Title
Theoretical Advances in Gravitational Microlensing Guided by Artificial Intelligence

Author
Zhang, Keming

Publication Date
2023

Peer reviewed|Thesis/dissertation

Theoretical Advances in Gravitational Microlensing Guided by Artificial Intelligence

By

Keming Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Astrophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Joshua S. Bloom, Chair
Professor Jessica R. Lu
Professor Uroš Seljak

Spring 2023

Theoretical Advances in Gravitational Microlensing Guided by Artificial Intelligence

Abstract

Theoretical Advances in Gravitational Microlensing Guided by Artificial Intelligence

by

Keming Zhang

Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Joshua S. Bloom, Chair

Three decades have passed since the technique of gravitational microlensing was proposed as a means for exoplanet detection, and nearly 200 microlensing exoplanets have been discovered to date. Previously, theoretical studies of the two-body point-mass gravitational lens have primarily focused on the properties of caustics, which are the singularities in the magnification map. The invariances and symmetries of microlensing caustics have led to the identification of physical degeneracies that cause distinctly different lens configurations to give rise to nearly identical observations. Nevertheless, inconsistencies in the application of existing degeneracy theories to observed events indicate that our current theoretical understanding of binary-lens gravitational microlensing, which was largely laid out in the late twentieth century, may in fact be incomplete.

This thesis introduces a novel approach to utilizing Artificial Intelligence as a driver for theoretical explorations, which represents a departure from its traditional role in accelerating empirical discoveries. First, a scalable inference framework is developed for binary-lens microlensing using the technique of Neural Posterior Estimation, which is then applied to model hundreds of simulated microlensing light-curve observations. By examining the large numbers of multi-modal modeling solutions, I propose and subsequently prove the offset degeneracy, which is shown to be ubiquitous in the interpretation of planetary microlensing observations, and unifies two leading types of caustic degeneracies as limiting cases. Motivated by properties of the offset degeneracy, I subsequently propose the generalized perturbative picture for planetary microlensing, which states that the planet can be considered to act as a variable-shear Chang-Refsdal lens on one of the images produced by the host star, leaving the other image largely unaffected. The analytic nature of the Chang-Refsdal lens indicates that the proposed formalism would allow for full magnification maps of the planetary lens to be derived analytically, thereby facilitating the accelerated modeling of observed events. The methodologies and results presented in this thesis may substantially benefit the analysis of the deluge of data expected from the first space-based microlensing survey of the Roman Space Telescope.

I dedicate this dissertation to my parents

Contents

List of Figures

List of Tables

Acknowledgments

I have been an avid stargazer and astrophotographer since the age of twelve, and for this reason, I have always wanted to become a professional astronomer. My most treasured childhood memories are, without doubt, the countless 1.5-hour road trips along road G-108 towards Mt. Baihua (elev. 6500 ft) near my hometown Beijing, where my parents would stay up all night with me to photograph the gorgeous nebulae and galaxies with my 8-inch reflector telescope. This would have been no easy task for any parent, as the temperature often drops to as low as $-20°$C ($0°$F). Therefore, I would like to first and foremost thank my parents, whose unwavering support for my pursuit of astronomy from an early age has been instrumental to my decision to pursue a Ph.D. and career in astronomy.

Given my avid pursuit of astrophotography, I have naturally aspired to become an observational astronomer. However, towards the end of college, I had hoped to pursue something off the beaten path for my Ph.D. This is when I met my advisor Josh Bloom, who had just had a successful exit from his start-up company and returned full-time to academia. Josh's entrepreneurship was reflected in his daring to conceptualize and bring to fruition the boldest ideas in science, which made us an ideal match. Josh has been a great source of inspiration, and our conversations have always been filled with new ideas, sometimes moonshot ideas. As a result, there have indeed been many turning points over the years that would have led me to pursue a completely different yet equally exciting thesis. Towards finishing Chapter 3, I had initially been hesitant to pursue the idea of searching for new microlensing degeneracies, arguing that even if I made any discoveries, it would have required the caliber of a theorist to make a proper explanation. Josh nevertheless pushed me to pursue the idea, which essentially turned my initial reservation into a self-fulfilling prophecy. In a twist of events, the solution to my initial reservation turned out to be for me to become a theorist myself, and I'm deeply grateful to Josh for supporting my pursuits toward the end of my Ph.D. that have taken a remarkable detour from our initial course. Over the years, Josh has always been available to meet with me for hours at a time, and this level of personal attention and guidance may be the biggest luxury and privilege to have as a graduate student.

I would also like to express my deepest gratitude to Scott Gaudi and Jessica Lu, who first introduced me to the exciting field of microlensing. Scott's generous mentorship over the years has been indispensable for my growth from a microlensing novice to working at the frontiers of this field. We have shared countless moments of exhilaration of making one surprising discovery after another. I can still vividly recount what Scott referred to as a

"magic trick" that I did to persuade him that a longstanding theory has always been inexact. Jessica has patiently guided me through the minute details of observing with the largest observatories in the world, which has been my childhood dream. Despite her unmistakable disinterest in exoplanets, our interactions have nevertheless nurtured within me a burgeoning fascination for black holes.

I would also like to thank the many other scientists whom I have had the pleasure of interacting with and learning from during graduate school, including Eric Agol, Vanessa Böhm, Sihao Cheng, Miles Cranmer, Courtney Dressing, Dan Foreman-Mackey, Sara Jamal, Tharindu Jayasinghe, François Lanusse, Shude Mao, Jorge Marínez-Palomera, Ben Nachman, Guy Nir, Uroš Seljak, Sean Terry, Stéfan van der Walt, Weicheng Zang, and Ruiqi Zhong.

As is the case for many of my peers, my graduate studies have been substantially impacted by the COVID pandemic. I'm particularly grateful to Shude Mao and colleagues, who have generously hosted me at Tsinghua University during this challenging period. My unexpected visit to Tsinghua has gifted me new friendships, which has also led to new collaborations as I returned to Berkeley. I'd also like to thank the archery teams at both Tsinghua and Berkeley, which have been the center of my social life at both institutions and have also unquestionably contributed to my better science.

Last but not least, it is important to acknowledge the funding agencies for their financial support. This thesis work has been primarily supported by private agency funding, including a Data-Driven Discovery grant from the Gordon and Betty Moore Foundation, and a faculty research award from the Two Sigma Cooperation. The generous and unrestricted funding have granted this thesis work substantial freedom to explore, and I'm deeply indebted to Josh for securing these funding. I thank Josh for leading the NSF proposal titled "Accelerating Astrophysical Insight at Scale with Likelihood-Free Inference," which was partly based on the initial results of this thesis. I thank the National Science Foundation for subsequently funding this project under award #2206744.

Chapter 1

Introduction

1.1 Historical Background of Gravitational Lensing

In his seminal 1936 paper "Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field," Einstein (1936) introduced a simple model describing how light from a distant star could be bent by the gravitational field of a massive foreground object, leading to the magnification of the background star's brightness. In this paper, Einstein presented a simple formula relating the magnification factor to the projected separation between the two objects:

$$q = \frac{l}{x}\frac{1 + x^2/2l^2}{\sqrt{1 + x^2/4l^2}}, \tag{1.1}$$

where x is the projected lens-source separation and

$$l = \sqrt{\frac{4GM}{D_{\rm rel}c^2}}, \tag{1.2}$$

where G is the gravitational constant, M is the lens mass, c is the speed of light, and $D_{\rm rel}^{-1} = D_{\rm lens}^{-1} - D_{\rm source}^{-1}$ is related to the relative distance between the lens and source. This quantity is now commonly referred to as the Einstein ring radius of the lens star.

The simple expression of Equation 1.1 has been frequently re-derived over the past century, including in Refsdal (1964), Liebes (1964), and notably later in Paczyński (1986b), Equation 5:

$$A = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} = \frac{1}{u}\frac{1 + u^2/2}{\sqrt{1 + u^2/4}}, \tag{1.3}$$

where u is the projected lens-source separation in units of the lens-star Einstein radius, and thus $u = l/x$. The point-lens point-source (PSPL) light curve is often attributed to Equation 1.3 and therefore commonly referred to as the "Paczyński light curve" in the literature. However, as I have shown, the correct PSPL magnification was first given in Einstein (1936) in essentially the identical form. Therefore, I contend that despite its popularity, the term "Paczyński light curve" may not be warranted.

In the same paper, Einstein also famously stated that "there is no great chance of observing this phenomenon," arguing that the required alignment of the background source star, the foreground lens star, and the observer was highly improbable. One year later, Zwicky (1937) presented the case of gravitational lensing by distant galaxies, where he showed that such a phenomenon is very likely to be observed on an extragalactic scale. Given the limitation of observing techniques at the time, the topic of gravitational lensing then faded into oblivion for nearly three decades, and was only briefly picked up by Refsdal (1964) and Liebes (1964), who presented detailed and systematic treatments of gravitational lensing by stars as well as the feasibility of observing them. Refsdal (1964) concluded that "[due] to progress in experimental technique we find, contrary to Einstein, that the effect may be of practical interest," a conclusion that was shared by Liebes (1964).

A relevant development at the time was the discovery of quasars in the late 1950s, with the increasing popularity of radio astronomy. As anticipated by Zwicky (1937), the first observational evidence of gravitational lensing came as QSO 0957+561 (Walsh et al. 1979; Pooley et al. 1979), which, at the time of its discovery, was correctly speculated as two images of a quasar produced by a massive foreground galaxy acting as a foreground lens. The discovery of QSO 0957+561 then marked the transition of gravitational lensing from a theoretical concept into an observable phenomenon. Shortly after its discovery, Chang & Refsdal (1979) suggested that individual stars in the halo of the lensing galaxy could perturb one of the two QSO images into two or four images, which will cause short-term perturbations with timescales of-order months or years that would manifest in only one of the quasar images. The collective effect of multiple stars within the lensing galaxy was explored in later works such as Gott (1981) and Chang & Refsdal (1984).

The first observational evidence of this time-variable lensing effect caused by stars—nowadays referred to as microlensing (Paczyński 1986a)—came a decade later in the case of QSO 2237+0305 (Huchra et al. 1985), where significant brightness variations in the brightest of the four quasar images were considered the first observed microlensing event (Irwin et al. 1989). It should be noted that the term "microlensing" was first introduced in Paczyński (1986a), which also employs the term "macrolensing" to distinguish between the lensing effects of a galaxy as a whole and the effects of individual stars within that galaxy. The scope of the term microlensing was later extended to the time-variable lensing phenomenon produced by stars in our own galaxy in Paczyński (1986b), who proposed to utilize microlensing to test the hypothesis that the dark matter halo of the Milky Way Galaxy is composed of MAssive Compact Halo Objects (MACHOs) — a class of hypothetical dark matter constituents consisting of massive, non-luminous objects, such as brown dwarfs, black holes, or even rogue planets. Such a scenario would suggest a microlensing optical depth[1] of $\tau \sim 10^{-6}$ (Paczyński 1986b), requiring the monitoring of millions of stars to yield meaningful statistical constraints.

While the simultaneous monitoring of millions of stars would have been improbable when Einstein (1936) first conceived of this phenomenon, the advent of CCD technology and the

[1]The microlensing optical depth is defined as the fraction of solid angle that is covered by the Einstein rings of all lensing objects along the line of sight to the source.

development of large-scale astronomical surveys allowed *Galactic* microlensing to become a practicality. The Large Magellanic Cloud (LMC) and the Small Magellanic Cloud (SMC) were then identified to be favorable targets to detect MACHOs given their high stellar surface densities. Massive microlensing surveys took off in the early 1990s, including the Optical Gravitational Lensing Experiment (OGLE; Udalski et al. 1993), the MACHO project (Alcock et al. 1996), and the Expérience pour la Recherche d'Objets Sombres (EROS; Aubourg et al. 1993). In particular, MACHO discovered 13 – 17 microlensing events by monitoring 11.9 million stars in the LMC during its operation from 1992 to 1998 (Alcock et al. 2000). OGLE-II, which operated from 1996 to 2000, detected only two candidate events (Wyrzykowski et al. 2009) towards the LMC. The number of events observed was far fewer than what would be expected if MACHOs were the primary source of dark matter in the Milky Way's halo. The microlensing optical depth derived from OGLE observations towards the SMC was $\tau \sim 1.3 \pm 1.01 \times 10^{-7}$ (Wyrzykowski et al. 2011), which was consistent with the expected contribution from Galactic disc and SMC self-lensing, leading to the conclusion that "there is no need for introducing any special dark matter compact objects in order to explain the observed event rates."

An alternative scenario of microlensing of Galactic bulge stars by galactic disk stars in our own galaxy was envisioned by Paczyński (1991), which was used by OGLE-I as a proof of concept because the distribution of galactic disk stars is well understood, thus creating a scenario where microlensing must operate. In the same year, Mao & Paczyński (1991) considered the effects of a secondary body of the galactic disk lens and suggested that a "massive search for microlensing of the Galactic bulge stars may lead to a discovery of the first extra-solar planetary systems." The practicality of microlensing as a method for planet detection was studied in more detail in later works such as Gould & Loeb (1992), Gaudi & Gould (1997), and Griest & Safizadeh (1998).

As the focus of this thesis concerns the theoretical aspects of gravitational microlensing, the exciting history of microlensing planet discovery is omitted, and the interested reader is referred to Gaudi (2012) for a comprehensive review. Instead, let us now focus on the theoretical foundations for interpreting observations of planetary microlensing events. A prominent feature of the two-body point-mass gravitational lens, including star-planet lenses, is the existence of extended caustics, which are the singularities in the source plane associated with infinite magnification. The properties of binary-lens caustics have been extensively studied in (e.g., Schneider et al. 1992), which have led to the identification of physical degeneracies (Gaudi & Gould 1997; Griest & Safizadeh 1998) under planetary mass ratios ($q = M_\odot/M_\oplus \ll 1$).

In particular, Gould & Loeb (1992) and Gaudi & Gould (1997) considered the similarity between the planetary lens and the lens formalism first introduced in Chang & Refsdal (1979), later referred to as the Chang-Refsdal lens. In both cases, the secondary object is orders of magnitude less massive than the primary object, where the vicinity of the secondary object may be described by a point-mass lens perturbed by uniform external shear. Therefore, a planetary lens companion would effectively act as a Chang-Refsdal lens on one of the two images produced by the host star. This physical picture is subsequently referred to

as the perturbative picture of planetary microlensing. In this context, Gaudi & Gould (1997) then pointed out that the Chang-Refsdal formalism would lead to an ambiguity in the interpretation of observed planetary microlensing events as to "whether the planet lies closer to or farther from the star than does the position of the image that it is perturbing," which is commonly referred to as the "inner-outer degeneracy" for planetary caustics, following the nomenclature of Han et al. (2018). Such an effect has been invoked to interpret the bi-modality of the parameter posterior of a multitude of observed planetary events. Fortunately, the perturbative picture requires that the image being perturbed lies close to the secondary object in order of its Einstein radius (Gould & Loeb 1992), and thus the resulting degenerate parameters would only have a marginal difference in the projected star-planet separation. The reader may refer to Chapter 6 for details regarding this topic.

However, an alternative degeneracy first identified by Griest & Safizadeh (1998) results in solutions to the projected star-planet separation that could differ by several astronomical units, and thus is of greater practical concern. Such degeneracies occur for high-magnification events, where planets with projected separations of s and $1/s$ in units of the angular Einstein radius of the primary star give rise to similar light-curve features associated with central caustics. This ambiguity is commonly referred to as the "close-wide" degeneracy, whose theoretical origins as the invariance of central caustics under the $s \leftrightarrow 1/s$ transformation was studied in detail in Dominik (1999) and An (2005), among others.

While the two aforementioned types of degeneracies have been invoked to interpret the great majority of multi-modality of model parameter posteriors for observed planetary microlensing events, many empirical degeneracies in observed events are in fact outside the regime in which these degeneracies are derived (Yee et al. 2021). In brief, both types of degeneracies require that the planet exist far from the Einstein radius of the host star, which is equivalent to requiring that the central and planetary caustics be well separated. Nevertheless, this requirement is not even satisfied in MOA-2016-BLG-319 (Han et al. 2018), where the reference to the two solutions as inner and outer are first used. In fact, the outer solution had the planetary and central caustics merged as a single resonant caustic.

The primary focus of the field of planetary microlensing promptly transitioned from theory to observation at the turn of the century, with the definitive detection of the first microlensing planet in 2004: OGLE-2003-BLG-235Lb (Bond et al. 2004). To date, nearly 200 exoplanets have been discovered through microlensing[2], which has been steadily increasing at a rate of two to three dozens of planet discoveries per year (e.g., Gould 2022). In the two decades of observational campaign, there is often an implicit understanding that our theoretical framework of microlensing by two point-masses has long been mature. The above discussion indicates that this assumption may be untrue, and it is now opportune to undertake a renewed investigation of the problem at hand.

[2]NASA Exoplanet Archive (https://exoplanetarchive.ipac.caltech.edu). Retrieved May 2023.

1.2 Machine Learning in Time-Domain Astronomy

The novel approach taken in this thesis work is to apply Artificial Intelligence (AI) and Machine Learning (ML) methods to guide theoretical studies in microlensing as a time-domain phenomenon. Precedents of adapting AI for theoretical studies were rather rare, as AI and ML have traditionally been applied to accelerate empirical discoveries. Therefore, it is necessary to first review the role AI and ML has played in the time domain.

The initial use of ML in the time domain has been focused on variable stars, which early microlensing surveys have identified in large numbers as a by-product. In particular, the OGLE experiment alone has identified over half a million eclipsing or ellipsoidal binary stars (Soszyński et al. 2016), ∼25000 RR Lyrae Stars (Soszyński et al. 2009), among others. These numbers and the phenomenological diversity of variable stars greatly outweigh those for microlensing. More recent surveys, such as the Palomar Transient Factory (PTF; Law et al. 2009) and the Zwicky Transient Facility (ZTF; Bellm et al. 2018), have continued to expand the catalog of variable stars. Nearly a million periodic variables have been identified in the ZTF Data Release 2 alone, including around 350,000 eclipsing binaries, 100,000 long-period variables, and about 150,000 rotational variables (Chen et al. 2020).

Initial efforts for the systematic classification of periodic variable stars employed color-magnitude diagrams (CMDs), period-luminosity relations, and period-color diagrams to classify them based on their specific properties (e.g., Soszyński et al. 2008, 2009). Given that only basic light-curve properties such as color and period are employed, some manual inspection and visual analysis of light curves remained necessary to filter out mis-classifications. As the volume of data continued to grow, it became increasingly apparent that relying on limited sets of features and manual inspection was not sustainable for efficient and accurate classification of the dozens of types of variable stars. Researchers began to recognize the need for more comprehensive feature sets and automated techniques that could better capture the underlying properties of variable stars. This marked the beginning of a new era in time-domain astronomy, where machine learning methods were introduced to revolutionize the process of variable star classification.

A pioneering contribution to this new approach was Debosscher et al. (2007), which was the first work to tackle many-class (>20) classification with supervised machine learning. Following this work, Richards et al. (2011) introduced tree-based methods for variable star classification for the first time, and demonstrated their superior performance based on a comprehensive set of 52 features carefully chosen to represent various aspects of light curves (Naul et al. 2016). Random forests (Breiman 2001) are ensembles of decision trees, which is rather analogous to traditional methods imposing boundaries in phase diagrams. Unlike traditional methods that require manually determining boundaries based on limited features, random forests take advantage of large sets of features, where decision boundaries are learned from the training data automatically.

Following Richards et al. (2011), numerous studies have adopted random-forest-based classification for variable stars (e.g., Dubath et al. 2011; Nun et al. 2014; Miller et al. 2015; Kim & Bailer-Jones 2016). However, the random forest and feature engineering approach

has its own set of challenges and limitations. One of the main issues is the need for domain expertise to create relevant and effective features that capture the essential properties of variable stars. This can be time-consuming and does not guarantee optimal performance, as the features may still miss some crucial information contained within the light curves, and with the addition of more features also comes the increase in computation time.

Around the same time as the Richards et al. (2011) work, there was tremendous progress in the field of AI and computer vision, which are primarily driven by the advancement of deep convolutional neural networks (CNNs), such as AlexNet (Krizhevsky et al. 2012). AlexNet is a neural network architecture that achieved a top-5 error rate of 15.3% in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, which was a significant improvement over the previous best performance of a top-5 error rate of 25.8% — based on traditional CV algorithms and shallow neural networks — that has been long plateauing. The success of AlexNet marked a turning point in the field of computer vision, as it demonstrated the capabilities of deep neural networks for representation learning, allowing for low-dimensional feature representations to be acquired from high-dimensional raw data automatically without the need for explicit feature engineering.

This progress in deep learning eventually found its way into variable star classification, as first demonstrated by the work of Naul et al. (2018), who proposed a recurrent neural network (RNN) for feature extraction on variable-stars light curves. It was demonstrated that the feature space automatically learned by the RNN enables comparable, and often superior random forest classification results as compared to multiple sets of previously published hand-crafted features, including the Richards et al. (2011) and Kim & Bailer-Jones (2016) sets of features. A critical issue in applying deep learning to light curves is astronomical light curves are characterized by periodicity, heterogeneous noise, irregular time sampling, and multiple channels that do not align with each other, which are uncommon in real-world applications of ML. Naul et al. (2018) addressed some of these challenges, but other issues remain wide open (e.g., Jamal & Bloom 2020). A specialized network architecture for periodic irregular light curves is developed in Chapter 2.

1.3 Likelihood-Free Inference

The growing maturity of efficient classification techniques for light-curve data obtained from time-domain surveys has led to the compilation of extensive, well-classified time-domain datasets, which have enabled the study of time-domain phenomena at scale. One of the fundamental challenges to the modeling of observed phenomena is the lack of inverse models. Indeed, the intricate physical forward models, which describe both the astrophysical processes and the instrumental effects underlying observed phenomena, are rarely invertible. As a result, statistical inference algorithms such as MCMC (e.g., Goodman & Weare 2010) and Nested Sampling (Skilling 2006) are required to extract model parameters and the associated uncertainties from observed datasets by means of iterative data-model comparisons, sometimes requiring millions of iterations. Here, the high information content from the high

S/N data acquired by modern facilities has necessitated the development of detailed forward models, at the expense of much increased computational cost. Moreover, the scale of current and future survey experiments indicates that such analysis would be performed for millions of stellar objects simultaneously, which often renders our current tool-set intractable.

Likelihood-Free Inference (LFI), also known as Simulation-Based Inference (SBI), presents a powerful solution to this imminent inference crisis faced by time-domain astronomy. LFI comprises a diverse set of methods that circumvent the need for explicit likelihood calculations, making them particularly advantageous for inference problems where exact likelihood evaluation is infeasible or computationally intensive. These methods were initially developed to address inference problems with intractable likelihood functions, often due to the presence of a high-dimensional nuisance parameter space, situations that are especially common in cosmology. In comparison, the inference problems in the time domain are often blessed with tractable likelihood functions, where the primary obstacle is the computational cost of the forward model and effective strategies to fully explore the parameter space. For this reason, LFI methods have been less relevant for the time domain until very recently.

Nevertheless, a classical LFI method known as Approximate Bayesian Computation (ABC) should be briefly mentioned for the sake of the completeness of this discussion. The reader is referred to Cranmer et al. (2020) for a complete overview of LFI methods and their use cases. In the ABC framework, summary statistics are calculated for both the observed data and the forward model, which are compared in terms of distance metrics in place of the raw, high-dimensional data. For example, summary statistics for light-curve data could be the mean, median, standard deviation, skew, etc. These summary statistics effectively reduce the dimensionality of the data while still retaining the essential information needed for parameter estimation.

Thus, the use of summary statistics in ABC is rather analogous to the use of engineered features for random-forest classification of variable star data, as previously discussed. However, the primary goal of the summary statistics is to enable an effective marginalization over the intractable parameter space. For example, the two-point correlation function (2PCF) is a measure of the excess probability of finding two galaxies at a given separation compared to a random distribution, which serves to quantify the clustering of galaxies in the universe. By using the 2PCF or the power spectrum (its Fourier transform) as a summary statistic for cosmological inference, one is effectively marginalizing over the intractable space of the initial cosmological conditions. Applications of ABC include supernova cosmology, (Weyant et al. 2013), galaxy demographics (Cameron & Pettitt 2012), exoplanet occurrence rates (Hsu et al. 2018), studies of the galaxy-halo connection (Hahn et al. 2017), among others.

While inference problems with tractable likelihoods do not generally benefit from ABC methods, new types of LFI methods driven by advances in deep learning have become increasingly relevant to problems that are otherwise amenable to asymptotically exact inference. First, representation learning with deep neural networks can substitute for expert-crafted summary statistics for automatic "featurization." Second, Neural Density Estimators (NDEs; e.g., Papamakarios et al. 2017) have been developed as specialized neural networks that are capable of modeling probability distributions with arbitrary and complex shapes and co-

variances. These two aspects combined have given rise to a new technique called Neural Posterior Estimation (NPE), where the NDE is employed to learn the Bayesian posterior parameter distribution purely as a conditional distribution (Papamakarios & Murray 2016). The conditional is given to the NDE as the output of the "featurizer network" in the form of a low-dimensional feature vector. By training on parameters and simulations produced using the forward model, NPE essentially enables the creation of surrogate inverse models for otherwise non-invertible physical models. The inverse models parameterized by neural networks can generally be evaluated in less than a few seconds, thus enabling the *amortization* of inference.

The concept of amortization refers to a computational approach in probabilistic modeling where the cost of performing inference is spread out or "amortized" over multiple tasks, rather than being computed individually for each instance. For NPE, the cost of inference is largely incurred upfront, which includes the computational cost of generating a training set using the physical forward model, as well as the cost of training the neural network. Once the inverse model is acquired, one can then produce accurate Bayesian model posteriors for any number of observations at marginal cost. In contrast to traditional inference approaches, where the cost of inference grows linearly with the number of tasks, the computational cost for NPE—training a surrogate inverse model—is largely independent of the scale of the downstream task.

The amortization of inference is particularly advantageous in the context of large-scale time-domain surveys with high data fidelity. In the context of microlensing, the Galactic-Bulge time-domain survey of the future Roman Space Telescope (Spergel et al. 2015) is expected to observe approximately 50,000 microlensing events, including around 1,400 two-body planetary events (Penny et al. 2019) and at least a few thousand stellar-binary events. Thus, Roman is expected to push the number of planets detected by microlensing by nearly an order of magnitude. The prevalence of degeneracies (Section 1) indicates that the microlensing inverse problem is intrinsically difficult even at the current scale. Thus, an NPE framework will provide a powerful solution to solving the Roman microlensing inverse problem, and such a framework is presented in Chapter 3. As the scale of astronomical surveys expands, the development of NPE analysis approaches will be crucial for fully leveraging the wealth of data these surveys provide.

1.4 AI-Guided Theoretical Exploration

With the development of the NPE microlensing framework in Chapter 3, the natural next step would have been to adopt such a framework for the analysis of current datasets, potentially leading to new statistical constraints on certain microlensing populations. Indeed, the use of NPE for theoretical studies in microlensing represents an unexpected yet fortuitous development of this thesis. The genesis of this idea came from an attempt to test the limits of the NPE framework to produce complex multi-modal posteriors. The events OGLE-2011-BLG-0950 and OGLE-2011-BLG-0526 (Choi et al. 2012) became excellent test cases, each of

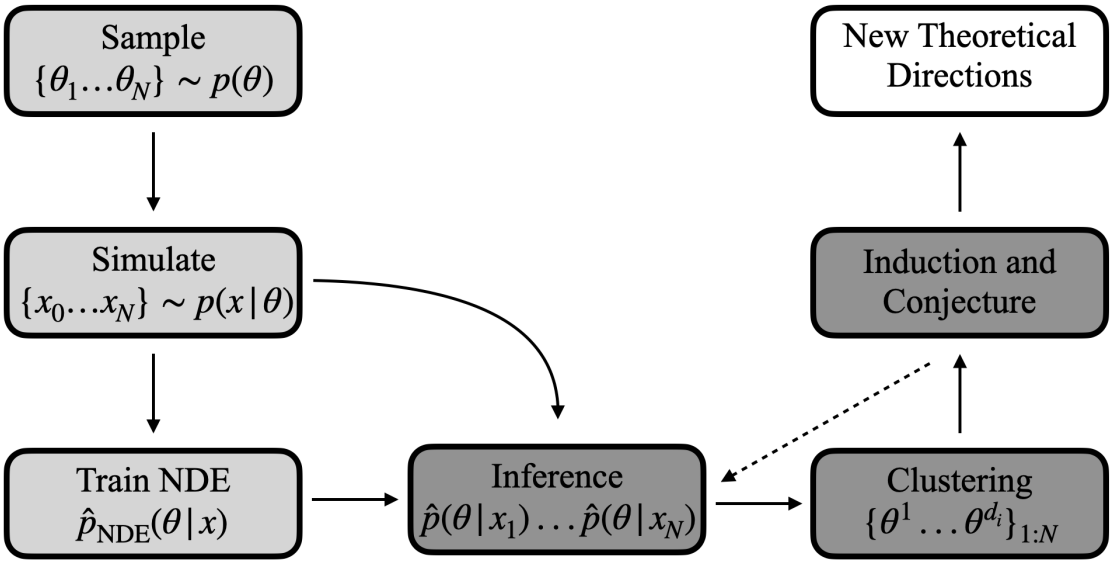


Figure 1.1: Schematic illustration of the use of Neural Posterior Estimation for phenomenological and theoretical studies.

which has a complex four-fold degenerate posterior that includes both stellar-binary solutions and planetary solutions.

In brief, the NPE framework not only recovered the expected posterior modes, but also produced an additional stellar-binary solution (Figure 3.5) that has not been reported in the original analysis. The question of the completeness of the Choi et al. (2012) analysis aside, this serendipitous revelation immediately pointed to a novel pathway for searching for new types of degeneracies, simply by using NPE to produce large numbers of examples and finding out the interesting ones. Indeed, new types of degeneracies are occasionally reported in the analysis of observed events, including the four-fold degeneracy[3] in Choi et al. (2012), along with a more recent discovery reported by Yang et al. (2022). The use of NPE to produce the parameter posterior distributions for a large number of simulated observations as examples of degeneracies, would then potentially produce these serendipitous discoveries systematically and at scale.

In Figure 1.1, I illustrate how neural surrogate inverse modeling may serve to guide theoretical studies. The first step is to acquire the surrogate inverse model by training a conditional NDE on simulated data, which is illustrated as the three boxes on the left-hand side of the figure. The lower-right corner of Figure 1.1 containing three boxes is the key to this AI-guided approach and is where the human agent interfaces with the AI agent. Here, a clustering algorithm automatically identifies the discrete solution modes for the NDE-produced parameter posterior distribution of each simulated observation. Those with two

[3]While Choi et al. (2012) reported on the degeneracy between stellar-binary and planetary lenses as "a new type ambiguity" in its title, it should, in fact, be seen as an extension of the Han (2008) degeneracy. In both cases, stellar-binary and planetary lenses can produce suppressed magnification in between central caustic cusps, which manifest either as a double-peak or a flat-peak light-curve.

or more posterior modes are then visually inspected by the human agent with the question: is this multi-modality predicted by existing theories of degeneracies? Those that do not conform to existing theories are then collected, which would then serve as the basis for the conjecture of new theories and further follow-up studies, as indicated in the top right box of Figure 1.1. In this process, one may also find certain regions of parameter space that are of particular interest, where one may then acquire additional degeneracy examples in this targeted region to clarify different hypotheses (arrow with dashed lines in Figure 1.1).

The execution of this new framework then quickly lead to the discovery that the close-wide and inner-outer types of degeneracies are in fact limiting cases of a general theory, presented as the *offset* degeneracy in Chapter 4. In retrospect, inconsistencies with the close-wide and inner-outer degeneracies were already apparent in the three example NDE posteriors in Chapter 3, which are purposefully chosen to be representative of the diversity of degeneracies in the literature. In Figure 3.3, the two solution modes were stated to follow the $s \leftrightarrow 1/s$ relationship of the close-wide degeneracy (Chapter 3.4.1). However, a retrospective closer look reveals that $s_{\rm wide} > 1/s_{\rm close}$ instead, which is exactly expected from the offset degeneracy as the source trajectory passes to the right of the caustic. The initial NDE posterior example shown in the earliest appearance of Chapter 3 as a NeurIPS ML4PS workshop paper (Zhang et al. 2020) shared a similar deviation from $s \leftrightarrow 1/s$.

On the other hand, the example of Figure 3.4 has been intended to replicate OGLE-2018-BLG-0677 (Herrera-Martín et al. 2020) for which both solutions to the actual event have $s < 1$, allowing an interpretation with the inner-outer degeneracy. However, the actual test light curve used for Figure 3.4 was set to exactly $s = 1$ (as opposed to $s = 0.985$), for was chosen simply for convenience. For Figure 3.4, the caustic topologies of the two degenerate solutions are rather similar to OGLE-2018-BLG-0677 and thus I have also stated it as a manifestation of the inner-outer degeneracy, but this bi-modality have nearly already violated the premise of the inner-outer degeneracy, which was that the both solutions should be either in the close regime ($s < 1$) or in the wide regime ($s > 1$) (see the Appendix of Han et al. 2018). Had I picked another value ever slightly greater than $s = 1$ with the remaining parameters held constant, then perhaps this inconsistency would have already been impossible to miss in this earlier Chapter.

Specifics aside, I now discuss how the approach of AI-guided phenomenological study as illustrated in Figure 1.1 and exemplified by the offset-degeneracy discovery may be more broadly applicable to other sub-fields. Prior to designing observational experiments, one would often wish to first conduct feasibility studies to answer questions such as: to what extent could my experiment place meaningful constraints on the desired properties of the physical system of interest? However, the actual question that gets answered is more often: does the desired property of my physical system generate a detectable signal? These questions could often be interchangeable, but sometimes the second could not substitute for the first, especially in the presence of continuous and discrete degeneracies that are not well understood. The reason that the first question is cast into the second is precisely the non-invertibility of our physical forward models. Therefore, the use of NPE to develop surrogate inverse models may in fact be a useful tool for such phenomenological studies of broad

appeal.

I end this Chapter with an overview of this thesis. This thesis is composed of two components. The first component (Chapters 2 – 3) is the development of AI methodologies that led to the theoretical results of the second component (Chapters 4 – 6). The development of cyclic-permutation invariant neural networks is presented in Chapter 2, which is shown to achieve state-of-the-art performance for the classification of variable stars. A Likelihood-Free Inference framework for binary-lens microlensing is presented in Chapter 3, which partially utilizes the neural network architecture developed in Chapter 2.

In the second component, Chapter 4 reports on the discovery of the offset degeneracy. A follow-up analytical study is presented in Chapter 5, which also explored the generalization of the offset degeneracy to higher-order lenses. These theoretical insights then led to a careful reexamination of the seminal works for planetary microlensing published in the 1990s, particularly Gould & Loeb (1992), Gaudi & Gould (1997), and Dominik (1999), which led to the conclusion that the scope of the perturbative picture and the Chang-Refsdal lens approximation is much broader than laid out in these papers. In Chapter 6, I propose a generalized perturbative picture for planetary microlensing, which states that the planet can be considered to act as a variable-shear Chang-Refsdal lens on one of the images produced by the host star, leaving the other image largely unaffected. The analytic nature of the Chang-Refsdal lens indicates that the proposed formalism would allow full magnification maps of the planetary lens to be derived analytically, thereby allowing for the accelerated modeling of observed events.

Acknowledgments

I would like to thank Joshua Bloom, Jessica Lu, and Uroš Seljak for insightful discussions and comments on a draft of this Chapter.

Chapter 2

Cyclic-Permutation Invariant Neural Networks for Periodic Irregular Light Curves

Neural networks (NNs) have been shown to be competitive against state-of-the-art feature engineering and random forest (RF) classification of periodic variable stars. Although previous work utilising NNs commonly operated on period-folded light-curves, no approach to date has taken advantage of the fact that network predictions should be invariant to the initial phase of the period-folded sequence. Initial phase is exogenous to the physical origin of the variability and should thus be immaterial to the downstream application. Here, we present cyclic-permutation invariant networks, a novel class of NNs for which the output is invariant to phase shifts by construction. We implement this invariance by means of "Symmetry Padding." Across three different datasets of variable star light curves, we show that two implementations of the cyclic-permutation invariant network: the iTCN and the iResNet, consistently outperform non-invariant baselines and reduce overall error rates by between 4% to 22%. Over a 10-class OGLE-III sample, the iTCN/iResNet achieves an average per-class accuracy of 93.4%/93.3%, compared to RNN/RF accuracies of 70.5%/89.5% in a recent study using the same data. Finding improvement on a non-astronomy benchmark, we suggest that the methodology introduced here should also be applicable to a wide range of science domains where periodic data abounds due to physical symmetries.

2.1 Introduction

Periodic variability arises across the Hertzsprung-Russell diagram and manifest through stellar pulsation, rotation, and/or binarity. The identification of dozens of distinct phenomenological sub-classes (e.g., Gaia Collaboration et al. 2019) reflects the richness of the underlying physical processes giving rise to observable changes in brightness and colour. Periodic variables can also serve as precision probes of distance (Paczyński 1997), line-of-sight

dust extinction (Kunder et al. 2008), and Galactic structure (Skowron et al. 2019). As such, the systematic discovery and classification of periodic variables in large time-domain surveys, some with billions of stars monitored, remains paramount.

At scale, human expert labelling of variability catalogues of light curves has naturally, in the past decade, given way to automated classification approaches with machine learning. Random forest (RF; Breiman 2001) classification, while performant, requires computationally expensive, hand-crafted feature engineering as part of data preprocessing (Richards et al. 2011; Kim & Bailer-Jones 2016). More recently, deep representation learning has further pushed the boundaries by learning not only decision rules on features of raw data, but also the low-dimensional feature representation itself. This approach has advanced many fields in astronomy (e.g., Kim & Brunner 2017; Shallue & Vanderburg 2018; Agarwal et al. 2020; Zhang & Bloom 2020).

For variable star classification, both convolutional neural networks (CNNs; LeCun et al. 2015) and recurrent neural networks (RNNs; Hochreiter & Schmidhuber 1997; Cho et al. 2014) have been shown to be competitive to the traditional RF-based methods. Naul et al. (2018) used an RNN autoencoder network to learn low-dimensional representations of period-folded light-curves in an unsupervised fashion. This representation was then, in a supervised context, used as feature inputs to a RF classifier. They showed that the learned features are at least as good as, and often better than, two sets of state-of-the-art hand-crafted features (Richards et al. 2011; Kim & Bailer-Jones 2016), in terms of downstream classification accuracy. Becker et al. (2020) used an RNN for which instead of period-folding, each input light curve is grouped with a moving window of size 50 and stride 25. Although period-folding improves performance (Naul et al. 2018), Becker et al. (2020)'s time-space RNN does not require the period to be calculated, and is thus less computationally expensive in terms of preprocessing. Again, they found similar performance to a RF classifier with the Nun et al. (2015) features over three datasets, although lower accuracy was seen for many sub-classes with the OGLE dataset (Table 2.2; see Section 2.3.3 for data description). More recently, Jamal & Bloom (2020) systematically benchmarked the performance of different configurations of RNN and CNN network architectures on variable star classification. Aside from other work (e.g., Tsang & Schultz 2019; Aguirre et al. 2018) evaluating neural network performance retrospectively on previously labeled datasets, Dékány & Grebel (2020) used an RNN classifier to identify a new sample of fundamental-mode RR Lyrae (RRab) stars. Similarly, Dékány et al. (2019) found Classical and Type II Cepheids with a CNN classifier, also using the VISTA Variables in the Via Lactea (VVV) survey (Minniti et al. 2010) and using period-folded light curves.

While the compact phase-space (i.e., period-folded) light-curve representation has been widely adapted in the aforementioned studies, none of the neural networks used therein guarantees the same prediction under phase-shifts, or cyclic-permutations, of the same period-folded light curve. Since the initial-phase of the phase-space sequence is experimentally determined and exogenous to the physical origin of the variability, it could be seen as a nuisance parameter which should not affect classification. In the limit where a classification task is non-trivial—either due to the inherent difficulty of class separability or low signal-to-noise

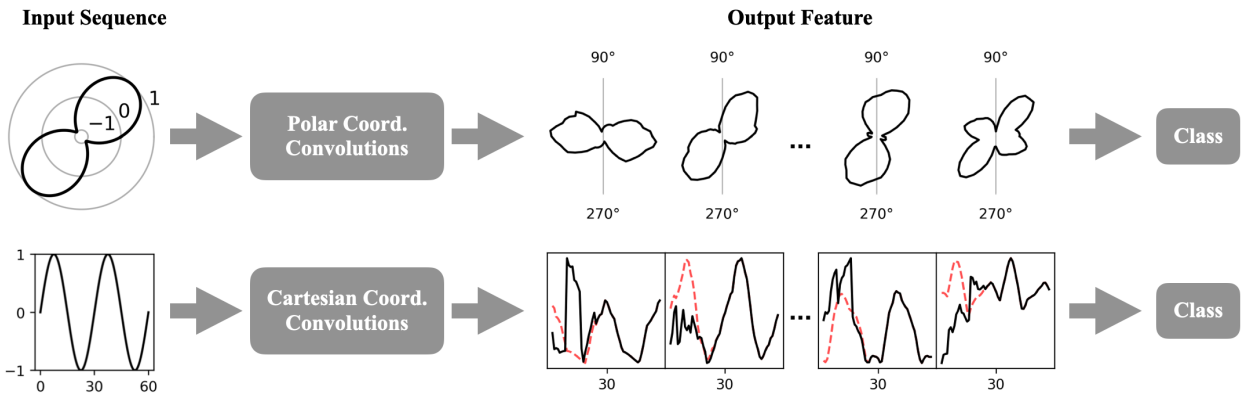


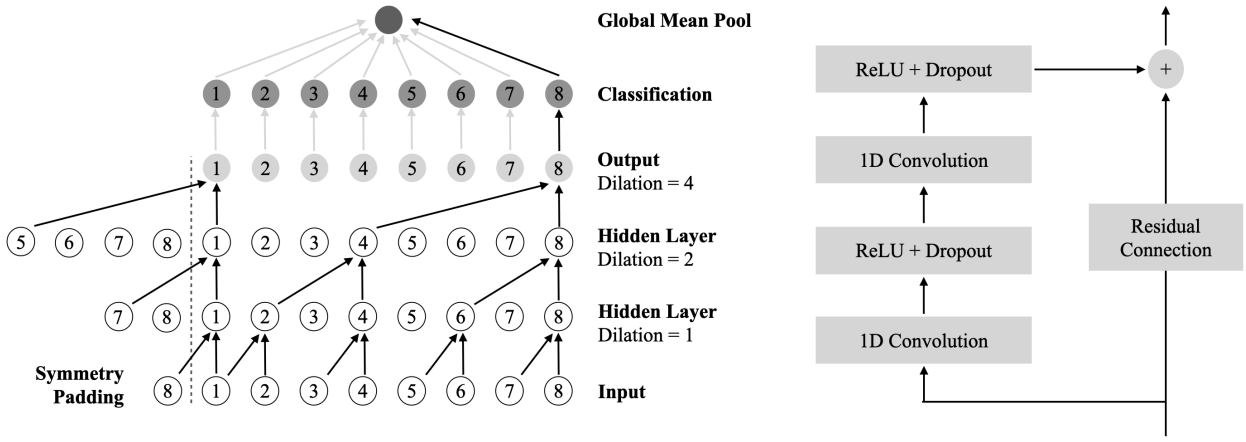
Figure 2.1: Schematic illustration of the effect of polar coordinate convolutions in preserving cyclic-permutation invariance. The input and output sequences are shown in polar coordinates for iTCN (top), and in Cartesian coordinates for TCN (bottom). The input sequence is a sine curve with two full oscillations in both cases. In the upper diagram, 1-D feature maps of the periodic input remains periodic; rotational symmetry is preserved. These periodic feature maps are also shown in Cartesian coordinates of the lower plots in red dashed lines for comparison. As demonstrated by the discrepancy, feature maps are distorted for the first full oscillation in the non-invariant network, which is shown in solid black lines.

data—some degree of domain knowledge can, in principle, be injected into the network architecture through known symmetries and conservation laws (Carleo et al. 2019; Mattheakis et al. 2020). Neural networks for computer vision tasks, for example, have been developed that are scale, rotation, and translation invariant (Jaderberg et al. 2015). Specialised networks for particle physics inference preserve known properties of quantum chromodynamics (Louppe et al. 2019). For periodic time series, we seek a network architecture with built-in invariance to cyclic-permutation to improve performance.

Here, we present cyclic-permutation invariant convolutional networks. We describe specific implementations with 1-D residual convolutional networks (ResNets), and with dilated 1-D convolutional networks (TCN) that have been shown to achieve state-of-the-art for a variety of sequence modeling tasks (Bai et al. 2018). The cyclic-permutation invariant network is descried in Section 2.2, whereas the variable star datasets used in benchmarking the invariant network against previous methods are discussed in Section 2.3. Finally, the performance of the invariant networks in various scenarios are discussed in Section 2.4. To facilitate applications of the cyclic-permutation invariant networks, we are releasing code at `https://github.com/kmzzhang/periodicnetwork`.

2.2 Method

The cyclic-permutation invariant networks that we introduce here refer to any neural network satisfying the following condition. Given an input sequence $\mathbf{x} \in \mathbb{R}^N$, a neural



(a) iTCN **(b)** Residual Block

Figure 2.2: (a) Simplified illustration of the cyclic-permutation invariant Temporal Convolutional Network (iTCN). Numbers refer to the ordering of the period-folded sequence. Dilated convolutions are represented by arrows where the dilation factor is indicated to the right of each layer. Gray arrows in the final two layers represent operations which are present only in the iTCN not the TCN. The classification layer consists of two convolutions of `kernel` size 1. (b) The residual block, which is the actual hidden layer used in the iTCN. Residual connections are to be replaced with $k = 1$ convolutions when consecutive layers have different hidden dimensions.

network $f : \mathbf{x} \rightarrow \mathbf{y}$ is invariant to cyclic-permutations if

$$\forall i \in [2, N], f(\mathbf{x}_{1:N}) = f(\text{concat}(\mathbf{x}_{i:N}, \mathbf{x}_{1:i-1})) \tag{2.1}$$

We first offer a high-level overview of cyclic-permutation invariant networks before discussing implementation details. Under the cyclic-permutation invariant network framework, the multi-cycle periodic time series is first period-folded into a single cycle by transforming from temporal space $(\mathbf{t}, \mathbf{m})$ into phase space $(\boldsymbol{\phi}, \mathbf{m})$: $\boldsymbol{\phi} = \mathbf{t} \mod p$, where m_i is the magnitude (or flux) measurement at phase ϕ_i and p, the period, is determined with periodogram analysis (Lomb 1976; Scargle 1982). The period is first used to fold the light-curve into phase-space and then concatenated to the output of the last convolution layer as an auxiliary input. We then make the observation that under polar coordinates, the period-folded sequence is essentially wrapped in a "closed ring" (Figure 2.1: Input Sequence, top row) where phase shifts simply become "rotations" which allows outputs to remain periodic (Figure 2.1: output feature). Phase-averaging the output feature map then results in a feature vector that is invariant to the initial phase (ϕ_0), rendering it a nuisance parameter. On the other hand, for the usual Cartesian-coordinate CNNs, phase shifts result in different input sequences and therefore different outputs. Polar coordinate convolution is implemented by replacing zero-padding of length (`kernel` size $-$ 1) in ordinary Cartesian-coordinate CNNs with "Symmetry Padding," which pads the input or hidden sequence not with zeros, but with the sequence itself (Figure 2.2a).

Based on the above framework, we present two particular implementations of the cyclic-permutation invariant network: the invariant Temporal Convolutional Network (iTCN) and the invariant Residual Convolutional Network (iResNet). The iTCN is based on the Temporal Convolutional Network (TCN; Bai et al. (2018)), and is composed of "residual blocks" (Figure 2.2b) of 1D dilated convolutions (Figure 2.2a), where the input to each "residual block' is concatenated to the output, creating a "gradient highway" for back-propagation, thus allowing for improved network optimization. Dilated convolutions refer to convolutions where the convolution `kernel` is applied over a region larger than the `kernel` size by skipping input values with a step of 2^{n-1} for the n-th layer. This dilation allows the network to achieve an exponential increase in the receptive field — the extent of input data accessible with respect to a particular output neuron — with network depth. The receptive field is calculated as

$$\mathcal{R} = (K-1) \times \sum_{n=1}^{D} 2 \times 2^{n-1} = (K-1) \times (2^{D+1} - 2), \tag{2.2}$$

where K is the kernel size, D the number of layers, 2^{n-1} the dilation factor for the n^{th} layer, and the additional factor of 2 due to the fact that each residual block consists of two dilated convolutions. Network depth is required to be large enough for the receptive field to be larger than the input sequence length, such that each feature vector in the output layer has complete information over the input sequence. Simultaneous predictions are then made for every possible initial phase of the input sequence (Figure 2.2a: "classification" layer) by first concatenating the period to each vector in the output layer, which serves as the feature vector for each phase. Each feature vector is then fed into a simple 2-layer feed-forward network that returns a vector with the same dimension as the number of classes. The outputs for the different phases are finally averaged with a `global mean pooling` layer as input to the `softmax` function for normalized class probabilities. By averaging predictions from all possible initial phases, the invariant network makes more robust predictions, as compared to non-invariant CNNs and RNNs, which can only predict for one particular initial phase with one network forward pass.

As a demonstration, for the toy iTCN network shown in Figure 2.2a, the last time-step of the output sequence (gray circle "8"; forth row bottom to top) is connected by arrows across the layers to the first time-step of the input sequence, and therefore has a receptive field of $\mathcal{R} = 8$. Applying a cyclic-permutation to the input sequence (e.g. 2, 3, 4, 5, 6, 7, 8, 1) would result in the same cyclic-permutation to the output sequence, which does not change the final classification because classification from each time-step is averaged, thus making the network invariant to such permutations.

To visualise the effects of cyclic-permutation invariance on modelling periodic sequences, we compare output sequences produced by the iTCN and the TCN in Figure 2.1. We create an iTCN and a TCN with the same weights and the same receptive field of $\mathcal{R} = 30$ at a network depth of 4 with kernel size 2. The input sequence is a length-60 sine function with two full oscillations (0 to 4π radian). As seen in the figure, while output feature maps produced by the iTCN remain symmetrical in polar coordinates, the first half of the output

sequence produced by the TCN is distorted by zero-padding, thus degrading the fidelity of output feature maps.

The second implementation, the iResNet, is also composed of stacks of "residual blocks," but is different from the iTCN in that the exponential receptive-field increase is achieved through `max pooling` layers, instead of dilated convolutions. A `max pooling` layer of `kernel` size 2 and `stride` 2, which combines every two adjacent feature vectors into one by selecting the maximum value, is added after every "residual block" to distil information extracted. After every such operation, the temporal dimension is reduced by half, while the number of hidden dimension is doubled until a specified upper limit. Unlike the iTCN, the feature vectors in the iResNet output layer do not have a one-to-one correspondence to the input sequence because the temporal dimension of the output feature map is reduced by a factor of 2^{D-1}, where D is the network depth. Because of this discreteness of featurisation, the iResNet is only invariant to phase shifts of 2^{D-1} steps (when the input sequence length is divisible by the same factor). Nevertheless, the iResNet potentially benefit from data augmentation of the input-sequence initial phase, as could non-invariant networks (see Appendix 2.8).

2.3 Benchmark Data

We assembled benchmarking datasets from three publicly available datasets of variable star light curves: All-Sky Automated Survey for Supernovae (ASAS-SN; Jayasinghe et al. 2018, 2019), Massive Compact Halo Object (MACHO; Alcock et al. 1996), and Optical Gravitational Lensing Experiment (OGLE-III; Udalski 2003). The datasets are described below.

2.3.1 All-Sky Automated Survey for Supernovae (ASAS-SN) Data

The ASAS-SN dataset consists of 282,795 light curves from eight classes of variable stars: 288 W Virginis ($p > 8$ day), 102 W Virginis ($p < 8$ day), 941 Classical Cepheids, 297 Classical Cepheids (Symmetrical), 1,631 Delta Scuti, 25,314 Detached Eclipsing Binaries, 12,601 Beta Lyrae, 43,151 W Ursae Majoris-type, 2,149 High Amplitude Delta Scuti, 9,623 Delta Scuti, 14046 Rotational Variables, 26,956 RR Lyrae type A/B, 7,469 RR Lyrae type C, 364 RR Lyrae type D, and 137,847 Semi-regular Variables. The class label of each variable star is classified by Jayasinghe et al. (2019) and only those with class probability greater than 99% are used. The maximum number of full light curve per class is capped at 20,000 to reduce the number of light curves of the dominant classes. Finally, segmenting into $L = 200$ chunks results in 106,005 fixed-length light curves.

2.3.2 Massive Compact Halo Object (MACHO) Project data

The MACHO dataset, taken directly from Naul et al. (2018), consists of 21,470 red band light curves from eight classes of variable stars: 7,403 RR Lyrae AB, 6,833 Eclipsing Binary,

3,049 Long-Period Variable Wood (sub-classes A–D were combined into a single super-class), 1,765 RR Lyrae C, 1,185 Cepheid Fundamental, 683 Cepheid First Overtone, 315 RR Lyrae E, and 237 RR Lyrae/GB Blend. Segmenting into $L = 200$ chunks has resulted in 80,668 fixed-length light curves.

2.3.3 Optical Gravitational Lensing Experiment: OGLE-III

The OGLE-III dataset is identical to that used in Becker et al. (2020), except for the selection of OSARGs. The OGLE-III data consists of 357,748 light curves from ten classes of variable stars: 6862 Eclipsing Contact Binaries, 21503 Eclipsing Detached Binaries, 9475 Eclipsing Semi-detached Binaries, 6090 Miras, 234,932 OGLE Small Amplitude Red Giants (OSARG), 25943 RR Lyrae type A/B, 7990 RR Lyrae type C, 34835 Semi-regular Variables, 7836 Classical Cepheids, 2822 Delta Scuti. Of the 234,932 OSARGs, 40,000 random ones are selected. Absent of a fixed random seed in their relevant code section, we have not been able to procure their exact selection, although the number selected is large enough for the difference to be small. Finally, segmenting into $L = 200$ chunks results in 540,457 fixed-length light-curves.

2.4 Results

We first study the evaluation metrics of the iTCN/iResNet architectures compared to the TCN/ResNet architectures to identify the gains made solely by cyclic-permutation invariance. Such "ablation" studies—applying a single change to the network architecture during training and testing to isolate the effect of that change—are common in deep learning. Input data is given in phase-space in all cases as the advantages compared to time-space have been demonstrated in previous studies (e.g. Naul et al. 2018). RNN baselines of GRU (Cho et al. 2014) and LSTM (Hochreiter & Schmidhuber 1997) are also included as additional baseline methods for comparison. We also note that cyclic-permutation invariance is forbidden in RNNs because of its acyclic topology.

The networks are trained on fixed-length light-curve segments ($L = 200$) from the three datasets described in Section 2.3. We apply randomised, stratified 60/20/20 train/validation/test splits for each dataset. To properly account for dataset boot-strapping noise—accuracy variations due to the particular choices of data splits—the same random splits are used to test every network, whose accuracies are compared pairwise in splits. We emphasise that the standard deviation of the accuracy differences, rather than the standard deviation of the accuracy themselves, should then serve as the basis for comparison of the accuracies. This is because accuracy variations for a given network and dataset are largely dominated by the boot-strapping noise due to train/test partitioning, and thus would be an overestimation of the variances solely attributed to the networks. Within each split, each full light-curve is divided into sequences of length 200 in temporal order and transformed into phase-space with respect to the period provided in the catalogs. Compared to random sampling, subdividing

Model	MACHO	OGLE-III	ASAS-SN
iTCN	**92.7%** $\pm$ 0.43%	**93.7%** $\pm$ 0.09%	**94.5%** $\pm$ 0.14%
TCN	92.0% $\pm$ 0.35%	92.9% $\pm$ 0.11%	93.0% $\pm$ 0.20%
diff [1]	$(-0.58\%^{+0.05\%}_{-0.17\%})$	$(-0.75\%^{+0.04\%}_{-0.02\%})$	$(-1.52\%^{+0.12\%}_{-0.09\%})$
iResNet*	92.6% $\pm$ 0.45%	**93.7%** $\pm$ 0.09%	93.9% $\pm$ 0.14%
ResNet	92.1% $\pm$ 0.34%	93.4% $\pm$ 0.11%	93.2% $\pm$ 0.22%
diff [1]	$(-0.48\%^{+0.05\%}_{-0.13\%})$	$(-0.25\%^{+0.03\%}_{-0.02\%})$	$(-0.64\%^{+0.01\%}_{-0.07\%})$
GRU	92.3% $\pm$ 0.37%	92.8% $\pm$ 0.19%	93.6% $\pm$ 0.42%
diff [2]	$(-0.34\%^{+0.05\%}_{-0.02\%})$	$(-0.86\%^{+0.17\%}_{-0.13\%})$	$(-0.71\%^{+0.07\%}_{-0.12\%})$
LSTM	91.7% $\pm$ 0.53%	92.6% $\pm$ 0.61%	93.5% $\pm$ 0.23%
diff [2]	$(-0.93\%^{+0.45\%}_{-0.16\%})$	$(-0.85\%^{+0.17\%}_{-0.48\%})$	$(-1.00\%^{+0.11\%}_{-0.06\%})$

Table 2.1: **Ablation study test accuracies demonstrating gains afforded by cyclic-permutation invariance**. The network with the top accuracy for each dataset is shown in bold. Test accuracies are the mean values for 8 different data splits. Median test accuracy differences of the different data partitions are shown in parentheses with the uncertainty interval corresponding to 1-σ range of test accuracy differences calculated pair-wise for the same random partitions of data. Negative accuracy differences indicate better performances of the invariant network.

[1]Compared to the invariant version of the same network.

[2]Compared to the best performing network.

*Semi-invariant due to use of discrete max-pooling layers.

by temporal order preserves the irregular samplings which resemble how data is accumulated. Each segment is then individually normalised (zero-mean and unit variance) while measurement times are rescaled by the period into phase ([0, 1]). The measurement phase intervals $\Delta\boldsymbol{\phi}$ between successive data points are fed together with the rescaled light curve as inputs to the network. The mean and standard deviation of each light curve segment, along with $\log(p)$, are concatenated to the network output layer as auxiliary inputs. We perform extensive hyperparameter optimisation for each pair of network and dataset (see Appendix 2.7).

As shown in Table 2.1, the improvements of the iTCN and the iResNet from their respective non-invariant baselines, as well as the RNNs, are significant by more than 5-σ in most cases, demonstrating the advantages of enforcing cyclic-permutation invariance. The improvements in classification accuracies correspond to reductions in overall error rates by between 4% to 22%, depending upon the non-invariant baseline and the dataset.

2.4.1 Comparison to published methods and results

We first consider the time-space RNN and RF results recently published in Becker et al. (2020). Becker et al. (2020) presents OGLE-III classification results with their time-space GRU and an RF baseline with the Nun et al. (2015) features. The Becker et al. time-space GRU work groups each full OGLE-III light curve with a moving window of size 50 and stride 25, whereby the effective sequence length is reduced by a factor of 25. This reduction alleviates the so-called vanishing gradient problem (Hochreiter & Schmidhuber 1997) which limits the sequence length that the RNN could be effectively trained on. To facilitate this comparison, we have used the same OGLE-III data selection as their work (Section 2.3.3). Since a $L > 300$ requirement has been applied to their OGLE-III data selection, we trained the iTCN/iResNet on $L = 300$ segments, and average classifications on $L = 300$ segments for each full light curve during testing. As seen in Table 2.2, the cyclic-permutation invariant networks outperform both results. The invariant network accuracies are significantly higher for most classes, reducing error rates by as much as 69% for the minority classes. We find this result to be critically important, as the hard-to-classify minority classes tend to be the least well-understood and often are the most interesting to identify for further study. In particular, the largest error rate reductions against RF are seen in Eclipsing Binaries, Delta Scuti, and Semi-Regular Variables, which are important both for accurate tests of stellar evolution models (e.g. Guinan et al. 2000; Torres & Ribas 2002) and for precision probes of distance (e.g. Bonanos et al. 2006; McNamara et al. 2007; North et al. 2012).

Additionally, Naul et al. (2018) published RF benchmark accuracies for the MACHO dataset of 90.50% with the Richards et al. (2011) features and 88.98% with the Kim & Bailer-Jones (2016) features. While we have use the same MACHO dataset as Naul et al. (2018), our results are not directly comparable because Naul et al. (2018) preformed randomised train/test split on the $L = 200$ segmented light curves, which have caused different versions of the same light curve to exist in both training and test split, resulting in information leakage and thus a higher accuracy.

Class	iTCN	iResNet	GRU	RF
Cep	**98.3%**±0.3%	**98.4%**±0.7%	72%	97%
RRab	**99.7%**±0.1%	**99.7%**±0.4%	85%	99%
RRc	**99.0%**±0.2%	**99.1%**±0.1%	30%	98%
Dsct	**97.6%**±0.8%	**97.8%**±0.6%	72%	93%
EC	**87.9%**±0.9%	**87.8%**±0.7%	54%	79%
ED	**95.0%**±0.3%	**94.8%**±0.4%	93%	92%
ESD	68.7%±1.0%	**70.7%**±0.9%	24%	61%
Mira	**97.1%**±0.6%	**96.8%**±0.3%	92%	**97%**
SRV	**96.0%**±0.4%	**95.9%**±0.2%	93%	82%
OSARG	93.2%±0.4%	93.4%±0.2%	90%	**97%**
Mean	**93.4%**	**93.3%**	70.5%	89.5%

Table 2.2: **Test accuracies for OGLE-III full-length light curves compared to classifications results in** Becker et al. (2020). For all but one subclass, the cyclic-permutation invariant networks outperform previous results. Similar to Table 2.1, we note that uncertainties are dominated by the bootstrapping noise arising from randomized data partitioning, and as such, are only upper limits to uncertainties in the accuracy differences for each class.

2.4.2 Adapting to variable-length sequences

Although none of the networks tested are restricted to fixed-length inputs, we emphasise that fixed-length sequence trained networks should not be naively applied to test sequences of different lengths because doing so results in degraded accuracy: different sequence lengths correspond to a different effective sampling frequency in phase space. The neural network is essentially asked to extrapolate, not interpolate, beyond the training function domain.

In Table 2.2, we showed a segment-and-classify scheme which is shown to be effective for OGLE-III full light curves. Here, we provide examples to show how the invariant networks could be directly trained on variable length sequences. A random sequence length in between $16 < L < 200$ is selected for each mini-batch during training. The optimal hyper-parameters for $L = 200$ networks are used, though each network could potentially benefit from increased complexity due to the increased task difficulty. As seen in Figure 2.3, high accuracy is maintained across a wide range of sequence lengths within the training range of $16 < L < 200$. Beyond the training range $16 < L < 200$, the ability of the networks to generalise is dataset dependent.

Furthermore, we note that the optimal range of training sequence length depends on the ratio of the period to the cadence. If the cadence is short compared to the periods, then the training sequence length should have a longer upper limit for each training light curve to cover at least one oscillation period. Figure 2.3 also suggests a way by which the training sequence length upper limit could be determined. As accuracy only increase marginally for MACHO beyond $L \sim 100$, a shorter upper limit could be selected whereby each full-length light curves is cut into more segments whose results are combined. On the other hand, the training sequence length upper limit could be increased for ASAS-SN, as classification

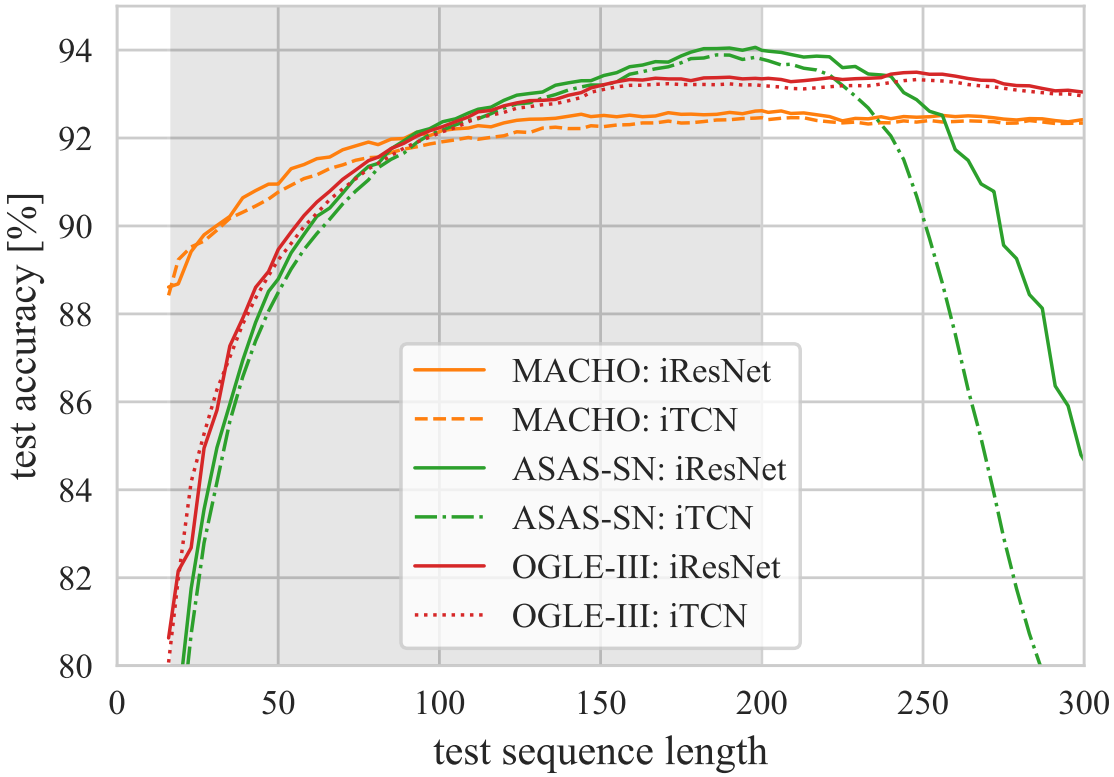


Figure 2.3: iResNet/iTCN test accuracy as a function of test sequence length for MACHO, ASAS-SN, and OGLE-III. Shaded region indicates the range of training sequence length: $16 < L < 200$

accuracy is still on the rise at $L = 200$, which suggests that the networks are still gaining additional information with increasing sequence length near the $L = 200$ cutoff.

2.5 PP-MNIST: Periodic Permuted MNIST

To examine the effectiveness of the cyclic-permutation invariant networks for classification tasks in other domains, we have created an additional benchmarking dataset, "periodic permuted MNIST" (hereafter PP-MNIST), which is derived from the "sequential MNIST" and "permuted MNIST" classification tasks (Figure 2.4). MNIST is a classic image dataset (LeCun et al. 1998) consisting of 70,000 28×28 images of hand-written digits in 10 classes (0 to 9). Under the sequential MNIST task, the 2D MNIST images are unwrapped into a 1D sequence of $L = 784$. Sequential MNIST is frequently used to test a recurrent network's ability to retain long-range information (Le et al. 2015; Wisdom et al. 2016; Krueger et al. 2017). For the more challenging permuted MNIST (P-MNIST) task, a fixed random per-

Table 2.3: Periodic permuted MNIST (PP-MNIST) classification accuracies.

iResNet	**96.0%**	iTCN	**94.8%**
ResNet	95.1%	TCN	77.4%

mutation is applied to each sequence (Le et al. 2015; Wisdom et al. 2016; Krueger et al. 2017; Arjovsky et al. 2016) so that any spatial/temporal structure is removed. It has been shown in Bai et al. (2018) that TCNs outperform RNN baselines for both sequential MNIST and P-MNIST. Here, we introduce periodicity to P-MNIST by introducing a random cyclic-permutation to each P-MNIST sequence. Because any of the 784 locations could be the zero index after permutation, only the relative, cyclic ordering of the sequence remains meaningful. Just as the case of periodic variable star classification, doing so essentially wraps each P-MNIST sequence in a ring whereby the "initial phase" of the sequence becomes a nuisance parameter and is no longer relevant for the classification.

We test the iResNet and the iTCN against their non-invariant counterparts to show improvements enabled by cyclic-permutation invariance. A hyper-parameter search is done for iResNet/ResNet over depth (9, 10), initial hidden dimension (24, 48), maximum hidden dimension (120, 200), and for iTCN/TCN over depth (8, 9), kernel size (3, 7), and hidden dimension (24, 48, 96). We present the PP-MNIST test accuracies in Table 2.3. Both invariant networks outperform their non-invariant counterparts, especially in the case of the iTCN/TCN. The poor performance of TCN can be partially attributed to the exceptionally large (784) number of possible initial phases for each sequence, four times more than the $L = 200$ sequences for periodic variable star classification. On the other hand, the regular ResNet performed relatively well. This is not surprising as the ResNet is by design different from the TCN — the ResNet is based on localised feature extraction where features are condensed through `pooling` layers, but the TCN is a sequential model subject to the causal condition, which requires it to memorize features extracted in temporal order.

2.6 Conclusions

Large scale time-domain surveys have both generated the need for, and enabled the training of, effective data-driven classification techniques for both periodic and non-periodic variable sources. In this work, as in other fields with established benchmark datasets, we have decoupled methodology from data and shown that the cyclic-permutation invariant networks achieve state-of-the-art accuracies for periodic variable star classification on datasets previously acquired. While the networks perform well on light curves with few data-points, we did not test the efficacy of such networks in a streaming context, where the period is not known *a priori*. Future work could explore how the invariant networks can be used in a streaming context, as well as efficient neural and non-neural ML methods for non-periodic data (Tachibana et al. 2020; Möller & de Boissière 2020; Narayan et al. 2018), which when

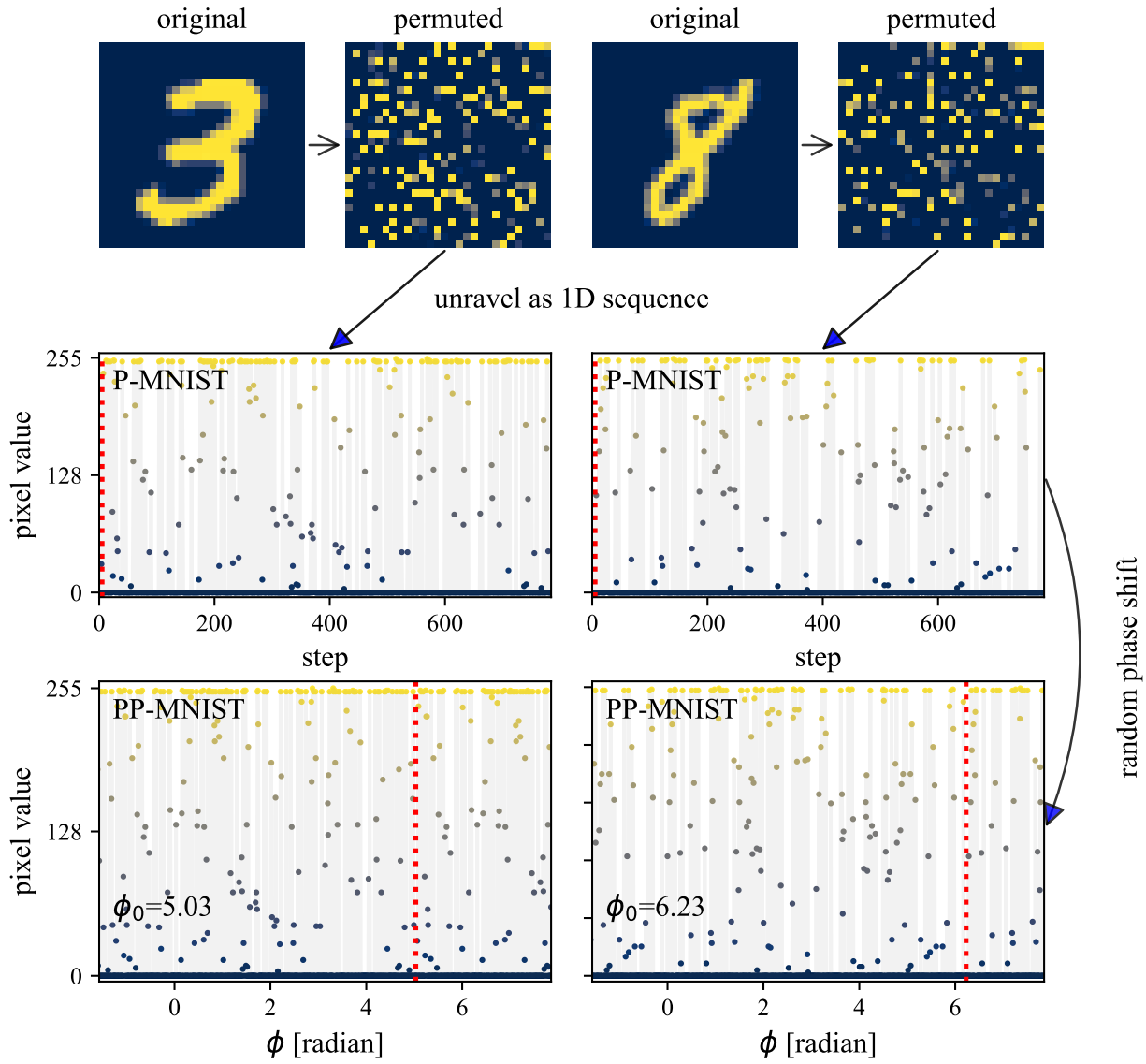


Figure 2.4: Construction of the PP-MNIST experiment. Pixel value is color coded with blue (0) transitioning to yellow (255). (top) The original 28×28 MNIST image and the same image with a fixed pixel order permutation. (middle) The P-MNIST 1D sequence, where the red vertical dashed lines indicate that no phase-shift is applied. (bottom) The PP-MNIST 1D sequence, which is the same sequence as the middle row, but with a phase-shift. The vertical red dashed lines indicate the initial-phase of the PP-MNIST sequence, whose numerical value is indicated in the bottom left corners.

combined with the methodology for periodic sources introduced here, can serve as the basis of a generalised classification framework for modern time-domain surveys.

2.7 Neural Network Hyper-Parameter Optimization

We search for optimal hyper-parameters independently for each network and for each dataset with the validation set in a fixed train/validation/test split. For all networks, among possible combinations of input features: phase interval ($\Delta\boldsymbol{\phi}$), magnitude ($\mathbf{m}$), magnitude change ($\Delta\mathbf{m}$), and gradient ($\Delta\mathbf{m}/\Delta\boldsymbol{\phi}$), we find the combination of ($\Delta\boldsymbol{\phi}, \mathbf{m}$) to yield the highest validation accuracy. For iTCN/TCN, we perform a hyper-parameter search over network depth (6, 7), hidden dimension (12, 24, 48), dropout (0, 0.15, 0.25), and kernel size (iTCN/TCN: 2, 3, 5). For iResNet/ResNet, we perform a grid search over initial hidden dimension (16, 32), maximum hidden dimension (32, 64), network depth (4, 5, 6), and kernel size (3, 5, 7). For GRU/LSTM, we search over network depth (2, 3), hidden dimension (12, 24, 48), and dropout rate (0, 0.15, 0.25). We find that a dropout rate of 0.15 works best for both GRUs and LSTMs across all three datasets, while no dropout works best for all other networks.

All networks are trained with the ADAM optimiser (Kingma & Ba 2015) with initial learning rates of 0.005, which are scheduled to decrease by a factor of 0.1 when training loss does not decrease by 10% for 5 epochs. Models are saved at the best validation accuracy for testing.

2.8 Data augmentation

Both the semi-invariant iResNet and non-invariant baseline networks potentially benefit from data augmentation of the initial-phase during training. Using cyclic-permutations of the input sequence as training-time data augmentation, we trained iResNets and ResNets on the three datasets, after redoing hyperparameter optimisation. As seen in Table 2.4, classification accuracies of both the iResNet and the ResNet are increased in most cases; the iResNets still hold a statistically significant advantage over the ResNets.

Acknowledgments

A version of this chapter was published in the Monthly Notices of the Royal Astronomical Society as Zhang & Bloom (2021). An extended abstract of this Chapter appeared in the "Machine Learning for the Physical Sciences" workshop at the 2020 Neural Information Processing Systems (NeurIPS) Conference, and the "Fundamental Science in the Era of AI" workshop at the 2020 International Conference on Learning Representations (ICLR).

I thank Benny T. H. Tsang for assistance with ASAS-SN data and Jorge Martínez-Palomera for assistance with OGLE-III data. I thank Sara Jamal for comments on a draft of

Table 2.4: Classification accuracies for networks with and without data augmentation. Accuracies without data augmentation is identical to Table 2.1.

Model	MACHO	OGLE-III	ASAS-SN
without phase data-augmentation			
iResNet	$92.6\% \pm 0.45\%$	$\mathbf{93.7\%} \pm 0.09\%$	$93.9\% \pm 0.14\%$
ResNet	$92.1\% \pm 0.34\%$	$93.4\% \pm 0.11\%$	$93.2\% \pm 0.22\%$
diff	$(-0.48\%^{+0.05\%}_{-0.13\%})$	$(-0.25\%^{+0.03\%}_{-0.02\%})$	$(-0.64\%^{+0.01\%}_{-0.07\%})$
with phase data-augmentation			
iResNet	$\mathbf{92.9\%} \pm 0.33\%$	$\mathbf{93.7\%} \pm 0.11\%$	$\mathbf{94.4\%} \pm 0.18\%$
ResNet	$92.4\% \pm 0.29\%$	$93.5\% \pm 0.12\%$	$94.2\% \pm 0.23\%$
diff	$(-0.39\%^{+0.08\%}_{-0.11\%})$	$(-0.19\%^{+0.02\%}_{-0.04\%})$	$(-0.32\%^{+0.14\%}_{-0.01\%})$

the MNRAS manuscript, which have helped improve this chapter. The experiments in this chapter is partially enabled by the Amazon Web Services (AWS) Cloud Credits for Research program. During the writing of this chapter, I am supported by a Gordon and Betty Moore Foundation Data-Driven Discovery grant.

Chapter 3

Likelihood-Free Inference of Roman Binary Microlensing Events with Amortized Neural Posterior Estimation

Fast and automated inference of binary-lens, single-source (2L1S) microlensing events with sampling-based Bayesian algorithms (e.g., Markov Chain Monte Carlo; MCMC) is challenged on two fronts: high computational cost of likelihood evaluations with microlensing simulation codes, and a pathological parameter space where the negative-log-likelihood surface can contain a multitude of local minima that are narrow and deep. Analysis of 2L1S events usually involves grid searches over some parameters to locate approximate solutions as a prerequisite to posterior sampling, an expensive process that often requires human-in-the-loop domain expertise. As the next-generation, space-based microlensing survey with the *Roman Space Telescope* is expected to yield thousands of binary microlensing events, a new fast and automated method is desirable. Here, we present a likelihood-free inference (LFI) approach named amortized neural posterior estimation, where a neural density estimator (NDE) learns a surrogate posterior $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ as an observation-parametrized conditional probability distribution, from pre-computed simulations over the full prior space. Trained on 291,012 simulated *Roman*-like 2L1S simulations, the NDE produces accurate and precise posteriors within seconds for any observation within the prior support without requiring a domain expert in the loop, thus allowing for real-time and automated inference. We show that the NDE also captures expected posterior degeneracies. The NDE posterior could then be refined into the exact posterior with a downstream MCMC sampler with minimal burn-in steps.

3.1 Introduction

When the apparent trajectory of a foreground *lens* star passes close to a more distant *source* star, the gravitational field of the *lens* will perturb the light rays from the *source*

which results in a time-variable magnification. Such are single-lens, single-source (1L1S) microlensing events. Binary microlensing events occur when the *lens* is a system of two masses: either a binary star system or a star-planet configuration. Observation of such events provides a unique opportunity for exoplanet discovery as the planet-to-star mass ratio may be inferred from the light curve without having to detect light from the star-planet *lens* itself (see Gaudi 2010 for a review). A next-generation microlensing survey with the Roman Space Telescope (Spergel et al. 2015; hereafter *Roman*) is estimated to discover thousands of binary microlensing events over the duration of the 5-year mission span, many with planetary-mass companions (Penny et al. 2019), which is roughly an order of magnitude more than events previously discovered (see Gaudi 2012 for a review).

While single-lens microlensing events are described by a simple analytic expression, binary microlensing events require numerical forward models that are computationally expensive. In addition, binary microlensing light-curves exhibit extraordinary phenomenological diversity, owing to the different geometrical configurations for which magnification could take place. This translates to a parameter space for which the likelihood surface suffers from a multitude of local minima that are disconnected, narrow, and deep; this issue significantly hampers any attempt of direct sampling-based inference such as MCMC where the chains are initialized from a broad prior. As a result, binary microlensing events thus far have generally been analyzed on a case-by-case basis.

For some planetary-mass-ratio events, heuristics could be used to "read off" an approximate solution from the planetary anomaly in the light curve (Gaudi & Gould 1997; Gould & Loeb 1992). Khakpash et al. (2019) applied the heuristics described in Gaudi & Gould (1997) on simulated *Roman* light-curves and found that the projected binary separation can be recovered very well for low-mass-ratio events, and the binary mass-ratios within an order of magnitude for events with wide and close caustic topologies.

More generally, an expensive grid search is usually conducted over a subset of parameters to which the magnification pattern is hyper-sensitive: i.e., binary separation, mass ratio, and the source trajectory angle of approach (e.g. Herrera-Martín et al. (2020)). At each grid-point, the remaining parameters are searched for with simple Nelder-Mead optimization (Nelder & Mead 1965) or MCMC. The fixed-grid solutions are then used to seed full MCMC samplings to refine solutions and sample the posteriors. This status quo approach, which is both computationally expensive and requires domain expertise in the loop, thus presents a great challenge to analyze the thousands of binary microlensing events expected to be discovered by *Roman*.

Recent progress in deep learning provides a promising path for a solution. In particular, both Convolutional (CNN; LeCun et al. 2015) and Recurrent Neural Networks (RNN Hochreiter & Schmidhuber 1997; Cho et al. 2014) have emerged as powerful alternatives to feature engineering of astronomical time-series (e.g. Naul et al. 2018). Given sufficient training data, CNN/RNNs could learn to compress the "high-dimensional" raw observations into "low-dimensional" feature vectors—automatically learning to produce features that are useful for downstream tasks such as classification or regression. Vermaak (2003) applied a more basic form of the neural network — the multilayer perceptron (MLP) — to predict

for 2L1S parameters on simulated noise-free light-curves, and achieved a success rate of 68% when the MLP results were further refined with Nelder-Mead optimization (Nelder & Mead 1965). However, there remains a large gap between the proof-of-concept work of Vermaak (2003) and application to real data due to the omission of noise and restrictions in parameter space. Additionally, machine learning has also been previously applied to *discover* and *classify* microlensing events (Wyrzykowski et al. 2015; Godines et al. 2019; Mróz 2020).

In addition to advances in this "representation learning," neural networks have also enjoyed significant progress in modeling probability distributions, otherwise known as neural density estimation, where the fundamental task is learn distributions from samples of that distribution. Both autoregressive models (Germain et al. 2015; Oord et al. 2016) and flow-based models (Papamakarios et al. 2017; Dinh et al. 2017) are NDEs that are highly capable of modeling complicated and multi-modal distributions, which can not only evaluate probability densities, but also sample from that distribution. NDEs thus allow for flexible uncertainty quantification and degenerate solutions which were not possible in Vermaak (2003).

The advancement in feature learning and NDE has allowed for accelerated progress in the field of likelihood-free inference (LFI), also known as simulation-based inference (SBI), which has been motivated by inference problems without a tractable likelihood. LFI is an umbrella term that encompasses a wide range of inference algorithms that do not require explicit evaluation of the likelihood. Under our particular LFI approach called amortized neural posterior estimation, an NDE learns a surrogate posterior as an observation-parametrized conditional probability distribution, from pre-computed simulations over the full prior space. A "featurizer" neural network is employed to compress raw observation into a feature vector which parametrizes the NDE. Inference is amortized in that all of the computation cost of simulation is paid upfront—likelihood evaluation with the slow forward simulator is no longer required, thus allowing for fast inference. For other neural LFI instances, neural networks could learn the likelihood (Papamakarios et al. 2019) or the likelihood-ratio (Thomas et al. 2022) as surrogates to accelerate sampling-based inference algorithms like MCMC (see Cranmer et al. 2020 for an overview).

In this paper, we present a likelihood-free inference approach for binary microlensing where an NDE learns a surrogate posterior $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ as an observation-parametrized conditional distribution from $(\mathbf{x}^i, \boldsymbol{\theta}^i)$ samples of simulated microlensing light-curves with the associated microlensing parameters. After training, the NDE can automatically generate posterior samples for future observations effectively in real-time. Because of the speed and performance without supervision by domain experts, the approach we introduce here has great potential for batch inference tasks such as those posed by *Roman*. Our preliminary results were reported as an extended abstract in Zhang et al. (2020). The work herein supersedes and expands upon that work.

We first lay out our inference framework in Section 3.2. Training set construction under the context of *Roman* is discussed in Section 3.3. In Section 3.4, we demonstrate the ability of the NDE to capture degenerate solutions and also present a systematic evaluation of the NDE performance over a large number of test events. In Section 3.5, we suggest future directions including a potential addition of a down-stream MCMC algorithm to refine the

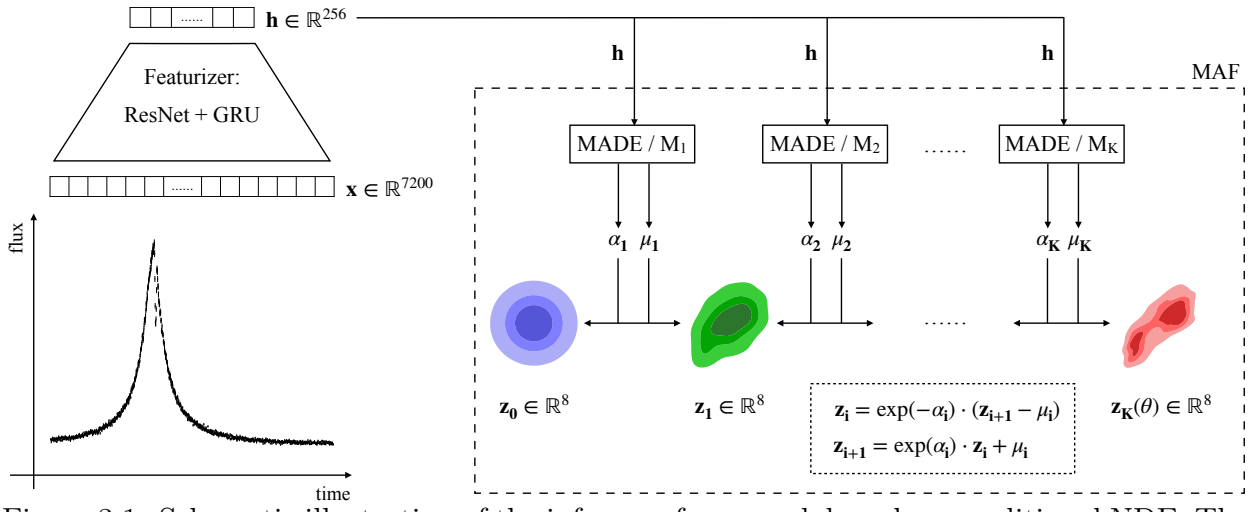


Figure 3.1: Schematic illustration of the inference framework based on conditional NDE. The bottom left shows a microlensing light-curve in arbitrary units which is abstracted into the length-7200 vector ($\mathbf{x}$) above. The featurizer composed of a combination of ResNet and GRU, shown in the trapezoid, compresses the light-curve into a low-dimensional feature vector $\mathbf{h}$. The masked autoregressive flow (MAF), composed of K blocks of masked autoencoder for density estimation (MADE), is shown in the dashed box. Each MADE block takes in the feature vector $\mathbf{h}$ and predicts scaling ($\boldsymbol{\alpha}$) and shifting ($\boldsymbol{\mu}$) factors, which parameterizes an invertible affine transformation between adjacent random variables (e.g., $\mathbf{z}_0$ and $\mathbf{z}_1$) shown in the dotted box. The left-most random variable is the mixture-of-Gaussian base distribution whereas the right-most random variable ($\mathbf{z}_K$) is the posterior ($\boldsymbol{\theta}$).

NDE posterior into the exact posterior, with minimal additional computation time.

3.2 Method

NDEs are neural networks that are capable of learning distributions from samples. We train an NDE to learn a surrogate posterior $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ as an observation-parametrized conditional distribution from $(\mathbf{x}^i, \boldsymbol{\theta}^i)$ samples of simulated microlensing light-curves, where $\boldsymbol{\theta}^i$ are the physical parameters and $\mathbf{x}^i \in \mathbb{R}^{\mathrm{N}}$ is the light curve with N data-points. The training objective is to minimize the Kullback–Leibler (KL) divergence (D_{KL}), or relative entropy, which is a measure of how one probability distribution (Q) is different from a reference probability distribution (P):

$$D_{\mathrm{KL}}(P||Q) = \mathbb{E}_{x\sim p(x)}\left[\log\left(\frac{p(x)}{q(x)}\right)\right] \tag{3.1}$$

In this case, we would like to minimize the KL divergence from the NDE surrogate posterior $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ to the true posterior $p(\boldsymbol{\theta}|\mathbf{x})$:

$$\begin{aligned}\phi &= \text{argmin}(D_{\text{KL}}(p(\boldsymbol{\theta}|\mathbf{x}))||\hat{p}_\phi(\boldsymbol{\theta}|\mathbf{x})))\\ &= \text{argmin}(\mathbb{E}_{\boldsymbol{\theta}\sim p(\boldsymbol{\theta}),\mathbf{x}\sim p(\mathbf{x}|\boldsymbol{\theta})}[\log(p(\boldsymbol{\theta}|\mathbf{x})) - \log(\hat{p}_\phi(\boldsymbol{\theta}|\mathbf{x}))])\\ &= \text{argmax}(\mathbb{E}_{\boldsymbol{\theta}\sim p(\boldsymbol{\theta}),\mathbf{x}\sim p(\mathbf{x}|\boldsymbol{\theta})}[\hat{p}_\phi(\boldsymbol{\theta}|\mathbf{x})]),\end{aligned} \tag{3.2}$$

where ϕ represents the neural network parameter, and $\mathbb{E}$ denotes the mathematical expectation over the specified distribution.

In light of Equation 3.2, the NDE is therefore trained through Maximum Likelihood Estimation (MLE) on a training set with physical parameters drawn from the prior $p(\boldsymbol{\theta})$ and light-curves drawn from the likelihood function, which is the Poisson measurement noise model on top of the noise-free microlensing light curve $g(\boldsymbol{\theta})$ (in the number of photons) which, for simplicity, is assumed to be in the Gaussian limit:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \mathcal{N}\left(\mu = g(\boldsymbol{\theta}), \sigma = \sqrt{g(\boldsymbol{\theta})}\right). \tag{3.3}$$

The noise-free light-curve, in turn, is determined by the baseline *source* flux (F_{source}), the magnification time-series produced by the microlensing physical forward model $A(\boldsymbol{\theta})$, and the constant *blend* flux, which is the flux from the *lens* star and any other star that is unresolved from the source star:

$$g(\boldsymbol{\theta}) = A(\boldsymbol{\theta}) \cdot F_{\text{source}} + F_{\text{blend}}. \tag{3.4}$$

We use a 20-block Masked Autoregressive Flow (MAF) (Papamakarios et al. 2017) to model $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$, and a ResNet-GRU network to extract features ($\mathbf{h}$) from the light curve ($\mathbf{x}$). We do not distinguish between $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ and $\hat{p}(\boldsymbol{\theta}|\mathbf{h})$ where the former is meant to refer to the "featurizer+NDE" model and the latter is meant to refer to the NDE model alone that is explicitly conditioned on $\mathbf{h}$. Figure 3.1 presents a diagram of our neural posterior estimation framework. The ResNet-GRU network is comprised of a 18-layer 1D ResNet (Residual Convolutional Network; He et al. 2016) and a 2-layer GRU (Gated Recurrent Network; Cho et al. 2014). We describe the neural networks in detail below.

3.2.1 Masked Autoregressive Flow

The masked autoregressive flow (MAF) belongs to a class of NDE called normalizing flows, which models the conditional distribution $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ as an invertible transformation f from a base distribution $\pi_z(\mathbf{z})$ to the target distribution $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$. The base density $\pi_u(\mathbf{z})$ is required to be fast to evaluate and is typically chosen to be either a standard Gaussian or a mixture of Gaussians for the MAF. The basic idea is that if the MAF, conditioned on the observation $\mathbf{x}$, could learn to map the posterior to a standard Gaussian, then the inverse transformation could enable sampling of the posterior by simply sampling from that standard Gaussian.

As binary microlensing events often exhibit degenerate, multi-modal solutions, we use a mixture of eight standard multivariate Gaussians, each with 8 dimensions, as the base distribution. The posterior probability density $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$ is evaluated by applying the inverse transformation f^{-1} from $\boldsymbol{\theta}$ to $\mathbf{z}$:

$$\hat{p}(\boldsymbol{\theta}|\mathbf{x}) = \pi_z\big(f^{-1}(\boldsymbol{\theta})\big) \left|\det\left(\frac{\partial f^{-1}}{\partial \boldsymbol{\theta}}\right)\right|, \tag{3.5}$$

where $\pi_z(f^{-1}(\boldsymbol{\theta}))$ represents the probability density for the base distribution (π_z) evaluated at $f^{-1}(\boldsymbol{\theta})$, while the second term—the determinant of the Jacobian—corresponds to the "compression" of probability space.

The MAF is built upon blocks of affine transformations where the scaling and shifting factors for each dimension are computed with a Masked Autoencoder for Distribution Estimation (MADE; Germain et al. 2015). For a simple 1-block case, the inverse transformation from $\boldsymbol{\theta}$ to $\mathbf{z}$ is expressed as:

$$z_i = (\theta_i - \mu_i) \cdot \exp\left(-\alpha_i\right), \tag{3.6}$$

In the above equation,

$$\mu_i = f_{\mu_i}(\boldsymbol{\theta}_{1:i-1}; \mathbf{x}) \tag{3.7}$$

$$\alpha_i = f_{\alpha_i}(\boldsymbol{\theta}_{1:i-1}; \mathbf{x}) \tag{3.8}$$

are the scaling and shifting factors modeled by MADE subject to the autoregressive condition that the transformation of any dimension can only depend on those prior to it according to a predetermined ordering. This allows the Jacobian of f^{-1} to be triangular, whose absolute determinant can be easily calculated as:

$$\left|\det\left(\frac{\partial f^{-1}}{\partial \boldsymbol{\theta}}\right)\right| = \exp\Big(-\sum\nolimits_i \alpha_i\Big), \tag{3.9}$$

where $\alpha_i = f_{\alpha_i}(\boldsymbol{\theta}_{1:i-1}; \mathbf{x})$.

To sample from the posterior, the forward transformation $\boldsymbol{\theta} = f(\mathbf{z})$ where $\mathbf{z} \sim \pi_z$ is applied:

$$\theta_i = z_i \cdot \exp \alpha_i + \mu_i, \tag{3.10}$$

where μ_i and α_i are computed in the same manner as the inverse transformation.

The MAF is built by stacking many such affine transformation blocks, $M_1, M_2, \ldots, M_K$, where M_K models the invertible transformation f_K between the posterior ($\mathbf{z}_K$) and intermediate random variable $\mathbf{z}_{K-1}$, M_{K-1} models that between intermediate random variables $\mathbf{z}_{K-1}$ and $\mathbf{z}_{K-2}$ and so on, and finally the base random variable $\mathbf{z}_0$ is modeled with the mixture-of-Gaussian distribution. M_1 also computes the mixture weights. The composite transformation can be written as $f = f_1 \circ f_2 \circ \ldots \circ f_K$ and the posterior probability density is now:

$$\hat{p}(\boldsymbol{\theta}|\mathbf{x}) = \pi_z\big(f^{-1}(\boldsymbol{\theta})\big) \prod_{i=1}^{K} \left|\det\left(\frac{\partial f_i^{-1}}{\partial \mathbf{z}_{i-1}}\right)\right| \tag{3.11}$$

where it is understood that $\mathbf{z}_K := \boldsymbol{\theta}$. The log-probability of the posterior is then, by Equation 3.9,

$$\begin{aligned}\log \hat{p}(\boldsymbol{\theta}|\mathbf{x}) &= \log[\pi_z\big(f^{-1}(\boldsymbol{\theta})\big)] + \sum_{i=1}^{K} \log \left| \det \left(\frac{\partial f_i^{-1}}{\partial z_{i-1}} \right) \right| \\ &= \log[\pi_z\big(f^{-1}(\boldsymbol{\theta})\big)] - \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_j^i,\end{aligned} \tag{3.12}$$

where α_j^i is the jth component of the scale factor in M_i, as in Equation 3.6. This serves as the optimization objective (see Section 3.3.3).

Autoregressive models are sensitive to the order of the variables. The original MAF paper uses the default order for the autoregressive layer closest to $\boldsymbol{\theta}$ and reverses the order for each successive layer. In this work, we adopt fixed random orderings for each MAF block which we find to allow for better expressibility. The random seed of the ordering serves as a hyper-parameter to be optimized on.

3.2.2 Featurizer Network

A custom 1D ResNet with a down-stream 2-layer GRU is used as the light curve featurizer which takes preprocessed light curves ($\mathbf{x}$) as input and outputs a low-dimensional feature vector ($\mathbf{h}$). The ResNet used in this study shares the identical architecture as Zhang & Bloom (2021) (Chapter 2; except for hyper-parameters) and consists of 9 identical residual blocks, each of which is composed of two convolutions followed by layer normalization (Ba et al. 2016). A residual connection is added between each adjacent residual block, which acts as a "gradient highway" to assist network optimization. A `MaxPool` layer is applied in between every two ResNet layers, where the sequence length is reduced by half and the feature dimension doubled until a specified maximum. This results in an output feature map of length $L = 56$ and dimension $D = 256$, when is then fed into the GRU network that sequentially processes information across the temporal dimension and outputs a single vector of $D = 256$ which then serves as the conditional input to the MAF.

3.3 Data

Training data is generated within the context of the Roman Space Telescope Cycle-7 design (see Penny et al. 2019). We first simulate 10^6 2L1S magnification sequences with the microlensing code `MulensModel` (Poleski & Yee 2019); each sequence contains 144 days at a cadence of 0.01 day, corresponding to the planned Roman cadence of 15 minutes (Penny et al. 2019). These sequences are chosen to have twice the length of the 72-day Roman observation window to facilitate sampling from a $t_0 \sim \text{Uniform}(0, 72)$ prior (see Section 3.3.1). We then

fit each simulated magnification time-series with a Paczyński single-lens-single-source (1L1S) model (assuming $S/N_{\rm base} = 200$ and $f_s = 1$; see Section 3.3.2) and discard those that are consistent with 1L1S ($\chi^2/{\rm dof} < 1$). This results in a final dataset of 291,012 light curves, among which 95% (276,461) are used as training set and the remaining 5% (14,551) as test set.

3.3.1 Prior

Assuming rectilinear relative motion of the observer, lens, and source, binary microlensing (2L1S) events are characterised by eight parameters: binary lens separation (s), mass ratio (q), angle of the source trajectory with respect to the projected binary lens axis (α), impact parameter (u_0), time of closest approach (t_0), Einstein ring crossing timescale (t_E), finite source size (ρ), and source flux fraction (f_s). α is the angle between the vector pointing from the primary to the secondary and the source trajectory vector, measured counterclockwise in degrees. u_0 and t_0 are defined with respect to the binary lens center-of-mass (COM). Where applicable, the parameters are normalized to the Einstein ring length-scale or the Einstein ring crossing time-scale of the total mass of the lens system. t_0 and t_E are in units of days. We simulate 2L1S events based on the following analytic priors:

$$\begin{aligned} s &\sim \text{LogUniform}(0.2, 5) \\ q &\sim \text{LogUniform}(10^{-6}, 1) \\ \alpha &\sim \text{Uniform}(0, 360) \\ u_0 &\sim \text{Uniform}(0, 2) \\ t_0 &\sim \text{Uniform}(0, 72) \\ t_E &\sim \text{TruncLogNorm}(1, 100, \mu = 10^{1.15}, \sigma = 10^{0.45}) \\ \rho &\sim \text{LogUniform}(10^{-4}, 10^{-2}) \\ f_s &\sim \text{LogUniform}(0.1, 1) \end{aligned} \tag{3.13}$$

We note that because of the $\chi^2_{1L1S}/{\rm dof} < 1$ cutoff, the effective prior is the parameter distribution for the 276,461 training set simulations, different from the prior above. As shown in Figure 3.2, large $\log q$ and small u_0, which otherwise have flat priors, are strongly preferred.

During training, a random 72-day segment is chosen on the fly from each 144-day magnification sequence, equivalent to prescribing a uniform prior on t_0. The truncated normal distribution for t_E is an approximation of a statistical analysis based on OGLE-IV data (Mróz et al. 2017). The lower limit of $q = 10^{-6}$ corresponds to the mass ratio between Mercury and a low-mass ($M \sim 0.1 M_\odot$) M-dwarf star, highlighting the superb sensitivity of Roman. The source flux fraction is defined as the ratio between the *source* flux and the total baseline flux

$$f_s = \frac{F_{\rm source}}{F_{\rm source} + F_{\rm blend}}. \tag{3.14}$$

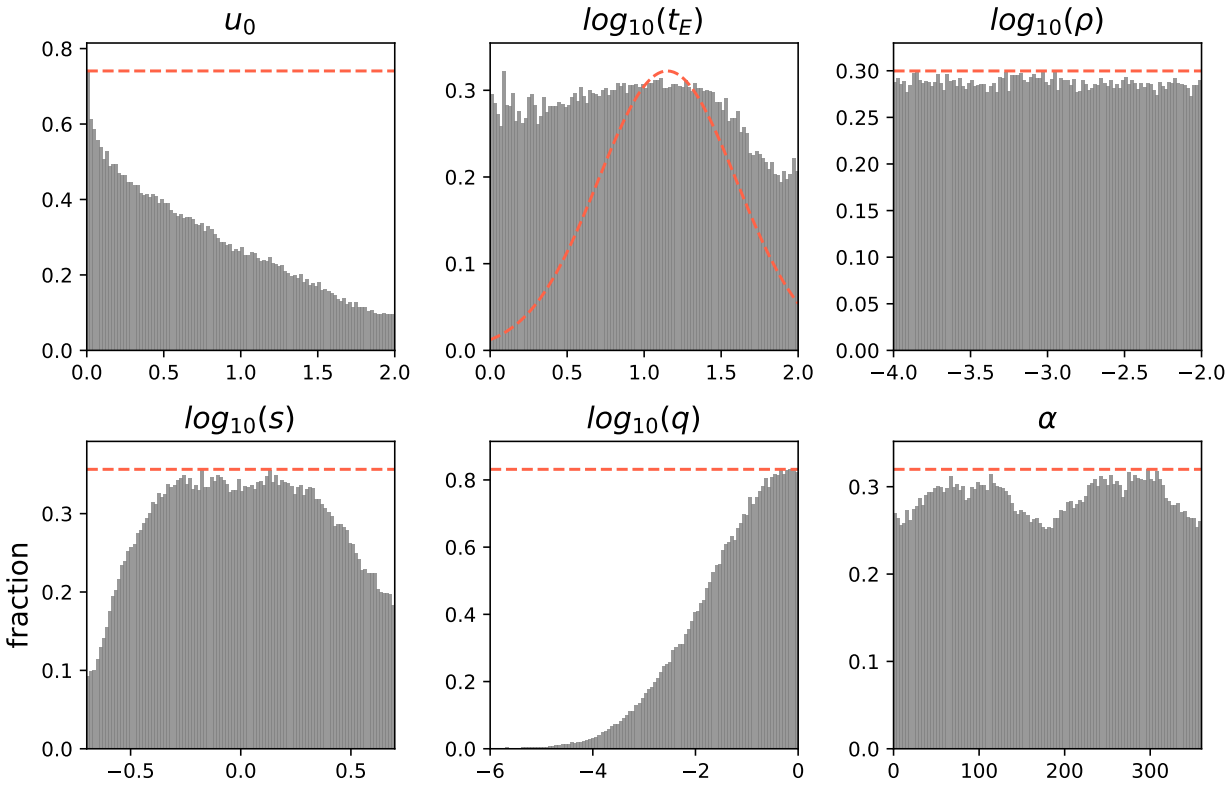


Figure 3.2: Fraction of the 10^6 simulations passing the $\chi^2_{1L1S}/\mathrm{dof} > 1$ cutoff as a function of each parameter, shown in the gray histograms. The original analytic priors used to generate the 10^6 simulations are shown in red-dashed lines up to a normalization factor. For parameters with a flat original prior, the gray histogram is also the effective training set prior up to a normalization factor. The t_0 and f_s distributions follow the original priors as they are sampled on the fly during training.

3.3.2 Light-curve realization

The magnification sequences are converted into light-curves during training on the fly by multiplying with the baseline pre-magnification *source* flux before adding the constant *blend* flux and applying measurement noise. For simplicity, we only consider photon-counting noise from the lens and fixed blend flux, assumed to be in the Gaussian limit of the Poisson noise (Equation 3.3), where the standard deviation of each photometric measurement is the square root of flux measurement in photon counts. Studies of the bulge star population show that the apparent magnitude largely lies within the range of 20 mag to 25 mag (Penny et al. 2019: Figure 5). The Roman/WFIRST Cycle 7 design has the zero-point magnitude (1 count/s) at 27.615 mag for the W149 filter. With exposure time at 46.8 s, the aforementioned magnitude range corresponds to signal-to-noise ($S/N_{\rm base}$) ratios between 230 and 23 for the baseline flux, which we randomly and uniformly sample during training. On-the-fly sampling of $S/N_{\rm base}$ and f_s also serves as data augmentation, which refers to the process of expanding the effective size of the training set.

3.3.3 Pre-processing and Training

Network optimization is performed with ADAM (Kingma & Ba 2015) at an initial learning rate of 0.001 and batch size 512, which decays to 0 according to a cosine annealing schedule (Loshchilov & Hutter 2017) for 250 epochs, at which point the training terminates. To ensure that there is no over-fitting, we first reserved 20% of the training set as a validation set. After confirming the absence of over-fitting, we then proceed with the full training set. We apply data augmentation on α by changing the direction of the source trajectory: the temporal order of each sequence is reverted and α becomes $-(\alpha+180) \mod 360$. Each training epoch takes ~ 6 minutes on four NVidia GTX 2080 Ti GPUs with a total training time of around 25 hours. As an evaluation metric, the final average negative log-likelihood (NLL) is -16.316 on the training set and -16.177 on the test set, where a lower value represents a better model fit to the data.

3.4 Results

The trained model is able to generate accurate and precise posterior samples at a rate of 10^5 per second on one GPU, effectively in real-time. This is much faster compared to the ~ 1 per second simulation speed of the forward model `MulensModel` on one CPU core. In this section, we first highlight the ability of the NDE to capture multi-modal solutions by providing NDE posteriors of representative events where we set the baseline $S/N_{\rm base} = 200$. Then, the quality of NDE posteriors is systematically analyzed by examining the accuracy and calibration properties on a test set of 14,511 simulated light-curves.

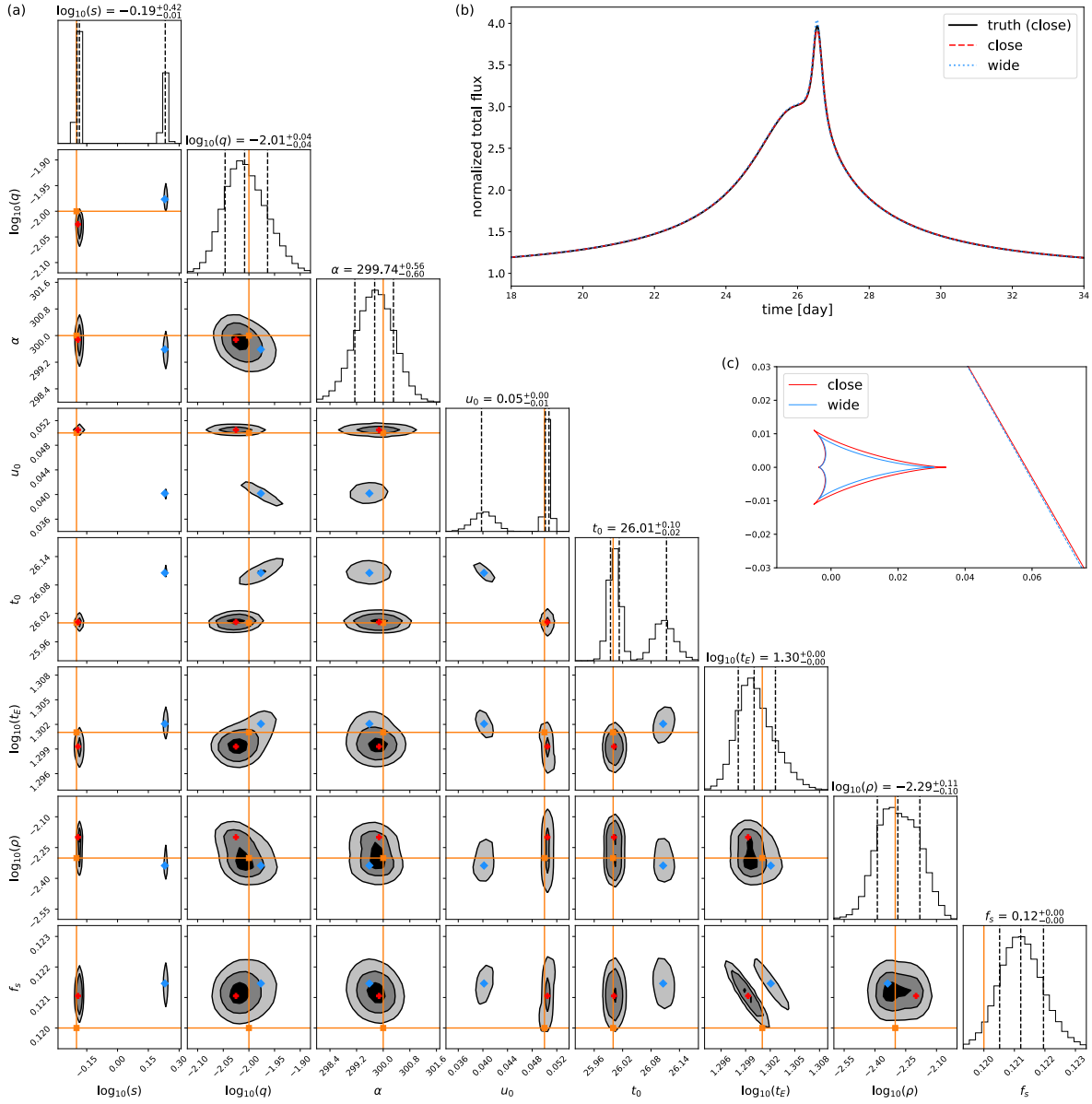


Figure 3.3: (a) NDE posterior for a central-caustic passing event. t_E and t_0 are in units of days, α in degrees, u_0, s, and ρ in units of θ_E. Filled contours show $1/2/3\sigma$ regions. The ground truth close solution is marked with orange cross-hairs. The close and wide solutions are marked with a red cross and a blue diamond, respectively. (b) Close-up view of the light-curve realizations normalized to the minimum fluxes for both solutions, in the same color-coding as the left panel. The 0.01 day cadence and measurement noise is negligibly small on the scale of the figure, and therefore not shown. (c) Caustic structures as well as trajectories for the two solutions in the same color-coding, centered on the center of caustic.

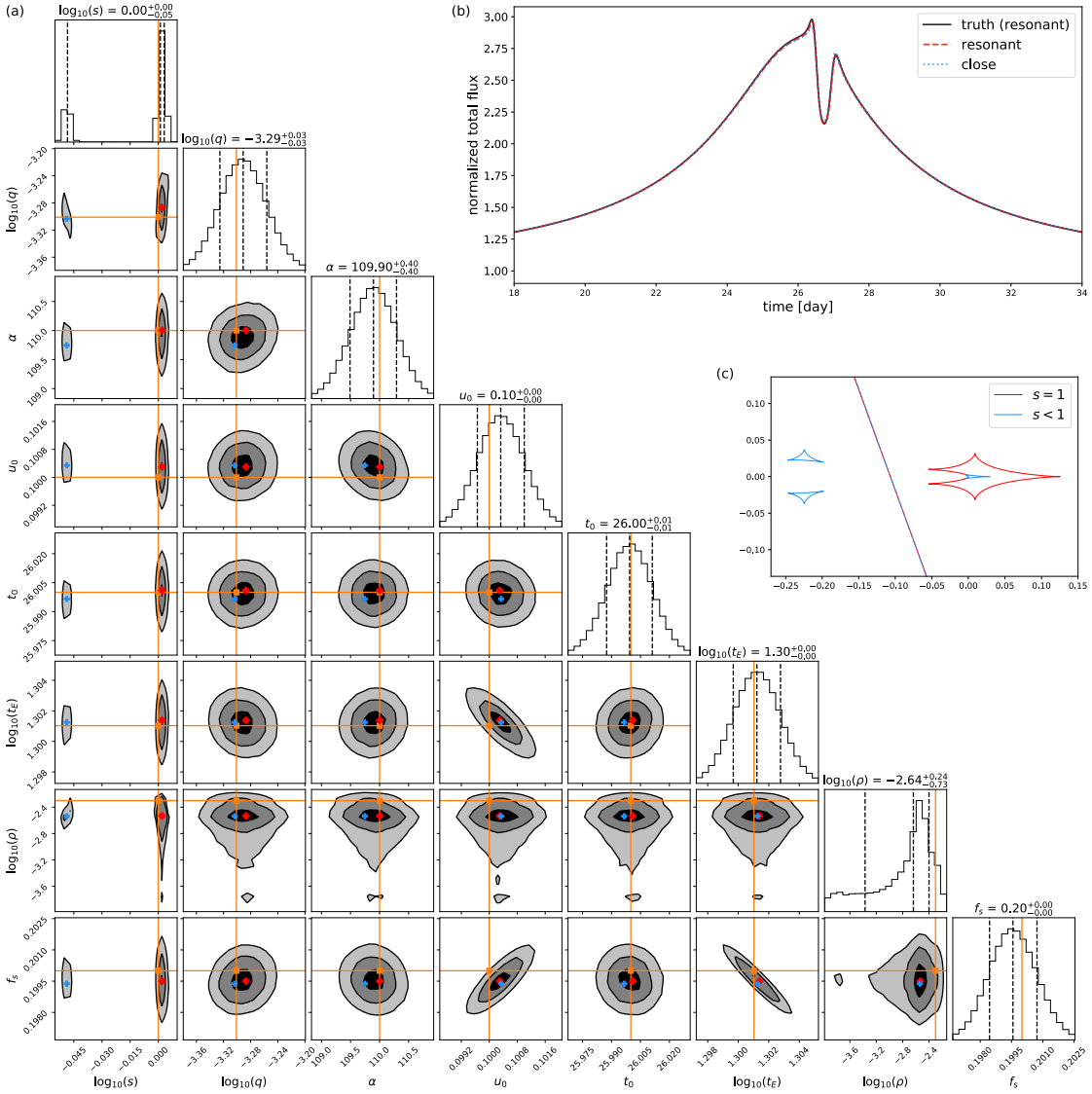


Figure 3.4: Resonant-caustic-passing event; same figure caption as Figure 3.3. Here, a degenerate solution is seen at $s < 1$, whose two triangular caustics cause a similar suppression pattern as the resonant caustic.

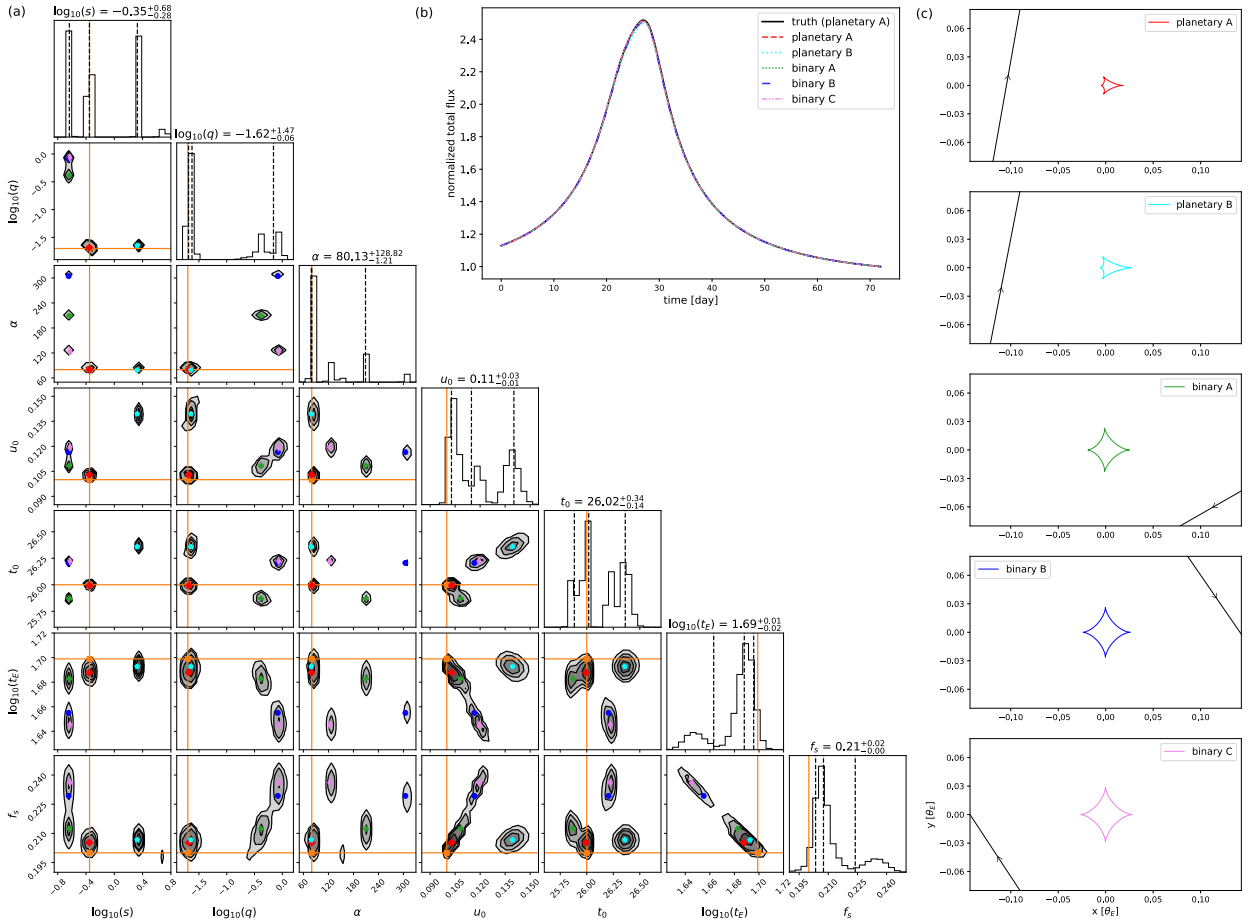


Figure 3.5: Example event exhibiting a blunt and flat light-curve near the peak, which has a 5-fold degenerate NDE posterior; same figure caption as Figure 3.3 for (a) and (b). (c) Caustic structures and source trajectories for the five solutions. The same color-coding is shared across the three panels.

Table 3.1: Solutions for the example central-caustic passing event. t_E and t_0 are in units of days, α in degrees, u_0, s, and ρ in units of θ_E. Uncertainties are 1σ marginal uncertainties.

	truth	close	wide
$\log_{10}(s)$	-0.200	$-0.1923^{+0.0034}_{-0.0036}$	$0.2301^{+0.0049}_{-0.0040}$
$\log_{10}(q)$	-2.000	$-2.0252^{+0.0144}_{-0.0147}$	$-1.9761^{+0.0177}_{-0.0155}$
α	300.000	$299.8524^{+0.2457}_{-0.2658}$	$299.5919^{+0.2469}_{-0.3139}$
u_0	0.050	$0.0505^{+0.0002}_{-0.0002}$	$0.0403^{+0.0007}_{-0.0010}$
t_0	26.000	$26.0027^{+0.0051}_{-0.0063}$	$26.1054^{+0.0131}_{-0.0075}$
$\log_{10}(t_E)$	1.301	$1.2993^{+0.0007}_{-0.0008}$	$1.3020^{+0.0010}_{-0.0009}$
$\log_{10}(\rho)$	-2.301	$-2.2075^{+-0.0044}_{-0.1133}$	$-2.3421^{+0.0737}_{-0.0210}$
f_s	0.120	$0.1211^{+0.0003}_{-0.0003}$	$0.1214^{+0.0004}_{-0.0003}$

Table 3.2: Solutions for the example resonant-caustic passing event. Same units as Table 3.1. Uncertainties are 1σ marginal uncertainties.

	truth	close	resonant
$\log_{10}(s)$	0.000	$-0.0484^{+0.0017}_{-0.0006}$	$0.0018^{+0.0010}_{-0.0003}$
$\log_{10}(q)$	-3.301	$-3.3008^{+0.0098}_{-0.0201}$	$-3.2858^{+0.0194}_{-0.0121}$
α	110.000	$109.7839^{+0.2044}_{-0.1526}$	$109.9666^{+0.1657}_{-0.2093}$
u_0	0.100	$0.1004^{+0.0003}_{-0.0003}$	$0.1003^{+0.0003}_{-0.0003}$
t_0	26.000	$25.9980^{+0.0048}_{-0.0064}$	$26.0014^{+0.0047}_{-0.0062}$
$\log_{10}(t_E)$	1.301	$1.3012^{+0.0007}_{-0.0008}$	$1.3012^{+0.0008}_{-0.0007}$
$\log_{10}(\rho)$	-2.301	$-2.5335^{+0.0492}_{-0.2505}$	$-2.5325^{+-0.0041}_{-0.4117}$
f_s	0.200	$0.1996^{+0.0006}_{-0.0005}$	$0.1995^{+0.0006}_{-0.0005}$

Table 3.3: Degenerate solutions for the binary-planetary degenerate event shown in Figure 3.5. Same units as Table 3.1. Uncertainties are 1σ marginal uncertainties.

	truth	planetary A	planetary B	binary A	binary B	binary C
$\log_{10}(s)$	-0.350	$-0.3520^{+0.0037}_{-0.0049}$	$0.3242^{+0.0035}_{-0.0035}$	$-0.6373^{+0.0037}_{-0.0047}$	$-0.6450^{+0.0026}_{-0.0030}$	$-0.6267^{+0.0046}_{-0.0043}$
$\log_{10}(q)$	-1.700	$-1.6849^{+0.0275}_{-0.0140}$	$-1.6464^{+0.0190}_{-0.0110}$	$-0.3729^{+0.0250}_{-0.0273}$	$-0.0813^{+0.0297}_{-0.0250}$	$-0.0609^{+0.0162}_{-0.0244}$
α	80.000	$80.0207^{+0.2170}_{-0.3531}$	$79.0411^{+0.2682}_{-0.3430}$	$209.7867^{+0.8607}_{-0.8053}$	$304.7107^{+0.6110}_{-0.6557}$	$123.4429^{+0.2513}_{-1.3218}$
u_0	0.100	$0.1027^{+0.0009}_{-0.0006}$	$0.1390^{+0.0022}_{-0.0016}$	$0.1082^{+0.0010}_{-0.0011}$	$0.1160^{+0.0012}_{-0.0011}$	$0.1197^{+0.0009}_{-0.0013}$
t_0	26.000	$25.9926^{+0.0084}_{-0.0070}$	$26.3604^{+0.0245}_{-0.0169}$	$25.8701^{+0.0089}_{-0.0080}$	$26.2091^{+0.0065}_{-0.0128}$	$26.2230^{+0.0102}_{-0.0101}$
$\log_{10}(t_E)$	1.699	$1.6874^{+0.0035}_{-0.0016}$	$1.6927^{+0.0024}_{-0.0022}$	$1.6824^{+0.0038}_{-0.0025}$	$1.6550^{+0.0037}_{-0.0030}$	$1.6443^{+0.0045}_{-0.0022}$
f_s	0.200	$0.2048^{+0.0018}_{-0.0010}$	$0.2065^{+0.0015}_{-0.0011}$	$0.2114^{+0.0027}_{-0.0014}$	$0.2280^{+0.0030}_{-0.0015}$	$0.2357^{+0.0023}_{-0.0021}$

3.4.1 Central-Caustic Passing Event

Figure 3.3a shows the NDE posterior for an example central-caustic-passing event where a classic "close-wide" degeneracy is clearly exhibited by the s-$1/s$ behavior (Griest & Safizadeh 1998; Dominik 1999). Table 3.1 presents the ground truth 2L1S parameters of this event as well as the "close" and "wide" solutions, calculated as the modes of their respective distributions. The fact that f_s is slightly underestimated is related to a systematic effect as discussed in Section 3.4.4. Although the source is expected to pass the caustic center at the same distances for the two cases, Figure 3.3a shows a bimodal solution for u_0 as well because u_0 has been defined with respect to the center-of-mass (COM), rather than the caustic center. While the caustic center is centered on the COM for close-separation events, for wide-separation events there is an offset from the COM of

$$\delta = \frac{s \cdot q}{1+q} - \frac{q}{s \cdot (1+q)} \tag{3.15}$$

where the first term accounts for the offset of the caustic center from the location of the primary (Han 2008), and the second term, the offset of the primary from the center of mass. Positive offsets are directed toward the companion and vice versa. Plugging in the wide solution, we expect an offset of $\Delta u_0 = 0.0116$, which is close to the actual $\Delta u_0 = 0.0099$. Magnification curves of the two solutions, as well as the ground truth are plotted in Figure 3.3b, which are hardly distinguishable from one another. Figure 3.3c shows the caustic structures of the two degenerate solutions.

3.4.2 Resonant-Caustic Passing Event

We also highlight an example of a resonant-caustic passing event, whose parameters and solutions are shown in Table 3.2. As illustrated in Figure 3.4, the NDE finds an additional solution at $s < 1$, whose triangular caustics are causing a similar weak de-magnification as the resonant caustics (also see Figure 7 in Gaudi 2010). This type of degeneracy has been previously observed in the microlensing event OGLE-2018-BLG-0677Lb (Herrera-Martín et al. 2020). Additionally, strong covariances are seen among u_0, t_E, and f_s, as is also seen in the previous example (Section 3.4.1). As first observed by Woźniak & Paczyński (1997), in the $f_s \ll 1$ and $u_0 \ll 1$ regime where the baseline flux is dominated by the blend flux, there is strong degeneracy between the three parameters for 1L1S events. While the binary perturbations break some of that degeneracy, strong covariances remain.

3.4.3 Binary-Planetary Degeneracy

We also provide a fascinating 5-fold-degenerate example that is similar to the degeneracy reported in Choi et al. (2012) where a light curve that is blunt and flat near the peak can be explained by either a binary case or a planetary case. Here, we simulate a close-topology, planetary mass ratio ($q = 10^{-1.7}$) event where the source trajectory passes through the negative perturbation region towards the back end of the arrowhead-shaped central caustic

as in the case of the "planetary A/B" caustic in Figure 3.5. Choi et al. (2012) noted that a similar perturbation can occur for the binary case when the source trajectory passes through the negative perturbation region between two adjacent cusps of the astroid-shaped central caustic, as in case of the "binary A/B/C" caustics in Figure 3.5; also see Figure 1 in their paper.

As shown in Figure 3.5, all five degenerate solutions cause magnification patterns that are hardly distinguishable. The two planetary solutions exhibit a close-wide degeneracy. For the three binary solutions, the "binary B/C" solutions suggest two possible trajectories ($\sim 90/270$ deg) for the same lens system configuration whereas "binary A" solution exhibits a smaller mass ratio and a wider binary separation than "binary B/C". We note that this additional degeneracy in the mass ratio for the binary case was not reported in Choi et al. (2012). It is not clear if this is a discrete or continuous degeneracy, nor if it is an "accidental degeneracy" that arises because of the relatively weak perturbation, or is due to some underlying symmetry in the binary lens equation (e.g. Dominik 1999).

On the other hand, wide solutions for the binary case are largely absent from the NDE posterior, apart from an inkling of density near $\log_{10}(s) \sim 0.69$ which points to the expected close-wide degeneracy for the binary solution. We note that the reason those degenerate solutions are excluded is that, because of the offset between the COM and the central caustic (Equation 3.15), wide-binary solutions would require $t_0 < 0$, which has a prior probability of zero.

3.4.4 Evaluating Performance

We present a systematic evaluation of all 14,551 test set events in the form of predicted vs. truth scatter plots (Figure 3.6). Each test event light-curve is realized in the same fashion as training time. As the NDE returns potentially multi-modal posteriors of arbitrary shape, we compute the mode(s) for the marginal 1D distributions of the posterior and consider the mode closest to the ground truth as the "predicted" value. The mode(s) is computed by first fitting each with a 1D histogram of 100 bins and then searching for local maxima defined as any bin count higher than that of the 20 adjacent bins. This limits the number of modes to 5. Considering the purpose of the NDE posterior is to allow ultra-fast convergence of a downstream sampling-based algorithm like MCMC to determine the exact posterior, as long as the correct solution has substantial density in the NDE posterior, it should not raise alarm if an alternative mode is mistakenly favored. Any degeneracies can be easily resolved downstream. Therefore, it is sensible to allow the correct mode to be used as the predicted value, even if another degenerate mode is incorrectly preferred.

As shown in the upper-left corner of each subplot in Figure 3.6, all parameters are constrained at a rate of close to 100% except for the finite source effect for which only 14.2% is constrained, as the source trajectory is required to either cross or pass close to a caustic for ρ to be determined.[1] We consider a parameter to be constrained if the probability density of

[1]Formally, effects on the light curve due to the finite size of the source are only significant if the gradient of the magnification across the source has a significant second derivative. In practice, this condition is only

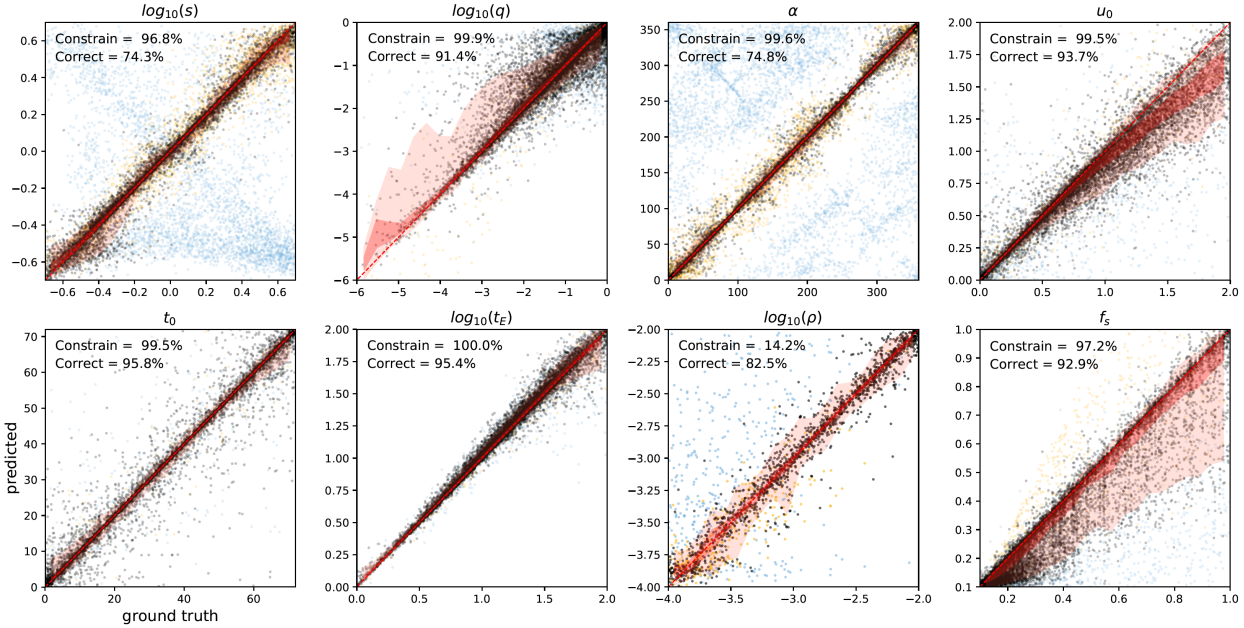


Figure 3.6: Predicted vs. ground truth 2L1S parameters for 14,551 test-set 2L1S events. t_E and t_0 are in units of days, α in degrees, u_0, s, and ρ in units of θ_E. Single-mode NDE posteriors are shown in black dots. For multi-model NDE posteriors, we color-code the solution as follows: those for which the global mode is closest to the ground truth are plotted in black; for cases where a minor mode is closest to the true value, this correct, minor mode is plotted in orange whereas the incorrect global mode is plotted in blue. Red shadows indicate 32–68th percentile (1σ) and 5–95 percentile (2σ) regions. Red-dashed lines show the diagonal. In the upper left of each subplot, "constrain" refers to the percentage of events whose NDE posterior poses sufficient constraint—the peak posterior probability must be at least twice the prior probability. "Correct" refers to the percentage of constrained events whose true parameter lies closest to the global mode.

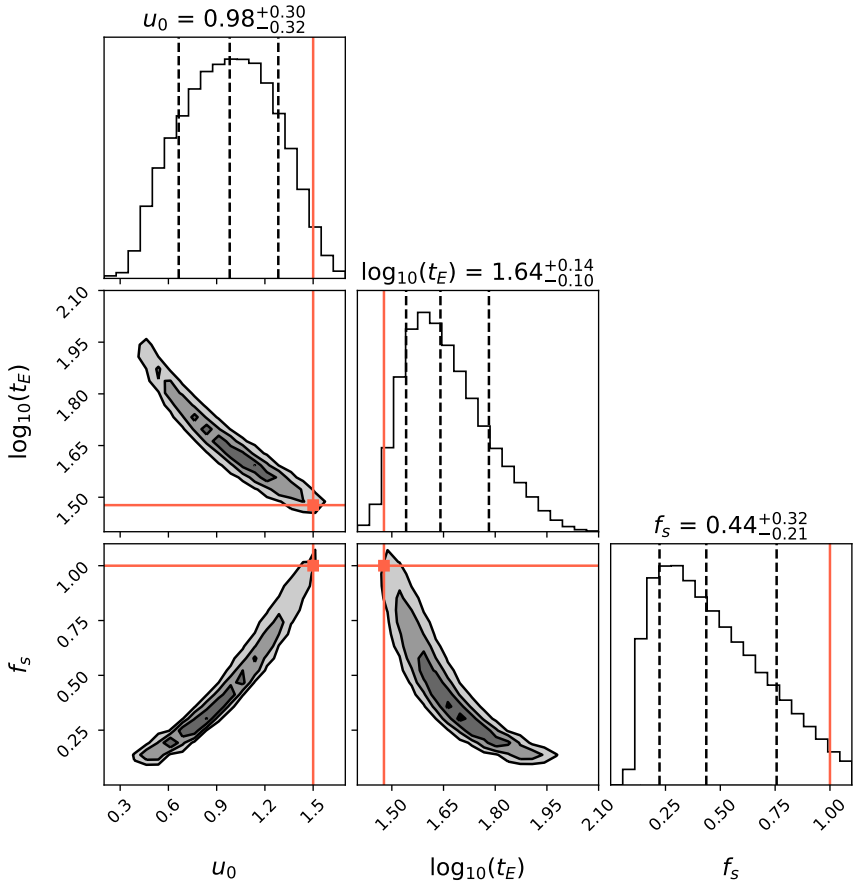


Figure 3.7: Corner plot for the marginal NDE posterior of an 1L1S event showing strong degeneracy among the three 1L1S parameters: u_0 in units of θ_E, t_E in units of days, and f_s. Filled contours show $1/2/3/4\sigma$ regions. Small u_0 and f_s are strongly favored because of the effective priors (Section 3.3.1) and a marginally informative likelihood.

the 1D marginal distribution is more than twice the prior probability density at the global mode.

The second row in each upper-left corner shows the frequency for which the correct mode is preferred by NDE, that is, the ground truth is the closest to the major mode compared to the minor modes(s), if any. If the ground truth is closer to a minor mode, the major mode is plotted in blue while the major mode is shown in orange. We see clear degeneracy patterns in $\log_{10}(s)$ and α. For $\log_{10}(s)$, the "wide-close" degeneracy is exhibited by the cluster around the upper-left to lower-right diagonal. For α, there is also a cluster of events along the same diagonal, indicating a degeneracy between α and $-\alpha$. Such a degeneracy may happen for nearly symmetrical central caustics along the direction perpendicular to the lens axis.

The 1-σ and 2-σ ranges of prediction, shown in red shadows, are clustered closely around the diagonal for most parameters. We emphasize that the loose 1L1S-fitting cut-off ($\chi^2_{1L1S}/\mathrm{dof} \sim 1$) means many of the test-set light-curves are only weakly perturbed by the binary nature of the lens, and should explain a number of cases in which the mass-ratio is poorly constrained. Interestingly, we find that there is a tendency to overestimate the mass ratio in these cases. In addition, we notice that u_0 and f_s are underestimated for a large number of cases while t_E is correspondingly overestimated, though hardly visible in Figure 3.6. This bias could be explained by the combined effect of a known degeneracy for 1L1S events and a distribution mismatch.

First, there exists a well-known degeneracy between u_0, f_s, and t_E for single-lens events which in our case, applies to events that are only weakly perturbed by the binary nature of the lens. As demonstrated by Woźniak & Paczyński (1997), this degeneracy is most severe for low magnification events ($u_0 \gg 1$), which is precisely where the biases occur as seen in Figure 3.6. Indeed, restricted to test events with $u_0 < 0.15$, the bias in f_s and t_E is largely removed. Figure 3.7 shows the NDE posterior for an example $u = 1.5$ 1L1S event which demonstrates the strong degeneracy among u_0, f_s, and t_E.

In the presence of strong degeneracies as such, the effective likelihood implicitly provided by the featurizer is only marginally informative. In other words, the featurizer cannot distinguish among solutions within the continuous degeneracy, and only prescribes a region in parameter space where the observation is about equally likely. Therefore, the posterior is essentially dominated by the prior, which strongly favors small u_0 and f_s, as seen in Figure 3.7. Had the parameters for the weakly perturbed events in the test-set been drawn from the same effective prior as the full training set, there would be little bias (under/over-estimation) at all in Figure 3.6. However, quite the contrary, the distribution of the weakly-perturbed is weighted towards the exact opposite direction of effective prior, e.g., towards large u_0 and small $\log_{10}(q)$—those more likely to be excluded from the $\chi^2_{1L1S}/\mathrm{dof} > 1$ cutoff. Because of this distribution mismatch, large u_0 and small $\log_{10}(q)$ occur much more often than expected by prior belief, thus resulting in the under/overestimation bias. And because of the strong covariances among u_0, f_s, and t_E, the under-estimation of u_0 translates into an under-estimation of f_s and an over-estimation for t_E (Figure 3.7), which explains the biases

satisfied if the source passes within a few angular source radii of a caustic.

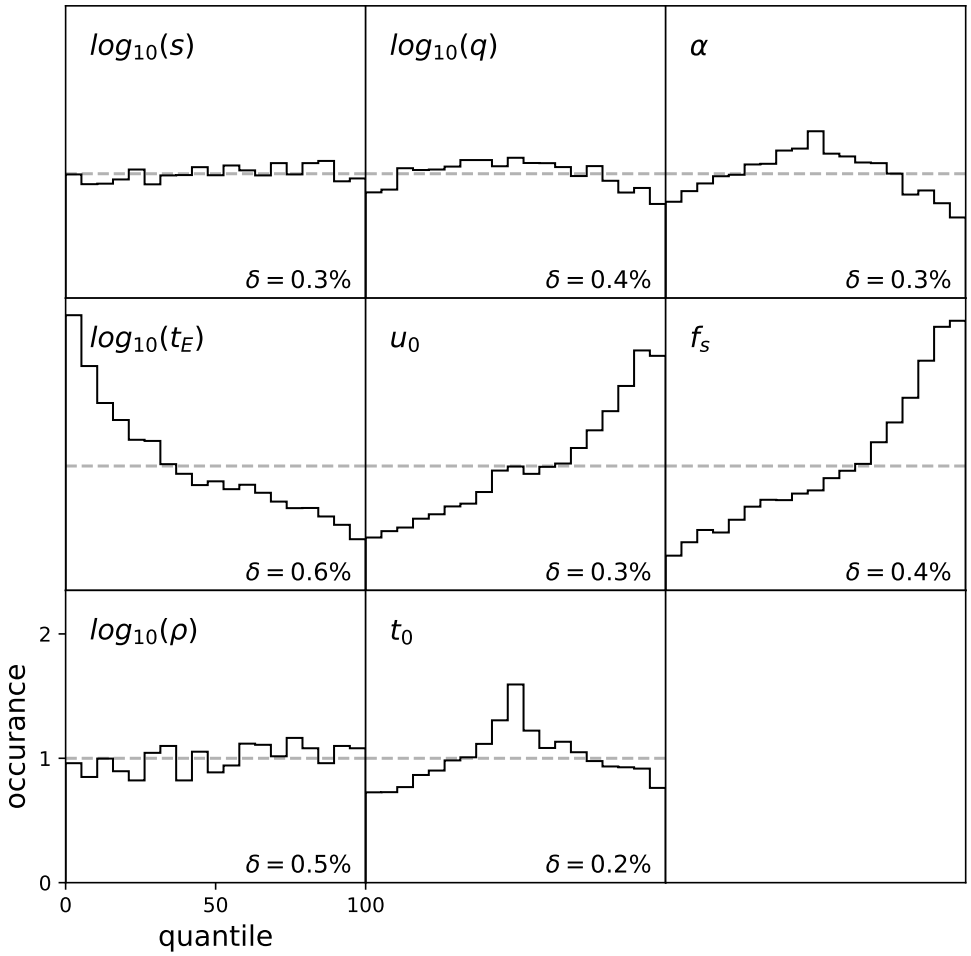


Figure 3.8: Calibration plot showing the test-set distributions of the ground truth quantile for the 1D marginal NDE posteriors. Dashed lines indicated the uniform distribution as expected for a perfectly calibrated posterior.

seen in Figure 3.6.

3.4.5 Calibration Properties

A perfectly calibrated posterior knows how often it is right or wrong. In other words, the quantile of the ground truth parameter under the NDE posterior should be expected to be distributed uniformly. Figure 3.8 shows the quantile distribution for the 1D marginal NDE posterior distributions for the same 14,551 test set inferences. The quantile distribution for $\log_{10}(q)$, $\log_{10}(\alpha)$, and t_0 is concave-up, indicating that the NDE uncertainty is overestimated and the true value lies closer to the center of the posterior more often than expected. This suggests that the NDE finds it hard to contract the posterior in those dimensions, possibly due to numerical optimization difficulties or insufficient neural network expressibility. On

the other hand, distributions for the three parameters in the second row—$\log_{10}(t_E)$, u_0, and f_s—demonstrate the systematic under/over-estimation as seen in Figure 3.6, where $\log_{10}(t_E)$ is systematically overestimated and u_0 and f_s are underestimated. The quantile distributions for $\log(s)$ and $\log_{10}(\rho)$ are consistent with uniform distributions and are thus well-calibrated.

3.5 Discussion and Conclusions

We have demonstrated that amortized neural posterior estimation, a likelihood-free inference method which uses a conditional NDE to learn a surrogate posterior, $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$, greatly accelerates binary microlensing inference—an approximate posterior could be produced in seconds without the need for an expert in the loop. Our new approach is capable of capturing a variety of degeneracies. For future work, it is straightforward to extend to higher-level effects such as parallax and binary motion by introducing additional parameters. Application to more complex systems such as 3L1S may be fruitful, where the physical forward model is orders of magnitude slower. In addition, the photometric noise model adapted in our study is somewhat simplistic, and future work can explore how to adapt models trained with ideal noise properties to fully realistic data with the help of image-based simulation pipelines such as ones used in Penny et al. (2019). We discuss two additional aspects of our work below.

3.5.1 A hybrid NDE-MCMC framework

The NDE posterior is easily validated and/or refined by a downstream MCMC sampler. While the NDE posterior is precise enough to allow for fast convergence of downstream MCMC typically within hundreds of steps, we do notice that the precision of the exact MCMC posterior could be more than order-unity higher in many cases. The precision of the NDE posterior is determined by two kinds of uncertainty: data uncertainty and the model uncertainty of the inference algorithm, the latter of which is negligible for MCMC. As neural networks in practice are not infinitely expressive, in the limit of the highest-quality data, the NDE model uncertainty is expected to dominate over data uncertainty. This is the case for *Roman* data. Applied to much noisier and more sparsely sampled ground-based data, we expect that data uncertainties will dominate over model uncertainties, thus allowing the NDE posterior to converge towards the exact posterior.

3.5.2 Choice of Coordinate System

For all events in this work, we have adopted the center-of-mass (COM) coordinate system, which is the default in `MulensModel` but not the most efficient reference frame in the sense that more than 70% of the 1 million simulations turn out to be consistent with a 1L1S model. For example, most 2L1S configurations with large u_0 do not pass close to either the central caustics or the planetary caustics. For parts of the parameter space, alternative reference coordinates may be more descriptive or useful. For example, the caustic-center

frame is preferred for binary and/or wide-separation events for which there is an offset of the caustic-center from the COM. Doing so recovers the missing wide/binary solution in Section 3.4.3 without the need to expand the prior to include negative t_0. Additionally, planetary-caustic passing events are also rare; for source trajectories far from the central caustics, most do not pass close to the planetary caustic and as a result, the magnification is frequently indistinguishable from 1L1S. For future work, a hybrid and self-consistent coordinate system could be used.

Acknowledgments

A version of this chapter was published in the Astronomical Journal as Zhang et al. (2021). An extended abstract of this Chapter appeared in the "Machine Learning for the Physical Sciences" workshop at the 2020 Neural Information Processing Systems (NeurIPS) Conference.

I thank Shude Mao and Tsinghua University for their hospitality during the COVID-19 pandemic, during which this chapter was written. Work in this chapter is supported by the AWS Cloud Credits for Research program. I thank Yang Gao and Yu Sun for helpful discussions, and Weicheng Zang for pointing out issues with the initial figure presentation. I am supported by a Gordon and Betty Moore Foundation Data-Driven Discovery grant. I thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant 1829740, the Brinson Foundation, and the Moore Foundation; my participation in the program has benefited the works presented in this chapter.

Chapter 4

A Ubiquitous Unifying Degeneracy in Two-Body Microlensing Systems

While gravitational microlensing by planetary systems (Mao & Paczyński 1991; Gould & Loeb 1992) provides unique vistas on the properties of exoplanets (Gaudi 2012), observations of a given 2-body microlensing event can often be interpreted with multiple distinct physical configurations. Such ambiguities are typically attributed to the *close–wide* (Griest & Safizadeh 1998; Dominik 1999) and *inner–outer* (Han et al. 2018) types of degeneracies that arise from transformation invariances and symmetries of microlensing caustics. However, there remain unexplained inconsistencies (e.g., Yee et al. 2021) between aforementioned theories and observations. Here, leveraging a fast machine learning inference framework (Zhang et al. 2021) (Chapter 3), we present the discovery of the *offset* degeneracy, which concerns a magnification-matching behaviour on the lens-axis and is formulated independent of caustics. This *offset* degeneracy unifies the *close–wide* and *inner–outer* degeneracies, generalises to resonant topologies, and upon reanalysis, not only appears ubiquitous in previously published planetary events with 2-fold degenerate solutions, but also resolves prior inconsistencies. Our analysis demonstrates that degenerate caustics do not strictly result in degenerate magnifications and that the commonly invoked *close–wide* degeneracy essentially never arises in actual events. Moreover, it is shown that parameters in *offset* degenerate configurations are related by a simple expression. This suggests the existence of a deeper symmetry in the equations governing 2-body lenses than previously recognised.

4.1 Discovery

In search for new types of microlensing degeneracies, we analysed the posterior parameter distribution of a large number of simulated 2-body microlensing events that exhibited multi-modal solutions. With over 100 planetary microlensing events observed so far, new degeneracies have indeed been serendipitously found in routine data analysis (e.g., Choi et al. 2012). However, while an exhaustive search on examples of multi-modal event pos-

teriors to constrain the existence of unknown degeneracies is plausible, such an endeavour has been computationally prohibitive with the current *status-quo* microlensing data analysis approaches. Thankfully, the recent application of likelihood-free inference (LFI) (see Cranmer et al. 2020 for an overview) to 2-body microlensing (Zhang et al. 2021) (Chapter 3) has accelerated calculation of microlensing posteriors to a matter of seconds, thus allowing posteriors for a large number of simulated events to be acquired with minimal computational cost.

The key to the accelerated inference is the use of a *Neural Density Estimator* (NDE), which is a particular type of neural network capable of modelling distributions that are complex and multi-modal. Here, the NDE learns a mapping from microlensing light-curves directly to posteriors, allowing future inferences to be done with the NDE alone in mere seconds. Following Zhang et al. (2021) (Chapter 3), we trained an NDE on 691,257 events simulated in the context of the *Roman Space Telescope* microlensing survey (Penny et al. 2019) so that our results would be directly relevant. The posteriors for a large number of randomly generated events are then produced with the NDE. To identify events with multi-modal solutions, we applied a clustering algorithm (Campello et al. 2013) which separates each posterior into discrete modes. The exact maximum likelihood solution within each posterior mode is then calculated with an optimisation algorithm (see Methods).

Visual inspection of multi-modal NDE posteriors revealed three apparent regimes of degeneracy: the *inner–outer* degeneracy, the *close–wide* degeneracy, and degeneracies that involve the resonant caustic which have also been previously observed (e.g., Herrera-Martín et al. 2020; Yee et al. 2021) and studied (An 2021). The *close-wide* degeneracy states that the central caustic shape is invariant under the $s \leftrightarrow 1/s$ transformation for $|1-s| \gg q^{1/3}$ (An 2021) and $q \ll 1$ (Figure 4.4a;c), where q refers to the planet-to-star mass ratio, and s refers to their projected separation normalised to the angular Einstein radius ($\theta_E = \sqrt{\kappa M \pi_{rel}}$), which is the characteristic microlensing angular scale. Here, $\kappa = 4G/(c^2\mathrm{AU})$, M is the total lens mass, and $\pi_{rel} = \mathrm{AU}/D_{rel}$ is the lens-source relative parallax. Interestingly, we found that most cases of apparent *close-wide* degeneracies do not exactly abide by the expected $s \leftrightarrow 1/s$ relation even though most are in the $|1 - s| \gg q^{1/3}$ regime where it is expected to hold. We also noticed that for degenerate events involving one resonant caustic, the source trajectory always passed to the front end of the resonant caustic for *wide-resonant* degenerate events, and the back end for *close-resonant* degenerate events.

To explore potential connections among these apparently discrete regimes of degeneracies, and to better understand the reason why the expected $s \leftrightarrow 1/s$ relation of the *close-wide* degeneracy is almost never satisfied, we examined maps of magnification differences between pairs of lenses with the same mass-ratio ($q = 2 \times 10^{-4}$), keeping lens B fixed at $s_B = 1/1.1$ and changing the projected separation s_A of lens A. The sequence of magnification difference maps in Fig. 4.1a–h immediately reveals the continuous evolution of a vertically-extended ring structure where the magnification difference vanishes (also see Figure 4.5, 4.6). This *null* ring originates near the primary star and grows increasingly large with increasing deviation from the *close-wide* degenerate configuration of $s_A = 1/s_B$, at which point the *null* contracts to a singular point (see Extended Figure 4.7 for a zoom-in). We may thus expect null-

passing trajectories (cyan arrows in Fig. 4.1a–h) to have degenerate magnifications, which is confirmed by light curves shown in Fig. 4.1i–p.

It is also immediately clear from Fig. 4.1f why the *close-wide* pair of configurations ($s_A = 1/s_B$) does not result in degenerate magnifications for any trajectory shown: the magnification differs everywhere on the lens-axis except for the singular null point. Thus for any given trajectory, close to or far from the central caustic, one can always move the *null* to the location of the source by shifting the planet location, to have the magnifications match exactly on the lens axis. For caustic crossing trajectories, the vertical extension of the *null*, located within the caustic (Figure 4.7c), also allows the width of the caustic to be matched (Figure 4.1f). We also found that both location and shape of the *null* are independent of q for $q \ll 1$, thus allowing the above discussion to also hold in the $|1 - s| \gg q^{1/3}$ regime (see Figure 4.8) of the *close-wide* degeneracy. This demonstrates that the above localised degeneracy does not arise due to the imperfect matching of the central caustic shapes, but is an fundamental behaviour of the lensing system in the limit of $q \ll 1$.

We name this phenomenon the *offset* degeneracy to refer to the source-*null* matching principle where the *null* is created by an *offset* of the planet location on the binary axis. Notably, we found that the location of the *null* on the star-planet axis is well described by a simple expression:

$$x_{\rm null} = \frac{1}{2}\left(s_A - 1/s_A + s_B - 1/s_B\right), \tag{4.1}$$

Numerically determined $x_{\rm null}$ (Figure 4.2) shows that deviations from this analytic prescription is consistently less than 5% except for extreme separation ($|\log_{10}(s)| \gtrsim 0.5$) cases where sources do not pass close to either caustic and therefore do not yield substantial planetary perturbation to be of practical interest. This expression can be interpreted as the midpoint between the locations $x_c = s_{A,B} - 1/s_{A,B}$ of the planetary caustics, which arises from the perturbative picture of planetary microlensing (Gould & Loeb 1992). However, the fact that such an expression holds well into the resonant regime for which there are no planetary caustics at all, and persists through caustic topology changes, likely suggests the existence of much deeper symmetries in the gravitational lens equation for mass ratios of $q \ll 1$ than had previously been appreciated, and should be explored in future work.

We now consider the relationship between the *offset* degeneracy and the two previously known mathematical degeneracies. Firstly, the *offset* degeneracy is a magnification degeneracy while the two previous degeneracies are caustic degeneracies. Our analysis demonstrates that degenerate caustics do not strictly result in degenerate magnifications. Furthermore, by setting $x_{null} = 0$ in Equation 4.1, one immediately recovers the $s_A = 1/s_B$ relation of the *close-wide* degeneracy. This suggests that the *close-wide* degeneracy is more suitably viewed as a transition point of the *offset* degeneracy where the central caustics *happen* to be degenerate. On the other hand, while the *inner-outer* degeneracy implies an expression similar to Equation 4.1 (Han et al. 2018), it arises from the symmetry of the Chang-Refsdal (Chang & Refsdal 1979) approximation to the planetary caustics (Gaudi & Gould 1997). However, cases attributed to the *inner-outer* degeneracy are often not in the pure Chang-Refsdal regime (Yee et al. 2021) in which case the planetary caustics are asymmetrical.

Also, even in the Chang-Refsdal regime, in observed events the source trajectory is fixed and passes equidistant to two different planetary caustics, rather than two sides of the same caustic. Therefore, the *offset* degeneracy not only resolves inconsistencies and unifies the two previously known degeneracies into a generalised regime, but also relaxes the $|1 - s| \gg q^{1/3}$ condition required by both cases.

Because of this unifying feature, we expected the *offset* degeneracy to be ubiquitous in past events with 2-fold degenerate solutions and speculate that a large number of cases may have been mistakenly attributed to the *close-wide* degeneracy. Therefore, we systematically searched for previously-published events with two-fold degenerate solutions satisfying $q_A \simeq q_B \ll 1$ (see SI). We found 23 such events, and then first compared the intercept of the source trajectory on the star-planet axis to the location of the *null* predicted with Equation 4.1. We also invert Equation 4.1 to predict one degenerate s_A from the other s_B:

$$s_A = \frac{1}{2}\left(2x_0 - (s_B - 1/s_B) + \sqrt{[2x_0 - (s_B - 1/s_B)]^2 + 4}\right), \tag{4.2}$$

where $x_0 = u_0/\sin(\alpha)$ is the intercept of the source trajectory on the binary axis, u_0 is the impact parameter, and α is the angle of the source trajectory with respect to the binary axis. As shown in Figure 4.3, the source trajectory always passes through the *null* location on the star-planet axis as predicted by Equation 4.1. Additionally, Equation 4.2 accurately predicts one degenerate solution from the other. The fact that Equation 4.1 applies for a wide range of α confirms that the *offset* degeneracy accommodates oblique trajectories, although proximity to planetary caustics might break the degeneracy (e.g., KMT-2016-BLG-1397 (Zang et al. 2018)). Thus we conclude that Equations 4.1,4.2 will be useful in the analysis of future events with *offset*-degenerate solutions.

Given its apparent ubiquity, it is reasonable to ask why the *offset* degeneracy has only been discovered over two decades after the first in-depth explorations of degeneracies in two-body microlensing events (Gaudi & Gould 1997; Griest & Safizadeh 1998; Dominik 1999). One reason may be the early strategic focus on high-magnification ($u \ll 1$) events (Griest & Safizadeh 1998; Gould et al. 2010), where deviations from $s \leftrightarrow 1/s$ were small, whose cause was not explored in detail. Recently, deviations from $s \leftrightarrow 1/s$ in semi-resonant topology events have led to explicit discussions on the applicability of the *close-wide* degeneracy in the resonant regime and potential connections to the *inner-outer* degeneracy (Yee et al. 2021; An 2021). Nevertheless, as we have shown, the resonant condition itself does not cause the deviation from $s \leftrightarrow 1/s$, but only allows it to be noticeable (see Methods). To our advantage, the novel ML-based technique of Zhang et al. (2021) (Chapter 3) presented us with a large number of degenerate events in non-resonant $|1-s| \gg q^{1/3}$ regime that deviated from the $s \leftrightarrow 1/s$ expectation, but also did not conform to the *inner-outer* degeneracy. These 'intermediate' *offset*-degenerate events ultimately allowed us to recognise the continuous and unifying nature of the *offset* degeneracy, showcasing another instance of ML-guided discovery of new theoretical insight (c.f. Davies et al. 2021). As the next-generation surveys further expand the sensitivity limit from space (Bennett & Rhie 2002), the *offset* degeneracy will increasingly manifest.

4.2 Methods

4.2.1 The Z21 fast inference technique

Zhang et al. (Zhang et al. 2021) (Chapter 3; Z21 hereafter) presented a likelihood-free inference (LFI) approach to binary microlensing analysis that allowed an approximate posterior for a given event to be computed in seconds on a consumer-grade GPU, compared to the hours-to-days timescales on CPU clusters that are typically required for *status-quo* approaches. We summarise the Z21 approach at the high level here, and refer the reader to the original paper for details.

The Z21 method is likelihood free in that it does not iteratively perform simulations to compute the likelihood, which is typical for sampling-based inference methods. Instead, Z21 directly learns the posterior probability as a conditional distribution $\hat{p}_\phi(\theta|x)$ with an NDE, where ϕ are the NDE parameters, θ the binary microlensing (2L1S) parameters, and x the input light curve. The NDE is essentially a mapping that takes a light curve as input and produces a specified number of discrete posterior samples. Such a mapping is trained on a large number of simulations (x_i, θ_i) with parameters drawn from a wide prior, and the NDE parameters (ϕ) are optimised to maximise the expectation of that conditional probability under the training set data distribution. The mapping learned can thus be applied to any given event unseen during training as long as it is within the pre-specified prior.

This specific approach to LFI is called *amortised neural posterior estimation*, where "amortised" refers to the process of paying all simulation cost upfront so that inferences of future events do not require additional simulations. After training, the NDE alone generates posterior samples for any future event at a rate of $\sim 10^6$ s^{-1} on a consumer grade GPU, or $\sim 10^5$ s^{-1} on a 8-core CPU, effectively doing inference in real time. Z21 demonstrated that, although not exact, the neural posterior places accurate constraints on all parameters nearly 100% of the time, except for the parameter that quantifies the effect of a finite-sized source. This is because substantial finite source effects only occur when the source approaches sufficiently close to the caustics, which is satisfied by only a small subset of events.

With a focus on the next-generation, space-based (Bennett & Rhie 2002) microlensing survey planned on the *Roman Space Telescope* (Penny et al. 2019), here we generated a training set in a similar fashion as the Z21 training set, but with a caustic-centred coordinate system rather than a centre-of-mass (COM) coordinate system. This is because the COM coordinate system is highly inefficient for producing planetary-caustic passing events with randomly drawn source trajectories with respect to the COM. In addition, for wide binary $(s > 1; q \sim 1)$ events, the time-to-closest-approach (t_0) to the COM could have an arbitrarily large offset from the time of peak magnification, which can lead to the missing of solution modes (see Section 4.3 of Z21). The caustic-centred coordinate system, on the other hand, efficiently spans the entire 2L1S parameter space that allows for substantial deviation from a single-lens light curve.

We generated a total of 228,892 events centred on the planetary caustic and 960,000 events centred on the central caustic, and further remove those that are consistent with a

single lens model by fitting each light curve to such a model and adopting a $\Delta\chi^2 = 140$ cutoff (see Z21). This resulted in a training set of 691,257 simulations, including 137,644 planetary caustic events and 553,863 central caustic events.

For planetary caustic events, u_0 is randomly sampled from 0 to 50 times the caustic size. For central caustic events, u_0 is randomly sampled from 0 to 2. Compared to Z21, we expanded the source flux fraction, defined as $f_s = \frac{F_{\rm source}}{F_{\rm source} + F_{\rm blend}}$, to $f_s \sim \text{LogUniform}(0.05, 1)$, to probe deeper into the severely blended regime. Other aspects of event simulation are the same with Z21 and the reader is referred to Section 3 of Z21 for details.

4.2.2 Identifying degeneracies in Z21 posteriors

Z21 provided three example events with degenerate posteriors where light curve realisations from each degenerate mode are almost indistinguishable from one another, a confirmation of the effectiveness in modelling light curves with degenerate solutions. While the posterior modes in Z21 were identified manually, in this work we automate the degeneracy-finding process.

To work with posterior distributions that vary in scale, position, and shape, we first fit and apply a parametric, monotonic "power" transformation (Yeo 2000) to the LFI-generated posterior samples for each simulated light curve. This transformation normalises each marginal parameter distribution to an approximate Gaussian. To automatically identify degenerate posteriors, we used the HDBSCAN algorithm (Campello et al. 2013) to perform clustering on the transformed posterior samples. The HDBSCAN algorithm is a density-based, hierarchical clustering method which required, for our task, minimal hyperparameter tuning. The output of HDBSCAN is a suggested cluster label for each posterior sample, including the labelling for outlier/noise samples. Events with more than one cluster are identified as degenerate events.

Although the NDE posteriors are accurate enough for a qualitative study of degeneracies, we nevertheless refined each solution mode to the maximum likelihood value. The approximate posterior allows us to make use of bounded optimisation algorithms to quickly locate the exact solution. We use a parallel implementation (Gerber & Furrer 2019) of the L-BFGS-B optimisation algorithm (Byrd et al. 1995) to quickly solve for the best fit solutions. The entire process from light curve to degenerate exact solutions takes a few minutes for each event, with the last refinement step costing the most time.

4.2.3 Comparison to events in the literature

We demonstrate the ubiquity of the offset degeneracy by performing a thorough investigation of 2L1S events in the literature with reported degenerate posteriors. We first filter through events on the NASA microlensing exoplanet archive which contains 112 planets and 306 entries with reported 2L1S parameters (retrieved August 23rd, 2021). Each entry reports one solution for a given event.

Entries from adaptive-optics follow-up papers of published events, as well as duplicate entries with identical 2L1S solutions are first removed. Triple lens events with detections of two planets — OGLE-2006-BLG-109 and OGLE-2018-BLG-1011 — are also removed. Planets with reported higher-order effects (parallax, xallarap) are also removed, as such effects often exhibit additional degeneracies and may complicate the application of the offset degeneracy. We further remove 2-fold degenerate events with $\Delta\chi^2 > 10$ where one solution is significantly favoured. This leaves us with 20 planets with exactly two solutions and 12 with more than two solutions.

Among the 20 planets with exactly two solutions (Skowron et al. 2018; Janczak et al. 2010; Hirao et al. 2016; Nagakane et al. 2017; Suzuki et al. 2013; Dong et al. 2009; Herrera-Martín et al. 2020; Rattenbury et al. 2017; Bond et al. 2017; Han et al. 2018; Bennett et al. 2017; Hirao et al. 2017; Han et al. 2017; Hwang et al. 2019; Han et al. 2020a; Ranc et al. 2019; Nucita et al. 2018; Han et al. 2021; Kim et al. 2021; Han et al. 2020c) six are excluded: KMT-2016-BLG-1107 (Hwang et al. 2019) because it is a different type of degeneracy: two distinct source trajectories crossing the $s < 1$ planetary caustic, one of which is parallel to and does not intersect with the binary axis, OGLE-2017-BLG-0373 (Skowron et al. 2018) because it is an accidental degeneracy without complete temporal coverage of the caustic entrance/exit, and KMT-2019-BLG-0371 (Kim et al. 2021) because of the large mass-ratio ($q \sim 0.1$) and that the *offset* degeneracy only *strictly* manifests when $q \ll 1$. We also exclude OGLE-2016-BLG-1227 (Han et al. 2020c) and OGLE-2016-BLG-0263 (Han et al. 2017) because in both cases $s_{\rm min,max} \sim 4$ makes difficult to include in Figure 3 scale-wise, and because both cases are deep in the $|1 - s| \gg q^{1/3}$ limit, and are thus already well-characterised by the *inner-outer* degeneracy. Similarly, MOA-2007-BLG-400 (Dong et al. 2009) is also deep in the $|1 - s| \gg q^{1/3}$ limit and represents one of the few instances where the source passes almost exactly the location of the primary star, thus allowing a degenerate pair of central caustics to manifest. However, the large uncertainty of $s_{\rm wide} = 2.9 \pm 0.2$ translate into an uncertainty in $x_{\rm null}$ that is orders-of-magnitude larger than the size of the central caustic, and makes it uninformative to include here.

We also inspected events with more than two degenerate solutions, and found that the solutions of KMT-2019-BLG-1339 (Han et al. 2020b) and MOA-2015-BLG-337 (Miyazaki et al. 2018) both consist of two pairs of degeneracies, each with their distinct shared mass-ratios. For both events, we include the pairs of solutions with planetary mass-ratios ($q \ll 1$).

Beyond the total 16 degenerate events retrieved from the NASA microlensing exoplanet archive and discussed above, we further looked for relevant events in the literature that are not included in the NASA exoplanet archive. Additions include the pairs of solutions with planetary mass-ratios for OGLE-2011-BLG-0526 (Choi et al. 2012) and OGLE-2011-BLG-0950 (Choi et al. 2012), as well as the four events with degenerate solutions recently reported in Hwang et al. (2022). We also include OGLE-2019-BLG-0960 (Yee et al. 2021). This results in a final sample of 23 degenerate events.

4.2.4 Range of applicability of the offset degeneracy

When considering larger mass ratios q, we find the qualitative structure of the *null* persists through $q \to 1$ (Figure 4.6, 4.8), suggesting that some form of the *offset* degeneracy may manifest even for $q \gtrsim 0.1$ events. In this regime, there should also be a transition point similar to the *close-wide* degeneracy that results in $x_{\rm null} = 0$, but $q_A = q_B$ may not hold, nor $s_A = 1/s_B$. For example, in the quadrupole and pure-shear approximation, the analogy to the *close-wide* degeneracy requires $\hat{Q} = \gamma$, where $\hat{Q} = s_c^2 \cdot q_c/(1+q_c)^2$ is the quadrupole moment of the close central caustic, and $\gamma = (1/s_w)^2 \cdot q_w/(1+q_w)$ is the shear of the wide central caustic (Dominik 1999). Furthermore, it is not clear if the values of $q_{A,B}$ at the $x_{\rm null} = 0$ *close-wide-equivalent* transition point remains constant when one of s_A and s_B undergoes *offset*. A notable example in the literature is KMT-2019-BLG-0371 (Kim et al. 2021) where the source trajectory passes through the null created by the two degenerate solutions but $q_A = 0.123$ and $q_B = 0.079$ are substantially different. The exact behaviour of the *offset* degeneracy for $q \to 1$ should be studied in future work.

We also note that *offset*-degenerate, caustic crossing events usually require nearly-vertical trajectories because of the additional constraint on the caustic-crossing length. However, oblique trajectories are allowed if the change in caustic width near $x_{\rm null}$ is small for both solutions (e.g., OGLE-2019-BLG-0960 Yee et al. 2021).

4.2.5 Relevant prior work

Inconsistencies of the *close-wide* and *inner-outer* degeneracies with degeneracies in observed events have recently been pointed out in the literature. In the analysis of the semi-resonant topology event OGLE-2019-BLG-0960, Yee et al. (2021) noticed that while the *close-wide* degeneracy is expected to break down as $s \to 1$, there are large numbers of resonant and semi-resonant topology events invoking the *close-wide* degeneracy, where one solution has $s_{\rm close} > 1$ and the other $s_{\rm wide} < 1$, but do not satisfy $s_{\rm close} = 1/s_{\rm wide}$. They further noted the conceptual similarity to the *inner-outer* degeneracy for these events, but again noted that this type of degeneracy too is expected to break down in the resonant regime. Based on these observations, they speculated that the two degeneracies merge as $s \to 1$.

While Yee et al. (2021) pointed out inconsistencies for resonant events ($|1-s| \lesssim q^{1/3}$), here we found that inconsistencies with $s_{\rm close} = 1/s_{\rm wide}$ persists even within the $|1-s| \gg q^{1/3}$ regime in which the two degeneracies are derived and the caustics are well separated. We claim that this inconsistency is fundamentally because caustic degeneracies are only approximately correct in describing magnification degeneracies, irrespective of caustic topology. While small deviations from $s_{\rm close} = 1/s_{\rm wide}$ in early high-magnification events tend to go unnoticed, resonant events do allow the asymmetry from $\log(s) = 0$ to be immediately noticeable. For OGLE-2019-BLG-0960, $\log_{10}(s_{close}) \simeq -0.001$ differs from $\log_{10}(s_{wide}) \simeq 0.01$ by an order of magnitude.

The theoretical follow up work of An (2021) studied the behaviour of the *close-wide*

degeneracy in the resonant regime. They first clarified that rather than $|\log(s)| \gg 0$, the exact condition of the *close-wide* degeneracy is $|1 - s| \gg q^{1/3}$, which is dependent on the mass ratio. Furthermore, even for $|1 - s| \lesssim q^{1/3}$, the central caustic could still be locally invariant under $s \leftrightarrow 1/s$ for parts of the caustic satisfying $|1 - se^{i\phi}| \gg q^{1/3}$, where ϕ is a parametric variable that describes the position along the caustic. We note that this fact has also been observed in the earlier work of Bozza (1999). They concluded by suggesting that slight changes to $s_{A,B}$ and $q_{A,B}$ may create a local pair of degenerate models, which in some sense anticipated our discovery.

Acknowledgments

Material from: 'K. Zhang, B. S. Gaudi & Joshua S. Bloom, A ubiquitous unifying degeneracy in two-body microlensing systems, Nature Astronomy, published 2022, Springer Nature.' A version of this chapter is published in Nature Astronomy as Zhang et al. (2022).

K.Z. thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. K.Z. and J.S.B are supported by a Gordon and Betty Moore Foundation Data-Driven Discovery grant. Work by B.S.G. is supported by NASA grant NNG16PJ32C and the Thomas Jefferson Chair for Discovery and Space Exploration. We thank Eric Agol and Jessica Lu for helpful comments on a draft of this Chapter.

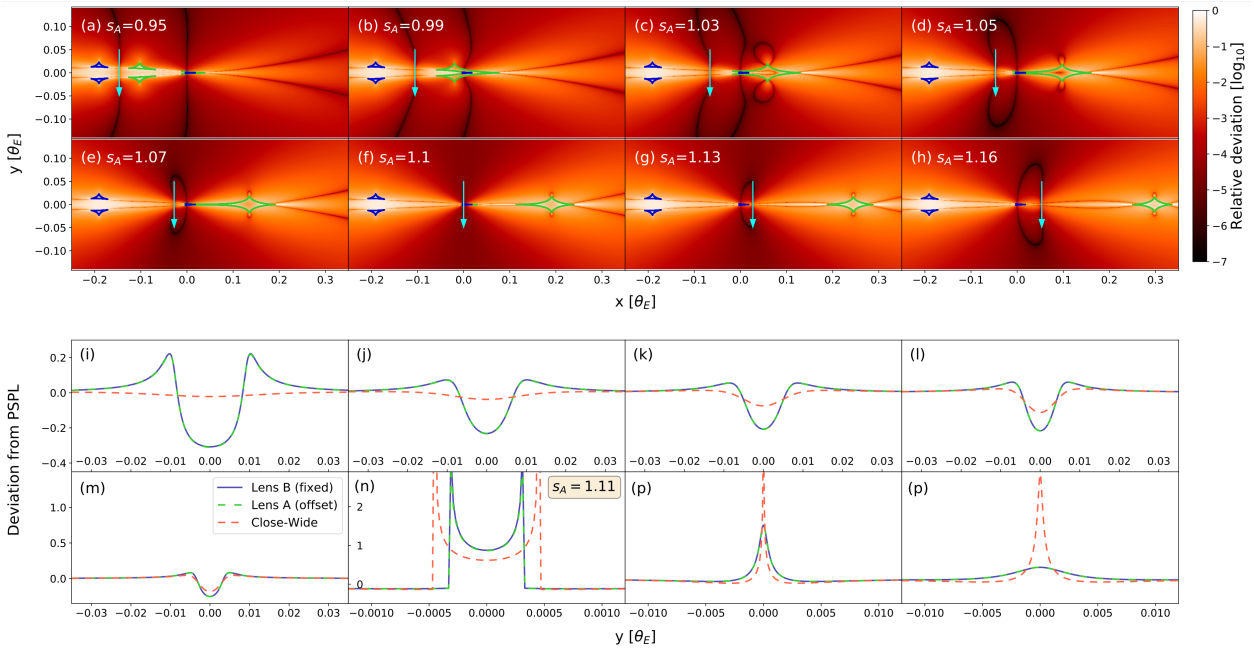


Figure 4.1: The manifestation of the *offset* degeneracy in source-plane magnification differences maps (top) and light curves (bottom). (a)–(h): Maps of magnification differences from lens B with fixed $s_B = 1/1.1$ to lens A with changing s_A specified in each subplot. The mass-ratio is fixed at $q = 2 \times 10^{-4}$ for all configurations. All magnification difference maps are shown on the same scale, specified in the colour-bar to the right. Lens A caustics are shown in green and lens B caustics are shown in blue. The black, oval-shaped ring with first decreasing and then increasing sizes in (a)–(h) is the *null* where the magnification difference between lens A/B vanishes. The evolution of the null ring is continuous with the progression of the lens A caustic into the resonant regime (e, f, g) and further into a wide topology (h). (i)–(p): Light curves for *null* crossing trajectories (cyan arrows in (a)–(h)), under lens A (blue), lens B (green), and the $s_A = 1/s_B = 1.1$ solution (red) expected from the close-wide degeneracy. Light curves are shown as relative deviations from the corresponding point-source point-lens (PSPL) model. Subplot (n) is shown for $s_A = 1.11$ instead of the $s_A = 1/s_B$ value of (f) to demonstrate the *offset* degeneracy for caustic crossing events: both caustic-crossing length and magnification patterns are matched for the *offset* solution but not for the *close-wide* solution.

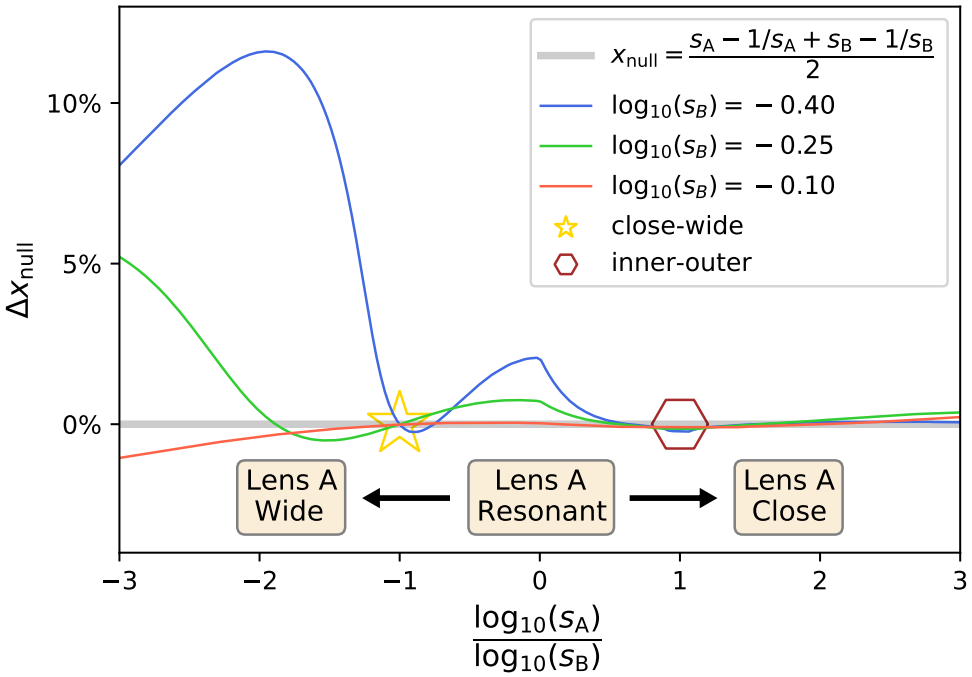


Figure 4.2: Deviation ($\Delta x_{\rm null}$) of numerically-derived, exact null position from the analytic form (Equation 4.1) for changing s_A against three values of fixed $s_B < 1$, normalised to the separation between the two (implied) planetary caustics: $|(s_A - 1/s_A) - (s_B - 1/s_B)|$. $\Delta x_{\rm null}$ is calculated for $q = 2 \times 10^{-4}$ but was found to be independent of q for $q \ll 1$ (Figure 4.8). The x-axis shows $\log_{10}(s_A)$ scaled to $\log_{10}(s_B)$ such that -1 corresponds to the *close-wide* degenerate case of $s_A = 1/s_B$ (gold star), 0 corresponds to $s_A = 1$, and 1 corresponds to the asymptotic *inner-outer* degenerate case where $s_A = s_B$ (brown hexagon). The coordinate origin is set to $sq/(1+q)$ from the primary for $s < 1$ and $s^{-1}q/(1+q)$ for $s > 1$, which describe the location of the central caustic and accounts for the non-differentiability at $s_{\rm A} = 1$.

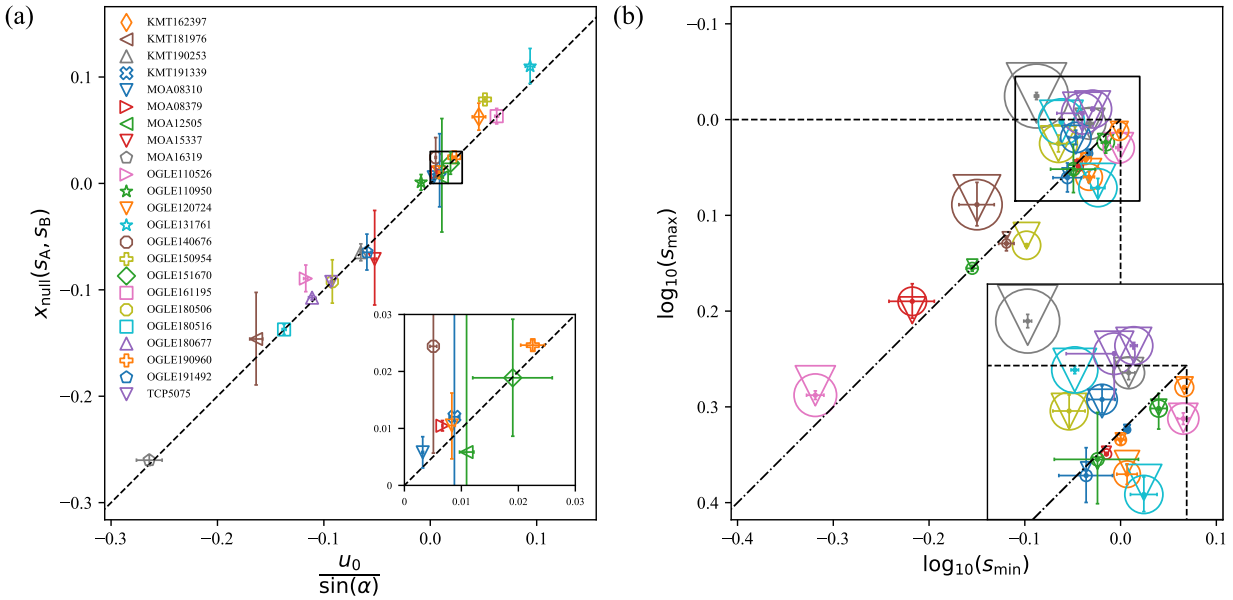


Figure 4.3: *Offset* degeneracy reanalysis of 23 systematically selected events in the literature with two-fold degenerate solutions. (a) confirms that the source trajectory always passes close to the null intercept on the star-planet axis ($x_{\rm null}$) as predicted by Equation 4.1. The x-axis shows the source trajectory intercept on the star-planet axis, calculated from the impact parameter (u_0) and trajectory angle (α). The y-axis shows the prediction for $x_{\rm null}$ using Equation 4.1 and reported values of s_A and s_B. Event labels as shown in the legend are the event abbreviations: for example, KMT162397 means KMT-2016-BLG-2397. The inset shows zoom-in of the central boxed region. (b) The x and y-axis show the smaller and larger value of the degenerate solutions referred to as $s_{\rm min,max}$. Circles are reported values of $s_{\rm min,max}$ whereas triangles are $s_{\rm max}$ values predicted with Equation 4.2 of the *offset* degeneracy and $s_{\rm min}$, α, and u_0. The same colour coding follows from the legend in (a). Circles and triangles largely coincide for all cases, demonstrating the predictive power of the *offset* degeneracy. Sizes of circles and triangles are scaled to the expected *null* location: $x_0 = u_0/\sin(\alpha)$ to show the correlation between larger size and greater distance from the dash-dotted diagonal line that represents the exact *close–wide* degeneracy where $s_{\rm min} = 1/s_{\rm max}$. Cases typically understood as *inner–outer* —$s_{A,B} > 1$ or $s_{\rm A,B} < 1$ — are found outside the box bounded by the dashed lines. Cases close to the dashed lines but far from their conjunction correspond to *resonant–close/wide* degeneracies. Cases within the dashed box and not on the diagonal line do not belong to either *close–wide* or *inner–outer* degeneracies. The inset shows zoom-in of the region boxed by solid lines. Error-bars are marginalised 1–σ posterior intervals. Uncertainties for the predicted $x_{\rm null}$ are propagated from the uncertainties of only one of s_{min} and s_{max} that give rise to a smaller uncertainty on $x_{\rm null}$.

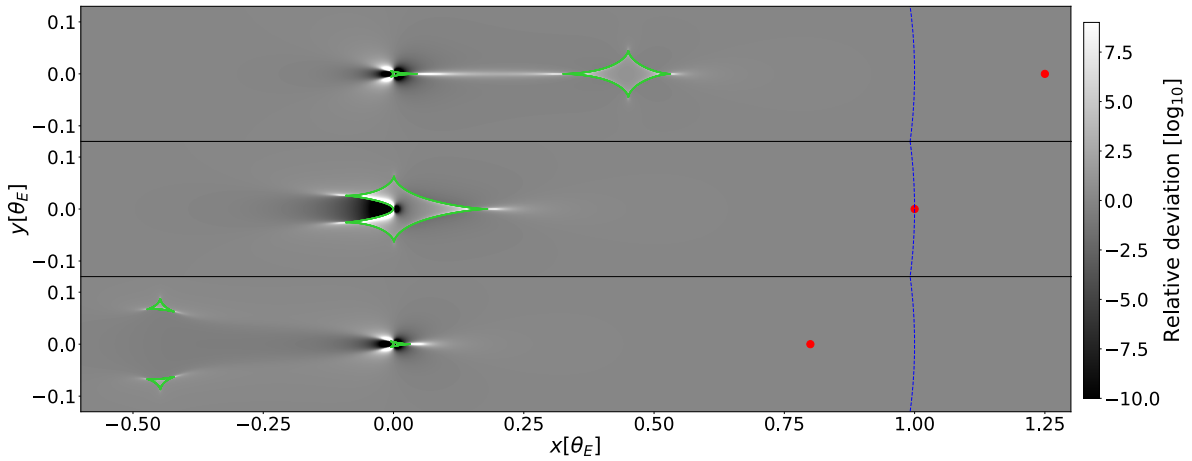


Figure 4.4: Caustics shown in green atop of maps of magnification differences from a 1-body lens, for wide (top), resonant (middle), and close (bottom) caustic topologies. Red dots indicate locations of the planet, with separations $s = 1/0.8, 1, 0.8$ from the host star, located at the origin. Blue dashed lines represent the Einstein ring $\theta_{\rm E}$, the angular size to which the projected separation (s) is normalised. Caustic topologies are delineated by values of s for a given q. In the wide regime ($s \gtrsim 1 + (3/2)q^{1/3}$), there is one central caustic located near the host star and one asteroid-shaped "planetary" caustic towards the location of the planet. In the close regime ($s \lesssim 1 - (3/4)q^{1/3}$), there are two small, triangular shaped "planetary" caustics in addition to the central caustic that appears similar to the wide central caustic, due to the *close-wide* degeneracy. For values of s in between these regimes, there is one six-cusped "resonant" caustic. For all cases, there are lobes of excess magnification compared to a point lens near caustic cusps, and lobes of de-magnification towards the back-end of the central/resonant caustic.

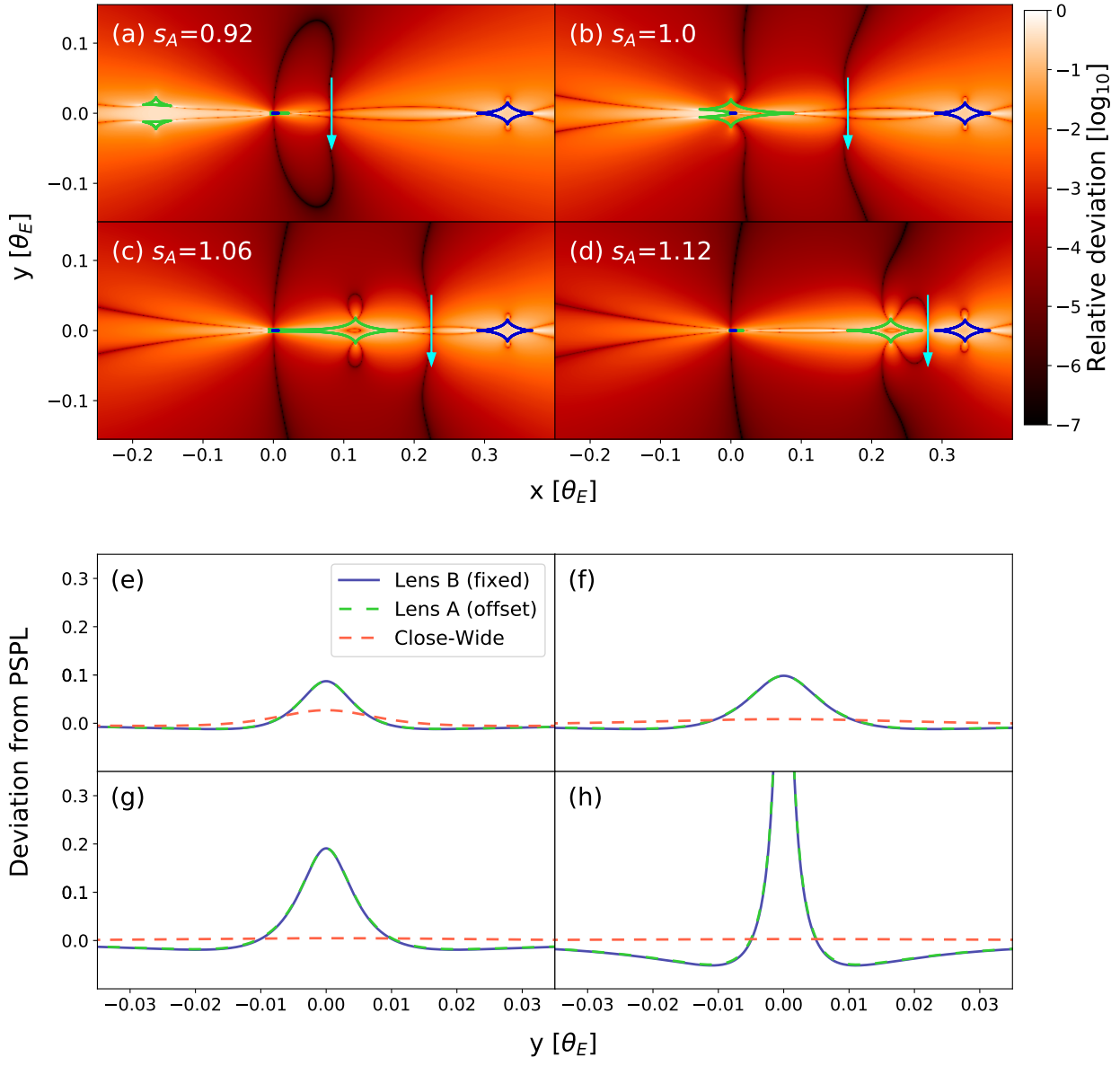


Figure 4.5: Similar to Figure 4.1, but for fixed $s_B = 1.18 > 1$. This completes the *resonant-close* (b) and wide-topology *inner-outer* (d) cases.

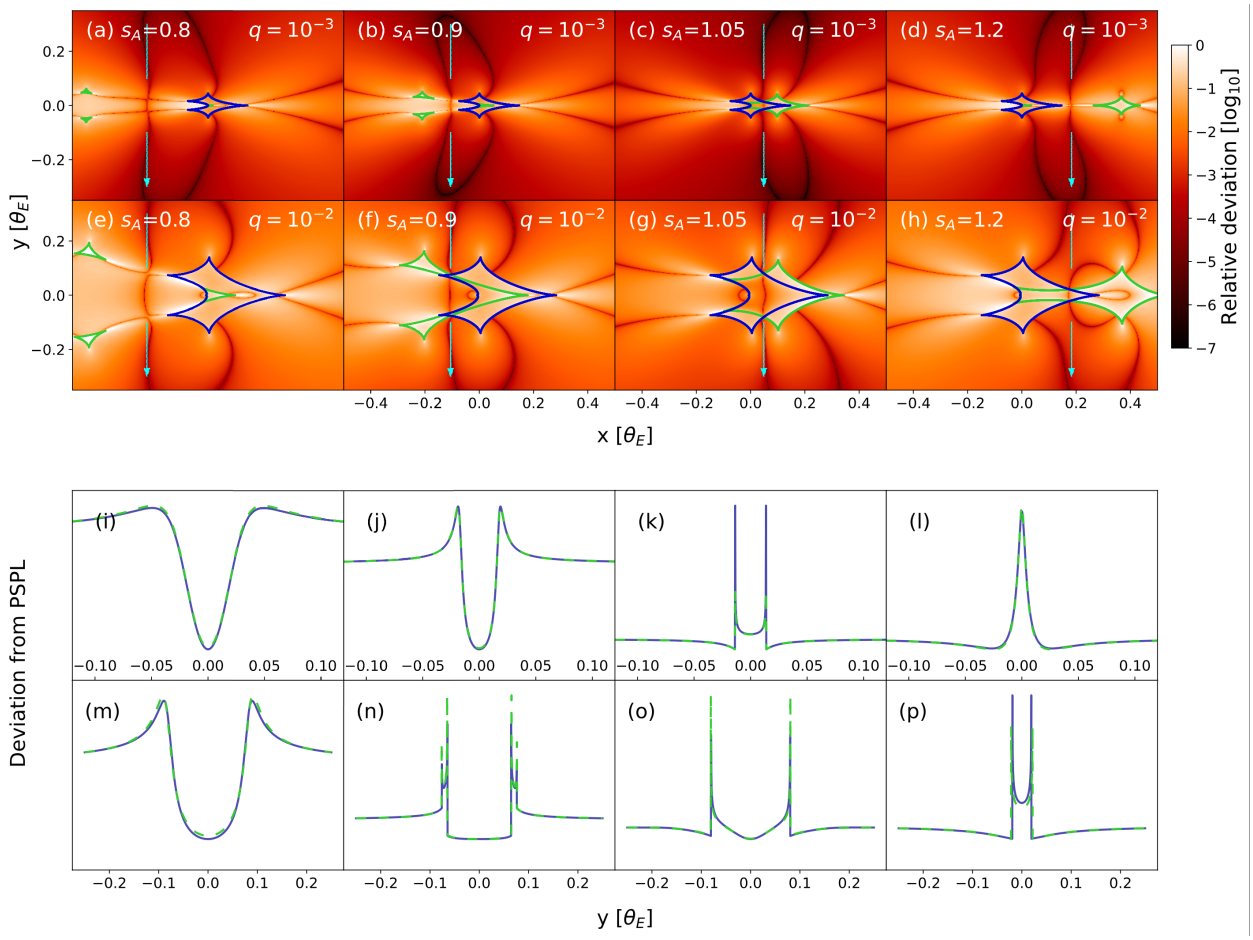


Figure 4.6: Magnification difference maps similar to Figure 4.1, but for fixed $s_B = 1$. (i)–(p) shows logarithmic deviations from PSPL on arbitrary scales, where green dashed curves are the changing lens A and sold blue curves are for fixed lens B. (a)–(d) and (e)–(h) show the same sequence of s_A but for $q = 10^{-3}$ and $q = 10^{-2}$ to illustrate how the *offset* degeneracy generalises to larger mass-ratios. (a,e) reveals that the ring structure of the null is composed of two distinct null segments, where one appears to originate from the centre of the central/resonant caustic and the other from the left two cusps of the same caustic. Closer inspection shows that the null rings for (a) and (e) have different topologies: for (a) it is the left part of the null that intersects on the star-planet axis but for (e) it is the right part. This disjoint topology of the null is also seen in Figure 4.1 and Figure 4.7 & 4.8. The topology transition point, presumably a function of s and q, may have mathematical implications for the *offset* degeneracy. Furthermore, we observe that the null segment near the star-planet axis becomes increasingly curved for $|\log(s)| \gg 0$ and $q \to 1$, which may explain how Equation 4.1 and the *offset* degeneracy in general, may break down in those limits.

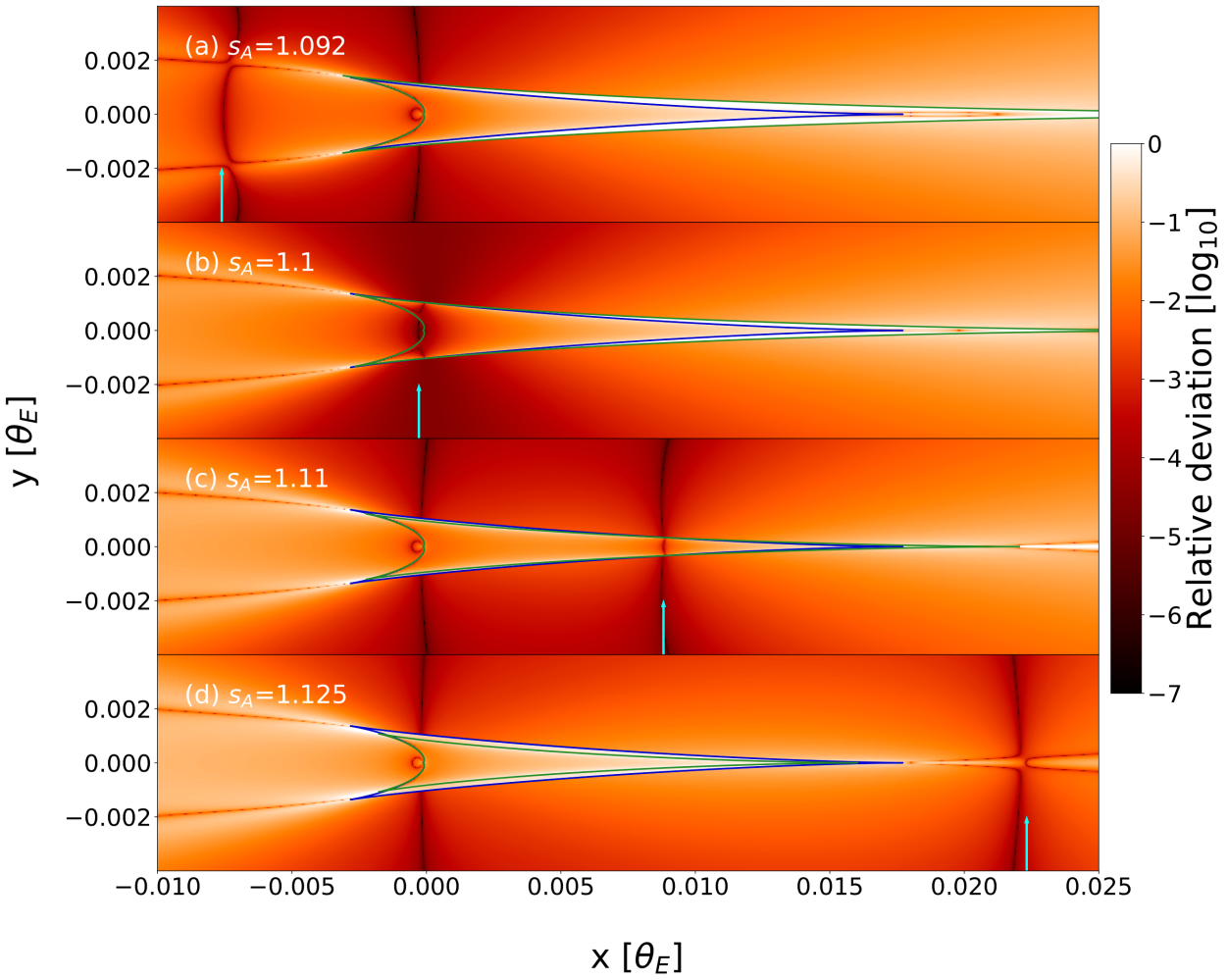


Figure 4.7: Magnification difference maps zoomed-in on the central caustic. Same $s_B = 1/1.1$ as Figure 4.1. Cyan arrows indicate the location of the null. For (b)–(c), the *null* always crosses the two caustics at their intersection.

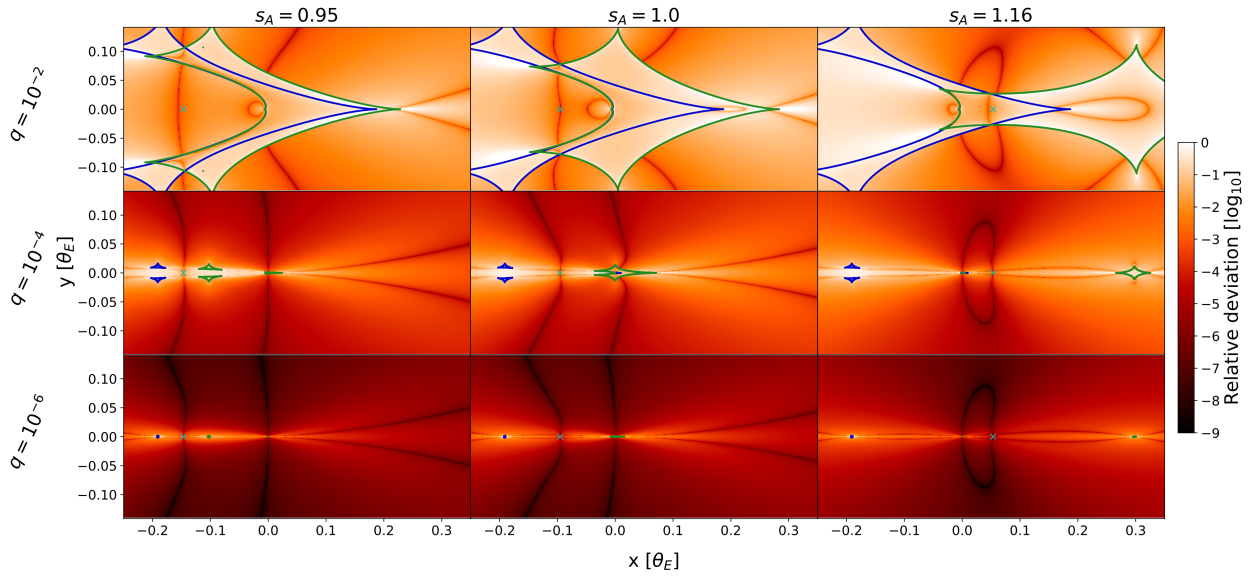


Figure 4.8: Magnification difference maps which demonstrates the *offset* degeneracy independence on q for $q \ll 1$. Lens B shares the same fixed $s_B = 1.1$ as in Figure 4.1. Each row shows cases of $s_A = 0.95, 1, 1.16$ for $q = 10^{-2}, 10^{-4}, 10^{-6}$. The null location predicted from Equation 4.1 is shown in cyan crosses. For $q = 10^{-4}$ and $q = 10^{-6}$, the null shape largely remains constant where the null intersection on the star-planet axis is well predicted by the analytic prescription (Equation 4.1). The three cases of $q = 10^{-2}$ demonstrate how the behaviour of the null changes as $q \to 1$. In the case of $s_A = 1.16$, the null is split into two disconnected segments inside and outside of the caustic, where the analytic prediction is close to their mean location. For $s_A = 0.95$, the discrepancy from the analytic prediction may be attributed to the curvature of the null near the star-planet axis.

Chapter 5

A Mathematical Treatment of the Offset Microlensing Degeneracy

The offset microlensing degeneracy, recently proposed by Zhang et al. (2022) (Chapter 4), has been shown to generalize the close-wide and inner-outer *caustic* degeneracies into a unified regime of *magnification* degeneracy in the interpretation of 2-body planetary microlensing observations. While the inner-outer degeneracy expects the source trajectory to pass equidistant to the planetary caustics of the degenerate lens configurations, the offset degeneracy states that the same mathematical expression applies to any combination of the close, wide, and resonant caustic topologies, where the projected star-planet separations differ by an offset ($s_{\rm A} \neq s_{\rm B}$) that depends on where the source trajectory crosses the star-planet axis. An important implication is that the $s_{\rm A} = 1/s_{\rm B}$ solution of the close-wide degeneracy never strictly manifests in observations except when the source crosses a singular point near the primary. Nevertheless, the offset degeneracy was proposed upon numerical calculations, and no theoretical justification was given. Here, we provide a theoretical treatment of the offset degeneracy, which demonstrates its nature as a mathematical degeneracy. From first principles, we show that the offset degeneracy formalism is exact to zeroth-order in the mass ratio (q) for two cases: when the source crosses the lens-axis inside of caustics, and for $(s_{\rm A} - s_{\rm B})^6 \ll 1$ when crossing outside of caustics. The extent to which the offset degeneracy persists in oblique source trajectories is explored numerically. Finally, it is shown that the superposition principle allows for a straightforward generalization to N-body microlenses with $N-1$ planetary lens components ($q \ll 1$), which results in a 2^{N-1}-fold degeneracy.

5.1 Introduction

Photometric observations of planetary microlensing events are commonly subject to a 2-fold-degenerate interpretation where the projected planet location differs ($s_A \neq s_B$) but the planet-to-star mass ratio remains the same ($q_A = q_B$). The close-wide degeneracy (e.g., Griest & Safizadeh 1998; Dominik 1999; An 2005) is commonly invoked for such events with

source stars passing close to the central caustic, while the inner-outer degeneracy (Gaudi & Gould 1997; Han et al. 2018) is cited for events which have source stars passing close to the planetary caustic. The close-wide degeneracy arises from the invariance of the shape and size of the central caustic under the $s \leftrightarrow 1/s$ transformation for $|1-s| \gg q^{1/3}$, a condition which is equivalent to the lens system being far from the resonant regime (An 2021). The inner-outer degeneracy arises from the Chang-Refsdal (Chang & Refsdal 1979) approximation to the planetary caustics (Gaudi & Gould 1997; Dominik 1999), which describes a point-mass lens with uniform shear. Chang-Refsdal caustics are symmetric both along the star-planet axis (referred to as the lens axis hereafter), and along the line perpendicular to the star-planet axis that runs through the center of the caustic.

Recently, Yee et al. (2021) and Zhang et al. (2022) (Chapter 4) noted various inconsistencies of the two aforementioned degeneracies with those seen in real and simulated events. Yee et al. (2021) noted the large number of semi-resonant topology events that cite the close-wide degeneracy, for which the degenerate solutions do not exactly follow $s \leftrightarrow 1/s$ nor satisfy $|1 - s| \gg q^{1/3}$. They went on to suggest that there may be a continuum between the close-wide and inner-outer degeneracies in the resonant regime. Subsequently, Zhang et al. (2022) (Chapter 4) pointed out that the $s \leftrightarrow 1/s$ relationship is also not exactly followed even within the $|1-s| \gg q^{1/3}$ regime in which the close-wide degeneracy is expected to hold. They pointed out that the close-wide and inner-outer degeneracies are fundamentally *caustic* degeneracies which do not necessarily translate to *magnification* degeneracies that manifest in light-curves.

The offset degeneracy (Zhang et al. 2022; Chapter 4) is then proposed independently of caustics as a magnification degeneracy, which both relaxes the non-resonant condition ($|1 - s| \gg q^{1/3}$) and resolves the aforementioned inconsistencies. A key observation in the offset degeneracy is that for two planetary ($q \ll 1$) lenses that differ only by an *offset* to the projected star-planet separation ($s_A \neq s_B$) on the same lens-axis, their locus of equal magnification — referred to as the *null* — intersects with the lens-axis at

$$\xi_{\rm null,0} = \frac{s_A - 1/s_A + s_B - 1/s_B}{2}, \tag{5.1}$$

where the subscript "0" indicates to zeroth-order in q, which we prove to be the correct form in Section 5.2. The intersection between the null and the lens-axis is referred to as the *lens-axis null* hereafter as a shorthand. Given that planetary anomalies primarily occur on and near the lens-axis, source trajectories crossing the lens-axis null

$$\frac{u_0}{\sin(\alpha)} = \xi_{\rm null,0} \tag{5.2}$$

are then expected to result in similar light-curves under the null-forming lens configurations. In the above equation, $u_0/\sin(\alpha) \equiv u_{\rm anom}$ is where the source crosses the lens-axis, which is usually also the source-star separation around the midpoint of the planetary anomaly, u_0 is the impact parameter to the coordinate origin (see Section 5.2.1 for detailed considerations), and α is the angle between the source trajectory and the lens axis.

Crucially, the above formalism is continuous over caustic topology transitions for $q \ll 1$, and thus generalizes the close-wide and inner-outer degeneracies to the resonant regime. One major implication is that the close-wide degeneracy only strictly manifests for the singular case of $u_0 = 0$, and elsewhere the offset degeneracy predicts a deviation from $s \leftrightarrow 1/s$. We thus refer to the close-wide degeneracy as the central caustic degeneracy, in line with An (2021). While Zhang et al. (2022) (Chapter 4) verified that the above formalism accurately describes the degenerate solutions in 23 observed events in the referred literature, it was found numerically and no theoretical justification was given. Subsequently, an alternative formalism for the unification of degeneracies was proposed in Gould et al. 2022, whose the relationship to the offset degeneracy will be discussed in Section 5.5.

In this work, we provide a mathematical treatment of the offset degeneracy. In Section 5.2, the location of the lens-axis null is derived from the lens equation, which proves the formalism proposed in Zhang et al. (2022) (Chapter 4). In Section 5.3, conditions on the source trajectory orientation is discussed. Finally, a generalized N-body offset degeneracy based on the superposition principle is discussed in Section 5.4, whereas Section 5.5 concludes our work.

5.2 Derivations

The goal of this section is to answer the question: given two planetary lenses with the same mass-ratio ($q_{\rm A} = q_{\rm B} \ll 1$) but different projected star-planet separations ($s_{\rm A} \neq s_{\rm B}$), where on the lens axis does their magnifications equal?

Let us begin by defining the lens equation. With the primary star on the origin and the planet on the real-axis at a distance s from the primary, the two-body complex lens equation (Witt 1990) states

$$\zeta = z - \frac{1-m}{\bar{z}} - \frac{m}{\bar{z}-s}, \tag{5.3}$$

where $\zeta = \xi + i\eta$ and $z = z_1 + iz_2$ are the complex source and image locations, m is the planetary mass normalized to the total lens mass ($M_{\rm tot}$), and s is the projected star-planet separation normalized to the angular Einstein radius $\theta_E = \sqrt{4GM_{\rm tot}/(D_{\rm rel}c^2)}$ where $D_{\rm rel}$ is the source-lens relative distance defined as $D_{\rm rel}^{-1} = D_{\rm lens}^{-1} - D_{\rm source}^{-1}$.

Witt & Mao (1995) showed that the lens equation can be transformed into a 5th-order polynomial in z by substituting the conjugate of Equation 5.3,

$$\bar{z} = \bar{\zeta} + \frac{1-m}{z} + \frac{m}{z-s}, \tag{5.4}$$

back into itself, whereby conjugates in $\bar{z}$ are cleared. The resulting polynomial is

$$p_5(z;\zeta,m,s) = \sum_{i=0}^{5} a_i(\zeta,m,s)\cdot z^i = 0, \tag{5.5}$$

where

$$\begin{aligned}
a_0 =&(1-m)^2 s^2 \zeta \\
a_1 =&(1-m)s[ms-(2+s^2)\zeta+2s\zeta\bar{\zeta})] \\
a_2 =&\zeta+2s^2\zeta-ms(1+s\zeta) \\
&-s(s-2ms-2(m-2)\zeta+s^2\zeta)\bar{\zeta}+s^2\zeta\bar{\zeta}^2 \\
a_3 =&-s(ms+\zeta)+(-2(m-1)s+s^3+2\zeta+2s^2\zeta)\bar{\zeta} \\
&-s(s+2\zeta)\bar{\zeta}^2 \\
a_4 =&ms-(1+2s^2+s\zeta)\bar{\zeta}+(2s+\zeta)\bar{\zeta}^2 \\
a_5 =&(s-\bar{\zeta})\bar{\zeta}.
\end{aligned}$$

The magnification of each individual image j located at z_j is given by the absolute value of the inverse of the Jacobian determinant of the lens equation:

$$\mu_j = \frac{p_j}{\det J|_{z=z_j}} \tag{5.6}$$

$$= p_j \left(1-\frac{\partial\zeta}{\partial\bar{z}}\frac{\overline{\partial\zeta}}{\partial\bar{z}}\right)^{-1}\Bigg|_{z=z_j}, \tag{5.7}$$

where $p_j = \pm 1$ denotes the parity of the image.

Witt & Mao (1995) further demonstrated how one may acquire the individual image magnifications μ_j without solving for the image locations z_j. Evaluating $\partial\zeta/\partial\bar{z}$ with Equation 5.3, clearing conjugates in z with Equation 5.4, and clearing fractions, one obtains a 8th-order polynomial in z whose coefficients are parameterized by μ_j. From here on, let us restrict our discussion to the lens-axis, i.e., the real-axis ($\zeta = \xi$). The common variable z in this 8th-order polynomial and 5th order polynomial associated with the lens equation (Equation 5.5) can be eliminated by calculating their resultant, which results in a lengthy 5th-order polynomial in μ:

$$p_5(\mu;\xi,m,s) = \sum_{i=0}^{5} b_i(\xi,m,s)\cdot\mu^i = 0. \tag{5.8}$$

whose coefficients are parametrized by ξ, m, and s. The above polynomial can be further factored into linear and cubic polynomials:

$$p_5(\mu;\xi,m,s) = \left(\sum_{i=0}^{1} c_i\cdot\mu^i\right)^2\cdot\left(\sum_{i=0}^{3} d_i\cdot\mu^i\right) = 0. \tag{5.9}$$

Of the five solutions μ_j, the equal-magnification solutions ($\mu_1 = \mu_2 = -c_0/c_1$) for the linear equation correspond to the two off-axis images that only exist when the source is inside of a caustic and are positive in parity. The cubic polynomial has three real roots which correspond to three negative parity images ($\mu_{3,4,5} < 0$) when the source is inside of caustics, but one positive and two negative parity images when the source is outside of caustics (Witt & Mao 1995). Let us now consider these two cases separately.

5.2.1 Inside Caustics

When the lens-axis null — the intercept of the locus of equal magnification on the lens axis — is located inside of caustics (Figure 5.1), images for each of the two polynomials in Equation 5.9 are respectively equal in parity and the total magnification can be derived directly from the polynomial coefficients:

$$\begin{aligned}\mu_{\rm tot,in}(\xi, m, s) =& (\mu_1 + \mu_2) - (\mu_3 + \mu_4 + \mu_5) \\ =& -2c_0/c_1 + d_2/d_3 \\ \mu_{\rm tot,in}(\xi, m, s) =& \frac{3m^2s^2 - \xi^2A^2 + 2msB}{m^2s^2 + \xi^2A^2 - 2ms\xi C},\end{aligned} \tag{5.10}$$

where

$$\begin{aligned}A =& 1 - s^2 + s\xi \\ B =& -2s + (1 + s^2)\xi - 3s\xi^2 + 2\xi^3 \\ C =& 1 + s^2 - 3s\xi + 2\xi^2.\end{aligned}$$

The location of the lens-axis null can be derived by solving $\mu_{\rm tot,in}(s_A) = \mu_{\rm tot,in}(s_B)$. Since for planetary microlenses $m \ll 1$, the m^2s^2 term can be dropped in both the numerator and the denominator, and we can substitute the planet-to-star mass ratio $q = m/(1 - m)$ for m. Clearing fractions in $\mu_{\rm tot,in}(s_A) - \mu_{\rm tot,in}(s_B) = 0$, we obtain a quadratic polynomial in ξ. Taking the zeroth-order Taylor expansion in q, one of the roots simplifies to

$$\xi_{\rm null,in} = \frac{s_A - 1/s_A + s_B - 1/s_B}{2} + \mathcal{O}(q), \tag{5.11}$$

where the other root is reduced to 0. We have thus shown that the empirically derived $\xi_{\rm null,0}$ (Equation 5.1) is exact for null-in-caustic to zeroth-order in q.

To see how $\xi_{\rm null,in}$ may deviate from the zeroth-order term ($\xi_{\rm null,0}$) for finite value of q, let us now consider the first-order term in q and its dependence on $s_{\rm A,B}$. In particular, for $s_{\rm A} = 1/s_{\rm B}$, we should expect the first-order term to not diverge to infinity in the $s_{\rm A,B} \to \{0, \infty\}$ limit, in order to be consistent with the central caustic degeneracy. Here, it is important to adapt a coordinate origin that is consistent with caustic degeneracies. An (2021) noted that while the central caustic degeneracy breaks down near the resonant regime, a pair of resonant caustics with $s_{\rm A} = 1/s_{\rm B}$ still resembles each other locally towards the back end of the caustic (near the primary star). This suggests that one should choose a coordinate origin that consistently aligns the back-end of the central/resonant caustic for a pair of lenses with an arbitrary difference in separation ($s_{\rm A,B}$).

We therefore opt to use the effective primary star location (Di Stefano & Mao 1996; An & Han 2002; Chung et al. 2005) as the coordinate origin, which is given by

$$\xi \to \xi + \frac{q}{(1 + q) \cdot (s + s^{-1})}, \tag{5.12}$$

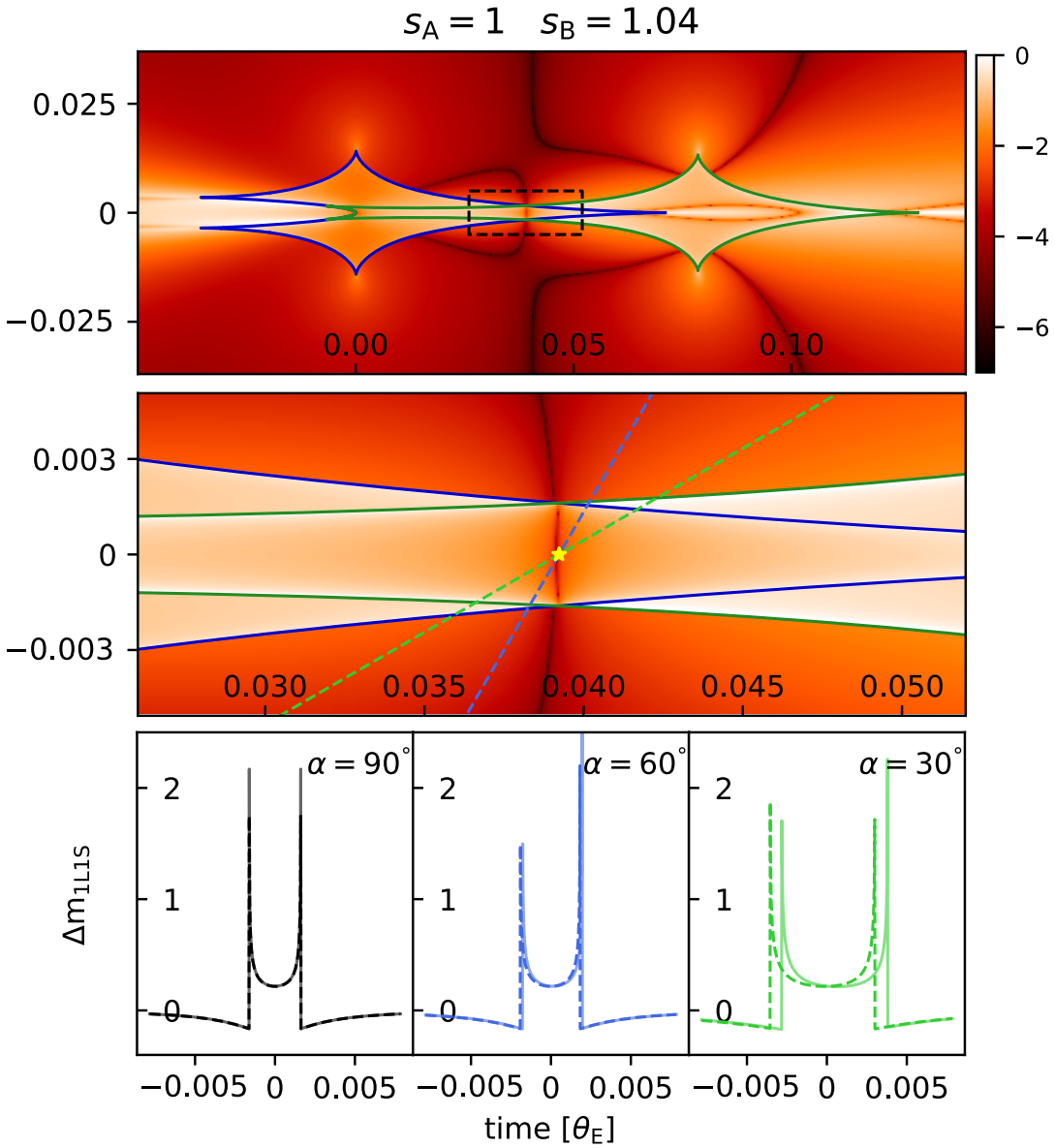


Figure 5.1: Top: fractional magnification difference between ($s_A = 1$, $q = 10^{-4}$) and ($s_B = 1.04$, $q = 10^{-4}$), with color-scale shown to the right in $\log_{10}$. Black contours illustrate the locus of equal magnification. The x and y axes are in units of θ_E. Middle: a zoom-in of the dashed-line boxed region in the top panel. The location of the lens-axis null expected from $\xi_{\rm null,0}$ is marked with the gold star in the center. Source trajectories with $\alpha = 30°, 60°$ are shown in green and blue dashed lines. Bottom: differences to single-lens light-curves for null-crossing trajectories. Dashed lines corresponds to $s_A = 1$ whereas solid lines are for $s_B = 1.04$. Trajectory orientation is marked in the subplot upper-right corners with the same color coding as the middle plot. The $\alpha = 30°$ case is seen to have different caustic entry-exit times but similar caustic-crossing durations.

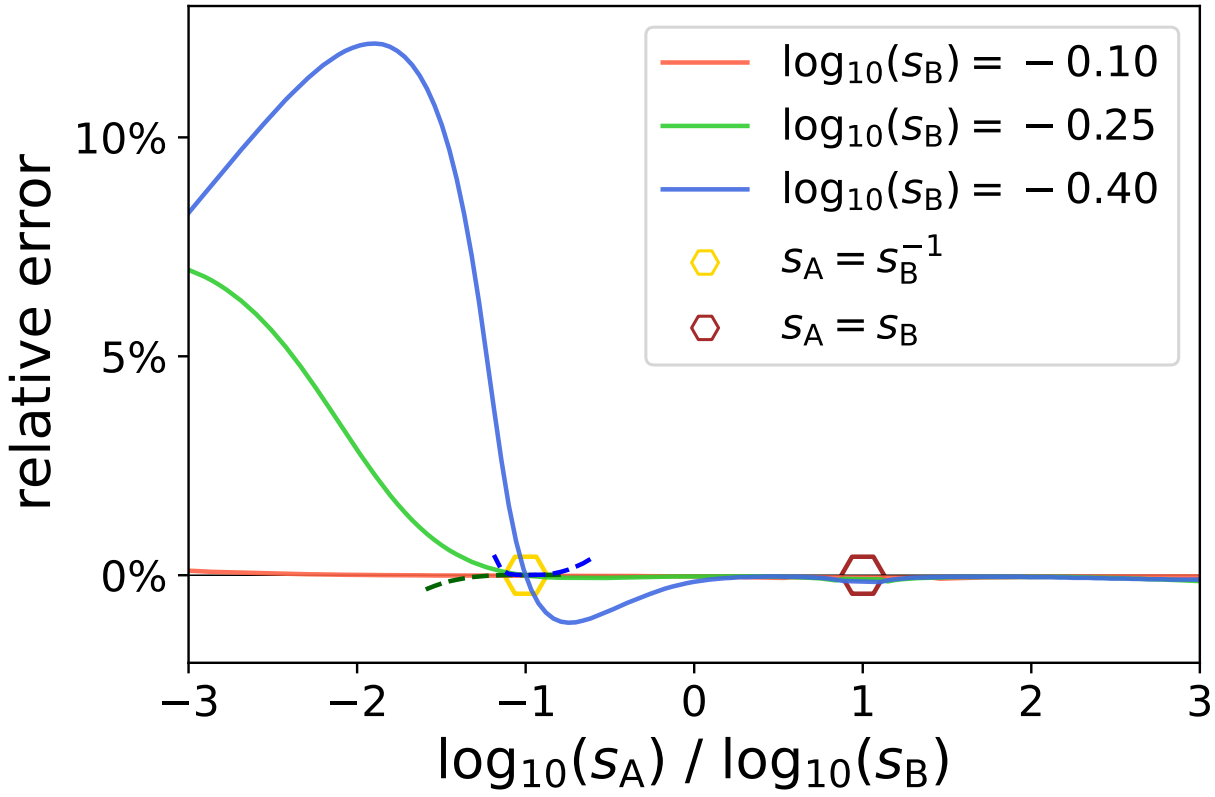


Figure 5.2: Deviation of $\xi_{\rm null,0}$ from the exact null location, normalized to $|(s_A - 1/s_A) - (s_B - 1/s_B)|$, where the exact null location is derived numerically with $q = 10^{-4}$. Three solid curves show this relative error for changing s_A against three values of fixed $s_B \simeq (1/1.3, 1/1.8, 1/2.5)$. The two dashed lines with darker colors show the alternative expression $\xi_{\rm null,hm}$ which is exact for $\xi_{\rm null} \ll 1$ (see Section 5.2.2), or equivalently $s_{\rm A} \sim 1/s_{\rm B}$, shown only for $|\xi_{\rm null}| < 0.5$ and $|s_{\rm A} - s_{\rm B}| > 1$.

and indeed achieves the aforementioned alignment. Note that the effective primary location reduces to

$$\xi \to \begin{cases} \xi + sq/(1+q) & s \ll 1 \\ \xi + s^{-1}q/(1+q) & s \gg 1, \end{cases}$$

which are the central caustic locations (Han 2008) that were used in Zhang et al. (2022) (Chapter 4) as the coordinate origin for their numerical calculations. We point out that the $\sim 2\%$ error at $s_{\rm A} = 1$ and $s_{\rm B} = 0.4$ in Figure 2 of Zhang et al. (2022) (Chapter 4) is a direct result of their coordinate choice, which is inaccurate in describing resonant caustic locations and causes a misalignment between the resonant and central caustics. Figure 5.2 reproduces that same figure, but with the effective primary (Equation 5.12) as the origin, and shows that the error of $\xi_{\rm null,0}$ at $s_{\rm A} = 1$ and $s_{\rm B} = 0.4$ is reduced to 0.1% and remains $< 0.1\%$ for $|\log(s_{\rm A,B})| < 0.25$, or $1/1.8 < s_{\rm A,B} < 1.8$.

Applying the above coordinate transformation to the previous derivation, we find that while the zeroth-order term remains $\xi_{\rm null,0}$ as expected, the first-order term $(f \cdot q)$ is rather involved. There are only two special cases that are relevant here.

If the null is located within the central caustic, we should expect $s_A \sim 1/s_B$, which simplifies the first order term $f \cdot q$ to

$$f \sim -\frac{s(3+2s^2+3s^4)}{(1+s^2)^3}. \tag{5.13}$$

Note that the above expression is symmetrical under $s \leftrightarrow s^{-1}$. Since $f \to 0$ for $s \to \{0, \infty\}$, f does not diverge and is typically of order unity. However, if we had defined the lens-equation (Equation 5.3) in units of the Einstein radius of the primary mass, then f diverges to infinity for both $s \to \{0, \infty\}$, justifying our choice of parameterization with the Einstein radius of the total mass.

On the other hand, if the null is within the resonant or the wide-planetary caustic, we should expect $s_A \simeq s_B \gtrsim 1$, which results in

$$f \sim -\frac{2}{s+s^3}, \tag{5.14}$$

and is also order unity. One may thus expect $\xi_{\rm null,in} \simeq \xi_{\rm null,0} - q$, that is, a deviation of order q, which is in agreement with the slight deviation seen in the middle panel of Figure 5.1.

5.2.2 Outside Caustics

For sources outside caustics (Figure 5.3 & 5.4), there are three images which are different in parity, and we can no longer obtain the total magnification directly from the polynomial coefficients. The sum of the absolute value of the cubic roots is also difficult to simplify. However, keeping coefficients up to first order in q, the cubic part of Equation 5.9 is reduced to a quadratic polynomial with two roots that are in a much simpler form compared to the cubic roots. The total magnification is then the absolute difference between the two roots representing one positive and one negative parity image. Indeed, when the source is away from the planetary caustic, the image closest to the planet typically has negligible magnification. As for the alternative scenario, we should already expect $\xi_{\rm null,0}$ to hold in the immediate vicinity of planetary caustics, given that the location of the lens-axis null transitions continuously from inside to outside of caustics.

Equating the total magnification for s_A and s_B, clearing fractions, further taking the first order expansion in q and simplifying, we acquire a quartic polynomial

$$p_{\rm null}(\xi; s_{\rm A}, s_{\rm B}) = \sum_{i=0}^{4} e_i(s_{\rm A}, s_{\rm B}) \cdot \xi^i = 0, \tag{5.15}$$

whose coefficients are

$$
\begin{aligned}
e_0 =& -16(s_A s_B - 1)(s_A^2 + s_A s_B + s_A^3 s_B + s_B^2 + s_A^4 s_B^2 + s_A s_B^3 + s_A^3 s_B^3 + s_A^2 s_B^4)\\
e_1 =& -2(s_A + s_B)(3 - 4s_A^2 + s_A^4 - 16 s_A s_B - 4 s_B^2 + 8 s_A^2 s_B^2 - 4 s_A^4 s_B^2 - 16 s_A^3 s_B^3 + s_B^4 - 4 s_A^2 s_B^4 + 3 s_A^4 s_B^4)\\
e_2 =& -4(s_A s_B - 1)(s_A^4 - 3 s_A s_B + 5 s_A^3 s_B + 6 s_A^2 s_B^2 + 5 s_A s_B^3 - 3 s_A^3 s_B^3 + s_B^4)\\
e_3 =& -(s_A + s_B)(1 + s_A^2 - 8 s_A^3 s_B + s_B^2 - 14 s_A^2 s_B^2 + s_A^4 s_B^2 - 8 s_A s_B^3 + s_A^2 s_B^4 + s_A^4 s_B^4)\\
e_4 =& 2 s_A s_B (s_A s_B - 1)(1 + s_A^2 + 2 s_A s_B + s_B^2 + s_A^2 s_B^2).
\end{aligned}
$$

This polynomial could be solved for the lens-axis null outside of caustics for any arbitrary pair of $s_{\rm A,B}$ satisfying $q \ll 1$.

To examine the conditions for $\xi_{\rm null,0}$ to be the exact form to zeroth-order in q, let us directly plug $\xi_{\rm null,0}$ into $p_{\rm null}$ as an ansatz, which reduces the polynomial to

$$
-\frac{(s_A - s_B)^6 (s_A s_B - 1)(s_A s_B + 1)^2}{4 s_A^2 s_B^2} = \mathcal{O}\big((s_A - s_B)^6\big). \tag{5.16}
$$

Given non-zero first order derivative $p'_{\rm null}$ and bounded higher order derivatives, $p_{\rm null} \to 0$ implies $\xi \to \xi_{\rm null,0}$, that is, the ansatz is indeed a root. Thus $\xi_{\rm null,0}$ is exact for $(s_{\rm A} - s_{\rm B})^6 \ll 1$ to zeroth-order in q. Note that this condition is substantially more relaxed than the $|s_{\rm A} - s_{\rm B}| \ll 1$ condition (e.g., $0.5^6 \simeq 0.015$). Furthermore, the condition of the lens being near the resonant regime ($|1 - s| \lesssim q^{1/3}$) is a sufficient condition for $(s_{\rm A} - s_{\rm B})^6 \ll 1$, allowing $\xi_{\rm null,0}$ to be essentially exact for semi-resonant events.

Numerical calculations (Figure 5.2) show that the error on $\xi_{\rm null,0}$ remains less than 1% for $1/2.5 < s_{\rm A,B} < 2.5$ and should be sufficiently accurate for practical purposes. Larger deviations of a few percent are found near $s_{\rm A} \sim 1/s_{\rm B}$ where $|s_A - s_B| \gtrsim 3$. As a theoretical exercise, an alternative expression for these high-magnification ($\xi_{\rm null} \ll 1$) events can be immediately acquired by linearizing $p_{\rm null}$ in $\xi_{\rm null}$, which results in:

$$
\xi_{\rm null,hm} = -e_0/e_1, \tag{5.17}
$$

where the coefficients can be found in Appendix A. Figure 5.2 shows $\xi_{\rm null,hm}$ for $|\xi_{\rm null}| < 0.5$ (dashed lines), which verifies that $\xi_{\rm null,hm}$ indeed describes the local behavior at $s_A \sim 1/s_B$.

5.3 Source trajectory orientation

Technically, the above derivation only guarantees exact magnification matching on the lens-axis. It was shown in Zhang et al. (2022) (Chapter 4) that vertical null-crossing trajectories result in nearly identical light-curves, which was also noted in Gaudi & Gould (1997) for the inner-outer degeneracy. Indeed, Figures 5.1, 5.3, 5.4 all demonstrate that the locus of equal magnification is vertically extended near the lens-axis. Here, we consider the extend to which oblique trajectories could remain degenerate.

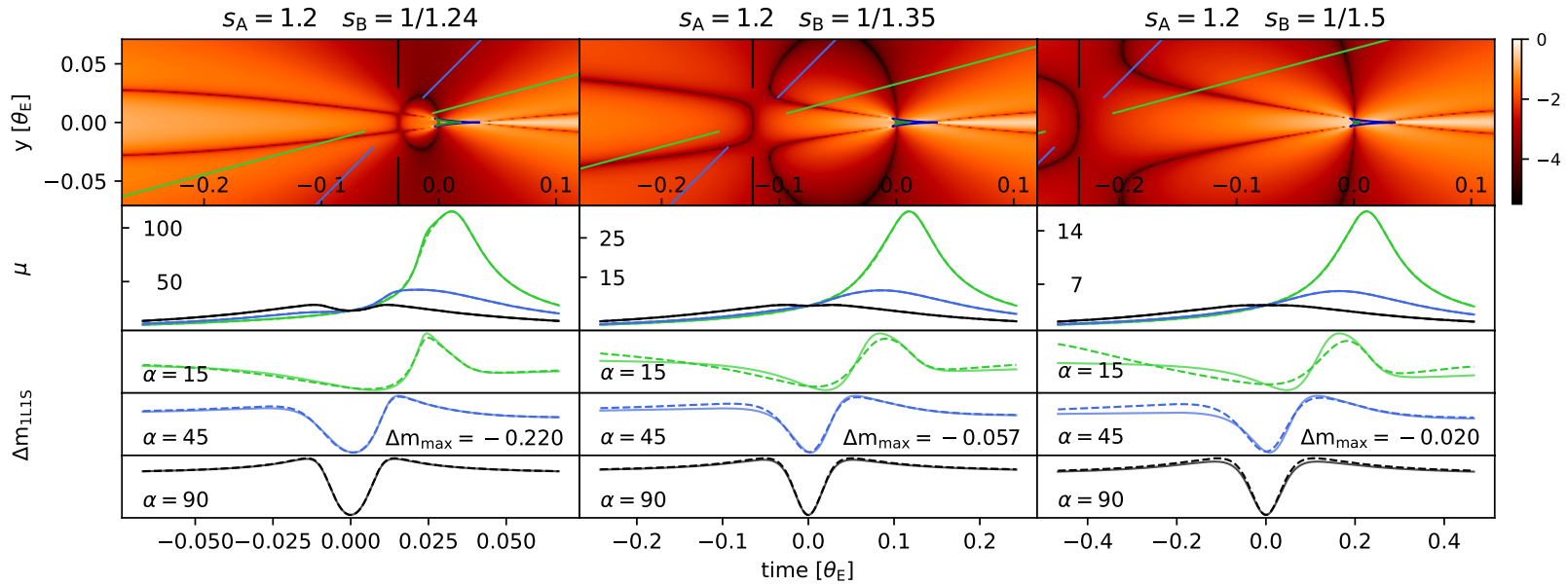


Figure 5.3: Top row: magnification difference in log-scale for three pairs of lens configurations indicated in the subplot titles. $q = 10^{-3}$ for all cases. Color-bar to the right shows the difference scale in $\log_{10}$. The oval-shaped contours are the loci of equal magnification (null). Three null-crossing source trajectories with $\alpha = 15°, 45°, 90°$ are shown with the two-segment solid lines, with direction going from upper-right to lower-left. The green central caustics are for the changing s_B. Second row: magnifications (μ) for null-crossing trajectories in the same color coding as the top row. Solid lines are for s_A and dashed lines for s_B. The x-axis (time) is centered on the lens-axis null and scaled to $|\xi_{\rm null}|$. Bottom three rows: planetary perturbation shown as the difference to a single lens model in unit of magnitudes. The maximum deviation is indicated in the second-to-last row.

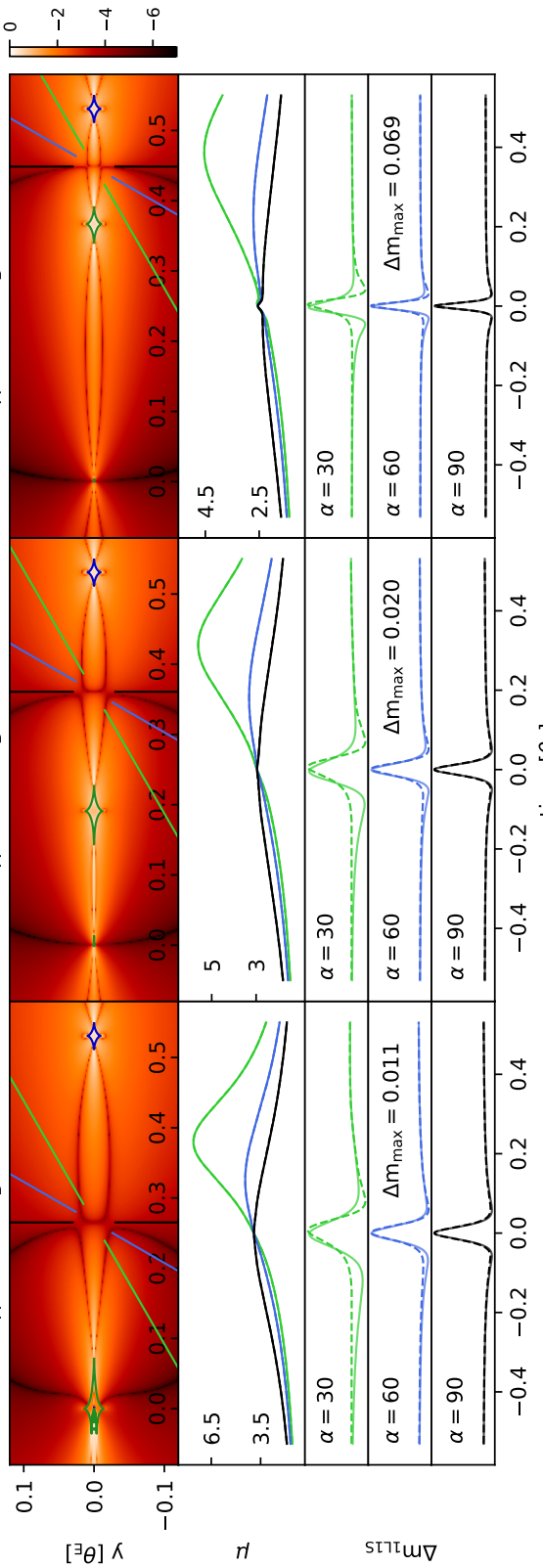


Figure 5.4: Same as Figure 5.3 but for three different configurations.

Let us first consider the case where the lens-axis null is located outside of caustics. Figure 5.3 shows three examples where the null gradually moves away from the central caustic. Figure 5.4 shows three additional cases where s_B approaches s_A from $s_B = 1$. Note how in Figure 5.4 $|\xi_{\rm null}|$ is greater than the examples in Figure 5.3. In both cases, vertical trajectories essentially give rise to identical light-curves. As the trajectory becomes more oblique, the magnifications under the two degenerate lenses begin to differ in the "wings" of the planetary perturbation, and thus sufficiently precise photometry can break the degeneracy. By comparing Figure 5.3 and 5.4, one may see that the trajectory angle can be as oblique as $\alpha = 15^\circ$ while the light-curves remain largely the same when the null is close to the central caustic ($|\xi_{\rm null}| \ll 1$). Elsewhere, the differences on the perturbation "wings" become a significant fraction of the peak planetary perturbation for $\alpha \lesssim 45^\circ$. While not shown, close approaches to the off-axis cusps of the planetary caustic with oblique trajectories will decisively break the degeneracy, as the time-of-approach will be either before or after crossing the lens-axis.

For the lens-axis null inside of caustics, there is notably an additional constraint on the caustic entry-exit times and duration. Figure 5.1 illustrates how the vertical null directionality implies that the caustic height is automatically matched at the lens-axis null, allowing the caustic entry-exit times and duration to be the same for vertical null-crossing trajectories. Essentially, intersections of caustics are the set of points in the source plane where magnifications for the two lenses diverge simultaneously, and by definition, must occur on the locus of equal magnification.

For oblique trajectories, note how the two resonant caustics are approximately the reflection of one another along the vertical null (black broken line in Figure 5.1) and appears like large planetary caustics. Because of this symmetry, the caustic-crossing duration remains approximately the same, but the caustic entry-exit times begin to differ, the extent of which depends on how quickly the caustic height changes ($\mathrm{d}\eta_{\rm caus}/\mathrm{d}\xi_{\rm caus}|_{\xi=\xi_{\rm null,0}}$) near the lens-axis null. Fine tuning of the lensing parameters (e.g., the event timescale) may reduce the difference in the caustic entry-exit times. Additionally and similarly to non-caustic-crossing events, close approaches to the off-axis cusps (not shown in Figure 5.2) will be asymmetrical for oblique trajectories and would categorically break the degeneracy. Finally, for the lens-axis null inside of central caustics ($|1 - s| \ll q^{1/3}$), the central caustics are close to identical due to the central caustic degeneracy and thus the aforementioned constraints on the caustic entry-exit times are less relevant.

Recent examples in the literature of caustic-crossing offset-degenerate events include, among others, KMT-2019-BLG-0371 (Kim et al. 2021), KMT-2019-BLG-1042 (Zang et al. 2022), and OGLE-2019-BLG-0960 (Yee et al. 2021). In the case of OGLE-2019-BLG-0960, the trajectory was quite oblique ($\alpha \simeq 15$), yet still resulted in very degenerate solutions because the caustic height in this particular case changes slowly near the null ($|\mathrm{d}\eta_{\rm caus}/\mathrm{d}\xi_{\rm caus}|_{\xi=\xi_{\rm null,0}} \ll 1$), allowing the caustic entry-exit times to remain approximately the same even for very oblique trajectories.

5.4 Generalization to N-body lens

The superposition principle (Bozza 1999; Han et al. 2001) states that planetary perturbations from an N-body lens satisfying $q_i \ll 1$ is well approximated by the superposition of perturbations from each individual planet. This allows a straightforward generalization of the offset degeneracy to N-body lenses, which has $N-1$ number of lens-axes, and thus the number of null to match, resulting in a 2^{N-1} number of degenerate configurations.

Figure 5.5 shows an example of the offset degeneracy generalized to triple lens systems, where the source passes close to the back end of the self-intersecting central caustics. We have adapted the same configuration in Figure 2 of Song et al. (2014) to facilitate comparison to the extension of the central caustic degeneracy to triple-lens discussed therein. The magnification difference between the wide/wide and close-close configurations is shown to be the sum of the residuals from the two singly-offset (close/wide and wide/close) configurations, which confirms the superposition picture. Additionally, as expected the 3-body offset degeneracy also serves as a correction to the 3-body central caustic degeneracy. The light-curve difference between the close/close and wide/wide configurations is greater near the null on the horizontal lens-axis (s_1) than the other because the source crosses the horizontal axis at $\alpha = 30$ but $\alpha = 90$ for the s_2 axis. This is in agreement with discussions in Section 5.3.

Interestingly, a detailed inspection of Figure 5.5 reveals that the central caustic cusps at the 'tips' of the central caustics are actual slightly off the two lens-axes, which can be attributed to the influence of one planet on the other's caustic. This indicates that technically one may have to apply the source-null matching principle to an "effective lens axis." Moreover, the superposition principle is expected to break down when the planets are close to being aligned on the same axis. Indeed, for a triple lens for which the two planets are aligned on the same axis, there is only one null that depends on the offset of both planets. We suggest that the simplest case of the axis-aligned triple planetary lens with equal mass-ratios may be analytically tractable by studying the following lens equation:

$$\zeta = z - \frac{1-2m}{\bar{z}} - \frac{m}{\bar{z}-s_1} - \frac{m}{\bar{z}-s_2}. \tag{5.18}$$

Details of the generalized N-body offset degeneracy should be explored in future work.

5.5 Discussion

In this work, we have provided a mathematical treatment of the offset degeneracy by deriving the intercept of the equal-magnification locus on the lens-axis — the lens-axis null — directly from the lens-equation in the limit of $q \ll 1$. The numerically found $\xi_{\rm null,0}$ expression (Zhang et al. 2022; Chapter 4) is shown to be the exact form of the lens-axis null location inside of caustics, and outside of caustics subject to $(s_{\rm A} - s_{\rm B})^6 \ll 1$, to zeroth-order in q. The derivations in this work demonstrate the nature of the offset degeneracy as a mathematical degeneracy deeply rooted in the lens equation itself.

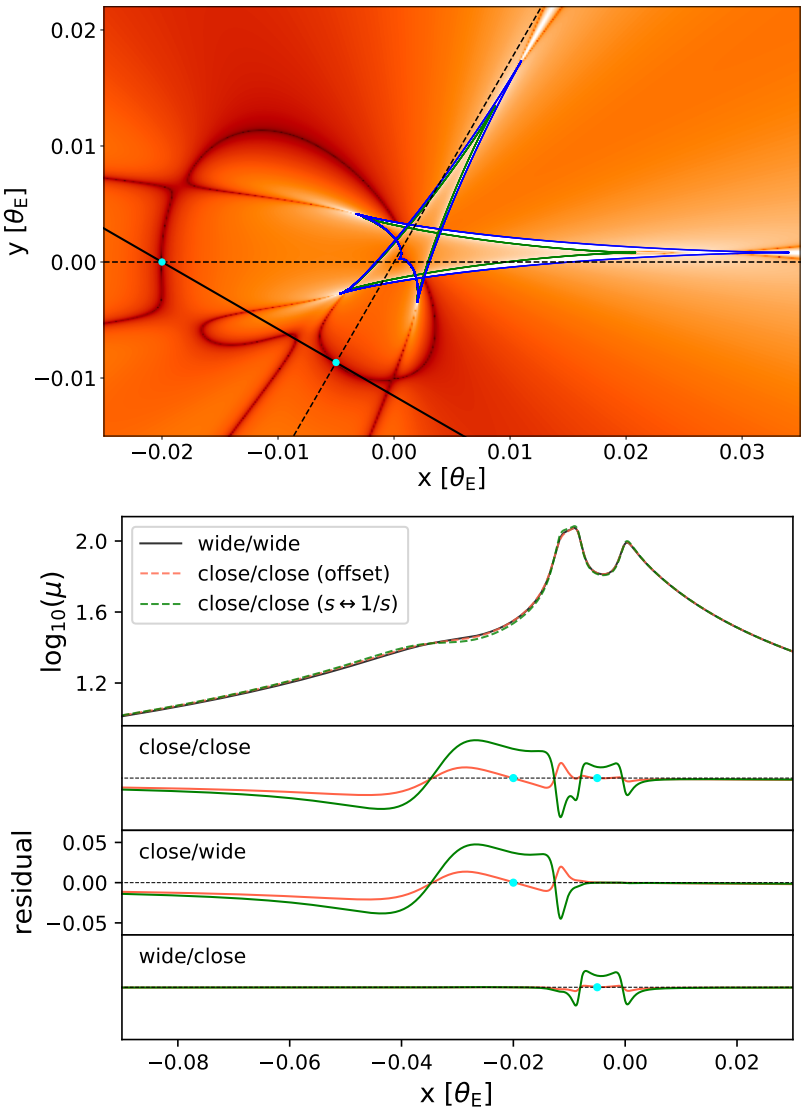


Figure 5.5: Example of the offset degeneracy generalized to triple lens systems. Top: magnification difference between triple lens configurations of (s_1, s_2, ϕ)=(1.2, 1.25, 60°), referred to as the wide/wide configuration whose central caustic is shown in blue, and the close/close configuration of (0.8189, 0.7938, 60°) whose central caustic is shown in green. ϕ is the angle between the two lens-axes (dashed lines), with the horizontal one corresponding to s_1. The two resulting lens-axis nulls are marked with cyan dots, which coincide with the source trajectory (solid line). Bottom: light-curves for the null-crossing trajectory. In the legend, $s \leftrightarrow 1/s$ refers to the (1/1.2, 1/1.25, 60°) configuration expected from the central caustic degeneracy. The designations "close" and "wide" refer to the caustic topology rather than the close-wide degeneracy. The bottom panels show light-curve residuals of the degenerate configurations to the wide/wide configuration in units of magnitudes. Light-curves resulting from the central caustic degeneracy (green curves) are shown to have greater residual than that from the offset degeneracy (red curves). The horizontal axis is the source location projected to the x-axis and the cyan dots indicate the nulls allowing for a straightforward comparison to the top figure.

The relationship between the offset degeneracy and the central caustic (close-wide) and inner-outer degeneracies has been discussed in Zhang et al. (2022) (Chapter 4). To summarize, the offset degeneracy relaxes the non-resonant ($|1 - s| \gg q^{1/3}$) condition required by the two caustic degeneracies and generalizes them to a unified regime of magnification degeneracy. For sources passing close to central caustics, the offset degeneracy serves as a correction to the $s \leftrightarrow 1/s$ relationship of the central caustic degeneracy, which only strictly manifests when $u_0 = 0$. For this reason, we advocate that the close-wide degeneracy should be more appropriately referred to as the central caustic degeneracy (e.g., An 2021), which also serves to discourage its misuse as a magnification degeneracy.

On the other hand, the inner-outer degeneracy expects the source star to pass equidistant to the planetary caustics located at $\xi_p = s_{\rm A,B} - 1/s_{\rm A,B}$, and thus results in the same mathematical expression as the offset degeneracy. However, the Chang-Refsdal approximation to planetary caustics fails near the resonant regime (Dominik 1999), and thus the offset degeneracy provides a more accurate conceptual explanation. In a subsequent paper, Zhang (2023) (Chapter 6) offered an alternative interpretation by showing how planetary lenses can be decomposed into Chang-Refsdal lenses with variable shear, which results in the offset degeneracy as a direct consequence. While the terms inner and outer were originally coined to refer to "the inner[/outer] region of the planetary caustic with respect to the planet host" (Han et al. 2018), the idea of a generalized perturbative picture (Zhang 2023) (Chapter 6) suggests that they remain meaningful labels for the offset degeneracy if they refer to the lens-plane instead — the location of the planet being inside or outside of the image being perturbed, with respect to the primary star.

The applicability of the central caustic degeneracy to the resonant regime was previously studied in An (2021), which found that the back-end of the central/resonant caustic remains locally degenerate into the resonant regime ($|1 - s| \lesssim q^{1/3}$) but the front end becomes different. They further suggested that in this case, slight adjustments to the $q_A = q_B$ and $s_A = 1/s_B$ pair of solutions may result in a locally degenerate model. This work directly responds to their suggestion: $q_A = q_B$ should remain the same whilst $s_{A,B}$ should be adjusted such that the location of the lens-axis null coincides with the source trajectory. Strictly speaking, the $q_A = q_B$ condition is an assumption made in this work which is known to be true for the caustic degeneracies. The fact that vertical trajectories give rise to identical light-curves (Figures 5.1, 5.3, 5.4) validates the $q_A = q_B$ assumption, but a formal proof would require examining the magnification off the lens-axis.

While examining the magnification-matching behavior on the lens-axis is a direct way of deriving the offset degeneracy formalism, there is a potential pathway to derive the $\xi_{\rm null,0}$ formalism for the null-in-caustic case by studying caustic resemblances, which was proposed by An (2021). In Section 5.3, we found that the caustic height for the offset-degenerate pair of lenses matches exactly at the lens-axis null, but such a claim is based on the observation that the null is vertically-directed near the lens-axis. Therefore, studying the intersection between caustics of lenses with equal mass-ratios may be not only be an independent pathway to deriving the offset degeneracy formalism, but also a verification of the equal mass-ratio condition.

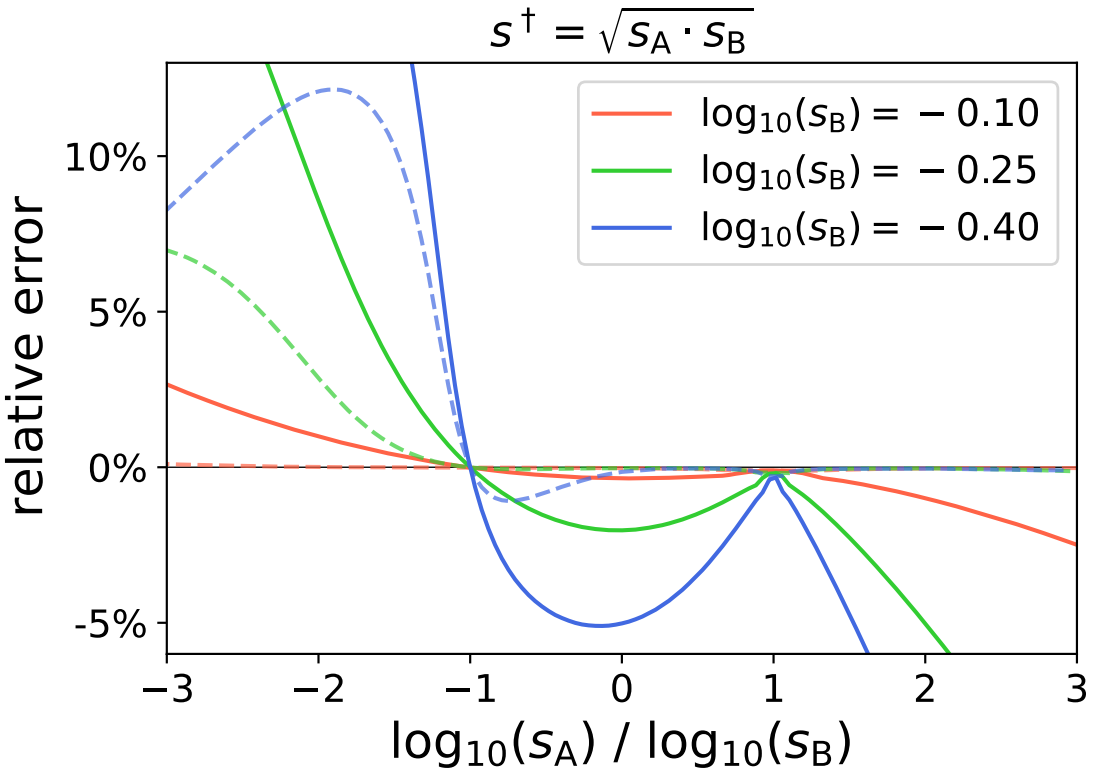


Figure 5.6: Error on the $s^\dagger = \sqrt{s_A \cdot s_B}$ heuristic, defined as the difference between the predicted value of $u_{\rm anom} = s^\dagger - 1/s^\dagger$ from $s_{A,B}$, and the exact location of equal magnification on the lens-axis. Solid curves are for the $s^\dagger$ heuristic and dashed curves are for the offset degeneracy ($u_{\rm anom} = \xi_{\rm null,0}$) for comparison. Quantities are defined similarly to Figure 5.2.

Subsequent to the proposal of the offset degeneracy, Ryu et al. (2022) and Gould et al. (2022) proposed an alternative formalism for unifying the close-wide and inner-outer degeneracies, referred to as the "$s^\dagger$ heuristic". The quantity $s^\dagger$ is defined by

$$s^\dagger = (\sqrt{u_{\rm anom}^2 + 4} + u_{\rm anom})/2, \tag{5.19}$$

which is a solution to $u_{\rm anom} = s^\dagger - 1/s^\dagger$, and thus the solution for planetary-caustic-crossing events. Here, we have defined $u_{\rm anom}$ as the *signed* location of where the source crosses the binary axis to avoid a sign ambiguity in the original expression. This quantity was initially used in Hwang et al. (2022) for the heuristic analysis of events subject to the inner-outer degeneracy, where the solutions are approximately related by $s_{\rm A,B} = s^\dagger \pm \Delta s$. More recently, Gould et al. (2022) proposed that an alternative expression, $s^\dagger = \sqrt{s_{\rm A} \cdot s_{\rm B}}$, would lead to the unification of the two degeneracies.

The derivations in this work show that the $s^\dagger = \sqrt{s_{\rm A} \cdot s_{\rm B}}$ expression does not correctly unify the close-wide and inner-outer degeneracies, but nevertheless provides approximate solutions in the $s \to 1$ limit. By substituting $\xi_{\rm null,0}$ for $u_{\rm anom}$ in Equation 5.19, we find that the first order Taylor expansion of $(s^\dagger)^2$ at $s_{\rm A,B} = 1$ is indeed $s_{\rm A} \cdot s_{\rm B}$. Figure 5.6 shows that although the $s^\dagger = \sqrt{s_{\rm A} \cdot s_{\rm B}}$ heuristic captures the boundary cases of $s_{\rm A} = 1/s_{\rm B}$ with $s^\dagger = 1$ (and $u_{\rm anom} = 0$), and $s_{\rm A} = s_{\rm B} = s^\dagger$, it is only approximately correct in the intermediate regime. Finally, we note that both the $s^\dagger$ heuristic and the offset degeneracy formalism require solving one quadratic equation to derive one solution from the other based on the source trajectory, which indicates that the exact form given by Equation 5.1 & 5.2 is equally convenient to use for heuristic analysis.

Acknowledgments

A version of this Chapter was published in the Astrophysical Journal Letters as Zhang & Gaudi (2022).

K.Z. is supported by a Gordon and Betty Moore Foundation Data-Driven Discovery grant. K.Z. thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. Work by B.S.G. is supported by NASA grant NNG16PJ32C and the Thomas Jefferson Chair for Discovery and Space Exploration. We thank Joshua Bloom, Shude Mao, and Jin An for helpful discussions, and Joshua Bloom and Shude Mao for comments on a draft of this paper.

Chapter 6

On the Perturbative Picture and the Chang-Refsdal Lens Approximation for Planetary Microlensing

Under the perturbative picture of planetary microlensing, the planet is considered to act as a uniform-shear Chang-Refsdal lens on one of the two images produced by the host star that comes close to the angular Einstein radius of the planet, leaving the other image unaffected. However, this uniform-shear approximation is only valid for isolated planetary caustics and breaks down in the resonant regime. Recently, the planetary-caustic degeneracy arising from the above formalism is found to generalize to the regime of central and resonant caustics, indicating that the perturbative picture and Chang-Refsdal lens approximation may have been under-explored in the past. Here, I introduce a new variable-shear Chang-Refsdal lens approximation, which not only supports central and resonant caustics, but also enables full magnification maps to be calculated analytically. Moreover, I introduce the generalized perturbative picture, which relaxes the required proximity between the planet and the image being perturbed in the previous work. Specifically, the planet always perturbs the image in the same half of the lens plane as the planet itself, leaving the other image largely unaffected. It is demonstrated how this new framework results in the offset degeneracy as a consequence of physical symmetry. The generalized perturbative picture also points to an approach to solve the two-body lens equation semi-analytically. The analytic and semi-analytic microlensing solutions associated with this work may allow for substantially faster light-curve calculations and modeling of observed events.

6.1 Introduction

In the simplest microlensing scenario, a foreground lens star splits a background source star into two images that are located inside and outside the Einstein radius of the lens star,

$$\theta_{\rm E} = \sqrt{\frac{4GM}{D_{\rm rel}c^2}}, \tag{6.1}$$

where G is the gravitational constant, M is the lens mass, c is the speed of light, and $D_{\rm rel}^{-1} = D_{\rm lens}^{-1} - D_{\rm source}^{-1}$ is related to the relative distance between the lens and source. The image outside the Einstein ring is usually referred to as the major image and the inside image as the minor image. The locations of the major/minor images, along with their magnifications, can also be expressed as simple closed-form expressions of the source location.

A two-body lens, on the other hand, splits a source star into either three or five images, depending on whether the source is inside or outside of caustics. The locations of the images are found by solving the lens equation in its complex form (Witt 1990)

$$\zeta = z - \frac{1}{\bar{z}} - \frac{q}{\bar{z} - s}, \tag{6.2}$$

where $\zeta = \xi + {\rm i}\eta$ is the true source location, $z = z_1 + {\rm i}z_2$ is the image location, q is the mass ratio between the two lens components, and s their projected separation in units of the Einstein ring radius of the more massive lens component. The above equation can be transformed into a quintic polynomial that can only be solved numerically. As pointed out in Witt & Mao (1995), the fact that the binary lens equation is not analytically tractable presents a major obstacle in further analytical studies. Additionally, when finite source effects are considered, this quintic polynomial generally has to be solved repeatedly to account for the variance of magnification over the source area, thereby creating a computationally non-trivial problem for the modeling of observed events.

The resemblance between the two-body lens with planetary mass ratios ($q \ll 1$) and the Chang-Refsdal lens has provided one pathway toward analytic studies of planetary microlensing. Both the planetary lens and the Chang-Refsdal lens consist of two components with extreme mass ratios. The Chang-Refsdal lens describes a point-mass lens perturbed by uniform external shear, and was introduced by Chang & Refsdal (1979) to describe the action of an individual star on the outskirts of a massive galaxy acting as a gravitational lens on a background quasar. For a Chang-Refsdal lens, a star lying close to a given quasar image could produce a time-variable magnification to that image due to the relative proper motion between the galaxy and quasar. This has led to the "perturbative picture" of planetary microlensing (Gould & Loeb 1992; Gaudi & Gould 1997), where a planetary-mass body acts as a uniform-shear Chang-Refsdal lens on one of the major/minor images produced by the primary star that comes close to it of order its angular Einstein radius, leaving the other image unaffected.

One advantage of the Chang-Refsdal lens approximation is that the Chang-Refsdal lens equation can be transformed into a quartic polynomial, which is the highest-order polynomial

that can be solved analytically. The Chang-Refsdal lens equation in its complex form is written as

$$\zeta = z - \frac{1}{\bar{z}} + \gamma \bar{z}, \tag{6.3}$$

where γ denotes the shear. This property has allowed for extensive analytic studies of the Chang-Refsdal lens, notably in An & Evans (2006).

However, there are important differences between the planetary lens and the Chang-Refsdal lens, which substantially limit the validity of the approximation of the former by the latter. While the star-galaxy mass ratio for the Chang-Refsdal lens is often below $q = 10^{-12}$, the mass ratio for planetary lenses could range anywhere between $q \sim 10^{-2}$ for Jovian planets and $q \sim 10^{-5}$ for terrestrial planets. The $\sqrt{M}$ scaling (Equation 6.1) of the Einstein ring radius suggests the Einstein radius of the planet may be a substantial fraction of that of the primary star. Thus, the effects of the primary star often can hardly be considered as a uniform background shear over the sphere of influence of the planet. Indeed, Dominik (1999) pointed out that the Chang-Refsdal approximation is only valid for planets sufficiently far from the Einstein ring of the primary star, where the effect of the planetary caustics can be considered in isolation and as a Chang-Refsdal caustic.

Moreover, the appreciable mass ratio of the planetary microlens has allowed for the existence of central and resonant caustics, which are not allowed under the Chang-Refsdal lens formalism. Additionally, planetary caustics in practice are usually elongated towards the host star along the real axis, whereas the Chang-Refsdal caustic is completely symmetrical. There exist other analytical studies that take advantage of the planetary mass ratio ($q \ll 1$), which has led to interesting results (e.g. Bozza 1999, 2000; An 2005). Nevertheless, to date, there has been an absence of an analytic framework for planetary microlensing that holds for all types of caustic topologies. As a result, modeling of current observations still relies on numerically solving the full lens equation, for which optimized quintic solvers have been developed that provide order-unity speed up (Skowron & Gould 2012; Fatheddin & Sajadian 2022) compared to a baseline `ZROOTs` routine from *Numerical Recipes*.

Recent results in microlensing degeneracy indicate that the Chang-Refsdal approximation and the perturbative picture may have been under-explored in the past. Specifically, an important consequence of the Chang-Refsdal approximation is the existence of light-curve degeneracies for planetary caustic perturbations (Gaudi & Gould 1997), commonly referred to as the inner-outer degeneracy (Han et al. 2018). Here, the source trajectory is expected to pass equidistant to the planetary caustics (located at $s - 1/s$) of the degenerate lens configurations, owing to its symmetry under the Chang-Refsdal lens approximation. However, observed degeneracies that reference the inner-outer degeneracy rarely have well-isolated planetary caustics (Yee et al. 2021), although the degenerate light curves often have excellent resemblance. Recently, the offset degeneracy proposed by Zhang et al. (2022) found the equidistance relationship underlying the inner-outer degeneracy to also apply to bi-modal solutions usually attributed to the close-wide degeneracy for central caustics, along with degeneracies involving two resonant topology solutions (cf. Gould et al. 2022).

The condition that the source trajectory shall pass equidistant to the locations $s -$

$1/s$ of the degenerate solutions regardless of the caustic topology was recently proved in Zhang & Gaudi (2022). However, the interpretation of the $s - 1/s$ term as the location of the planetary caustic is rather unsatisfactory, especially for light-curve anomalies primarily associated with central and resonant caustics. Under the perturbative picture, the $s - 1/s$ term describes the coordinate origin of the Chang-Refsdal lens approximation. The fact that the equidistance relationship with respect to $s - 1/s$ persists for central and resonant caustics therefore suggests the possibility that the perturbative picture and the Chang-Refsdal lens approximation may also be generalized beyond the planetary caustic, which is studied in the current work.

This paper is organized as follows. In Section 6.2, I introduce the generalized perturbative picture, which relaxes the required proximity between the planet and the image being perturbed in the previous work. Specifically, the planet always perturbs the image in the same half of the lens plane as the planet itself, leaving the other image largely unaffected. In Section 6.3, I propose a new variable-shear Chang-Refsdal lens approximation that quantifies the generalized perturbative picture. I show that the offset degeneracy becomes a consequence of physical symmetry under this variable-shear approximation. Crucially, the proposed Chang-Refsdal formalism enables full magnification maps to be derived analytically, whose accuracy is examined in Section 6.4. Section 6.5 shows that this variable-shear Chang-Refsdal lens formalism recovers known caustic properties of the planetary lens. In Section 6.6, I introduce a semi-analytic approach to solve the lens-equation exactly that is associated with the generalized perturbative picture. The results of this paper are reviewed in Section 6.7, where I discuss how they may be employed to substantially accelerate the modeling of observed events. A `Python` implementation of the exact semi-analytic and approximate analytic microlensing solutions presented in Sections 6.4 & 6.6 is provided[1].

6.2 The Perturbative Picture

The perturbative picture for planetary microlensing was initially laid out in Gould & Loeb (1992), which states "[a] planet of mass m affects appreciably the microlensing image only if the planet and the unperturbed image are separated by or order the planet's own Einstein radius," and that "at most one image is significantly affected and that the perturbed images lie near the unperturbed image[.]" Subsequent work then showed that the planet could also appreciably affect the microlensing image even if the planet lies far from the unperturbed image[2], namely via central caustics for high magnification events (Griest & Safizadeh 1998). The question that remains is whether the condition "at most one image is significantly affected" holds when the source passes close to central and resonant caustics.

To answer this question, it is illuminating to consider the lens plane of microlensing as opposed to the source plane. The lens plane describes the perturbing lens masses, the result-

[1] https://github.com/kmzzhang/analytic-lensing

[2] In this work, the term perturbation is used solely with respect to the planet. Thus, the unperturbed image refers to the major/minor images resulting from the primary star alone.

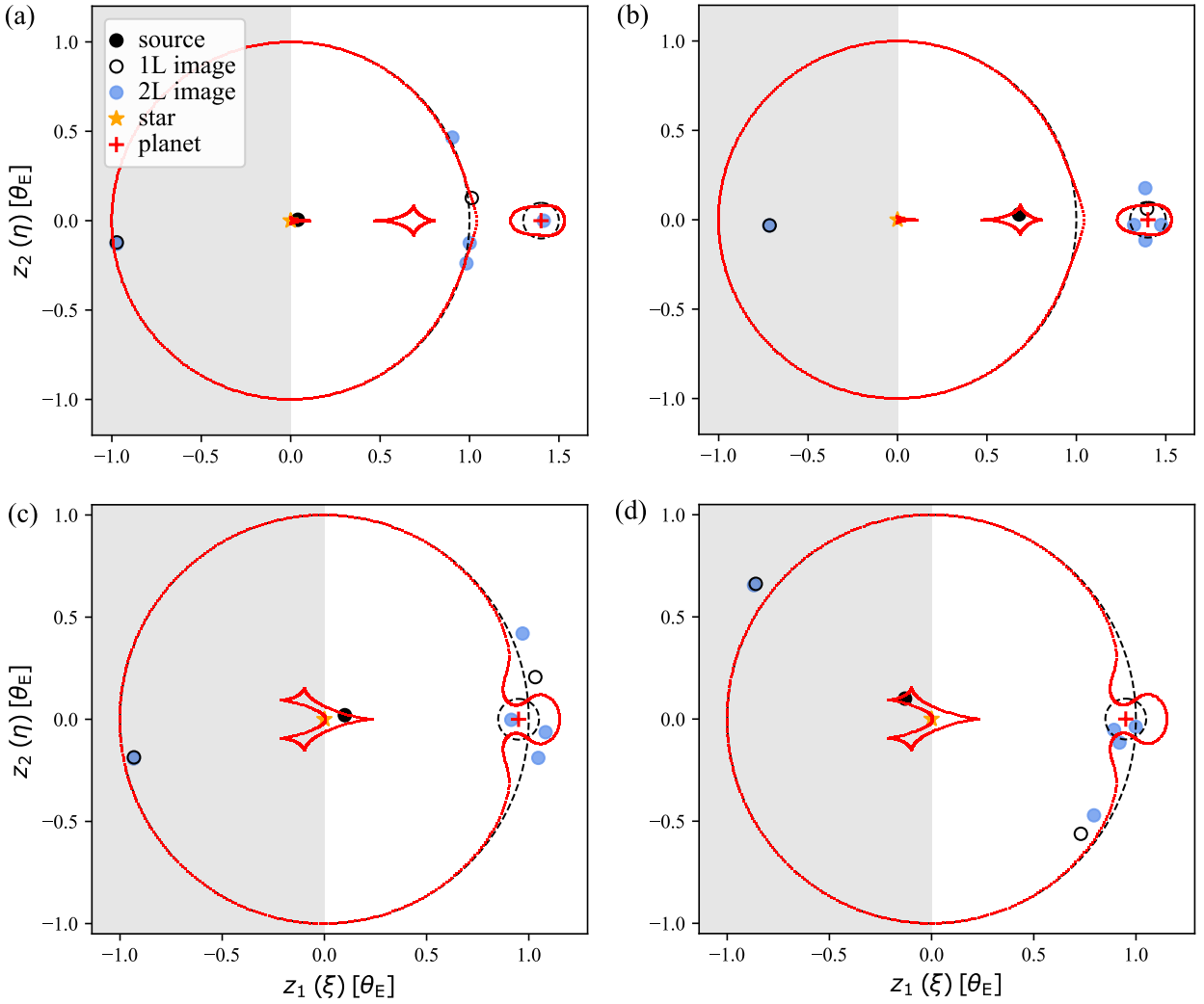


Figure 6.1: Illustration of the generalized perturbative picture for (a) central, (b) planetary, and (c,d) resonant caustic perturbations. (a,b,c) show major image perturbations and (d) shows minor image perturbation. The sphere of influence of the planet in the lens plane is indicated by the non-shaded region to the right with $z_1 > 0$. In the legend, 1L refers to the single-lens major/minor images resulting from the primary star alone, and 2L refers to the five images produced by the star-planet binary lens. In each case, the image in the shaded region is shown to be largely unaffected by the presence of the planet, as the 1L and 2L images coincide. The critical curves and caustics are shown in red, and the angular Einstein ring radius for the star and planet are shown in black dashed lines for comparison. The source star (black dot) is inside the caustics for all three cases. The mass ratio is $q = 0.01$ and the projected separation is $s = 1.4$ for (a,b) and $s = 0.95$ for (c,d).

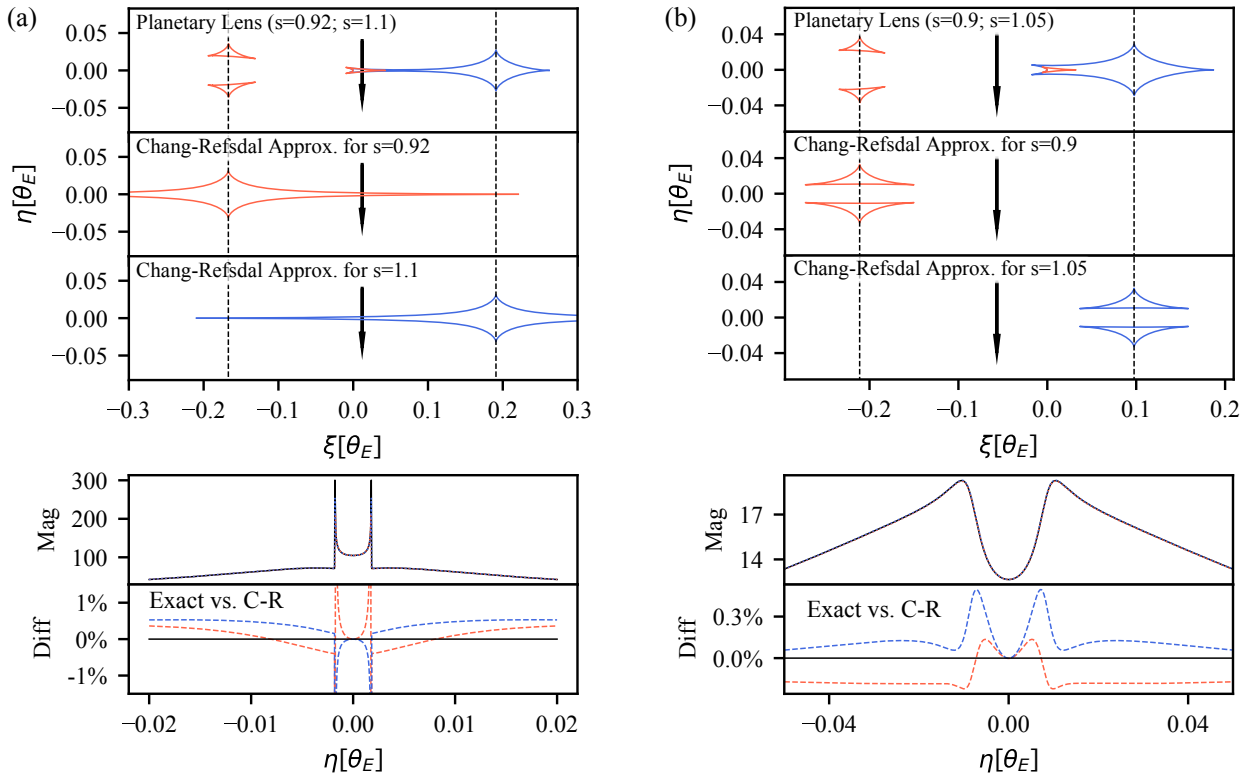


Figure 6.2: Illustration of the Chang-Refsdal lens approximation in the context of the Offset Degeneracy for (a) the generalized major image perturbation and (b) the generalized minor image perturbation. The top panel of (a,b) shows the caustics of two lens configurations overlaid in red ($s < 1$; left) and blue ($s > 1$; right), which give rise to degenerate light curves (second-from-bottom panel) for the source trajectory (vertical arrow) that crosses equidistance to the locations $s - 1/s$ (dashed lines) of the two lens configurations. The 2nd and 3rd panels from the top illustrate the Chang-Refsdal lens approximation for each lens configuration, which are shown to be exactly symmetrical with respect to the source trajectory. The bottom panels show the differences from the approximate Chang-Refsdal light curve to the two exact point-source light curves, shown in the same color coding. The mass ratio is $q = 5 \times 10^{-4}$ for both subplots.

ing images, and the critical curves, whereas the source plane describes magnification maps along with caustics resulting from all contributing images. Thus, the lens plane describes the *cause* and the source plane describes the *effect*. In the source plane, the proximity between the source and caustics correlates with higher magnifications. Analogously, the proximity between images and critical curves serves a similar purpose. For example, for a single lens, there exists one point-like caustic exactly at the lens mass itself. As the source approaches this singularity, both major/minor images approach the critical curve, leading to increased magnification. When the source coincides with the caustic, the image also coincides with the critical curve as an Einstein ring, which results in infinite magnification for a point source.

Let us now consider how the presence of the planet as an additional lens mass causes the critical curve to deviate from the Einstein ring of the primary mass. As illustrated in Figure 6.1(a, b), in the non-resonant regime, there is an isolated "planetary" critical curve centered on the planet with spatial scale $\theta_{\rm E,p}$. Parts of the "primary" critical curve ($\theta_{\rm E,\star}$) near its intersection with the positive real axis are elongated towards the planet, which is associated with the existence of central caustics. In the resonant regime (Figure 6.1c,d), the primary and planetary critical curves merge.

From these examples, it can be seen that the planet only affects parts of the critical curve in the positive lens plane, leaving the negative lens plane (shaded regions in Figure 6.1) largely unaffected. By considering the proximity of the unperturbed image locations to the critical curves as a proxy for magnification, we may then conclude that the planet also only perturbs the single-lens image in the positive lens plane, leaving the image in the negative lens plane unaffected. As illustrated in Figure 6.1, this is indeed the case regardless of the proximity between the planet and the image being perturbed.

While the original perturbative picture considers the sphere of influence of the planet as limited to its angular Einstein radius, the above discussion indicates a generalized perturbative picture: the sphere of influence of the planet is constrained to one-half of the lens plane, where it splits one of the major/minor images into two or four images, leaving the other image largely unaffected. This generalized perturbative picture then allows for a unified classification of planetary perturbations into major-image perturbations and minor-image perturbations. Under the original perturbative picture, the distinction between major and minor-image perturbations has been restricted to planetary-caustic perturbations (Figure 6.1b). Thus, major-image perturbations have been restricted to wide-separation planets ($s > 1$) and vice versa[3]. The present discussion shows that the distinction between major/minor image perturbations should be made not by the location of the planet, but by the location of the source instead. Specifically, major-image perturbations occur when the source is in the positive source plane ($\zeta > 0$), and vice versa. This would then allow major-image perturbations to be generalized to $s < 1$ planets, as illustrated in Figure 6.1(c).

[3]See the Appendix in Han et al. (2018): "Types of Planetary Perturbations"

6.3 The Chang-Refsdal Lens Approximation

The uniform-shear Chang-Refsdal lens approximation quantifies the action of the planet under the original perturbative picture of Gould & Loeb (1992). Here, I introduce a variable-shear Chang-Refsdal lens approximation that quantifies the generalized perturbative picture, which accurately describes all of the central, resonant and planetary caustics.

6.3.1 Uniform-Shear Approximation

Since past works have adapted slightly different conventions on the uniform-shear Chang-Refsdal lens approximation, let us first re-examine the relevant works using a uniform notation of the complex two-body lens equation (Equation 6.2). To consider the lensing behavior near the planetary lens companion, let us first transform the complex lens equation from the *primary* frame (ζ, z) to the *planetary* frame ($\zeta^{[2]}$, $z^{[2]}$), which has units of the planetary Einstein radius $\theta_{\rm E,p} = \sqrt{q}\theta_{\rm E,\star}$, and coordinate origins at the location of the planet ($z = s$) for the lens plane, with the corresponding location in the source plane ($\zeta = s - 1/s$). The latter is often interpreted as the location of the planetary caustic. Applying the coordinate transformation

$$\begin{aligned} \zeta &= \sqrt{q}\zeta^{[2]} + s - 1/s \\ z &= \sqrt{q}z^{[2]} + s, \end{aligned} \tag{6.4}$$

and rearranging, the two-body lens equation becomes

$$\zeta^{[2]} = z^{[2]} - \frac{1}{\bar{z}^{[2]}} + \frac{\bar{z}^{[2]}}{s \cdot (\sqrt{q}\bar{z}^{[2]} + s)}. \tag{6.5}$$

In the limit of $q \to 0$, the above equation is reduced to the Chang-Refsdal lens with uniform shear $\gamma = 1/s^2$ (Dominik 1999). For finite $q \ll 1$, the Chang-Refsdal lens with $\gamma = 1/s^2$ is the first order Taylor expansion of Equation 6.5 around $\bar{z}^{[2]} = 0$ (Dominik 1999; Bozza 2000), which can also be interpreted as a power-series in $\sqrt{q}$

$$\zeta^{[2]} = z^{[2]} - \frac{1}{\bar{z}^{[2]}} + \sum_{i=1}^{\infty}(-1)^{i+1} \cdot q^{(i-1)/2} \cdot \frac{(\bar{z}^{[2]})^i}{s^{i+1}}. \tag{6.6}$$

On the other hand, the original Chang-Refsdal approximation of the earlier work of Gaudi & Gould (1997) adopted a slightly different shear definition. Instead of the planet location, the shear is evaluated at the location of the image being perturbed at the mid-point of the perturbation, which occurs when the source crosses the star-planet axis. Recall that the original perturbative picture requires the image being perturbed to pass the planet closer than $\mathcal{O}(\theta_{E,\rm p})$. Therefore, the location of the image being perturbed would approach the planet location for $q \to 0$, and the two shear definitions would become equivalent.

6.3.2 Variable-Shear Approximation

In this subsection, I introduce a new variable-shear Chang-Refsdal lens approximation that holds for all three caustic topologies. The shear is defined to be real positive $\gamma = 1/z_+^2$, where

$$z_+ = \frac{\sqrt{\xi^2 + 4} + \xi}{2}. \tag{6.7}$$

For sources on the real axis ($\zeta = \xi$), the shear definition is identical to Gaudi & Gould (1997), which is nevertheless formulated within the original perturbative picture that requires $|z_+ - s| \lesssim \mathcal{O}(\theta_{E,\mathrm{p}})$. Under the generalized perturbative picture, Equation 6.7 corresponds to the unperturbed location of the image that is assumed to be perturbed by the planet, which is the major image for $\xi > 0$ and the minor image for $\xi < 0$. The image being perturbed is always in the positive lens plane, and thus the "+" subscript. For sources off the real axis ($\eta \neq 0$), the shear is evaluated by projecting the source location onto the real axis. Thus the lines of constant shear (LCS) are perpendicular to the real axis by construction.

The proposed approximation is different from the variable-shear approximation of Gould & Loeb (1992), which evaluates the shear directly at unperturbed image location rather than its projection on the real axis. The formalism of this earlier work was derived by Taylor expanding the time-delay surface at the unperturbed image location, which was motivated by the condition where "the perturbed images lie near the unperturbed image[.]" Again, this assumption only holds for isolated planetary caustics, as can be seen in Figure 6.1. As noted in footnote 3 of Gould & Loeb (1992), this approximation also results in a leftward arching of the planetary caustics that is not present in the exact calculation, nor the new variable-shear formalism proposed here (see Section 6.4).

In contrast, the requirement of the shear to be real and the LCS to be perpendicular to the star-planet axis is motivated by conditions of the offset degeneracy. As illustrated in Figure 6.2, vertical source trajectories result in nearly identical light curves under the degenerate lens configurations (Zhang et al. 2022). Now, if one were to apply a literal reading of the uniform-shear approximation of Gaudi & Gould (1997) and hold the shear fixed on the star-planet axis, one would find that the resulting light curve under the Chang-Refsdal lens approximation nearly perfectly resembles both of the degenerate light curves. In fact, the extent to which the Chang-Refsdal light curves deviate from the exact light curves is similar to the extent to which the two degenerate light curves deviate from one another.

The above findings may appear rather surprising, since the uniform-shear approximation of Gaudi & Gould (1997) is known to fail for resonant and central caustics (Dominik 1999). However, an implicit assumption made in the previous works is that the uniform-shear approximation fails as a *global* approximation. As we have seen from Figure 6.2, the uniform-shear Chang-Refsdal lens serves as an excellent *local* approximation along the vertical direction in the source plane, despite the absence of *global* caustic resemblance. Consequently, for oblique trajectories, one can derive equally accurate light-curve approximations by evaluating the shear at the projection of the source onto the lens axis.

As I will show, this variable-shear Chang-Refsdal lens approximation not only leads

to accurate magnification maps (Section 6.4), but also recovers known caustic properties of the planetary lens (Section 6.5). However, the accuracy of the proposed variable-shear approximation in the source plane contrasts sharply with the fact that it does not recover the correct image locations. This can be easily seen by considering a hypothetical source at infinity, where the major image overlaps with the true source location. Here, the shear goes to zero and the Chang-Refsdal lens is reduced to the point lens, but the origin of the Chang-Refsdal lens is located at $1/s$, thus placing the major image at the wrong location. This behavior indicates that the variable-shear lens should be considered degenerate with the exact planetary lens because they share similar source-plane but not lens-plane behavior. This intriguing behavior deserves further analytical study in future works.

6.4 Analytic Magnifications

Given that the Chang-Refsdal lens equation can be transformed into a quartic polynomial, we may now calculate full magnification maps for the planetary lens analytically. To acquire analytic magnifications, one first takes the complex conjugate of Equation 6.3, and substitutes the expression for $\bar{z}$ back into Equation 6.3 itself. After clearing fractions, we arrive at a quartic polynomial,

$$p(z^{[2]}) = \sum_{i=0}^{4} a_i(\zeta^{[2]}, \bar{\zeta}^{[2]}, \gamma) \cdot (z^{[2]})^i = 0, \tag{6.8}$$

where, with $\gamma = 1/z_+^2$ (Equation 6.7),

$$\begin{aligned}
a_0 =& \gamma \\
a_1 =& -\zeta^{[2]} + 2\gamma\bar{\zeta}^{[2]} \\
a_2 =& -2\gamma^2 - \zeta^{[2]}\bar{\zeta}^{[2]} + \gamma\bar{\zeta}^{[2]^2} \\
a_3 =& \gamma\zeta^{[2]} + \bar{\zeta}^{[2]} - 2\gamma^2\bar{\zeta}^{[2]} \\
a_4 =& -\gamma + \gamma^3.
\end{aligned}$$

Note that not all roots of the quartic polynomial are solutions to the original lens equation, and each solution should be verified by plugging back into Equation 6.3. The total magnification is the sum of the magnification of each individual image, which is given by the absolute value of the inverse Jacobian determinant,

$$\mu_\gamma = \sum_j \left| 1 - \frac{\partial\zeta^{[2]}}{\partial\bar{z}^{[2]}} \frac{\overline{\partial\zeta}^{[2]}}{\partial\bar{z}^{[2]}} \right|^{-1}_{z_j^{[2]}}, \tag{6.9}$$

where the derivatives are evaluated using Equation 6.3 at the valid image solutions $z_j^{[2]}$.

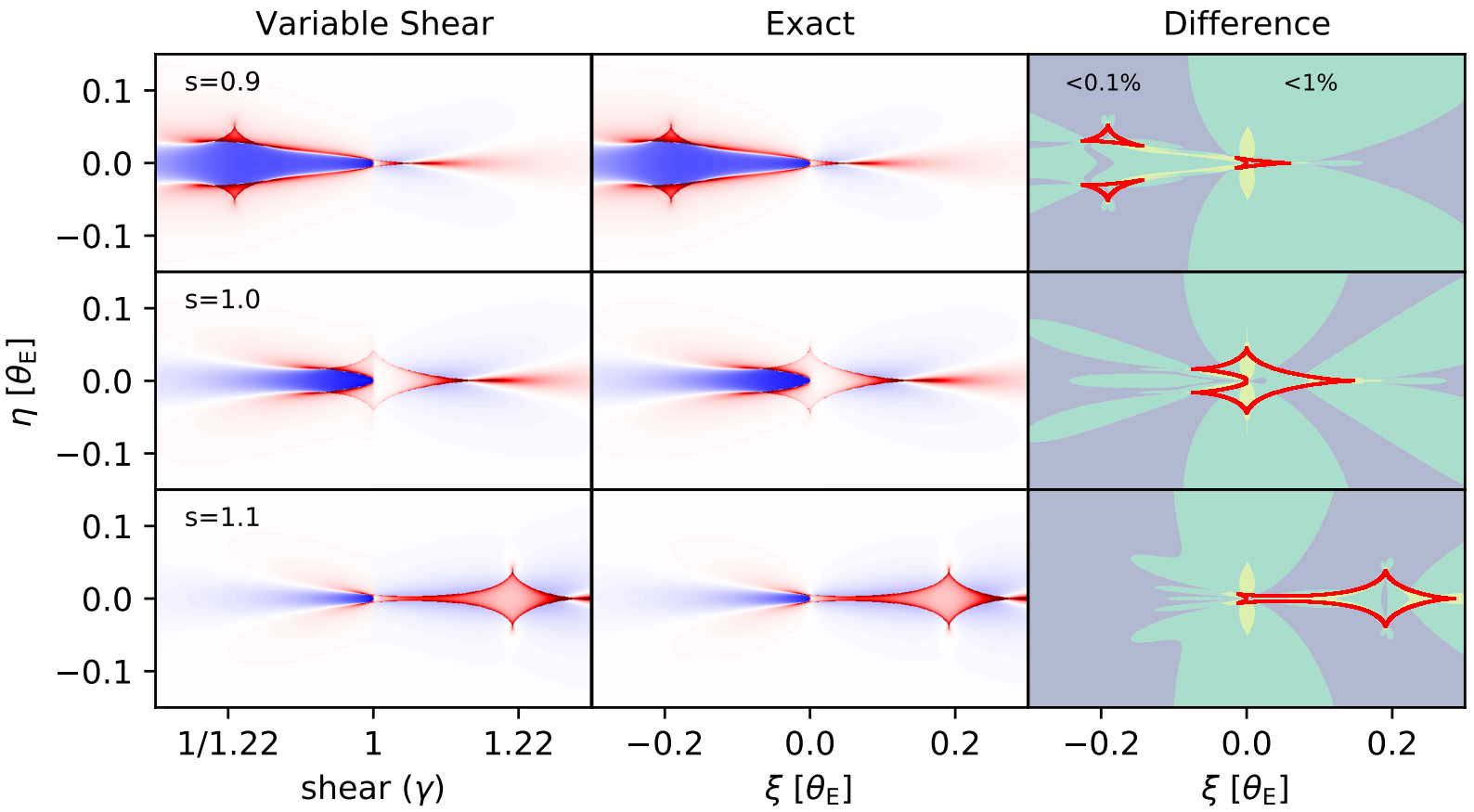


Figure 6.3: Magnification maps calculated using the variable-shear Chang-Refsdal lens approximation and the exact lens formalism (with the semi-analytic solver of Section 6.6), as well as their fractional differences. The magnification maps are visualized as the deviation from the point-lens point-source magnification map, where excess magnification is shown in red and suppressed magnification is shown in blue, with consistent color scale across subplots. The horizontal axis for variable shear is re-parameterized with shear (γ) using the definition in Equation 6.7. The coordinate origin for the exact calculation is offset to the primary cusp. The color coding in the difference maps indicates differences of less than 0.1%, 1%, 10%, as labeled in the middle panel except for the light-green central region of <10%. The caustics are overlaid in red in the difference maps for reference. The mass ratio is $q = 10^{-3}$ for all subplots.

The variable-shear lens provides the magnification perturbation by the planet through,

$$\Delta\mu = \mu_\gamma - \mu_\infty \tag{6.10}$$

$$= \mu_\gamma - \frac{1}{|\gamma^2 - 1|}, \tag{6.11}$$

where μ_∞ is the terminal magnification ($|\zeta^{[2]}| \to \infty$). With $u = |\zeta|$, the full magnification for the total of three or five images can be found by adding back the single-lens magnifications,

$$\mu = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} + \Delta\mu. \tag{6.12}$$

Although lengthy when expressed as a function of (ζ, s, q), Equation 6.12 is indeed closed-form and may offer substantial speed-up in the calculation of planetary microlensing light curves and the modeling of observed events. Let us now examine the accuracy of magnification maps under the variable-shear Chang-Refsdal approximation. It is already known from previous works that the Chang-Refsdal lens provides excellent approximation near isolated planetary caustics. Therefore, let us first examine the accuracy of the variable-shear magnification maps near resonant and semi-resonant caustics, before examining magnification maps near central caustics.

Figure 6.3 shows the variable-shear and the exact calculations of magnification maps for lenses in or near the resonant regime, which appear nearly identical. The magnification difference maps reveal two major regimes where the two calculations differ by $> 1\%$. First, there is a dumbbell-shaped structure along the imaginary axis of size $\Delta\eta \sim 0.1$, the interpretation of which will be clear with the discussion of Figure 6.4 in the next paragraph. Deviations greater than 1% also occur along the excess magnification ridges straddling the suppressed magnification zone between the close-planetary caustic and the central caustic, as seen in the top panel of Figure 6.3. This type of deviation concerns the exact shape of the excess magnification ridges. One example is already seen in Figure 6.2(b) where, with a mass ratio of $q = 5 \times 10^{-4}$, the maximum deviation is merely 0.5%. Therefore, the variable-shear approximation is in excellent agreement with the exact calculation outside of a small central region, which we turn our attention to now.

Both of the aforementioned types of discrepancies become more pronounced in the high-magnification regime. In Figure 6.4, one immediately notices the discontinuity across the imaginary axis for the variable-shear calculation, which accounts for the dumbbell-shaped structure seen in Figure 6.3. For sources near the imaginary axis and the primary star, the two unperturbed image locations are about equidistant to the planet and both images are substantially affected by the planet. In other words, the generalized perturbative picture is no longer accurate in the high-magnification regime near the imaginary axis. The magnification near the imaginary axis is overestimated for minor-image perturbations and underestimated for major-image perturbations. Given that this discontinuity is known in advance and does not usually coincide with true planetary features, one may apply *post-hoc* corrections when applying the analytic magnification to the modeling of observed events, for example, by reducing the weights of photometric data points near the imaginary axis.

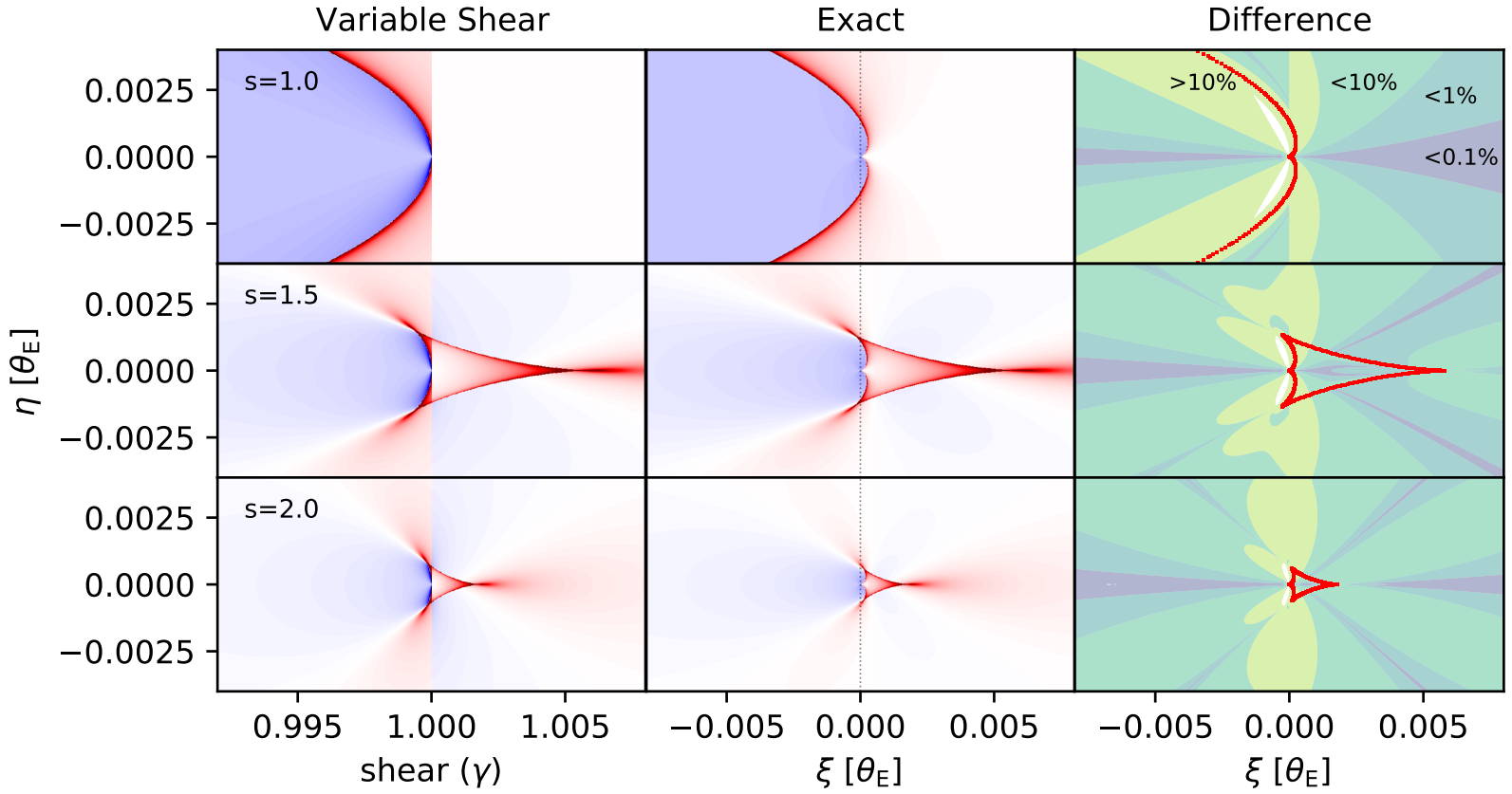


Figure 6.4: Similar to Figure 6.3, but zoomed into the central region and shown for $s = (1, 1.5, 2)$. The color scale is consistent across subplots, but different from Figure 6.3. The coordinate origin for the exact calculation is offset to the primary cusp. The uncolored regions in the difference maps indicate negative magnifications resulting from the variable-shear approximation.

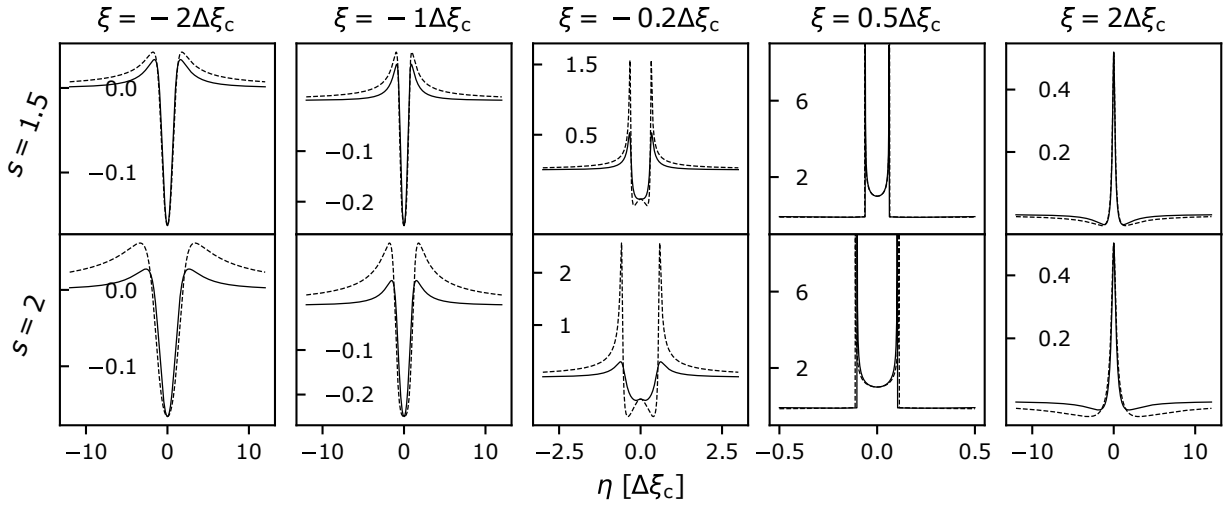


Figure 6.5: Magnification slices (light curves) along the vertical direction for $s = (1.5, 2)$, which are associated with the lower two panels of Figure 6.4. As shown in the subplot titles, the impact parameters of the magnification slices are in units of the central-caustic size, which is $\Delta\xi_c \sim 0.006$ for $s = 1.5$ and $\Delta\xi_c \sim 0.002$ for $s = 2$. The exact calculation is shown in solid lines and the variable-shear calculation in dashed lines. The vertical axes are in units of the fractional deviation from the point-source point-lens magnification.

On the other hand, the discrepancy along the excess magnification ridge becomes prominent for high-magnification minor-image perturbations, where high magnification means the immediate vicinity of central caustics with impact parameters of $u_0 \sim q$. Figure 6.5 shows magnification slices along the vertical direction with various impact parameters in units of the central-caustic size (Equation 6.15). The left two columns show that the variable-shear calculation substantially overestimates the strength of the excess magnification ridge. The middle panel of Figure 6.5 shows that this deviation diverges in the immediate vicinity of the two off-axis central-caustic cusps, where minuscule differences in the caustic shape becomes important. The magnification difference maps in Figure 6.3 also reveal limited regions of spatial scale $\mathcal{O}(q)$ with unphysical negative magnifications under the two caustic folds towards the back end, which is also reflected in the double dips in the middle panels of Figure 6.5. The interpretation of these behaviors will become clearer with the discussion of caustics in Section 6.5.

In contrast, the right two panels of Figure 6.5 show that variable-shear magnifications are much more accurate in this ultra high-magnification regime for major-image perturbations, including inside of central and resonant caustics (also see Figure 6.2a). Lastly, despite these anomalous behaviors in the regime of $u_0 \sim q$, Figure 6.6 shows that the variable-shear calculation is nearly identical to the exact calculation on the real axis in the high-magnification regime. This behavior may be straightforwardly derived using the resultant method (Witt & Mao 1995), which was adopted in Zhang & Gaudi (2022) to derive closed-form magnifications for the planetary lens on the real axis.

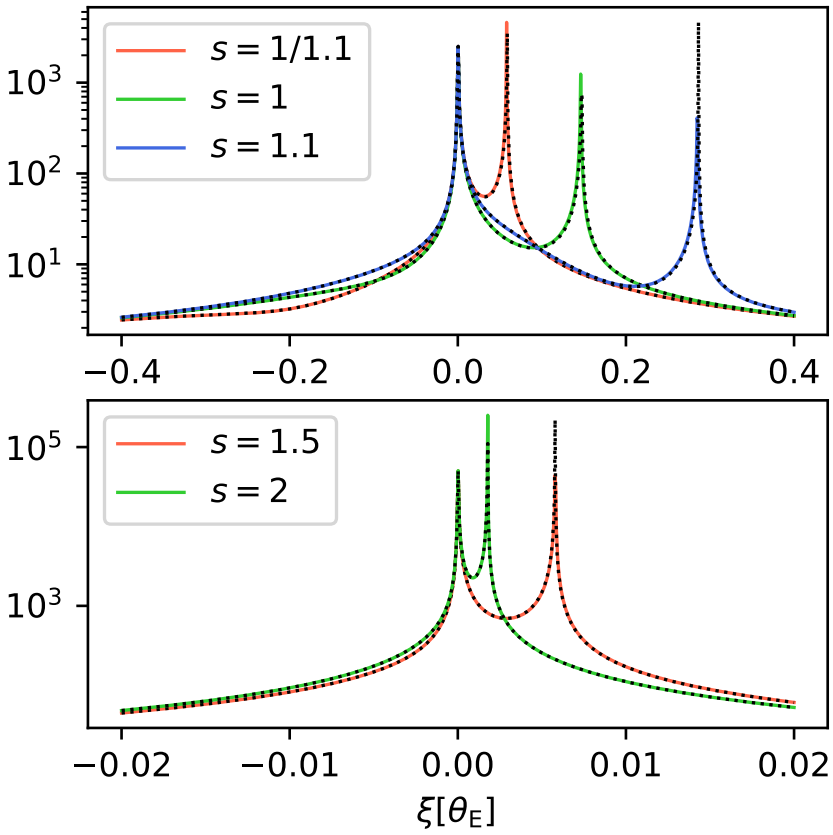


Figure 6.6: Real-axis magnifications under the variable-shear (solid lines) and exact (dotted lines) calculations. The top panel corresponds to the lens configurations in Figure 6.3 and the bottom panel corresponds to the configurations in Figure 6.4. The vertical axes show magnification on log scale.

6.5 Caustics

Let us first examine the interpretation of central and resonant caustics under the variable-shear Chang-Refsdal lens approximation, which will also assist the interpretation of Figure 6.4. Imagine a hypothetical source on the real axis moving across the primary star from $\xi > 0$ to $\xi < 0$. Here, the shear changes from $\gamma < 1$ to $\gamma > 1$, and the underlying Chang-Refsdal caustic splits into two (compare Figure 6.2a/b). Since the horizontal size of the $\gamma > 1$ Chang-Refsdal caustic diverges to infinity in the limit of $\gamma \to 1$, this hypothetical source transitions from the inside to the outside of caustics in the above scenario, thus resulting in a cusp exactly at the primary star, which is referred to the primary cusp. Moreover, since the two caustic folds originating from the primary cusp are parts of $\gamma > 1$ Chang-Refsdal caustics, the back ends of the central/resonant caustic are restricted to the negative source plane. In comparison, there is always a small offset between the primary cusp and the primary star under the exact calculation, where the primary folds are also allowed to traverse into the positive source plane (see Figure 6.4).

For the remainder of this section, I will examine how the variable-shear Chang-Refsdal lens approximation recovers known caustic properties of the two-body planetary lens. Caustic cusp locations can be derived under the variable-shear approximation by recognizing that Chang-Refsdal cusp locations are expressed as simple analytic expressions of the shear (e.g., An & Evans 2006), which are related back to the cusp locations themselves via Equation 6.7. Here, I will examine two special cases: the central caustic and the $s = 1$ resonant caustic.

Since the primary cusp is always located at the primary star and coordinate origin under the variable-shear formalism, the length of the central caustic is given by the location of its other cusp on the real axis located at ξ_c. Since $\xi_c \ll 1$, the shear at ξ_c is

$$\gamma_c = \frac{1}{z_+^2(\xi_c)} \simeq 1 - \xi_c, \tag{6.13}$$

which is illustrated in the variable-shear axis labels in Figure 6.4. The real-axis cusps for $\gamma < 1$ Chang-Refsdal caustics are located at $\pm 2\gamma/\sqrt{1-\gamma}$ in the planetary frame. Equating its locations in the primary and planetary frames (Equation 6.4) and substituting in the shear in Equation 6.13, we arrive at

$$\sqrt{q}\left(s - \frac{1}{s} - \xi_c\right) = \pm\frac{2\gamma_c}{\sqrt{1-\gamma_c}} = \pm\frac{2(1-\xi_c)}{\sqrt{\xi_c}}, \tag{6.14}$$

where the plus sign corresponds to the wide topology and the minus sign for the close topology. The above equation can be rearranged into a cubic polynomial in $\sqrt{\xi_c}$. We may then Taylor expand the valid cubic root in q and acquire

$$\xi_c = \frac{4q}{(s-1/s)^2} - \frac{32q^2s^4(s^2-s-1)}{(s^2-1)^5} + \mathcal{O}\big(q^3\big). \tag{6.15}$$

The first order q term is invariant under $s \leftrightarrow 1/s$ and is in agreement with the exact planetary lens (e.g., An 2005; Chung et al. 2005). However, the second-order term disagrees

(e.g., An 2021; Eq. 13), indicating higher-order differences. Note that by clearing fractions in Equation 6.14 and dropping the highest order term in ξ_c, Equation 6.14 itself becomes invariant under $s \leftrightarrow 1/s$, allowing one to directly acquire the first-order-q term without the Taylor expansion.

For the $s = 1$ resonant caustic, the origins for the primary and primary coordinates coincide ($s - 1/s = 0$), indicating that the "planetary cusps" are exactly on the imaginary axis of the primary frame, where the shear becomes $\gamma = 1$. Therefore, the imaginary-axis cusps are located at $\eta^{[2]} = \pm 2\gamma/\sqrt{1+\gamma} = \pm\sqrt{2}$ in the planetary frame, and $\eta_r = \pm\sqrt{2q}$ in the primary frame. The vertical size of the $s = 1$ caustic is therefore $\Delta\eta_r = 2\sqrt{2q}$.

The horizontal size of the $s = 1$ caustic may be derived in a similar manner as the central caustic via Equation 6.14. With the location of the real-axis resonant-caustic cusp written as ξ_r with shear γ_r,

$$\sqrt{q} \cdot \xi_r = \frac{2\gamma_r}{\sqrt{1-\gamma_r}}. \tag{6.16}$$

Substituting in Equation 6.7 and expanding up to first order in ξ_r, we have

$$\frac{2}{\sqrt{\xi_r}} - \frac{3\sqrt{\xi_r}}{2} - \frac{\xi_r}{\sqrt{q}} = 0. \tag{6.17}$$

For $\xi_c \ll 1$ and $q \ll 1$, the $\sqrt{\xi_c}$ term may be dropped, which result in $\xi_r = \sqrt[3]{4q}$, and this is the length of the resonant caustic. The above results also show that the vertical-to-horizontal width ratio of the resonant caustic scales as

$$\eta_r/\xi_r \propto q^{1/6}, \tag{6.18}$$

which does not appear to be well-known in the literature.

6.6 Semi-Analytic Solutions

In this section, I demonstrate how the generalized perturbative picture allows the full two-body lens equation to be solved semi-analytically. Given that the image in the negative lens plane is only weakly affected by the planet, its unperturbed location

$$z_{\rm PSPL} = \frac{\zeta}{2} \cdot \left(1 \pm \sqrt{1 + 4|\zeta|^{-2}}\right) \tag{6.19}$$

can be used as an initial guess to Newton's or Laguerre's method to quickly solve for one quintic root of the lens equation. Here, PSPL refers to point-source point-lens. In the above equation, the plus sign represents the major image location that is chosen for minor image perturbations, and vice versa. Once one quintic root is found and divided out, the resulting quartic polynomial can be solved in closed form. The quartic roots can then be verified with the full quintic equation and, depending on the requested precision, may be optionally refined by Newton's method to reduce the numerical noise from the initial root division.

This noise is nevertheless expected to be small for well-isolated roots (e.g. Skowron & Gould 2012). Indeed, the weakly perturbed image is also generally the most isolated image (Figure 6.1), and thus the final refinement may not be necessary.

Note that the closed-form quartic solution discovered by Lodovico Ferrari is known to suffer from certain round-off errors for cases with large root spread (e.g. Strobach 2010), defined as the ratio between the largest and smallest root magnitudes. Therefore, the coordinate origin is defined at the primary star, as only the minor image becomes close to the origin for very faraway sources. In comparison, other frameworks such as VBBL (Bozza et al. 2018) have also considered coordinate origins at the planetary location, which may induce large root spread as one image is usually very close to the planet. In future work, an improved quartic solver proposed by Orellana & Michele (2020) may also be explored, which is robust against these errors but only costs twice the computational time as Ferrari's solution.

The initial root-refinement step may benefit from a combination of Newton's and Laguerre's methods depending on the polynomial residual (Skowron & Gould 2012), which is larger for sources near the imaginary axis in our case. As an illustration, in the case of the $s = 1$ magnification map in Figure 6.3, it only took 2/4 iterations with Laguerre's/Newton's method to refine from the PSPL location of the weakly affect image location to a polynomial residual less than 10^{-14} for 99.9% of the pixels. In comparison, using the PSPL location of the strongly perturbed image, the planet location, and the primary location takes 7/16, 5/16, and 9/23 iterations[4] with Laguerre's/Newton's method to locate the first quintic root subject to the same precision requirements, which demonstrates the comparative advantage of starting from the weakly affected image.

A basic benchmark test of a vectorized python implementation provided in the code repository of this paper shows that $s = 1$ magnification map in Figure 6.3 with 2.5×10^5 pixels is calculated in merely 0.6s with the semi-analytic method, with the two steps of finding the initial root and solving the quartic polynomial in closed-form taking around 0.3s each. Since the total cost is only twice the cost of a few iterations of Newton's method to solve for the initial root, we may expect the semi-analytic method to be substantially faster than the standard numerical approach (e.g. Skowron & Gould 2012). However, we do not attempt to quantify the factor of speed-up here, which involves the delicate task of holding the level of optimization consistent across all methods tested. In comparison, the analytic variable-shear solution of Section 6.4 costs only 0.2s for the same $s = 1$ magnification map. The semi-analytic solver also applies to binary mass ratios, which takes a slightly longer 1s for $q = 0.9$ and $s = 1$ given the reduced accuracy of Equation 6.19 as the initial guess. The above numbers will be dependent on the computational device and may be rerun with the code-base provided.

[4]These numbers also suggest an alternative and potentially useful approach to first find and divide out three roots initializing from the two PSPL locations and the planet location, where the resulting quadratic equation after root division could then be solved in closed form.

6.7 Conclusions

In this paper, I have introduced the idea of a generalized perturbative picture for planetary microlensing, which states that the planet acts as a variable-shear Chang-Refsdal lens on one of the unperturbed images in the positive lens plane, leaving the other image largely unaffected. The proposed framework generalizes upon the original perturbative picture of Gould & Loeb (1992) and the uniform-shear Chang-Refsdal lens approximation of Gaudi & Gould (1997), by relaxing both the required proximity between the planet and the image being perturbed and the condition of isolated planetary caustics.

Under the generalized perturbative picture, the action of the planet can be classified into major image perturbations and minor image perturbations, which are distinguished by whether the source is in the positive ($\xi > 0$) or negative ($\xi < 0$) source plane, as opposed to the location of the planet being inside ($s < 1$) or outside ($s > 1$) the Einstein ring of the primary star (cf. Appendix in Han et al. 2018). Moreover, the generalized perturbative picture demonstrates that the existence of a unified regime of light-curve degeneracy independent of caustic topologies can be explained by the symmetry of the Chang-Refsdal lens. Therefore, the offset degeneracy can be interpreted as a generalization of the inner-outer degeneracy for planetary caustics, both of which describe an ambiguity as to "whether the planet lies closer to or farther from the star than does the position of the image that it is perturbing" (Gaudi & Gould 1997).

It should be noted that the variable-shear Chang-Refsdal lens approximation is proposed without formal derivation in this paper. Instead, I have demonstrated that the proposed formalism not only produces accurate magnification maps (Section 6.4), but also recovers known caustic properties of the full planetary lens (Section 6.5). An interesting property of the variable-shear lens is that it does not recover the correct image positions, despite its accuracy in the source plane. This intriguing behavior deserves further analytical study in future works.

Moving forward, it is beneficial for the two analytic prescriptions (Section 6.4 & 6.6) together with finite-source algorithms (e.g. Dominik 1998; Gould 2008; Bozza 2010) to be implemented[5] in automatic-differentiation frameworks such as `jax` (Bradbury et al. 2018) or `julia` (Bezanson et al. 2017), which allows the gradient of the likelihood function to be acquired without deriving explicit expressions. This allows for the use of gradient-based inference algorithms, particularly Hamiltonian Monte Carlo (HMC) methods including the No-U-Turn Sampler (NUTS; Hoffman & Gelman 2014), which utilize gradient information to avoid the random walking behavior of common Markov chain Monte Carlo (MCMC) samplers such as Metropolis-Hastings. For the exact semi-analytic approach, it is also not necessary for the gradient to be "back-propagated" through the root-refinement step for planetary mass ratios, where the location of the weakly affect image is insensitive to the planetary parameters.

[5]I note that one such differentiable microlensing code named `caustics` (Bartolić & Dominik, in prep), is currently under development.

Acknowledgments

A version of this chapter has been accepted for publication in the Monthly Notices of the Royal Astronomical Society. A pre-print is available at Zhang (2023).

I would like to thank Scott Gaudi and Jin An for helpful discussions, and Joshua Bloom, Scott Gaudi, Jessica Lu, Sean Terry, and Weicheng Zang for comments on the manuscript. K.Z. thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. This work is partially supported by funding from a Two Sigma faculty fellowship and the Gordon and Betty Moore Foundation.

Bibliography

Agarwal, D., Aggarwal, K., Burke-Spolaor, S., Lorimer, D. R., & Garver-Daniels, N. 2020, Monthly Notices of the Royal Astronomical Society, 497, 1661

Aguirre, C., Pichara, K., & Becker, I. 2018, Monthly Notices of the Royal Astronomical Society, 482, 5078

Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1996, The Astronomical Journal, 111, 1146

Alcock, C., Allsman, R. A., Alves, D. R., et al. 2000, The Astrophysical Journal, 542, 281

An, J. 2021, arXiv:2102.07950 [astro-ph]

An, J. H. 2005, Monthly Notices of the Royal Astronomical Society, 356, 1409

An, J. H., & Evans, N. W. 2006, Monthly Notices of the Royal Astronomical Society, 369, 317

An, J. H., & Han, C. 2002, The Astrophysical Journal, 573, 351

Arjovsky, M., Shah, A., & Bengio, Y. 2016, in Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML'16, 1120

Aubourg, E., Bareyre, P., Bréhin, S., et al. 1993, Nature, 365, 623

Ba, J. L., Kiros, J. R., & Hinton, G. E. 2016, arXiv:1607.06450 [cs, stat]

Bai, S., Kolter, J. Z., & Koltun, V. 2018, arXiv:1803.01271 [cs]

Becker, I., Pichara, K., Catelan, M., et al. 2020, Monthly Notices of the Royal Astronomical Society, 493, 2981

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2018, Publications of the Astronomical Society of the Pacific, 131, 018002

Bennett, D. P., & Rhie, S. H. 2002, The Astrophysical Journal, 574, 985

Bennett, D. P., Bond, I. A., Abe, F., et al. 2017, The Astronomical Journal, 154, 68

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAM Review, 59, 65

Bonanos, A. Z., Stanek, K. Z., Kudritzki, R. P., et al. 2006, The Astrophysical Journal, 652, 313

Bond, I. A., Udalski, A., Jaroszyński, M., et al. 2004, The Astrophysical Journal, 606, L155

Bond, I. A., Bennett, D. P., Sumi, T., et al. 2017, Monthly Notices of the Royal Astronomical Society, 469, 2434

Bozza, V. 1999, Astronomy and Astrophysics, 348, 311

—. 2000, Astronomy and Astrophysics, 355, 423

—. 2010, Monthly Notices of the Royal Astronomical Society, 408, 2188

Bozza, V., Bachelet, E., Bartolić, F., et al. 2018, Monthly Notices of the Royal Astronomical

Society, 479, 5157
Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: composable transformations of Python+NumPy programs
Breiman, L. 2001, Machine Learning, 45, 5
Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM Journal on Scientific Computing, 16, 1190
Cameron, E., & Pettitt, A. N. 2012, Monthly Notices of the Royal Astronomical Society, 425, 44
Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in Advances in Knowledge Discovery and Data Mining, ed. J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu, Lecture Notes in Computer Science, 160
Carleo, G., Cirac, I., Cranmer, K., et al. 2019, Reviews of Modern Physics, 91, 045002
Chang, K., & Refsdal, S. 1979, Nature, 282, 561
—. 1984, Astronomy and Astrophysics, 132, 168
Chen, X., Wang, S., Deng, L., et al. 2020, The Astrophysical Journal Supplement Series, 249, 18
Cho, K., van Merriënboer, B., Gulcehre, C., et al. 2014, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1724
Choi, J.-Y., Shin, I.-G., Han, C., et al. 2012, The Astrophysical Journal, 756, 48
Chung, S.-J., Han, C., Park, B.-G., et al. 2005, The Astrophysical Journal, 630, 535
Cranmer, K., Brehmer, J., & Louppe, G. 2020, Proceedings of the National Academy of Sciences, 117, 30055
Davies, A., Veličković, P., Buesing, L., et al. 2021, Nature, 600, 70
Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, Astronomy & Astrophysics, 475, 1159
Di Stefano, R., & Mao, S. 1996, The Astrophysical Journal, 457, 93
Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2017, in 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings
Dominik, M. 1998, Astronomy and Astrophysics, 333, L79
—. 1999, Astronomy and Astrophysics, 349, 108
Dong, S., Bond, I. A., Gould, A., et al. 2009, The Astrophysical Journal, 698, 1826
Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, Monthly Notices of the Royal Astronomical Society, 414, 2602
Dékány, I., & Grebel, E. K. 2020, The Astrophysical Journal, 898, 46
Dékány, I., Hajdu, G., Grebel, E. K., & Catelan, M. 2019, The Astrophysical Journal, 883, 58
Einstein, A. 1936, Science, 84, 506
Fatheddin, H., & Sajadian, S. 2022, Monthly Notices of the Royal Astronomical Society, 514, 4379
Gaia Collaboration, Eyer, L., Rimoldini, L., et al. 2019, Astronomy and Astrophysics, 623, A110
Gaudi, B. S. 2010, arXiv:1002.0332 [astro-ph]

—. 2012, Annual Review of Astronomy and Astrophysics, 50, 411
Gaudi, B. S., & Gould, A. 1997, The Astrophysical Journal, 486, 85
Gerber, F., & Furrer, R. 2019, The R Journal, 11, 352
Germain, M., Gregor, K., Murray, I., & Larochelle, H. 2015, in International Conference on Machine Learning, 881
Godines, D., Bachelet, E., Narayan, G., & Street, R. A. 2019, Astronomy and Computing, 28, 100298
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Gott, III, J. R. 1981, The Astrophysical Journal, 243, 140
Gould, A. 2008, The Astrophysical Journal, 681, 1593
—. 2022
Gould, A., & Loeb, A. 1992, The Astrophysical Journal, 396, 104
Gould, A., Dong, S., Gaudi, B. S., et al. 2010, The Astrophysical Journal, 720, 1073
Gould, A., Han, C., Zang, W., et al. 2022, Astronomy & Astrophysics, 664, A13
Griest, K., & Safizadeh, N. 1998, The Astrophysical Journal, 500, 37
Guinan, E. F., Ribas, I., Fitzpatrick, E. L., et al. 2000, The Astrophysical Journal, 544, 409
Hahn, C., Vakili, M., Walsh, K., et al. 2017, Monthly Notices of the Royal Astronomical Society, 469, 2791
Han, C. 2008, The Astrophysical Journal, 691, L9
Han, C., Chang, H.-Y., An, J. H., & Chang, K. 2001, Monthly Notices of the Royal Astronomical Society, 328, 986
Han, C., Udalski, A., Gould, A., et al. 2017, The Astronomical Journal, 154, 133
Han, C., Bond, I. A., Gould, A., et al. 2018, The Astronomical Journal, 156, 226
Han, C., Udalski, A., Kim, D., et al. 2020a, Astronomy and Astrophysics, 642, A110
Han, C., Kim, D., Udalski, A., et al. 2020b, The Astronomical Journal, 160, 64
Han, C., Udalski, A., Gould, A., et al. 2020c, The Astronomical Journal, 159, 91
Han, C., Udalski, A., Kim, D., et al. 2021, Astronomy and Astrophysics, 650, A89
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770
Herrera-Martín, A., Albrow, M. D., Udalski, A., et al. 2020, The Astronomical Journal, 159, 256
Hirao, Y., Udalski, A., Sumi, T., et al. 2016, The Astrophysical Journal, 824, 139
—. 2017, The Astronomical Journal, 154, 1
Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735
Hoffman, M. D., & Gelman, A. 2014, The Journal of Machine Learning Research, 15, 1593
Hsu, D. C., Ford, E. B., Ragozzine, D., & Morehead, R. C. 2018, The Astronomical Journal, 155, 205
Huchra, J., Gorenstein, M., Kent, S., et al. 1985, The Astronomical Journal, 90, 691
Hwang, K.-H., Ryu, Y.-H., Kim, H.-W., et al. 2019, The Astronomical Journal, 157, 23
Hwang, K.-H., Zang, W., Gould, A., et al. 2022, The Astronomical Journal, 163, 43
Irwin, M. J., Webster, R. L., Hewett, P. C., Corrigan, R. T., & Jedrzejewski, R. I. 1989, The

Astronomical Journal, 98, 1989
Jaderberg, M., Simonyan, K., Zisserman, A., & kavukcuoglu, k. 2015, in Advances in Neural Information Processing Systems, Vol. 28
Jamal, S., & Bloom, J. S. 2020, The Astrophysical Journal Supplement Series, 250, 30
Janczak, J., Fukui, A., Dong, S., et al. 2010, The Astrophysical Journal, 711, 731
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 3145
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, Monthly Notices of the Royal Astronomical Society, 486, 1907
Khakpash, S., Penny, M., & Pepper, J. 2019, The Astronomical Journal, 158, 9
Kim, D.-W., & Bailer-Jones, C. A. L. 2016, Astronomy & Astrophysics, 587, A18
Kim, E. J., & Brunner, R. J. 2017, Monthly Notices of the Royal Astronomical Society, 464, 4463
Kim, Y. H., Chung, S.-J., Yee, J. C., et al. 2021, The Astronomical Journal, 162, 17
Kingma, D. P., & Ba, J. 2015, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, ed. Y. Bengio & Y. LeCun
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems, Vol. 25
Krueger, D., Maharaj, T., Kramár, J., et al. 2017, in 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings
Kunder, A., Popowski, P., Cook, K. H., & Chaboyer, B. 2008, The Astronomical Journal, 135, 631
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, Publications of the Astronomical Society of the Pacific, 121, 1395
Le, Q. V., Jaitly, N., & Hinton, G. E. 2015, arXiv:1504.00941 [cs]
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278
Liebes, S. 1964, Physical Review, 133, B835
Lomb, N. R. 1976, Astrophysics and Space Science, 39, 447
Loshchilov, I., & Hutter, F. 2017, in 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings
Louppe, G., Cho, K., Becot, C., & Cranmer, K. 2019, Journal of High Energy Physics, 2019, 57
Mao, S., & Paczyński, B. 1991, The Astrophysical Journal, 374, L37
Mattheakis, M., Protopapas, P., Sondak, D., Di Giovanni, M., & Kaxiras, E. 2020, arXiv:1904.08991 [physics]
McNamara, D. H., Clementini, G., & Marconi, M. 2007, The Astronomical Journal, 133, 2752
Miller, A. A., Bloom, J. S., Richards, J. W., et al. 2015, The Astrophysical Journal, 798, 122

Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, New Astronomy, 15, 433
Miyazaki, S., Sumi, T., Bennett, D. P., et al. 2018, The Astronomical Journal, 156, 136
Mróz, P. 2020, Acta Astronomica, 70, 169
Mróz, P., Udalski, A., Skowron, J., et al. 2017, Nature, 548, 183
Möller, A., & de Boissière, T. 2020, Monthly Notices of the Royal Astronomical Society, 491, 4277
Nagakane, M., Sumi, T., Koshimoto, N., et al. 2017, The Astronomical Journal, 154, 35
Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, The Astrophysical Journal Supplement Series, 236, 9
Naul, B., Bloom, J. S., Pérez, F., & Walt, S. v. d. 2018, Nature Astronomy, 2, 151
Naul, B., Walt, S. v. d., Crellin-Quick, A., Bloom, J. S., & Pérez, F. 2016, in Proceedings of the 15th Python in Science Conference, ed. S. Benthall & S. Rostrup, 27
Nelder, J. A., & Mead, R. 1965, The Computer Journal, 7, 308
North, P., Gauderon, R., Barblan, F., & Royer, F. 2012, Astronomy and Astrophysics, 540, C1
Nucita, A. A., Licchelli, D., De Paolis, F., et al. 2018, Monthly Notices of the Royal Astronomical Society, 476, 2962
Nun, I., Pichara, K., Protopapas, P., & Kim, D.-W. 2014, The Astrophysical Journal, 793, 23
Nun, I., Protopapas, P., Sim, B., et al. 2015, arXiv:1506.00010 [astro-ph]
Oord, A. v. d., Dieleman, S., Zen, H., et al. 2016, in Proc. 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9), 125
Orellana, A. G., & Michele, C. D. 2020, ACM Transactions on Mathematical Software, 46, 1
Paczyński, B. 1986a, The Astrophysical Journal, 301, 503
—. 1986b, The Astrophysical Journal, 304, 1
—. 1991, The Astrophysical Journal, 371, L63
Paczyński, B. 1997, in The Extragalactic Distance Scale, ed. M. Livio, M. Donahue, & N. Panagia, 273
Papamakarios, G., & Murray, I. 2016, in Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, 1036
Papamakarios, G., Pavlakou, T., & Murray, I. 2017, in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 2335
Papamakarios, G., Sterratt, D., & Murray, I. 2019, in Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, 837
Penny, M. T., Scott Gaudi, B., Kerins, E., et al. 2019, The Astrophysical Journal Supplement Series, 241, 3
Poleski, R., & Yee, J. C. 2019, Astronomy and Computing, 26, 35
Pooley, G. G., Browne, I. W. A., Daintree, E. J., et al. 1979, Nature, 280, 461
Ranc, C., Bennett, D. P., Hirao, Y., et al. 2019, The Astronomical Journal, 157, 232
Rattenbury, N. J., Bennett, D. P., Sumi, T., et al. 2017, Monthly Notices of the Royal Astronomical Society, 466, 2710

Refsdal, S. 1964, Monthly Notices of the Royal Astronomical Society, 128, 295
Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, The Astrophysical Journal, 733, 10
Ryu, Y.-H., Jung, Y. K., Yang, H., et al. 2022, The Astronomical Journal, 164, 180
Scargle, J. D. 1982, The Astrophysical Journal, 263, 835
Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses
Shallue, C. J., & Vanderburg, A. 2018, The Astronomical Journal, 155, 94
Skilling, J. 2006, Bayesian Analysis, 1, 833
Skowron, D. M., Skowron, J., Mróz, P., et al. 2019, Science, 365, 478
Skowron, J., & Gould, A. 2012, arXiv:1203.1034 [astro-ph]
Skowron, J., Ryu, Y. H., Hwang, K. H., et al. 2018, Acta Astronomica, 68, 43
Song, Y.-Y., Mao, S., & An, J. H. 2014, Monthly Notices of the Royal Astronomical Society, 437, 4006
Soszyński, I., Poleski, R., Udalski, A., et al. 2008, Acta Astronomica, 58, 163
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, Acta Astronomica, 59, 1
Soszyński, I., Pawlak, M., Pietrukowicz, P., et al. 2016, Acta Astronomica, 66, 405
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757 [astro-ph]
Strobach, P. 2010, Journal of Computational and Applied Mathematics, 234, 3007
Suzuki, D., Udalski, A., Sumi, T., et al. 2013, The Astrophysical Journal, 780, 123
Tachibana, Y., Graham, M. J., Kawai, N., et al. 2020, arXiv:2003.01241 [astro-ph]
Thomas, O., Dutta, R., Corander, J., Kaski, S., & Gutmann, M. U. 2022, Bayesian Analysis, 17
Torres, G., & Ribas, I. 2002, The Astrophysical Journal, 567, 1140
Tsang, B. T.-H., & Schultz, W. C. 2019, The Astrophysical Journal, 877, L14
Udalski, A. 2003, Acta Astronomica, 53, 291
Udalski, A., Szymanski, M., Kaluzny, J., et al. 1993, Acta Astronomica, 43, 289
Vermaak, P. 2003, Monthly Notices of the Royal Astronomical Society, 344, 651
Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381
Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, The Astrophysical Journal, 764, 116
Wisdom, S., Powers, T., Hershey, J. R., Roux, J. L., & Atlas, L. 2016, in Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, 4887
Witt, H. J. 1990, Astronomy and Astrophysics, 236, 311
Witt, H. J., & Mao, S. 1995, The Astrophysical Journal Letters, 447, L105
Woźniak, P., & Paczyński, B. 1997, The Astrophysical Journal, 487, 55
Wyrzykowski, , Kozłowski, S., Skowron, J., et al. 2009, Monthly Notices of the Royal Astronomical Society, 397, 1228
Wyrzykowski, , Skowron, J., Kozłowski, S., et al. 2011, Monthly Notices of the Royal Astronomical Society, 416, 2949
Wyrzykowski, , Rynkiewicz, A. E., Skowron, J., et al. 2015, The Astrophysical Journal Supplement Series, 216, 12
Yang, H., Zang, W., Gould, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 516, 1894
Yee, J. C., Zang, W., Udalski, A., et al. 2021, The Astronomical Journal, 162, 180

Yeo, I.-K. 2000, Biometrika, 87, 954
Zang, W., Hwang, K.-H., Kim, H.-W., et al. 2018, The Astronomical Journal, 156, 236
Zang, W., Yang, H., Han, C., et al. 2022, Monthly Notices of the Royal Astronomical Society, 515, 928
Zhang, K. 2023, arXiv:2207.12412 [astro-ph]
Zhang, K., & Bloom, J. S. 2020, The Astrophysical Journal, 889, 24
—. 2021, Monthly Notices of the Royal Astronomical Society, 505, 515
Zhang, K., Bloom, J. S., Gaudi, B. S., et al. 2020, arXiv:2010.04156 [astro-ph, physics:physics]
—. 2021, The Astronomical Journal, 161, 262
Zhang, K., & Gaudi, B. S. 2022, The Astrophysical Journal Letters, 936, L22
Zhang, K., Gaudi, B. S., & Bloom, J. S. 2022, Nature Astronomy, 6, 782
Zwicky, F. 1937, The Astrophysical Journal, 86, 217