.

and $\Delta t_{2,3} = \Phi_3 - \Phi_2$. But we can use the "clean" curve $f_1(t)$ to build an artificial blend curve $\Gamma(t) = f_1(t) + \epsilon f_1(t - \Delta_s)$, where ϵ and Δ_s are free parameters to be determined. In the terms of the source curve we get $\Gamma(t) = a_1 g(t - \Phi_1) + \epsilon a_1 g(t - \Phi_1 - \Delta_s)$. It is not hard to see that in a fortunate case when $\Delta_s = \Delta t_{2,3}$ and $\epsilon = a_3/a_2$, the observed blend can be matched with the artificial blend by shifting the $\Gamma(t)$ in time by $\Delta t_{1,2}$ and amplifying it by a_2/a_1. The factor a_2/a_1 corresponds to the m parameter in the real sampled cases, that is calculated for every set of the trial parameters using the least squares fit (see Section 2.2.1). Then we will have three basic parameters to vary: the two artificial blend parameters ϵ and Δ_s, and Δ_l to estimate the time delay $\Delta t_{1,2}$.

If the data is in magnitudes, we can not form the replica of the observed blend directly from the unblended curve. Then we have to switch our light curves into the relative flux units for the fist step, using the definition of apparent magnitude

$$m_x = -2.5 \log_{10}(F_x/F_x^0), \tag{2.10}$$

where m_x is the apparent magnitude in the band x and F_x/F_x^0 is the relative flux in the band x compared to a standard star.

As the real light curves are discretely sampled, not continuous functions, we must know, how to add them to build artificial blends and how to subtract them to calculate the dispersion spectrum.

2.2.1 Subtraction of sampled curves

Below we will define *statistical distance* or *statistical difference* between any two sampled curves $t_i, a_i, W_i, i = 1, 2, \ldots, N_A$ and $t_j, b_j, W_j, j = 1, 2, \ldots, N_B$. Input data sets can be original data tables, data with shifted time arguments, or artificial blends computed from input data by adding time-shifted variants of it.

In the case of one resolved image and a blend of two subimages the two curves to be subtracted are: the artificial blend $t_i, \Gamma_i, W_i, i = 1, 2, \ldots, N_\Gamma$ and the observed blend $t_j, \Theta_j, W_j, j = 1, 2, \ldots, N_\Theta$. As these curves are in relative flux units, we assume that the artificial blend can be amplified by an unknown magnification coefficient m to achieve the match. The baseline shift is not needed in this case, and we may fix $h \equiv 0$. To be still more general in our treatment, we include h into our equations as well.

For a particular set of input parameters we can form a table of triples:

$$\frac{t_i + t_j}{2}, (ma_i + h - b_j)^2, W_{i,j}, \tag{2.11}$$

where $W_{i,j}$ are the statistical weights for every row. The actual values for the m and h parameters are to be estimated using the least-squares routine, and they are always calculated for every set of free parameters in our method. (In the case of one resolved image and a blend we have $\Delta_l, \Delta_s, \epsilon$ for the free parameters.) All the rows in the table of triples are not equally significant. If it happens that $t_i = t_j$, then we can assign a full weight to the corresponding row. But if the time difference between the two points is quite large (say larger than a certain pregiven value ν), then comparing the values for different curves does not make sense. Following these heuristics we introduce slightly modified version of inclusion condition or *downweighting function* that takes into account the change of the strength of correlation between different data points according to their distance in time:

$$S_{i,j} = \begin{cases} 1 - \frac{|t_i - t_j|}{\nu}, & \text{if } |t_i - t_j| \leq \nu, \\ 0, & \text{if } |t_i - t_j| > \nu \end{cases}. \tag{2.12}$$

Finally, the combined statistical weights for every row in the table of squared differences (Eq. 2.11) can be written as:

$$W_{i,j} = S_{i,j} \frac{W_i W_j}{W_i + m^2 W_j}. \tag{2.13}$$

(There is no need to compute rows for too distant pairs of points that would have zero weights. We write the Eq. 2.13 for the differences and not for the squared differences, as we will use it in the next Eq., which is in fact a generalisation of the standard χ^2 statistic. In the standard formulation of the χ^2 statistic the weights are calculated just for the differences. Note, that fixing $h \equiv 0$ does not change Eq. 2.13.) The normalised estimator of the dispersion of the difference between the two curves is now

$$DS = \frac{1}{2} \min_{m,h} \frac{\sum_{i,j} (m a_i + h - b_j)^2 W_{i,j}}{\sum_{i,j} W_{i,j}}, \tag{2.14}$$

and we may call it *statistical distance* or *statistical difference.* From the point of view of the time delay estimation scheme, it is also a *merit function* to compare different sets of trial parameters.

Because one of the parameters we search for, m, is included in the weight system, the minimisation proceeds iteratively. We first fix $m = 1$, and compute the weights using this value. Then we use a standard weighted least-squares routine to estimate both the parameters m and h. By inserting the estimated m back into the weights, we can proceed iteratively until convergence is achieved. Fortunately, only a small number (about four for 0.1% precision) of iterations is needed.

For particular shifts the overlapping part of one curve can be approximately matched with the mirror image of the other, and then the least squares fit finds negative m value. The distance computation procedure must take this possibility into account and exclude "mirror matches". We simply invert the negative m values. Then the corresponding set of trial parameters can not be recognised as the best one.

2.2.2 Addition of sampled curves

Below we compute a discrete analogue to the combined curve $\Gamma(t) = f_1(t) + \epsilon f_1(t-\Delta_s)$. Let us denote the discrete versions of $f_1(t)$ and $f_1(t-\Delta_s)$ as $t_i, a_i, W_i, i = 1, 2, \ldots, N_A$ and $t_j, b_j, W_j, j = 1, 2, \ldots, N_B$. Note, that here $N_A \equiv N_B$, but the discrete addition procedure can also be used for any pair of sampled data sets with overlapping time domain. If we use it for building the artificial blend in the case of one resolved image and a blend of two subimages, we work with data in relative flux units and include the magnification coefficient ϵ into our equations. Using similar considerations as above for the case of subtraction, we can form the triples

$$\frac{t_i + t_j}{2}, a_i + \epsilon b_j, W_{i,j}, \tag{2.15}$$

where the combined weights

$$W_{i,j} = S_{i,j} \frac{W_i W_j}{\epsilon^2 W_i + W_j}, \tag{2.16}$$

consist of appropriately propagated weights and downweighting function.

The total number of selected triples (we do not need these with zero weights) depends on the sampling and the downweighting parameter ν. In Fig. 2.1 we have added an original time series (lower part of the figure) and its shifted version to form a combined curve (upper part of the figure). In the case of a dense sampling, our constructed data set is quite redundant, especially for larger values of ν. If the sampling step is longer than ν, we may get sparser time series as well. It is very hard to choose a proper value for ν from purely theoretical considerations. However, the proper range of usable values can be established by using model or trial calculations. The sets of triples Eq. 2.11 or Eq. 2.15 can be looked upon as a new input data set for further operations.

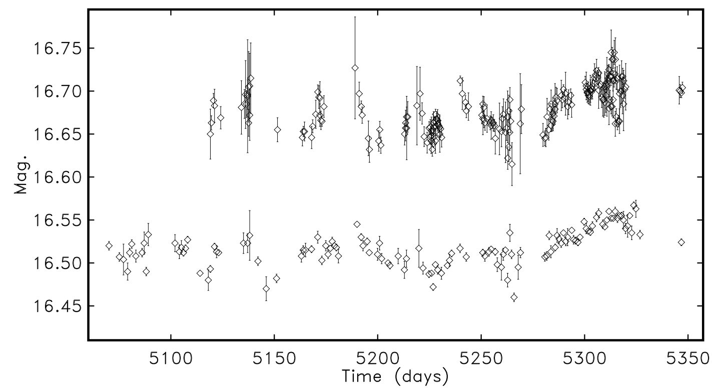

Figure 2.1: Combining an original time series with its shifted version. Depending on the spacing of time points and the downweighting parameter ν, the resulting series can be sparser (middle part of the series) or denser (right part). The combined error estimates are larger than the original (given) values.

Note, in the case of a clean image and a blend it is, in principle, possible to use the optimal prediction technique on the shifted unblended curve. First, we can shift the clean curve by trial time delays Δ_l and $\Delta_l + \Delta_s$ to make the two components of the artificial blend. Then we can (in parts overlapping in time) predict values of these components at the time points of the observed blend and after applying the trial magnification coefficient ϵ, compose the artificial blend by pointwise summation of the predicted values. When computing the dispersion spectrum, we can now subtract directly the corresponding points (at the same time moments only) of the observed and artificial blend. Then we would not need the downweighting parameter. An alternative way to apply the optimal prediction is to interpolate the intensity values of the unblended curve with a rather dense step. Then we can use simple linear interpolation in composing the artificial blend and computing the pointwise dispersion spectrum. However, we have not studied the possibility to use the optimal prediction in the case of two blended images. In principle, it is also possible to include the polynomial correction against microlensing into the blended algorithms, as it was done in calculation of the combined dispersion spectrum (Eq. 2.1).

2.2.3 Finding the best matching parameters

For every set of trial parameters $\Delta_l, \Delta_s, \epsilon$ we statistically subtract the observed and artificial blend and evaluate their distance DS. We search for the global minimum of the DS that corresponds to the best parameter combination. For some sets of trial parameters the distance computation can reveal "mirror matches". These sets actually cause "perfect" mismatches of the selected regions of the light curves and are discarded as discussed in Section 2.2.1.

There is one interesting aspect in this global search procedure – it is essentially degenerate. The degeneracy comes from the fact that the long and short delays between the blend components can be computed differently. The short delay depends on how we assign names to hypothetical parts of the blend. In one case the delay is $\Phi_3 - \Phi_2$, but in another case $\Phi_2 - \Phi_3$. And corresponding long delays will be also different: $\Phi_2 - \Phi_1$ and $\Phi_3 - \Phi_1$. This degeneracy results in symmetrically placed minima on the grid of the time delays (see for instance Fig. 3.10). For finite sequences both solutions can give slightly different values for the merit function because of the boundary effects. Sometimes physical considerations can define the proper order of total flight times and then we do not need to compute full grids, but can restrict our computations to only one half of them. However, as a sanity check, it is worth to compute merit function values for a larger parameter grid. Then the overall pattern of symmetrically shaped and mirrored minima allow us to get the general impression of the validity of our solutions.

It is also important to check how the actual distribution of time moments in input data sets influences the the number of data point pairs that are used in computing the final expression for dispersion. Even in the case of the comparison of two pure (one of them shifted) light curves, it is not ruled out that for a particular time delay the observations of one curve occur just in the gaps of the other. The well-known controversy on the time delay of the classical double quasar QSO 0957+561 was just a result of this kind of accident (see for details Press et al. 1992a,b; Pelt et al. 1994). In this case, the number of pairs that can be used to compute the merit function may be reduced too much and we may get statistically unstable DS value. Multidimensional graphs of the parameter dependent pair counts or sums of weights from Eq. 2.14 can reveal regions where there may not be sufficient information to estimate the parameters of the model.

2.2.4 Choosing the downweighting parameter

Sampling determines the shortest possible delay we are able to find from the particular data. If the downweighting parameter ν (see Eq. 2.12) is too small for a given sampling, we will have too few pairs in the calculation of the merit function and the map of DS will be poor due to noise and boundary effects. On the other hand, enlargement of ν is limited because of the smoothing effect of this parameter – using larger ν reduces the possibility of finding shorter time delays.

We can use an interactive simulation for estimating the suitable downweighting parameter for real observational data. First, we generate artificial noise-free curves with some pregiven time delays, using the sampling of our real data. Then starting from small downweighting parameter (say $\nu = 0.5$), we move on towards larger ones and recalculate the plot of DS and recover the time delays for each ν. In general, there is an optimal ν for a given sampling which recovers the time delays correctly and produces clearest minimum on the DS surface. Once we have found the optimum, further enlargement of downweighting parameter will not improve the results. It is also possible, that for a given sampling and time-delay system, there is no working downweighting parameter at all. Even for a correctly estimated value of ν the overall success of the algorithm depends on the length of the time series, noise level and absolute values of actual time delays.

As a rule of thumb ν should be kept equal to or smaller than half of the shortest possible time delay Δt_s we are trying to find. For sound statistics, we should have on the average at least 3–5 pairs for every observed time point when combining and subtracting the time series. In practice it would be useful to carry on computations with varying values of ν to check statistical stability and robustness. See for instance Pelt et al. (1996) where such an analysis was used for a simple case of delay estimation.

2.3 Estimating time delays from two blends of two subimages

A specific case of lensed images, when we have two blended light curves of two subimages in each, was developed and tested by Hirv et al. (2007b). Below we will describe the method in some detail.

Let us have a quasar image split into four components by an intervening gravitational lens. Formally we have four functions of the quasar source variability $g(t)$: $f_r(t) = a_r g(t - \Phi_r), \quad r = 1, \ldots, 4$, where a_r are the magnification coefficients and Φ_r are the flight times. We ignore again the microlensing effect (variability of a_r in time) and also other possible distortions. Our observational equipment can supposedly record only two images as the close pairs of f_1, f_2 and of f_3, f_4 are blended together due to insufficient resolution. Thus the corresponding signals $\Theta_1(t)$ and $\Theta_2(t)$, that we are going to observe, are the following functions of the source variability:

$$\Theta_1(t) = a_1 g(t - \Phi_1) + a_2 g(t - \Phi_2), \tag{2.17}$$

$$\Theta_2(t) = a_3 g(t - \Phi_3) + a_4 g(t - \Phi_4). \tag{2.18}$$

As the spatial separation of f_1 and f_2 is small, we may assume, that $a_1 \approx a_2$ and similarly $a_3 \approx a_4$ for f_3 and f_4. The amplification ratio between $\Theta_1(t)$ and $\Theta_2(t)$ is then $\approx a_1/a_3$. Let the time delay between $f_1(t)$ and $f_2(t)$ be $\Delta a \equiv \Delta t_{1,2} = \Phi_2 - \Phi_1$, and the time delay between the components of the second image $\Delta b \equiv \Delta t_{3,4} = \Phi_4 - \Phi_3$. These delays are typically rather short due to nearby flight paths for the component images. As the paths of $f_1(t)$ and $f_3(t)$ differ significantly (larger spatial separation), the corresponding delay $\Delta c \equiv \Delta t_{1,3} = \Phi_3 - \Phi_1$ is the longest one. Now we can rewrite the Eqs. (2.17) and (2.18) in terms of the first subimage $f_1(t)$ and relative time delays:

$$\Theta_1(t) = f_1(t) + f_1(t - \Delta a), \tag{2.19}$$

$$\Theta_2(t) = f_1(t - \Delta c) + f_1(t - \Delta c - \Delta b). \tag{2.20}$$

To keep things easier to follow we did not divide the right side of Eq. (2.20) by the amplification ratio $\approx a_1/a_3$. The fact, that $\Theta_1(t)$ and $\Theta_2(t)$ may have different amplitudes (or baselines in logarithmic scale) is taken into account in our matching algorithm. As a schematic example of the initial variability, the $f_1(t)$ is shown as a single-peaked function in Fig. 2.2. Shifting it by delays Δa, Δb and Δc and adding the results as in the Eqs. (2.19) and (2.20) we get the double peaked blends $\Theta_1(t)$ and $\Theta_2(t)$ of the source variability.

To recover all the three independent time delays Δc, Δa and Δb hidden in the light curves $\Theta_1(t)$ and $\Theta_2(t)$, we will combine the data using three trial delays δc, δa and δb into artificial blends $\Gamma_1(t)$ and $\Gamma_2(t)$:

$$\Gamma_1(t) = \Theta_1(t - \delta c) + \Theta_1(t - \delta c - \delta b), \tag{2.21}$$

$$\Gamma_2(t) = \Theta_2(t) + \Theta_2(t - \delta a). \tag{2.22}$$

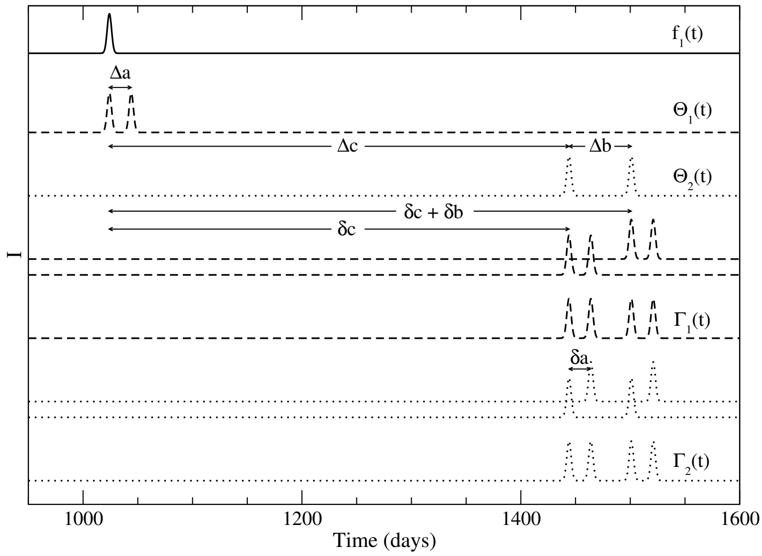

Figure 2.2: Graphical explanation of the method. See text for details.

If it happens, that $\delta c = \Delta c$, $\delta a = \Delta a$ and $\delta b = \Delta b$, the difference of $\Gamma_1(t)$ and $\Gamma_2(t)$ vanishes to zero and this is the situation we are going to search for. The composition of the artificial blends $\Gamma_1(t)$ and $\Gamma_2(t)$, when the trial delays match the initial delays, is also shown in Fig. 2.2. For clarity we plotted the components of the artificial blends before and after adding. Blends and components that have the same origin are plotted using the same line type. As we can see, artificial blends have the same profile, when trial delays correspond to the initial ones, and the difference between $\Gamma_1(t)$ and $\Gamma_2(t)$ vanishes. This is the idea of our method in terms of the continuous and noise-free light curves.

For real sampled and noisy data we use the algorithms described in Sections 2.2.2, 2.2.1, 2.2.4.

Taking into account assumptions made before, we may compose our two artificial blends to be matched either from observational data given in magnitudes or in relative flux units. In order to compose an artificial blend, we denote the time shifted versions of the same observed blend as $t_i, a_i, W_i, i = 1, 2, \ldots, N_A$ and $t_j, b_j, W_j, j = 1, 2, \ldots, N_B$, and add them as described in Section 2.2.2. Note, that $N_A \equiv N_B$ here. In composing the artificial blends Γ_1 and Γ_2 we do not need the magnification coefficient and we fix $\epsilon \equiv 1$ in the Eqs. 2.15, 2.16.

Next, by varying trial delays δc, δa and δb we recalculate the artificial blends Γ_1, Γ_2 and compute the statistical distance DS between them (see Eq. 2.14). In the case of logarithmic light curves we use the baseline shift h as regression parameter and fix amplification $m \equiv 1$. If we prefer to work in relative flux units we have m as regression parameter and we fix $h \equiv 0$ in the computation of the dispersion spectrum.

By varying trial delays δc, δa and δb over pre-given grids, we are searching for the global minimum of statistical distance DS which corresponds to the recovered time-delay system. Keeping in mind that negative m values in the computation of the statistical difference are useless we exclude corresponding values of DS from the search.

Recovering the time-delay system is still a degenerate problem. The mirrored values of short delays Δa and Δb are also valid. For a single data set we can get four equally correct solutions: Δc, Δa and Δb; $\Delta c + \Delta b$, Δa and $-\Delta b$; $\Delta c - \Delta a$, $-\Delta a$ and Δb; and $\Delta c - \Delta a + \Delta b$, $-\Delta a$ and $-\Delta b$. (Interchanging Θ_1 and Θ_2 gives us four additional sets of solutions, where Δc is mirrored and Δa and Δb are interchanged.) All the four solutions form detectable minima in the three-dimensional grid of DS values. For finite sequences these minima may have slightly different merit function values. Our method just finds formally the deepest minimum and corresponding time-delay system. The recovered set of time delays may be considered real, if it shows up as a visually noticeable minimum in the two-dimensional slice of statistical distance values. Formal error estimation is possible using the bootstrap-type techniques and ideas from Pelt et al. (1996).

Our method does not work if $|\Delta a| = |\Delta b|$. Both observed blends are then similar, and we can recover only the largest delay Δc using simplest "one-dimensional" dispersion spectrum. Having a value for the long delay it is then in principle possible to recover the short delay (the same for both blends) from the combined data using the methods described in Geiger & Schneider (1996). (The combining of two photometric series with estimated long delay allows sometimes – if microlensing effect is negligible – to get a data set with twice the original mean sampling rate.)

The case of $|\Delta a| = |\Delta b|$ may be promptly recognised from the plot of DS values – one of the four possible solutions has a characteristic distribution along straight line of DS values (see for instance Fig. 3.17). We may also hit an arbitrary solution corresponding to mirrored arbitrary short delays, which shows up as a normal minimum in the two-dimensional plot of DS values. Hence the solutions where $|\delta a| \approx |\delta b|$ should be handled with care. A three-dimensional plot of DS values would be useful here.

Our method for two blends recovers the time delays correctly also for a clean image and a blend. However, as it will be discussed later, the sensitivity to observational noise of the method for two blends is higher. It is sensible to use proper method for the nature of a given problem. For unknown nature, it is worth trying both algorithms.

2.4 Evaluation of the methods

2.4.1 Resolved images

There are two different ways to evaluate algorithms for time delay estimation. First, we can use simple random walk for generating artificial light curves with known time delays and sample them randomly or use sampling of some real time series. Then we apply different time delay finding methods to the generated data.

However, algorithm that works best on generated light curves, may not always perform well on real observational data. As we are unable to simulate all observational effects, method that is trained to work on artificial data may give wrong results in real case. Hence we decided to use real observational data for evaluating time delay estimation procedures for resolved images.

If the time series is well sampled, has sufficient time coverage and reasonably small observational errors, the results of applying all useable time delay finding methods should converge on the same value. The situation changes when we use them for time series of lower quality and shorter durability. The method, that works well and gives the same answer with data of both higher and lower quality of the same object, should be recognised as more stable and consistent.

2.4.2 Blended images

Currently we do not have blended observational sequences at our disposal, that are long enough, sampled well and have noise level our methods can work with. So, to test the blended methods, we had to build artificial sequences.

2.4.2.1 Generating test curves

The generation of simulated data is simplified by the fact that model curves for different images can be computed from the same source curve. We used a simple random walk procedure to generate the source variability curves.

The time steps for the curves were selected according to two principles: they must be shorter than typical sampling intervals and they must be longer than typical photometric integration times. So we generated the initial time points by using random step sizes from the interval $[0.2, 1.8]$ days. The simulated intensities were obtained by adding a random value of ± 1.0 cumulatively in each step. It was interesting to observe that sometimes the generated curve was quite poor in features (minima and maxima, etc.). In these cases we discarded them. There is a similar effect when dealing with actual lens systems. A quasar can be "quiet" for a long time and its photometry is not sufficient for time delay estimation.

The generated source variability curve was then used to read off (using appropriate time shifts and linear interpolation) all the image components f_r. In the case of one resolved image and a blend we have to work in the relative flux units. Hence, to simulate different magnifications of the lensed images, we amplified one component of the simulated blend by the amplification coefficient a_3. In the case of two blends we also chose the relative flux units and magnified one blend against the other by a_3. (The inherently important assumption of the method for two blends is that both components of a given blend have nearly equal magnification coefficients.) We used linear interpolation for composing our test blends.

The simulated light curves can be resampled using generated or real observed sequences of time points and linear interpolation. To take into account daylight and randomly changing observational conditions, part of the time points were discarded in generating the simulated samplings.

Different levels of Gaussian noise may be added to the simulated light curves to check our methods' stability against noise. We can use the weight systems of real observed time series for our simulated data. In this case we have to scale the simulated curves or observed standard errors appropriately. We chose to scale standard errors given by observer according to the ratio of the full amplitudes of the real and simulated data. Next these scaled errors were used as standard deviations for Gaussian noise components, which were added to each point of simulated data.

One of the generated curves and the blend constructed from it is shown in Fig. 3.3.

2.5 Estimating the precision of the results

In the case of resolved images we may obtain robust confidence intervals for time delays using bootstrap technique (see Pelt et al. 1996). We can take the optimal prediction of the light curve as a model and resample the residuals between observed and model curves to get bootstrap estimations of the time delay. As such procedure may be very time-consuming and giving the most accurate error bars was not the aim of our work, we used another idea by PRH instead: the interval of the trial time delay Δt^*, that increases the $\chi^2(\Delta t^*)$ curve by 4 units from its minimum $\chi^2(\Delta t_{AB})$, corresponds approximately to the 95% formal confidence interval of the time delay Δt_{AB}. In order to use this approach, we have to rescale the minimum of our $CDS(\Delta t^*)$ curve to the value of $\chi^2(\Delta t_{AB})$ that can be obtained from the *not* normalised version of Eq. 2.1:

$$\chi^2(\Delta t_{AB}) = \min_{p_1,p_2,\ldots,p_P} \left[\sum_k [\hat{a}_k - b_k - h(p_1,p_2,\ldots,p_P,t_k)]^2 W_k + \right.$$
$$\left. + \sum_l [a_l - \hat{b}_l - h(p_1,p_2,\ldots,p_P,t_l)]^2 W_l \right]. \qquad (2.23)$$

After rescaling, the $CDS(\Delta t^*)$ curve has the same normalisation in the proximity of the minimum point as it is for $\chi^2(\Delta t^*)$. A parabola can be fitted into the neighbourhood of the $CDS(\Delta t^*)$ minimum to make estimating the error bars easier.

In the case of blended images and simulated light curves we can estimate the errors of the time delays and magnification coefficient using Monte Carlo type calculations. We add appropriately scaled Gaussian noise components to the noise-free model curves, so that the expected signal-to-noise ratios were the same as those for observational data. We repeat this, say, $3000 \ldots 5000$ times and store the obtained time delays and magnification coefficients. From this resulting table we calculate the standard deviations for the estimated parameters.

Chapter 3

Discussion and results

3.1 Requirements for observations

If a galaxy acts as a lens, the separation between lensed images is only a few arcseconds (see Refsdal 1964a; Refsdal & Surdej 1994). We need high angular resolution and good seeing for observations of lensed quasars.

If the quasar is not in quiet state, we can measure the time delays between light curves of different images. As the time delays may be in order of several days to hundreds of days for real gravitational lenses (see Refsdal & Surdej 1994; Schneider et al. 2006), longtime monitoring of quasar images is needed. The total time base of observations must be determined from the expected length of the longest time delays. We cannot sensibly find longer time delays from the observed time series, than about half of the length of the time series itself. Moreover, if we want to take the possible microlensing into account in the case of resolved images, the longest sensible trial time delay is about 1/2.5 of the duration of the time series. This results from the fact that using the simple polynomial $h(p_1, p_2, \ldots, p_P, t)$ we start to fit and reduce the variability of the quasar in addition to the microlensing, if the overlapping part of the two light curves is too short for the given polynomial degree. Using trigonometric polynomials is probably a better choice, when large shifts are analysed.

The source must have brightness variations on time scale that is shorter than the monitoring period. Sampling should be dense enough to match the characteristic periods of variability of the source quasar. The pair counts or sums of weights for every parameter combination to be compared must be large enough to avoid statistically unstable merit function values (see Sections 2.2.3, 2.2.4 and Fig. 3.15) – *i.e.* sufficiently dense sampling is needed again. This is especially important for the blended methods where interpolation is not used. For the CDS statistic, recovering the characteristic periods of the source variability with sampling, is still crucial. Although our methodology does not fit data into gaps, a sufficiently dense and truly random distribution of the time points is still the best choice. Unfortunately, ground-based astronomical observations tend to have (nearly periodic) gaps. To lessen the effect of gaps, combined observations from different sites can be used.

3.1.1 The tolerable noise level of the observational data

The more inhomogeneous and larger are the observational errors for a given sampling, the noisier is the minimum of the dispersion spectrum and the more insecure is the result of the time delay estimation.

To see, whether the noise level of the observational data is tolerable for our time delay estimation algorithm, we can build simulated noise-free light curves using the sampling of the observational data and add gradually varying levels of Gaussian noise to them. Next we try to recover the foreknown time delays from these simulated data sets. We look for the signal to noise ratio (S/N) of the simulated data when we loose the solution and compare it with the S/N of the real observed data. The observational data can be considered usable if the S/N of it is higher than for the usable test data.

Let us take the case of a clean image and a blend of two subimages as an example here. We analysed the light curves of the QSO 0957+561 provided by Schild[1]. We selected for further calculations a 4202 day long sampling-interval from the original time series. We composed simulated data with pregiven $\Delta t_{1,2} = 420.15$ days, $\Delta t_{2,3} = 20.21$ days, $a_3 = 0.8$ and added different amount of Gaussian noise to it. The trial amplification coefficient was held fixed ($\epsilon \equiv 0.9$) at the value recovered for the noise level corresponding to the original data. The trial delays were varied in the ranges $[360, 480] \times [-40, 40]$ with 0.1 day step. The different noise levels and corresponding results are given in Table 3.1. The plots for four noise levels are given in Fig. 3.1. It is clearly seen how the minima are smeared out when the noise level rises. As it was mentioned above, sometimes the solution can jump to its mirror place (and in this case we essentially do not lose it). For higher noise values we can completely lose the correct solution. As we can see, noise level must be kept under 5% of the amplitude of light

[1] http://cfa-www.harvard.edu/~rschild/fulldata2.txt

curve (or the S/N above 20) for the given sampling in order to use the method for a clean image and a blend. Fortunately, the mean noise levels for the selected sampling interval of the observed data were 4% and 3% for the A and B curves respectively.

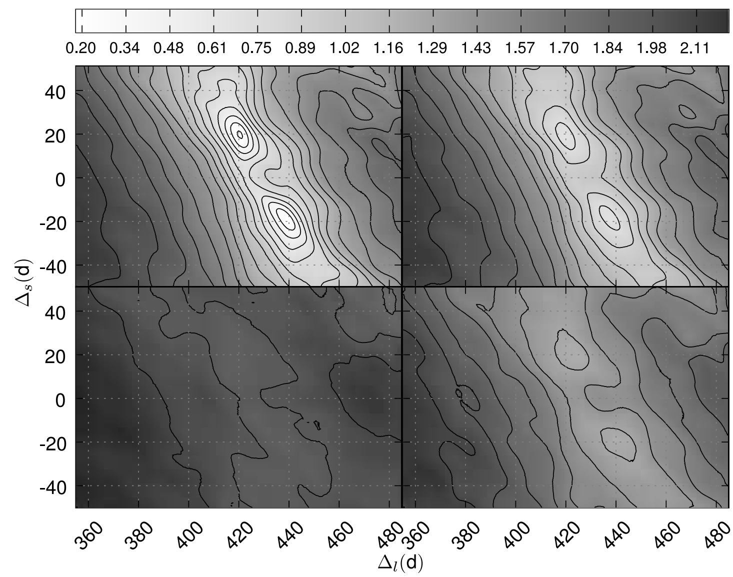

Figure 3.1: Vanishing of the characteristic minima of the merit function values due to observational errors in the case of a clean image and a blend of two subimages. The noise levels are (clockwise from upper left): 0%, 2%, 5%, and 10%. Values on the colour key represent the $log(DS)$ and spacing of the contours. The same type of colour key is used in all two-dimensional plots.

The process of calculating the DS values in the algorithm for two blends is different from its analog for a clean curve and a blend. The calculation of the DS involves here differences of the observed data sums. In the case of a clean image and a blend, we have differences of original data points and combined sums. From what follows that total scatter of the differences in the method for two blends is somewhat higher and statistical stability is lower. Consequently, the two blend method demands data with higher quality. We found that for Schild's sampling the $S/N \geq 50$ or the noise level under 2% was needed for the method for two blends to work properly. Consequently, we could not use the method for two blends for searching the time delays from Schild's data.

Table 3.1: Location of global minima depending on added Gaussian noise in the case of a clean image and a blend of two subimages.

Noise (percent)	Δ_l	Δ_s
0	420.6	19.0
1	420.5	19.9
2	419.9	19.1
3	438.5	−19.0
4	423.4	20.0
5	419.5	26.9
6	417.5	80.9
10	480.0	46.9
15	499.4	80.9
20	481.4	43.9

3.2 Data sets

In Table 3.2 we present the real observed data sets used for testing the time delay finding algorithms for resolved images, and some simulated data sets used for testing the methods for blended images. The reference, object, number of time points and duration of each light curve are given. In the same table we also present the time delays for real data found by original authors; the foreknown time delays and magnification coefficients for simulated data; and our results as well. For the real data we give also the range of accepted polynomial degrees. Note, that $\Delta t_{A,B}$ is positive, if variability of the A image is preceding the variability of the B image.

Table 3.2: Data sets and results.

Object	Points	Duration	Original Delays (days)			Our Delays (days)		
		(days)	$\Delta t_{A,B}$	$\Delta t_{D,A}$		$\Delta t_{A,B}$	$\Delta t_{D,A}$	P
HE 0435-1223[1]	143	606		$-14.37^{-0.85}_{+0.75}$			(-16...-14)±2	1..10
HE 1104-1805[2]	236	1630	-157±21			(-160...-156)±8		0..10
HE 1104-1805[3]	245	1763	-161±7			(-160...-159)±6		0..7
HE 1104-1805[4]	383	3279	$-152.2^{-2.8}_{+3.0}$			(-161...-156)±6		0..9
QSO 0957+561[5]	97	581	417±3			417±2		0..5
QSO 0957+561[6]	1233	6805	416.3±1.7			(417...426)±2		0..9
QSO 0957+561[7]	131	2926	415±20			(412...416)±6		1..9
SDSS J1004+4112[8]	104	259	-40.6±1.8			-40±2		4..10
			$\Delta t_{1,2}$	$\Delta t_{2,3}$	a_3	$\Delta t_{1,2}$	$\Delta t_{2,3}$	a_3
TEST-1	306	734	50.2	10.6	1.3	50	11	1.1
TEST-2	1032	4202	420.15	20.21	0.8	420.2±0.8	20.0±1.2	0.9±0.1
TEST-3	1233	6806	412	0	1.0	410	2	1.0
TEST-4	1233	6806	412	22	1.0	412	25	1.0
			Δa	Δb	Δc	Δa	Δb	Δc
TEST-5	2740	4300	20.2	56.5	420.2	19	59	417
TEST-6	932	1898	56.5	50.1	420.2	56	50	420.2
TEST-7	932	1898	20.2	20.2	420.2	–	–	420.2

[1] Kochanek et al. (2006); [2] Wyrzykowski et al. (2003); [3] Ofek & Maoz (2003); [4] Poindexter et al. (2007); [5] Kundic et al. (1997); [6] Schild (Pelt et al. 1998b); [7] Vanderriest et al. (1989); [8] Fohlmeister et al. (2008). Note that a_3 is the dimensionless amplification factor and P is the accepted polynomial degree.

As an example of a real observed time series the light curves of QSO 0957+561 by Vanderriest et al. (1989) are shown in Fig. 3.2.

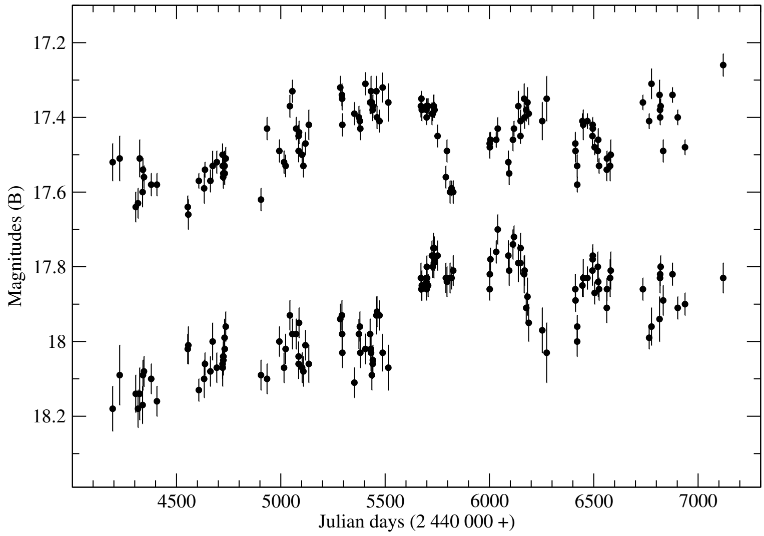

Figure 3.2: A 2926 day long photometric time series of QSO 0957+561 by Vanderriest et al. (1989). B colour of Johnson, 131 points, A curve is upper and B curve is lower (B is shifted by +0.4 magnitudes).

The TEST-1 simulated time series for a clean image and a blend of two subimages has daylight caps and additional random gaps to take into account changing observational conditions. 5% of Gaussian noise was added to the both curves. The light curves of TEST-1 are depicted in Fig. 3.3.

The TEST-2 simulated light curves for a clean image and a blend were sampled using the time points of the observational time series of the QSO 0957+561 by Schild. The parameters used in this simulation are mostly the same as in Section 3.1.1, but the noise part was generated using the observational standard errors. The standard errors given by Schild were scaled according to the ratio of the full amplitudes of the real and simulated data. Next these scaled errors were used as standard deviations for Gaussian noise components, which were added to each point of simulated data.

The TEST-3 and TEST-4 simulated light curves for a clean image and a blend have the sampling and weights according to the full 6806 day time series of the QSO 0957+561 by Schild.

In the TEST-5, TEST-6 and TEST-7 simulated data sets for the case of two blended images we have multiplied the blend Θ_2 by the amplification

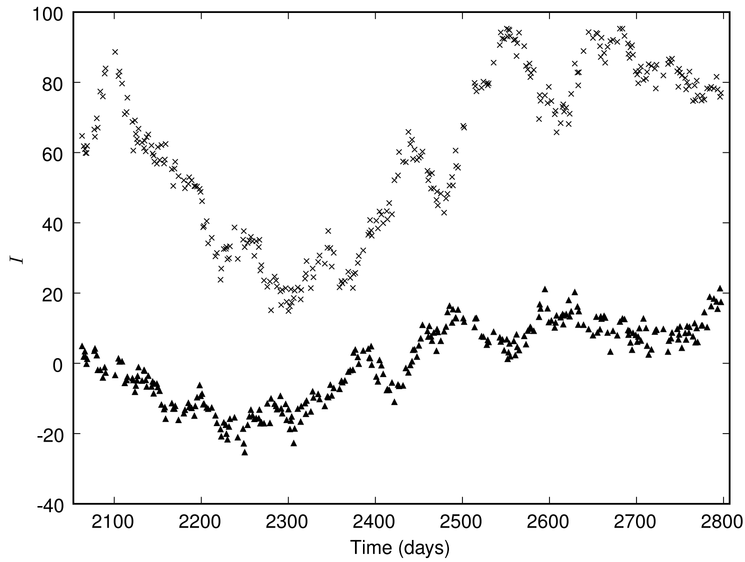

Figure 3.3: The TEST-1 data – a basic computer-generated random walk f_1 (lower curve) and a computer generated blend curve Θ ($\Delta t_{1,2} = 50.2$, $\Delta t_{2,3} = 10.6$, the amplification coefficient $a_3 = 1.3$; we added 5% noise to both curves and shifted the blend up by 60 units).

factor $a_3 = 0.8$. These sets have random sampling with gaps in day time only. The level of Gaussian noise for the particular run of set TEST-5 given in Table 3.2 is 3%. The TEST-6 and TEST-7 sets have no noise added. The TEST-5 light curves are depicted in Fig. 3.4.

3.3 Using the methods for time delay estimation

3.3.1 Resolved images

The results of using our combined method for time delay estimation from real observed data is given in the first 8 lines of the Table 3.2. Uncertainties of order 3%–5% or less in the time delay measurements are needed to estimate H_0 with sufficient precision (Schneider et al. 2006). We can see from Table 3.2 that for some data sets the precision of the time delay estimate falls well into that range.

The important point about Table 3.2 is that all computations for the real observed light curves are done with one and the same algorithmic set up. Even the only free parameter of the structure function building procedure (the number of observation pairs in bin) was set to be $P_{bin} = 85$

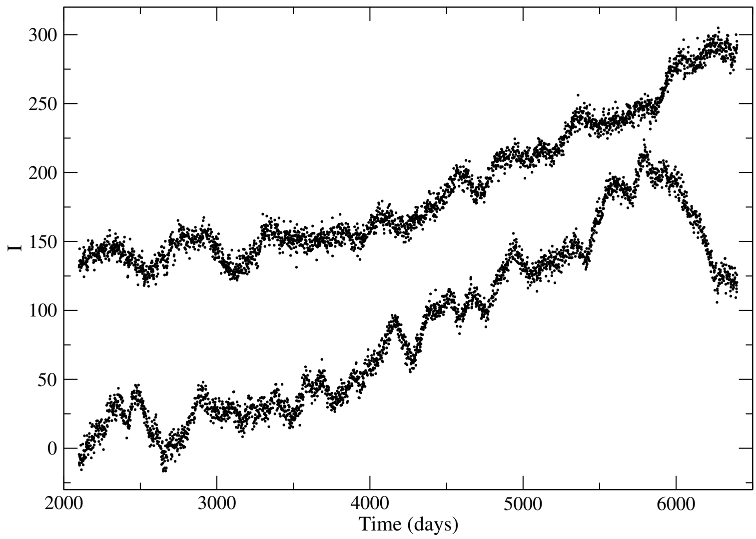

Figure 3.4: The TEST-5 data – the computer-generated blends Θ_1 (lower curve) and Θ_2 (shifted up by 130 units). $\Delta c = 420.2$, $\Delta a = 20.2$, $\Delta b = 56.5$ days and $a_3 = 0.8$. The standard deviation of the added Gaussian noise is 3% of the amplitude of the light curves.

for all tests. The step of trial time delay Δt^* was 1.0 days. In general case we can vary Δt^* in the range from 0.0 to $\pm(timecoverage)/2.5$. For a single sharp event in the light curves this may not be the case, especially if a predicted feature is observed to establish the time delay value (see for an example Kundic et al. 1995, 1997). We chose to vary Δt^* in the ranges used by original authors.

For QSO 0957+561 and HE 1104-1805 we have three separate time series. As we got consistent results for these light curves of different time coverages, samplings and weight systems, we can say that our method is quite stable against variable observational quality.

Details about particular data sets follow.

3.3.1.1 QSO 0957+561

As the first test data for our method we used the three photometric time series of the most well known lens system QSO 0957+561, published by Vanderriest et al. (1989), by Schild[2] and by Kundic et al. (1997). All the three data sets were analysed previously multiple of times.

[2] http://cfa-www.harvard.edu/~rschild/fulldata2.txt

Vanderriest et al. (1989) got initially 415 ± 20 days for the time delay between the A and B light curves. They used cross-correlation method on interpolated light curves and also cross-covariance method with discrete Fourier transform to obtain that value. PRH analysed the same data set and found 536 ± 14 days instead. In Pelt et al. (1994) the both delays were obtained by using different schemes of analysis. Later Pelt et al. (1998b) analysed a much longer and detailed data set provided by Schild and reported 416.3 ± 1.7 days for the time delay. In the interesting project Kundic et al. (1995) found a significant drop in the photometry of the A component light curve. The observed event was used to predict time moment for a similar drop in the B curve. The follow up observations one and half year later confirmed the value 417 ± 3 days (Kundic et al. 1997). The currently accepted time delay value for this system is still around 417 days (see Colley et al. 2003; Shalyapin et al. 2008, and references therein). However there are another probable values around $422 \ldots 426$ days which are supported by some authors (Oscoz et al. 2001; Goicoechea 2002; Ovaldsen et al. 2003).

Together with combined method we implemented the PRH method and applied it to the Vanderriest et al. (1989) data. As in original paper we got 536 ± 10 days for the time delay (see Fig. 3.5). But, using our new approach (ignoring possible microlensing), we got 440 ± 6 days (see Fig. 3.6). This value is somewhat nearer to the currently accepted value but still off target. Much more clearer picture is revealed when we perform delay search using polynomial trend models. In Table 3.3 we listed our results for a range of polynomial degrees. Three specific spectra are also depicted in Fig. 3.7.

From the results of our fully automatic combined method (see Table 3.2, Table 3.3 and Figs. 3.6-3.7) we can conclude the following:

- *Vanderriest data, no trend.* The pointwise matching using Eq. 2.1 gives somewhat more realistic delay estimate if to compare with PRH method (440 against 536, true value assumed to be around 417 days). The effect of data fitting into the gaps is not so pronounced, but result is still off target.

- *Vanderriest data, with trend model.* For degrees $P = 1 \ldots 9$ we got consistent set of delay values well inside of error bars of the current best estimates and also similar to the value obtained in the original paper. From what follows that original implementation of the PRH method was unsuccessful because of two reasons – data fitting into the gaps due to the use of global χ^2 matching criterion and also due to

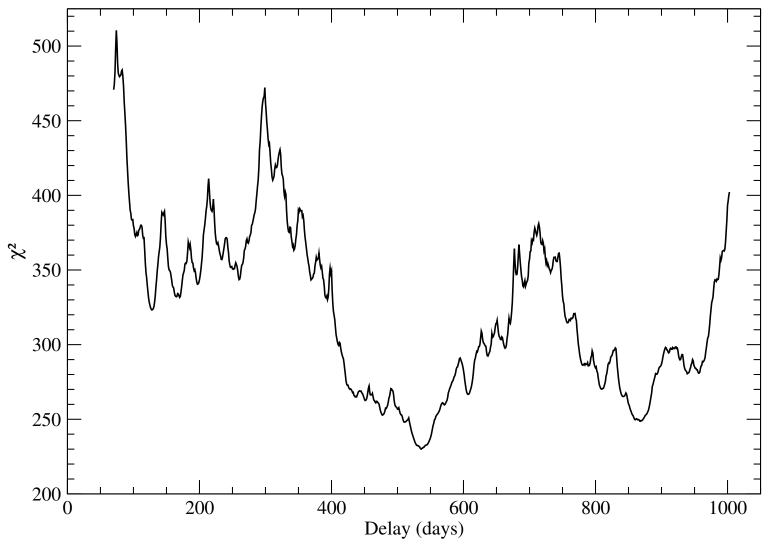

Figure 3.5: The χ^2 curve of the PRH method applied to the Vanderriest et al. (1989) data. Delay estimate $\Delta t_{AB} = 536$ days.

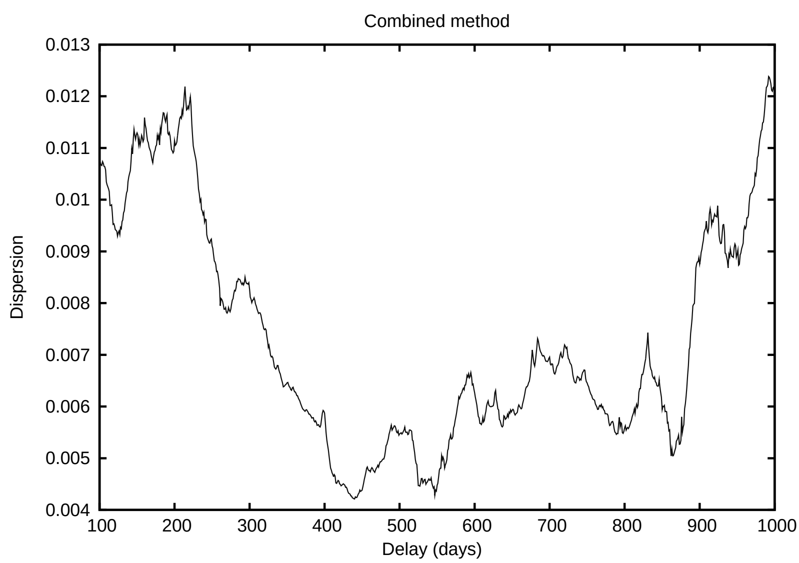

Figure 3.6: The output curve of the combined method applied to the Vanderriest et al. (1989) data. Extrinsic variability ignored ($P = 0$), delay estimate $\Delta t_{AB} = 440$ days.

Table 3.3: Trend effect on time delays.

Polyn. deg. P	QSO 0957+561[1] A,B (days)	HE 1104-1805[2] A,B (days)	HE 0435-1223[3] A,D (days)	SDSS J1004+4112[4] A,B (days)
0	440	-158	-18	-40
1	412	-159	-16	-40
2	412	-158	-15	-43
3	415	-160	-15	-38
4	415	-160	-15	-40
5	416	-160	-15	-40
6	413	-160	-15	-40
7	416	-160	-15	-40
8	416	-160	-14	-40
9	412	-160	-14	-40
10	531	-156	-14	-40

[1] Vanderriest et al. (1989); [2] Wyrzykowski et al. (2003); [3] Kochanek et al. (2006); [4] Fohlmeister et al. (2008).

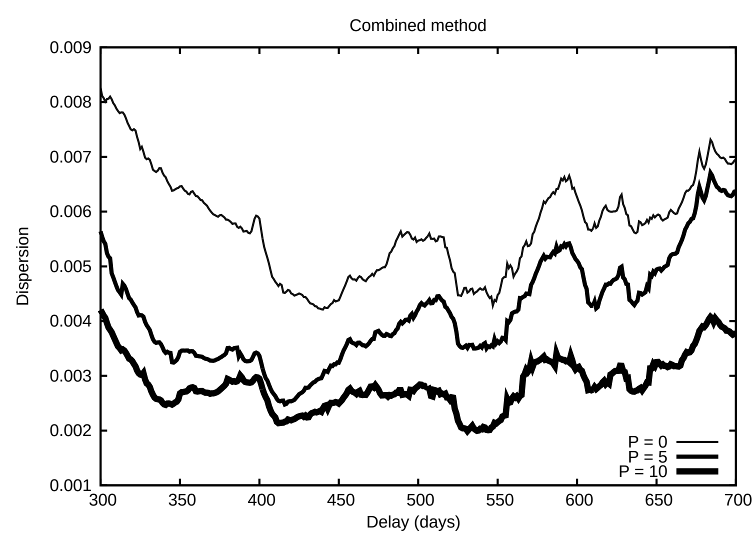

Figure 3.7: Combined dispersion spectra for three different polynomial degrees. Vanderriest et al. (1989) data. See Table 3.3 for delay estimates for different P values.

the leaving off possibility of microlensing. This important conclusion is of separate interest (see Press & Rybicki 1997).

- *Schild's data.* Paradoxically, the most abundant and longest data series for the double quasar does not help us to fix time delay finally and sharply. We are not going to solve here this so called *small controversy* of the QSO 0957+561 time delay (see Goicoechea 2002; Hirv et al. 2007a; Shalyapin et al. 2008) and leave it for further studies.

- *Kundić's (g-filter) data.* The time delay 417 days (also confirmed by use of the combined method) between two sharp features in the A and B light curves is used by many authors as the definitive value. However, because of *short controversy* we are not so convinced. Long time statistical behaviour of the light curves is quite complex and final word is not said.

3.3.1.2 Other data sets

From Table 3.2 and Table 3.3 the results for other five observational data sets can be read off. One of the particular solutions is also illustrated in Fig. 3.8.

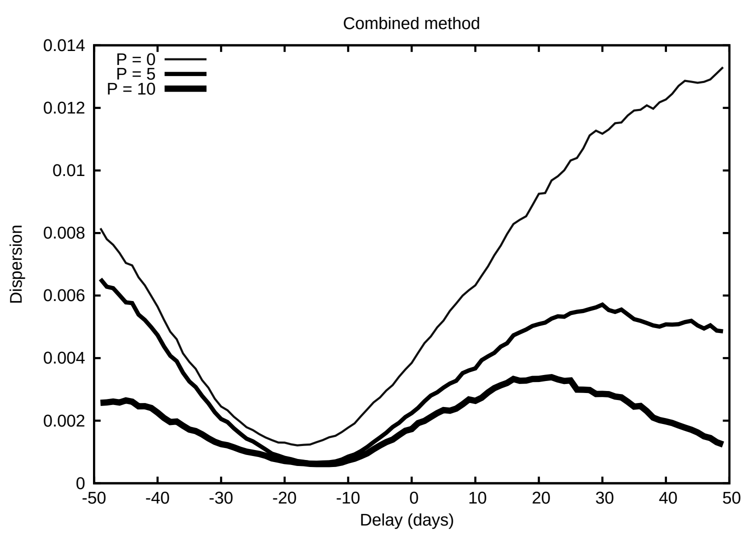

Figure 3.8: Combined dispersion spectra for three different polynomial degrees. Kochanek et al. (2006) data. See Table 3.3 for delay estimates for different P values.

We take an opportunity to stress once more – all the results obtained are computed by using our software as a black-box. No manual nudging, fixing certain free parameters or extra selection among different variants. Typical output of our code is just a list of delays for different trend degrees and as it is seen from Table 3.3 this is enough – the best estimate reveals itself as a sequence of similar or absolutely equal values in the list. Of course, there are some important restrictions. We must have enough observations, the sampling must have a reasonably good coverage, the observational errors should not be exceedingly large etc. But these are just standard demands for a good photometry.

3.3.2 One resolved image and a blend of two subimages

3.3.2.1 Random sampling

First we used the TEST-1 simulated light curves (see Section 3.2 for details) to show the different capabilities of the method for resolved images and the method for one resolved image and a blend. If the TEST-1 curves are used as input for the resolved algorithm, the delay $\Delta t_{1,2}$ can be recovered, but with remarkable error due to blending. The resolved algorithm reported $\Delta t_{1,2} \approx 59$ days. The resulting dispersion curve is depicted in Fig. 3.9. However, the blend was generated by using $\Delta t_{1,2} = 50.2$ days, $\Delta t_{2,3} = 10.6$ days and $a_3 = 1.3$.

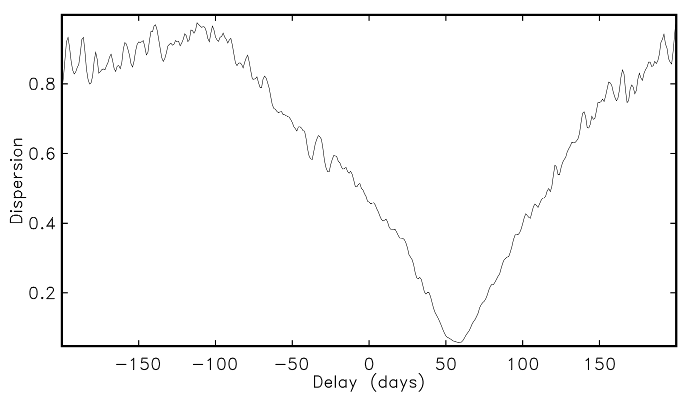

Figure 3.9: A dispersion spectrum computed for the TEST-1 simulated light curves using the resolved algorithm. It reveals a shift $\Delta t_{1,2} \approx 59$ days. However, the blend was generated by using $\Delta t_{1,2} = 50.2$ days and $\Delta t_{2,3} = 10.6$ days. Consequently, blending can mask proper time delay values. The fully resolved case is shown in Fig. 3.10.

To recover both delays, we have to apply our method with artificial blend. A three-parameter search grid $[-360, 360 : 1.0] \times [-90, 90 : 1.0] \times [0.6, 1.6 : 0.1]$[3] was used. The value for downweighting parameter ν was taken at 2.5 days. (See Section 2.2.4 for discussion on choosing the value of parameter ν.) Our method reported the general minimum of the merit function at $\Delta_l = 50$ days, $\Delta_s = 11$ days, and $\epsilon = 1.1$. The two-dimensional slice at $\epsilon = 1.1$ of the search grid is given in Fig. 3.10. In this plot we can clearly see the degenerate character of our procedure.

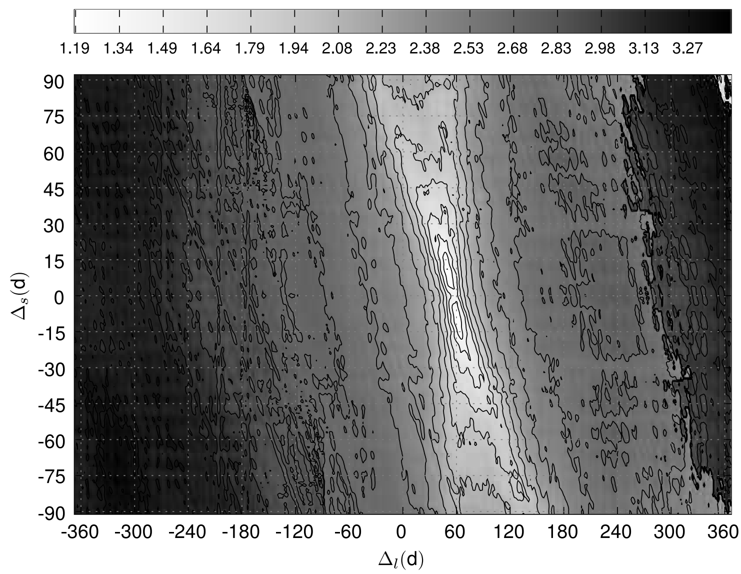

Figure 3.10: The two-dimensional grid of merit function values for a computer generated random walk and a blend computed from it (TEST-1). The general minimum must indicate the true pair of long and short delay values. The plot demonstrates degeneracy in full-scale computations well – there is obvious symmetry between the areas for positive and negative values of short delays.

We can see that our algorithm did not exactly recover the amplification ratio parameter (we found 1.1 instead of 1.3). This is quite typical – for every particular pair of delay values the merit function dependence on the parameter ϵ is quite weak and the corresponding curve has a wide minimum around the correct value.

For high quality data (with small errors) we can look for a final solution with higher precision. For each strong minima found during the rough analysis, we can build refined local parameter grids in the vicinities of the

[3]Here and below we use a systematic notation for search grids. Inside the square brackets we give the minimum and maximum values for the parameter in question, followed by the grid step.

preliminary solutions. An example of such local refinement is given in the next subsection.

3.3.2.2 Real sampling

As it was discussed in Section 3.1.1, the noise level of Schild's light curves of the QSO 0957+561 is tolerable for our algorithm for a clean image and a blend, and we can use their sampling and weight system for generating test data. So, the TEST-2 simulated light curves (see Section 3.2 for details) with the observational time points and weight system of the Schild's time series were used to test our method under real sampling conditions. The values of parameters used in TEST-2 simulation were: $\Delta t_{1,2} = 420.15$ days, $\Delta t_{2,3} = 20.21$ days, $a_3 = 0.8$. A crude search grid $[-150, 1000 : 1.0] \times [-200, 200 : 1.0] \times [0.6, 1.6 : 0.1]$ was used to estimate the amplification parameter a_3. The grid slice with the best trial value $\epsilon = 0.9 \pm 0.1$ was then used to refine other two parameters. (The value for the downweighting parameter ν was taken at 7 days here.) Finally we got the best estimates for the delays $\Delta_l = 420.2 \pm 0.8$ days, $\Delta_s = 20.0 \pm 1.2$ days. The plot of merit function values in the slice $[-150, 1000 : 0.1] \times [-200, 200 : 0.1]$ is shown in Fig. 3.11. The error bars of the result were computed using Monte Carlo type calculations as it was described in Section 2.5.

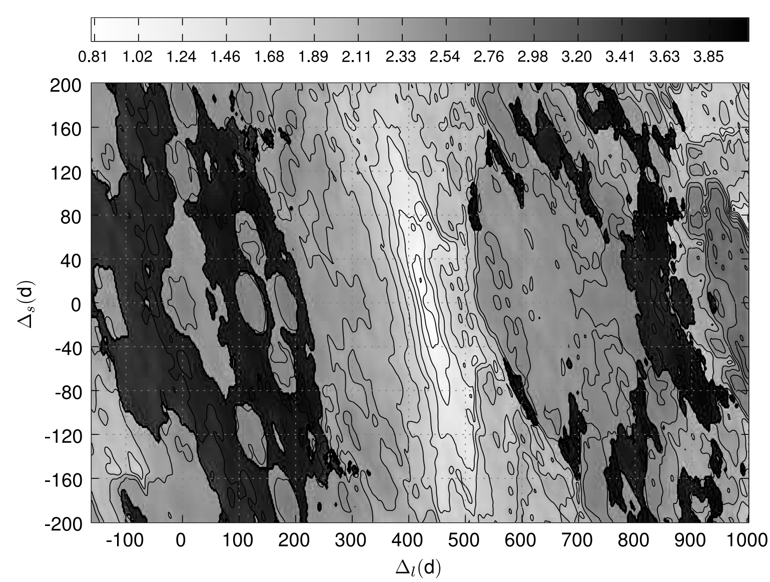

Figure 3.11: Merit function values for simulated data and real sampling (TEST-2).

As the results recover the initial parameters well and we can see the two symmetrically placed minima, we can say that our method works correctly under the real sampling conditions.

3.3.2.3 Results for a real system

To evaluate the new method in a more realistic context, we used the master data set for the double quasar QSO 0957+561 A,B kindly provided by Rudy Schild (R-band optical CCD photometry; see Table 3.2 for details). As far as we know, the components of the system itself cannot be considered as blends and consequently we used this data set as a model for time point spacing and observational error distribution for a long and realistic monitoring programme, as was discussed in Section 3.3.2.2. In the course of experimentation we also performed some calculations with the full Schild's data set and got unexpected results. Assuming that B is a blend, we indeed got a distribution characteristic to the blended case, shown in Fig. 3.12. We found the estimates for the time delays in real data at $\Delta_l = 412, \Delta_s = 22$ days. In the current section, the value for the downweighting parameter ν was taken 10 days and the amplification factor was held fixed at $\epsilon \equiv 1.0$. ($a_3/a_2 \approx 1.0$ is the expected value for short $\Delta t_{2,3}$).

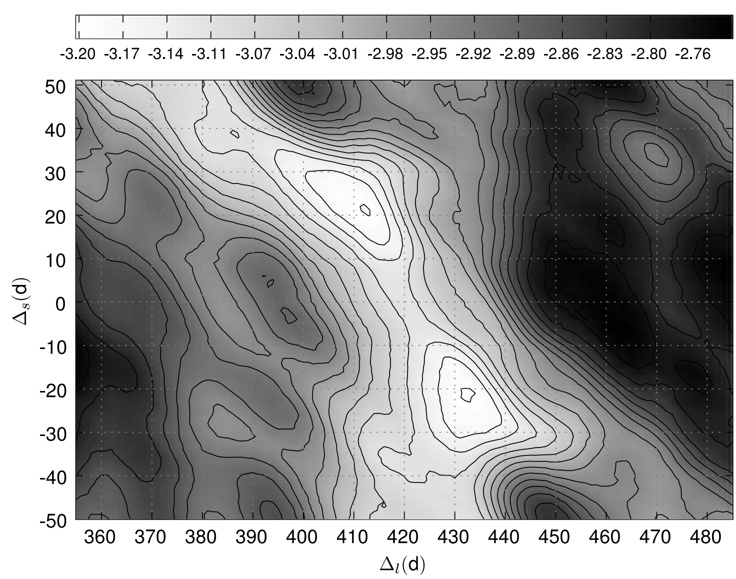

Figure 3.12: Merit function values for the actual double quasar QSO 0957+561 data. It differs significantly from Fig. 3.13.

To convince ourselves that the symmetric minima in Fig. 3.12 are not caused by boundary effects of our computational algorithm, we performed some additional tests. First we used the real time moments and error estimates from the Schild's data set and built a pair of artificial curves TEST-3 (see Section 3.2 for details) with a given single time delay between the two curves $\Delta t_{1,2} = 412$ days. The resulting two-dimensional slice of the merit function is shown in Fig. 3.13. It is well seen that there is one unique global minimum near the true delay value, indicating that we do not have a blend here; the delay applied is recovered and the estimated short shift value (if we assume that the B curve is a blend) is zero. Consequently, our method does not generate symmetrically placed minima just as an artifact of the procedure.

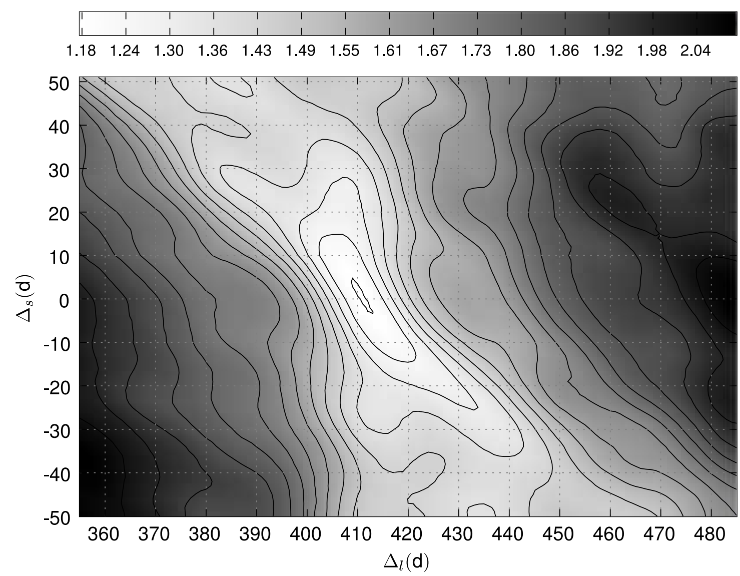

Figure 3.13: Merit function values for a random walk and its shifted version – TEST-3 ($\Delta t_{1,2} = 412$ days). Time points and standard errors are from Rudy Schild's monitoring programme.

Finally we built an artificial blended model TEST-4 with the long and short delays found from the real curves A and B. The resulting grid is shown in Fig. 3.14. From the last simulation we found the time delay estimates $\Delta_l = 412$, $\Delta_s = 25$ days, indicating that the time delay values found from the real data are real. The three-day long estimation error in short shift characterises the precision of the algorithm at the given level of observational accuracy.

To check how statistically stable the merit function values are for dif-

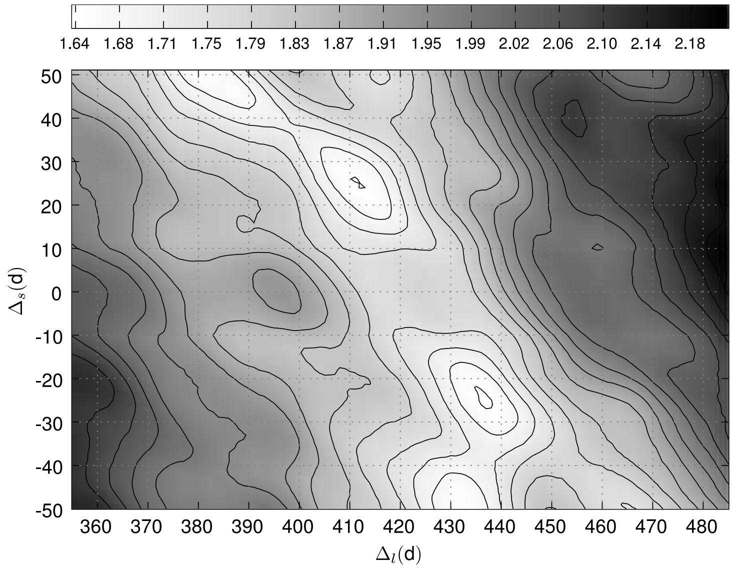

Figure 3.14: Merit function values for a computer generated random walk and the blend computed from it using the parameters found from the real observational data – TEST-4.

ferent parameter combinations, we calculated sums of weights for the time delay grid of Schild's data. As it is seen from Fig. 3.15, $\Delta_l = 412$ and $\Delta_s = 22$ days fall into the region of higher weights and should be considered as a reliable result. It is currently very difficult to tell why the B curve of the classical double quasar behaves as a blend. It is known that there is something wrong with the estimated time delays and magnification ratios (if optical data is compared with radio data). The peculiar form of microlensing proposed in Press & Rybicki (1998) can solve the problem of magnification ratios. However, it is hard to expect that the spacing of microlensing events in time can mimic a proper blend. In another development, Goicoechea (2002) singles out the different features in the double quasar light curves which give different values for time delays. As a possible explanation he uses a quasar model with spatially distant flares, as discussed also in Yonehara (1999) and Yonehara et al. (2003). Similar and even more radical ideas can be found in Schild (2005). Our computations show that not only single events, but the full B curve of the system can be decomposed into a sum of two similar and shifted curves. What theoretical interpretation can be given to this phenomenon remains an open question.

Note, we also got similar results when the A curve was assumed to be a blend. However, applying the method for two blends was not successful due to insufficient quality of the data.

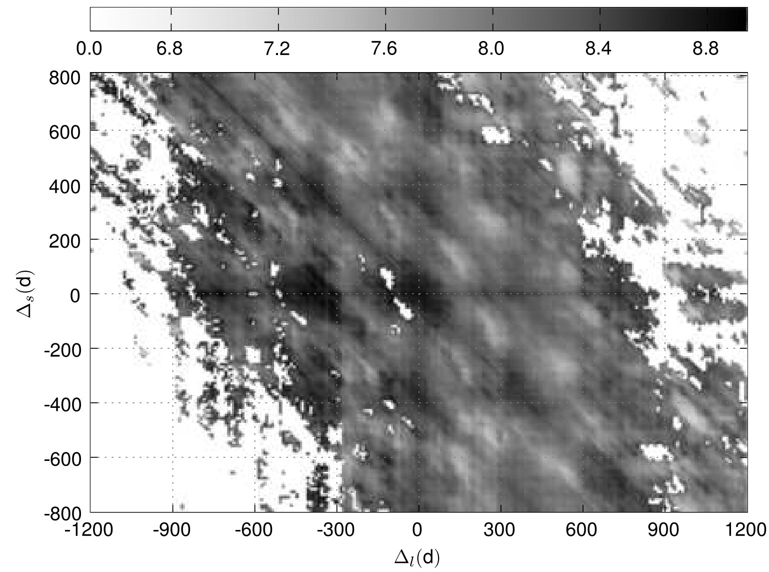

Figure 3.15: The sums of weights for the time delay space of Schild's data. Zero values represent areas of "mirror matches" with negative parameter m values. (See Section 2.2.1 for the discussion about m.)

3.3.3 Two blends of two subimages

First we analysed the TEST-5 simulated data set which is described in Section 3.2 and depicted in Fig. 3.4. The time delays used in this simulation were $\Delta c = 420.2$, $\Delta a = 20.2$, $\Delta b = 56.5$ days. From the experiments with different noise levels for the given sampling we found that we should keep the $S/N \geq 30$ for the method to work properly. So, the Gaussian noise with standard deviation of 3% of the amplitude of light curves is acceptable for our method for the TEST-5 sampling. For the TEST-5 (as well as for TEST-6 and TEST-7) data the optimal downweighting parameter was $\nu = 1.5$. We performed a three-dimensional search for time delays, using the one day step size and the following limits for trial delays: $\delta c = 370...470$, $\delta a = -30...70$, $\delta b = 6...106$ days. We found a global minimum of the DS at $\delta c = 417$, $\delta a = 19$, $\delta b = 59$ days.

Our method does not work, if $|\Delta a| = |\Delta b|$. However, having well sampled data and low noise, it is still possible to get a solution for very close short delays. For example, we made a noiseless simulation TEST-6 (see Section 3.2 for details) where $\Delta c = 420.2$, $\Delta a = 56.5$, $\Delta b = 50.1$ days. We

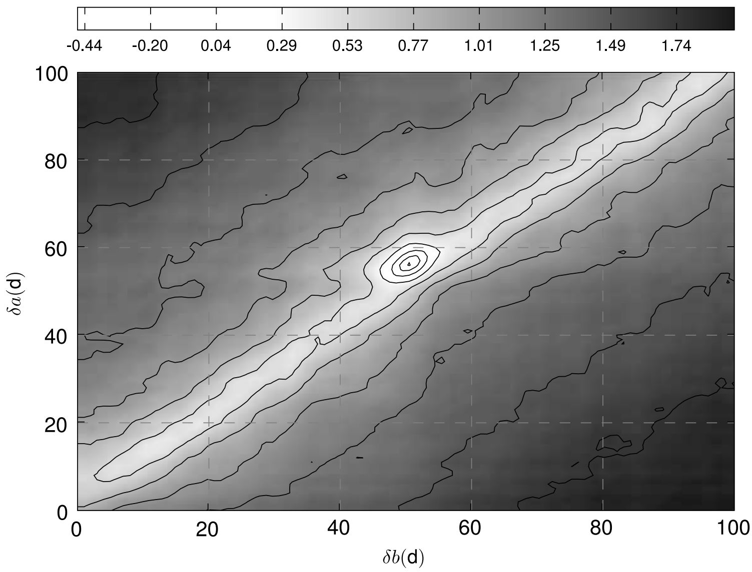

Figure 3.16: DS values for very close short delays. $\Delta c = 420.2$, $\Delta a = 56.5$, $\Delta b = 50.1$ days (TEST-6).

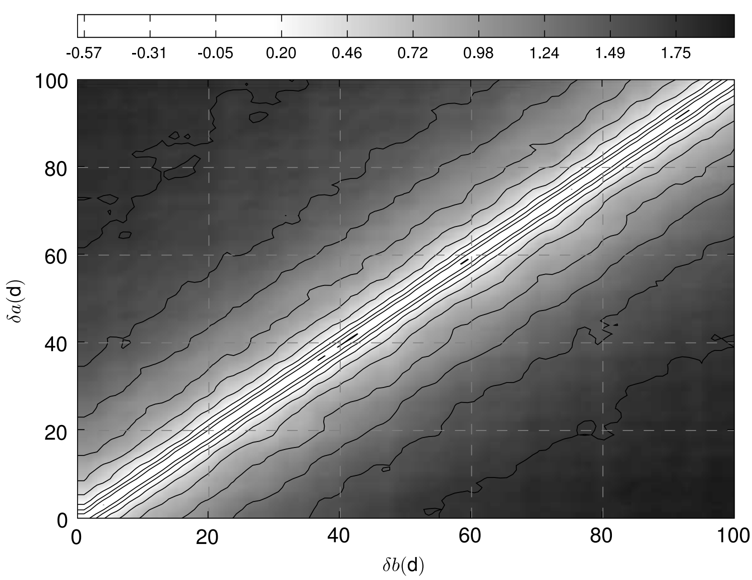

Figure 3.17: DS values for $\Delta a = \Delta b = 20.2$, $\Delta c = 420.2$ days (TEST-7).

fixed the long trial delay at foreknown value $\delta c \equiv 420.2$ days when the data was analysed. Then the given time delays were recovered correctly. The resulting plot of DS values is shown in Fig. 3.16. Even a one day difference between Δa and Δb in a noiseless case is still tolerable for the method, but then the minimum on the DS surface is not very convincing indeed.

The singular situation of TEST-7 where initial $\Delta c = 420.2$, $\Delta a = 20.2$, $\Delta b = 20.2$ days is shown in Fig. 3.17. We can see a characteristic distribution along straight line of DS values and no minima. To get the characteristic picture depicted in Fig. 3.17 the long trial delay was fixed again at foreknown value $\delta c \equiv 420.2$ days when analysing the TEST-7 data set. Without fixing the δc, we may also hit an arbitrary solution corresponding to mirrored arbitrary short delays and get nice but spurious pictures of DS minima as it was discussed in Section 2.3.

Chapter 4

Conclusions

To turn the gravitational lensing into a precise tool for measuring H_0, we need to overcome the degeneracies and uncertainties associated with the modelling of the mass profiles of the lenses. Using a large number of lenses with measured time delays and simultaneous non-parametric modelling of the mass profiles with shared H_0 is a promising approach (see Paraficz & Hjorth 2010). Knowing the value of H_0 we can use the time delay gravitational lenses to measure the masses of galaxies and establish the corresponding mass distribution in the Universe. We can match the light curves of a given source by the estimated time delay and magnification difference to get a combined time series with better sampling, which in turn, can be used in studies of the source. The time delays, found between light curves of nearby images in the data of large photometry programs, can be used to discover new lens systems. For these purposes we need a time delay estimation method that has as few as possible user set parameters and performs well in real situations. *As a result of our study, such an automatic method was developed and tested with different sets of observed data. We have also removed many problems from methodology, which could lead to wrong time delay estimations (fitting data into gaps, ignoring of extrinsic variation etc).*

As one important result, we can conclude that the original algorithm by PRH gave wrong result with the Vanderriest et al. (1989) data because of two reasons – data fitting into the gaps due to the use of global χ^2 *matching criterion and also due to the leaving off possibility of microlensing.*

The proposed method for fully automatic time delay estimation is actually a combination of good sides of different previously well known approaches. First we use cross-interpolation scheme which was introduced in

Gaskell & Sparke (1986) and Gaskell & Peterson (1987). Then we compute actual interpolated values using linear prediction scheme introduced in PRH. We added to this scheme only minor improvements - rules for excluding certain bins of the structure function, computing the variance level, as well as using the estimated variance in building the model of the structure function. The use of pointwise $\chi^2(\Delta t^*)$ matching criteria is ubiquitous, but not always with correct normalisation. And finally, the trend component fitting into the differences is implementation of ideas from Pelt et al. (1996). In this way the step undertaken is relatively small. However, we were somewhat amazed how persistently the combined method landed at or very near to the already established delay values. We can say, that the method works automatically and gives correct results for light curves that are long enough, have sufficiently low noise level and are sampled according to the requirements of the given time delay estimation task.

As longer photometric time series can be often obtained using telescopes with modest apertures at not very good seeing conditions, *two special algorithms were built, which allow us to estimate integral time-delay systems for blended (not fully resolved) light curves.* The two new methods were developed before the combined one and then the automatic use of them was not considered; also the possible microlensing effect was ignored. Hence they are based on computing of the standard dispersion spectrum, introduced in Pelt et al. (1994) and refined in Pelt et al. (1996). These methods were tested to work with simulated data, as we do not have long enough and well sampled blended data sets yet. However, applying the blended algorithms allow us to get interesting results, which would have been unnoticed otherwise. Using the method for a resolved image and a blend of two subimages, we showed that the light curve of the B image of the QSO 0957+561 can be considered as a blend.

The automatic method can be modified for use with the data of one resolved image and a blend of two subimages. The treatment of the microlensing can be included into both blended methods.

The developed software modules can be also used to plan new observations. We can model situations that may occur in real long-time monitoring programmes. By varying model parameters we can estimate sufficient sampling and durations for observational sessions as well as the accuracy of observation needed. On the other hand, even accurately planned sessions can result in a failure because the source quasars themselves can show persistent stationarity or the time series observed can be contaminated by strong microlensing which is unacceptable for the present formulation of the blended algorithms.

We believe that the automatic method for estimating time delays between resolved images can be used to analyse data which will flow out from the extensive photometric programs planned. It is also worth considering to apply blended algorithms in processing this data. We hope that the availability of the new methods for blended images gives an extra motivation for astronomers observing at telescopes with modest resolutions to carry out long monitoring programmes for gravitational lens systems.

There are also some problems left to be solved in subsequent studies. First, we should establish more robust constraints of the significance of the estimated (or discovered) time delays. Getting a stable time delay with different polynomial degrees within the detection range of our method may not be sufficient to suppress false alarms if the real time delay is much longer, or there is no time delay at all. Using Monte Carlo type calculations may be considered here. Second, it is worth implementing the singular value decomposition method for possible cases of ill conditioned covariance matrices. Third, the small controversy of the time delay of the QSO 0957+561 ($412 \ldots 417$ against $422 \ldots 426$ days) needs to be explained.

Acknowledgements

I am indebted to my supervisor Jaan Pelt for initiating, motivating and guiding the study. I am grateful for his ideas on developing the time delay estimation methods and for help on publishing our results. I am also thankful for his advices on improving my manuscript.

I wish to thank my coauthors and colleagues at the Tartu Observatory Tõnis Eenmäe, Tiina Liimets, Lauri Juhan Liivamägi, Nigul Olspert and also my colleague Tõnu Viik for their help on writing the papers. My gratitude goes to Jaan Pelt and to my colleagues Tõnis Eenmäe, Lauri Juhan Liivamägi and Enn Saar, who helped me to overcome many computational problems. I am grateful to my colleague Peeter Tenjes for patiently helping me to manage my Ph.D. studies at the University of Tartu. I appreciate the help of my colleagues Anna Aret, Antti Tamm and Elmo Tempel on drawing up my manuscript.

Many thanks go to my colleague Kalju Annuk for supervision during my first years at the Tartu Observatory, and also for managing the computational resources I have used.

I wish to thank all the staff of the Tartu Observatory for providing good atmosphere to work in. Special thanks go to my friends for unforgettable hikes.

My gratitude goes to my parents (Asta Hirv and Leon Hirv) and to my teachers Johannes Mangus and Arvo Kase and also to my colleagues Enno Ruusalepp, Peeter Tenjes and Uno Veismann for bringing up my interest in nature and universe.

I thank my parents, Aime Lobja, Lembit Lobja, Mare Kurs, Hugo Kurs, Helju Padjus, Jaan Padjus and Ulve Tammerik for kindly supporting me during my studies at the University of Tartu.

I acknowledge the financial support by the Estonian Science Foundation (the grants Nos. 6810 and 6813).

Finally, my greatest gratitude goes to Margit for her loving support, inspiration and love.

References

Barkana, R. 1997, ApJ, 489, 21

Bozza, V. 2010, General Relativity and Gravitation, 79

Burud, I., Magain, P., Sohy, S., & Hjorth, J. 2001, A&A, 380, 805

Chwolson, O. 1924, Astronomische Nachrichten, 221, 329

Coe, D. & Moustakas, L. A. 2009, ApJ, 706, 45

Coles, P. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 252, Historical Development of Modern Cosmology, ed. V. J. Martínez, V. Trimble, & M. J. Pons-Bordería , 21–+

Colley, W. N., Schild, R. E., Abajas, C., et al. 2003, ApJ, 587, 71

Cooke, J. H. & Kantowski, R. 1975, ApJ, 195, L11+

Courbin, F., Chantry, V., Revaz, Y., et al. 2010, e-print arXiv:1009.1473

Courbin, F. & Minniti, D., eds. 2002, Lecture Notes in Physics, Berlin Springer Verlag, Vol. 608, Gravitational Lensing: An Astrophysical Tool

Cuevas-Tello, J. C., Tiňo, P., & Raychaudhury, S. 2006, A&A, 454, 695

Dyson, F. W., Eddington, A. S., & Davidson, C. 1920, Royal Society of London Philosophical Transactions Series A, 220, 291

Eddington, A. S. 1920, Space, time and gravitation. an outline of the general relativity theory (Cambridge University Press)

Edelson, R. A. & Krolik, J. H. 1988, ApJ, 333, 646

Einstein, A. 1911, Ann. Phys. (Leipzig), 35, 898

Einstein, A. 1936, Science, 84, 506

Emmanoulopoulos, D., McHardy, I. M., & Uttley, P. 2010, MNRAS, 404, 931

Fohlmeister, J., Kochanek, C. S., Falco, E. E., Morgan, C. W., & Wambsganss, J. 2008, ApJ, 676, 761

Gaskell, C. M. & Peterson, B. M. 1987, ApJS, 65, 1

Gaskell, C. M. & Sparke, L. S. 1986, ApJ, 305, 175

Geiger, B. & Schneider, P. 1996, MNRAS, 282, 530

Gil-Merino, R., Wisotzki, L., & Wambsganss, J. 2002, A&A, 381, 428

Goicoechea, L. J. 2002, MNRAS, 334, 905

Haarsma, D. B., Hewitt, J. N., Lehar, J., & Burke, B. F. 1997, ApJ, 479, 102

Hirv, A., Eenmäe, T., Liimets, T., Liivamägi, L. J., & Pelt, J. 2007a, A&A, 464, 471

Hirv, A., Eenmäe, T., Liivamägi, L. J., & Pelt, J. 2007b, Baltic Astronomy, 16, 241

Hirv, A., Olspert, N., & Pelt, J. 2011, e-print arXiv:1105.5991

Hjorth, P. G., Villemoes, L. F., Teuber, J., & Florentin-Nielsen, R. 1992, A&A, 255, L20

Hovatta, T., Tornikoski, M., Lainela, M., et al. 2007, A&A, 469, 899

Hughes, P. A., Aller, H. D., & Aller, M. F. 1992, ApJ, 396, 469

Kayser, R. & Refsdal, S. 1983, A&A, 128, 156

Kochanek, C. S., Morgan, N. D., Falco, E. E., et al. 2006, ApJ, 640, 47

Kundic, T., Colley, W. N., Gott, III, J. R., et al. 1995, ApJ, 455, L5+

Kundic, T., Turner, E. L., Colley, W. N., et al. 1997, ApJ, 482, 75

Lehar, J., Hewitt, J. N., Burke, B. F., & Roberts, D. H. 1992, ApJ, 384, 453

LSST Science Collaborations, Abell, P. A., Allison, J., et al. 2009, e-print arXiv:0912.0201

Ofek, E. O. & Maoz, D. 2003, ApJ, 594, 101

Oguri, M. & Marshall, P. J. 2010, MNRAS, 405, 2579

Oscoz, A., Alcalde, D., Serra-Ricart, M., et al. 2001, ApJ, 552, 81

Ovaldsen, J. E., Teuber, J., Schild, R. E., & Stabell, R. 2003, A&A, 402, 891

Paraficz, D. & Hjorth, J. 2010, ApJ, 712, 1378

Pelt, J., Hjorth, J., Refsdal, S., Schild, R., & Stabell, R. 1998a, A&A, 337, 681

Pelt, J., Hoff, W., Kayser, R., Refsdal, S., & Schramm, T. 1994, A&A, 286, 775

Pelt, J., Kayser, R., Refsdal, S., & Schramm, T. 1996, A&A, 305, 97

Pelt, J., Schild, R., Refsdal, S., & Stabell, R. 1998b, A&A, 336, 829

Pijpers, F. P. 1997, MNRAS, 289, 933

Poindexter, S., Morgan, N., Kochanek, C. S., & Falco, E. E. 2007, ApJ, 660, 146

Press, W. H. & Rybicki, G. B. 1997, in Astrophysics and Space Science Library, Vol. 218, Astronomical Time Series, ed. D. Maoz, A. Sternberg, & E. M. Leibowitz, 61–+

Press, W. H. & Rybicki, G. B. 1998, ApJ, 507, 108

Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992a, ApJ, 385, 404

Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992b, ApJ, 385, 416

Refsdal, S. 1964a, MNRAS, 128, 307

Refsdal, S. 1964b, MNRAS, 128, 295

Refsdal, S. 1966, MNRAS, 132, 101

Refsdal, S. & Surdej, J. 1994, Reports on Progress in Physics, 57, 117

Renn, J., Sauer, T., & Stachel, J. 1997, Science, 275, 184

Rybicki, G. B. & Kleyna, J. T. 1994, in Astronomical Society of the Pacific Conference Series, Vol. 69, Reverberation Mapping of the Broad-Line Region in Active Galactic Nuclei, ed. P. M. Gondhalekar, K. Horne, & B. M. Peterson, 85–+

Rybicki, G. B. & Press, W. H. 1992, ApJ, 398, 169

Schild, R. E. 2005, AJ, 129, 1225

Schmidt, M. 1963, Nature, 197, 1040

Schneider, P. 1985, A&A, 143, 413

Schneider, P. 1995, e-print arXiv:astro-ph/9512047

Schneider, P., Kochanek, C. S., & Wambsganss, J. 2006, Gravitational Lensing: Strong, Weak and Micro (Springer-Verlag)

Shalyapin, V. N., Goicoechea, L. J., Koptelova, E., Ullán, A., & Gil-Merino, R. 2008, A&A, 492, 401

Soldner, J. G. 1804, Berliner Astronomisches Jahrbuch, 161

Treu, T. 2010, ARA&A, 48, 87

Vanderriest, C., Schneider, J., Herpe, G., et al. 1989, A&A, 215, 1

Virbhadra, K. S. & Ellis, G. F. R. 2000, Phys. Rev. D, 62, 084003

Virbhadra, K. S. & Keeton, C. R. 2008, Phys. Rev. D, 77, 124014

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Wambsganss, J. 1998, Living Reviews in Relativity, 1, 12. URL (cited on June 17, 2011): http://www.livingreviews.org/lrr-1998-12

Wyrzykowski, L., Udalski, A., Schechter, P. L., et al. 2003, Acta Astronomica, 53, 229

Yonehara, A. 1999, ApJ, 519, L31

Yonehara, A., Mineshige, S., Takei, Y., Chartas, G., & Turner, E. L. 2003, ApJ, 594, 107

Zwicky, F. 1937a, Physical Review, 51, 290

Zwicky, F. 1937b, Physical Review, 51, 679

Summary in Estonian

Gravitatsiooniläätsest põhjustatud kvasarite mitmikkujutiste vaheliste ajanihete mõõtmine

Kauge kvasari ja maise vaatleja vahelise vaatekiire lähedale jäävad massiivsed objektid toimivad gravitatsiooniläätsena. Käesolevas töös käsitletakse nn tugevaid gravitatsiooniläätsi, kus tekivad mitmikkujutised. Tugevas läätses on valguskiiri kallutavaks massiivseks objektiks enamasti (hiid)galaktika. Valgus (või muu elektromagnetkiirgus) saab läätse tõttu tulla vaatlejani eri teid pidi. Eri teid mööda kulgemine võtab kvasari signaalil erinevalt aega; ka Shapiro efektist tingitud viivitus võib eri kujutistel olla erinev. Kujutiste heleduskõverate vahel on seega ajanihked, mida me saame mõõta.

Milleks mõõta ajanihkeid? Sjur Refsdal (vaata Refsdal 1964b,a) näitas, et me saame määrata Hubble parameetrit H_0, teades allika ja läätse punanihkeid, kujutiste heledusi ja nurkkaugusi, läätse massijaotust ning mõõdetud ajanihet. (Suuremate punanihete korral peame lisaks arvesse võtma kosmoloogiliste kõverus- ja tihedusparameetrite väärtused.) Seejuures on oluline, et H_0 hindamiseks ei pea me tarvitama nn kosmilist redelit ja pääseme sellega seotud ebatäpsustest. Tegelikkus pole siiski nii lihtne – osutub, et vaatlusandmetega saavad klappida erinevad läätse massijaotuse mudelid, mis annavad erinevad H_0 väärtused. Kauge galaktika massijaotuse ja vaatekiire lähedusse jääva täiendava massi paigutuse ühene vaatluslik mõõtmine pole aga enamasti piisava täpsusega võimalik.

Läätse massijaotuse modelleerimisega seotud määramatust H_0 hinnangus saab maha suruda, kui kasutame suurt arvu mõõdetud ajanihetega läätsesüsteeme (vaata Coe & Moustakas 2009; Oguri & Marshall 2010; Paraficz & Hjorth 2010). Teades H_0 väärtust, saab omakorda määrata läätsgalaktikate masse ja galaktikate masside jaotust. Ka see eeldab suure hulga ajanihete mõõtmist. Me peame hindama suurt arvu ajanihkeid ka siis, kui tahame nende abil avastada uusi gravitatsiooniläätsi massilise fotomeetria andmetest. Ajas nihutatud heleduskõveraid saab mõõdetud ajanihke abil kokku nihutada. Sedasi saame allikast parema vaatlusrea kui vaid üksikuid kujutisi uurides.

Suure arvu ajanihete usaldusväärseks mõõtmiseks on vaja meetodit, mis ei nõua kasutaja poolset sekkumist ja subjektiivsete parameetrite sisestamist. Üks käesoleva töö eesmärk oligi automaatse ajanihete leidmise metoodika välja töötamine ja katsetamine reaalsete vaatlusandmetega. Töö käigus koostati algoritm (vaata Hirv et al. 2011), mis vähemalt kasutatud vaatlusandmete korral töötas automaatselt ja leidis teiste autorite tulemustega kooskõlalised ajanihete väärtused. Meetod pandi kokku juba kaua kasutusel olnud ideedest, ühendades tuntud algoritmide head küljed. Esiteks rakendame heleduskõveratesse punktide ristinterpoleerimise skeemi, mida tarvitasid Gaskell & Sparke (1986) ja Gaskell & Peterson (1987). Interpoleeritud väärtuste arvutamiseks kasutame lineaarse ennustamise metoodikat, mille võtsid kasutusele Press et al. (1992a). Viimati mainitud metoodikas täiustasime struktuurifunktsiooni koostamist, dispersioonitaseme hindamist. Vaatlusandmetega sobivaima ajanihke leidmiseks tarvitame nn kombineeritud dispersioonispektrit, mis on sisuliselt laialt levinud punktipaariviisilise $\chi^2(\Delta t^*)$ korrektsema normeeringuga analoog. Mikroläätseefekti mõju kõrvaldamiseks rakendame polünoomi sobitamist heleduskõverate vahesse, mida kasutasid Pelt et al. (1996).

Kuigi tehtud samm on selles valguses suhteliselt väike, oleme me üllatunud, kui kindlalt uus meetod andis juba tunnustatud ajanihete väärtustega väga lähedasi või identseid vastuseid. Usume, et meetod töötab korrektselt ka teiste piisavalt pikkade aegridadega, mille müratase ja vaatluspunktide tihedus vastavad ülesande püstitusele.

Näitasime, et Press et al. (1992a) said QSO 0957+561 jaoks vale ajanihke väärtuse kahel põhjusel: nende kasutatud globaalne χ^2-statistik ei ole stabiilne ligikaudu perioodiliste vaatluspauside suhtes; nad jätsid arvestamata mikroläätseefekti.

Kui gravitatsiooniläätsena toimib (hiid)galaktika, on tekkinud kujutiste nurkkaugused vaid mõne kaaresekundi suurusjärgus. Kasutades vaatlusteks väiksemaid teleskoope halvemates vaatlustingimustes, võivad vähem eraldatud kujutised jääda lahutamata ehk kokku sulada. Koostasime kaks meetodit kokku sulanud kujutiste heleduskõveratest ajanihete mõõtmiseks. Esiteks, kui vaatlejani jõuab kvasari signaal ühe lahutatud kujutise ja ühe kahest kujutisest kokku sulanud kujutise kaudu (vaata Hirv et al. 2007a). Teiseks, kui vaatleja registreerib kaks kokku sulanud kujutist, mis kumbki koosnevad kahest kujutisest (vaata Hirv et al. 2007b). Algoritmid testiti simuleeritud andmetel, sest piisavalt kvaliteetseid liitunud kujutistega aegridu ei olnud veel saada. Kuna mitte lahutatud kujutistele mõeldud meetodid koostati enne automaatset, siis on nendes kasutatud standardset dispersioonispektrit, mille võtsid kasutusele Pelt et al. (1994, 1996). Ka mikroläätseefekt jäetakse nende praeguses formuleeringus arvestamata. Samas võime neid rakendades saada huvitavaid tulemusi – esimese meetodi abil leidsime, et QSO 0957+561 B kujutist võib vaadelda kui kahe kokku sulanud kujutise summat. Märgime, et automaatset meetodit saab kohandada ühe lahutatud ja ühe kokku sulanud kujutise juhtumi tarbeks. Kahe kokku sulanud kujutise korral pole see korrektne. Mikroläätseefekti saab põhimõtteliselt arvesse võtta mõlemal juhul.

Koostatud metoodikat saab kasutada ka uute vaatluste planeerimiseks. Me saame simuleerida reaalseid vaatlusandmeid ning hinnata vajalikku vaatlustäpsust, aegrea pikkust ja vaatluspunktide tihedust oodatavate ajanihete piisava täpsusega mõõtmiseks. Usume, et meie automaatset ajanihete leidmise meetodit saab rakendada planeeritavatest suurtest fotomeetriaprojektidest tulevate andmete töötlemiseks. Loodame, et kokku sulanud kujutiste jaoks loodud meetodid julgustavad vaatlejaid tegema pikki vaatlusridu teleskoopidega, mis muidu jääks gravitatsiooniläätsede kujutiste vaatlemiseks sobimatutesse vaatlustingimustesse. Ka kvaliteetsematele massilise fotomeetria aegridadele võib rakendada kokku sulanud kujutistele loodud algoritme, et avastada võimalikke lahutamata jäänud kujutisi.

Edasise töö käigus on plaanis tegeleda järgmiste probleemidega. 1. Meetod peaks hindama leitud ajanihete statistilist olulisust. 2. Võimalike raskesti pööratavate kovariatsioonimaatriksite tarbeks tuleb rakendada töökindlamat algoritmi. 3. Nn väike vastuolu QSO 0957+561 erinevate autorite poolt saadud ajanihete väärtuste vahel vajab lahendamist.

Attached original publications

Curriculum vitae

Name:	Anti Hirv
Date and place of birth:	April 7, 1979, Kamari village, Jõgevamaa, Estonia
Citizenship:	Estonian
Marital status:	Married with Margit Hirv
Current position:	Tartu Observatory (research associate)
Address:	Tartu Observatory 61602 Tõravere Tartumaa, Estonia
E-mail:	anti@aai.ee

Education:

The Elementary School of Kamari, 1989
The Secondary School of Põltsamaa (silver medal), 1997
University of Tartu, B.Sc. in physics, 2001
University of Tartu, M.Sc. in astrophysics, 2003

Professional training:

11.08.–25.08. 2002	Summer school "Nordic – Baltic Research School 2002: Astrophysics of Interacting Stars", Molėtai, Lithuania
03.09.–07.09. 2007	Summer school "The Finnish Graduate School in Astronomy and Space Physics 2007: Time Series Analysis", Elva, Estonia
19.05.–21.05. 2008	Summer school "Scientific Writing for Young Astronomers 2008", Blankenberge, Belgium

Conference presentations:

15.08.–19.08. 2005	Conference "Stellar Evolution at Low Metallicity: Mass Loss, Explosions, Cosmology", Tartu, Estonia. *Poster presentation*: Nugis T., Annuk K., Hirv A., "Observational evidences for the dependence of WNE-star mass loss on metallicity"

Research and professional experience:

2003–2009	University of Tartu, PhD student
2007–...	Tartu Observatory, research associate

Research grants and stipends:

2002	E. J. Öpik stipend (Tartu Observatory)

Field of research:

Observational investigation of luminous highly evolved stars; theory and applications of dispersion spectra

Publications:

1. Hirv A., Olspert N., Pelt J., 2011. Towards the automatic estimation of gravitational lenses' time delays. Baltic Astronomy, accepted for publication on 23 May 2011, e-print arXiv:1105.5991

2. Nugis T., Annuk K., Hirv A., Niedzielski A., Czart K., 2008. Near Infrared Spectra of Galactic Wolf-Rayet Stars. Baltic Astronomy, 17, 39-49.

3. Hirv A., Eenmäe T., Liimets T., Liivamägi L. J., Pelt J., 2007. Estimation of time delays from unresolved photometry. Astronomy and Astrophysics, 464, 471-478.

4. Hirv A., Eenmäe T., Liivamägi L. J., Pelt J., 2007. Estimation of Time Delays from Two Blended Light Curves of Gravitational Lenses. Baltic Astronomy, 16, 241-250.

5. Nugis T., Annuk K., Hirv A., 2007. Mass Loss Parameters of WNE Stars: Dependence on Metallicity. Baltic Astronomy, 16, 227-239.

6. Nugis T., Annuk K., Hirv A., 2006. Observational evidences for the dependence of WNE-star mass loss on metallicity. Stellar Evolution at Low Metallicity: Mass Loss, Explosions, Cosmology, 353, 203.

7. Hirv A., Annuk K., Eenmäe T., Liimets T., Pelt J., Puss A., Tempel M., 2006. Orbital Elements and Mass-Loss Rate of V 444 Cyg. Baltic Astronomy, 15, 405-412.

8. Kolka I., Eenmäe T., Hirv A., Tuvikene T., Kama M., 2005. Emission Line Stars in the Framework of Gaia. The Three-Dimensional Universe with Gaia, 576, 543.

9. Kipper T., Klochkova V. G., Annuk K., Hirv A., Kolka I., Leedjärv L., Puss A., Skoda P., Slechta M., 2004. The peculiar variable V838 Monocerotis. Astronomy and Astrophysics, 416, 1107-1115.

10. Dubkova D. N., Kudryavtseva N. A., Hirv A., 2003. The Light Curve of the New Cataclysmic Variable SDSS J015543.40 +002807.2. Information Bulletin on Variable Stars, 5389, 1.

Elulugu

Nimi:	Anti Hirv
Sünniaeg ja koht:	7. aprill 1979, Kamari küla, Jõgevamaa, Eesti
Kodakondsus:	Eesti
Perekonnaseis:	Abielus Margit Hirvega
Praegune töökoht:	Tartu Observatoorium (teadur)
Aadress:	Tartu Observatoorium 61602 Tõravere, Tartumaa
E-post:	anti@aai.ee

Haridus:

Kamari Algkool, 1989
Põltsamaa Ühisgümnaasium (hõbemedal), 1997
Tartu Ülikool, B.Sc. füüsika erialal, 2001
Tartu Ülikool, M.Sc. astrofüüsika erialal, 2003

Erialane enesetäiendus:

11.08.–25.08. 2002	Suvekool "Nordic – Baltic Research School 2002: Astrophysics of Interacting Stars", Molėtai, Leedu
03.09.–07.09. 2007	Suvekool "The Finnish Graduate School in Astronomy and Space Physics 2007: Time Series Analysis", Elva, Eesti
19.05.–21.05. 2008	Suvekool "Scientific Writing for Young Astronomers 2008", Blankenberge, Belgia

Konverentside ettekanded:

15.08.–19.08. 2005	Konverents "Stellar Evolution at Low Metallicity: Mass Loss, Explosions, Cosmology", Tartu, Eesti. *Posterettekanne*: Nugis T., Annuk K., Hirv A., "Observational evidences for the dependence of WNE-star mass loss on metallicity"

Teenistuskäik:

2003–2009	Tartu Ülikool, doktorant
2007–...	Tartu Observatoorium, teadur

Uurimistoetused ja stipendiumid:

2002	E. J. Öpiku nim. stipendium (Tartu Observatoorium)

Teadustöö põhisuunad:

Heledate kaugele evolutsioneerunud tähtede vaatluslik uurimine; dispersioonispektrite teooria ja rakendused

DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS

1. **Tõnu Viik.** Numerical realizations of analytical methods in theory of radiative transfer. Tartu, 1991.
2. **Enn Saar.** Geometry of the large scale structure of the Universe. Tartu, 1991.
3. **Maret Einasto.** Morphological and luminosity segregation of galaxies. Tartu, 1991.
4. **Urmas Haud.** Dark Matter in galaxies. Tartu, 1991.
5. **Eugene A. Ustinov.** Inverse problems of radiative transfer in sounding of planetary atmospheres. Tartu, 1992.
6. **Peeter Tenjes.** Models of regular galaxies. Tartu, 1993.
7. **Ivar Suisalu.** Simulation of the evolution of large scale structure elements with adaptive multigrid method. Tartu, 1995.
8. **Teimuraz Shvelidze.** Automated quantitative spectral classification of stars by means of objective prism spectra: the method and applications. Tartu, 1999.
9. **Jelena Gerškevitš.** Formation and evolution of binary systems with compact objects. Tartu, 2002.
10. **Ivan Suhhonenko**. Large-scale motions in the universe. Tartu, 2003.
11. **Antti Tamm.** Structure of distant disk galaxies. Tartu, 2006.
12. **Vladislav-Veniamin Pustynski.** Modeling the reflection effect in pre-cataclysmic binary systems. Tartu, 2007.
13. **Anna Aret.** Evolutionary separation of mercury isotopes in atmospheres of chemically peculiar stars. Tartu, 2009.
14. **Mari Burmeister.** Characteristics of the hot components of symbiotic stars. Tartu, 2010.
15. **Elmo Tempel.** Tracing galaxy evolution by their present-day luminosity function. Tartu, 2011.