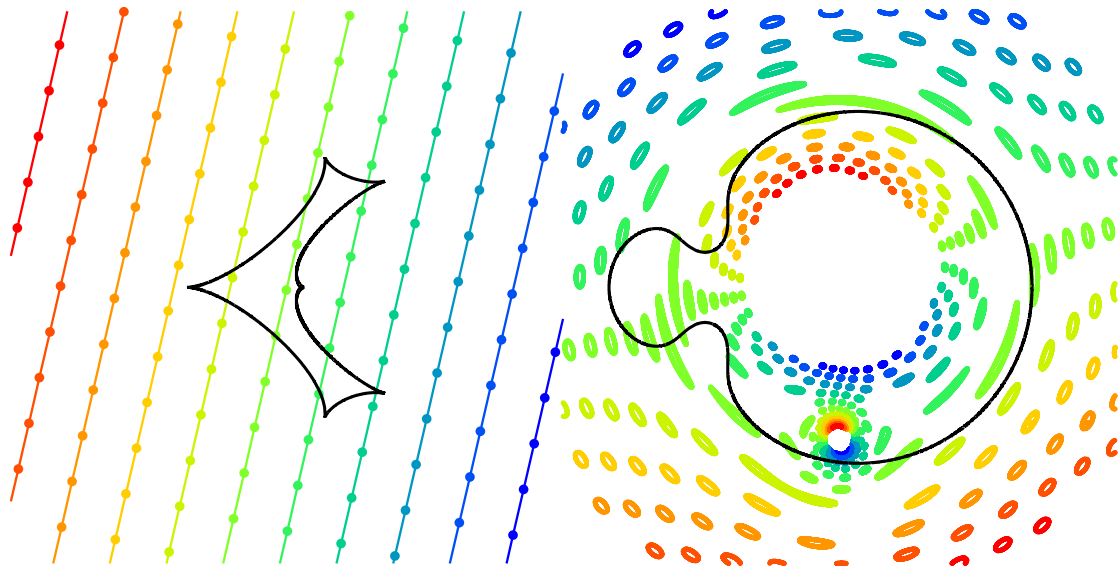Markus P. G. Hundertmark

Gravitational Microlensing: GPU-based Simulation Algorithms and the Information Content of Light Curves

Dissertation

Gravitational Microlensing: GPU-based Simulation Algorithms and the Information Content of Light Curves

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen

Doktorgrades

"Doctor rerum naturalium"

an der Georg-August-Universität Göttingen

vorgelegt von

Markus Peter Gerhard Hundertmark

aus Göttingen

Göttingen, den 18.05.2011

Referent: Prof. Dr. Stefan Dreizler
Korreferent :   Dr. Ansgar Reiners
Tag der mündlichen Prüfung: 20.06.2011

Abstract

The simulation of gravitational microlensing light curves is a time-consuming effort, but essential for confirming potential planetary microlensing events. The basic equations of lensing, describing the expected magnification of a source star affected by a binary lens, comprise transcendental functions whose roots cannot be determined analytically. Two major approaches to simulating light curves can be identified: ray-tracing simulating magnification maps or root-finding leads to the required image positions implying that the magnification can be simulated for arbitrary source positions.

The use of graphics processing units (GPUs) has become increasingly common for accelerating the simulation of light curves. Ray-tracing seems to be optimally suited for graphic cards, and thus the existing implementations have focused on this approach. In the context of this thesis, I present a different implementation of a point-wise binary lens model, parallelized on GPUs. As most graphic cards are equipped with machines with single-precision floating point numbers, special care was taken to understand and control the numerical accuracy.

In order to increase the efficiency of the process, I have exploited information theoretical concepts to speed-up the fitting process. For this purpose, the parameter space to be investigated, is probed using a Monte Carlo Markov Chain. Based on the insight, that observations differently contribute to the total information content, subsets of the light curves are selected according to the current state of the model.

The assessment of the information content of a light curve can also be useful for planning observations, particularly for following-up microlensing events with a high cadence. Prioritization schemes for selecting the most promising targets rely on the chance of detecting a planet. For this purpose the analytical point source point lens model is considered. I have addressed this approach from the point of view of optimal experimental design. After testing different ways of distributing observations, an information modulated approach offers a convenient compromise between sampling points with highest information content and preventing gaps.

In order to search for extrasolar planets, I have participated in follow-up observations as part of the MiNDSTEp consortium and prepared a data reduction pipeline for the Göttingen MONET telescope. The analysis of an anomalous event, observed in the season 2010, was used to test the newly developed GPU-model for fitting light curves. The best solution from a Markov Chain Monte Carlo fit was used for inferring the event parameters of the binary system. Based on the Besançon population synthesis model of the Galaxy, a physical interpretation of the system was given, indicating that one component is likely to be a substellar object. A planetary system can be excluded for both components. The second component is either a low-mass star or a heavy brown dwarf, which motivates further analyses. The estimated source star radius can only be consistently achieved, if the distance between lens and source star is $\approx 0.2\,\mathrm{kpc}$ and the corresponding stars are located around $6\,\mathrm{kpc}$.

I conclude this work, by providing estimates for the limiting magnitudes for two space missions,

namely the Heliospheric Imagers 1 (HI 1) onboard the STEREO satellites and the planned PLATO mission. According to these estimates, only extremely bright stars can be seen from (HI 1) and due to the short coverage of the bulge, this is a less rewarding endeavor. The prospects for PLATO are much better and depending on the final mission characteristics, high-magnification events can be observed with an appropriate signal-to-noise ratio contributing satellite parallax measurements from the Lagrangian point L2.

Contents

List of Figures

List of Tables

Only the ideas that we actually live are of any value.

Hermann Hesse, Demian

Chapter 1

Introduction

The deflection of light in the gravitational field of an astronomical body, predicted by Einstein (1911), has lead to a new field of research, namely the study of gravitational lensing: the observable distribution of electromagnetic radiation emitted by distant sources is changed by the presence of a mass-energy distribution along the line-of-sight. Advances in the instrumentation of telescopes turned these theoretical considerations into an active observational research field, with the discovery of multiple images of lensed Quasars starting by Walsh et al. (1979). If the deflecting mass distribution is a galaxy, Chang & Refsdal (1979) showed that its stellar constituents lead to a detectable change of brightness for each image. While the total mass of the lensing galaxy leads to resolved macroscopic images, the action of individual stars does not, which lead Paczynski (1986a) to introduce the concept of *gravitational microlensing*. In the following thesis, this system used in the context of Galactic gravitational microlensing, indicating that lens and source stars originate from our galaxy (Paczynski, 1986b).

In recent years, gravitational microlensing has impressively proven its capability for detecting distant low-mass objects, such as planetary systems in our galaxy. The sensitivity of these detections has reached several Earth masses (Beaulieu et al. 2006; Bennett et al. 2008) and theoretical estimates indicate that even moons (Liebig & Wambsganss 2010) or debris disks around host stars (Hundertmark et al. 2009) are detectable. On the theoretical side, higher order corrections for the deflection angle are accessible, but require a substantially higher data quality. Gravitational microlensing is subject to all kinds of distortions of space-time, and, as was recently shown by Abe (2010), can be applied even to detecting the presence of Ellis wormholes. For achieving further milestones beyond the detection of extrasolar planets and for improving the sensitivity of existing follow-up efforts, this thesis tries to exploit the modern computing capabilities of on graphics processing units (GPUs) for speeding-up the simulations of light curves. Furthermore, I discuss the benefits of applying basic principles of information theory, adding extra worth with respect to the dynamic estimation of microlensing parameters and their uncertainties and illustrating how data compression can support the analysis.

For testing the outlined improvements, observations were taken as part of the international MiNDSTEp consortium's search for extrasolar planets. Microlensing is an interesting detection technique for planets, due to its unique capabilities for detecting rocky extrasolar planets far away from their host stars, as microlenses are not required to be bright. In fact, lenses do not have to be observable at all, as long as a distant source star is sufficiently well-aligned with the lens star; For a typical configuration the angular separation is of the order microarcseconds (μas). Consequently,
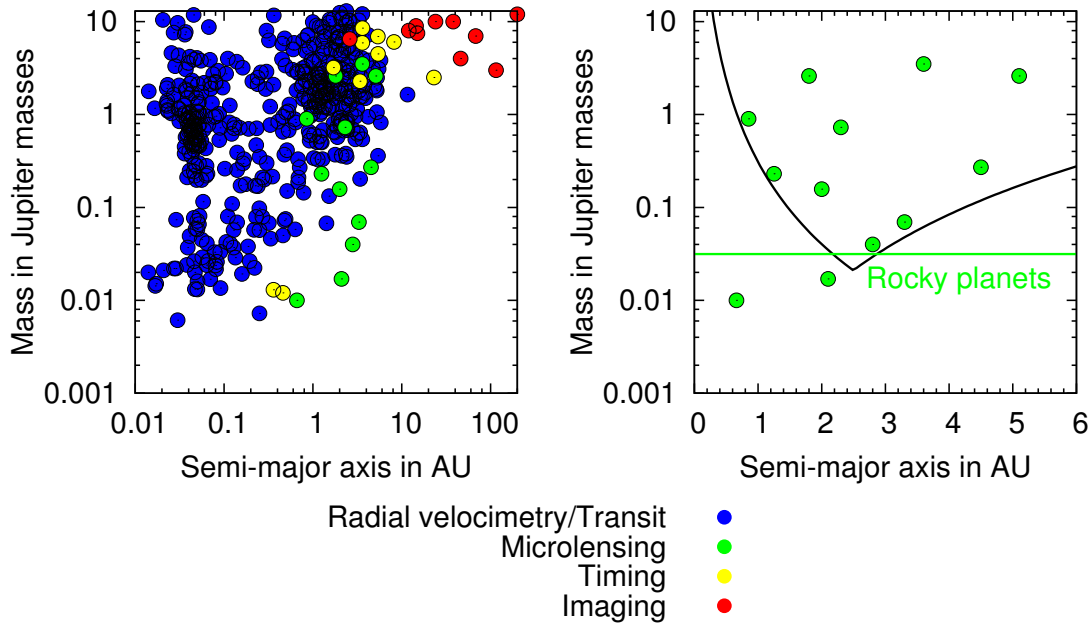
Figure 1.1: Mass and semi-major axis of extrasolar planets for different detection methods. The triangle on the left frame indicates where highly magnified microlensing events can certainly be detected.

microlensing probes a complementary parameter space in mass and semi-major axis[1] as illustrated in Fig. 1.1. Longer radial velocity measurements lead to an improving coverage of the semi-major axis especially for objects beyond Neptunian-mass, but the ground-based detections of Earth-mass planets around 2 AU achieved by microlensing, justifies further efforts. The distance from our Solar System, due to the detectability of faint lenses, is a second important aspect as illustrated in Fig. 1.2, where planetary systems are shown along with their estimated position in our galaxy. Thus, microlensing probes the large-scale distribution of extrasolar planets in the Milky Way. Actually, the true lens positions can deviate from the reported position, if the distances to lens and source are insufficiently constrained, for instance, because the mass estimate relies exclusively on statistical inference using an underlying Galactic model.

The structure of this work is motivated by an extrasolar planet follow-up scheme. The first chapter provides a derivation of the microlensing formalism based on the derivation of the deflection angle from General Relativity. Essential for microlensing is the lens equation encoding how light, subject to the attractive deflection of a mass-energy distribution, is redistributed. The simple analytical solutions for a point lens are given and examples of binary light curves are shown, whereas the binary lens model is predominantly applied for characterizing planetary events. Finally, a numerical simulation of a perturbed point lens model is shown, that is used for constructing an interpolant, enabling the assessment of higher order effects and their feasibility.

The second chapter introduces the benefits of modeling finite source point lens and finite source binary lens light curves by exploiting the capabilities of graphics processing units (GPUs)

---

[1]based on `exoplanet.eu` (22th April 2011)
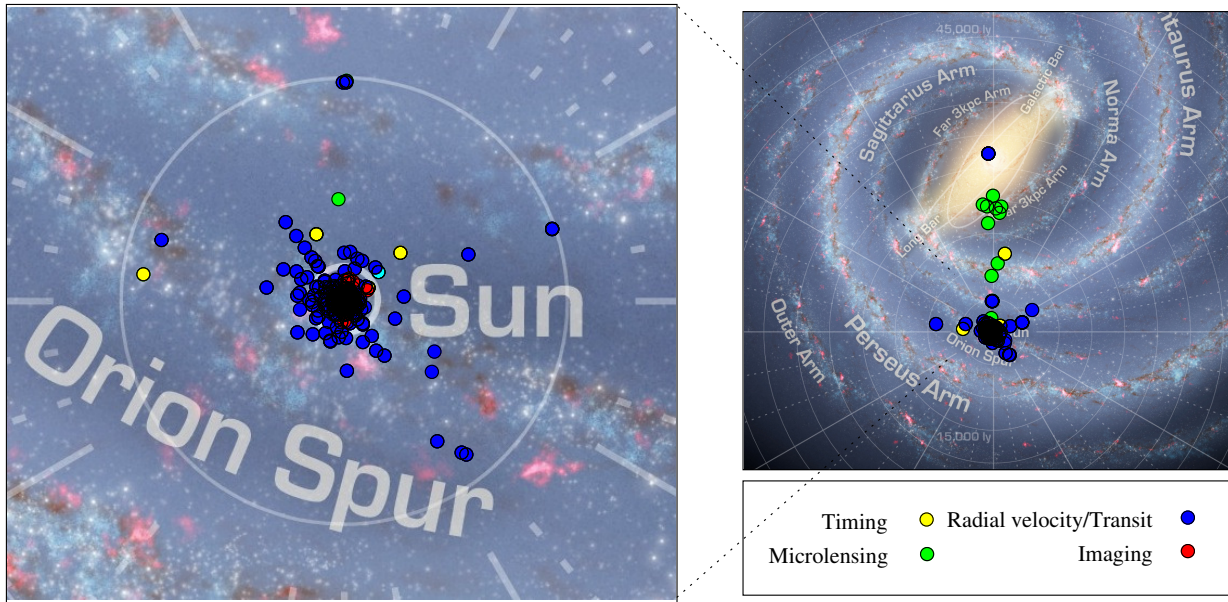
Figure 1.2: Artist's impression of the top-down view of the Milky way (Image credit: NASA) overlaid with distributions of known extrasolar planets.

and describes the implementation of a point-wise GPU-assisted contouring algorithm. The implementation was tested on GPUs, where processing units carry out computations with ≈ 7 digit precision. Accordingly, estimates of the resulting numerical accuracy of the model were made and compared with results from CPU-based ray-tracing simulations.

After discussing the prerequisites for modeling microlensing events, the information-driven sampling of microlensing events is discussed in the third chapter. An emphasis is put on the possible application of methods from information theory. For this purpose, Shannon's and Fisher's definition of information are compared in the context of Galactic microlensing. The former considers each single brightness measurement as independent carrier of information, while the latter approach shows how observations should be sampled for increasing the total information content, interpreted as the inverse variance-covariance matrix. These suggestions indicate how a fixed number of observations needs to be distributed in the course of an event, leading to an information modulated observing strategy.

For analyzing data with a Metropolis-Hastings Markov Chain Monte Carlo, the GPU-assisted contouring method demands modifications due to the limited model accuracy. In addition, it turns out that the Fisher information can be used for selecting subsets of the data-set, modifying the likelihood function at runtime but being able to neglect irrelevant parts of the observations for the given state of the Markov chain. A first implementation of such an approach is shown in chapter four. In the following chapter, five, the results of following up and modeling the event MOA 2010-BLG-406 are shown. For this purpose, relevant aspects for setting-up a data reduction pipeline are discussed and a physical interpretation of the best fit is given, based on a Galactic model.

The final chapter illustrates the capabilities of the MONET telescopes for a test observation in 2010, where MONET/North has achieved a good coverage for MOA 2010-BLG-477. In addition, the prospects of using auxiliary telescopes as part of space missions are discussed. For this purpose, signal-to-noise estimates and limiting magnitudes for microlensing observations towards

the Galactic center are given for the Heliospheric Imager 1 onboard the STEREO mission and compared with estimates for the future use of the PLATO spacecraft, which is one of ESA's planned M-class missions.

Chapter 2

Gravitational microlensing

Gravitational Microlensing is the science of modeling and measuring the effects of light deflection in metric theories of gravitation by studying unresolved images of Galactic and Extragalactic source objects. In this respect gravitational microlensing is a special case of strong gravitational lensing, where galaxies or galaxy clusters produce resolved multiple images of whole galaxy or quasar source objects which can be seen from ground based observations. In the following thesis, the special case of Galactic microlensing will be considered, where lens and source are stellar objects located in the Milky Way Galaxy.

2.1   The deflection angle

Before being able to predict how light is deflected, it is inevitable to start with the conception of space-time in the sense of special relativity (Einstein, 1905) where time is considered to behave like one of the other spatial coordinates. The constancy of the speed of light in all reference frames implies in the flat space of special relativity that massive test particles move with light speed through the four dimensional space-time. The time in a comoving reference frame denoted as proper time depends on its relative motion and the closer it gets to the speed of light the smaller the eigen time will be. The remarkable insight of Einstein (1915) that space and time are curved by energy and momentum leads to a modification of that picture - time and space intervals have to be changed according to the curved geometry giving rise to changed equations of motion. The quantitative description of how space-time is curved is encoded in Einstein's field equations, where curvature, expressed as a Ricci tensor $R_{\mu\nu}$, the scalar curvature R, the stress-energy tensor $T_{\mu\nu}$, and the cosmological constant Λ are connected to each other:

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^2}T_{\mu\nu}. \tag{2.1}$$

These field equations provide the framework for determining the metric tensor $g_{\mu\nu}$[1] describing how lengths and times can be locally measured. A vacuum solution for spherical objects, which is particularly relevant for gravitational lensing, was derived by Schwarzschild (1916): its line element ds, shown in the form of Weinberg (1972), for spherical coordinates is

$$\mathrm{d}s^2 = -A(r)c^2\mathrm{d}t^2 + B(r)\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2, \tag{2.2}$$

---

[1]the chosen metric signature is $(-, +, +, +)$

Figure 2.1: Flat and curved two dimensional space is illustrated for a star and empty space-time; n.b. this simplified picture suggests only spatial coordinates are curved but the coordinate time is affected as well.

where the expressions $A(r)$ and $B(r)$ are given by

$$A(r) = 1 - \frac{2GM}{c^2r}, \qquad B(r) = \left(1 - \frac{2GM}{c^2r}\right)^{-1}. \tag{2.3}$$

The latter can be modified for higher order corrections as shown by Keeton & Petters (2005). For an appropriate description of motion of particles and light in curved space-time, equations of motion like the Euler-Lagrange equations in classical mechanics need to be found. The natural approach for this endeavor is to start searching for an extremal principle giving rise to the already known equations characterizing the problem. In general relativity, one can understand light deflection in the context of Fermat's principle where light takes the "shortest" path through space-time which is by definition a geodesic. The arc-length with a given metric tensor $g_{\mu\nu}$ takes the form

$$\int_{\lambda_1}^{\lambda_2} \left(g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\right)^{1/2} \mathrm{d}\lambda, \tag{2.4}$$

where λ parameterizes the path. Alternatively one can use the principle of stationary action (compare, for example, Gönner 1996)

$$\int_{\lambda_1}^{\lambda_2} \frac{1}{2} g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} \mathrm{d}\lambda \tag{2.5}$$

for finding the equations of motion, where the integrand is the Lagrangian L of the problem. The light deflection of a spherically symmetric body cannot depend on the angle θ, which can then be conveniently chosen to be $\theta = \pi/2$. In the latter case, the Lagrangian of the Schwarzschild metric Eq. 2.2 is

$$L = \frac{1}{2} \left(-A(r)c^2 \left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 + B(r) \left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + r^2 \left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2\right) \mathrm{d}\lambda. \tag{2.6}$$

The Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\partial L}{\partial (\mathrm{d}x^\mu/\mathrm{d}\lambda)}\right) = \frac{\partial L}{\partial x^\mu} \tag{2.7}$$

provide constants of motion k_ϕ and k_{ct} for the coordinates ϕ and ct, as these coordinates do not explicitly appear in Eq. 2.6:

$$k_{ct} = A(r)c\frac{\mathrm{d}t}{\mathrm{d}\lambda} \tag{2.8}$$

$$k_\phi = r^2\frac{\mathrm{d}\phi}{\mathrm{d}\lambda} \tag{2.9}$$

A third equation of motion can be obtained by using the fact that light propagates at zero eigen time and, thus, that the line element $g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}$ vanishes:

$$-A(r)c^2\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 + B(r)\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2 = 0 \tag{2.10}$$

For determining the deflection angle, a differential equation for $r(\phi)$ is required. Dividing by $\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2$ gives

$$-A(r)c^2\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^{-2} + B(r)\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 + r^2 = 0. \tag{2.11}$$

Inserting the constants of motion leads to

$$-\frac{k_{ct}^2}{k_\phi^2} + \frac{1}{r^4}\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 + \frac{A(r)}{r^2}, \tag{2.12}$$

which can be further simplified by substituting $p = r^{-1}$ (see Capozziello et al. 1997)

$$\left(\frac{\mathrm{d}p}{\mathrm{d}\phi}\right)^2 - \frac{2GM}{c^2}p^3 + p^2 - \frac{k_{ct}^2}{k_\phi^2} = 0. \tag{2.13}$$

The desired differential equation can be simplified by differentiating with respect to ϕ and dividing the result by $2\frac{\mathrm{d}p}{\mathrm{d}\phi}$:

$$\frac{\mathrm{d}^2p}{\mathrm{d}\phi^2} + p = \frac{3GM}{c^2}p^2, \tag{2.14}$$

which is an ordinary second-order nonlinear differential equation. The solution for the homogeneous solution is easily found to be

$$p(\phi) = C\sin(\phi + \phi_0) \tag{2.15}$$

where the integration constant can be determined for the impact parameter b as illustrated in Fig. 2.2. The angle ϕ can be freely chosen to cancel the phase angle ϕ_0. For a thin lens, the impact parameter can be used to determine the constant $C = 1/b$ and thus the homogeneous solution is

$$p(\phi) = b^{-1}\sin(\phi) \tag{2.16}$$

The solution for the inhomogeneous part (Eq. 2.14) is given by Schröder (2007) as

$$p_p(\phi) = \frac{GM}{c^2b^2}\left(1 + \cos(\phi)\right). \tag{2.17}$$
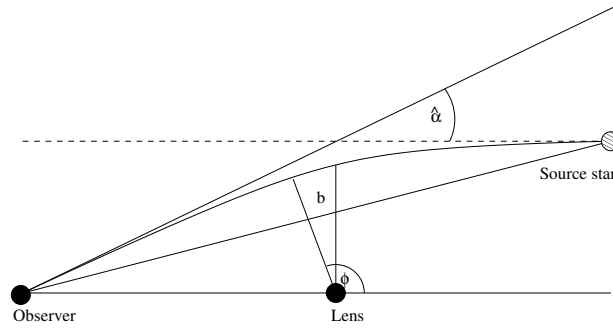
Figure 2.2: Definition of quantities for determining the deflection angle.$\hat{\alpha}$

The combination of both solutions can be expressed in Cartesian coordinates

$$y = b - \frac{GM}{c^2 b} \frac{2x^2 + y^2}{\sqrt{x^2 + y^2}} \approx b - \frac{2GM}{c^2 b} x \tag{2.18}$$

According to the definitions in Fig. 2.2, the total deflection angle can be determined as

$$\hat{\alpha} = \frac{4GM}{c^2 b}. \tag{2.19}$$

There are other ways of deriving this deflection angle using perturbative approaches fo Eq. 2.17 or by expanding the elliptic integrals solving Eq. 2.14, but approximations are necessary in the course of each derivation. The deflection angle Eq. 2.19 can be determined even with the same result, in the weak-field approximation and thus without the Schwarzschild metric. The deflection angle is not just robust concerning its derivation - another benefit from keeping the first order approach turns up when considering the lens equation.

2.2   The lens equation

In classical optics as well as in gravitational microlensing the creation of images is described by a lens equation under similar approximations. In both cases, the propagation of light is modeled as one dimensional ray propagating orthogonally to the wavefront of the electromagnetic wave and the thin lens approximation is applied. Recent calculations by Walters et al. (2010) have shown one way of superseding the approximations that have been made in the previous section. This approach would require numerical efforts for integrating the deflection angle which can be neglected for regular main-sequence stars acting as lenses. The standard lensing geometry is illustrated in Fig. 2.3. The corresponding lens equation for this case can be written as

$$\tan \beta = \tan \theta - \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}} \left(\tan \theta + \tan(\hat{\alpha} - \theta) \right) \tag{2.20}$$

If the lens star is not in the relativistic regime, i.e. the lens is not a black hole (cf. Virbhadra & Ellis 2000) the small angle approximation is satisfied and thus the lens equation can be written as

$$\boldsymbol{\beta}(\theta) = \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}} D_{\mathrm{s}}} \frac{4GM}{c^2 \boldsymbol{\theta}}. \tag{2.21}$$
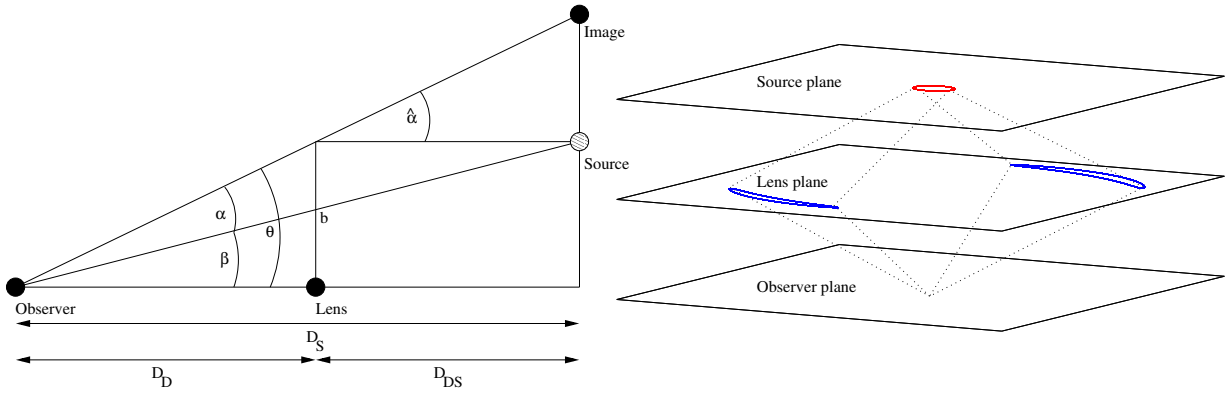
Figure 2.3: Side view of the lensing geometry (left) and the corresponding magnification of the source star area.

Assuming the number of stellar black holes in our galaxy is $\approx 10^9$ (Brown & Bethe, 1994) and the Milky Way contains of the order of 10^{11} stars, several microlensing events could have been black holes, but for sufficiently large separations these are well approximated by the small angle approximation. The lens equation (Eq. 2.21) can be used to model symmetric lenses. For systems with multiple components it is assumed that deflection angles can be superposed and thus the total deflection of N deflectors and an underlying mass-energy density can be obtained

$$\boldsymbol{\beta}(\theta) = \boldsymbol{\theta} - \sum_i^N \frac{D_{\mathrm{ds},i}}{D_{\mathrm{d},i}D_{\mathrm{s},i}} \frac{4GM_i\Delta\boldsymbol{\theta_i}}{c^2\Delta\theta^2}. \tag{2.22}$$

The superposition approach is valid as long as the deflectors are sufficiently separated and thus the curvature of one lens does not affect the curvature of another object, which is a first order consequence of the weak field approximation for

$$\frac{2GM}{c^2r} \ll 1. \tag{2.23}$$

2.3 Observable quantities

2.3.1 Einstein radius

In the commonly used point lens equation (Eq. 2.21), the constant quantities can be summarized by considering a collinear alignment of source star, lens star and observer, i.e. $\beta = 0$. In this case the solution corresponds to a ring with an angular radius of

$$\theta_{\mathrm{E}} = \sqrt{\frac{4GM}{c^2}\frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}}}. \tag{2.24}$$

This quantity can be used to obtain the dimensionless lens equation for $i \neq j$

$$\frac{\beta_j}{\theta_{\mathrm{E}}} = \frac{\theta_j}{\theta_{\mathrm{E}}} - \frac{\theta_{\mathrm{E}}\theta_j}{\theta_i^2 + \theta_j^2}, \tag{2.25}$$
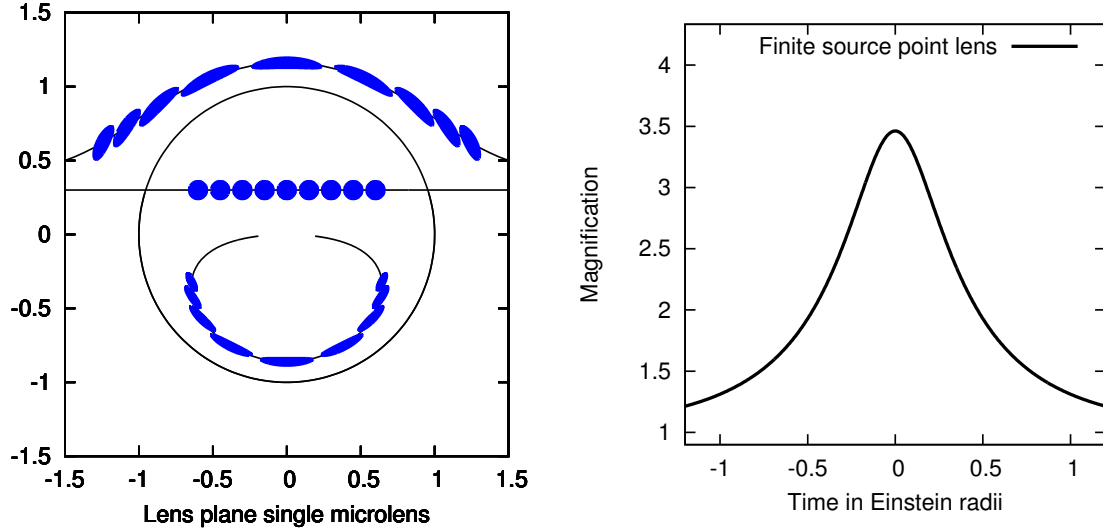
Figure 2.4: Images of a point lens (left) and the corresponding light curve shape (right).

where the angular Einstein radius does not have to be explicitly included if all angles are given in its unit. In the following, this assumption is made and the corresponding quantities are denoted as θ' and β'. Therefore Eq. 2.25 becomes

$$\beta'_j = \theta'_j - \frac{\theta'_j}{\theta'^2_i + \theta'^2_j} \tag{2.26}$$

The Einstein radius is of the order of a milliarcsecond (mas) for Galactic microlensing, preventing a direct measurement, whereas the Einstein radius can be resolved for extragalactic objects (e.g. Fig. 2.5).
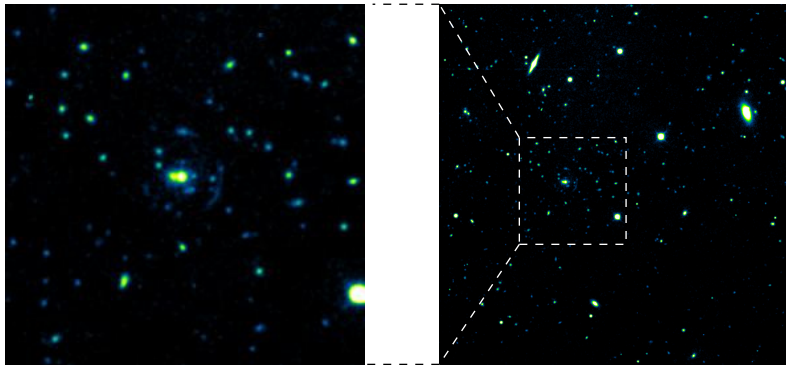
Figure 2.5:   Example for strong lensing of galaxies: Hall's arc (SDSS J104943.14+442035.6) as seen from the DANISH 1.54m telescope in the R-band. For detailed measurements see Estrada et al. (2007)

2.3.2 Brightness

Point source point lens

Currently the most common observational technique for Galactic microlensing is that proposed by Paczynski (1986b) and is based on the analysis of microlensing light curves requiring an adequate description how such a curve can be created. For the purpose of data fitting it is convenient to consider the simple model of a symmetric lens. The magnification can be obtained by solving the lens equation for the images $\boldsymbol{\theta}$, integrating the subtended angle of the solutions and, where appropriate, convolving the solutions with a suitable limb-darkening law. For idealized point source stars, the magnification of a single image of the lens equation is given by the modulus of the inverse Jacobian determinant (Schneider et al., 1992) constituting the infinitesimal change of an area element subject to the gravitational deflection. In an ordinary Cartesian representation, the magnification for a single solution θ of the lens equation is

$$\mu = \left(\frac{\partial \beta_i}{\partial \theta_i} \frac{\partial \beta_j}{\partial \theta_j} - \frac{\partial \beta_i}{\partial \theta_j} \frac{\partial \beta_j}{\partial \theta_i} \right)^{-1} \tag{2.27}$$

which can be calculated for every differentiable lens equation. For a point source point lens (PSPL) model, both components of the lens Eq. 2.26 can be combined and provide as corresponding image positions

$$\theta'_{j,\pm} = \frac{1}{2} \left(|\boldsymbol{\beta}| \pm \sqrt{\beta^2 + 4} \right) \frac{\beta'_j}{|\boldsymbol{\beta}|}. \tag{2.28}$$

For each solution in the lens plane the Jacobian determinant can be calculated according to Eq. 2.28 and the magnification becomes

$$\mu = \left(1 - \frac{1}{|\boldsymbol{\theta}'|^4} \right)^{-1}. \tag{2.29}$$

The total magnification for a given source position can be obtained by summing up the magnification at every given image position. For a PSPL model the magnification can be obtained from Eqs. 2.29 and 2.28 and depends only on the modulus of the source position $|\boldsymbol{\beta}'|$ which is denoted in the literature as $|\boldsymbol{\beta}'| = u$

$$\mu(u) = \frac{u^2 + 2}{u \left(u^2 + 4 \right)^{1/2}} \tag{2.30}$$

This equation serves as a fiducial model for further considerations. Light curves can be simulated according to this equation by parameterizing the changing source position. To first order, one can assume that a point lens and a point source move uniformly with different proper motions. Relevant for the magnification of an ongoing event is the relative motion of source and lens. The timescale of a microlensing event is characterized by the time t_E it takes to cross the Einstein ring while the maximal magnification depends on the closest separation u_0 at a time t_0:

$$u(t) = \left(u_0^2 + \frac{t - t_0}{t_E} \right)^{1/2}. \tag{2.31}$$

Among these parameters the Einstein time is the most interesting one, because it is related to the mass. Nevertheless, knowledge of the distances and the relative lens-source proper motion

$$\mu_r = \frac{\theta_E}{t_E} \tag{2.32}$$

is required for determining the mass. Furthermore, all observations are carried out in crowded fields and the flux of a source star F_S can be contaminated by blending flux from other objects F_B (Di Stefano & Esin, 1995) which, in turn, can exhibit variability. The observed light curves thus take the shape

$$F(t) = F_S(t)\mu(u(t)) + F_B(t). \tag{2.33}$$

If the blending can be considered to be insensitive with respect to time-dependent changes, it is convenient to introduce the blending parameter $g = \frac{F_B}{F_S}$ (e.g. Dominik 2009). The blending parameter can have a significant impact on the determination of the Einstein time, which is the observable quantity related to the mass of the lens. The effect can be seen if the full width at half maximum (FWHM) of a point source point lens (PSPL) model is considered as introduced by Gondolo (1999) and approximated for source star positions much smaller than the Einstein radius

$$t_{\mathrm{FWHM}} \approx 3.46 \frac{t_{\mathrm{E}}}{\mu(u_0)} \tag{2.34}$$

where it can be directly seen how a misinterpretation of the blending parameter alters the prospects of accurately determining the Einstein time for high magnification events.

The binary lens

Binary lens models with two components in the general lens equation (Eq. 2.22) play a major role for characterizing ongoing events deviating from the standard PSPL structure. Even if binary stars cannot be found in 1/2 but 1/3 of all potential stellar systems as suggested by Lada (2006), the number of binary stars serving as lenses for surveys observing 10^8 stars at any given time stays high.

Different magnification maps are shown in Fig. 2.6. These maps provide the magnification for a given source star position on the sky. In cases where the Jacobian determinant (Eq. 2.27) vanishes, the resulting point source magnification diverges, but this is a hypothetical scenario, since all source objects are extended structures. Points in the lens plane with infinite magnification are referred to as critical curves and, after being mapped to the source plane, they become caustics. In the case of a point lens the critical curve is the Einstein radius and the caustic is the diverging magnification for a collinear alignment of lens, source and observer. The angular Einstein radius for binary systems is

$$\theta_{\mathrm{E}} = \sqrt{\frac{4G\left(M_1 + M_2\right)}{c^2} \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}}}, \tag{2.35}$$

where M_1 and M_2 are chosen in a way that the mass ratio $q = \frac{M_1}{M_2} \leq 1.0$ and the angular separation d is given in units of the angular Einstein radius. In contrast to the single lens case, the angle between source trajectory and connecting line of both lenses arises as additional model parameter. Without loss of generality, a convenient coordinate system can be chosen: throughout this thesis the center of mass is chosen as origin and both lenses are placed on the first axis.

Depending on the combination of the parameters mass ratio and separation, different lensing regimes can be discriminated (Schneider & Weiss, 1986). Extended caustic structures can easily be detected as the number of images changes when a source star passes a caustic. For randomly oriented tracks, the chance of detecting a binary system in a given light curve increases with
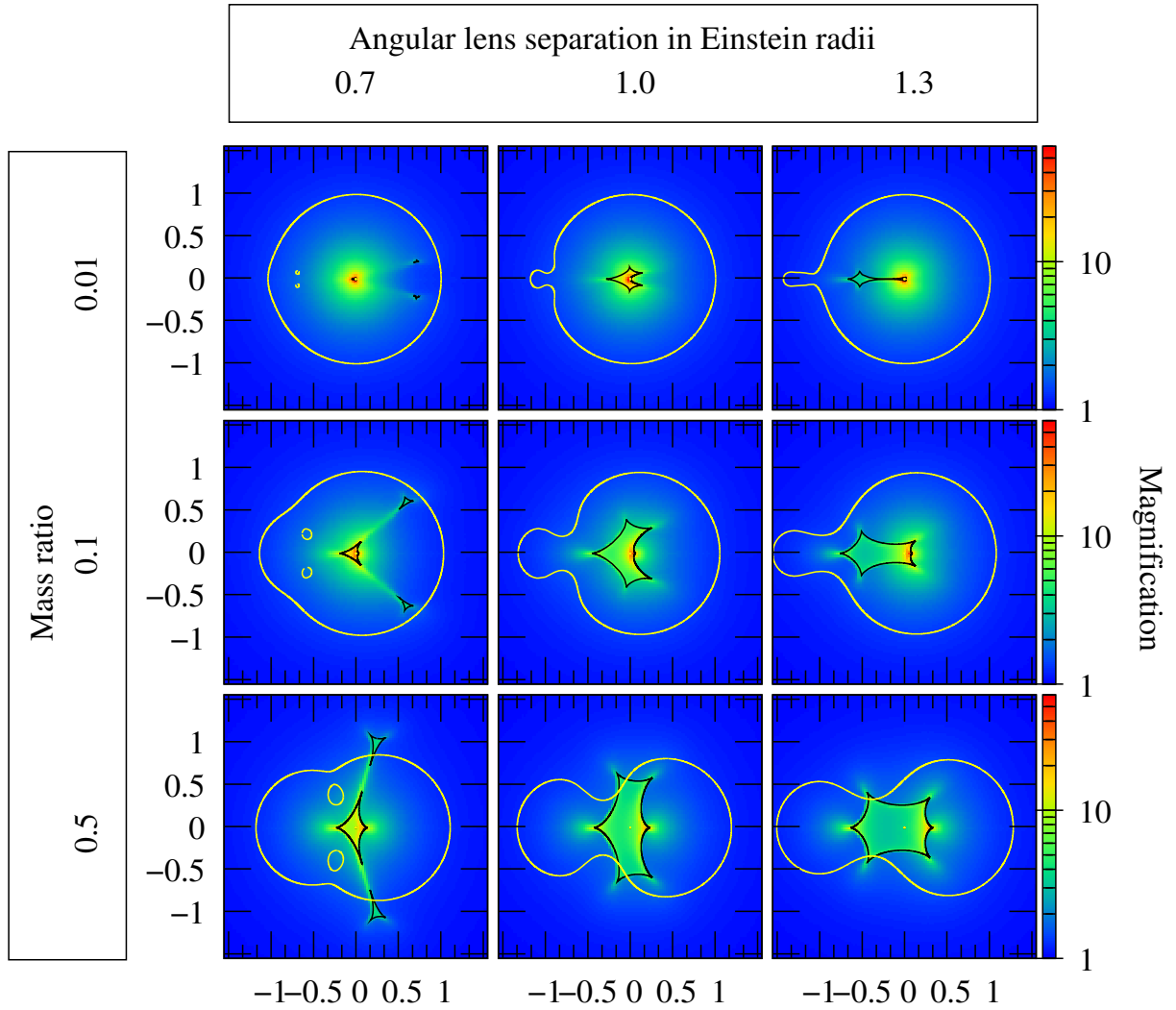
Figure 2.6: Magnification maps are shown along with critical curves and caustics for different mass ratios q and separations d .

caustic size while, in turn, the caustic size depends on mass ratio and separation. Fig. 2.6 illustrates qualitatively different trends such as an increase of the caustic size with mass ratio as well as an increase in caustic size for separations $d \approx 1 \; \theta_{\mathrm{E}}$. For widely separated binaries, the resulting caustic structure resembles the superposition of two point lens magnification patterns. In addition, it is worth noting that the number and position of cusps can serve as important constraints for selecting possible configurations. The drastic changes of magnification close to cusps and during caustic crossing events enable the detection of low-mass objects such as planets below $10 \, M_{\oplus}$ even from ground-based observations. Fig. 2.7 gives an example of the variety of light curves that can be obtained just for one combination of mass ratio and separation. Details on the calculation of these events by means of graphical processing units are given in chapter 3.
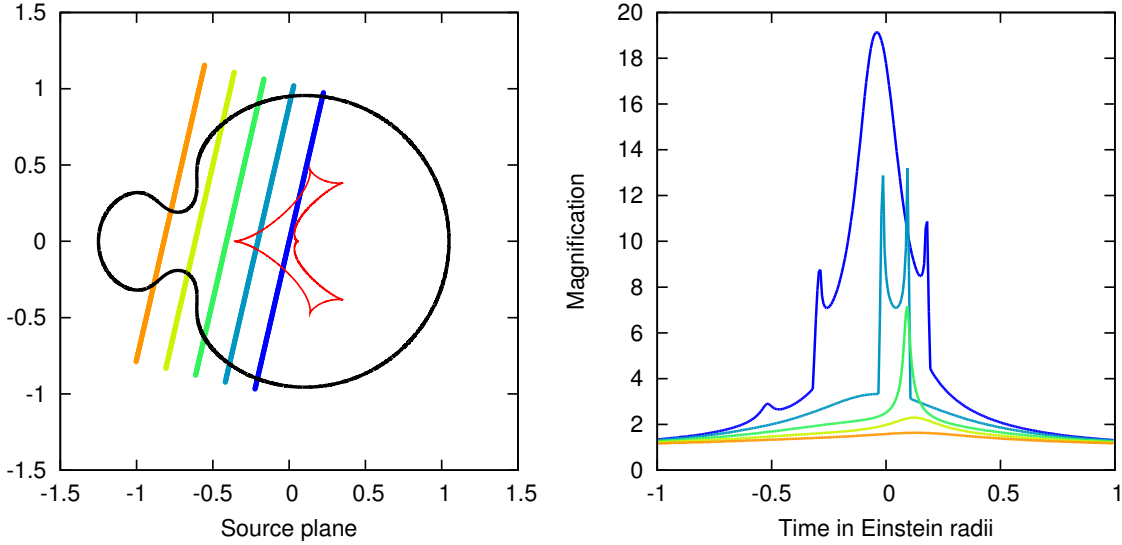
Figure 2.7: Binary light curves for different source tracks are shown along with critical curves and caustics. A mass ratio of 0.1, a separation of $0.9\,\theta_E$ and a source star radius of $0.007\,\theta_E$ have been chosen.

2.4   Optical depth and event rates

It was shown in Equations 2.31 and 2.35, that the duration of an event is related to the mass of a given lens distribution. However, to infer the lens mass, potential lens and source distances, as well as mass and velocity distributions are required. The first estimates for graviational microlensing were made by Paczynski (1986b) in his seminal paper.

The fundamental quantity for detecting gravitational microlensing events is the optical depth: it does not describe the transparency of a medium but provides a measure of the chance of a given source star to be lensed. Evidently, this is related to the measure of transparency contained in the classical optical depth, as it gives an estimate of how transparent the Galaxy is with respect to being lensed. For this purpose, a definition has to be made, when a source star has been lensed. A source star is considered to be lensed when the closest separation is smaller than $1\,\theta_E$ and thus the magnification exceeds 1.34. The optical depth τ depends on the sizes of the Einstein radii for a given distribution of deflectors, where each deflector contributes its Einstein ring area

$$\tau = \pi \int_0^{D_s} n(D_d)\theta_E^2 \mathrm{d}D_d \tag{2.36}$$

comparable to the cross section in nuclear and particle physics. The optical depth depends on the linear number density of lenses $n(D_d)$ and the Einstein radius for the given distance.

A further parameter of interest is the event rate Γ of microlensing events. Paczynski (1991) connected the event rate to the optical depth and the event duration t_E as the inverse time interval between two consecutive microlensing events for the same observer-source configuration

$$\Gamma = \frac{2\tau}{\pi t_E}, \tag{2.37}$$

14

which has to be integrated for different source positions but also for different relative proper motions μ_r depending on the given source and lens position, leading to an event rate

$$\Gamma = 2 \int_0^{D_s} n(D_d) \theta_E \mu_r dD_d. \tag{2.38}$$

2.4.1 Astrometry

For the confirmation of extrasolar planets in a binary microlensing event it is necessary to know the mass of the lens star: binary models can provide a direct measurement of the mass ratio but even the parameter space for determining mass ratio and separation is affected by degeneracies caused by the structure of the binary lens equation (Dominik, 1999). Astrometrical observations can help to discriminate between different lens configurations.

The centroid position of the source star is changed due to the multiplicity of solutions of the binary lens equation and their corresponding brightnesses (Hog et al., 1995; Miyamoto & Yoshii, 1995; Walker, 1995). For typical microlensing configurations with a lens of $0.3 \, M_\odot$ located at $6.5 \, \text{kpc}$ and source star located at $8.5 \, \text{kpc}$, the resulting Einstein radius corresponds to an angular size of $0.3 \, \text{mas}$. The centroid position $\boldsymbol{\theta}_c$ for a lens configuration with N solutions $\boldsymbol{\theta_i}$ is

$$\boldsymbol{\theta}_c = \frac{\sum_i^N k_i \boldsymbol{\theta}_i \mu_i}{\sum_i^N \mu_i} \tag{2.39}$$

where k_i denote possible transmission coefficients for light-rays passing through an attenuating medium. The transmission coefficients can be used to analyze an attenuating medium around a microlens as indicated by Hundertmark et al. (2009). Fig. 2.8 illustrates the maximal effect of an optically circumstellar thick disk around a microlens.
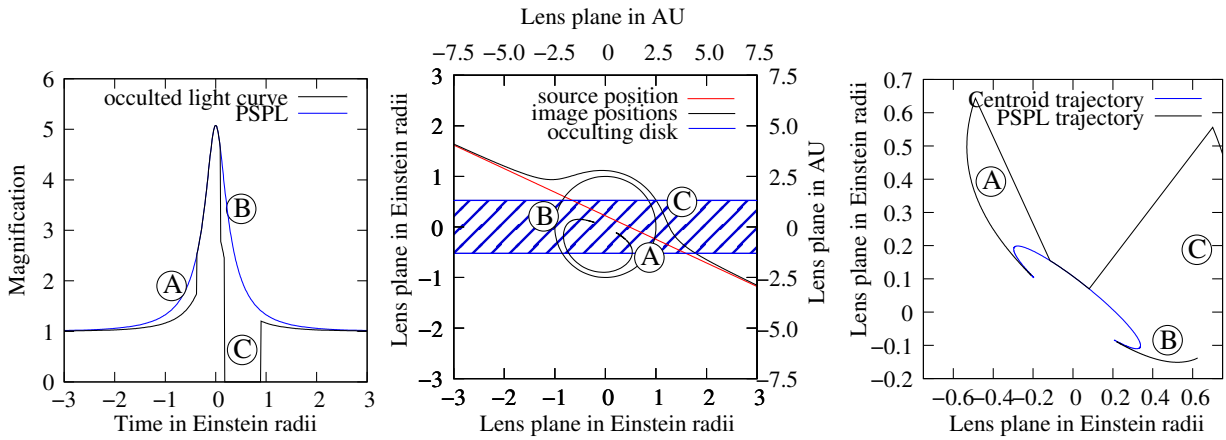
Figure 2.8: PSPL images occultated by an optically thick disk is shown for an inclination of 89.5 degrees (left). The features A and B are caused by the occultation of the less-magnified image μ_- and feature C corresponds to a total occultation. The length scale in AU assumes a typical Galactic microlensing event (right).

2.5   Classification of Microlensing applications
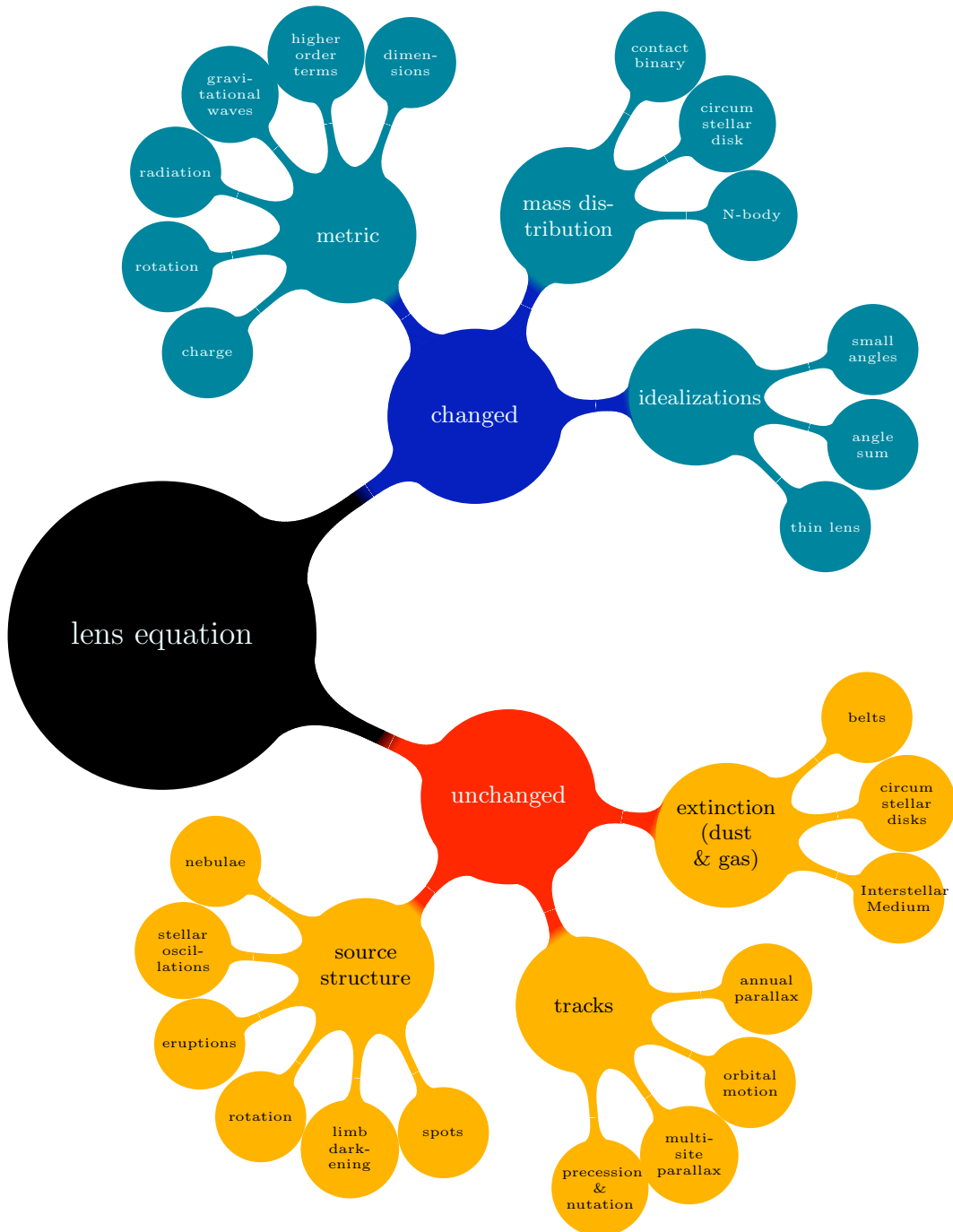
Figure 2.9: Mindmap for microlensing related effects from a modeler's point of view.

The observable shape of a point source point lens microlensing light curve can be altered by two different classes of effects; effects changing the lens equation and effects preserving the magnification pattern on the sky but modifying everything else. The former class comprises changes in the underlying metric as well as the mass and energy distribution of the deflection system,

whereas the latter class considers arbitrary source structures, blending, extinction and dynamical changes caused by the complicated structure of the lens-source-observer relative motion. Equally well one could have classified the effects according to different observational techniques, but from a modeler's point of view the given choice outlines the implementation. Fig. 2.9 illustrates examples of different detectable effects directly following from common astrophysical concepts and independent of potential detailed studies. A complete picture requires the treatment of several effects and all aspects shown in Fig. 2.9 can be present at the same time.

In the following, methods for quickly estimating the impact of these effects in comparison to the point lens will be discussed. Depending on the objectives, a detailed analysis can be a time consuming effort. Due to the existing surveys MOA (Microlensing Observations in Astrophysics; Bond et al. (2001); Sumi et al. (2003)) and OGLE (The Optical Gravitational Lensing Experiment; Udalski (2003)) an emphasis is put on brightness variations.

2.5.1 Perturbations of the PSPL model

As indicated, the PSPL model and its lens equation serve as a fiducial model for all further calculations and its lens equation (Eq. 2.21) can be modified by adding an angular perturbation vector $\boldsymbol{\delta}(\boldsymbol{\theta})$, where we will assume that the additional deflection is small in contrast to the PSPL approximation:

$$\boldsymbol{\beta}(\theta) = \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}} \frac{4GM}{c^2\boldsymbol{\theta}} + \boldsymbol{\delta}(\boldsymbol{\theta}) \tag{2.40}$$

In the most simple case, the perturbation does not depend on the image position and thus the general shape of the magnification map is conserved; $\boldsymbol{\delta}$ vanishes in the partial derivatives of Eq. 2.27. The PSPL model is adapted simply by changing the source positions $\boldsymbol{\beta}$ for each solution. One example of such a perturbation is given by Chiu et al. (2006), who treat gravitational lensing in Tensor-Vector-Scalar (TeVeS) gravity (Bekenstein, 2004), illustrating that the effect does not leave the light curve shape invariant. For all further considerations it is assumed that the perturbed light curve changes its shape and cannot be scaled or shifted for correcting the baseline and blending flux in Eq. 2.33.

Modifying the mass distribution of the lens or by using higher order approximations for different physical situations (e.g. Keeton & Petters 2006a,b) leads to changes depending on the lens position which generally cannot be solved analytically. In addition, it is not clear a priori how the number of solutions evolve. For treating all possible solutions, the squared deviation of the perturbed lens equation (Eq. 2.40)

$$sqd = \left| \boldsymbol{\beta}(\theta) - \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}} \frac{4GM}{c^2\boldsymbol{\theta}} + \boldsymbol{\delta}(\boldsymbol{\theta}) \right|^2 - r_\star^2 = 0 \tag{2.41}$$

as introduced by Schramm & Kayser (1987) is considered, permitting a quick inspection of the solution structure. The magnification in the lens plane can be evaluated numerically according to Eq. 2.27 with multi-precision packages[2]. The image position is corrected for the perturbed version by using a random walk starting at the point lens solutions and taking random steps scaled according to the squared deviation.

By definition, this approach fails to address finite source effects and thus certain parts of the parameter space must not be included. The achievable maximal magnification of a source star

---

[2]The calculations presented here were carried out in Python using the *Mpmath* module by Johansson et al. (2010).

modeled as a uniform disk with radius $r_\star$ can be directly calculated from Eq. 2.28. In this case, the images merge and form the Einstein ring with $\boldsymbol{\theta}_\pm$ as inner and outer radii. The maximal magnification becomes

$$\mu_{\max}(r_\star) = \sqrt{1 + \frac{4}{r_\star^2}}.$$
(2.42)

In the picture of the PSPL model this corresponds to a minimal separation of

$$u_0 = \sqrt{r_\star/2\left(1 + \left((1 + 4/r_\star^2)^2 - (1 + 4/r_\star^2)\right)^{1/2} - (1 + 4/r_\star^2)\right)}.$$
(2.43)

For a typical configuration[3] a sun-like star acting as source has a radius of $2 \cdot 10^{-3}\,\theta_{\mathrm{E}}$ while the disk of a giant star with $55\,R_\odot$ corresponds to $0.01\,\theta_{\mathrm{E}}$. Instead of imposing a cutoff in minimal separation, a cutoff is imposed for the maximal magnification which can be applied to prevent the evaluation at caustic crossings. For the intermediate case $r_\star = 0.005\,\theta_{\mathrm{E}}$, this implies a magnification limit of $\mu_{\max} \approx 400$.

The evaluation of the detectability requires knowledge of the sampling of points in the source plane. As long as no prior information is available, all tracks on the sky are equally likely: consequently an equidistant grid with sampling intervals of $0.02\,\theta_{\mathrm{E}}$ is chosen. For an Einstein time $t_{\mathrm{E}} = 20\,\mathrm{d}$ an observation is necessary every $0.4\,\mathrm{d}$, which is achieved by survey programs like OGLE and MOA.

Binary perturbation

Before looking at two classes of perturbations, I consider the canonical binary model. In Fig. 2.10 the effect of a perturbation caused by a binary lens is shown. The magnification in the lens plane displays lines of infinite magnification, where the Jacobian determinant vanishes - the aforementioned critical curves. If the source star is located inside or outside the caustic either three or five images are created. One way of estimating the maximal number of solutions is achieved by converting the lens equation to a complex polynomial, as shown by Witt (1990); according to the fundamental theorem of algebra, the number of solutions can be as high as the degree of the resulting polynomial. Consequently, the total magnification will be underestimated if just two solutions from the magnification pattern are used, but if these solutions already increase the total brightness by a detectable amount, detailed studies are justified. Fig. 2.11 gives an example of a perturbed binary model with source positions sufficiently far away from the caustic.

2.5.2   Higher order calculations

For directly estimating the order-of-magnitude deviation of higher order effects, two simple interpolating relations shall be given which directly provide a measure for the deviation and can be used for a sensitivity and detectability analysis. In the first case, the lens equation can be modified by including second order effects, leading to the lens equation

$$\beta_j' = \theta_j' - \frac{\theta_j'}{\theta_i'^2 + \theta_j'^2} - k_{\mathrm{sec}} \frac{\theta_j'}{\left(\theta_i'^2 + \theta_j'^2\right)^{3/2}}$$
(2.44)

---

[3]$D_S = 8.5\,\mathrm{kpc}, D_L = 6.5\,\mathrm{kpc}$ and $M = 0.3\,M_\odot$
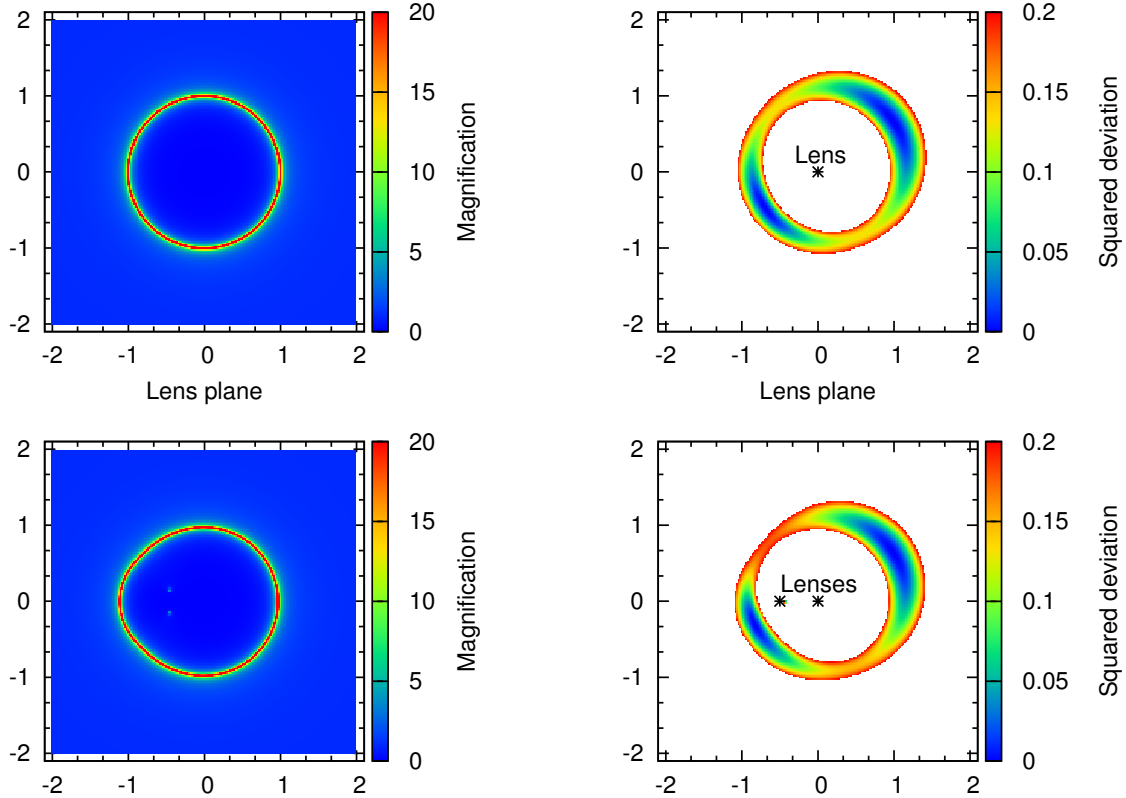
Figure 2.10: The magnification in the lens plane calculated for a point lens (top) and for a binary lens with a mass ratio of 1/9 (bottom). The points of infinite magnification form the critical curve and can be mapped to the source plane. Right: the squared deviation of the lens equation is plotted, illustrating the shape and position of the given images. For finite sources, the squared deviation corresponds to the squared source star radius $r_\star$ in units of Einstein radii.

where the coefficient k_{sec} depends on the origin of the correction which, in turn, depends in the Schwarzschild metric on the size of the Einstein radius. For theories introducing a constant offset like TeVeS, a second different relation can be used:

$$\beta_j' = \theta_j' - \frac{\theta_j'}{\theta_i'^2 + \theta_j'^2} - k_{\text{off}} \frac{\theta_j'}{\left(\theta_i'^2 + \theta_j'^2\right)^{1/2}} \tag{2.45}$$

The correction of the magnification is determined on a grid of values where the correction term can be determined sufficiently accurately. For this purpose, $\sqrt{sqd}$ provides a measure of uncertainty: for $\sqrt{sqd} < 2 \cdot 10^{-18}$, a finite source star with $2 \cdot 10^{-3}$, and a minimal achievable separation $u = 10^{-3}$, the magnification can be determined with an accuracy of 10^{-12}. Both limits are taken into account during interpolation by fitting a bivariate polynomial to the data, as illustrated in Fig. 2.12. As the order-of-magnitude is most relevant for these estimations, all quantities are considered on a logarithmic scale according to

$$\log_{10}(|\Delta\mu|) = p_1 \log_{10}(k) + p_2 \log_{10}(u) + p_3 \log_{10}(u) \log_{10}(k) + p_4, \tag{2.46}$$
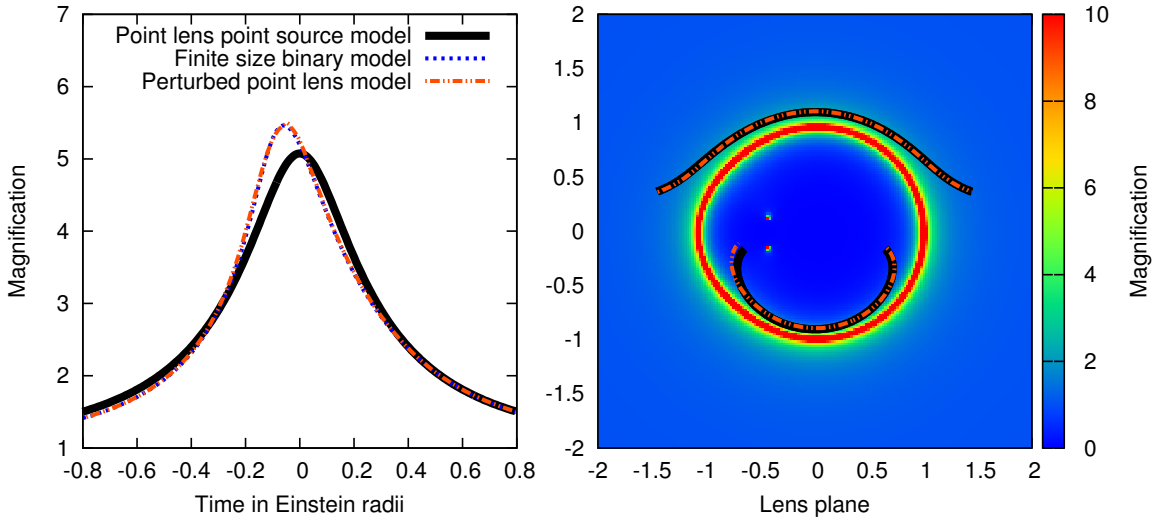
Figure 2.11: A perturbed binary light curve for a mass ratio of 0.1 and $0.5\,\theta_{\mathrm{E}}$ (left) is shown along with the corrected tracks on the lens plane (right).

and the corrected light curve is then

$$\mu = \frac{2+u^2}{u\sqrt{u^2+4}} \pm 10^{p_1 \log_{10}(k)+p_2 \log_{10}(u)+p_3 \log_{10}(u)\log_{10}(k)+p_4} \tag{2.47}$$

where the coefficients p_i are provided in Tab. 5.1 and the sign depends on the sign of $\Delta\mu$.

Coefficient	p_1	p_2	p_3	p_4	SSR
k_{sec}	-1.14355	1.96606	-0.02000	-0.67545	0.00264
k_{off}	-1.000051	1.000034	-7.7623e-06	0.0001751106	0

Table 2.1: Coefficients for the estimation of higher order effects according to Eq. 2.46.

Finally, the interpolation can be applied to estimate the detectability of such a deviation. For this purpose, we need the probability density for the separation u. Assuming a uniform distribution of u_0, the distribution of u is given as

$$P(u) = 2u, \tag{2.48}$$

on the interval $u \in [0,1]$. In combination with Eq. 2.46 the expected deviation can be calculated according to

$$\langle \log_{10}(|\Delta\mu|)\rangle = \int_1^0 2u \log_{10}(|\Delta\mu|)(u)\mathrm{d}u \tag{2.49}$$

$$= p_1 \log_{10}(k) - p_2 \frac{1}{2\log(10)} - p_3 \frac{1}{2\log_{10}(k)\log(10)} + p_4. \tag{2.50}$$

For studying events with potential second order effects, events with $k_{\mathrm{off}} > 6.3 \cdot 10^{-3}$ and $k_{\mathrm{sec}} > 2 \cdot 10^{-3}$ provide an expected deviation exceeding $1\,\%$.
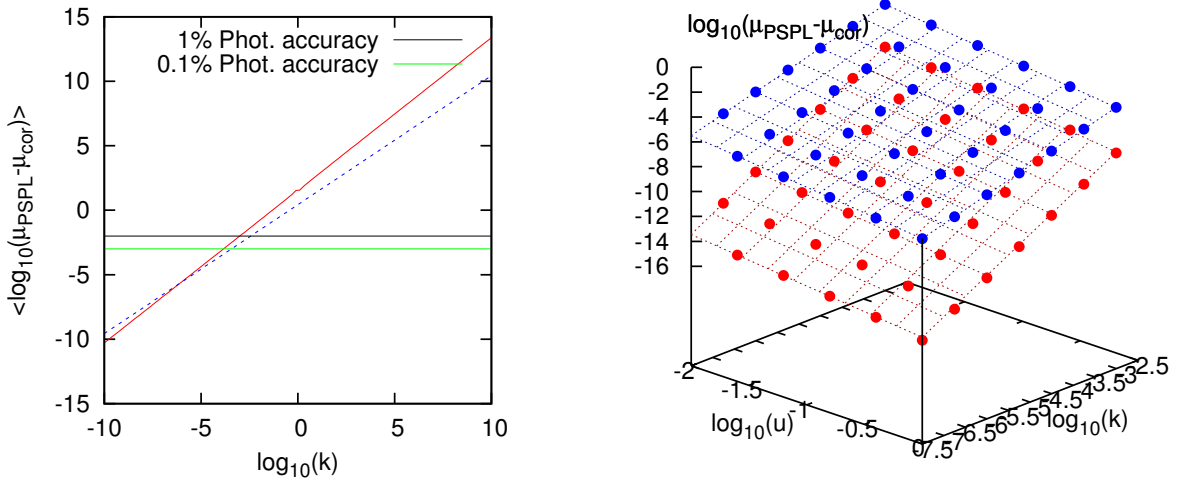
Figure 2.12: Expected magnitude of perturbation and simulation used for determining the coefficients of Eq. 2.46.

2.6   Conclusions

Gravitational microlensing is a versatile tool with a sheer variety of physical, geometrical and observational models - all seeking to overcome various limitations and approximations. The point is less whether they are currently observable but whether the requirements that have to be fulfilled to make them detectable are obtained. For this purpose, order-of-magnitude estimations provide helpful constraints in order to achieve an assessment, if a more detailed model is appropriate.

At the conclusion of this chapter, a simple test has been presented for estimating the order of magnitude of perturbations for higher order perturbations based on initial numerical simulations. The resulting power law approximation provides a tool for comparing the detectability of higher order effects. Potential theories are second order corrections as introduced by Fischbach & Freeman (1980) or the detailed version by Keeton & Petters (2005). For a constant offset, the relativistic generalization of MOND as introduced by Bekenstein (2004) offers a convenient testing ground. The numerical approaches introduced by Walters et al. (2010) could also be used for obtaining approximations. The possibility of ambiguous point source point lens models describing higher order effects can be addressed by fitting PSPL-models to the corrected light curves or by studying the information content and parameter correlations which will be shown later in this work.

Chapter 3

GPU-based models for gravitational microlensing

Before being able to infer physical parameters from observed light curves, an appropriate model is required. For fitting purposes, models need to be sufficiently precise, accurate and, for practical reasons, computationally inexpensive. Satisfying all these requirements can be challenging and overcoming these limitations often requires access to a computer cluster. This work shall illustrate how commercial off-the-shelf graphic cards can be used for modeling microlensing events, thereby significantly reducing the computational as well as the ecological burden. As introduced in preceding chapter, systems consisting of multiple lenses or asymmetrically distributed matter require numerical solutions of the lens equation. Existing strategies comprise approximative approaches, perturbative analyses, inverse ray-shooting simulations and contouring techniques based on integrating around the contour lines of images. The latter technique shall be studied in more detail and a first implementation on graphics processing units (GPUs) using the NVIDIA® CUDA™ technology will be presented.

3.1 GPU-computing

The rapid evolution of computation capabilities as predicted by Moore (1965) is restrained by the atomic structure of semiconductors. As long as further miniaturization is not achievable, the number processing units have to be increased either by using multiple CPUs or by devoting more space on existing chips for computation and less for flow control. Accidentally, such a concept exists and is used in modern graphic cards - strongly but not exclusively motivated by the consumer interest in elaborate computer games.

Graphical Processing Units act as mathematical coprocessors and consequently GPU-assisted software consists of parts executed on a CPU and parts executed on a GPU. Fig. 3.1 gives an overview of a single multiprocessor and indicates what steps have to be taken to parallelize an algorithm for running it on graphic cards. Starting from a single loop, a series of commands is executed, denoted as thread[1]; These threads are organized as blocks which in turn are arranged on a grid. Parallelization means preparing loops as lightweight threads for an execution on a grid of blocks. Depending on the thread structure, more or less registers and shared memory are needed and, considering that processors cannot arbitrarily communicate between themselves, the programmer has to optimize the code in advance for this structure. Despite having some

---

[1]in contrast to GPU-threads which are the smallest processed data elements
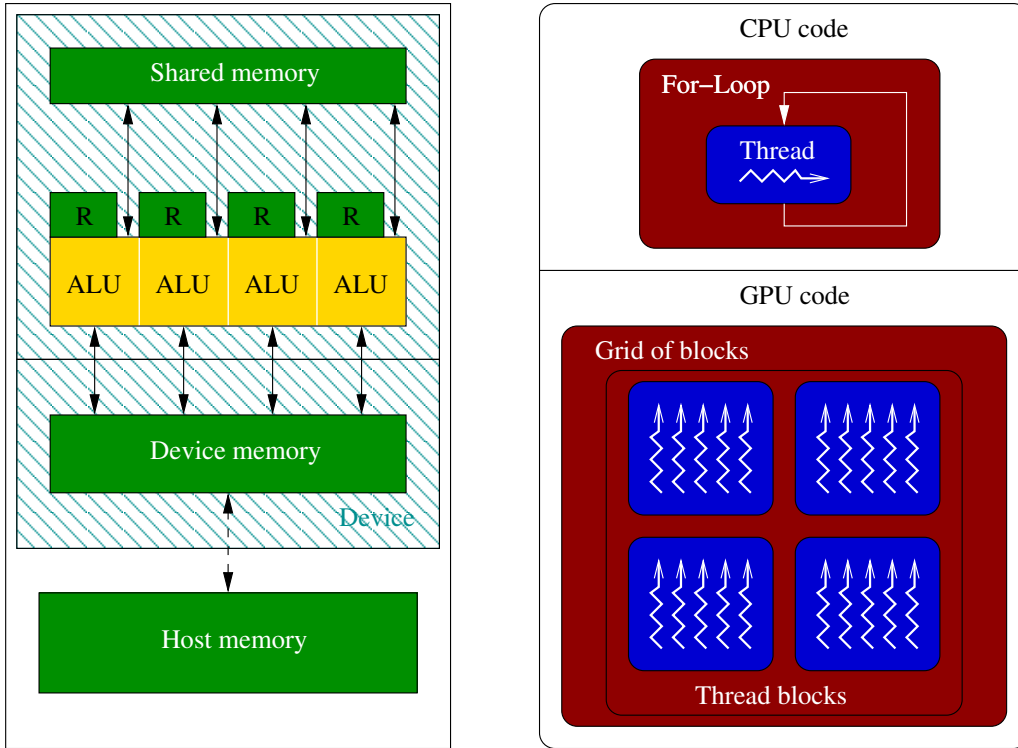
Figure 3.1: Simplified structure of a multiprocessor consisting of arithmetic logical units (ALU), registers (R) (left) and basic parallelization of GPUs.

global strategies for programming (e.g. Kirk 2010), the runtime optimization is a time-consuming effort. Despite being able to analyze the occupancy of the GPU, an implementation has to be rewritten several times by identifying the limiting quantities regarding precision and runtime. In the following, this is done for a single finite source point lens model.

3.2 The finite size model revisited

On the basis of the trivial case of a point source point lens, the next enhancement for creating a realistic multiple lens model considers the magnification of a finite source point lens, where the source star is included as uniform disk. One early model describing the light curve of such an event was developed by Witt & Mao (1994) and can be efficiently implemented on computers, because it is based on Elliptic integrals which can be quickly evaluated numerically, in principle as fast as the evaluation of a logarithm, since it can be expanded in a similar series. There are different strategies of how this approach can be optimized, e.g. by Lee et al. (2009).

For illustrative purposes and to get an understanding how precise and accurate an integration on a GPU can be made, the finite source effect of a point lens is modeled as a simple, i.e. not self intersecting, polygon as shown in Fig. 3.2. By consecutively mapping the segments of the source star contour the area A can be numerically integrated on the basis of the Gaussian trapezoidal rule

$$A = \left| \frac{1}{2} \sum_i^N \left(\theta_{1,i} - \theta_{1,i+1} \right) \left(\theta_{2,i} + \theta_{2,i+1} \right) \right|. \tag{3.1}$$

A similar contour integration approach was proposed by Gould & Gaucherel (1997). The basic implementation can be equally well achieved on CPUs but cannot compete with the analytical formulae of Witt & Mao (1994), as it relies on multiple evaluations of PSPL-solutions. Evidently, the outlined approach can be easily converted to the block structure outlined in the previous section. The integration of every point in the light curve is processed as a thread. This approach wastes some registers during the evaluation but reduces the memory transfers to the final light curve transfer. Despite the fact that the usage of the shared memory, available for each block, can be as fast as accessing registers, an exclusive usage of registers is a safe way to avoid performance losses which may arise in accessing shared memory.
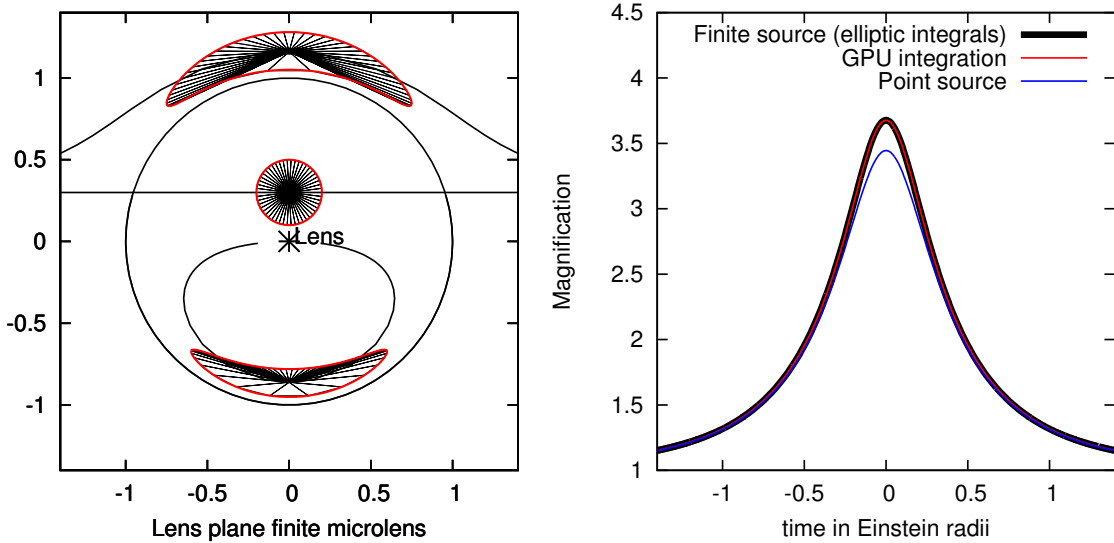
Figure 3.2: The implemented version of a finite source model calculated on GPUs is shown along with the results of the finite source model by Witt & Mao (1994), as well as a point source model. The central point of each solution is not required for the integration and is shown for indicating a non-equidistant size of integration elements.

3.2.1 Estimating the numerical uncertainties

For estimating the numerical uncertainties it is necessary to identify the limiting quantities in a given implementation. In case of the finite size model the dominant uncertainty is caused by the discretization of the contour line. Depending on its number of elements, relative and absolute uncertainties can be controlled. For obtaining an analytical estimate, the images of the source contour are assumed to be scaled versions of a circle and in that sense the effect of discretization can be approximated by comparing the area of a magnified circle

$$\mu = \frac{R^2}{r_\star^2} \tag{3.2}$$

with the area of the magnified simple polygon of the images. The discretization error δ_d for a given magnification can be estimated and expanded for $N \gg 1$ as

$$\delta_{\mu,\mathrm{d}} = \left(1.0 - \frac{N \sin^2{(\pi/N)}}{\pi \tan(\pi/N)}\right)\mu \approx \frac{2\pi^2}{3N^2}\mu, \tag{3.3}$$

In the given situation, this obstacle can be addressed by estimating the magnification ahead of time and increasing the resolution to reach the desired accuracy. The initial estimate can be achieved by evaluating a point source point lens model or alternatively by integrating a coarse grid of elements. In a second step the final integration is carried out by applying the trapezoidal integration. Comparing the results from the GPU with the achievable accuracy on a CPU for the same approach shows the drawbacks of the GPU implementation as illustrated in Fig. 3.3. An absolute precision of 10^{-4} was demanded but just a relative precision of $\approx 10^{-4}$ was achieved with the GPU approach.
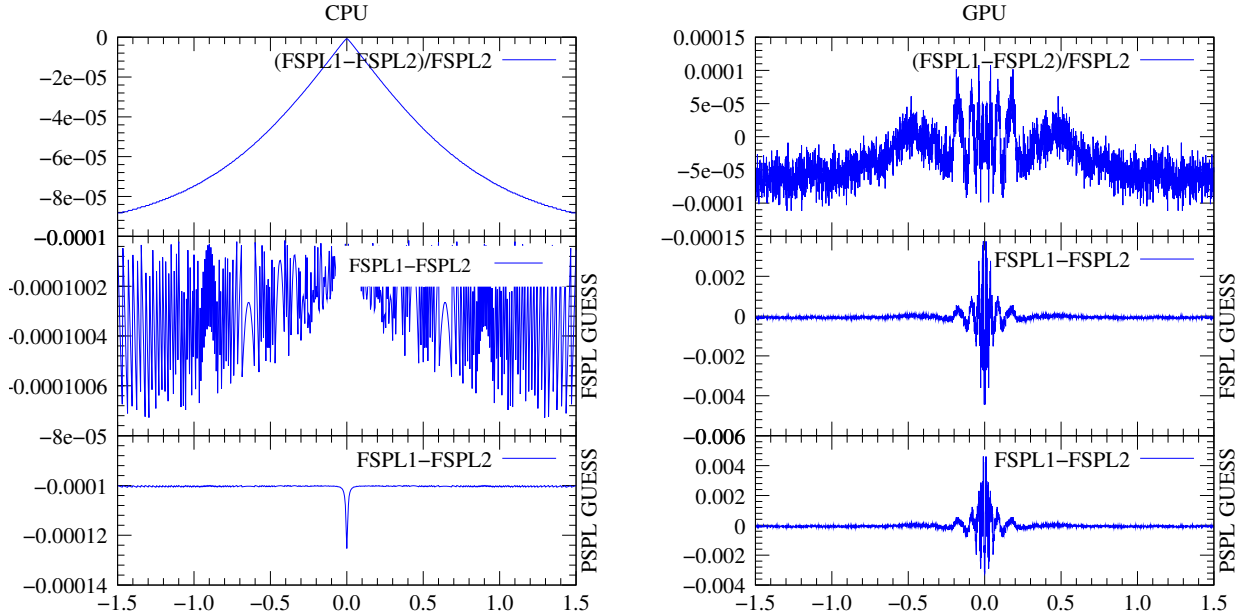
Figure 3.3: Accuracy and precision of a finite size point lens model implemented for GPUs are compared with a CPU version. The two lower panels show the absolute change in magnification of the reference model (Witt & Mao, 1994) with different initial estimates of the magnification in Eq. 3.3 and respective elements in the integration. The upper panel illustrates the relative deviation between the two finite models. The minimal separation parameter is $u_0 = 0.01\,\theta E$ and the source star radius is $0.01\,\theta_{\mathrm{E}}$.

All expressions evaluated in the model can be calculated up to machine precision[2] and for most arithmetic operations there is no simple software solution for reaching double-precision on single-precision machines, but there are techniques for compensating round-off errors in the calculation of sums. In the simulations shown before, the relative deviation was reduced by following the proposed compensation algorithm by Kahan (1965), keeping track of low-order bits by storing them in an additional variables as illustrated in Fig. 3.4. Most efforts reducing round-off errors cost computation time and cannot compete with a true double-precision implementation. The GPUs used in this thesis are operating at their maximal speed exclusively with standard single-precision variables and thus double precision variables were not used in the given implementation.

---

[2] predominantly IEEE 754 single-precision (32 bits)
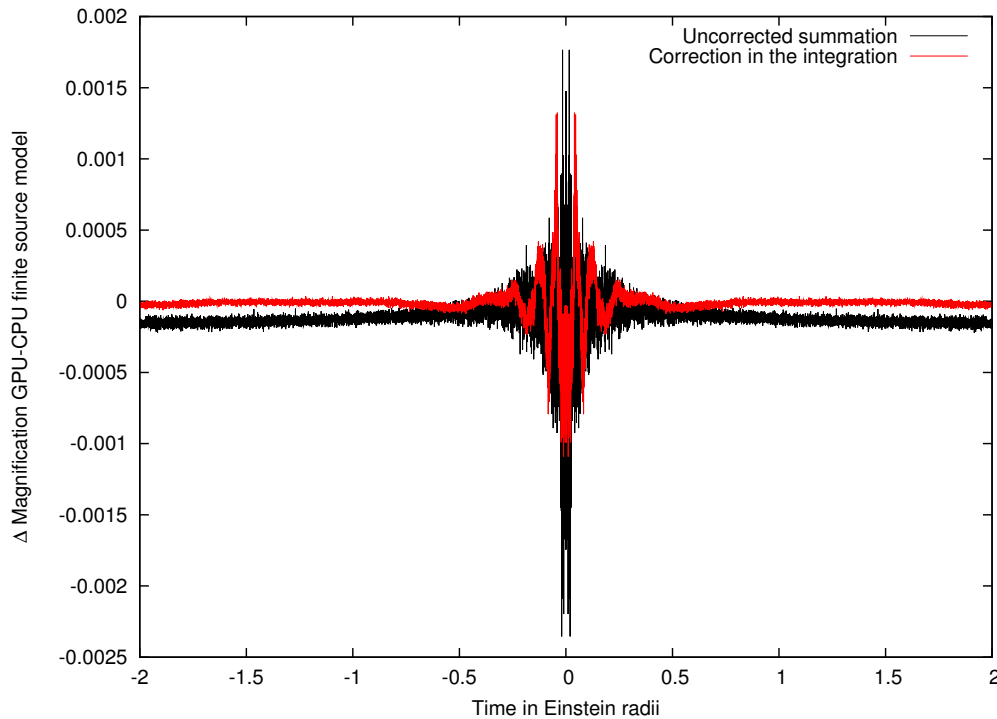
Figure 3.4: Compensation of numerical errors by using a running compensation according to Kahan (1965).

3.3   GPU-contouring for binary events

The contouring technique for calculating the magnification of a microlensing event relies on determining the image contour for a given source star contour and calculating the relative change in area expressed in angular units. For a given source star radius the lens equation needs to be solved either as complex polynomial providing the solutions as complex roots (cf. Witt & Mao 1995) which can be efficiently evaluated by using Laguerre's method (e.g. Vermaak 2000; Bozza 2010) or by using the squared deviation function as introduced by Schramm & Kayser (1987). The latter approach was used by Dominik (2007) in his adaptive contouring approach. Due to the low requirements concerning registers and the possibility of analytically calculating the first derivative, the latter approach will be followed, as illustrated in 3.5, where the images are given as lines of equidistant squared deviation and the function values themselves correspond to the squared source star radius.

gl For an efficient evaluation of the two-dimensional parameter space, the lens plane was parameterized by using two polar coordinate systems centered on each lens. This improves the accuracy, as the search direction is nearly orthogonal to the contour line and thus the residing deviation is small. Both grids are separated by a normal line at the center of mass. For each angular segment, an independent radial root-finding search is carried out, implemented as one thread. The starting point for each search has to be chosen sufficiently away from the singularity but also close enough to avoid missing a solution. In our implementation the starting point was chosen to be one source star radius away from the singularity, which readily fulfills both constraints. The polar parmeterization is not only useful for the contouring approach, but also for simulating light curves via ray-shooting as shown by Bennett (2010).

Squared deviation on a polar grid
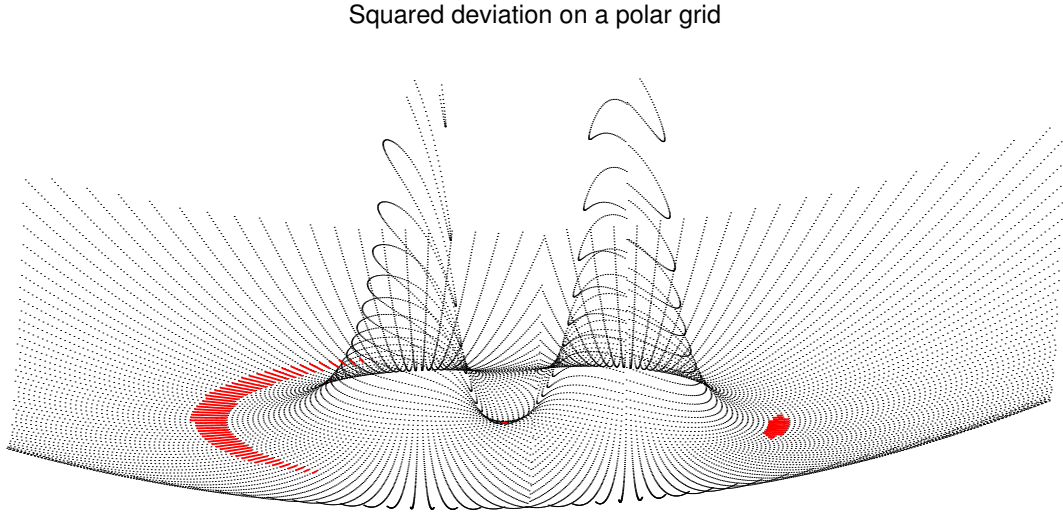
Figure 3.5: Polar parameterization of the lens equation as squared deviation function (cf. Schramm & Kayser 1987) for source stars located outside of the caustic.

3.3.1   Root-finding techniques

For obtaining an adequate numerical solution, an appropriate root-finding technique needs to be selected (Press, 1994). By selecting the contouring approach, techniques for finding the roots of a polynomial can be rejected. The analytical derivative offers the possibility of quickly probing the parameter space. Consequently, the first search is based on Newton's method and stopped if a given accuracy is reached, namely $5 \cdot 10^{-7}$. The second solution is found by using a bracketing technique which needs an initial pair of points bracketing the root. The initial points for bracketing the second solution are determined by using a parabolic approximating of the underlying function. The numerical accuracy limits for all roots need to be larger than 10^{-6} to ensure convergence. This can be directly understood as a consequence of single-precision variables, as solutions are located around $\approx 1\,\theta_{\mathrm{E}}$ and 7–8 digits are available.

Another advantage of a polar parameterization is a simple stopping criterion for searching the parameter space. A change in sign of the first derivative of the squared deviation serves as breaking condition, if no root has been detected. Although it is often desirable that all threads execute the same number of operations, stopping all threads in a block permits processing the next one. As the complicated root-finding cannot be unrolled further, it is safe to skip this option and break the execution of threads. Nevertheless, one can also set a limit for the maximal number of iterations which can be used as additional breaking criterion. For each grid the squared deviation decreases by $\approx r^{-2}$, where r is the distance from the grid center. Newton's method changes the step size by $2^n r_\star$ and thus convergence can be achieved in 20 iterations. The bracketing approach follows a similar convergence rate. In some cases, minor images are located far away from the lens, requiring more steps to be taken before discarding a radial direction.

Lens plane in Einstein radii
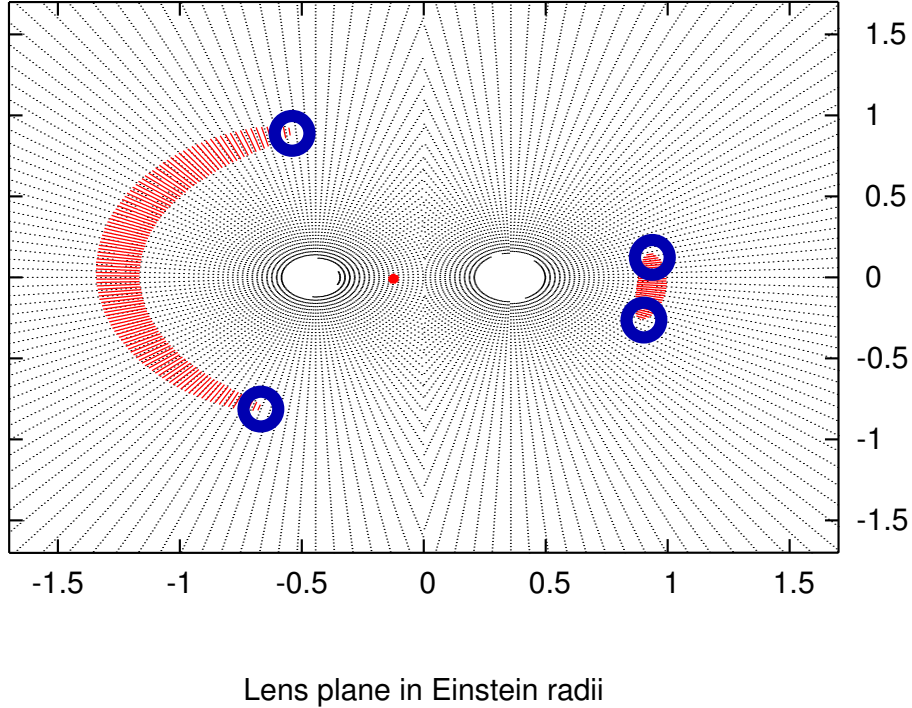
Figure 3.6: Precision control is achieved by rescaling the number of integration elements after a first search. Due to the independent evaluation of each radial direction at the end of each angular grid the total area underestimated.

3.3.2 Precision control for binary models

The numerical uncertainties of the binary model can be described by analogy with the finite size model in Eq. 3.3 providing the number of integration elements

$$N \approx 2.6 \sqrt{\frac{\mu}{\sigma_\mu}}, \tag{3.4}$$

where σ_μ is the requested accuracy. For most ground-based observations without adaptive optics and lucky-imaging cameras, a photometric accuracy of 1 % is achieved and thus the model needs to be good enough to provide results which are at least one order of magnitude below the photometric accuracy. This implies, that at least 100 integration elements have to be used for lensed events with $\mu > 1.34$ corresponding to a source-lens separation of $1\,\theta_{\mathrm{E}}$ and for high magnification events with $\mu \approx 200$ the requested number of elements should be > 1000.

In contrast to the single lens model, two modifications have to be taken into account, namely the finite accuracy of the root-finding without compensation of round-off errors, which is $\sigma_R = 10^{-6}$, and a loss of area caused by the segmented grid as depicted in Fig. 3.6. The former issue can be addressed by propagating the numerical error in radial direction under the assumption of

an uncorrelated and normally distributed σ_r which gives

$$\delta_{\mu,\mathrm{dr}} = \frac{2\sqrt{\mu}N\sin^2\left(\pi/N\right)}{\pi r_\star \tan\left(\pi/N\right)} \approx 2\sqrt{\mu}\frac{\sigma_r}{r_\star}, \tag{3.5}$$

approximated for $N \gg 1$. For low magnification events $\mu < 100$ one can reach the desired precision only for $\sigma_r < 5 \cdot 10^{-8}$ or by imposing a lower magnification limit. This is still feasible for 32 bit single-precision providing 7–8 digits if precision, as long as the image positions stay around $1\,\theta_{\mathrm{E}}$. For image positions around $10\,\theta_{\mathrm{E}}$, effectively one digit is lost for the residual deviation.

Finally, an estimate for the upper limit of the missing area can be given, because the maximal number of solutions is known to be five and if one complete segment can be missed on each side the uncertainty becomes

$$\delta_{\mu,\mathrm{d}\varphi} \approx \frac{10}{N}\mu. \tag{3.6}$$

The last component is a delicate one, as it is not obvious what the true size of the missing component really is. For modeling purposes, it is desirable to stay at least one order of magnitude below the scatter of datapoints. In the case under consideration, this means a relative precision of 10^{-3} is required with at least 10^4 segments for each integration.

In contrast to the finite source point lens model, the number of elements cannot be arbitrarily increased as long as the number of integration elements decays for solutions located far away from the lens positions. the precision decreases with a decreasing number of segments. If the numerical uncertainties of the model vary systematically with model parameters, the numerical and systematical uncertainties act as a prior distribution, compromising the fit results and particularly confusing the comparison of degenerate parameter configurations.
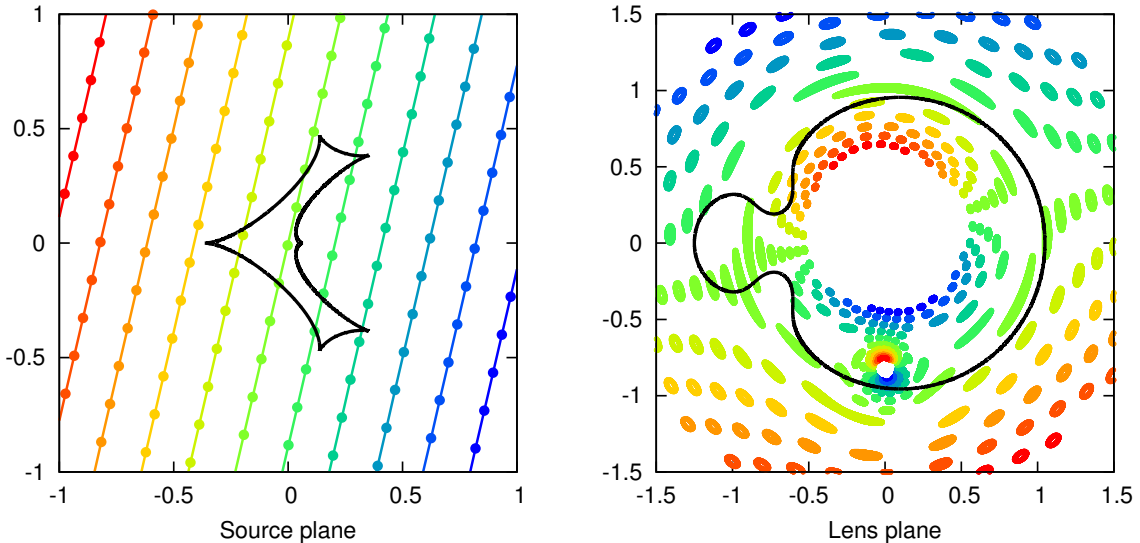
Figure 3.7: The solution structure of the lens equation is shown for different source tracks corresponding to different light curves. The binary mass ratio is 0.9 and lenses are separated by $0.9\,\theta_{\mathrm{E}}$.

All strategies aiming to conserve the precision discussed before are obsolete if complete images are missed, which is in principle possible whenever solutions are located sufficiently far away from the lens position. In order to address this problem properly, an initial grid search is carried out with a source star radius increased by a factor of two. In most cases, one can expect to find solutions at separations of the order of $1\,\theta_{\mathrm{E}}$ from the nearest lens as depicted in Fig. 3.7; the image positions for different source tracks stay around critical curves. For a main sequence source star with an angular radius of $10^{-3}\,\theta_{\mathrm{E}}$ solutions will be missed if they are located more than $5\,\theta_{\mathrm{E}}$ away from the closest lens which would also mean a separation of the source of $\approx 5\,\theta_{\mathrm{E}}$. Widely separated lenses, i.e. $d \gg 1\,\theta_{\mathrm{E}}$, can equally reduce the chance of detecting all solutions and thus the initial source star radius is also scaled with d as long as $d > 1$.

For optimally exploiting the available resources, an initial grid search is carried out with 16384 elements. After transferring the angular position back to the CPU the total number of segments is scaled to 16384 and contiguous solutions are send back to the GPU for refinement. The refinement is limited to a total number of 1024 elements as numerical uncertainties start affecting the results. The maximal precision is not reached as there is an overlap of detections in the initial search due to the scaled source star radius. The improvement reached by a refined integration is illustrated in Fig. 3.8. For a well sampled light curve it is even possible to recycle some segments as adjacent source positions provide adjacent solutions, as shown in Fig. 3.7.
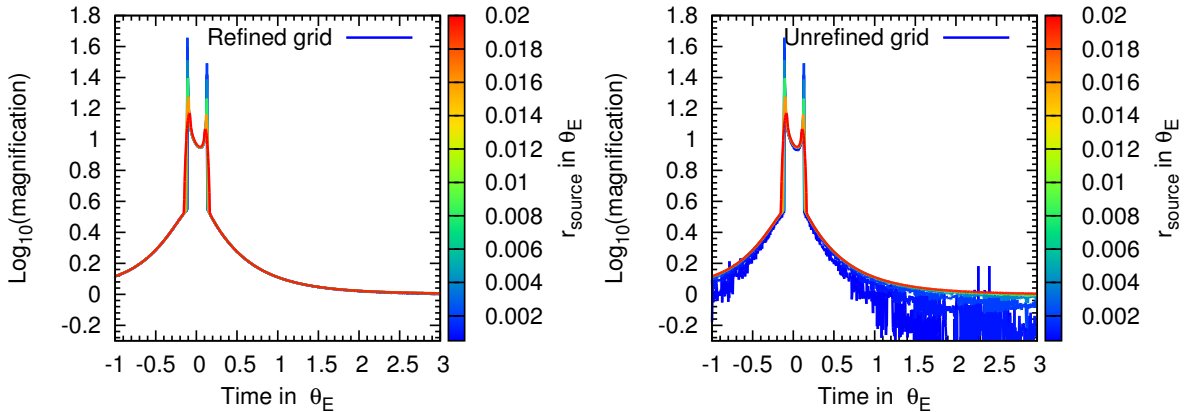
Figure 3.8: The disadvantages of a grid with fixed resolution are illustrated for source star radii between $5 \cdot 10^{-4}$ and $3 \cdot 10^{-2}$.

3.4 Inverse ray-shooting

For solving the lens equation Eq. 2.22 the number of solutions and the corresponding magnifications have to be determined. In order to compare the results of the new GPU-contouring approach, a different technique is used as a fiducial model. One successful approach is based on the inverse ray-shooting technique as introduced by Kayser et al. (1986) and applied to planetary microlensing events by Wambsganss (1997). The magnification map for a given configuration of mass ratio q and separation d is created by inserting a grid of rays into the lens equation and deflect them to the source plane. Each pixel in the source plane coincides with a given source position and the number of hits per pixel is proportional to the magnification.

The basic steps that have to be taken for simulating magnification maps are illustrated in
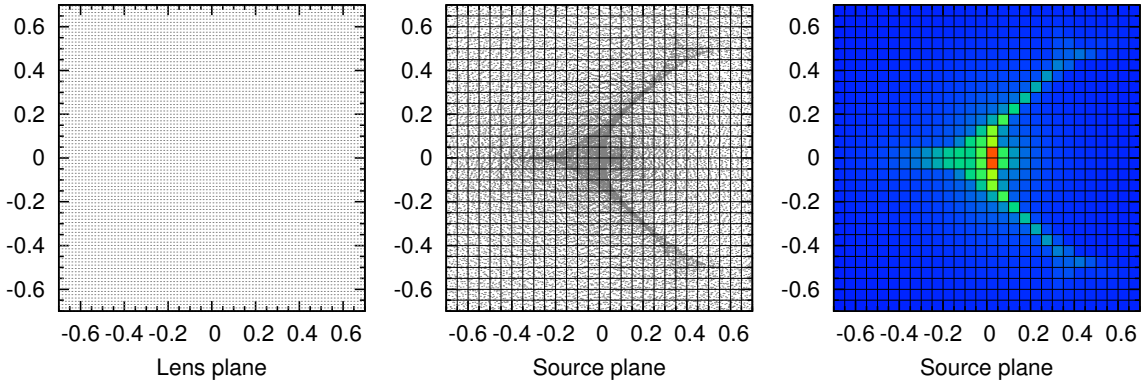
Figure 3.9: A grid of rays is on the lens plane (left) is deflected after jittering the ray position it is deflected according to the lens equation (middle) and binned to a grid of pixels or apertures for calculating the magnification(right).

Fig. 3.9. The lens plane is covered by rays probing the lens equation and in order to prevent aliasing in the source planes, each ray is jittered within a given pixel. The computational burden for calculating random numbers can be decreased by using a pre-calculated grid of random numbers which is restarted at a random position. For each target pixel in the lens plane, the relative numerical uncertainty $\delta_{\mu,\mathrm{rs}}$ can be estimated assuming the deflected rays are randomly distributed. If the theoretical number of hits for the true magnification is N_h^2, every ray at the border has a ≈ 50 % chance of missing the pixel. The binomial distribution tells us that the expected value will provide just half of the border elements with a standard deviation of $1/2\sqrt{N_\mathrm{h}}$ which is negligible for $N_\mathrm{h} \gg 2$

$$\delta_{\mu,\mathrm{rspx}} = \frac{2(N_\mathrm{h} - 1) + 1/2\sqrt{N_\mathrm{h}}}{N_\mathrm{h}^2} \approx \frac{2}{N_\mathrm{h}} \tag{3.7}$$

For an aperture the number of elements at the edge decreases to $2\sqrt{\pi}$ and the uncertainty is

$$\delta_{\mu,\mathrm{rsap}} \approx \frac{\sqrt{\pi}}{N_\mathrm{h}}, \tag{3.8}$$

which is not helpful as the number of hits is assumed to be constant but in praxis one cannot skip rays missing the aperture grid. In order to reach a relative precision of 10^{-3} one needs $4 \cdot 10^6$ rays per pixel and each deflection requires at least 2 divisions with 4 floating point operations (FLOP) per clock cycle. As ray-shooting requires complete pixel grids with typical dimensions of $4\,\mathrm{K} \times 4\,\mathrm{K}$ pixels just the divisions need 260 TFLOP; for a CPU capable of 0.1 TFLOP/s that means just the divisions will take 45 minutes.

3.4.1   Accelerating ray-shooting by interpolation

One of the most time consuming aspects of light deflection is the evaluation of divisions. For accelerating the computation it is helpful to use a bilinear interpolation for the deflection angle in each direction. According to Press (1994) a bilinear interpolation can be carried out with 16 multiplications and 8 additions/subtractions for a binary event, while a direct evaluation of the lens equation requires 10 additions/subtractions, 8 multiplications and 2 divisions. Under

the assumption that the latter counts as 8 regular operations, one expects to see just a slight acceleration but for the simulation of planetary systems consisting of multiple deflectors this means a significant improvement, because the number of operations stays constant for the bilinear approximation. One example is the deflection angle of an arbitrary solar system state as illustrated in Fig. 3.10.
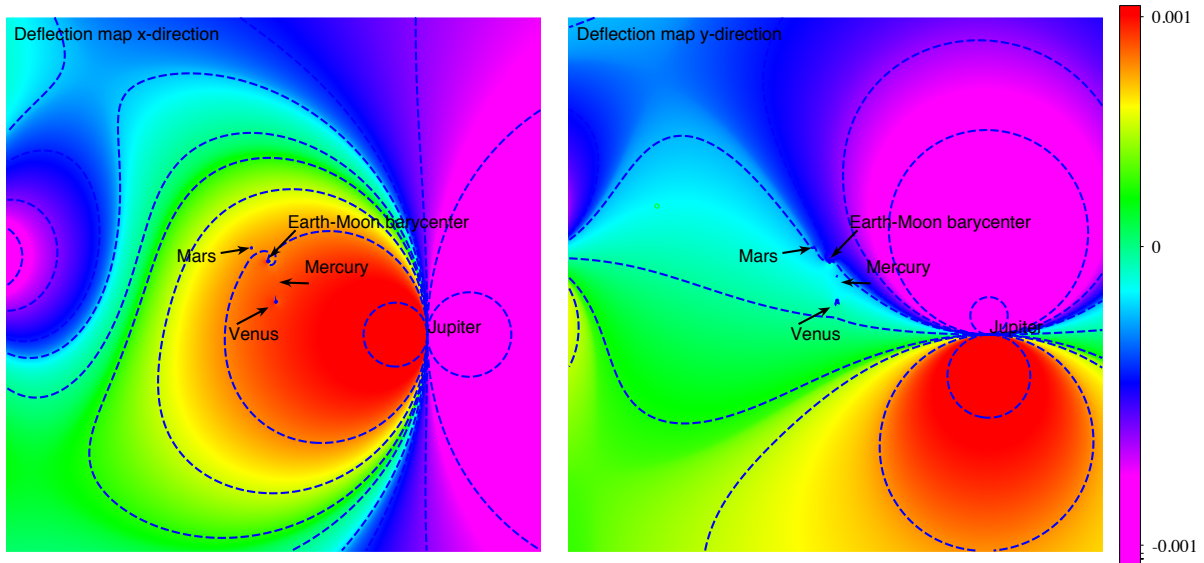
Figure 3.10: Interpolated components of the deflection angle caused by the presence of the planets in our solar system.

3.4.2 Existing implementations for graphic cards

GPU implementations of the ray-shooting technique have been shown by Bate et al. (2010) and Thompson et al. (2010) which is especially for large number of lenses a useful approach as the deflections consist of homogeneous operations. The simulation of binary events cannot directly exploit all these benefits, because the results of each deflection have to be stored temporarily and transferred back to device memory. In addition, magnification maps must be kept in the device memory, otherwise the data transfer of a typical grid of $4\,\mathrm{K} \times 4\,\mathrm{K}$ points means moving $\approx 64\,\mathrm{MB}$ from the device to host memory, i.e. the whole analysis needs to be carried out on the GPU. For a light curve where ≈ 200 points characterize and particularly exclude features the simulation of a whole magnification map is unnecessary and sophisticated optimization codes for fitting purposes exist on CPUs which cannot be used by this approach. In this thesis, a pointwise approach has been chosen, where the evaluation and calculation of the magnification for each single point is parallelized on GPUs.

3.5 Comparison

From a numerical point of view, comparing contouring and ray-shooting methods is artificial as in both cases the lens equation is evaluated for determining the magnification pattern and by definition both approaches require the same number and the same kind of mathematical evaluations for the lens equation. Both methods require an overhead of evaluations. A thorough

comparison is hard to achieve as every magnification map quickly permits the evaluation of arbitrary source tracks while inadequate magnification maps are wasting a substantial amount of the parameter space especially during a fitting routine based on a random walk. By considering the idealized approximations of the numerical uncertainties in Eq. 3.7 and Eq. 3.3 it is obvious that the uncertainties of the ray-tracing model require a factor of $\pi^2/3$ more evaluations to reach the same precision.

3.5.1   Comparing Simulations

For testing the accuracy and precision of the GPU-contouring model a ray-tracing model was evaluated for a source star with a radius of $r_\star = 0.0025\,\theta_{\rm E}$. A single magnification map for a mass ratio of $q = 0.5$ and a separation of $d = 0.7\,\theta_{\rm E}$ was created by using the ray-shooting method. More than 10^6 rays were deflected per pixel; according to Eq. 3.8, the accuracy was better than $2 \cdot 10^{-3}$. Indeed the standard deviation for the residuals $\mu_{N_h} - \mu_{2N_h}$ with $N_h = 500$ and $2N_h = 1000$ is $1.3 \cdot 10^{-4}$ and thus one magnitude below the expected accuracy of the GPU-model.

In Fig. 3.11 relative deviation and the deviation divided by $\sqrt{\mu}$ are shown which is motivated by Eq. 3.3. Statistically, the given histograms perfectly comply with the accuracy requirement 10^{-3}. Nevertheless, deviations can occur for single points of a given light curve if the solutions are located exactly on the line separating both polar search grids. For high magnifications the number of segments is limited, due to the occurring high values in the integration of Eq. 3.1.
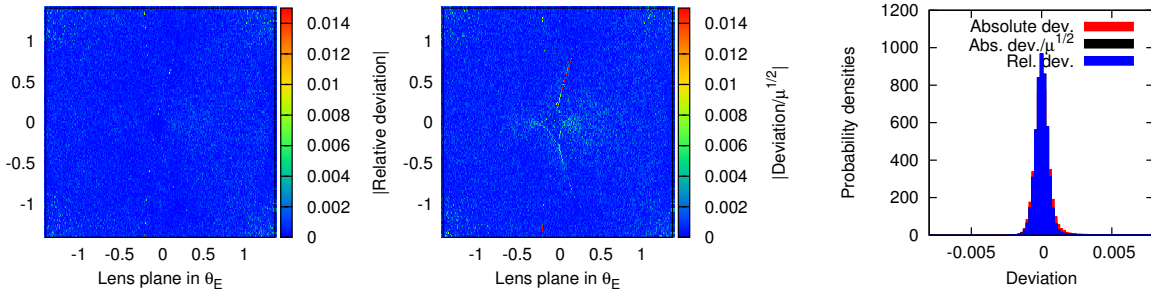
Figure 3.11:   Absolute and relative deviations between ray-shooting and GPU-contouring approach are shown for $q = 0.5$, $d = 0.7\,\theta_{\rm E}$ and $r_\star = 0.0025\,\theta_{\rm E}$ along with the corresponding probability density histogram of the deviation.

3.5.2   Runtime comparison

It is not obvious what the runtime improvement of the contouring model really is, as the solution is customized for graphic cards. Due to the possibility of rearranging the grid search after finding a first solution, one could start smaller polar sub-grids for a refinement as shown in Fig. 3.12. Moreover, most modern computers can be operated in a multithreading mode which can speed-up calculations even on a single processor and, last but not least, optimized mathematical functions are used on GPUs which are not directly accessible on CPUs, though architecture-specific compiler options could avoid some losses. Keeping all these caveats in mind, the code can be compiled as a forthright test for execution in a device-emulation mode taking into account some of the aforementioned optimizations. On a test-system[3] a speed-up by a factor of ≈ 60 has been achieved.

---

[3]Intel® Pentium® 4 CPU 3.40GHz and GeForce™ 8600 GT

Lens plane in Einstein radii
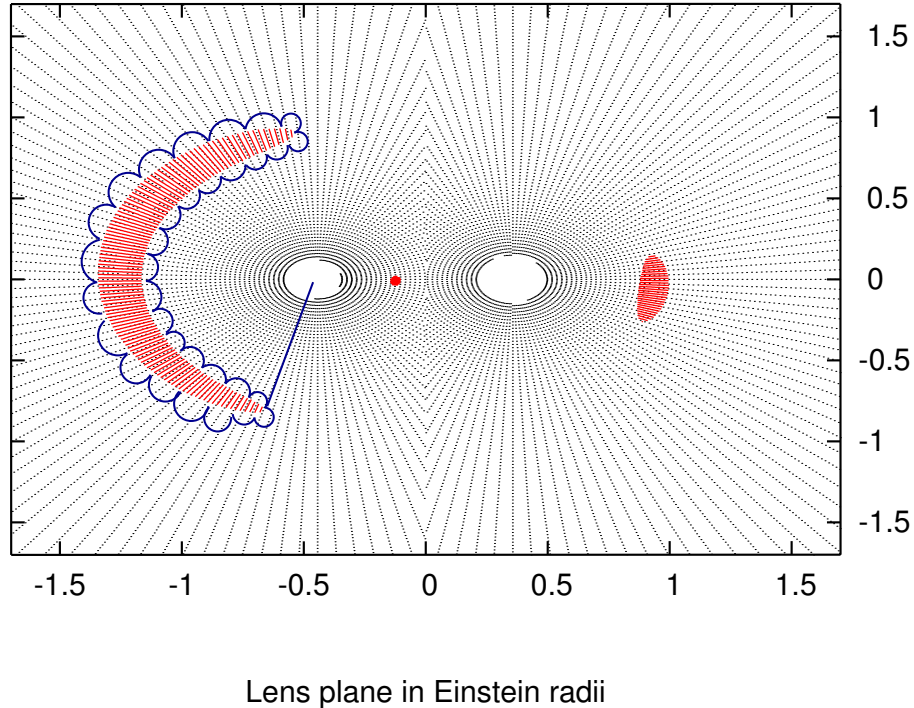
Figure 3.12: An illustration of a possible optimization of the grid search in polar coordinates for CPUs is shown for the major image. Instead of searching all angular directions in parallel, these searches are carried out until the first search provides a root and polar sub-grids integrate along the contour.

In addition, different versions of the given source code have been tested on three different graphic cards for a test configuration. The comparison of the GPU-contouring code with existing contouring implementations was based on the assumption that an integration with a relative precision of $10^{-3}\,\theta_E$ can be achieved. For the code presented by Dominik (2007) it is possible to simulate 1000 points of a caustic crossing light curve with an accuracy of $5 \cdot 10^{-3}$ on a CPU[4] in less than 3 minutes, albeit depending on the magnification and the given parameters. The refined GPU-contouring code requires 10 seconds on the same test-system at the cost of absolute precision but with a dependence on the magnification at a given point.

Comparing the performance of the code on three different test systems (Table 3.1) reveals that the refined solution runs particularly slow on standard GPUs due to the bandwidth limitations between GPU and CPU, while it benefits strongly from the specifications of a dedicated system where the initial analysis communicates quickly and accelerates later calculations without loosing performance in first place.

---

[4]Intel® Pentium® 4 CPU 3.40GHz

Testing environment	CPU	GPU	Number of processors	Runtime unrefined (s)	Runtime refined (s)
Laptop	Intel® Core 2 Duo 2.00 GHz	GeForce™ 9400 M	16	10	20
Desktop PC	Intel® Pentium® 4 3.40 GHz	GeForce™ 8600 GT	32	2.0	9.1
Tesla system	Intel® Xeon® E5520 2.27 GHz	TESLA™ C1060	240	1.7	1.1

Table 3.1: Runtime of the refined and the unrefined binary model for a light curve consisting of 1000 points.

3.6   Conclusions

GPU-assisted models for simulating binary events are an inevitable result of the ongoing miniaturization of computer hardware components. I have shown that a contouring code can run faster on existing graphic cards with a predictable runtime of the simulation but by giving up absolute accuracy for highly magnified source stars. A full double precision simulation can be used for compensating inaccuracies.

There is still a list of improvements that can be made for improving the precision of the model:

- the refinement can be started between two adjacent solutions where an additional bracketing can be carried out;

- for the first and last entry an additional sub-grid can be started;

- the required memory operations for refining the grid can be kept on the GPU and just the number and positions of the solution can be send back to the host;

- limb darkening approximations can be implemented as a grid of magnifications for different source star radii weighted according to a chosen limb darkening law; and

- if a crude solution structure has been found a ray-tracing scheme can be applied to refine the solutions

The ongoing development of new graphic cards enabling double-precision calculations without undermining the code acceleration is prepared to address issues which were not within the scope of this thesis. For points of a light curve requiring high accuracy it is probably a convenient approach to merge existing codes and switch between these solutions. Ways how to decide which points can be neglected and which points have to be kept are discussed in the next chapters.

Chapter 4

The Information Content of Light Curves

Applying the methods of information theory to gravitational microlensing can be rewarding for a variety of reasons. Even Einstein's field equations can be motivated by using an information theoretical approach, as has been shown by Frieden (2004). In the scope of this thesis, a different aspect is addressed. Global telescope networks following-up ongoing anomalous microlensing events have to decide how to distribute their observing time among different events. Naturally, these networks want to follow a strategy for detecting and characterizing as many planetary events, as well as possible. Other strategies are conceivable, such as the search for deviations from an underlying Schwarzschild metric. Planning ongoing observations requires an understanding of the worth of each data point with respect to the physically interesting parameters.

When the measurement is finished it can be helpful for the analysis to select the most important data points for characterising the system and reducing the computational burden, but how does one define what is important? All these complications, outlined by Dominik (2009), can be addressed by modeling the information content of quantities of interest in event prioritization systems and selecting the most valueable data points.

4.1   Shannon's and Fisher's definition of information

The term "information" expresses the gain in knowledge with respect to a certain quantity. Evidently, information can be characterized in different ways depending on the context where it is used. In the natural sciences, there are at least two convenient approaches for measuring and characterizing the information content in a given system. One of these methods was introduced by Shannon (1948) and is intended to measure the information content based on the frequency of measured values. The corresponding information I_μ for such a frequency provides the information content for each measurement μ as

$$I_\mu = -\log_2(p_\mu), \tag{4.1}$$

where p_μ is the chance of observing the magnification μ. The unit of Shannon's information definition in Eq. 4.1 is given in bits, being a consequence of its original application to describe human communication by means of telegraphy or digital computers. By choosing a basis of 10, the information can be expressed in digits, which is appropriate regarding scientific measurements. In Shannon's theory, measuring a microlensing light curve can be interpreted as a communication

process between nature and an observer, transmitting knowledge of the physical lensing properties encoded as change of brightness. Because the total information content is calculated by assessing each individual observation, i.e. neglecting that observations belong to one light curve, Shannon's entropy gives a global measure for the information content. Consequently, one can assess the information content by studying the frequency of observable values or deduced quantities following the maxime: the more frequent a value occurrs the less informative it is.

As an initial choice, the event magnification μ is considered as the observable quantity, which is achievable as long as baseline and blending flux can be determined. If single brightness measurements are regarded as carriers of information, prior information about the expected light curve shape is discarded; especially our knowledge about the parameter values. The latter issue is addressed by Fisher's definition of information (Fisher, 1922), which is requires a model describing an observed light curve. Given the likelihood $\mathscr{L}$ and a parameter vector p_i of an observation, the information content can be calculated according to the Fisher matrix:

$$F_{i,j} = \left\langle \left(\frac{\partial \log(\mathscr{L})}{\partial p_i} \right) \left(\frac{\partial \log(\mathscr{L})}{\partial p_j} \right) \right\rangle. \tag{4.2}$$

Assuming the measurement of some quantity subject to statistical and systematic fluctuation, it is obvious that the more measurements are taken, the higher the total information content will be. Furthermore considering the information instead of uncertainties is beneficial as the total information is the simple sum of the information content of each point. The Fisher-matrix offers a simple interpretation, using estimates of uncertainties obtained from light curve modeling: the higher the determined uncertainties are, the less informative the observation is. For this interpretation a connection has to be found between parameter uncertainties, correlations and the Fishers matrix in Eq. 4.1. This relation is given by the Cramér-Rao inequality (Radhakrishna Rao, 1945),(Cramér, 1946) which provides a lower bound for the variance-covariance matrix $\boldsymbol{C} = C_{i,j}$ assuming unbiased estimators:

$$\boldsymbol{C} \geq \boldsymbol{F}^{-1}, \tag{4.3}$$

where the difference between these matrices is a non-negative definite matrix. The equality in 4.3 is valid if and only if the underlying likelihood is Gaussian. A derivation of Eq. 4.3 is given, for instance, by Cover (2006). The inequality itself is a consequence of the Cauchy-Schwarz inequality for the given definitions under the assumption of an unbiased estimator of the Fisher matrix shown in Eq. 4.1.

4.2   Light Curves as language

Astronomical light curves are inspected by human observers to assess by eye if an ongoing microlensing event exhibits anomalous behaviour. In that sense, light curves can be compared to sentences in a regular language where every brightness measurement is considered as a character. The information content for any given magnification can be calculated according to the definition by Shannon (1948), where frequently occurring magnification values are supposed to be less informative than less frequently ones. It is also possible to consider combinations of consecutive measurements instead of single measurements, but it is pointless to use consecutive measurements without including the time interval between these measurements. As the latter remark is genuinely part of Fisher's information definition, the effect of different sampling rates will be discussed there. The frequency of possible magnification values depends on the number of possible

light curves. The corresponding probability density for the lens-source separation u is

$$P(u) = 2u, \tag{4.4}$$

for a unifrom distribution of u_0 and $u \in [0,1]$. This can be translated to a distribution of magnification values for perfectly covered light curves. For unblended events, it is possible to determine directly the magnification and solve for the separation u. As each measurement is done with a given finite accuracy Δu, there is a limited number of accesible magnification values. These magnification bins are the characters building the microlensing language. The probability of detecting a value around $u \pm \Delta u$ is given by

$$p(u) = 4\Delta u u \tag{4.5}$$

and the corresponding information content is given by

$$I_S(u) = -\log_2(\Delta u u) - 2 \tag{4.6}$$

Each separation u corresponds to a distinct magnification μ according to

$$u(\mu) = \sqrt{2\frac{-\mu^2 + \sqrt{\mu^2(\mu^2 - 1)} + 1}{\mu^2 - 1}} \tag{4.7}$$

with an adapted value for

$$\Delta u(\mu, \Delta\mu) = u_2(\mu + \Delta\mu) - u_1(\mu - \Delta\mu) \tag{4.8}$$

and thus the value for each magnification can be obtained from

$$I_S(\mu) = -\log_2\left(\Delta u\left(\mu, \Delta\mu\right) u\left(\mu\right)\right) - 2 \tag{4.9}$$

Expressing Eq. 4.1 in digits instead of bits confirms the expected results that more significant digits are available if the uncertainty in $\Delta\mu$ stays the same. It also quantifies how observers select ongoing events: by plotting and inspecting incoming light curves, human anomaly detectors are guided by values or light curve shapes which are rare and thus carry high information in Shannon's sense. In Fig. 4.1 the information value of binned magnification measurements are shown. Due to the a priori insight that information is an inverse uncertainty, it is clear that the information content rises as soon as more magnification values can be discriminated. Furthermore, the finite size affect limits the number of possible bins. Depending on the accuracy and the finite source size, it is possible to estimate the expected information content for an ensemble of measurements. The required relation is denoted as Shannon-entropy which can be used for the geometrically constrained values of $u \in [u_{\min}, 1]$

$$H(u) = -\int_{-\infty}^{+\infty} p(u) \log_2\left(p(u)\right) \mathrm{d}u, \tag{4.10}$$

which integrates to

$$H(u) = \frac{1}{2}\frac{2\log(2) - 1}{\log(2)} \approx -0.279 \tag{4.11}$$
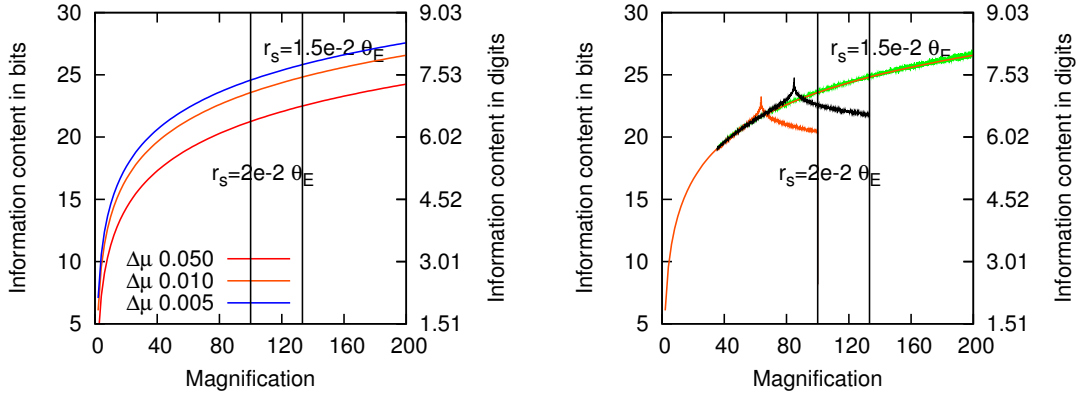
Figure 4.1: For a point lens, the information content can be calculated analytically according to Eq. 4.9 and is plotted for currently achievable accuracies (left). The number of possible "states" is limited by the finite source effect, which is numerically simulated for two different source star radii $r_\star = 1.5 \cdot 10^{-2}, 2 \cdot 10^{-2}$ (right).

but does not anticipate finite size effects and finite photometric accuracy. For this purpose it is more convenient to use the information content depending on the sampling rate

$$H(u) = -\sum_{i=0}^{N_\mu} 4\Delta u u \log_2(4\Delta u u). \tag{4.12}$$

As the maximal magnification for finite sources affects small separations $u \ll 1$, the magnification can be approximated as $\mu \approx 1/u$ and the maximal magnification due to the finite size effect can be expanded in a series $\sqrt{1 + 4/r_\star^2} \approx 2/r_\star$. Thus the number of accesible magnification bins is given by

$$N_\mu = \frac{\mu_{\max} - 1.0}{2\Delta\mu} \approx \frac{2/r_\star - 1.0}{2\Delta\mu} \tag{4.13}$$

The maximal achievable Shannon entropy for uniformly distributed magnification values is simply

$$H_{\max} \approx \log_2(\frac{2/r_\star - 1.0}{2\Delta\mu}), \tag{4.14}$$

and grows for smaller source star radii and higher photometric accuracy as the number of distinguishably states increases. Due to the rareness of high magnification events, the expected information is not sensitive to changes in the source star radius, as illustrated in Fig. 4.2.

4.2.1   Distribution of source stars

Motivated by the sensitivity of the magnification distribution, an exemplary study for a more realistic distribution of source star radii can be made. For estimating the change of information caused by the source star radius, a galactic model is required. In the following, the Besançon model
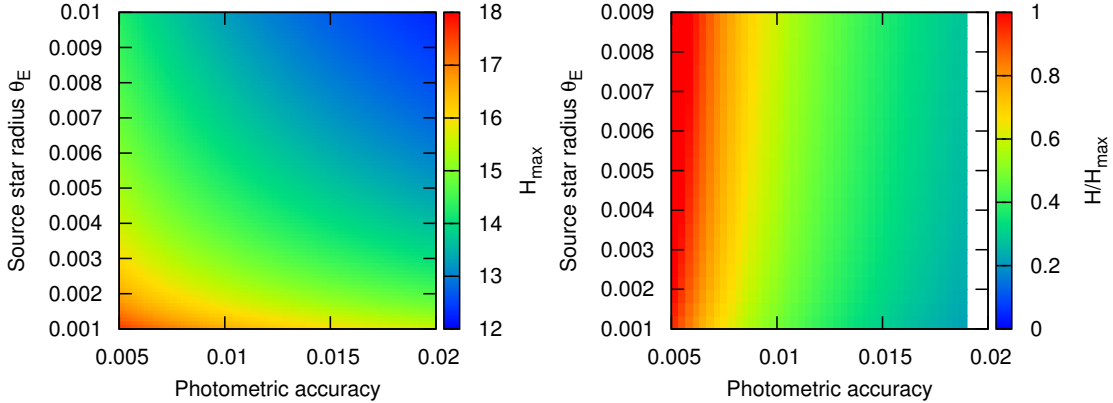
Figure 4.2: The maximal Shannon entropy depending on Photometric accuracy and source star radius illustrate the change of the information content of an ongoing microlensing event.

introduced by Robin et al. (2003) is applied for estimating the distribution of source star radii. For generating a sample of stars at a given field of view, diffuse extinction is included as a relative extinction value. Marshall et al. (2006) have introduced a three dimensional extinction distribution in the K_S-band which can be used for inferring parameters in gravitational microlensing as shown by Kerins et al. (2009). Unlike preceding studies, the Besançon model can also provide a theoretical distribution of source star radii. For this purpose, a catalogue simulation of the Besançon model is carried out in a 1x1 arcmin region in Baade's window at $l = 1.2°$, $b = -2.7°$ corresponding to the fields where Sumi et al. (2006) analyzed the optical depth. For converting the extinction A_{KS} appropriately to the visual extinction, the relation $A_{KS}/A_V = 0.062 \pm 0.005$, determined by Nishiyama et al. (2008) for the OGLE-fields has been used. The total extinction changes most rapidly in the first 4 kpc and remains constant afterwards. For the source star distributions, a diffuse extinction with $0.4\,\mathrm{mag\,kpc^{-1}}$ is used. Estimates for the given field relying on the alternative conversion $A_{KS}/A_V = 0.1$ would provide the commonly used visual extinction of $0.3\,\mathrm{mag\,kpc^{-1}}$ for fields within the Galactic plane. The apparent magnitude limit in the I-band $I < 19$ was chosen to match the limiting magnitude in the OGLE data.

The source star and lens star distributions for such a configuration were extracted by applying a rejection sampling to the simulated catalog. Fig. 4.3 shows a consistency check for the requested model with extinction rates of 0.3 and 0.4 mag kpc^{-1}. The optical depth τ is calculated for all lenses below a given value. For each lens in the catalog, all possible source stars were used for obtaining an average Einstein radius per lens. Given a source star position, multiplying the optical depth by the possible number of sources provides the number of observed source stars per distance bin, which can be understood as probability function of the source star position. Resampling from this distribution and drawing a lens star according to the probability of single lenses leads to the distribution of lens stars. The resulting distributions agree with a typical lens distance of $6.52^{+1.03}_{-0.98}$ kpc, as well as a typical source distance of $7.8^{+0.80}_{-1.91}$ kpc. The latter estimates are given as mode values and their uncertainties as differences with respect to the 16 % and 84 % quantiles. The average optical depth is $\bar{\tau} = 2.97 \pm 1.80 \cdot 10^{-6}$, which is consistent with the result

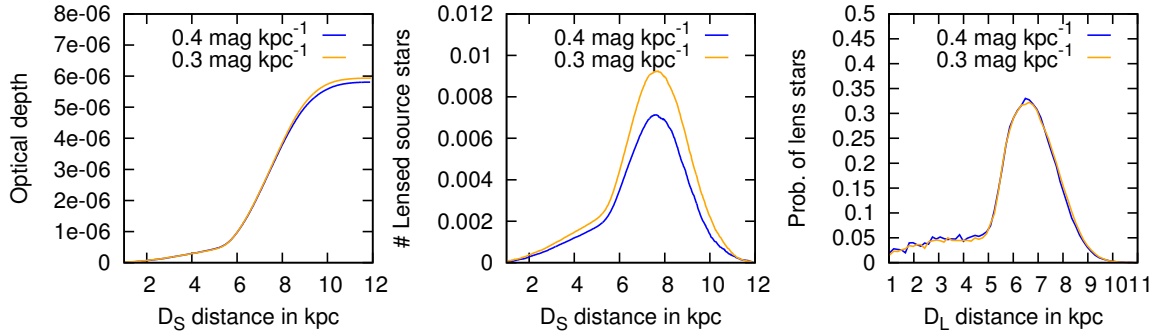**Figure 4.3:** Optical depth (left) and lens/source-star distributions depending on position, according to a catalog simulation using the Besançon model in a 0.001 square degree field pointing towards $l = 1.2°$, $b = -2.7°$

for 32 OGLE events: $\tau = 2.55^{+0.57}_{-0.46}10^{-6}$ (Sumi et al., 2006).

The given Galactic model yields more predictions about the underlying source star structure. Estimates for bolometric magnitude and effective temperature are part of the Besançon model and according to the Stefan-Boltzmann law the stellar radius can be calculated. For this purpose, a source-lens star pair is drawn from the simulated sample and converted to Einstein radii. In the preceding consistency check, the optical depth was determined assuming a magnitude limit of $I < 19$ but microlensing magnifies fainter stars beyond that limit. For estimating the distribution of source star radii, it is necessary to accept fainter stars if their maximal magnification exceeds the magnitude limit. Given that the minimal separations u_0 are uniformly separated, a larger number of source stars has to be taken into account. Since the simulation provides the Einstein radius for an accepted object, the event rate depending on the Einstein time is not required for obtaining the source star distribution. In Fig. 4.3 the expected number of source stars is shown on a logarithmic scale, indicating that a log-normal distribution appropriately describes the data

$$P(r_\star) = \frac{1}{\sqrt{2\pi}\sigma r_\star}e^{-(r_\star - \langle r_\star \rangle)^2/(2\sigma_{r_\star}^2)}, \tag{4.15}$$

where the parameters and uncertainties can be determined with a least square fit and the corresponding (logarithmic) coefficients are $\langle r_\star \rangle = -6.326^{+0.0053}_{-0.0057}$ and $\sigma_{r_\star} = 1.0074^{+0.0049}_{-0.0049}$. This result can be used for simulating finite source light curves but it also serves as prior distribution for binary events. Note that the GPU-model presented in the preceding chapter gives results with 0.1% accuracy for $\langle r_\star \rangle > -6.9$. The probability density of the source star magnitude covers the range of values observed by the OGLE survey (Udalski, 2003). The distribution of source star radii is not a smooth function, as different population of stars contribute, typically main sequence and giant stars, however the uncertainties in lens and source position combined with the stellar mass function smooth out the distribution if they are expressed in units of θ_E. Finally, the possibility of studying the source star distribution can be reassessed. Instead of providing the change in information content, it is more relevant to estimate the number of measurements, which may have been affected by finite source effects. For a source star radius of $r_\star \approx 5 \cdot 10^{-3}$ the finite source light curves will start to deviate by more than 1% for $u < 0.0179$ and thus for
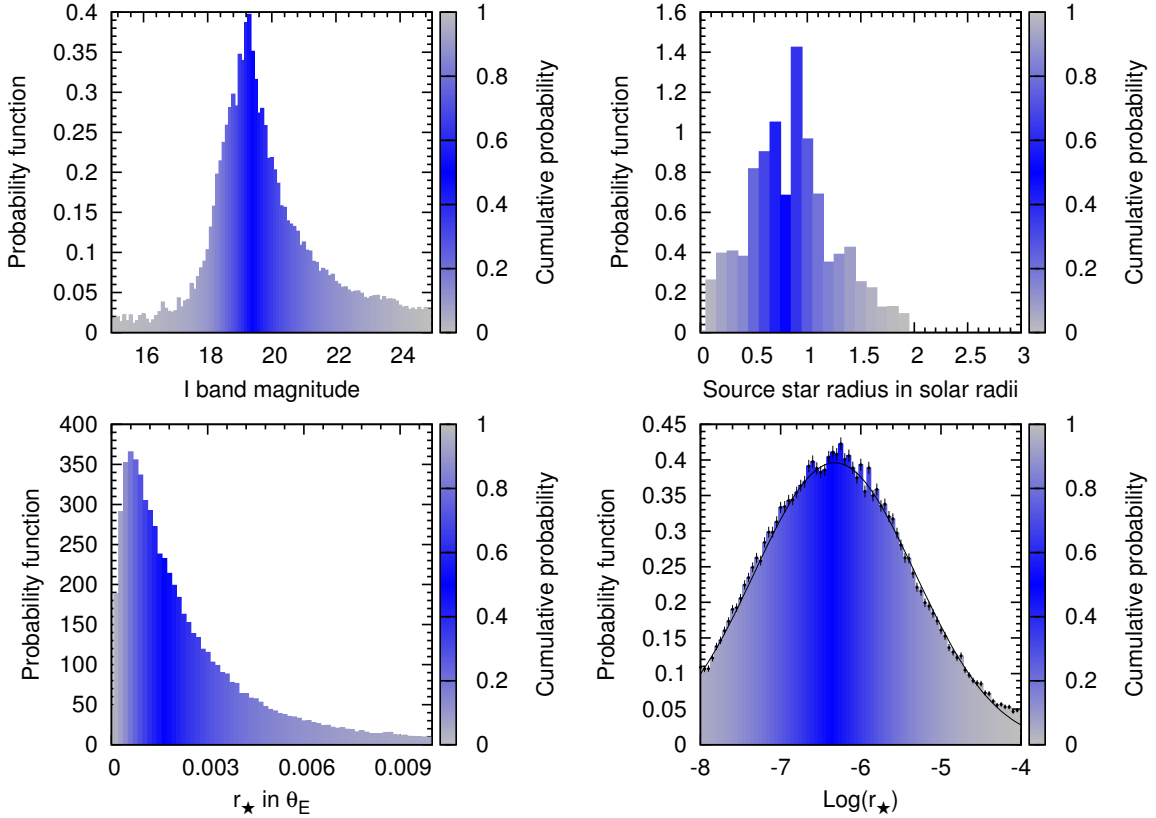
Figure 4.4: In the top row the unblended baseline magnitude of source stars is shown (left) along with the source star radius expressed in $R_\odot$. The lower row illustrates the source star distribution in units of θ_E (left) and an approximation as lognormal-distribution (right).

$\mu > 55.9$. Assuming that finite source effects cannot be reproduced by rescaling the model light curve, the distribution of relative deviations and its frequency can be studied. The results for the given source star radius distribution and $u < 0.02$ are shown in Fig. 4.5, along with a modified distribution with $\langle r_\star \rangle = -2.302$. As before, the high number of small deviations are caused by the transition from the PSPL dominated part of the light curve. The peak around 35% is caused by the local minimum of deviation where the source star radius starts to overlap the singularity. The small fraction of higher deviations is mostly informative for discriminating different source star radii distribution hypotheses. Due to the contamination by high magnification values by binary events, the prospects of such a study are limited, given that high magnification events are especially sensitive for binary events, such as planetary systems (Yoo et al., 2004).

4.2.2 Distribution of Super-Earth magnification patterns

As a final application, the distribution of relative deviations and its direct information content is shown for different planetary magnification patterns. The effect of source star radii was considered
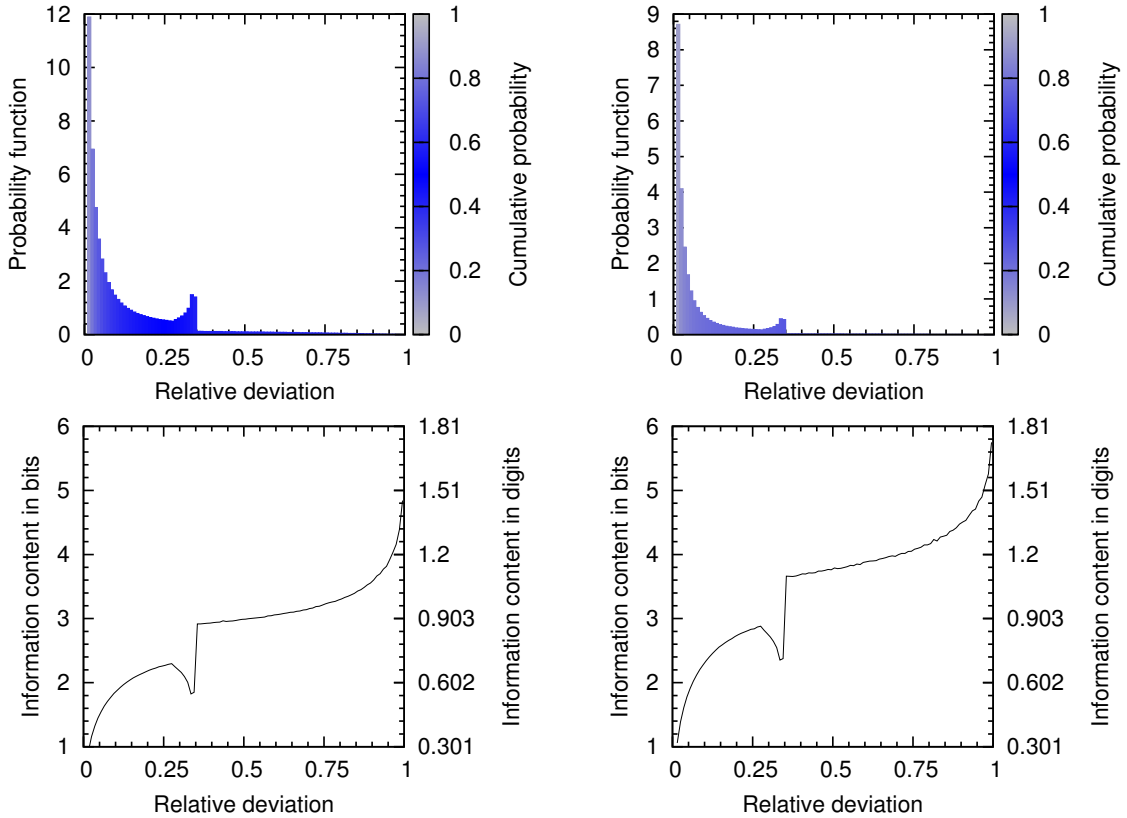
Figure 4.5: The top row shows the distribution of relative deviation values for finite source stars simulated according to Eq. 4.15 (left) and with $\langle r_\star \rangle = -2.302$ (right). The lower row illustrates the information content for the deviation considered as observable quantity.

to be independent of the sampling rate, assuming that a good coverage of observations within $1\,\theta_E$ will compensate these effects. Due to the given asymmetries and moreover the short perturbations, it is necessary to include the effects of typical sampling intervals. In the following, sampling intervals $\Delta t \in [15, 30, 60\,\mathrm{min}]$ are chosen for the simulation. In order to translate these to Einstein radii, the Einstein time is simulated using a prior distribution obtained from the Besançon model. The corresponding distribution is shown in Fig. 4.6 along with fits of a log-normal distribution with and without skewness sk.

$$P(t_E) = \frac{2}{\sqrt{2\pi}\sigma t_E} e^{-(t_E - \langle t_E \rangle)^2/(2\sigma_{t_E}^2)}, \tag{4.16}$$

where the best fit parameters are summarized in Tab. 4.1. Undeniably the given models does not fit the data as long as the whole parameter space is used for normalization. If long duration binary events occur, they would exhibit parallax effects (Hardy & Walker, 1995). In addition, their duration reduces the difficulty of properly characterizing the event. As a cut-off an Einstein-time of 100 days is used, i.e. $15\,\%$ of all events are not considered. Assuming that future microlensing surveys will be able to see 1000 microlensing events per year, only 15 anomalous events per year are not taken into account.

σ	$\langle t_{\mathrm{E}} \rangle$	$sk_{t_{\mathrm{E}}}$	$\chi^2/d.o.f.$
$3.310^{+0.012}_{-0.013}$	$1.1197^{+0.0088}_{-0.0088}$	0 (fixed)	14.3
$1.625^{+0.029}_{-0.032}$	$2.229^{+0.031}_{-0.027}$	$1.712^{+0.082}_{-0.087}$	2.7

Table 4.1: Parameters for modeling the prior distribution of Einstein times.
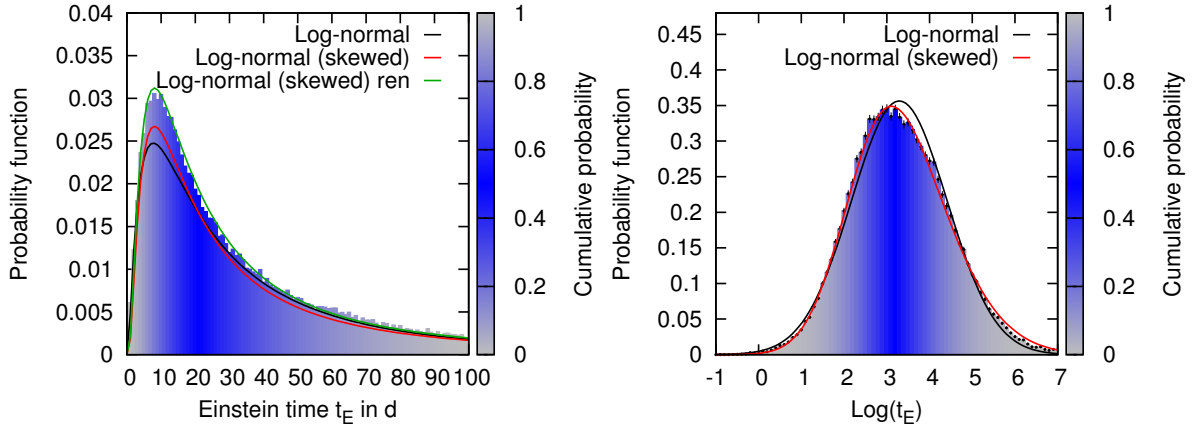
Figure 4.6: The distribution of Einstein times is modeled as log-normal distribution with and without skewness.

Depending on the binary lens parameters, a PSPL fit can partly compensate and redistribute the detectable deviation as illustrated in Fig. 4.7. Taking this into account, binary light curves have been extracted from a grid of magnification patterns. The PSPL model was fitted to each simulated binary light curves using the simplex method. If the fit did not converge, the intial PSPL model for the given parameters was assumed to be correct. For each magnification map 5000 light curves have been simulated whereas 1.6 % of all fits did not converge. The finite source star size was fixed at $\approx 6 \cdot 10^{-4}$ which is in the range of typical values. According to the prior estimate for the distribution in Fig. 4.4, 30 % of all source stars are below such a value. The detection efficiency of planetary events is affected by finite source effects as giant source stars smooth out distinct features, as shown by Bennett & Rhie (1996), but for main-sequence stars the following estimates are correct.

The resulting median $\Delta\chi^2$ values with variance one are shown in Fig. 4.8 for sampling rates of 15 and 30 min and maximal magnifications above and below ≈ 28.2. Assuming the baseline flux can be determined with an accuracy of 5 %, all values in Fig. 4.8 are increased by a factor of 400. For all light curves in the upper panel, the resulting $\Delta\chi^2$ value is around 40 and should decrease approximately linear with langer sampling intervals. The chance that these deviation occur by accident is of the order 10^{-8}. Demanding a significance level of 5 % leads to sampling intervals of $< 1\,\mathrm{h}$. If the errorbars are only known up to an order of magnitude, the resulting sampling interval for Super-Earths is $\approx 5\,\mathrm{min}$. Obviously, high magnification events are not subject to these issues, i.e. well sampled planetary events will be detected for a given main-sequence source
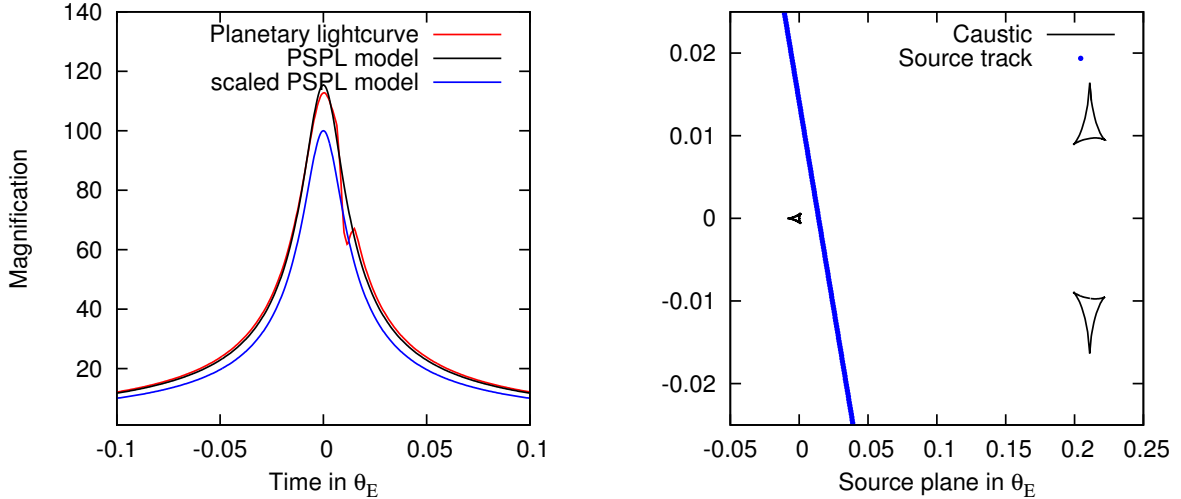
Figure 4.7: A planetary microlensing light curve (left) is shown along with the corresponding source track and caustic structure (right). A PSPL-fit to the binary model indicates the magnitude of the detectable deviation.

star (Griest & Safizadeh, 1998).

For the outlined scenario, the contribution of Earth-like planets to the detectable distribution of $\Delta\chi^2$ values and their corresponding information content can be calculated (Fig. 4.9). The logarithmic representation is a convenient choice for providing a shape-invariant distribution that can be shifted according to a given sampling rate and photometric accuracy. As expected, the information content is higher for larger $\Delta\chi^2$. In addition to the complete Super-Earth grid distribution, the high-mass end is shown separately indicating a higher contribution of larger deviations. Nevertheless the prospects of detecting such a deviation are small. The fraction of $\log(\Delta\chi^2)$ exceeding a 5 % significance level can be described by power laws depending on the requested frequency. Requesting that every fourth Super-Earth on the given grid is detected, the sampling intervals are given as

$$\Delta t \approx e^{-1.99}\sigma_\mu^{-4.96}. \tag{4.17}$$

Demanding a minimal time between two visits of $\approx 10\,\mathrm{min}$, requires a photometric accuracy of 2.6 % where the latter is achievable for most events, but the former needs to rely on an intensified coverage by follow-up teams.

4.3   Information driven sampling

After discussing the information content of observable parameters such as magnification and $\Delta\chi^2$, the worth of customized sampling stragegies for determining the event parameters can be addressed. Techniques for this purpose are commonly used for analyzing the likelihood sensitivity to parameters, but frequently the theoretical worth of optimal sampling and observation planning is not explicitly discussed. In order to obtain analytical estimates of the information content for each observed data point, an underlying model needs to be defined for the parameters denoted as p_i. Based on a given model, the effect of changing one or more parameters can be studied for the
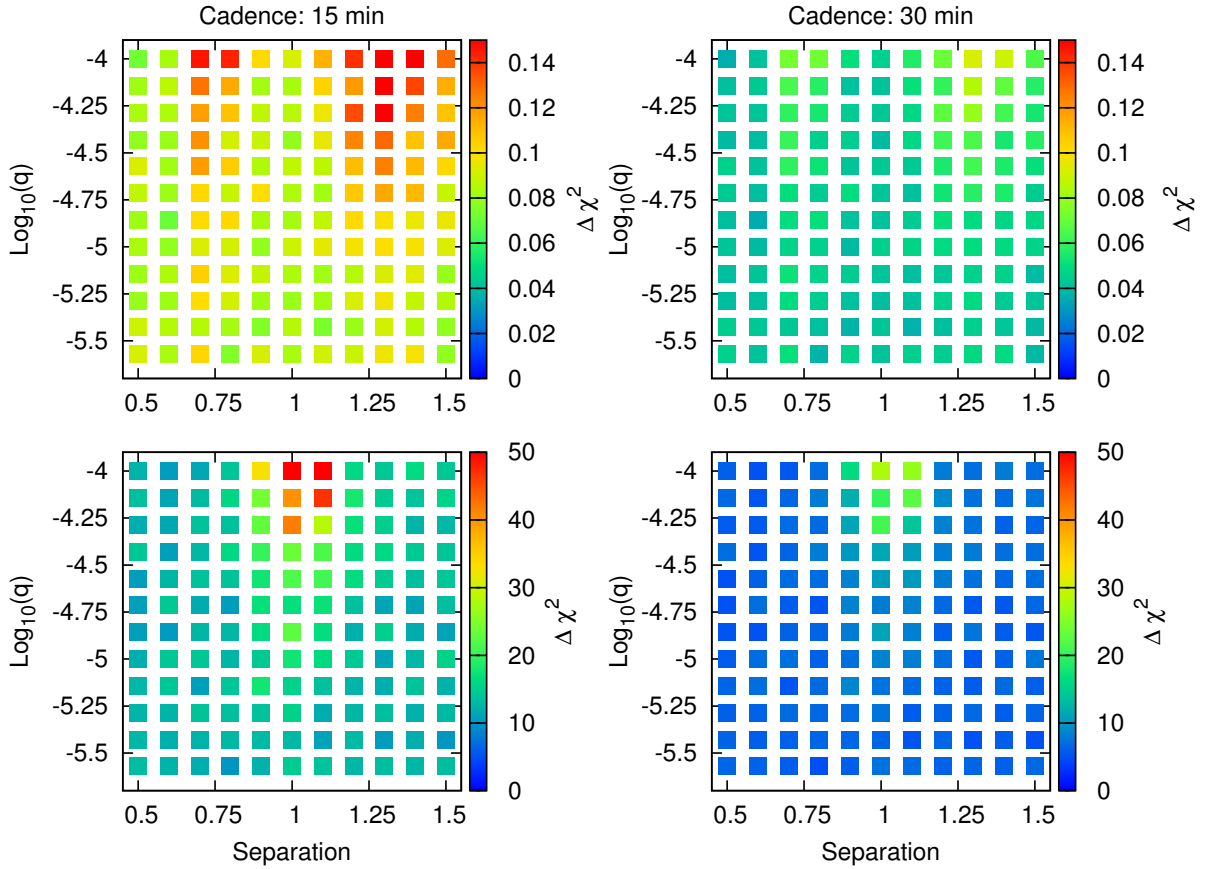
Figure 4.8: The detectability of binary microlensing light curves is shown for $u_0 \in [0, 1]$ (upper row) and $u_0 \in [0, 0.1]$ (lower row), with sampling rates of 15-20 min.

local Fisher (information) matrix as introduced in Eq. 4.1. This defintion can be interpretated in the context of a given likelihood $\mathscr{L}$ as a first order part of the multivariate Taylor expansion of the $-\log(\mathscr{L})$ surface. Qualitatively, it is clear that the uncertainties are related to $\partial \log(\mathscr{L})/\partial p_i$, as a higher rate of change in $\Delta\chi^2$ demands smaller parameter intervals in order to reach the requested significance level. For the best fit value, given as local minimum in χ^2, a difficulty arises: the first order terms vanish and thus the next order needs to be included, providing a modified version of $F_{i,j}$

$$F_{i,j} = \left\langle \left(\frac{\partial^2 \log(\mathscr{L})}{\partial p_i \partial p_j} \right) \right\rangle, \tag{4.18}$$

which is simply the curvature of the $\log(\mathscr{L})$ surface. By assuming Gaussian noise, the first order approximation Eq. 4.1 can provide a reasonable interpretation assuming that the best fit parameters are subject to fluctuations. This follows from $-\log(\mathscr{L})$ with normally distributed noise which is related to the χ^2 of N observations via

$$\log(\mathscr{L}) = \frac{1}{2}\chi^2 = \frac{1}{2}\sum_k^N \frac{(f_{\text{model}} - f_{\text{obs}})^2}{\sigma^2}. \tag{4.19}$$
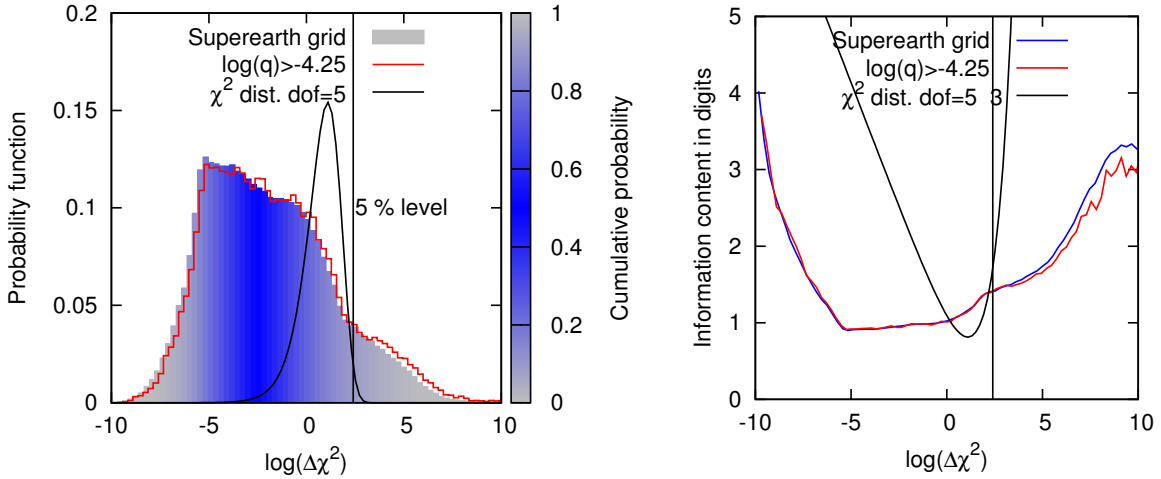
Figure 4.9: Distribution and information content of $\Delta\chi^2$ values in the Super-Earth regime.

If the model appropriately describes the data, the numerator is a random number, scaled according to a given observation. Depending on the number of observations and the sampling of these observations, the shape of the global minimum is altered as depicted in Fig. 4.10. The Fisher ma-
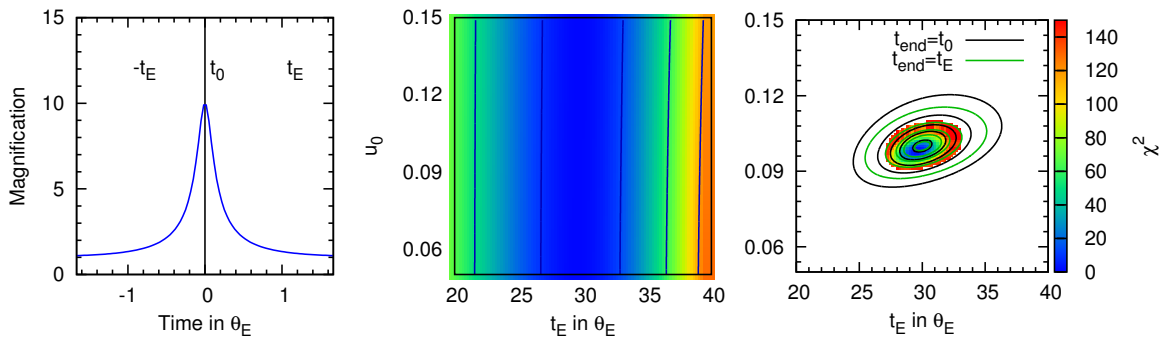
Figure 4.10: The curvature change of the χ^2-surface illustrates the change of the information content of an ongoing microlensing event.

trix itself does not depend on a given observation, which is expressed by calculating the expected value for random variables, ensuring the validity of the Cramér-Rao inequality (Frieden, 2004).

The observed value f_{obs} in Eq. 4.19 does not explicitly depend on the parameters p_i and the corresponding derivative vanishes but, in practice, the determination of best estimators depends on a given fit. For a complete measurement, different realizations should provide similar model parameters, as the likelihood is calculated for several points which includes averaging over random variables. Under this assumption, the Fisher information for a complete light curve can be obtained independent of a given realization according to

$$F_{i,j} = \left\langle \left(\frac{\partial\chi^2}{\partial p_i}\right) \left(\frac{\partial\chi^2}{\partial p_j}\right) \right\rangle \tag{4.20}$$

For the sake of clarity and without loss of generality, the contribution of a single measurement can be calculated providing the desired quantity. The observation does not change when the model parameters are altered and thus the derivative of the model remains

$$F_{i,j} = \left\langle \frac{1}{\sigma^4} \left(\frac{\partial f_{\mathrm{model}}}{\partial p_i} f_{\mathrm{model}} - \frac{\partial f_{\mathrm{model}}}{\partial p_i} f_{\mathrm{obs}} \right) \left(\frac{\partial f_{\mathrm{model}}}{\partial p_j} f_{\mathrm{model}} - \frac{\partial f_{\mathrm{model}}}{\partial p_j} f_{\mathrm{obs}} \right) \right\rangle \qquad (4.21)$$

Those terms exclusively depending on the model or on the observed value cancel each other out as $\langle f_{\mathrm{obs}} \rangle = f_{\mathrm{model}}$ for Gaussian noise. Suprisingly, the result still depends on the expected value of the variance and thus the Fisher information can be written as

$$F_{i,j} = \frac{1}{\sigma^2} \left(\frac{\partial f_{\mathrm{model}}}{\partial p_i} \right) \left(\frac{\partial f_{\mathrm{model}}}{\partial p_j} \right). \qquad (4.22)$$

This expression is analytically accesible for PSPL events and thus the local information content with respect to given parameters can be determined. Without the averaged fluctuation of different realizations, the first derivatives would vanish for the maximum likelihood estimators.

4.3.1 Information content of PSPL models
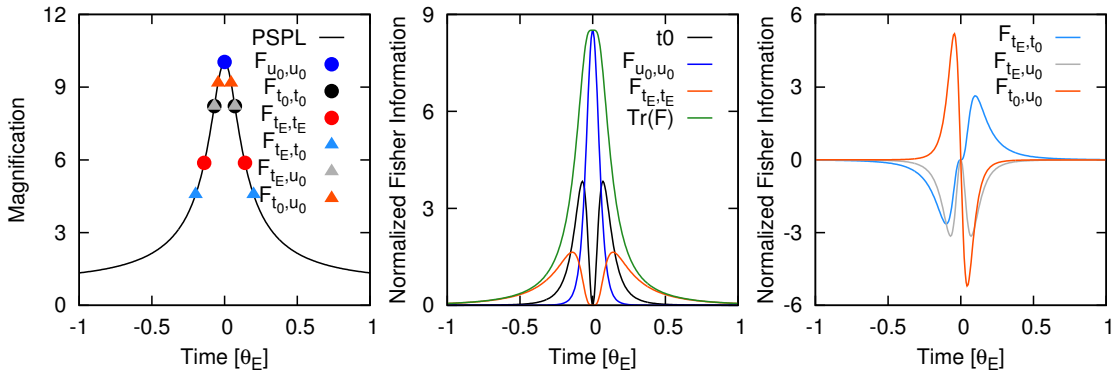
Figure 4.11: Information content according to the entries of the Fisher matrix normalized on the interval $[-1, 1]$ θ_{E}. Extremal points have been marked in the point lens light curve.

To first order, the Fisher-matrix can be used to estimate the parameter variances of a fit according to the Cramér-Rao inequality, where for the diagonal entries in Eq. 4.3 reduce to

$$\sigma_{p_i}^2 \geq \frac{1}{F_{i,i}} \qquad (4.23)$$

The resulting parameter uncertainties underestimate the true error and have to be replaced by uncertainties from Monte-Carlo simulations and a good error model if the parameter is biased or if the likelihood deviates from a Gaussian distribution. However, the Cramér-Rao inequality implies that the maximal achievable information content can be estimated. Fig. 4.11 shows the entries of the Fisher information depending on the standard lensing parameters. Blending effects are not shown, as the Fisher information of the blending flux is the same for every point, while

the information content of the baseline flux directly follows the magnification. Even if a blending ratio is introduced, the location of the maxima of each entry is not affected.

Due to the initially limited knowledge about the event parameters, it is difficult to give a general recommendation, on how to proceed before reaching the maximum of a PSPL light curve. As a first estimate, the first appearing maximum at

$$t_{\mathrm{os}} = t_0 - t_{\mathrm{E}}\sqrt{\sqrt{u_0^4 + 4u_0^2 + 1} - 1} \tag{4.24}$$

can be used as starting point. In $99.7\,\%$ of all cases t_{os} is covered when sampling is started for magnifications $\mu > 1.625$ assuming a uniform impact parameter u_0. For achieving a sampling rate offering the chance of covering local maxima, the given number of points have to be sampled in steps of

$$\Delta t_{\min} = \frac{1}{4}\sqrt{\sqrt{u_0^4 + 4u_0^2 + 1} - 1} \tag{4.25}$$

in units of the Einstein radii. Using the prior distribution of Einstein times as input, this result can be translated to a minimal sampling rate for a good characterization of events. In $99.7\,\%$ of all cases, a cadence of $\approx 20\,\mathrm{min}$ is sufficient, but in $95.4\,\%$ of all cases the requirements are achieved already for revisits every $\approx 5\,\mathrm{h}$, justifying the success of existing surveys. The given numbers are valid for PSPL models, but one can determine local maxima for every other desired quantity, such as astrometric centroid shift or planetary detection zone equally well.
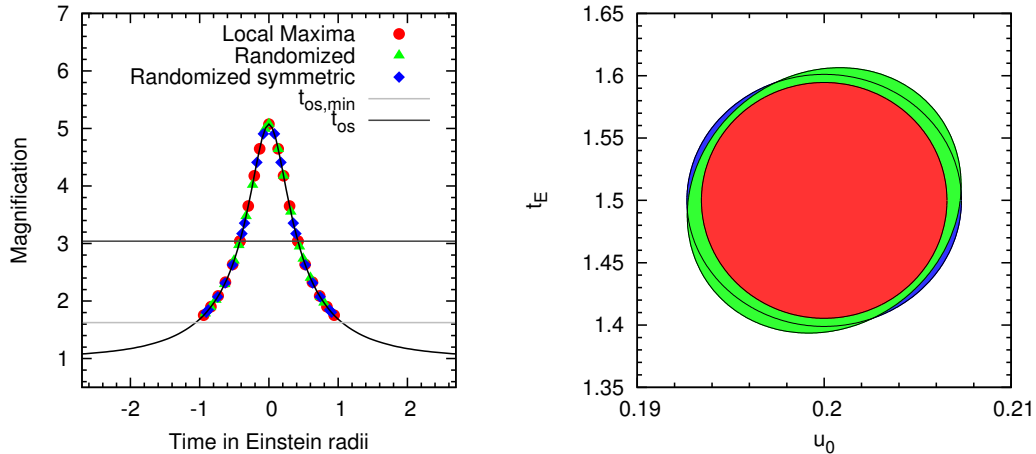
Figure 4.12: Impact of information driven sampling of a light curve (left) on the error ellipses (right) in contrast to randomized observations. The best points were chosen to be local maxima of the Fisher matrix entries with an assumed uncertainty of $5\,\%$ for each value μ.

Fig. 4.12 illustrates how the error ellipse for the parameter u_0 and t_{E} is changed for a set of randomized data points and observations taken at the local maxima of the Fisher matrix entries. If further points are supposed to be optimally placed, the pointwise description is insufficient, as all entries need to be addressed at the same time. In addition, the advantage of determining

the information locally turns into a disadvantage as the combined information content of a given set of measurements alters all entries of the Fisher-matrix and makes further proposed points dependend on the current state of observation. As there is no unique optimal sampling, different optimality criteria need to be studied.

4.3.2 Optimality criteria

Optimality per se needs a reference quantifying in what sense optimality is regarded. In many respects, gravitational microlensing studies can be subject to improvements. For instance, exposure times and technical specifications can be chosen to increase the number of detected events, the data reduction can be adapted to increase the signal-to-noise ratio of the light curve in crowded fields, and finally observing and modeling can be accelerated, if data points do not contibute to the parameter estimation. The latter approach can be applied for planning obervations and defines how frequently a given target should be observed.

A comprehensive overview of different kinds of optimality criteria are given by Atkinson (1992) and the nomenclature used here follows their definition, originally introduced and motivated by Kiefer (1959). Most of the given approaches seek to reduce the area and/or principle components of the generalized error ellipses as shown in Fig. 4.12. For practical purposes, the volume of the n_p-dimensional error ellipsoid can be minimized by maximizing the determinant of the Fisher matrix. This is a useful approach, as maximizing the determinant of $F_{i,j}$ minimizes all entries of the resulting covariance matrix due to the Cramer-Rao inequality and the relation

$$\mathbf{C} = \mathbf{F}^{-1} = \frac{1}{\det(\mathbf{F})}\mathbf{F}_{adj}^{T}, \tag{4.26}$$

which is referred to as D-optimality. A further approach relies on maximizing the sum of the eigenvalues λ of the inverse Fisher matrix

$$A_\lambda = \sum_{i=0}^{i=n_p} \lambda \tag{4.27}$$

denoted as A-optimality. Major and minor axes of the error ellipses are given as eigenvalues of the covariance matrix and consequently minimizing the eigenvalues of the covariance matrix reduces the uncertainties. According to Eq. 4.26, this can be directly expressed for the Fisher matrix itself, as the eigenvalues of $\mathbf{F}$ are proportional to the inverse eigenvalues of the covariance matrix. Maximizing the trace of those also leads to an improvement. In principle, it is desirable to improve all parameters, but if this is not achievable or does not lead to distinct recommendations, the minimal eigenvalue of $\mathbf{F}$ can be maximized acting as limiting quantity. Further enhancements can be made, which are not addressed here, but can be useful for an optimal discrimination between models. Due to the Gaussian uncertainty propagation

$$\sigma_f^2 = \mathbf{D_f}\mathbf{C}\mathbf{D_f}^{T}, \tag{4.28}$$

where the vector $\mathbf{D_f}$ carries the local derivatives of the model with respect to the parameters, the given optimality criteria also increase the predictive power of the model. Note, however, that the approach of locally optimizing a microlensing light curve cannot be achieved for D- and A-optimal designs: the determinant of the Fisher matrix vanishes and the minmal eigenvalue is zero - parameter independent. In the case of multiple measurements, this drawback can be overcome, due to the additivity of information.

Before showing examples of these criteria for μ_{PSPL}, the relevance for detecting extrasolar planets can be illustrated, en passant. Keeping in mind that the information content has been evaluated as sensitivity with respect to μ_{PSPL}, other functions characterizing the chance of detecting planets and their dependance on event parameter can be studied. Horne et al. (2009) have shown how follow-up teams can optimize their gain of detecting extrasolar planets. Their prioritization relies on the calculation of detection zones in the lens plane. In the context of the Shannon entropy, we have seen how the $\Delta\chi^2$ is distributed for a variety of light curves in the Super-Earth regime. Horne et al. (2009) have calculated $\Delta\chi^2$ maps around given image positions for an underlying PSPL model and calculate detection zone areas according to

$$D_{\mathrm{za}}(\mu) = (2\mu + 1)\left(\frac{\mu - 1}{\mu + 1}\right)^{1/2} \tag{4.29}$$

defining a weighting factor g characterizing the value of each observation depending on the mass ratio q, the exposure time for a signal-to-noise of one t_{snr1} and a $\Delta\chi^2$ threshold

$$g = \frac{q D_{\mathrm{za}}(\mu)}{\sqrt{t_{\mathrm{snr1}\Delta\chi^2}}}. \tag{4.30}$$

Under the assumption of a given distribution of planets these expressions may change (e.g. Dominik et al. 2010). Whatever exact shape is chosen, they all rely on a good characterization of the event parameters on the basis of an initial PSPL fit. Due to the uncertainties of the PSPL fit, the resulting Fisher matrix can be used for estimating the impact on the detection zones.

Fig. 4.13 shows how the information content changes for given light curves, assuming the underlying PSPL model is known. As expected, different methods indicate where the highest change is expected to be. Notably A- and E-optimality contribute contradicting recommendations, i.e. the minimal eigenvalue does not change at points, where the total standard deviation can be reduced. The recommendations and maxima for different types of optimality overlap with existing data points. D-optimality provides the most useful description of how to continue for limited initial information, as it is minimizes all entries of the covariance matrix. As expected, it recommends a good coverage at t_0 with a smooth increase for given initial observations. For each light curve different ending points were chosen implying that the position of maxima depends on the given observations. This behaviour suggests a constructive technique for determining an optimal sampling:

supposing the correctness of the given model and using an initial set of measurements, it can be assumed that an observation will be taken at a proposed point and first of all at $t = t_0$; In a second step the position of new local maxima can be determined and assumed to be observed; this process can be iterated until a certain number of points, $\Delta\chi^2$ or thresholds for functions of $\mathbf{F}$ are achieved. In Fig. 4.14 the first three steps of such an iteration are shown, indicating how quickly the remaining points converge towards an uninformative evolution. Our first approach of optimizing each matrix entry separately is not unreasonable. Precautions have to be made exclusively for light curves where data has been taken exactly at $t = t_0$, because the information content for the parameters t_0 and u_0 becomes infinity. Finite exposure times and imperfect timing, prevent this in practical cases.

4.3.3   Optimal sampling for unknown event parameters

In the rising part of the light curve, little is known about the event parameters. Nevertheless, estimates for possible future tracks can be given based on the distribution of u_0 and the current
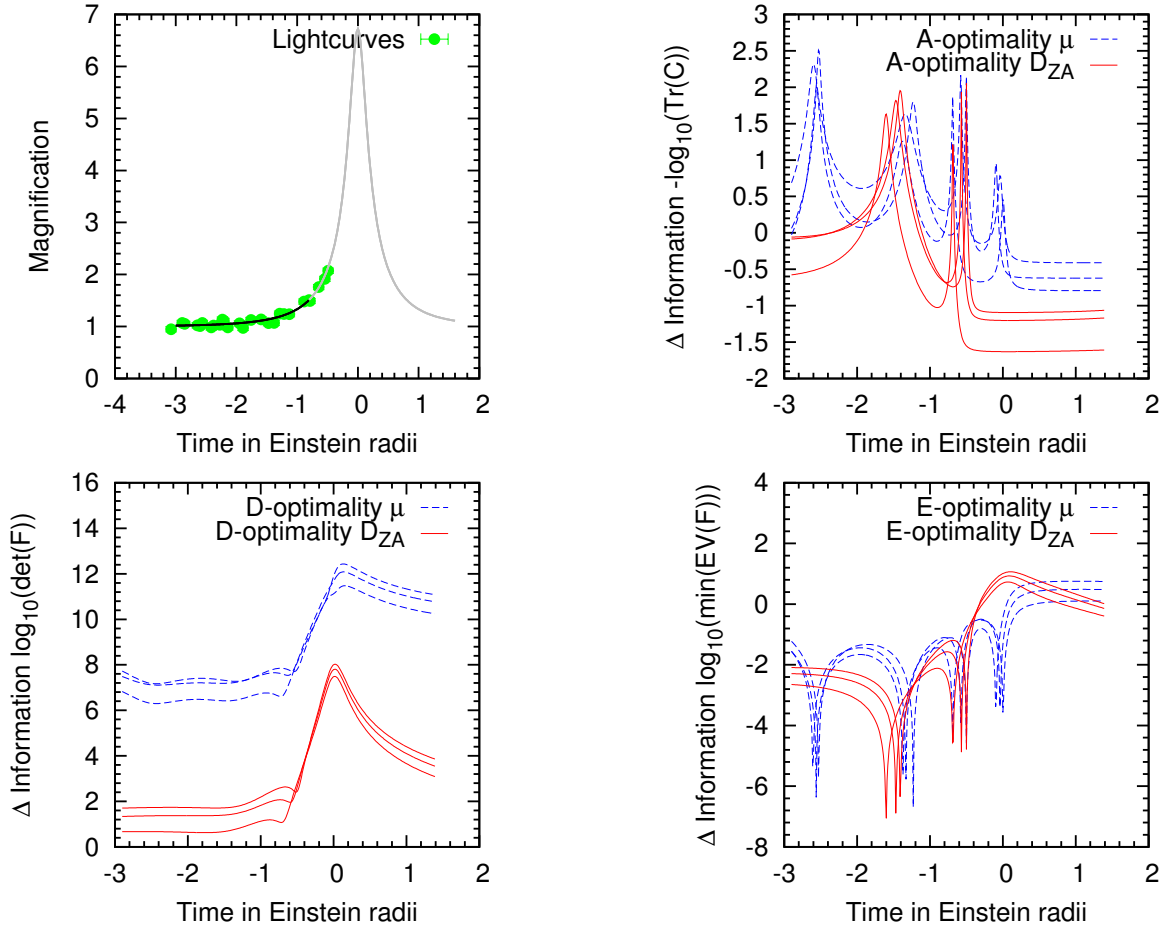
Figure 4.13: The expected change in information with respect to magnification and planetary detection zone after detecting an increase of brightness is shown for three different optimality criteria. The light curves were simulated with a noise of 5 % and three different ending points were chosen.

magnification $\mu \Rightarrow u_{0,\mathrm{max}}$. As before, an uninformative uniform distribution is assumed neglecting the limiting magnitude for a given field. The additional contribution of each data point cannot be determined independent of the complete light curve obtained for a given sampling rate. Obviously, it is better to spend more time on regions with a higher information content: if the sampling intervals Δt between two observations are chosen to be

$$\Delta t \propto \frac{1}{I}. \tag{4.31}$$

a better sampling of information carrying points is achieved. Here I represents the information content for a chosen optimality criterion and underlying sampling rate, yielding a total observing time of

$$t_{\mathrm{obs}} = \sum_{i=1}^{N_{\mathrm{obs}}} \Delta t, \tag{4.32}$$
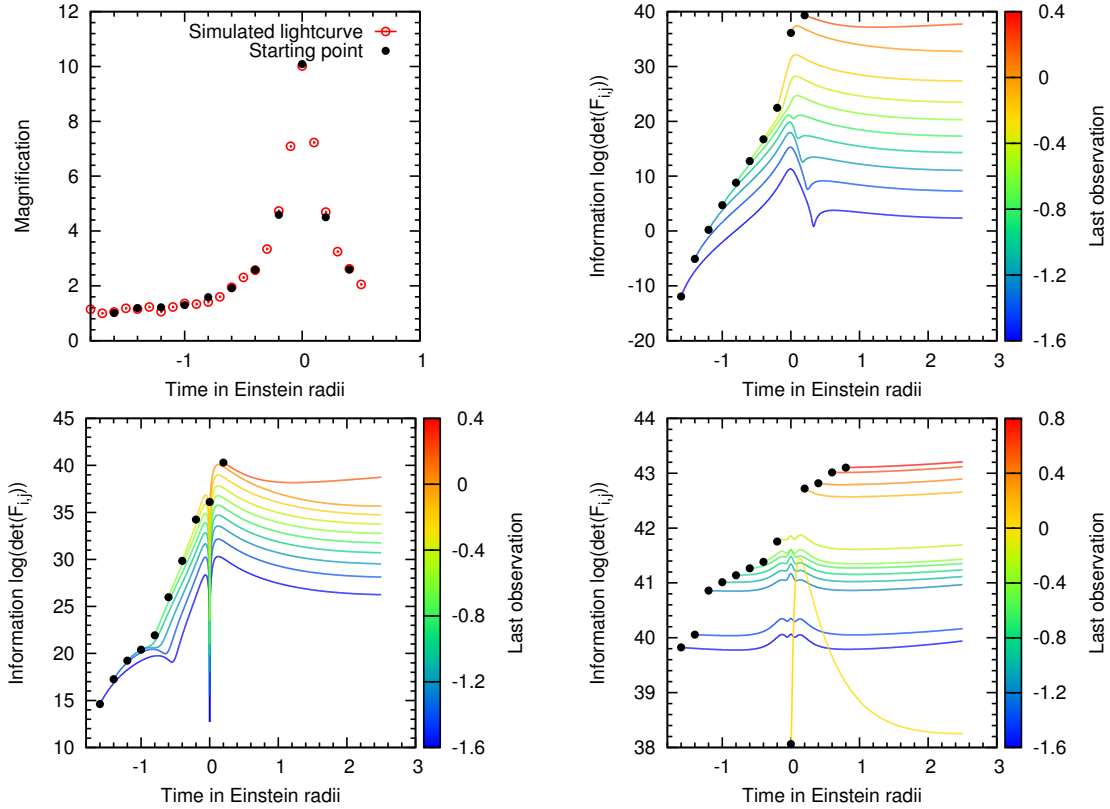
Figure 4.14: Iteratively improved sampling for D-optimal designs are shown for different starting points (top left). The initial recommendation is to observe at $t = t_0$ (top right). Assuming this point has be observed leads to local maxima (lower left) and assuming the two remaining maxima leads to the nearly flat change in information (lower right).

where the average sampling rate is supposed to be given as

$$\frac{t_{\text{obs}}}{N_{\text{obs}}} \tag{4.33}$$

D-optimal designs are more appropriate for such an approach, as different units of the eigenvalues contribute to the global scaling. In an A-optimal scenario, this would lead to differing weights for each parameter. The example in Fig 4.15 shows how such a directly modulated sampling can be realized for a given PSPL-light curve (the marginalized error ellipses are shown for the standard parameters). The number of regions with higher sampling rate depends on the impact parameter u_0 and exhibits a more complicated sampling structure for low magnification events. Here one sees now an information modulated approach has several advantages over the other approaches: in contrast to a sampling which is exclusively driven by local maxima, potential gaps are filled ensuring that possible anomalous deviations can be detected. An iterative scheme as shown before, where the local information content is used, is numerically costly as all possible future times have to be simulated and only when the maximal magnification is reached all the event parameters provide better predictions for the remaining light curve. As long as the light curve is incomplete, only typical values or constraints from expected values such as $\langle u_0 \rangle_{\text{uniform}} = u_{0,\text{max}}/2$

and the mode for a given line of sight $t_{\mathrm{E,mode}}(l, b)$ can be provided, which is $8.04\,\mathrm{d}$ for the earlier mentioned OGLE field at $l = 1.2°$, $b = -2.7°$.
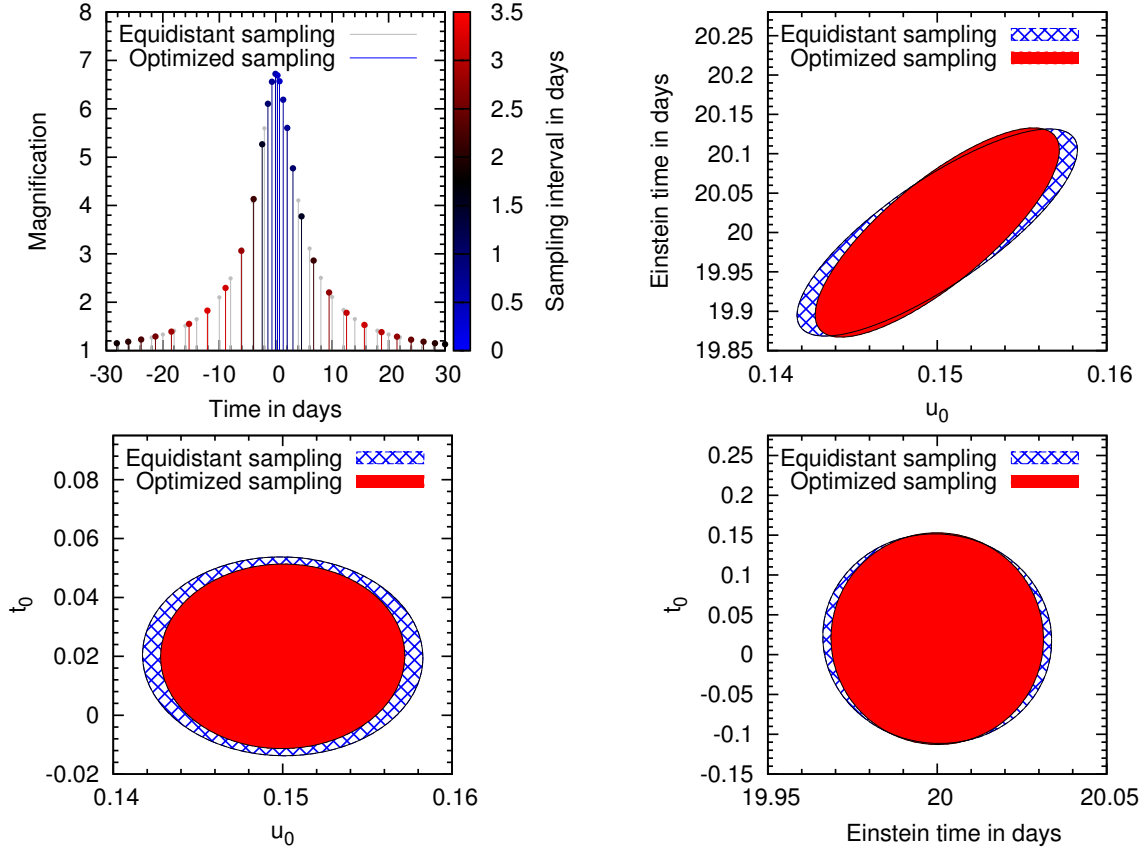
Figure 4.15: Information modulated sampling for a PSPL-light curve with $u_0 = 0.15$, $t_{\mathrm{E}} = 20.0\,\mathrm{d}$, $t_0 = 0.02$ in contrast to an equidistant sampling (top left) and the resulting marginalized error ellipses.

All these assumptions and issues lead to an interpolation formula for an optimized sampling. The proposed sampling depends not only on the knowledge of the parameters, but also on the definition when a microlensing event is supposed to be discovered. Given that the photometric accuracy is at least around $3\,\%$, required by the planet search, a detection of a microlensing event is by defintion beyond $3\,\sigma$ and thus the magnification needs to exceed 1.1 which implies $t \in [t_0 - 1.67\,t_{\mathrm{E}}, t_0 + 1.67\,t_{\mathrm{E}}]$ or for a $1\,\%$ limit $t \in [t_0 - 2.5\,t_{\mathrm{E}}, t_0 + 2.5\,t_{\mathrm{E}}]$. For both scenarios, different sampling strategies are required. Fig 4.16 illustrates how the proposed sampling recommendations are affected by u_0, photometric accuracy and the respective sampling interval. For low magnification events, multiple local minima contribute equally to the total information content, but for high magnification events a sampling rate increasing with magnification is dominant. The symmetry of the light curve and the initial sampling modify the requested intensified sampling. For practical purposes this function can be interpolated by using a bivariate polynomial

$$\Delta t = \frac{t_{\mathrm{E}}}{\langle \Delta t \rangle} \sum_{i=0}^{N} \sum_{j=0}^{N-i} \alpha_{i,j} u_0^i \left| \frac{t - t_0}{t_{\mathrm{E}}} \right|^j, \tag{4.34}$$

where 44 parameters $\alpha_{i,j}$ for $N = 8$ can be determined from a linear fit. Evidently the sampling intervals have to be adapted in the course of a microlensing event due to the aforementioned difficulties. The accuracy of the 8th order interpolant can be determined according to the scatter of the residuals and in both cases it stays around $10^{-3}t_{\rm E}$, implying for typical events an optimal placement up to 15 min. Comparing this with a 6th order approach with 27 parameters would have degraded the accuracy by a factor of two.
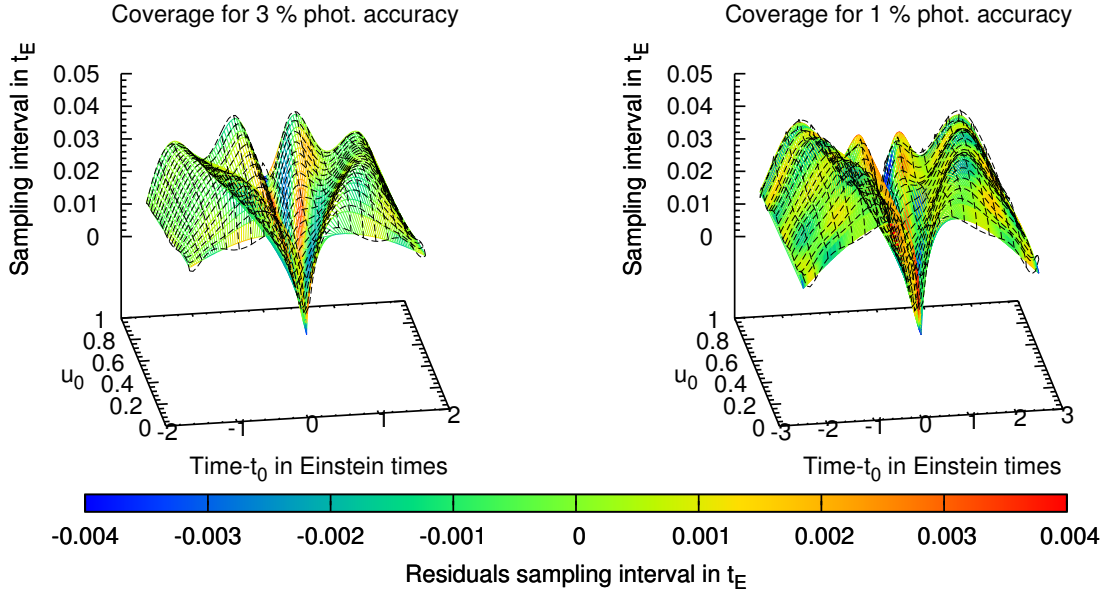
Figure 4.16: Modulated sampling recommendations depending on u_0 and photometric accuracy are shown along with a interpolant based on an 8th order bivariate polynomial. The sampling intervals are calculated for an underlying equidistant sampling intervals of $\langle \Delta t \rangle = 0.01$.

i,j	0	1	2	3	4	5	6	7	8
0	0.023	-0.135	-1.983	32.279	-139.874	301.582	-361.651	230.726	-60.752
1	1.930	-3.936	-19.410	77.872	-67.571	-6.974	18.019	-2.515	-
2	-6.400	31.357	-40.504	-71.108	150.293	-39.611	-3.755	-	-
3	11.163	-49.238	102.163	-74.864	-65.369	19.693	-	-	-
4	-11.210	37.923	-36.865	85.079	1.726	-	-	-	-
5	5.308	-25.936	-17.237	-18.593	-	-	-	-	-
6	0.448	14.355	8.097	-	-	-	-	-	-
7	-1.440	-3.323	-	-	-	-	-	-	-
8	0.388	-	-	-	-	-	-	-	-

Table 4.2: Interpolation parameters for 1 % photometric accuracy

i,j	0	1	2	3	4	5	6	7	8
0	0.028	-0.151	-3.014	38.339	-157.047	323.976	-364.446	212.509	-50.107
1	1.450	-2.932	-8.552	30.363	-20.458	-4.321	8.734	-3.573	-
2	-3.139	14.863	-14.880	-24.439	31.464	-5.598	1.191	-	-
3	3.366	-17.083	29.738	-5.931	-12.949	0.531	-	-	-
4	-1.819	8.849	-13.522	9.429	2.021	-	-	-	-
5	0.184	-3.071	0.973	-2.082	-	-	-	-	-
6	0.328	0.932	0.337	-	-	-	-	-	-
7	-0.166	-0.151	-	-	-	-	-	-	-
8	0.025	-	-	-	-	-	-	-	-

Table 4.3: Interpolation parameters for $3\,\%$ photometric accuracy

4.4 Conclusions

Information theory can provide interesting new results and different interpretations of existing studies. The considerations being illustrated here, rely on a frequentist statistical approach and prior information was exlusively introduced for creating possible parameter configurations. Nevertheless, most methods can be extended for Bayesian statistics; local and global information content can be modified by replacing frequency probabilities by posterior probabilities.

Both information theoretical approaches provide recommendations on how experiments should be designed for studying planetary populations or any other scientific objective. Shannon's information definition quantifies the worth of rarely occurring observables and motivates merging all microlensing observations to a single experiment. Instead of using the results of single light curve fits, ensembles of observable values based on the observed brightness change can be used. This approach requires the construction and understanding of background distributions. Such a description needs to include all potential explanations of anomalous events, imposing a major challenge for the future. Furthermore, the underlying planetary population may lead to ambigous results of uninformative results, as shown for the source star distribution. As PSPL model fits can be achieved rather easily, the distribution of χ^2 values seems to be a good criterion as deviations may provide the desired insights. Such an approach has one major drawback: the errorbars provided by using existing data reduction pipelines are not always convincing. Consequently, a good simulation starts with simulating frames according to chosen Galactic models and for each observing site. For different realizations of the observation, the total uncertainties of the observation and reduction could be deduced and applied to all preceding reductions. This would increase the confidence in all results so far achieved and it is inevitable if microlensing is used as a unified, global experiment.

The local information content, encoded as a Fisher matrix, leads to completely different results. It helps observers to improve planning their experiments in order to achieve more accurate parameter estimates spending the same observing time. Nevertheless the effect is limited if the true shape of the light curve is unknown. For PSPL events, a general strategy can be defined. After sampling the maximal magnification the knowledge of parameters is dramatically increased, as u_0 and t_0 are well known afterwards. The resulting recommendations require next points to be taken based on the current state of the light curve. This is especially helpful as symmetries can be exploited. An information driven sampling approach is a delicate system in the rising wing of the light curve, as the initial parameter estimates are not known, smoothing out the distinct

maxima of information content for different model parameters. By assuming typical event parameters an information modulated strategy was introduced, following a D-optimal design. This approach is not only useful for predict the next light curve points but also for optimizing the information about the planetary detection zone. In the theory of optimal experimental design, more optimality criteria have been proposed (cf. Atkinson 1992) and wait for being applied to gravitational microlensing.

Finally, the prospects for an optimal experimental design are lower as soon as survey telescopes are able to provide a high number of event detections and a high observational cadence simultaneously. The future application of optimally sampled light curve needs to be customized for characterizing anomalous events, such as the prospects of studying extrasolar planets and extrasolar moons Liebig & Wambsganss (2010).

A more relevant application for the Fisher-matrix can be achieved if it is used in the course of fitting a binary microlensing event, as it provides a local estimate of the sensitivity of a model with respect to model parameters. Consequently, it tells us which data points could temporarily detected and thus compress the amount of data.

Chapter 5

Data analysis techniques

In the previous chapter, I discussed how light curves should be sampled in order to increase the accuracy of parameter estimates. In the following chapter, the practical implications of information theory are used to efficiently fit models to real observations. One of the standard global optimizing techniques, namely Metropolis Hastings Markov Chain Monte Carlo (MHMCMC) method, is adapted for use with the GPU-assisted contouring model. The information content of a given light curve can help to reduce the amount of data being processed. This reduction can be achieved by processing a subset of the observed light curve or by using other information-driven compression techniques.

5.1   Massive compression

As indicated by Shannon (1948), the consideration of information content can lead to a more efficient communication process and thus to data compression. One historic example of such a lossless compression is the Morse code, where frequently used characters consist of less bits. For microlensing measurements, Fisher's information approach can also be relevant. Heavens et al. (2000) have shown how a massive lossless data compression of galaxy spectra can be achieved, where lossless by definition means without loss of information and here without loss of Fisher information. In principle this technique can be applied to gravitational microlensing for speeding up the fitting process. The following adaptation follows their work.

Starting with a vector of flux measurements $\boldsymbol{F}_{\mathrm{obs}}$, weighting vectors $\boldsymbol{w}_n$ are constructed for each parameter p compressing the measurement vector to a single value. If a model with n_p parameters describes the data well, the resulting compressed measurement consists of n_p data points. Demanding that the resulting data points are uncorrelated leads to an orthogonalization scheme for determining the compressed parameters:

$$\boldsymbol{w}_n = \frac{\boldsymbol{d}_n - \sum_i^{n_p-1} \left(\boldsymbol{c}_n \cdot \boldsymbol{w}_i\right) \boldsymbol{w}_i}{\sqrt{F_{n,n} - \sum_i^{n_p-1} \left(\boldsymbol{c}_n \cdot \boldsymbol{w}_i\right)^2}} \tag{5.1}$$

where the vectors $\boldsymbol{c}_n$ and $\boldsymbol{d}_n$ are defined as

$$\boldsymbol{c}_n = \frac{\partial F_i}{\partial p_n} \tag{5.2}$$

$$\boldsymbol{d}_n = \frac{1}{\sigma_i^2} \frac{\partial F_i}{\partial p_n}, \tag{5.3}$$

and the Fisher matrix entry $F_{n,n}$ is computed as defined in Eq. 4.1. Eq. 5.1 can be qualitatively understood as a weighting scheme where every observed data point is weighted according to its sensitivity and accuracy. By consecutively constructing compression vectors for all parameters, the contribution of already determined weights is adapted in order to conserve the Fisher matrix.

Eq. 5.1 is given in a less general form under the idealized assumption that consecutive observations are uncorrelated. For observations carried out in the same night, and thus with similar weather and seeing conditions, this assumption will inevitably be violated. For fitting purposes, the maximum likelihood fit can be determined as usual except for exchanging data set and model by its compressed constituents

$$F_{\mathrm{cmprs},p} = \boldsymbol{w}_n \cdot \boldsymbol{F}_{\mathrm{obs}}. \tag{5.4}$$

The compression requires prior knowledge of the event parameters and Heavens et al. (2000) have indicated that a fiducial model can be used for an initial estimate providing parameter estimates which can be iteratively improved, that means by updating the compression and reanalyzing the event.
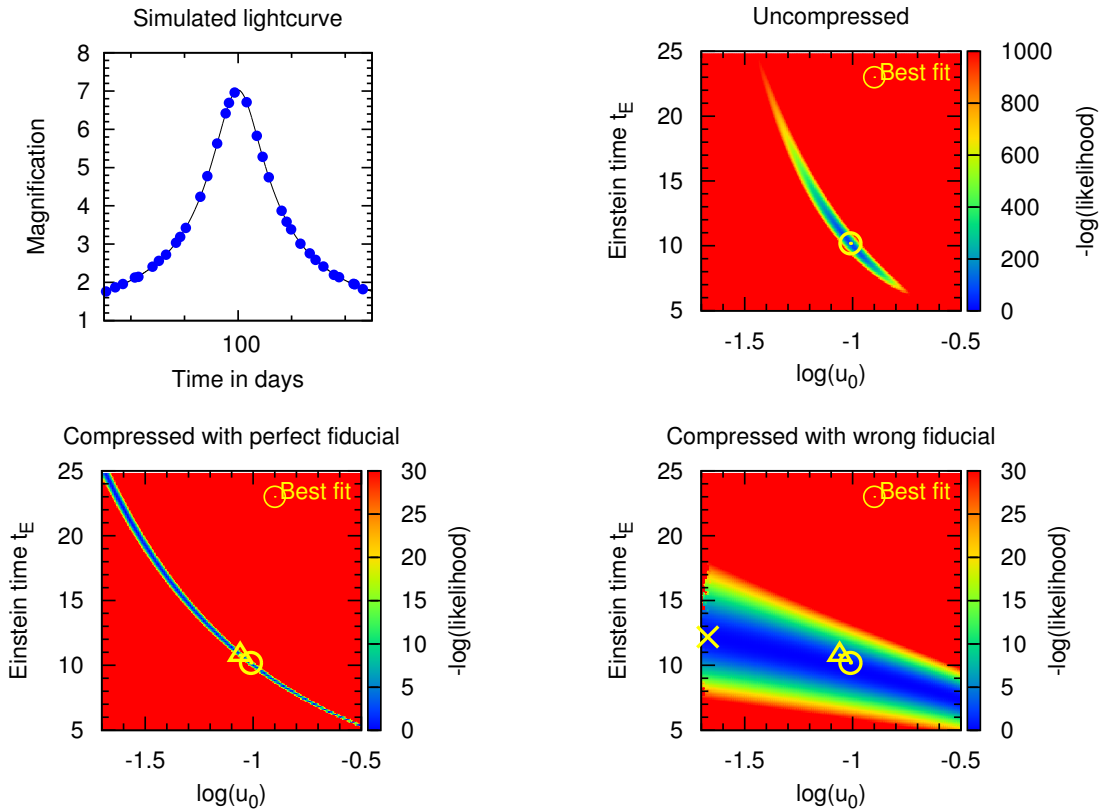
Figure 5.1: The potential of massive data compression is shown for a light curve simulated on a randomized grid with a photometric accuracy of 2% (upper left) as marginalized $-\log(\mathscr{L})$ surfaces for $\log_{10}(u_0)$ and t_{E} (lower row). The best fit is indicated for different modifications starting with the uncompressed χ^2 surface.

The feasibility of implementing this approach for gravitational microlensing can be assessed theoretically in the context of the outlined scheme. Originally, the compression technique was used for analyzing galaxy spectra (Reichardt et al., 2001), which are usually well sampled and feature

rich. As shown in the preceding chapter, the sampling clearly affects the Fisher-matrix which is used for compressing the data in Eq. 5.1. Consequently, the compressed data points $F_{\mathrm{cmprs},p}$ will not remain uncorrelated. However, use of Fisher-matrices for estimating uncertainties requires unbiased estimators. For a fixed fiducial model, used for compression this inevitably leads to biased parameter estimates. Nevertheless, just the possibility of speeding up the initial part of data-analysis justifies an implementation for a PSPL-model.

Fig. 5.1 illustrates a PSPL-model simulated for the parameters $t_{\mathrm{E}} = 1\,\mathrm{d}, t_0 = 100.0\,\mathrm{d}, u_0 = 0.1$ and a blend ratio $g = F_{\mathrm{B}}/F_{\mathrm{S}} = 0.5$. The χ^2 surface is plotted for the parameters t_{E} and u_0 for a fit using the whole data set and for a massively compressed version. The compression was tested with a perfect fiducial model, but also with a fiducial model increasing all parameters by $20\,\%$. The resulting best fits are indicated: for a perfect model the 4 parameter fit provides reasonable estimates but for a wrong compression iterations are required.

In principle, this technique can also be applied to binary microlensing events. The advantage of using more parameters is in general a higher information content. Nevertheless, the computation of the compression requires a complete model for a given light curve, the first derivatives with respect to all parameters and, last but not least, a good fiducial model. The latter issue is the most delicate one, as the parameter space is too large to be treated with a single fiducial model. Fig. 5.1 also gives an idea what happens if different weighting vectors are used. Preparing subspaces in parameter space with each compressed in a different way, is one option. Another difficulty for such an endeavor arises, as the compression depends on the sampling of a light curve. A pre-calculated grid of compressed models requires that unevenly sampled observations are rebinned and interpolated to match the reference model. Such an approach increases the uncertainty per data point and an overall lossless accuracy is not longer achievable. Finally, a standard technique is needed for determining the accurate solution and the corresponding uncertainties. The given requirements motivate following different ways of directly integrating an information driven compression in existing standard approaches.

5.2 Metropolis Hastings Markov Chain Monte Carlo

Different optimization algorithms are used for fitting single and binary lens microlensing events, one example is the simplex technique (Nelder & Mead, 1965): a simplex in n dimensions is a generalized triangle probing the parameter by modifying the vertices of the simplex, i.e. substituting the worst vertices. Alternatively, one can use the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) which is a modified version of Newton's method with additional step size control. Both techniques can be applied to the analysis of microlensing events, but require additional modifications to avoid getting stuck in local minima. The Levenberg-Marquardt algorithm demands accurate first derivatives, which is hard to guarantee for the GPU-contouring model as long as single precision GPUs are used.

I have chosen to use another common technique, namely the Metropolis Hastings Markov Chain Monte Carlo (MHMCMC) algorithm. In the seminal paper by Metropolis et al. (1953), it was shown how the equation of state can be efficiently calculated by carrying out a random walk and accepting new steps with a defined transitional probability, i.e. better steps are always accepted and worse steps are acceepteed according to the ratio of the system probability, in the original case the Boltzmann probability. The next proposed step in the random walk exclusively relies on the current state of the system and thus satisfies the Markov property, that is, a series of steps is a Markov chain. Hastings (1970) has generalized the original sampling approach for application to a wider range of numerical problems. The transitional probability can be calculated

for other processes such as the likelihood ratio in a fitting scheme. For this purpose, random steps are taken in the parameter space and accepted if the likelihood is increased. The likelihood $\mathscr{L}(y|p)$ (Fisher, 1922), for a given model is the probability of the parameter configuration p for any given observation or for a larger data set

$$\mathscr{L}(y|p) = \prod_{i=0}^{N} \mathscr{P}(y|p), \tag{5.5}$$

and depending on the probability density $\mathscr{P}(y|p)$ of each observation y. This technique is frequently applied for simulating the posterior probability function according to Bayes theorem

$$P(p|y) = \frac{\mathscr{L}(y|p)\mathscr{P}(p)}{\mathscr{P}(y)}, \tag{5.6}$$

which is relevant for inferring estimates including prior knowledge in the form of a distribution $\mathscr{P}(p)$. In data analysis, the probability $\mathscr{L}(y|p)$ represents the likelihood of an observation with model parameters p. The normalization constant $\mathscr{P}(y)$ needs to be determined properly for estimating probabilities, but can be neglected if exclusively the distribution of parameters is of interest.

In many practical cases, the problem reduces to the special case of a Gaussian likelihood function. The transitional probability in the MHMCMC scheme then reads

$$\mathscr{P}_{\text{transition}} = e^{-\Delta\chi^2/2}. \tag{5.7}$$

Compared with other fitting techniques, the MHMCMC approach can be effective at avoiding local minima, though at the price of computationally intensive sampling. In practice, the basic MHMCMC approach needs further modification for being used as fitting algorithm. Depending on the initial parameter estimates, the proposed steps may not match the scale of the current part of the parameter space, and as a consequence, inefficiently sample the parameter space. In order to compensate for such an behavior, the scale of the proposed steps can be increased to reach an empirical efficiency of around 20 %, i.e. every fifth step is accepted. Depending on the implementation and the given problem, a different range of values can be useful. Caldwell et al. (2009), for instance, have tried to keep the efficiency between 10 and 50 %. For the GPU-computing part, too small proposed steps have to be avoided, leading to pseudo-efficient sampling: Markov chains can reach formal convergence not only because of reaching the desired distribution but also because of parameter ranges where the walk only probes the numerical inaccuracies. For this purpose the sampling can be increased by a factor of 10 whenever the relative step size σ/p drops below a defined threshold, such as $10^{-4} \cdot p_i$, ensuring that 2-3 digits of the step size remain significant. This correction is especially relevant when a small number of data points is considered.

Whenever an efficient sampling has been reached, the proposal distribution needs to be adapted to match the structure of the distribution being analyzed. For Gaussian proposal functions, the standard deviation can be used for updating the proposal function in each dimension. Such an update can be done after a defined number of steps or with running updates the standard deviation and efficiency control. I implemented the latter approach in order to ensure that local minima can be left as soon as a better solution is found. The efficiency control supports this effort, as inadequate sampling is assessed after each step. One disadvantage of running updates can be ameliorated by the choice of proposal function shape: instead of using marginalized Gaussians for each dimension, a multivariate Gaussian proposal function helps to avoid wasting steps,

as correlations are included and better follow non-symmetric structures in the likelihood (e.g. Doran & Müller 2004). This implies that a principal component analysis needs to be done for every step, counteracting the GPU-acceleration effect, as the Graphics processor then has to wait for the CPU to propose the next MCMC step.

Another important aspect for configuring the MHMCMC is the question when the chain is sampling from a stationary distribution and thus when the proposal function can be fixed. A technique for evaluating the convergence of a chain was developed by Gelman & Rubin (1992): instead of considering single chains, the behavior of several chains in parallel is evaluated. The variance, of each parameter p within one of the chains W and the variance between all chains B can be compared using the statistic

$$R_p = \frac{n-1}{n} W_p + \frac{1}{n} B_p, \tag{5.8}$$

where $R \approx 1$ in case of convergence. The complete run is considered to be converged if all parameters provide R-values below a given threshold, usually $R < 1.2$ Doran & Müller (2004). In addition, it is important to ensure that the sampling efficiency stays around the aforementioned empirical limits, where the sampling efficiency is defined as the relative number of accepted steps. In the implementation for the GPU-model, the efficiency was kept between 15 and 35 % and running updates were based on the last 50 steps. The scale factor, rescaling all step sizes if the sampling is out of the efficiency limits, was chosen between $2^{1/50}$ and $3^{1/50}$.

The effect of the single precision limitation is apparent in the χ^2-value

$$\chi^2 = \sum_k^N \frac{(\Delta f + \Delta f_{\mathrm{num}})^2}{\sigma_{\mathrm{obs}}^2}, \tag{5.9}$$

which is used in the fitting process. In addition to the random residual Δf a random fluctuation Δf_{num} leads to a systematic broadening which is for two Gaussian distributed random numbers

$$\sigma_{\mathrm{eff}} = \sqrt{\sigma_{\mathrm{num}}^2 + \sigma_{\mathrm{obs}}^2}, \tag{5.10}$$

reducing the effect of the accuracy limit. For a given constant uncertainty σ_{obs} the effect on the $\Delta \chi^2$ can be assessed. Assuming the number of data points n is much greater than the number of parameters n_p, the change of χ^2 per degree of freedom *dof* yields

$$\langle \Delta \chi^2 \rangle / dof = \frac{1}{\sigma_{\mathrm{obs}}^2} \left(\langle 2 \Delta f \sigma_{\mathrm{obs}} \Delta f_{\mathrm{num}} \rangle + \langle \Delta f_{\mathrm{num}}^2 \rangle \right) = \frac{\sigma_{\mathrm{num}}^2}{\sigma_{\mathrm{obs}}^2}. \tag{5.11}$$

This equation summarizes why it is usually demanded that a numerical model needs to be at least one order of magnitude more accurate than the measurement itself, implying the resulting χ^2_{red} is accurate up to 1 %. The first term in Eq. 5.11 does not seem to contribute to the result but its fluctuation needs to be determined to define a safe limit for rejecting the effect of an inaccurate model. By calculating the variance, Eq. 5.11 reads

$$\langle \Delta \chi^2 \rangle / dof = \frac{\sigma_{\mathrm{num}}^2}{\sigma_{\mathrm{obs}}^2} \pm 2 \sigma_{\mathrm{obs}} \sigma_{\mathrm{num}} \tag{5.12}$$

The worst case for a 3σ-level is a change in χ^2 by $\sigma_{\mathrm{num}}^2 / \sigma_{\mathrm{obs}}^2 + 6 \sigma_{\mathrm{obs}} \sigma_{\mathrm{num}}$. Considering again a numerical accuracy one magnitude below the measurement implies a relative contribution of

less than $1\,\%$ to the $\Delta\chi^2$, which is negligible in most practical cases. A typical example with $\sigma_{\mathrm{obs}} = 0.01$ and $\sigma_{\mathrm{num}} = 0.001$ leads to an additional $\Delta\chi^2 \approx 15 \pm 0.03$ for a light curve consisting of ≈ 1500 points, which is relevant for the case of confirming the presence of a planetary perturbation. Due to the model inaccuracy, of the order of one data point is effectively discarded and the best fit is still achieved, but with a shifted χ^2.

5.3   Initial fit

Before discussing other ways of exploiting the Fisher information in fitting techniques, a simulated light curve is considered. The parameters were chosen in a way that difficulties will be encountered; more precisely, these difficulties are caused by symmetries, parameter degeneracies and imperfect coverage. In principle, sufficiently long MHMCMC chains should be able to reach the best fit, as the random walk can leave local maxima. However, depending on the structure of the likelihood function, many local distributions need be probed before the global minimum is reached.

An initial grid with multiple chains can be helpful for getting an understanding where more chains are needed, due to the fluctuations of the likelihood. The grid search comprises three parameters which are especially affected by fluctuations, namely q, d and φ (e.g. Dong et al. 2009). For each grid point the remaining PSPL parameters are fitted, initially with a higher source star radius and thus with reduced bookkeeping for GPUs. The resulting marginalized χ^2 surface for q,d is shown in Fig. 5.2 along with the initial configuration. The final fit, resulting from MHMCMC chains that were started at the best grid points, and the best fit in a least square sense are also shown. For this exemplary fit, no convergence was achieved even for the best fit result and thus no credible uncertainties can be deduced. The reason why the correct answer was not found is obvious when one considers the sampled light curve and the best fit as shown in the right panel of Fig. 5.2: a degenerate solution was found due to the poor coverage with respect to information carrying parts of the light curve leading to the parameters shown in Tab. 5.1. It is still remarkable that some parameters, particularly q and t_{E} serving as indication of the masses M_1, M_2, do provide a reasonable answer. In total 200 data points have been simulated.

The situation can be improved by using the aforementioned lower limits for the variances of the proposal function, but at the end, a good initial guess of parameters is needed. If a PSPL model can be fit to the light curve, parameters like t_0 can be accurately determined in advance. Nevertheless, a limited coverage can lead to wrong results, even with a perfectly sampled light curve and converged chains. Binning or rejecting data points in a microlensing light curve is a risky endeavor as long as the parts of the light curve carrying information are affected, leading to an approach for selecting the right points at runtime.

Model	u_0	t_{E}	t_0	φ	q	d	$r_\star$
Underlying model	0.1	10.0	100.0	135.0	0.5	0.5	0.0018
Best fit	0.07713	11.2680	100.97322	253.57241	0.50180	0.51468	0.00102

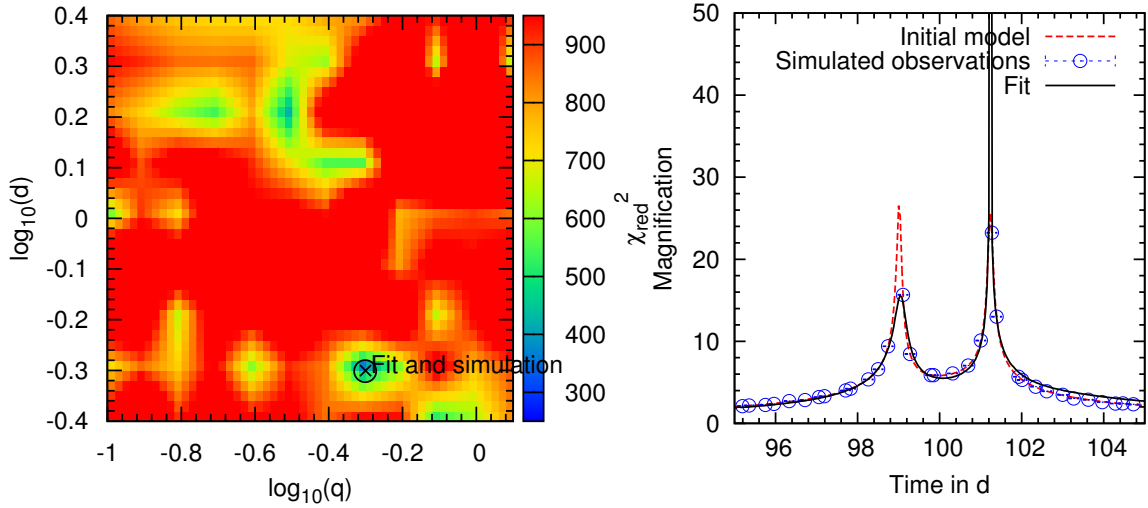**Table 5.1:** Example fit for a simulated binary light curve.

Figure 5.2: χ^2 surface for a simulated binary light curve (left) and the corresponding best fit along with the initial model.

5.4 Information-accelerated MHMCMC

In most practical cases, featureless parts of the light curve leave the likelihood unchanged when parameters are changed. As the simulation of each point of the light curve increases computation time, the best fit can be achieved more quickly if parts of the light curve can be neglected. The initial selection of relevant data points seems to be the easiest way for reducing the computational burden, but this may exclude important data points depending on the given state of a light curve. Before introducing s more flexible strategy, the calculation of the information content in a binary model needs to be discussed.

5.4.1 Information content in binary models

Recalling the findings of chapter 4, a considerable number of first derivatives have to be determined in order to compute the information content. Seven parameters particularly need to be addressed for fitting purposes. The parameters for such a standard binary event were introduced in chapter 2: their meaning is summarized in Fig. 5.3 for the simple case of a linear source star track. The source track itself requires three parameters: orientation φ; impact parameter u_0; time of closest approach of the center of mass t_0; and the source star radius in units of the Einstein radius $r_\star$. The minimal separation u_0 is given relative to the center of mass as well, which is defined to be the origin of the source plane. This definition is kept throughout this work, providing a good agreement with a PSPL model as long as the track is sufficiently far away from a caustic. A plethora of other choices can be found in the literature. All parameters are necessary for inferring the physical properties of the system, because parameter degeneracies arise due to an insufficient sampling rate, symmetries in a given light curve or symmetries in the lens equation itself. The masses M_1 and M_2 are not directly accessible, but their mass ratio q and the angular separation d are, because they correspond to a distinct caustic curve for a perfectly sampled light curve.

Analytically, the maximal information content for a binary model with a given source track
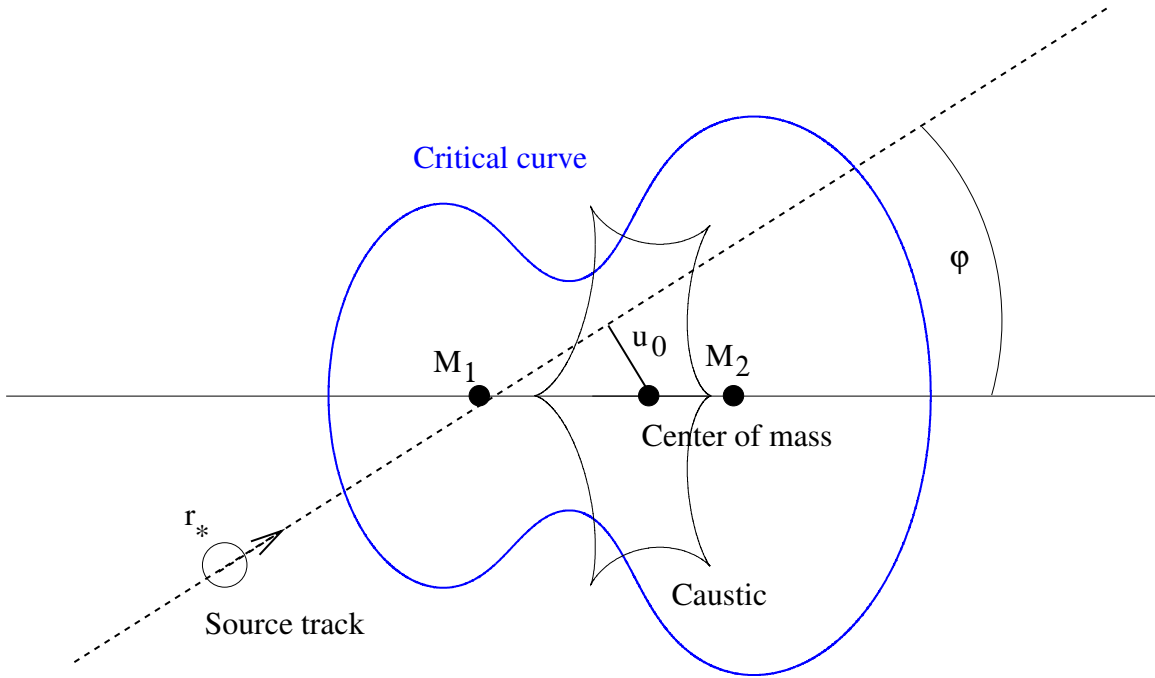
Figure 5.3: Parameter definition for a standard binary microlensing event.

parameterization $\mu\left(\boldsymbol{\beta}\right)$ can be obtained from

$$F_{i,j} = \left(\left(\nabla_{\boldsymbol{\beta}}\mu\right)\cdot\left(\partial_{p_i}\boldsymbol{\beta}\right)\right)\cdot\left(\left(\nabla_{\boldsymbol{\beta}}\mu\right)\cdot\left(\partial_{p_j}\boldsymbol{\beta}\right)\right) = \max. \tag{5.13}$$

Neglecting the effect of mixed terms, the information content follows the gradient of the magnification pattern. In general, mixed terms with potentially negative contributions cannot be neglected and thus numerical derivatives are inevitably neede for calculating the total information content.

Given that the accuracy of the GPU-model is limited by a numerical uncertainty of $\approx 0.1\,\%$, the step size for each numerical evaluation must be chosen to balance numerical inaccuracy and the discretization error. The optimal step size for each parameter p is (Press, 1994, p. 187.)

$$\Delta p \approx \left(10^{-3}\right)^{1/3} p. \tag{5.14}$$

Based on this estimate, numerical derivatives for determining the information content are calculated for each parameter. The total information content is calculated as volume of the error ellipsoid, based on the product of eigenvalues which is determined according to Galassi et al. (2009). The resulting vector of information serves as a sorting key where a predefined fraction of data points will be used for the fit. The total information content of a binary light curve is shown in Fig. 5.4 and, as expected, there is an increase in information close to the magnification maxima, but there are also points in the wings of the light curve carrying high information. At these positions the first derivatives predict changes caused by the morphology of the parameter space. Finally it is remarkable that the best initial fit and the original parameter configuration provide similar results. Keeping in mind that the caustic structure for a degenerate solution is similar compared to the true configuration, some gradients have to be similar as well. The overhead caused by calculating the derivative information selection can be minimized, but the Levenberg-Marquardt technique, for instance, requires complete derivatives for each step.
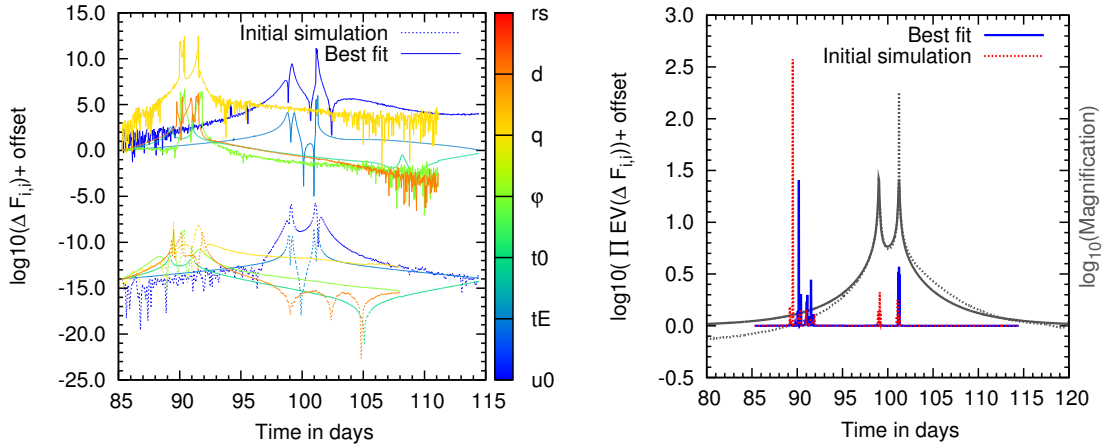
Figure 5.4: Information content $F_{i,i}$ for a given binary light curve (left) calculated for the best fit and the initially chosen parameter combination.

Using different subsets also implies that different likelihoods must be optimized. Whenever a new subset is selected, a new Markov Chain needs to be started. The results of the preceding chains do not need to be discarded; they contribute to the starting point and the initial proposal function. By definition, the new subset will react more sensitively to the proposed steps. Despite an initially lower likelihood, the chains can reach the global minimum quicker not only due to the smaller data set but also due to the ability of leaving local minima more quickly. A lower number of points leads to a less informative likelihood where the samples will provide larger steps. In addition, efforts avoiding pseudo-convergences due to too small steps are reduced.

For the test configuration shown in Tab. 5.1, the chains have been started close to the desired solution. In Fig. 5.5 the reduced χ^2 is shown for different ways of updating the chain and two different underlying light curves. The selected subset always comprises 100 data points while the corresponding complete data set consists of 200 and 500 data points. The regular MHMCMC is illustrated with and without lower limit for the variance of the proposal function preventing the chain to run into a deadlock. The exemplary chains illustrate that a variance control descends more quickly to the desired solution. Fig. 5.5 also reveals one of the drawbacks of the method: At the transition to a new chain, the $\Delta\chi^2$ is reset to the new starting point and thus neglecting the last achieved minimum. By keeping the $\Delta\chi^2$ value, the convergence rate can be enhanced because jumping to a new peak will be prevented. A disadvantage that may become evident is the adaptation of the proposal function, because proposed steps will be less frequently accepted, especially if the likelihood is dramatically changed by the new subset.

5.5 Conclusions

Compression techniques based on the Fisher information have been introduced for the use with microlensing models. The most radical compression reduces the light curve to a data set with n_p points. A PSPL model was studied in the context of this approach and illustrated two major issues when using a pre-calculated grid of compressed binary curves: the light curve needs to be binned and interpolated to match the sampling rate of the compressed binary model. In addition the success of the compression depends on the fiducial model used for compressing all models.
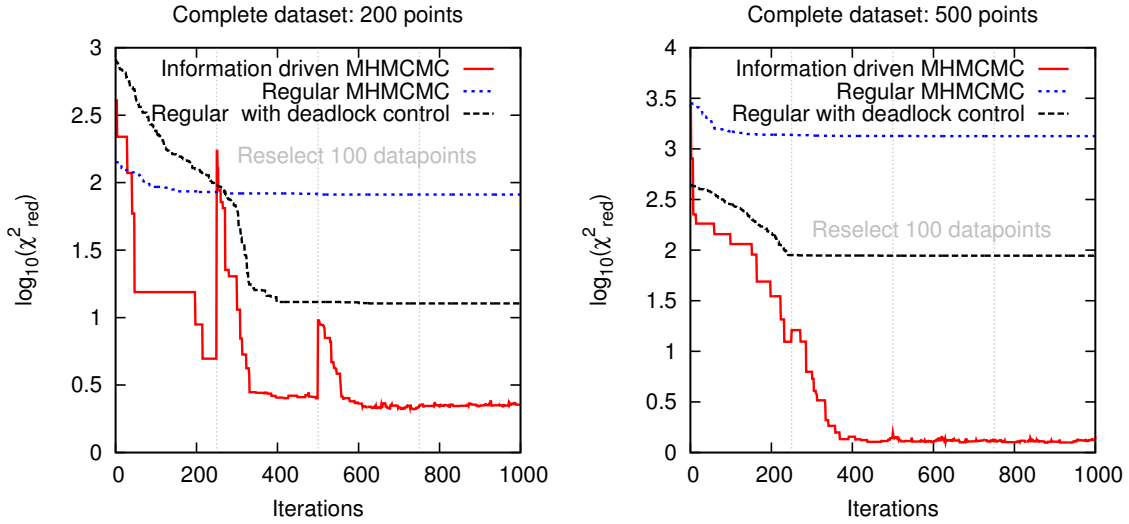
Figure 5.5: Reduced χ^2 for a MHMCMC accelerated by using an information-driven selection of subsets.

An approach that cannot be fulfilled easily. Alternatively, a less radical rejection of data points was developed and tested.

The alternative approach relies on the MHMCMC technique for fitting binary events. After fitting a PSPL model, a grid of short chains is started without formally checking the convergence. The final search for the best solution can be achieved by starting chains at the highest local minima in the initial search. Two complications were identified, both of them related to the finite accuracy of the model. The possibility of reaching a pseudo-convergence for parameter configurations, that are limited by numerical effects needs to be compensated for to avoid running in a deadlock. In addition the likelihood will be systematically underestimated due to the inaccuracy of the model. The result of a MHMCMC approach is a distribution of samples drawn from the parametere space under consideration. Histograms deduced from these samples can be marginalized and depending on the marginalization will provide different modal values also reflecting numerical limitations of the model.

The new idea tested in the context of using MHMCMC for searching the global optimum, is selecting the information carrying points for the current state of the search and continuing to fit the corresponding subset of the light curve. It was shown for a typical example, that such a technique can reach the minimum of the reduced χ^2 faster and with less computational efforts. Due to the difficulties of determining the convergence, regular chains are required for finally fulfilling convergence criteria. One relevant aspect for further studies, is the optimal configuration of an information driven fitting, which was not quantitatively accomplished in the scope of this thesis.

Chapter 6

Modeling observations

After discussing the prerequisites for interpreting data, such as modeling, optimal sampling and fitting techniques, I discuss the observations carried out in the scope of this thesis. This observational part describes follow-up observations using the MONET/North telescope in Texas, the DANISH 1.54m telescope in Chile and the SALT telescope in South Africa. The Follow-up of microlensing events is a highly collaborative effort; the taking and re-reduction of data as well as its analysis is usually carried out by different groups and individuals. Consequently, not all observed events are discussed in detail, as they are part of ongoing research. Here the complete analysis of the microlensing event MOA 2010-BLG-406, observed at the the DANISH 1.54m will be shown which was analyzed using the GPU-contouring model. A second analysis for the event MOA 2010-BLG-477, also observed with the MONET/North telescope, is shown in the following chapter. The latter observation can be used for estimating the expected capabilities of the MONET telescopes for future observations.

6.1   Crowded field photometry

For detecting and characterizing anomalous microlensing events, i.e. events which are not well described by a PSPL model, feedback strategies like those as shown in Fig. 7.2 have been established. The basic idea is to observe microlensing events and to report brightness measurements as quickly as possible to a central repository accessible to all other observers (Dominik et al. 2008). Incoming data points can be processed and potential deviations can be used for triggering additional observations (Dominik et al. 2007). According to the assessment of this initial analysis, further observations can be prioritized (Horne et al. 2009). Clearly, the feedback system depends crucially on the processing time, given that planetary or other low-mass perturbations are of short duration. For this purpose an adequate data reduction system needs to be prepared.

A series of astronomical images carries the information for a microlensing event, accessible as the brightness variation of a single object in a potentially crowded field. In order to extract the corresponding light curve, one has to address the issue of blending by other stars. The optimal image subtraction technique introduced by Alard & Lupton (1998) has become one of the standard approaches for this analysis. As microlensing light curves consist of observations covering several nights, the observing conditions between different images can differ substantially. In order to reduce the effect of blending, frames can be subtracted, demanding some preparatory work.

The basic idea of optimal image subtraction pipelines, which are predominantly used for crowded field photometry, is to construct a high quality reference frame based on parameters like seeing and background level. Afterwards a spatially varying convolution kernel (Alard 2000) is
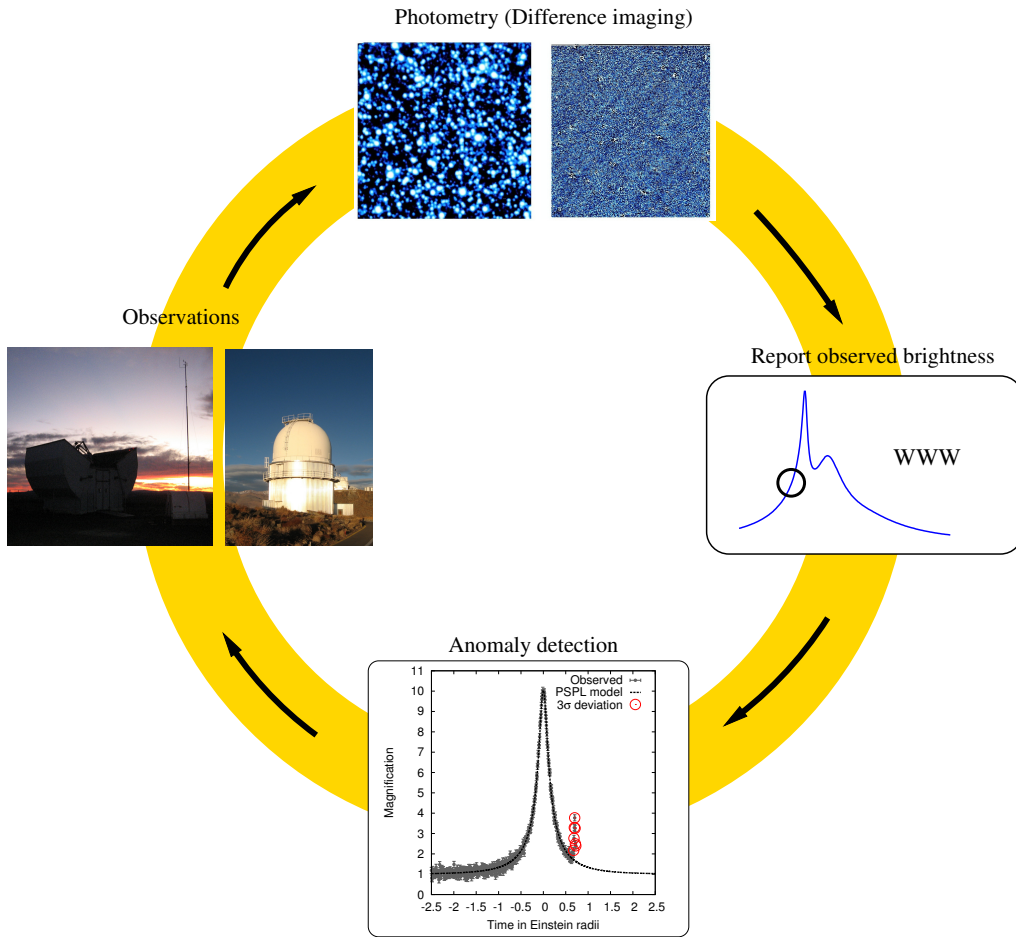
Figure 6.1: Follow-up scheme for gravitational microlensing.

determined and applied to each data-frame, thereby compensating for different observing conditions, and exposure times. In addition, potential gradients on each frame can be compensated for bivariate polynomial. In short, the difference image D is constructed from an image I to be analyzed and the reference image R according to

$$D = I - (K \otimes R) + B, \tag{6.1}$$

where the kernel K consists of a superposition of modulated Gaussian basis functions as chosen in the original approach (Alard & Lupton 1998). A more general and flexible approach was introduced by Bramich (2008), who considered each pixel of the kernel to be subject to the adaptation which is done in a least square sense for the residuals constituting each difference frame. The final subtraction effectively cancels out stars with constant brightness while variable stars remain. The final light curve is obtained by fitting a model of the Point-Spread-Function (PSF) to the residing PSF of the variable object on each difference frame. An example for such a subtraction can be seen in Fig. 6.2 showing MONET/North data of the event MOA 2010-BLG-406.

Most pipelines comprise a similar number of preparatory steps which shall be described in brief. Here it is assumed that all frames are taken with a charged coupled device (CCD) and
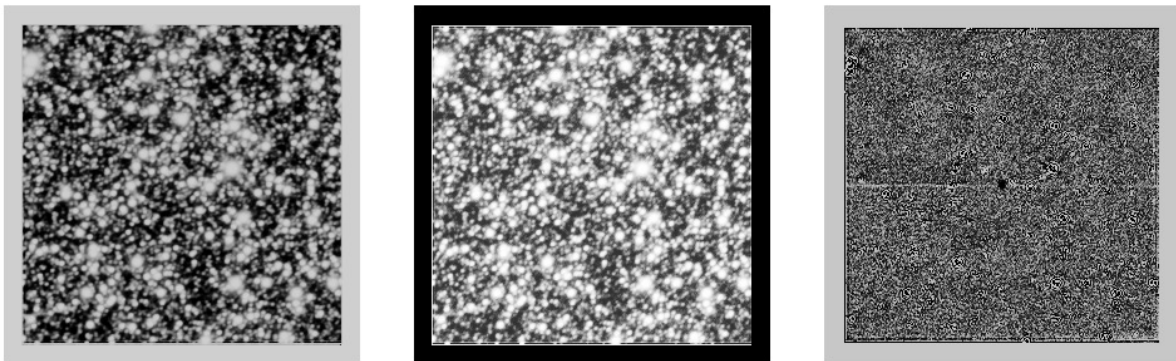
Figure 6.2: Example of a difference image (right) observed from the DANISH 1.54m telescope. An aligned, resampled and convolved version of the image (middle) is subtracted from the reference frame (left). The residual PSF at the center is caused by the microlensing event MOA 2010-BLG-406.

calibrated using at least bias and flat-field frames. The latter corrects for vignetting and pixel-to-pixel variations caused by dust on filters and the detector itself. Images taken with CCDs having limited cooling capabilities requiring the subtraction of a dark current, which is generated by thermal electrons. Depending on the cooling capabilities, dark frames taken with closed shutters can improve the image quality. The liquid nitrogen cooling of the Danish Faint Object Spectrograph and Camera (DFOSC) is sufficient for providing good results without dark frame correction ($T_{CCD} \approx -120°$ C), while the Peltier cooled Apogee Alta E47+ CCD, which was used by MONET/North, demands a dark correction ($T_{CCD} \approx -20°$ C).

Depending on the telescope, the pixel position of an object can differ due to pointing inaccuracies. In order to avoid subtracting and determining the convolution kernel with a varying amount of underlying stars, images are cropped to common subframes. The chosen technique for aligning the frames is already affected by the demanded properties of the resulting light curve. In this context especially two characteristics of the resulting frame are changed, namely flux conservation, needed for estimating the uncertainties using the camera gain and correlated noise in neighboring pixels (the latter is particularly discussed by Albrow et al. (2009). In the context of this work, two different pipelines are applied for re-reducing the frames: one pipeline (DIAPL) is based on the difference imaging package introduced by Wozniak (2000) relying on a kernel consisting of polynomial modulated Gaussian basis functions; in the second approach (DanDIA), the Gaussians in the basis function are replaced by δ-peaks creating a numerical kernel at every given position, by Bramich (2008). Both approaches can be time consuming, but the latter is especially so, so the reduction to 256×256 pixel subframes is also essential for reducing the computational burden and providing quick results.

6.2 Data reduction pipeline

For a real-time reduction of binary events the optimal image subtraction packages need to be integrated in an automatic system that takes care of the calibration and subsequent calls of all packages for reducing the data. For this purpose the DIAPL package based on the work of (Wozniak 2000) was chosen. Despite the advantages of the numerical kernel approach, quick re-reductions of a real-time light curve update can be achieved more quickly with the modulated
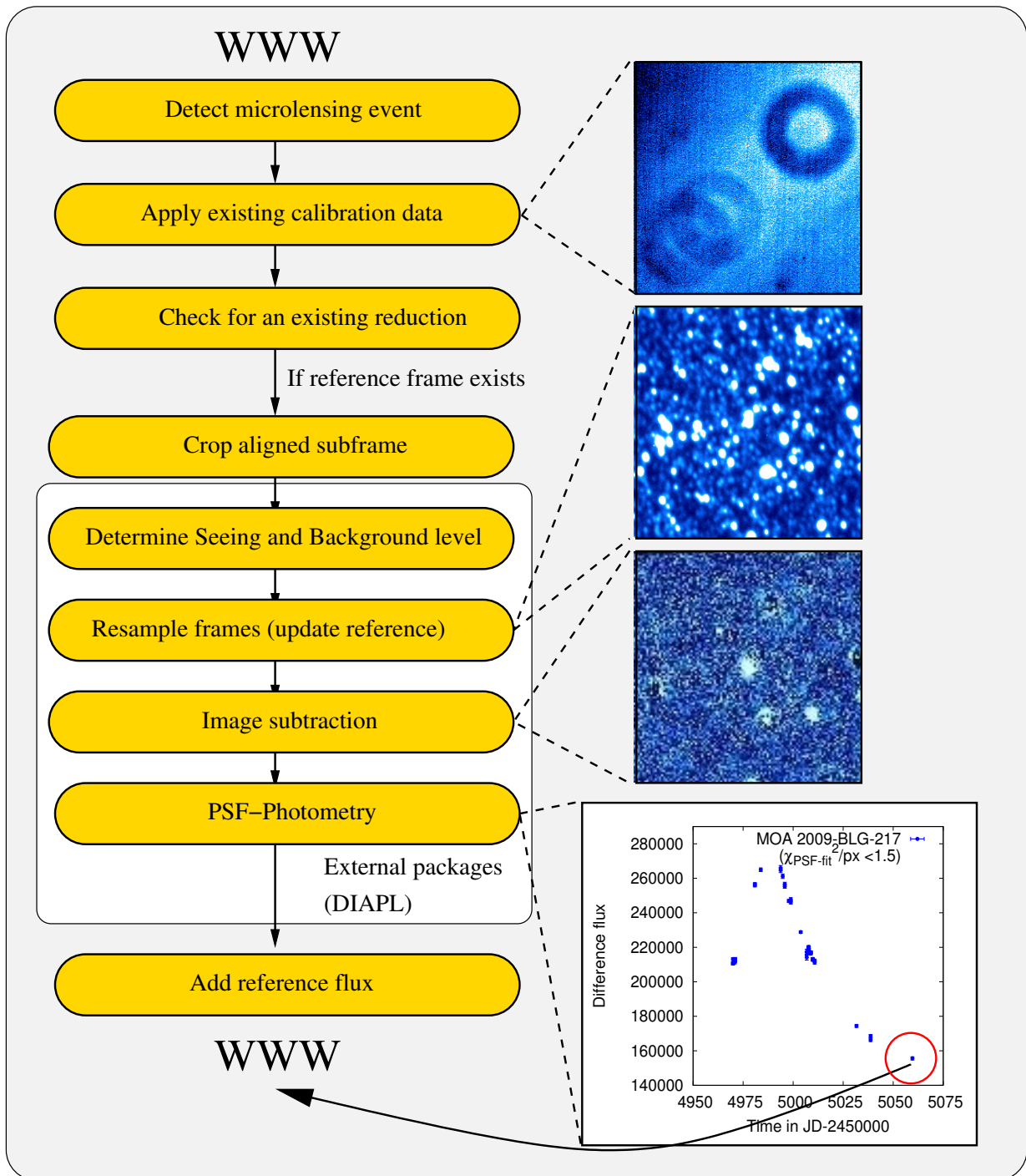
Figure 6.3: The data flow in the implemented reduction pipeline is shown starting from an web-based prioritization tool and ending with the submission of new data points.

Gaussian kernel. Data taken with the MONET/North telescope was processed remotely, as data access is limited due to its location: the Galactic Bulge is only visible for 1-2 hours per night at ≈ 2 airmasses, i.e. of the order 50 frames need to be processed. The MONET/South telescope will contribute more observations due to better visibility and thus shorter exposure times. The

automatic data reduction system for MONET/North was available in 2010 for a first test run. The data flow is depicted in Fig. 6.3, showing how the exemplary event MOA 2009-BLG-217 was processed.

As indicated in the illustration, the essential routines for subtracting the images were taken from the DIAPL package. For the use in an automated realtime reduction, they had to be arranged differently in contrast to the provided framework. The initial registration and resampling of frames does not have to be repeated for every incoming frame and is kept. If the initial seeing estimate cannot be made, e.g. because of a cloudy night, the entry is immediately removed from the list and cannot lead to numerical difficulties in the subtraction process. The nightly image subtraction is also performed for each single frame as long as the reference is still close to the best seeing frame. Following this approach, a brightness measurement can be reported within ≈ 1 min. Updates of all reference frames after a night with significantly improved seeing conditions, are carried out after daybreak.
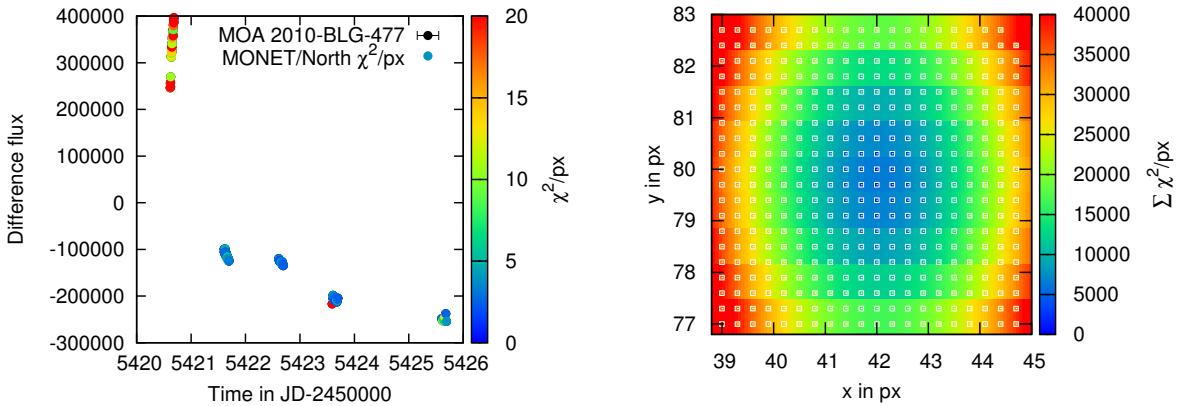

Figure 6.4: The χ^2/px obtained from the best fit of the PSF position is shown for each point of the light curve (left). The grid of the total least squares per pixel for all points of the reported light curve are shown along with a paraboloid fit to the surface (right).

Test observations have lead to insights for future improvements of the data reduction and shall be summarized in the order of their appearance. Usually, the best seeing frames are selected for the final-data reduction but, due to possible differing background gradients, the safer option is to use all the frames of the best seeing night. A completely empirical selection can be done by testing different combinations and minimizing the sum of the least squares of all difference frames, which can lead to optimized sorting keys[1]. In a similar way, the number and size of Gaussian components can be optimized. For very inhomogeneous observing conditions it can be beneficial to consider two parameter configurations of the pipeline as independent light curves due to the different blending ratio. Fig 6.4 show how the photometry can be enhanced if the PSF is relocated. Resampling frames for compensating image distortions may cause correlations, but this approach is kept for Gaussian basis functions as the kernel is shifted on a pixel grid contrary to the numerical kernel approach where resampling can be switched off.

The initial position of the microlensing events does not necessarily lead to the correct pixel coordinates of the residual PSF on the difference frames (Albrow et al. 2009). Nevertheless, all

---

[1]Corresponding tests were done in the Bachelor's thesis by Andreas Boesch: Optimierung einer Reduktionspipeline für Mikrogravitationslinsenereignisse (2010)

existing pipelines report the χ^2/px giving an insight how to center the PSF model. Given a grid of assumed PSF positions, the total least squares for all PSF-fits can be determined and indicates clearly where to put it; otherwise the light curve quality will deteriorate quickly. Assuming the fitting noise given by the χ^2 is independent of the photon noise, the contribution to the reported uncertainty is on average 0.2 % and in the worst PSF-fit 5 %, which is still negligible. If the PSF position is not accurate to one pixel, the relative contribution increases by an order of magnitude, which is alarming and must be included in the uncertainty estimation. Even better, it can be avoided by the use of a better PSF-model, e.g. by using an empirical PSF-model as in the DanDIA pipeline.

6.3   Observed events

In the context of this thesis different observing runs were carried out using telescopes on different sites. In 2008, data was taken by the South African Large Telescope (SALT) during its commissioning phase. A follow-up of the event OGLE 2010-BLG-510 was triggered as a target of opportunity and observed in service mode. The main mirror of SALT consists of 92 mirror segments and those were not optimally aligned: a point spread function (PSF) consisting of multiple images of all objects in the crowded fields limited the photometric accuracy to $\approx 100\,\text{mmag}$. In 2009, the MONET/North telescope in Texas was available for test observations. The quality and coverage of observations was limited by weather conditions combined with observations at low altitude, i.e. ≈ 2 airmasses. The best coverage was achieved for the anomalous event MOA 2009-BLG-217, illustrated in Fig. 6.3, where 70 data points were contributed. In the 2010 season, two weeks of observations with MONET/North and two weeks of observations at the DANISH 1.54m telescope were carried out as part of the MiNDSTEp consortium. In the following, the event MOA 2010-BLG-406 is discussed in detail and indications of the nature of the event MOA 2010-BLG-477 (see Fig. 6.4) are given on the basis of data from the DANISH 1.54m telescope and the MONET/North telescope, giving an insight into the current capabilities of the robotic MONET telescopes.

6.4   Modeling of the event MOA 2010-BLG-406

The gravitational microlensing event MOA 2010-BLG-406 was followed up in the 2010 observing campaign at the 1.54m DANISH telescope as part of the MiNDSTEp consortium. Data was taken after having been triggered by the SIGNALMEN anomaly detector (Dominik et al., 2007) starting on July 10th, 2010. The data was reduced using both the DIAPL and the DanDIA pipelines (Bramich, 2008). Despite the improved capabilities of the numerical kernel approach, the first part of the analysis relied on a reduction based on Gaussian kernel functions.

6.4.1   Preparing the data

The available online data from the MOA collaboration represents several years of observing and thus provides an accurate baseline for fitting purposes, but instead of simulating all these points they are all collapsed to one highly accurate baseline measurement. Their propagated weight ensures that the linear baseline and blending is well constrained, reducing potential ambiguities. From the Fisher matrix it is known that the information content of the blending flux is constant for each point of a light curve and for extremely low magnifications the same applies to the baseline flux; $5\,t_\mathrm{E}$ away from t_0 the magnification is changed by less than 0.3 %. The microlensing

event itself occurred between the Julian dates 2455370 and 2455410, and the weighted mean of all other sufficiently accurate data points was used for constraining the baseline and blending flux, where sufficiently accurate means the reported uncertainties were demanded to be below three times the median uncertainty. In Fig. 6.5 the steps for preprocessing the MOA light curve for a binary fit are shown.
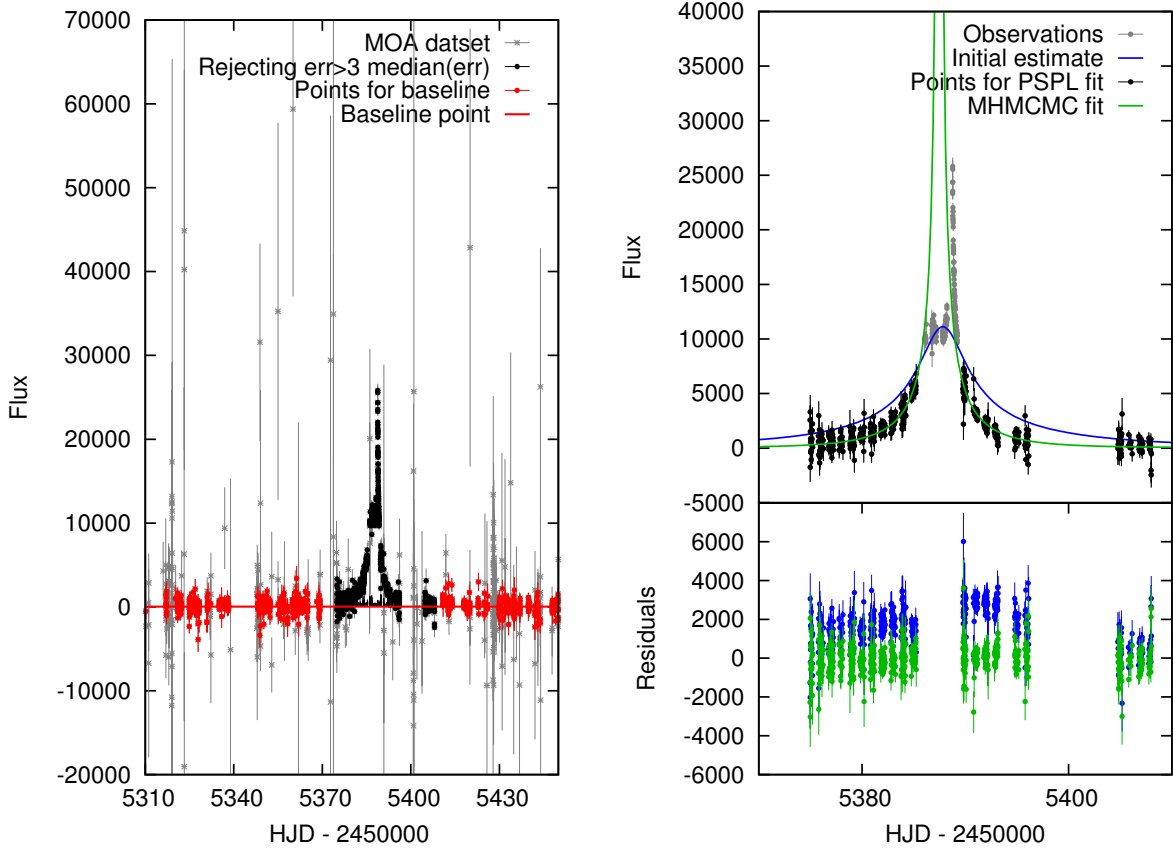

Figure 6.5: Data preparation for binary fitting by rejecting data points exceeding three times the median and by averaging all baseline points (left). After an initial parameter estimation a MHMCMC fit is carried out.

After rejecting inaccurate data points, the standard parameters of a PSPL event u_0, t_E and t_0 are determined assuming an unblended event. The maximal observed magnification can be translated to the parameter u_0 as $u_0 \approx 1/\mu_{\max}$. Due to the anomalous nature of the light curve t_0 and t_E cannot be determined very well. For initial parameter estimates the light curve is supposed to be of Gaussian shape and by applying the rejection sampling method from von Neumann (1951), it is possible to sample from the observed times. The expected value yields the time of maximal magnification and the standard deviation σ_t can be applied for estimating the Einstein time via

$$t_E \approx 0.68 \cdot \mu_{max} \sigma_t, \tag{6.2}$$

according to Eq. 2.34. In contrast to other methods this approach is less sensitive with respect to the anomalous shape of the light curve and requires just a small number of evaluations. Based on this estimate, all points exceeding 3σ defined by the residuals, are rejected in the final PSPL

fit. Due to the anomalous shape of the light curve, the wings provide the requested information about the underlying PSPL-model. Unfortunately, the Einstein time can only be determined properly if the parameter u_0 is known; which cannot be unambiguously determined for the given measurements.

For obtaining an initial parameter estimate, a PSPL model is fitted by starting starting several Markov chains. The variance within and between different chains were used for checking if the chains have converged and samples have been used for determine the uncertainties. The result of this analysis is a multidimensional probability distribution, which can be used to simulate or calculate all other properties. If the finalized chain elements are kept, different Galactic models or distributions can be used for inferring the physical properties by drawing samples from the chain according to the prior probability. The initial estimates of the PSPL parameters shown in Tab. 6.1 serving as starting point for further analyses. The uncertainties are determined after reaching convergence and the resulting estimates are shown in Tab. 6.1.

u_0	t_{E} in d	t_0 in d
$6.6583^{+0.0032}_{-0.0034} \cdot 10^{-5}$	$7.606^{+0.023}_{-0.022}$	$5387.46681^{+0.00074}_{-0.00072}$

Table 6.1: PSPL fit parameters for the initial grid search.

6.4.2 Initial grid search

For investigating potential configurations reproducing the characteristic features of an anomalous microlensing light curve, the GPU-based binary model introduced in chapter 3 was applied to model the event. A preliminary grid search was carried out relying on 600 starting values of $\log(q)$ and $\log(d)$. At each of these grid points, 250 steps of an MHMCMC search were simulated, just without convergence control. At this stage, the finite source size is kept fixed at $0.015\,\theta_{\mathrm{E}}$: no prominent caustic crossing features can be seen, so by fixing the source star radius at a comparably large radius, the computationally faster unrefined GPU-model can be used.

The microlensing model parameters obtained in different studies make use of the freedom to define source-plane coordinate systems. As pointed out by Skowron et al. (2011), a common format for reporting results would ease comparing the results of different models. However, different coordinate systems chosen to support modeling efforts are following the variety and characteristics of binary microlensing light curves. For an initial estimate of the PSPL parameters of the given event, a close binary configuration with separations $d < 1\,\theta_{\mathrm{E}}$ can be efficiently determined if the center of mass is considered as basis for modeling. The case of widely separated binaries can be more appropriately addressed by using models centered on each of the companions. Thus, for this study, each random MCMC walk was repeated for three initial positions, namely center of mass and the positions of the companions.

The total computation time required for such an approach is still large and thus the amount of data was reduced by selecting a subsample from both data sets. As shown before, a massive data compression needs a model for appropriately rejecting data points. In order not to introduce a selection bias the subset used in this context has been selected randomly, by accepting data points according to their reported brightness as high magnification values are less likely and thus, according to Shannon's information definition, more informative. In addition, points close to gaps were given heigher weight, reducing parameter degeneracies caused by gaps. Fig. 6.6

illustrates snapshots in the fitting process with local minima caused by the limited coverage of the event. There are also other reasons for degenerate parameter configurations that may affect the analysis as, e.g. symmetric binary light curves, parameter degeneracies, but intrinsically degenerate solutions can emerge from symmetries in the lens equation itself (Dominik 1999).
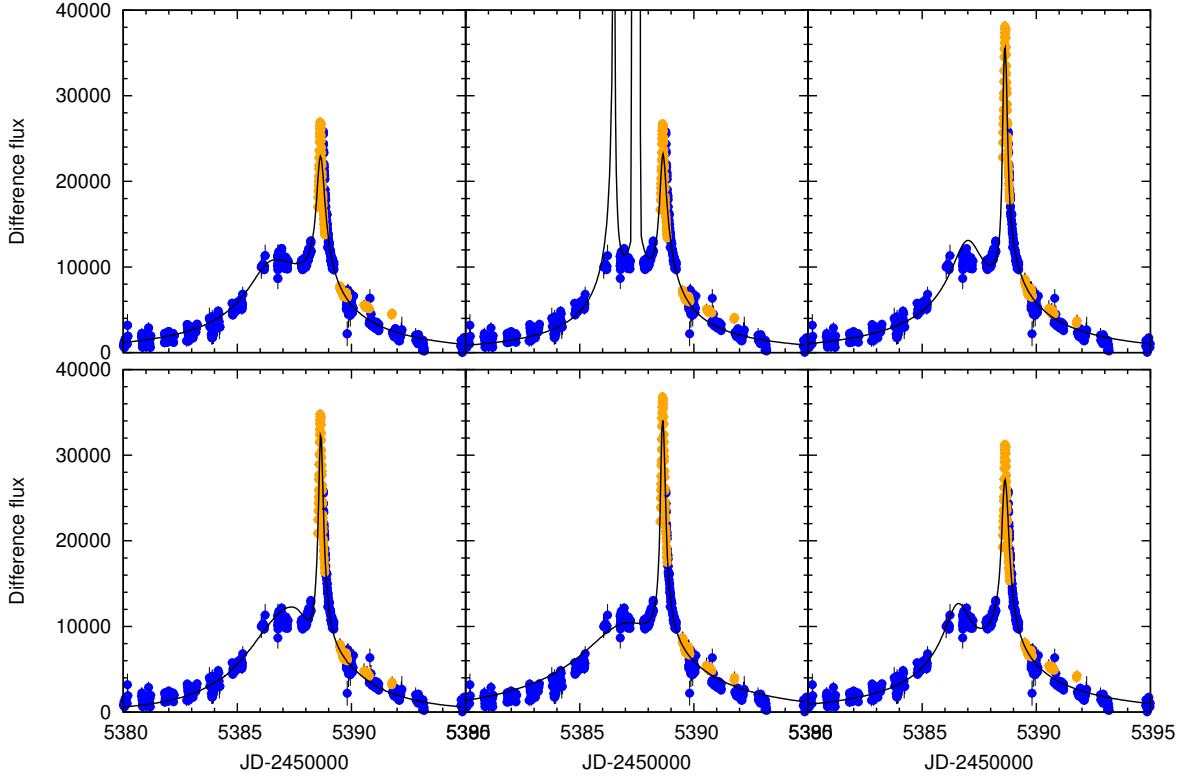

Figure 6.6: The coverage related parameter degeneracy, limiting the possibility of determining the best solution, is shown for local maxima in the course of the initial grid search.

The initial grid search depicted in the left plot of Fig. 6.7 provided local minima, that serve as convenient starting points for further Markov chains, as shown in the right plot. The χ^2 value is used to compare the solutions offered by these approaches and to decide if the measurements are well described, assuming that it is much easier to convolve the corresponding distributions with a prior which may need to be altered. The uncertainties of the photometry have been kept as provided: in the initial DIAPL reduction, a Poisson-like estimate is provided according to the total Flux F_{tot} measured by aperture photometry on the unsubtracted image (Wozniak, 2000) by

$$\sigma_F = \sqrt{\frac{F_{\text{tot}}}{\text{gain}}}, \tag{6.3}$$

This may result in an unbalanced view, as the reported uncertainties are not necessarily compatible with the scatter of the data points, but for the given case this seems to be suitable. The fitting radius was also adapted to be less affected by outliers, which may have been caused

by measurements in the nonlinear regime of the detector, where the number of electrons is not proportional to the number of collected photons. Nevertheless, the resulting χ^2-values are in a reasonable range, and thus attempts to rescale the reported uncertainties were skipped.
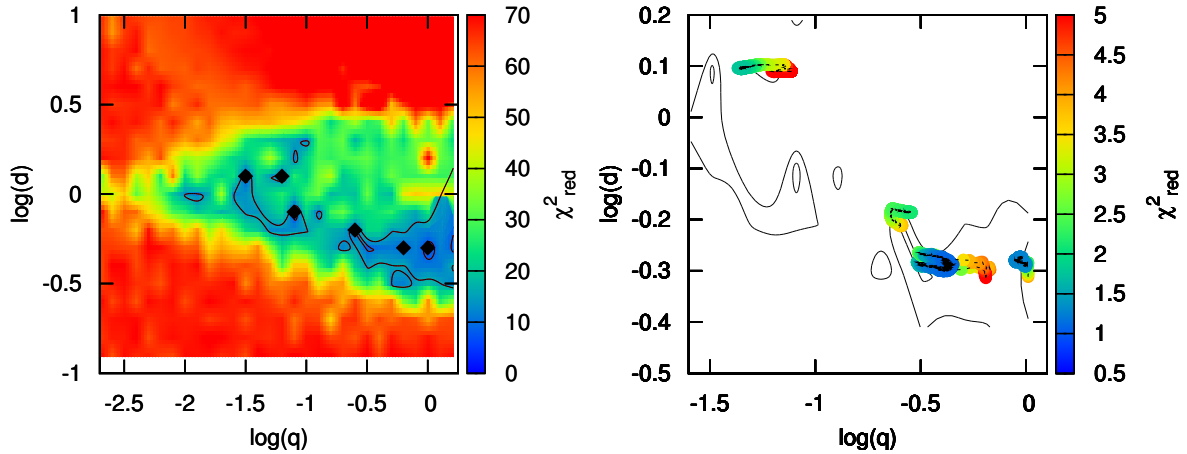

Figure 6.7: Grid search for determining appropriate starting points for subsequent Markov chains (left) and first results from a refined analysis of these events (right).

6.4.3 Refined study of local minima

Before starting longer chains for the selected measurements in Fig. 6.5 and observations from the DANISH 1.54m, accuracy issues with the reported time-stamps need to be discussed. The analysis relies on the reported Julian date which has been determined according to the reported coordinated universal time (UTC). The orbital motion of the Earth and the observatories separated by several 1000 km can produce an effect of several minutes. For an event lasting several days, the resulting change is of the size of a source star or of the exposure time. Timing accuracy should not affect the results, for MOA 2010-BLG-406, as it does not exhibit caustic crossing features, where uncorrected dates could have affected the estimates. Being able to include constraints from space-based observatories as discussed by Graff & Gould (2002) and Gould et al. (2003), would rely on more accurate time-stamps. To compensate Rømer's delay caused by the finite speed of light, different approaches are followed. Most commonly the Heliocentric Julian Date is calculated, transferring the time-stamp reference system to the center of the sun. A more accurate correction has been recommended by Eastman et al. (2010) and is used here for finalizing the parameter estimations: the final reported times are given as Barycentric Julian Date in the Barycentric Dynamical Time standard BJD_{TDB}. The correction applied to the given observations represents a maximal offset between the time-stamps of both sides of $\approx 30\,\mathrm{s}$ relative to the initially reported JD_{UTC}. This correction can be neglected for fitting a standard binary model without diurnal parallax, orbital motion or multi-site parallax effect.

The analysis was carried out by starting multiple chains from local minima of the χ^2-surface as shown in Fig. 6.7. For this purpose, the prepared MOA and complete DANISH 1.54m measurements are used and Markov chains were started for both rereductions of the observed DANISH 1.54m light curve. Fig. 6.8 illustrates that the chains end in similar regions of the parameter space for both reductions and, despite differently reported uncertainties, the χ^2 value provides a similar answer. The reported photometric uncertainty in the DANDIA pipeline particularly includes
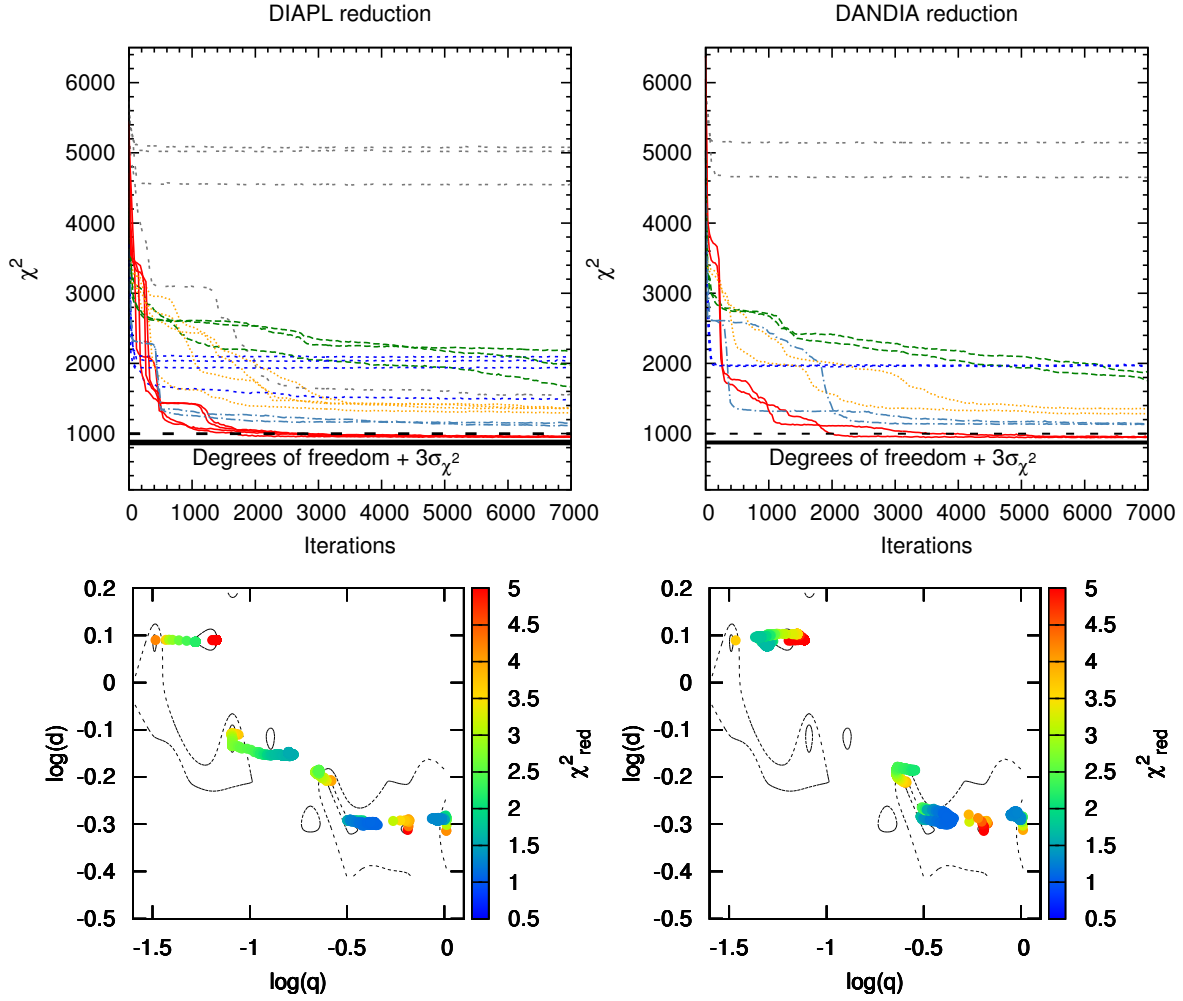
Figure 6.8: Resulting χ^2 for Markov chains started at local minima shown for two reductions of the DANISH 1.54m follow-up observation (top) and their evolution on the $\log(q) - \log(d)$ plane (bottom).

two contributions which were neglected in the DIAPL pipeline, i.e. readout noise and the noise contribution caused by the reference frame. Because of using a stacked reference frame in both cases, there is only a minor effect. The best fit in a least-squares sense is used for comparing how the residual noise is affected by different reductions and, according to Fig. 6.9 it is evident that systematic deviations at the peak were reduced in the numerical kernel approach. Consequently, the numerical kernel rereduction is used for inferring parameters.

In addition to reaching a low χ^2, it is important that the best solutions sample from the same distribution. In some cases different starting points leave the tested region immediately or step towards other solutions and even for chains close to the best solution, the formal convergence criterion was not achieved in the course of the fitting process. The PSPL-parameters especially did not comply with the r-value criterion Eq. 5.8. The adaptive step-size control is not sufficient as long as it relies exclusively on the variance of each chain. For achieving convergence, the variances of the proposal function were adjusted according to the variance between 5 chains, which were
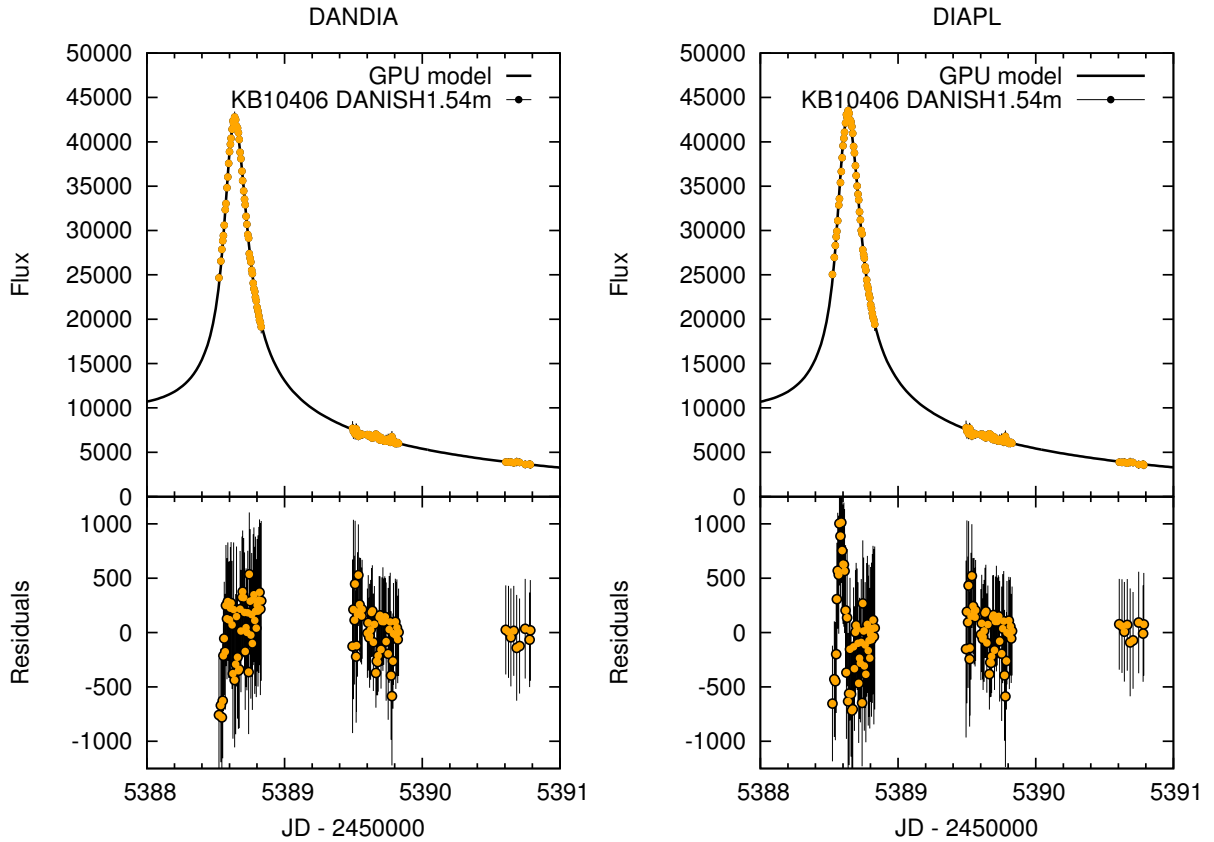
Figure 6.9: Comparison of light curve fits of the numerical kernel approach (left) and the kernel consisting of modulated Gaussian basis functions illustrating the resulting change in data quality.

started at the lowest χ^2 value of all chains. One possible explanation why the chains did not formally converge is the limited numerical accuracy of the input parameters, leading to pseudo-convergence as the numerical noise is sampled instead of the required distribution. By using the variance between different chains, the proposal function better reflects the local parameter space and converges in all dimensions efficiently. The best fit is achieved for the light curve shown in Fig. 6.10.

The result of the MHMCMC is a sample providing likelihood distributions for all parameters and all possible marginalizations. The important special case of the two dimensional marginalization is shown in Fig. 6.11, illustrating that no unimodal distribution is achieved. In principle this can be avoided by restarting chains with constraints on the parameter space or by selecting the global mode. In fact, each mode shown in the likelihood plot corresponds to the same χ^2 (Fig. 6.11) and thus these modes need to be included for getting an unbiased view. Each mode reflects the artifacts and structures of the observed data as well as accuracy limitations due to the model. Given the (single precision) GPU-model, these multimodal distributions provide the most sincere interpretation that can be given in that context.
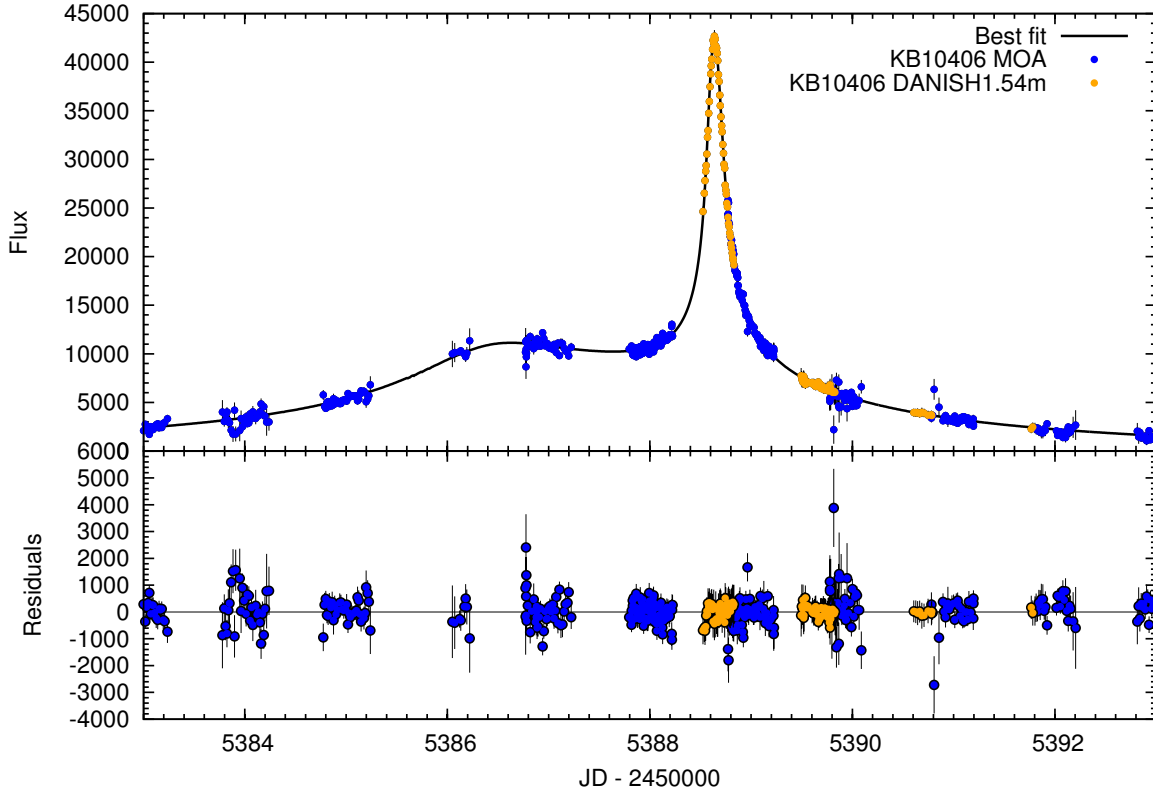
Figure 6.10: The best fit binary lens light curve simulated with the GPU-contouring model is shown for MOA 2010-BLG-406 along with the corresponding residuals.

6.4.4 Physical parameter estimates

If the observed data does not intrinsically provide estimates for the distances to lens and source star as well as the lens-source proper motion, a Galactic model is required to deduce the results by deducing the physical parameters given the fit parameters. Instead of directly converting the fit results based on defined parameter distributions, samples are drawn from a catalog simulation of the Besançon-model. Pairs of lens and source-stars are drawn from these catalog samples in

Reported	u_0	t_E	t_0	φ in $^\circ$	q
Median	$0.1377^{+0.00057}_{-0.0019}$	$6.763^{+0.100}_{-0.046}$	$5388.0297^{+0.0049}_{-0.0053}$	$56.03^{+0.22}_{-0.22}$	$0.4125^{+0.0082}_{-0.0072}$
Mode	$0.1371^{+0.0011}_{-0.0013}$	$6.784^{+0.079}_{-0.067}$	$5388.0216^{+0.013}_{-0.0028}$	$55.70^{+0.55}_{-0.11}$	$0.4149^{+0.0058}_{-0.0096}$

d	$r_\star$	g_{MOA}	g_{DANISH}	χ^2
$0.5297^{+0.0014}_{-0.0029}$	$9.2814^{+0.0049}_{-0.019} \cdot 10^{-3}$	$-0.98689^{+0.00020}_{-0.00021}$	$-12.393^{+0.18}_{-0.088}$	$944.9^{+3.2}_{-2.0}$
$0.5308^{+0.00031}_{-0.0040}$	$9.2830^{+0.0034}_{-0.021} \cdot 10^{-3}$	$-0.987125^{+3.0\cdot10^{-5}}_{-4.5\cdot10^{-4}}$	$-12.44^{+0.13}_{-0.13}$	$944.0^{+4.1}_{-1.1}$

Table 6.2: Sample parameters deduced from one dimensional marginalizations of the converged MHMCMC chains
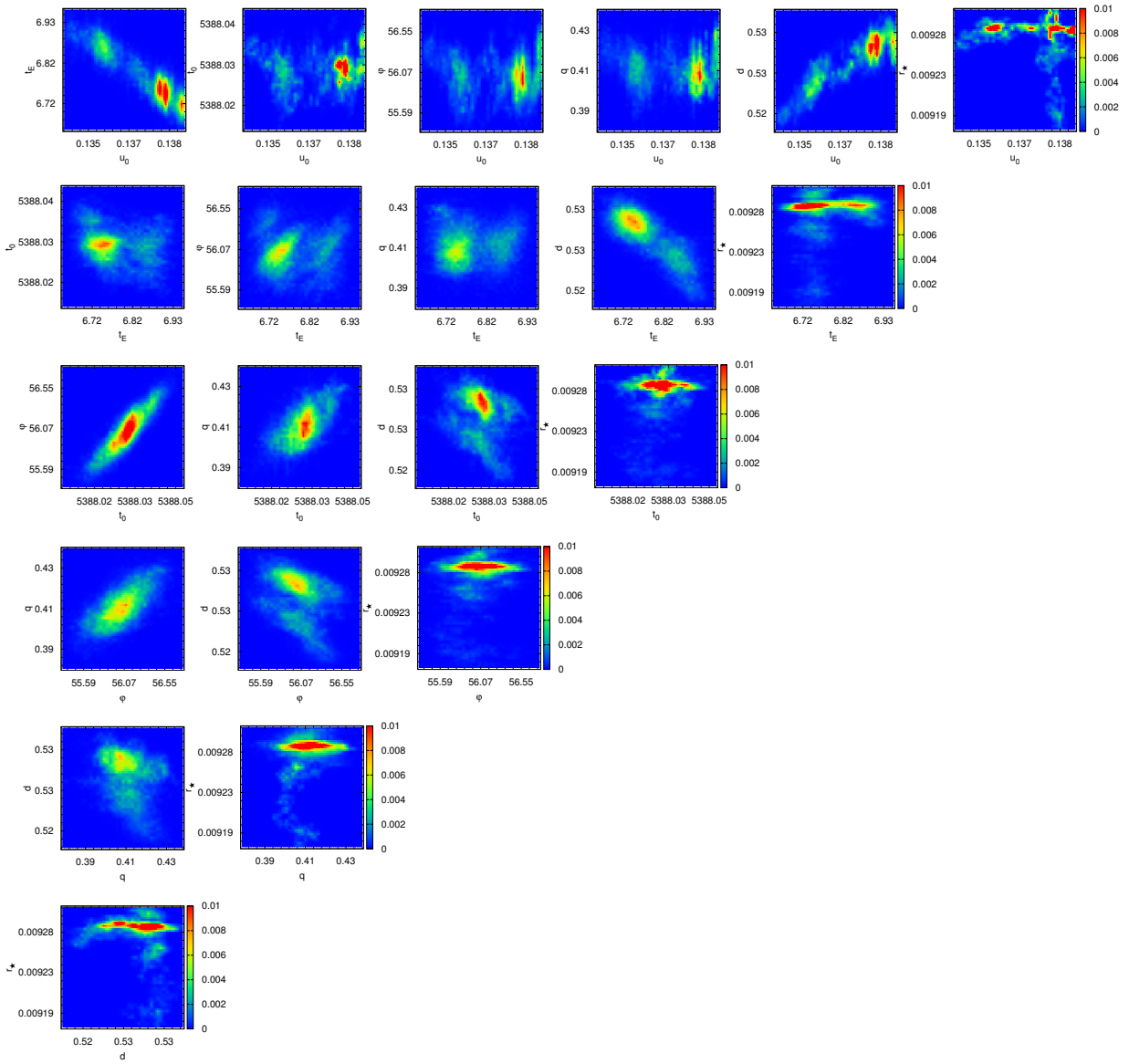
Figure 6.11: Probability of binary lens parameters and χ^2 as two dimensional marginalizations of the final Markov chains.

order to infer the mass of each component.

One relevant issue for such a resampling approach is to guarantee that a sufficiently large number of stars is available in the corresponding data set, supposed to be compatible with all cuts applied to the distribution. In addition to the parameters from the fit, position and baseline magnitude from MOA (Bond et al., 2001) can help to constrain potential configurations. Regarding these limitations leads to the setup for the catalog simulation which shall be briefly summarized.

A high blend ratio indicates that baseline and blend flux are of the same order of magnitude, so thus the reported baseline magnitude gives an indication what kind of source stars are consistent with the fit results. For converting the fit parameters, the apparent magnitude of the source star is
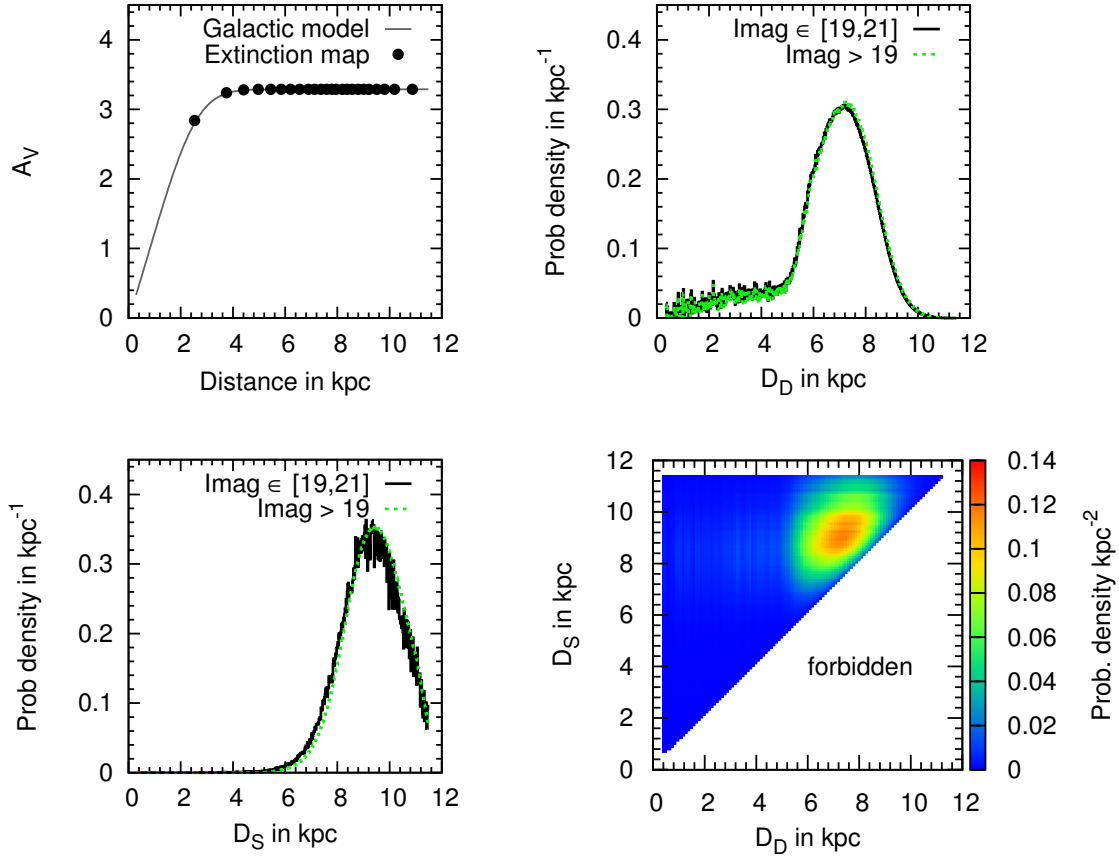
Figure 6.12: Control plots for the extinction in comparison to the extinction map (Marshall et al., 2006) are shown (left). In addition the 1D marginalized distributions for source and lens position for the Besançon model are shown for different apparent magnitude cuts (middle) as well as their 2D marginalized representation (right) . The best fit light curve simulated with the GPU-contouring model is shown for MOA 2010-BLG-406 along with the corresponding residuals.

assumed to be in the range of 20.5 ± 1.5 Imag. For properly determining the apparent magnitude, the extinction of the given field has to be known. The relative extinction was adapted for matching the closest distribution in the maps of Marshall et al. (2006) at $l = 358.00°$, $b = -3.25°$ with $A_{KS}/A_V = 0.1$. The corresponding relative extinction was set to 1.172 mag kpc^{-1} in the visual band. Stars were simulated in the range of $[0, 11.5]$ kpc, for constraining the size of the output files, which covers the expected range of stars as shown in chapter 4. The field of view was set to 0.002 square degrees and no cut was applied for the apparent magnitude, because faint stars can serve as lenses, but proper motions are required. For all other parameters the standard configuration was kept.

Control plots for the parameter distributions are given to ensure that the observed values can be analyzed; the corresponding results are shown in Fig. 6.12. In contrast to the OGLE field discussed in Chapter 4, the expected source position is shifted to positions of ≈ 9 kpc, which has

only a minor effect on the expected mass, as the Einstein radius is changed by 1.2% in comparison to the typical configuration $D_D = 6.5\,\mathrm{kpc}$, $D_S = 8.5\,\mathrm{kpc}$. The empirical distributions extracted from the catalogue simulation of the Besançon model are used to simulate lens-source distances by consecutively drawing potential parameter pairs $(D_\mathrm{D}, D_\mathrm{S})$. Such an initial distribution serves as the envelope distribution for drawing samples from the Galactic model. Applying this technique leads to the distributions shown in Fig. 6.12 indicating that the fitted parameter configuration does not occur frequently.
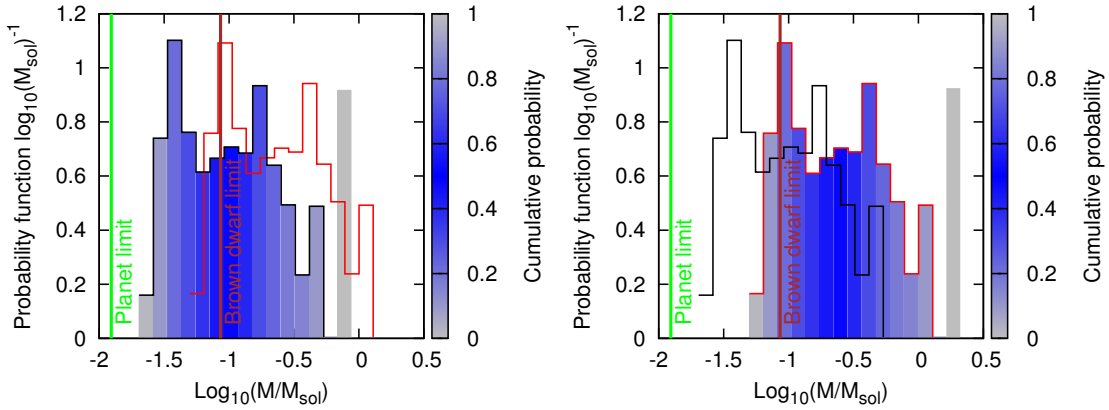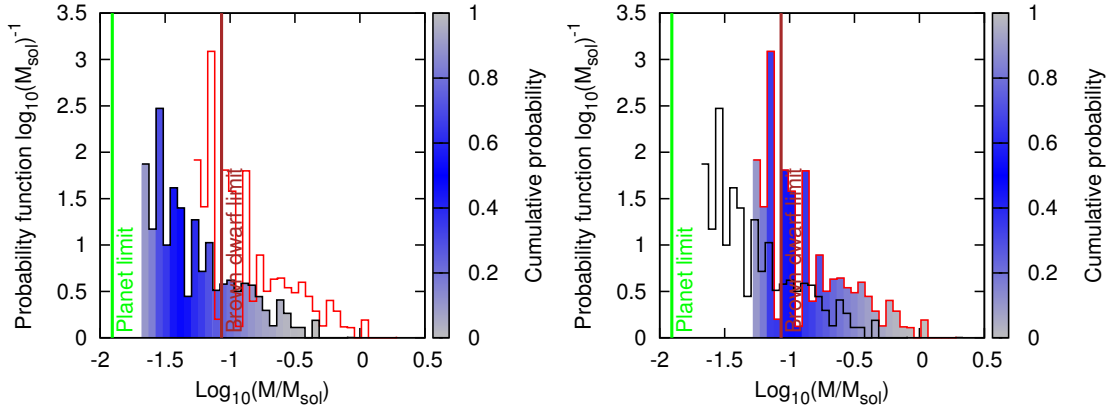

Figure 6.13: Mass of the binary components of MOA 2010-BLG-406 determined according to a catalog simulation based on the marginalized distribution of Einstein times.

If samples are drawn from the simulated star catalog according to the observed Einstein time, the potential mass distribution of the components of the binary system can be determined and leading to the distributions shown in Fig. 6.13. The planetary and brown dwarf limits are indicated as 13 and $90\,M_\mathrm{Jupiter}$ (Spiegel et al. 2011; Close et al. 2005) depending on the exact physical properties of the lens star, which are not accessible here. The t_E-based sample provides different alternatives for the given parameter configurations. The system is likely to be a binary star and not a planetary system. The low mass companion can also be explained by a brown dwarf. If the remaining samples are demanded to be consistent with t_E and the source star radius in θ_E it turns out, that no source star is consistent.

Recalling that the lens and source positions were drawn according to the lensing probability and the potential number of stars, one can consider an additional analysis for arbitrary lens and source star positions, selected exclusively to match the fit parameters t_E and $r_\star$. The range of source star radii is rather small and so is the number of stars complying with such a range. The (single precision) GPU-model is not able to address limb-darkening due to the required small annuli. The model accuracy itself is also affected by the computational limitations of the model and finally the reported stellar radius in the Besançon model is estimated based on the bolometric magnitude while the fit parameters are obtained from observations in a specific band. Altogether, the reported uncertainty seems to be to small and thus for selecting samples a $\pm 10\%$ range was accepted, resulting in a configuration that can be more easily achieved, as illustrated in Fig. 6.14. As expected, the lens-source positions are changed in comparison to the initial assumption. It is especially noteworthy that the lens-source distance D_DS is changed to relatively short distances around 0.18 kpc, according to the parameter estimates shown in Tab. 6.3. The mass distribution

Figure 6.14: Mass of the binary components of MOA 2010-BLG-406 determined according to a catalog simulation based on the marginalized distribution of Einstein times and the source star radius given in units of the Einstein radius.

supports more clearly the possibility of a brown-dwarf companion and even the possibility of a binary brown dwarf.

Constrained by	M_1 in M_{Jupiter}	M_2 in M_{Jupiter}	D_{D} in kpc	D_{S} in kpc	D_{DS} in kpc
t_{E}, q	131^{+272}_{-88}	319^{+657}_{-215}	6.8 ± 1.7	9.3 ± 1.1	2.5 ± 1.8
Constrained by	M_1 in M_{Jupiter}	M_2 in M_{Jupiter}	D_{D} in kpc	D_{S} in kpc	D_{DS} in kpc
t_{E}, q, $r_\star$	49^{+86}_{-19}	$119^{+208}_{-45.0}$	$5.9^{+2.0}_{-1.8}$	$6.2^{+1.8}_{-2.1}$	$0.18^{+0.22}_{-0.12}$

Table 6.3: Physical parameter estimates of mass and distances for the given event.

6.4.5 Alternative explanations

The short duration and the symmetry of the event make parallax effects, such as the first event discovered by Alcock et al. (1995), unlikely. This can be supported by a heuristic estimation of the maximal achievable effect, as illustrated in Fig. 6.15. The sum of angles provides a relation if lens and observers as well as source and observers form isosceles triangles, implying that the offset of the source position is

$$\Delta\beta_{\text{max}} = \pi - 2\arctan\left(\frac{B}{2D_S}\right) - 2\arctan\left(\frac{2D_{\text{D}}}{B}\right), \tag{6.4}$$

where B denotes the baseline between two observing sites or their orbital motion. As the arguments of the arc tangents are roughly of inverse order, replacing them by approximations is not advisable. For the given configuration, this implies that the resulting scale for a baseline of $8500\,\text{km}$ is $\Delta\beta_{\text{max}} \approx 5 \cdot 10^{-6}\,\theta_{\text{E}}$, which is clearly below the finite size of the source star of $\approx 10^{-2}\,\theta_{\text{E}}$ and thus no multi-site parallax is available. For the given Einstein time, the corresponding Earth's orbital motion is large enough to create an offset of $\Delta\beta_{\text{max}} \approx 10^{-2}\,\theta_{\text{E}}$ and thus
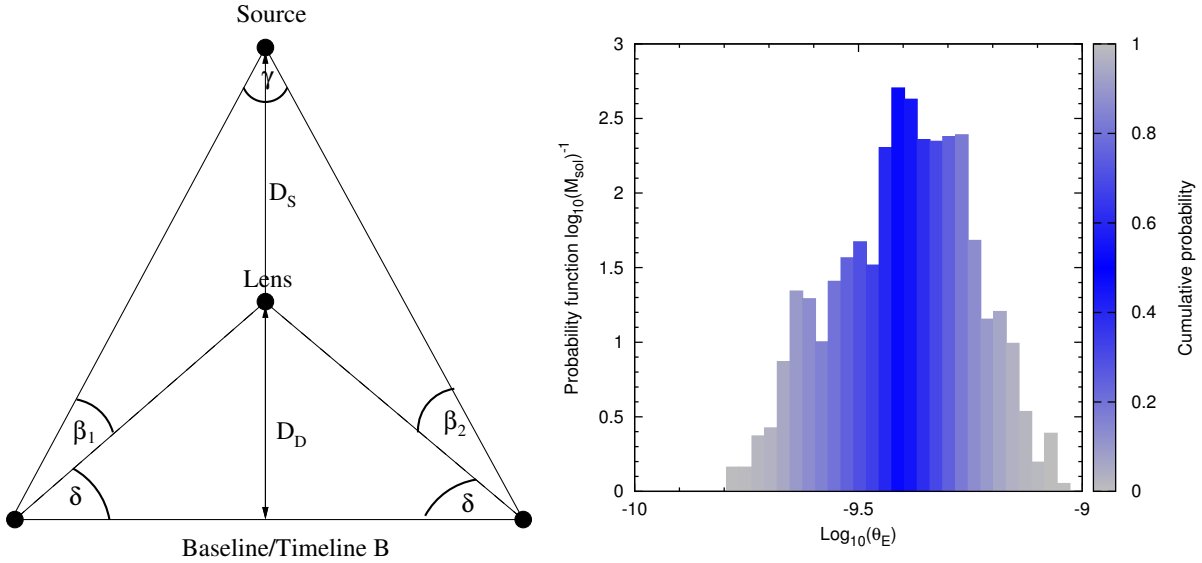
Figure 6.15: Geometrical, maximal change in the effective source position for $\beta_1 = \beta_2$ (left) shown along with the distribution of Einstein radii (right) for the interpretation of the data set including $t_E, r_\star$.

will have a negligible impact on the parameter estimates. In addition, the offset is not able to shift the source position to the caustic as shown in Fig. 6.16. The asymmetric light curve also excludes finite source point lens models. Finally, the effect of a binary source needs to be considered, which cannot directly by excluded, although the light curve shape and especially the even brightness between both maxima of magnification a priori supports a binary lens. The resulting light curve is a superposition of two PSPL light curves as introduced by Griest & Hu (1992). Assuming the separation of both companions is static leads to the parameterization introduced by Gaudi (1998)

$$F(t) = F_1(t)\mu_{\text{PSPL}}((u_1(t)) + F_2(t)\,\mu_{\text{PSPL}}(u_2(t)) + F_B(t) \tag{6.5}$$

For modeling the given event, the tracks are then parameterized with one source position centered on one of the potential source stars

$$u_1(t) = \left(u_0^2 + \frac{t - t_0}{t_E}\right)^{1/2}. \tag{6.6}$$

and a second source star displaced by $(\Delta u_0, \Delta u_t)$:

$$u_2(t) = (u_0 + \Delta u_0)^2 + \left(\frac{t - t_0}{t_E} + \Delta u_t\right)^{1/2}. \tag{6.7}$$

In the fitting process different baseline fluxes are represented by their flux ratio

$$f_r = \frac{F_2}{F_1}, \tag{6.8}$$

as introduced in the superposition of both magnifications in Eq. 6.5. For fitting such a model, a standard simplex algorithm was used and the reported uncertainties were estimated by applying
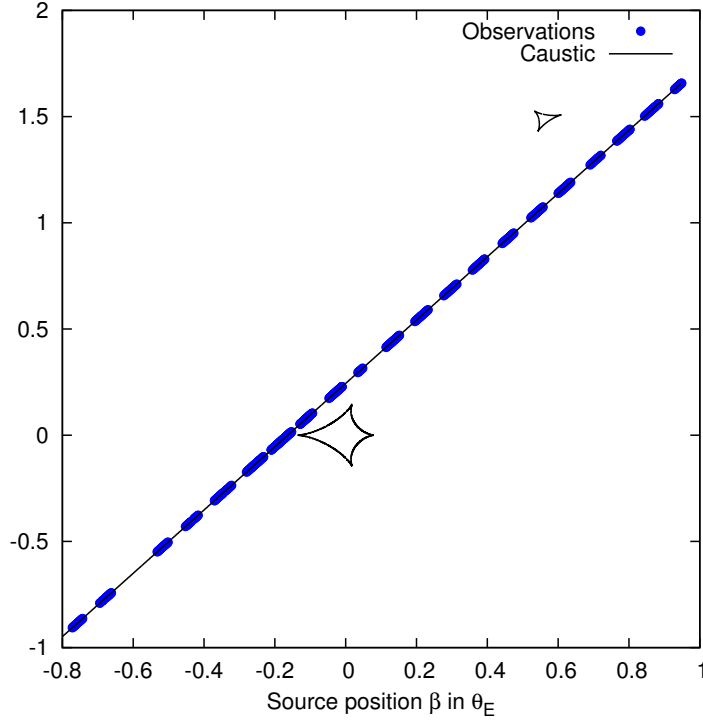
Figure 6.16: Source star track with respect to the caustic structure of the best fit.

the bootstrap technique, resampling distributions by drawing samples with replacement from the underlying data (Efron, 1979; Press, 1994). The resulting parameter configuration is shown in Tab. 6.4 and the corresponding best fit in Fig. 6.17. The parameter space does not provide a clear indication as the parameters t_E, u_0 are degenerate, which is a consequence of Eq. 6.2 and directs the fit to extremely large Einstein times, which is unlikely regarding the initial estimates which are consistent with the binary lens fit.

χ^2	u_0	t_0 in d	f_r
4064.54	$5.182^{+0.0849}_{-0.729} \cdot 10^{-5}$	$2365.284^{+41.9}_{-77.5}$	$1.0^{7.4\cdot10^{-6}}_{-7.4\cdot10^{-6}}$

t_E in d	Δu_0	Δu_t
$5388.6249^{+0.00042}_{-0.0060}$	$0.0001160^{+0.000025}_{-0.0000049}$	$0.0008693^{+0.0000083}_{-0.000073}$

Table 6.4: Best fit parameters for a binary point source model.

6.5 Conclusions and discussion

In the course of this chapter, the data reduction system implemented for the telescopes used by the MiNDSTEp consortium was introduced and applied to the anomalous microlensing event MOA 2010-BLG-406. The reduction was compared with the results of the numerical kernel approach.
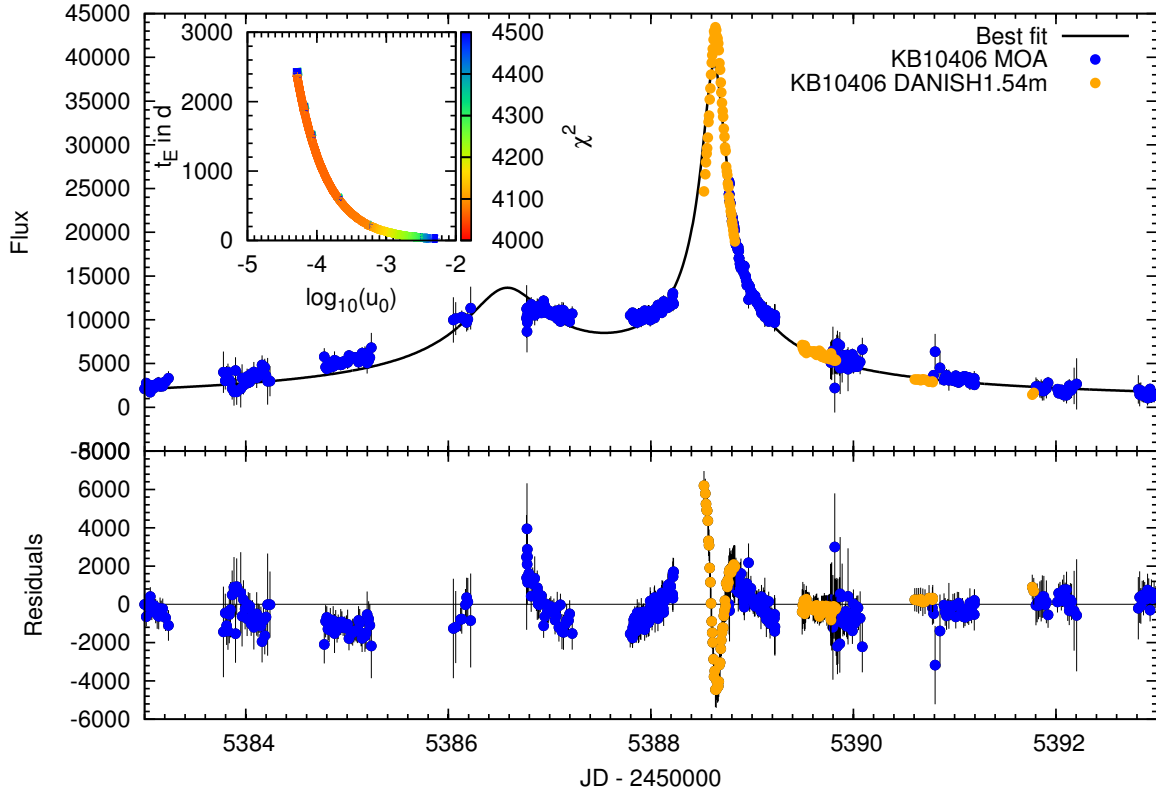
Figure 6.17: The best fit binary source light curve simulated with the GPU-contouring model is shown for MOA 2010-BLG-406 along with the corresponding residuals and the χ^2 in the course of the simplex fit.

Both systems were providing consistent results, but due to the reduced systematics the numerical kernel approach was used for an assessment of the physical nature of the event.

The GPU-model developed in the context of this work was applied to a MHMCMC scheme for fitting the corresponding light curve. The multimodal structure of the parameter estimates obtained from the fit indicate either systematic effects in the reduction or limitations caused by the limited parameter accuracy of the single-precision GPU-model. The resulting one-dimensional marginalized distributions were combined with a catalog simulation of the Besançon model. The most likely source and lens positions are not able to reproduce the fitted Einstein-time and source star radius consistently. By letting the source and lens position vary independent of the lensing probability, it turns out that pairs of close lens and source stars provide the required configuration. According to this analysis, one of the source stars is likely to be a brown dwarf, while the second component can be a brown dwarf or a low-mass star.

These results here, were achieved for a given Galactic model: as differing initial mass functions can alter the mass estimates, a more detailed study including recent brown dwarf populations is inevitable. Moreover, the system serves as a first test of the GPU-assisted contouring approach, though and a double precision version may lead to different parameter estimates as these are not subject to the same numerical inaccuracies. The sharp modes of the given fit illustrate how difficult the parameter space can look. Degeneracies set their footprint on the analysis and if the global minimum for the event is located in a sharp distinct feature of the total likelihood, there

is a chance, that the number of Markov chains miss such a solution.

Despite all these complications, the results of this analysis encourage a more detailed study of this event, as it indicates the possibility of measuring a rare brown-dwarf binary at several kpc with a source star a few hundred light-years away.

Chapter 7

Future observational prospects and summary

In this last chapter, I provide esimations for the detectability of planets using the MONET and space-based telescopes. The latter have the principle capability of contributing observations with large baselines. Gould (1995) and Graff & Gould (2002) showed that observations from space can provide a satellite parallax measurement if they are sufficiently far away from the Earth. Due to the altered optical axis, the new source position with respect to the lens changes the magnification and thus constraints the distance to lens and source. In this study, the feasibility of using the existing STEREO mission and the planned M-class mission PLATO are discussed. Both missions are not dedicated for carrying out microlensing observations, but use auxiliary telescopes capable of observing the Galactic bulge.

7.1   MONET/North observation of MOA 2010-BLG-477

So far, the best covered microlensing event observed with MONET/North is MOA 2010-BLG-477, which is also the first observation which was carried out semi-automatically by means of an automated system called Claude (Husser 2011), which processes target recommendations from ARTEMiS. The automatic data reduction pipeline based on the DIAPL software presented by Wozniak (2000) was run locally in Göttingen. As this event was highly prioritized, it was also followed by the DANISH 1.54m telescope, offering the possibility of directly comparing the two data sets which were taken with similar cadence and longitude. By aligning the data sets, estimates of the capabilities of both MONET telescopes can be given which optimize current recommendations and increase the chance for further detections.

The expected signal-to noise ratio of CCD-detectors (Howell 2000) can be approximated by

$$\frac{S}{N} \approx \sqrt{N_{\text{electrons}}}, \tag{7.1}$$

for photon noise dominated observation. This is a consequence of the standard deviation of a Poisson distribution expressed for the number of electrons - the countable quantity. In the linear regime of the detector, where the number of counted electrons is proportional to the number of photons, the signal-to-noise is $\frac{S}{N} \propto \sqrt{t_{\text{exp}}}$ assuming that the exposure is not affected by time-dependent perturbations, such as clouds. The ARTEMiS system estimates the exposure time according to the initially reported magnitudes. The reported baseline magnitude is used for estimating the exposure time, without assessing its true reliability. In practice, the estimate of

the baseline magnitude is subject to blending and on a thorough calibration of the instrumental flux. This implies that the magnitude cannot alway be properly determined, for instance, due to a time dependent baseline flux or a bright blend star which, in turn, changes the noise level. If photon noise and the blend ratio $q \geq 1$ are limiting the data quality, the signal-to-noise of the observable source star variation is

$$\frac{F_{S}\mu}{\sqrt{F_{S}\mu + F_{B}}} \approx g^{-1/2}\sqrt{F_{S}}\mu, \tag{7.2}$$

and thus reduced by a factor of $g^{-1/2}$. For a potential blend ratio of 10, this leads to a decrease in the signal-to-noise by a factor of 3, requiring to increase the exposure time by a factor of 10. In these cases a systematical error in the determination of F_{B} may easily affect the proposed exposure times. One of these systematic effects is the determination of the reference flux. Difference imaging pipelines provide a difference flux with respect to the chosen reference frame. For estimating the absolute flux and magnitude m, it is necessary to measure the brightness as reference flux F_{R} on the reference frame, yielding the magnitude m

$$m = m_0 - 2.5\log(F_{R} + \Delta F), \tag{7.3}$$

where m_0 is an offset, which can be calibrated to provide absolute magnitudes. Overlapping PSFs, for instance, can lead to bad estimates of the reference flux and hence to the reported blend ratio. The blending flux is not considered for determining the exposure time for MONET/North, as most events stay between blend ratios of $g \in [0,1]$ (Smith et al. 2007) and thus the worst case leads to a signal-to-noise reduction by $2^{-1/2}$ for magnifications $\mu \approx 1$.

The exposure time estimates for MONET are integrated in the automation tool *Claude* by Husser (2011), which receives the list of recommended targets from ARTEMiS, calculates the required exposure times for MONET, and submits the corresponding results back. Moving the telescope and starting the exposures is done automatically via *Claude*. Due to technical constraints, exposure times are kept between 20 and 300 s, providing a sufficient number of stars for the subtraction and ensuring that the telescope tracking does not lead to elongated PSFs.

According to the uncertain blending estimates leads to proposed events with exposure times between 300 and 600 s. The exposure time estimation itself relies on an inversion of the signal-to-noise calculation (Howell 1989) including the MONET telescope parameters, such as a sky brightness of $m_{I} \approx 20\,\mathrm{mag/arcsec^2}$ and a fixed extinction coefficient of 0.08 per airmass. Test observations have indicated, that the initially required $S/N \approx 100$ is not sufficient for the given parameter estimations and thus $S/N \approx 500$ was demanded. The reasons for this offset are the age of the camera, a decreasing mirror reflectivity, but also the assumption of applying an aperture photometry with a given aperture radius of 1 arcsec. Consequently, either empirical corrections are needed or a quality assessment of the reported light curve fit needs to be implemented.

In Fig. 7.1 the preliminary results of the GPU-assisted grid search for the event MOA 2010-BLG-477 are shown. For this purpose, observations from the DANISH 1.54m and the MONET/North telescope are used. Data sets from both sites were re-reduced using the same difference imaging package (DIAPL). Obviously, no convergence was achieved as the light curve is dominated by gaps. The corresponding χ^2_{red} stays around $\chi^2_{\mathrm{red}} \approx 8$, which is partly caused by the fit, but also by the reduction indicating a limited quality of the PSF fit (cf. Fig.6.4). As a consequence, it is likely that the reported errorbars are underestimated. The initial χ^2-surface does not reflect the final result, but in contrast to the event discussed in the preceding chapter, a potential lower mass ratio cannot be excluded. Directly comparing MONET/North and DANISH 1.54m observations illustrates that the outlined configuration was working well, considering the geographical location

of the telescope and observations through 2 airmasses. The exposure time chosen for the baseline was at its maximum and thus events fainter than $17\,m_I$ are hardly accesible for MONET/North.
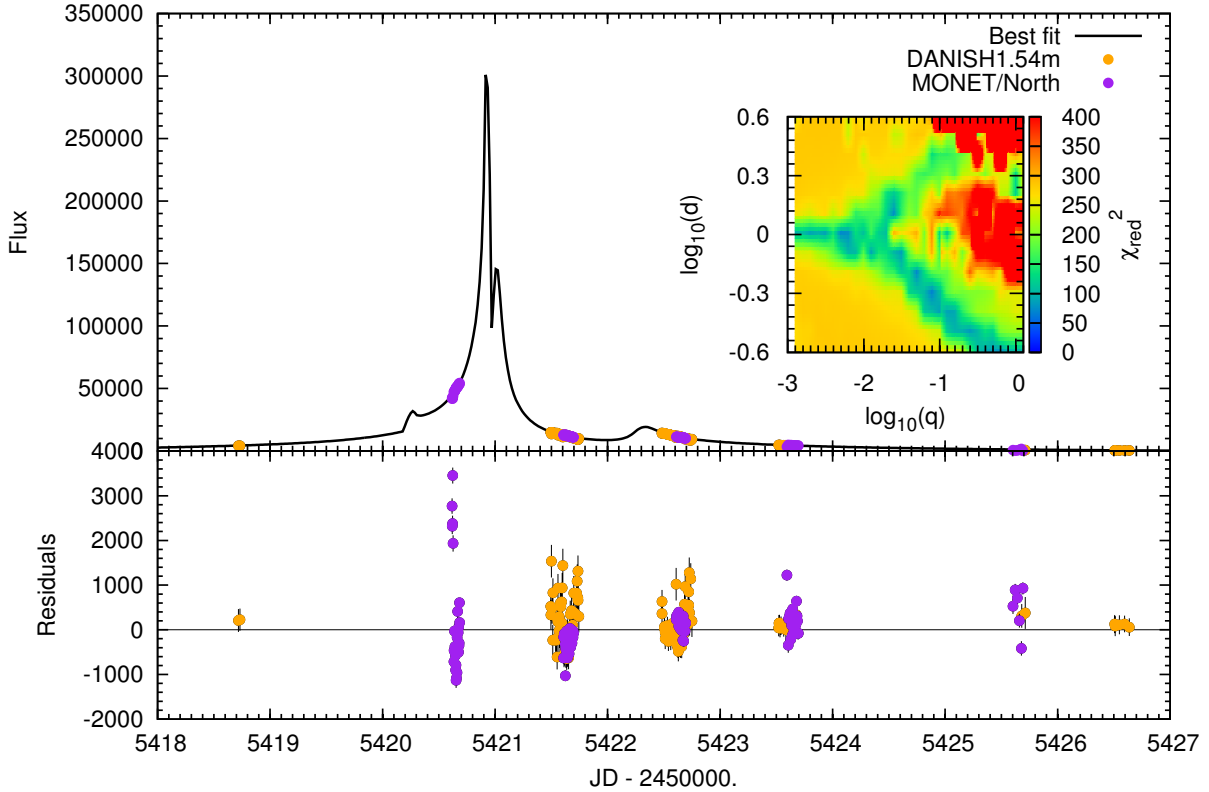

Figure 7.1: Preliminary binary lens fit for MOA 2010-BLG-477.

7.2   Space-born microlensing

The benefits of observing microlensing events from space are twofold: on the photometric side, highly accurate brightness measurements can be achieved as long as the spatial resolution is good enough to separate lens and source stars. On the geometric side, long baselines can constrain the distances to lens and source, which is important for constraining the mass of a potential planetary system. The first observation of that kind was carried out by Dong et al. (2007). In the following, prospects for the STEREO mission (Driesman et al., 2008) and its Heliospheric Imager 1 (HI-1), as part of the SECCHI instrument suite (Defise et al., 2001; Howard et al., 2002; Eyles et al., 2009), and the PLATO mission (Catala & the PLATO consortium, 2008; Catala, 2009) are compared with respect to observing microlensing events.

7.2.1   Distance constraints

The geometric effect is limited by the baseline of the corresponding space observatory. An initial assessment of the capabilities concerning a satellite parallax can be given according to Eq. 6.4. The STEREO, HI cameras are moving ahead and behind the Earth's orbit, changing their position with time but also staring at the space between the Sun and the Earth and thus the baseline can even

reach $\approx 1\,\mathrm{AU}$. The proposed PLATO mission is supposed to be located around the Lagrangian point L2, in a large amplitude libration orbit leading to a baseline of about 1.5 million kilometers. For a $0.5\,M_\odot$ lens at $D_\mathrm{D} = 6\,\mathrm{kpc}$ and a source at $D_\mathrm{D} = 8\,\mathrm{kpc}$ the effective change in source position is $\Delta\beta_{PLATO} \approx 10^{-3}\,\theta_\mathrm{E}$ making the observations relevant for smaller source star radii, such as main-sequence stars observed in high magnification events. For the STEREO mission the effect is more prominent: $\Delta\beta_{STEREO} \approx 0.1\,\theta_\mathrm{E}$ for $B = 1\,\mathrm{A}$. For a microlensing event with $t_\mathrm{E} = 10\,\mathrm{d}$ this would correspond to an offset of 2 days.

7.2.2 Photometric limitations

Microlensing observations are limited by detections in crowded fields. Auxiliary telescopes on space missions are often equipped with small aperture telescopes covering large fields of view. The corresponding pixel scale, usually given in arcseconds per pixel, is large for these telescopes leading to a higher blending flux contribution if a microlensing event within such a pixel is observed. This kind of pixel-lensing was originally motivated for surveys observing M31 (Crotts 1992; Baillon et al. 1993), but can be applied to microlensing observations of the Galactic bulge as well. The major disadvantage of this technique is the increased blending flux contributing to the background noise and reducing the signal-to-noise.

Assuming that microlensing events are detectable if a limiting magnitude is exceeded, leads to estimates for the expected number of events for a given brightness interval. A magnitude limit of $m_\mathrm{I} < 19$ serves as an initial selection criterion. As mentioned before, the number of observable source stars exceeds the number of stars below the magnitude limit, because the magnification can increase the brightness to make these sources observable. A simulation of the corresponding stellar statistics for the field $l = 1.2°$, $b = -2.7°$ as presented in chapter 4 is shown in Fig. 7.2 constrained by limiting magnitude, magnification and underlying Galactic model.
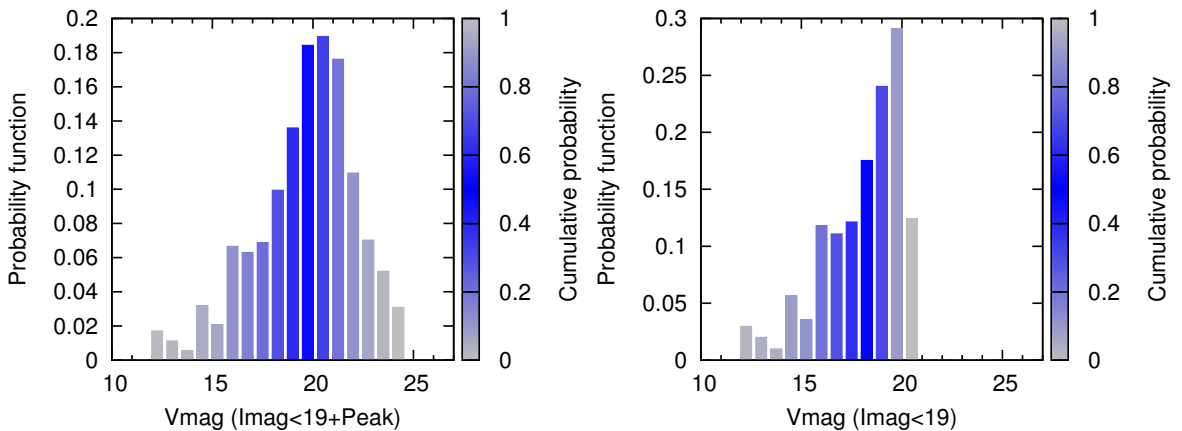
Figure 7.2: Expected visual magnitudes for a catalog simulation centered on $l = 1.2°$, $b = -2.7°$ with and without taking magnification into account.

For microlensing observations, a signal-to-noise ratio of $S/N \approx 100$ is required for detecting planetary deviations. For simulating the effective signal-to-noise per pixel, the frequency of magnitudes in the catalog simulation is used and scaled according to the field of view, given in

square degrees. The noise contribution of all stars can be assessed for both space missions. In the case of STEREO, the photon noise for HI-1 depending on different magnitudes is taken from the electron statistics presented by Howard et al. (2002) and for PLATO the photon noise estimates are relying on the expected number of electrons $N_e = 1.5 \cdot 10^8 h^{-1}$ (Catala 2009), where the noise contribution of fainter stars is extrapolated, assuming the number of photons is proportional to the flux and dominated by photon noise. The source star magnitude serves is defined to be the corresponding limiting magnitude for each mission. For PLATO, a magnitude limit of $14\,\mathrm{m_V}$ (Catala & the PLATO consortium 2008) is assumed for the PLATO mission and for STEREO $12\,\mathrm{m_V}$ (Eyles et al. 2009). Fig. 7.2 also provides an estimate of the fraction of observable stars for both missions. The magnitude limit for STEREO is $< 1.3\,\%$ and given that the Galactic bulge is visible only for 2-4 weeks, the total number of accesible microlensing events stays around one. Estimates for PLATO are more promising, because $2\,\%$ of the $3 \cdot 10^6$ stars comply with $14\,\mathrm{m_V}$ and assuming a staring phase, covering the Galactic Bulge for half a year. PLATO can contribute data to at least 20 events, neglecting caustic crossing events with an intrinsic high magnification, which are not included by recent estimates.
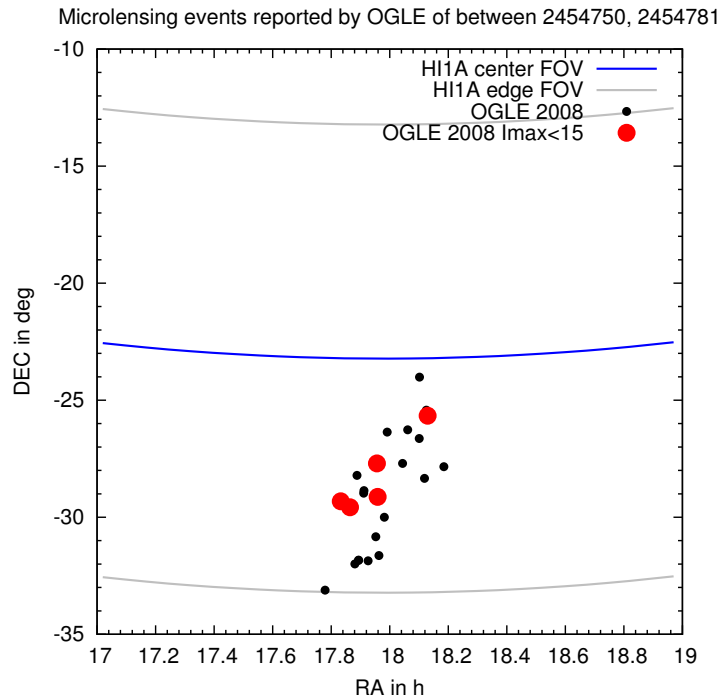
Figure 7.3: OGLE events observed by STEREO HI1A in 2008.

More accurate estimates can be given if the signal-to-noise is known. The resulting estimates, shown in Tab. 7.1, are based on the assumption that the source star flux is predominantly stored in one pixel and thus extraction techniques for pixel-lensing (Gould, 1996) need to be applied. The STEREO mission is only helpful if the unbinned HI1 frames are used and a microlensing event is magnified below $11\,\mathrm{m_V}$. Even in this case, systematical effects have to be considered, such as the contamination by the Frauenhofer corona (F-corona). The latter effect can be roughly estimated by considering the underlying count rate of each frame, implying that for most of the time (12 d) the contribution of the F-corona is below $\approx 2 \cdot 10^4$ photo electrons per pixel. In this scenario, the signal-to-noise is reduced by a factor of 2 and thus it is more advisable to consider

Mission/instrument	PLATO	STEREO HI1 (unbinned)	STEREO HI1 (2×2)
FOV in arcsec	12.4	35	70
exposure time in s	20.0	40	40
limiting magnitude	$14\,\mathrm{m_V}$	$12\,\mathrm{m_V}$	$12\,\mathrm{m_V}$
S/N per px	160	20	10

Table 7.1: Estimated signal-to-noise for the STEREO and PLATO mission based on photon noise. The signal-to-noise is provided for the limiting magnitude.

events below $10\,\mathrm{m_V}$. If such an event occurs, it needs to be checked if it was in the field-of-view of HI1-A or HI2-B. Fig. 7.3 illustrates where microlensing events were located in the 2008 season, after applying the pointing calibrations by Brown et al. (2009).

The signal-to-noise estimates for PLATO support the feasibility of using it for microlensing observations. Nevertheless, the assumption of considering one pixel can be violated, as the PSF can be distorted at the edge of the detector, where the Galactic Bulge is expected to be observable (Zima et al. 2010). If these kind of systematics affect the quality of the corresponding light curve, it is still possible to keep the signal-to-noise by stacking or drizzling potentially dithered frames (Fruchter & Hook 2002). In addition, the variability of all stars in a given pixel was neglected, contaminating the observation if the observed source has same brightness compared to the variable contaminant. Fig. 7.4 summarizes again the prospects including different observing times for the PLATO mission as well as differently binned frames for HI-1 with and without additional F-corona noise.
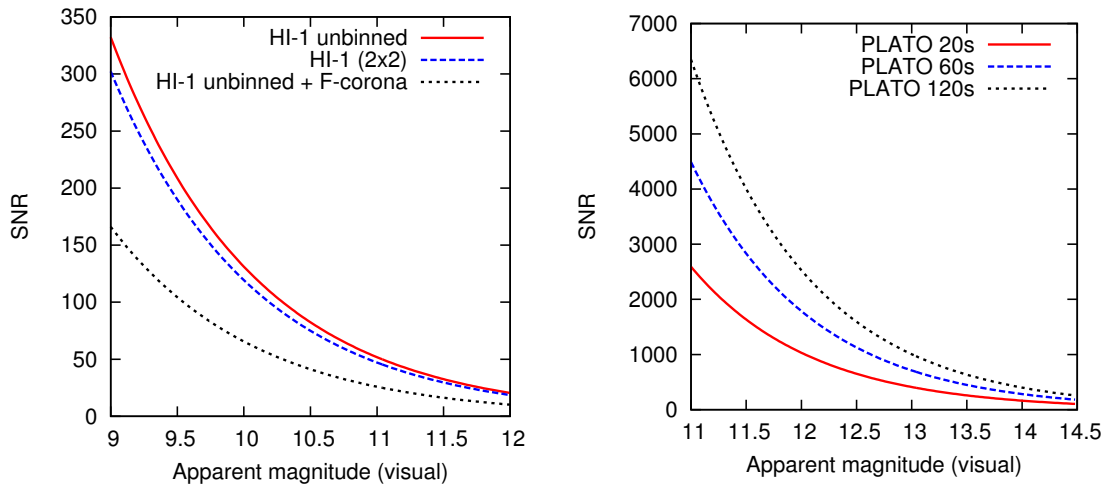
Figure 7.4: Expected signal-to-noise for different apparent magnitudes for the STEREO (left) and the PLATO mission (right), depending on different exposure times for PLATO and illustrating the change by using binned HI-1 observations with and without F-corona contribution.

7.3 Planetary detections

The possible number of detectable microlensing planets by the PLATO mission can be estimated for different magnification-scenarios, i.e. especially for high magnification events with $\mu > 100$ and low magnification events. In chapter 4 the sensitivity of detecting planets was assessed by fitting PSPL models to light curves and calculate the $\Delta\chi^2$. The following estimates are made by integrating the contours of a predefined relative deviation between binary magnification map and a PSPL map (Chung et al. 2005). The magnification maps were divided by the PSPL model centered on the host star, as illustrated in Fig. 7.5, and its detectable area was integrated and related to the area of a given maximal PSPL magnification. One Einstein radius was used as reference for low magnification events and thus planetary caustics were only considered within that radius
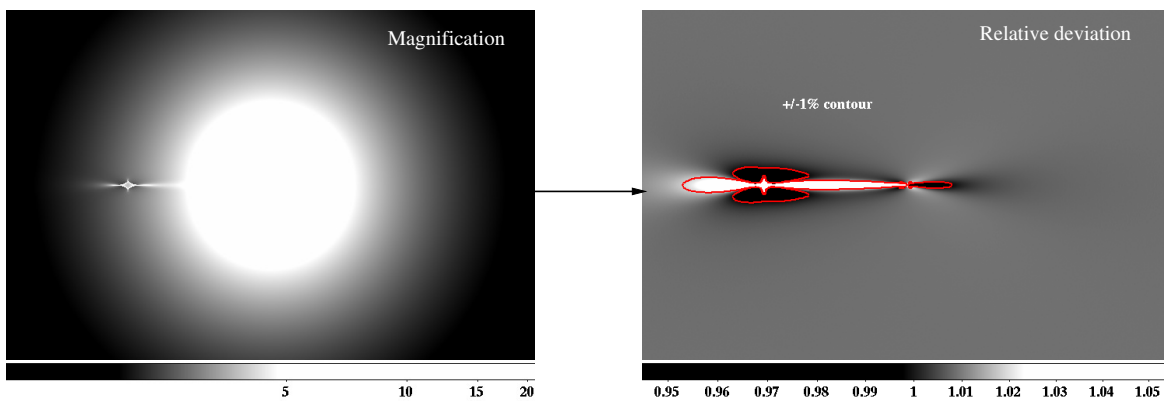
Figure 7.5: Example for the calculation of relative deviation from PSPL model and 1 % contours for $q = 1.39 \cdot 10^{-4}$, $d = 1.20$.

Detecting planets in high magnification events is more likely, because the deviation caused by the central caustic coincides with the central region, sampled by the source star (Griest & Safizadeh, 1998). For PLATO, a high magnification approach is the natural strategy, as it demands bright events. This constraint can be fulfilled either by a bright source star, with a potentially larger source star radius, which smoothes out the planetary signal, or by a highly magnified event. In order to estimate the chances of detecting Super-Earths and the prospected number of detections, a few assumptions need to be made. The number of planet-detections can be estimated using the number of microlensing events N_{events}, the frequency of planets in the Galaxy p_{planets} and the detection sensitivity for each event $p_{\mathrm{sensitivity}}$:

$$N_{\mathrm{planets}} = N_{\mathrm{events}} \cdot p_{\mathrm{planets}} \cdot p_{\mathrm{sensitivity}}. \tag{7.4}$$

The number of microlensing events per year is well known and for a given magnitude limit, one can estimate the number of detections. Despite efforts to improve the number of microlensing events, is is assumed that ≈ 1000 events can be detected per year. The sensitivity is calculated according to the area of the contour of the 1 % deviation and multiplied by the chance of having a planet in the detectable zone of microlensing events. Estimates for the frequency of planets can be provided by using all planets so far discovered or by using planetary evolution simulations. The latter take into account a variety of potential parameters, as shown by Ida & Lin (2005), for instance. Another example for such a simulation was presented by Mordasini et al. (2009) and is used here to estimate the chance of having a planet in the microlensing regime. For the

distributions of the semi-major axis a uniform distribution of planets ranging from 0 to 10 AU is assumed. For Super-Earths in that regime, a chance of 40 % can be assumed.
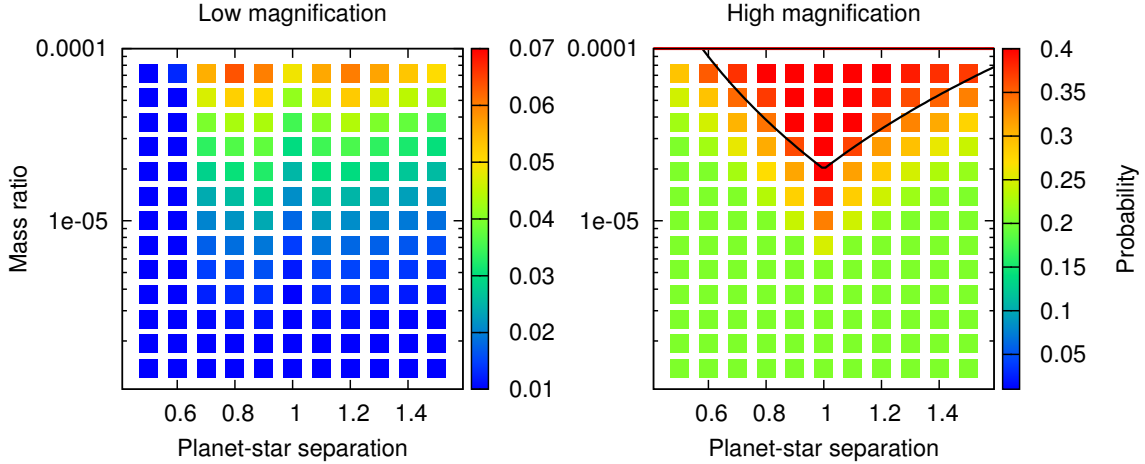

Figure 7.6: Planet detection probability for Super-Earths based on estimates from planet-formation calculations given for low magnification events (left) and high magnification events (right). The black line indicates the triangle of certain detection assuming a magnification of $\mu \approx 1000$ (Gould et al., 2010).

Combining these results with the sensitivity from the contour sensitivity, leads to a planet detection probability as shown in Fig. 7.6. For low magnification events we obtain an averaged probability of 1.3 % for detecting Super-Earths and for high magnification events a chance of 8.8 %. The number of events accesible for the PLATO mission is limited to ≈ 20 events per year, demanding a magnitude limit of 14 m_V. For a $S/N \approx 100$ the corresponding limiting magnitude is 14.5 m_V with ≈ 40 events per year. This would lead to a 0.4 Super-Earth detections per year and its feasibility depends on the final specification as well as the number of detected events in the year 2018 and beyond.

7.4   Summary

To conclude this thesis, a short review of its content shall be given. After reflecting the derivation of the deflection angle and the lens equation, the observable quantities and analytical limitations for simulating light curves were discussed. For assessing higher order contributions, a numerical perturbative approach was introduced for studying the detectability of deviations in light curves. In order to address the computationally demanding simulation of light curves, a GPU-assisted model was developed and its numerical limitations were studied and compared with CPU-based ray-tracing simulations.

Before applying this technique to real observations, the benefits of applying information theoretical properties were outlined, especially exploiting Shannon's and Fisher's definition of information. This has lead to the insight how sampling strategies can be optimized and that all observations could be used as one large data set. Furthermore, information theory leads directly to compression techniques, which was illustrated by implementing the compression technique introduced by Heavens et al. (2000). The benefits of using an information driven selection in the course of an MCMC fitting were introduced.

For the microlensing event MOA 2010-BLG-406 the Besançon model was combined with the GPU-model and it was estimated that the underlying event is a binary system, where at least one component is a brown dwarf. In addition, lens and source only comply with a rather small lens-source separation encouraging further studies of this event. Finally, the future prospects for using space-born telescopes, designed for different purposes are addressed for the ongoing STEREO mission and the planned PLATO satellite. The latter can contribute to planetary detections via microlensing, but only with the detection of ≈ 1 Super-Earth for 2 years of bulge observations.

There are a number of relevant issues for further studies. Higher order corrections and different metric theories of gravitation should be assessed to exploit observations of higher photometric precision. The GPU-technology can be applied to various stages of the analysis, e.g. to the image subtraction, light curve simulation, fitting process and to Galactic modeling. The results from information theory, especially those of optimal experimental design, can be used in a similar way to add extra worth to each observation. It also seems promising to define the detectability of planets as an optimization goal and to study the information theoretical implications. In the context of real-time binary modeling, the compression techniques can improve the applicability of these methods. Concerning the limitations of the Galactic model, two improvements need to be considered: the spiral arm structure of the Galaxy is not included, which may change some of the estimated planetary masses significantly; and for brown dwarf binaries and their host stars, suitable distributions are required for detections beyond the solar neighborhood.

Bibliography

ABE, F. (2010) Gravitational Microlensing by the Ellis Wormhole. *ApJ*, **725**, 787–793.

ALARD, C. (2000) Image subtraction using a space-varying kernel. *A&AS*, **144**, 363–370.

ALARD, C. & LUPTON, R. H. (1998) A Method for Optimal Image Subtraction. *ApJ*, **503**, 325–331.

ALBROW, M. D., HORNE, K., BRAMICH, D. M., FOUQUÉ, P., MILLER, V. R., BEAULIEU, J., COUTURES, C., MENZIES, J. ET AL. (2009) Difference imaging photometry of blended gravitational microlensing events with a numerical kernel. *MNRAS*, **397**, 2099–2105.

ALCOCK, C., ALLSMAN, R. A., ALVES, D., AXELROD, T. S., BENNETT, D. P., COOK, K. H., FREEMAN, K. C., GRIEST, K. ET AL. (1995) First Observation of Parallax in a Gravitational Microlensing Event. *ApJ*, **454**, L125–L128.

ATKINSON, A. (1992) *Optimum Experimental Designs*. Clarendon Press, Oxford.

BAILLON, P., BOUQUET, A., GIRAUD-HERAUD, Y. & KAPLAN, J. (1993) Detection of Brown Dwarfs by the Microlensing of Unresolved Stars. *A&A*, **277**, 1–9.

BATE, N. F., FLUKE, C. J., BARSDELL, B. R., GARSDEN, H. & LEWIS, G. F. (2010) Computational advances in gravitational microlensing: A comparison of CPU, GPU, and parallel, large data codes. *New A*, **15**, 726–734.

BEAULIEU, J.-P., BENNETT, D. P., FOUQUÉ, P., WILLIAMS, A., DOMINIK, M., JORGENSEN, U. G., KUBAS, D., CASSAN, A. ET AL. (2006) Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature*, **439**, 437–440.

BEKENSTEIN, J. D. (2004) Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D*, **70**(8), 083509–1–083509–28.

BENNETT, D. P. (2010) An Efficient Method for Modeling High-magnification Planetary Microlensing Events. *ApJ*, **716**, 1408–1422.

BENNETT, D. P., BOND, I. A., UDALSKI, A., SUMI, T., ABE, F., FUKUI, A., FURUSAWA, K., HEARNSHAW, J. B. ET AL. (2008) A Low-Mass Planet with a Possible Sub-Stellar-Mass Host in Microlensing Event MOA-2007-BLG-192. *ApJ*, **684**, 663–683.

BENNETT, D. P. & RHIE, S. H. (1996) Detecting Earth-Mass Planets with Gravitational Microlensing. *ApJ*, **472**, 660–664.

Bond, I. A., Abe, F., Dodd, R. J., Hearnshaw, J. B., Honda, M., Jugaku, J., Kilmartin, P. M., Marles, A. et al. (2001) Real-time difference imaging analysis of MOA Galactic bulge observations during 2000. *MNRAS*, **327**, 868–880.

Bozza, V. (2010) Microlensing with an advanced contour integration algorithm: Green's theorem to third order, error control, optimal sampling and limb darkening. *MNRAS*, **408**, 2188–2200.

Bramich, D. M. (2008) A new algorithm for difference image analysis. *MNRAS*, **386**, L77–L81.

Brown, D. S., Bewsher, D. & Eyles, C. J. (2009) Calibrating the Pointing and Optical Parameters of the STEREO Heliospheric Imagers. *Sol. Phys.*, **254**, 185–225.

Brown, G. E. & Bethe, H. A. (1994) A Scenario for a Large Number of Low-Mass Black Holes in the Galaxy. *ApJ*, **423**, 659–664.

Caldwell, A., Kollár, D. & Kröninger, K. (2009) BAT - The Bayesian analysis toolkit. *Computer Physics Communications*, **180**, 2197–2209.

Capozziello, S., de Ritis, R., Man'ko, V. I., Marino, A. A. & Marmo, G. (1997) Defocusing gravitational microlensing. *Phys. Scr*, **56**, 212–220.

Catala, C. (2009) PLATO: PLAnetary Transits and Oscillations of stars. *Communications in Asteroseismology*, **158**, 330–336.

Catala, C. & the PLATO consortium (2008) PLATO: PLAnetary Transits and Oscillations of stars. *Journal of Physics: Conference Series*, **118**(1), 012040, URL `http://stacks.iop.org/1742-6596/118/i=1/a=012040`.

Chang, K. & Refsdal, S. (1979) Flux variations of QSO 0957+561 A, B and image splitting by stars near the light path. *Nature*, **282**, 561–564.

Chiu, M., Ko, C. & Tian, Y. (2006) Theoretical Aspects of Gravitational Lensing in TeVeS. *ApJ*, **636**, 565–574.

Chung, S.-J., Han, C., Park, B.-G., Kim, D., Kang, S., Ryu, Y.-H., Kim, K. M., Jeon, Y.-B. et al. (2005) Properties of Central Caustics in Planetary Microlensing. *ApJ*, **630**, 535–542.

Close, L. M., Lenzen, R., Guirado, J. C., Nielsen, E. L., Mamajek, E. E., Brandner, W., Hartung, M., Lidman, C. et al. (2005) A dynamical calibration of the mass-luminosity relation at very low stellar masses and young ages. *Nature*, **433**, 286–289.

Cover, T. (2006) *Elements of Information Theory*. Wiley-Interscience, New York.

Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press.

Crotts, A. P. S. (1992) M31 - A unique laboratory for gravitational microlensing. *ApJ*, **399**, L43–L46.

Defise, J.-M., Halain, J.-P., Mazy, E., Rochus, P. P., Howard, R. A., Moses, J. D., Socker, D. G., Simnett, G. M. et al. (2001) Design of the Heliospheric Imager for the STEREO mission. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, edited by O. H. Siegmund, S. Fineschi, & M. A. Gummin, vol. 4498 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*.

DI STEFANO, R. & ESIN, A. A. (1995) Blending of Light in Gravitational Microlensing Events. *ApJ*, **448**, L1–L4.

DOMINIK, M. (1999) Ambiguities in FITS of observed binary lens galactic microlensing events. *A&A*, **341**, 943–953.

DOMINIK, M. (2007) Adaptive contouring - an efficient way to calculate microlensing light curves of extended sources. *MNRAS*, **377**, 1679–1688.

DOMINIK, M. (2009) Parameter degeneracies and (un)predictability of gravitational microlensing events. *MNRAS*, **393**, 816–821.

DOMINIK, M., HORNE, K., ALLAN, A., RATTENBURY, N. J., TSAPRAS, Y., SNODGRASS, C., BODE, M. F., BURGDORF, M. J. ET AL. (2008) ARTEMiS (Automated Robotic Terrestrial Exoplanet Microlensing Search): A possible expert-system based cooperative effort to hunt for planets of Earth mass and below. *Astronomische Nachrichten*, **329**, 248–251.

DOMINIK, M., JØRGENSEN, U. G., RATTENBURY, N. J., MATHIASEN, M., HINSE, T. C., CALCHI NOVATI, S., HARPSØE, K., BOZZA, V. ET AL. (2010) Realisation of a fully-deterministic microlensing observing strategy for inferring planet populations. *Astronomische Nachrichten*, **331**, 671–691.

DOMINIK, M., RATTENBURY, N. J., ALLAN, A., MAO, S., BRAMICH, D. M., BURGDORF, M. J., KERINS, E., TSAPRAS, Y. ET AL. (2007) An anomaly detector with immediate feedback to hunt for planets of Earth mass and below by microlensing. *MNRAS*, **380**, 792–804.

DONG, S., BOND, I. A., GOULD, A., KOZŁOWSKI, S., MIYAKE, N., GAUDI, B. S., BENNETT, D. P., ABE, F. ET AL. (2009) Microlensing Event MOA-2007-BLG-400: Exhuming the Buried Signature of a Cool, Jovian-Mass Planet. *ApJ*, **698**, 1826–1837.

DONG, S., UDALSKI, A., GOULD, A., REACH, W. T., CHRISTIE, G. W., BODEN, A. F., BENNETT, D. P., FAZIO, G. ET AL. (2007) First Space-Based Microlens Parallax Measurement: Spitzer Observations of OGLE-2005-SMC-001. *ApJ*, **664**, 862–878.

DORAN, M. & MÜLLER, C. M. (2004) Analyse this! A cosmological constraint package for CMBEASY. *J. Cosmology Astropart. Phys.*, **9**, 3–16.

DRIESMAN, A., HYNES, S. & CANCRO, G. (2008) The STEREO Observatory. *Space Sci. Rev.*, **136**, 17–44.

EASTMAN, J., SIVERD, R. & GAUDI, B. S. (2010) Achieving Better Than 1 Minute Accuracy in the Heliocentric and Barycentric Julian Dates. *PASP*, **122**, 935–946.

EFRON, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**(1), 1–26, URL http://links.jstor.org/sici?sici=0090-5364(197901)7:1<1:BMALAT>2.0.CO;2-6&origin=MSN.

EINSTEIN, A. (1905) Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, **322**, 891–921.

EINSTEIN, A. (1911) Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik*, **340**, 898–908.

EINSTEIN, A. (1915) Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*, 844–847.

ESTRADA, J., ANNIS, J., DIEHL, H. T., HALL, P. B., LAS, T., LIN, H., MAKLER, M., MERRITT, K. W. ET AL. (2007) A Systematic Search for High Surface Brightness Giant Arcs in a Sloan Digital Sky Survey Cluster Sample. *ApJ*, **660**, 1176–1185.

EYLES, C. J., HARRISON, R. A., DAVIS, C. J., WALTHAM, N. R., SHAUGHNESSY, B. M., MAPSON-MENARD, H. C. A., BEWSHER, D., CROTHERS, S. R. ET AL. (2009) The Heliospheric Imagers Onboard the STEREO Mission. *Sol. Phys.*, **254**, 387–445.

FISCHBACH, E. & FREEMAN, B. S. (1980) Second-order contribution to the gravitational deflection of light. *Phys. Rev. D*, **22**, 2950–2952.

FISHER, R. A. (1922) On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **222**, 309–368, URL http://dx.doi.org/10.2307/91208.

FRIEDEN, B. (2004) *Science from Fisher Information*. Cambridge University Press, Cambridge.

FRUCHTER, A. S. & HOOK, R. N. (2002) Drizzle: A Method for the Linear Reconstruction of Undersampled Images. *PASP*, **114**, 144–152.

GALASSI, M., DAVIES, J., THEILER, J., GOUGH, B., JUNGMAN, G., BOOTH, M. & ROSSI, F. (2009) *GNU Scientific Library Reference Manual - Third Edition*. 3rd edn., Network Theory Ltd., URL http://www.worldcat.org/isbn/0954612078.

GAUDI, B. S. (1998) Distinguishing Between Binary-Source and Planetary Microlensing Perturbations. *ApJ*, **506**, 533–539.

GELMAN, A. & RUBIN, D. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

GONDOLO, P. (1999) Optical Depth Evaluation in Pixel Microlensing. *ApJ*, **510**, L29–L32.

GÖNNER, H. (1996) *Einführung in die spezielle und allgemeine Relativitätstheorie*. Einführung in Die Spezielle und Allgemeine Relativitätstheorie (Spektrum Akademischer Verlag, Heidelberg, Berlin, 1996).

GOULD, A. (1995) MACHO parallaxes from a single satellite. *ApJ*, **441**, L21–L24.

GOULD, A. (1996) Theory of Pixel Lensing. *ApJ*, **470**, 201–210.

GOULD, A., DONG, S., GAUDI, B. S., UDALSKI, A., BOND, I. A., GREENHILL, J., STREET, R. A., DOMINIK, M. ET AL. (2010) Frequency of Solar-like Systems and of Ice and Gas Giants Beyond the Snow Line from High-magnification Microlensing Events in 2005-2008. *ApJ*, **720**, 1073–1089.

GOULD, A. & GAUCHEREL, C. (1997) Stokes's Theorem Applied to Microlensing of Finite Sources. *The Astrophysical Journal*, **477**(2), 580–584, URL http://stacks.iop.org/0004-637X/477/i=2/a=580.

GOULD, A., GAUDI, B. S. & HAN, C. (2003) Resolving the Microlens Mass Degeneracy for Earth-Mass Planets. *ApJ*, **591**, L53–L56.

GRAFF, D. S. & GOULD, A. (2002) Microlens Parallaxes of Binary Lenses Measured from a Satellite. *ApJ*, **580**, 253–260.

GRIEST, K. & HU, W. (1992) Effect of binary sources on the search for massive astrophysical compact halo objects via microlensing. *ApJ*, **397**, 362–380.

GRIEST, K. & SAFIZADEH, N. (1998) The Use of High-Magnification Microlensing Events in Discovering Extrasolar Planets. *ApJ*, **500**, 37–50.

HARDY, S. J. & WALKER, M. A. (1995) Parallax effects in binary microlensing events. *MNRAS*, **276**, L79–L82.

HASTINGS, W. K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**(1), pp. 97–109, URL http://www.jstor.org/stable/2334940.

HEAVENS, A. F., JIMENEZ, R. & LAHAV, O. (2000) Massive lossless data compression and multiple parameter estimation from galaxy spectra. *MNRAS*, **317**, 965–972.

HOG, E., NOVIKOV, I. D. & POLNAREV, A. G. (1995) MACHO photometry and astrometry. *A&A*, **294**, 287–294.

HORNE, K., SNODGRASS, C. & TSAPRAS, Y. (2009) A metric and optimization scheme for microlens planet searches. *MNRAS*, **396**, 2087–2102.

HOWARD, R. A., MOSES, J. D., SOCKER, D. G., DERE, K. P. & COOK, J. W. (2002) Sun earth connection coronal and heliospheric investigation (SECCHI). *Advances in Space Research*, **29**, 2017–2026.

HOWELL, S. B. (1989) Two-dimensional aperture photometry - Signal-to-noise ratio of point-source observations and optimal data-extraction techniques. *PASP*, **101**, 616–622.

HOWELL, S. B. (2000) *Handbook of CCD Astronomy*.

HUNDERTMARK, M., HESSMAN, F. V. & DREIZLER, S. (2009) Detecting circumstellar disks around gravitational microlenses. *A&A*, **500**, 929–934.

HUSSER, T.-O. (2011) Claude - An Automation Tool for the Monet Telescopes. *ArXiv e-prints*.

IDA, S. & LIN, D. N. C. (2005) Toward a Deterministic Model of Planetary Formation. III. Mass Distribution of Short-Period Planets around Stars of Various Masses. *ApJ*, **626**, 1045–1060.

JOHANSSON, F. ET AL. (2010) *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.14)*. http://code.google.com/p/mpmath/.

KAHAN, W. (1965) Pracniques: further remarks on reducing truncation errors. *Commun. ACM*, **8**, 40, URL http://doi.acm.org/10.1145/363707.363723.

KAYSER, R., REFSDAL, S. & STABELL, R. (1986) Astrophysical applications of gravitational micro-lensing. *A&A*, **166**, 36–52.

KEETON, C. R. & PETTERS, A. O. (2005) Formalism for testing theories of gravity using lensing by compact objects: Static, spherically symmetric case. *Phys. Rev. D*, **72**(10), 104006–1–104006–16.

KEETON, C. R. & PETTERS, A. O. (2006a) Formalism for testing theories of gravity using lensing by compact objects. II. Probing post-post-Newtonian metrics. *Phys. Rev. D*, **73**(4), 044024–1–044024–9.

KEETON, C. R. & PETTERS, A. O. (2006b) Formalism for testing theories of gravity using lensing by compact objects. III. Braneworld gravity. *Phys. Rev. D*, **73**(10), 104032–1–104032–13.

KERINS, E., ROBIN, A. C. & MARSHALL, D. J. (2009) Synthetic microlensing maps of the Galactic bulge. *MNRAS*, **396**, 1202–1210.

KIEFER, J. (1959) Optimum Experimental Designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, **21**(2), pp. 272–319, URL http://www.jstor.org/stable/2983802.

KIRK, D. (2010) *Programming Massively Parallel Processors*. Morgan Kaufmann, San Diego.

LADA, C. J. (2006) Stellar Multiplicity and the Initial Mass Function: Most Stars Are Single. *ApJ*, **640**, L63–L66.

LEE, C., RIFFESER, A., SEITZ, S. & BENDER, R. (2009) Finite-Source Effects in Microlensing: A Precise, Easy to Implement, Fast, and Numerically Stable Formalism. *ApJ*, **695**, 200–207.

LEVENBERG, K. (1944) A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math*, **2**, 164–168.

LIEBIG, C. & WAMBSGANSS, J. (2010) Detectability of extrasolar moons as gravitational microlenses. *A&A*, **520**, 1–13.

MARQUARDT, D. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math*, **11**, 431–441.

MARSHALL, D. J., ROBIN, A. C., REYLÉ, C., SCHULTHEIS, M. & PICAUD, S. (2006) Modelling the Galactic interstellar extinction distribution in three dimensions. *A&A*, **453**, 635–651.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

MIYAMOTO, M. & YOSHII, Y. (1995) Astrometry for Determining the MACHO Mass and Trajectory. *AJ*, **110**, 1427–1432.

MOORE, G. E. (1965) Cramming More Components onto Integrated Circuits. *Electronics*, **38**(8), 114–117, URL http://dx.doi.org/10.1109/JPROC.1998.658762.

MORDASINI, C., ALIBERT, Y., BENZ, W. & NAEF, D. (2009) Extrasolar planet population synthesis. II. Statistical comparison with observations. *A&A*, **501**, 1161–1184.

NELDER, J. & MEAD, R. (1965) A simplex method for function minimization. *Computer Journal*, **7**, 308–313.

NISHIYAMA, S., NAGATA, T., TAMURA, M., KANDORI, R., HATANO, H., SATO, S. & SUGITANI, K. (2008) The Interstellar Extinction Law toward the Galactic Center. II. V, J, H, and K_s Bands. *ApJ*, **680**, 1174–1179.

PACZYNSKI, B. (1986a) Gravitational microlensing at large optical depth. *ApJ*, **301**, 503–516.

PACZYNSKI, B. (1986b) Gravitational microlensing by the galactic halo. *ApJ*, **304**, 1–5.

PACZYNSKI, B. (1991) Gravitational microlensing of the Galactic bulge stars. *ApJ*, **371**, L63–L67.

PRESS, W. (1994) *Numerical Recipes in C*. Cambridge University Press, Cambridge.

RADHAKRISHNA RAO, C. (1945) Information and the accuracy attainable in the estimation of statistical parameters. **37**, 18–91.

REICHARDT, C., JIMENEZ, R. & HEAVENS, A. F. (2001) Recovering physical parameters from galaxy spectra using MOPED. *MNRAS*, **327**, 849–867.

ROBIN, A. C., REYLÉ, C., DERRIÈRE, S. & PICAUD, S. (2003) A synthetic view on structure and evolution of the Milky Way. *A&A*, **409**, 523–540.

SCHNEIDER, P., EHLERS, J. & FALCO, E. (1992) *Gravitational Lenses*. Springer, Berlin, Germany; New York, U.S.A.

SCHNEIDER, P. & WEISS, A. (1986) The two-point-mass lens - Detailed investigation of a special asymmetric gravitational lens. *A&A*, **164**, 237–259.

SCHRAMM, T. & KAYSER, R. (1987) A simple imaging procedure for gravitational lenses. *A&A*, **174**, 361–364.

SCHRÖDER, U. E. (2007) *Gravitation*. Verlag Harri Deutsch, Frankfurt am Main.

SCHWARZSCHILD, K. (1916) On the Gravitational Field of a Mass Point According to Einstein's Theory. *Abh. Konigl. Preuss. Akad. Wissenschaften Jahre 1906,92, Berlin,1907*, 189–196.

SHANNON, C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379–423, 625–56.

SKOWRON, J., UDALSKI, A., GOULD, A., DONG, S., MONARD, L. A. G., HAN, C., NELSON, C. R., MCCORMICK, J. ET AL. (2011) Binary microlensing event OGLE-2009-BLG-020 gives a verifiable mass, distance and orbit predictions. *ArXiv e-prints*.

SMITH, M. C., WOZNIAK, P., MAO, S. & SUMI, T. (2007) Blending in gravitational microlensing experiments: source confusion and related systematics. *Monthly Notices of the Royal Astronomical Society*, **380**(2), 805–818, URL http://dx.doi.org/10.1111/j.1365-2966.2007.12130.x.

SPIEGEL, D. S., BURROWS, A. & MILSOM, J. A. (2011) The Deuterium-burning Mass Limit for Brown Dwarfs and Giant Planets. *ApJ*, **727**, 57–66.

SUMI, T., ABE, F., BOND, I. A., DODD, R. J., HEARNSHAW, J. B., HONDA, M., HONMA, M., KAN-YA, Y. ET AL. (2003) Microlensing Optical Depth toward the Galactic Bulge from Microlensing Observations in Astrophysics Group Observations during 2000 with Difference Image Analysis. *ApJ*, **591**, 204–227.

SUMI, T., WOŹNIAK, P. R., UDALSKI, A., SZYMAŃSKI, M., KUBIAK, M., PIETRZYŃSKI, G., SOSZYŃSKI, I., ŻEBRUŃ, K. ET AL. (2006) Microlensing Optical Depth toward the Galactic Bulge Using Bright Sources from OGLE-II. *ApJ*, **636**, 240–260.

THOMPSON, A. C., FLUKE, C. J., BARNES, D. G. & BARSDELL, B. R. (2010) Teraflop per second gravitational lensing ray-shooting using graphics processing units. *New A*, **15**, 16–23.

UDALSKI, A. (2003) The Optical Gravitational Lensing Experiment. Real Time Data Analysis Systems in the OGLE-III Survey. *Acta Astron.*, **53**, 291–305.

VERMAAK, P. (2000) The effects of resolved sources and blending on the detection of planets via gravitational microlensing. *MNRAS*, **319**, 1011–1019.

VIRBHADRA, K. S. & ELLIS, G. F. R. (2000) Schwarzschild black hole lensing. *Phys. Rev. D*, **62**(8), 084003.

VON NEUMANN, J. (1951) Various techniques used in conjunction with random digits. *National Bureau of Standards Applied Math Series*, **12**, 36–38.

WALKER, M. A. (1995) Microlensed Image Motions. *ApJ*, **453**, 37–39.

WALSH, D., CARSWELL, R. & WEYMANN, R. (1979) 0957+561 A, B: twin quasistellar objects or gravitational lens? *Nature*, **279**, 381–384.

WALTERS, S. J., FORBES, L. K. & JARVIS, P. D. (2010) A kinematical approach to gravitational lensing using new formulae for refractive index and acceleration. *MNRAS*, 953–962.

WAMBSGANSS, J. (1997) Discovering Galactic planets by gravitational microlensing: magnification patterns and light curves. *MNRAS*, **284**, 172–188.

WEINBERG, S. (1972) *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity, by Steven Weinberg, pp. 688. ISBN 0-471-92567-5. Wiley-VCH , July 1972.

WITT, H. J. (1990) Investigation of high amplification events in light curves of gravitationally lensed quasars. *A&A*, **236**, 311–322.

WITT, H. J. & MAO, S. (1994) Can lensed stars be regarded as pointlike for microlensing by MACHOs? *ApJ*, **430**, 505–510.

WITT, H. J. & MAO, S. (1995) On the Minimum Magnification Between Caustic Crossings for Microlensing by Binary and Multiple Stars. *ApJ*, **447**, L105–L108.

WOZNIAK, P. R. (2000) Difference Image Analysis of the OGLE-II Bulge Data. I. The Method. *Acta Astronomica*, **50**, 421–450.

YOO, J., DEPOY, D. L., GAL-YAM, A., GAUDI, B. S., GOULD, A., HAN, C., LIPKIN, Y., MAOZ, D. ET AL. (2004) Constraints on Planetary Companions in the Magnification A=256 Microlensing Event OGLE-2003-BLG-423. *ApJ*, **616**, 1204–1214.

ZIMA, W., ARENTOFT, T., DE RIDDER, J., SALMON, S., CATALA, C., KJELDSEN, H. & AERTS, C. (2010) The PLATO End-to-End CCD Simulator – Modelling space-based ultra-high precision CCD photometry for the assessment study of the PLATO Mission. *ArXiv e-prints*.

Appendix A

GPU-contouring

Code details

The implementation of the root-finding is illustrated in Fig. A.1 and indicates how each radial direction is used for finding the contour. If less than 20 directions provide roots, the source star radius is increased in order to fulfill this requirement. The refinement for the initially chosen radius is started afterwards.
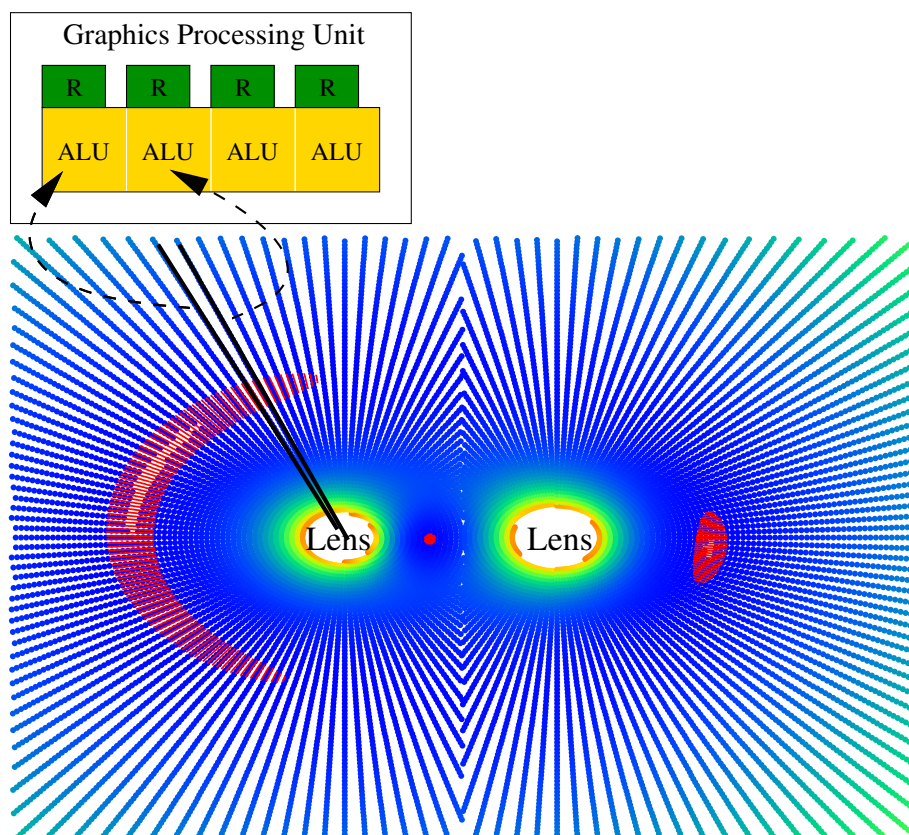
Figure A.1: GPU-contouring implementation details.

Runtime issues

A runtime analysis of the unrefined GPU-model relying exclusively on one grid-search, reveals the limiting factor in the simulation. Despite efforts to accelerate the root-finding, it is responsible for the highest contribution to the total runtime, as illustrated in Fig. A.2. One reason is the limited occupancy (33%), which will can be resolved by GPUs with more registers.
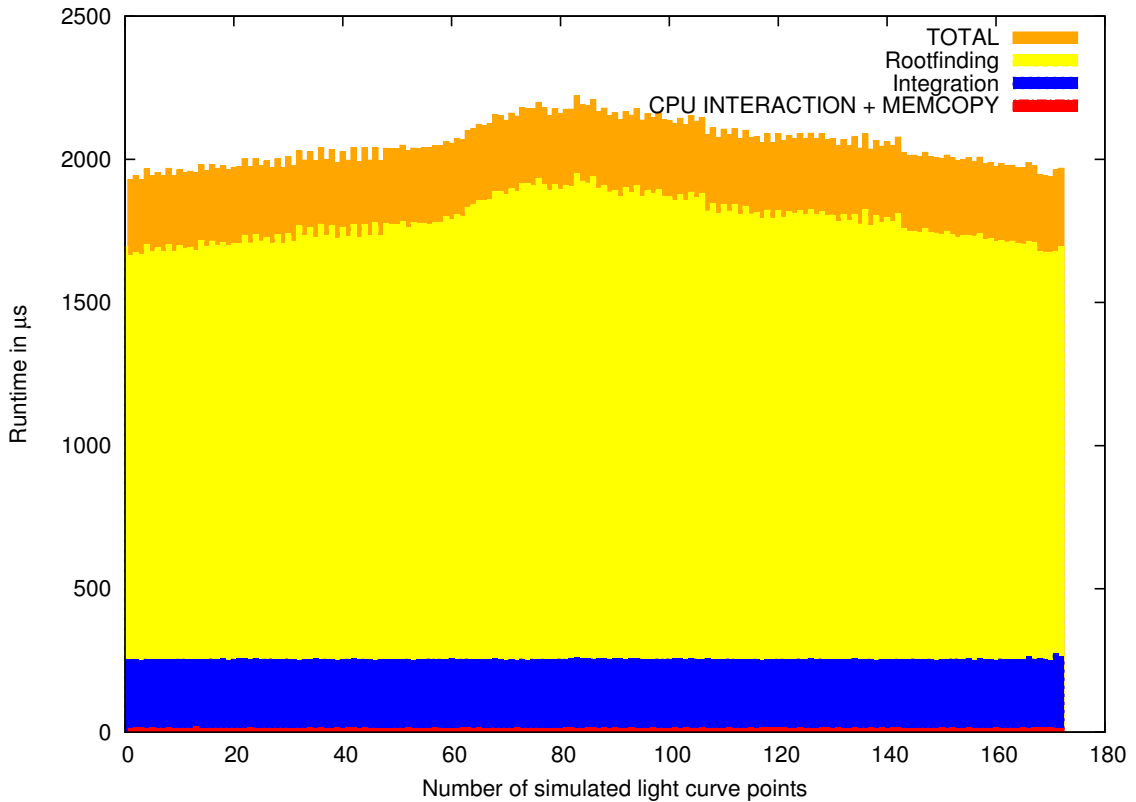
Figure A.2: Runtime of the unrefined GPU-contouring model.

Attempts of rearranging the computation of the squared deviation function, did not reduce the total runtime, due to higher latencies. In addition, a limit for the maximal number of root-finding iterations was tested. This can speed-up the computation drastically, because some multiprocessors have to wait for others for finishing the search. Unfortunately, the systematic noise increases, especially in the wings of a lightcurve, where solutions are more than $\approx 1\,\theta_E$ away from the critical curve and thus the search stops before reaching the image contour.

The CPU interaction cannot be optimized for the unrefined model, but for the refined model flags for indicating the convergence are exchanged. Finally, there is only the integration time which can be reduced, but given that it is contributing only 10-15 %, the expected acceleration is negligible.

Appendix B

Data analysis

Multimodal likelihood

The distributions of binary parameters for the microlensing event MOA 2010-BLG-406 are multi-modal. The multimodal structure of the target distribution is a consequence of adjacent parameter distributions with the same χ^2. This fact is illustrated in Fig. B.1, where these
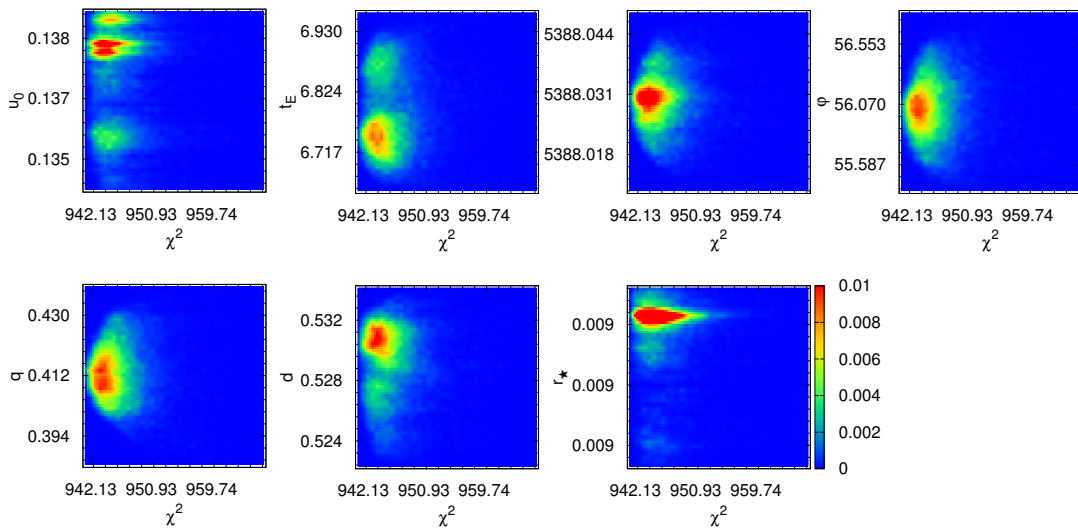
Figure B.1: Target distribution parameters and χ^2.

B.1   Extinction and Galactic model effects

The physical interpretation of the binary fit for MOA 2010-BLG-406 relies on the Galactic coordinates of the catalog simulation and on appropriate extinction coefficients. In order to illustrate what happens if the catalog simulation for $l = 1.2°$, $b = -2.7°$ and an extinction of $0.3\,\mathrm{mag/kpc}$ is used, Fig. B.2 shows how the resulting masses of the components are affected. The interpretation given in this context, crucially depends on the correctness of the Galactic model and the extinction.
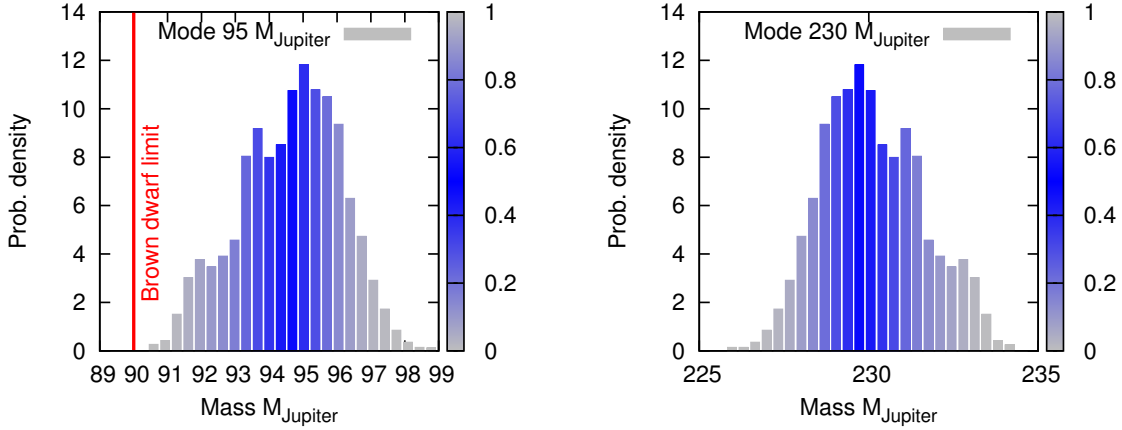
Figure B.2: Binary mass inferred for $l = 1.2°$, $b = -2.7°$ and $0.3\,\mathrm{mag/kpc}$.

Acknowledgements / Danksagung

Curriculum Vitae

Dipl.-Phys. Markus P. G. Hundertmark
Humboldtallee 16
37073 Göttingen

Geburtstag:	15.02.1982
Geburtsort:	Göttingen
Staatsangehörigkeit:	Deutsch

08/1987 – 07/1991	Grundschule Reinhausen
08/1991 – 07/1993	Orientierungsstufe Groß Schneen
08/1993 – 07/1998	Felix-Klein-Gymnasium Göttingen
23.6.2001	allgemeine Hochschulreife
10/2002 – 09/2007	Physikstudium (Georg-August-Universität Göttingen)
22.10.2004	Vordiplom Physik (Georg-August-Universität Göttingen)
13.09.2007	Diplom Physik (Georg-August-Universität Göttingen)
10/2007 – 06/2011	Promotionsstudent (Georg-August-Universität Göttingen)