



School of Science  
Department of Physics and Astronomy  
Master Degree Programme in Astrophysics and Cosmology

# **Applications of automatic differentiation in gravitational lensing**

Graduation Thesis

March 15, 2024

Presented by:  
**Simone La Porta**

Supervisor:  
**Chiar.mo Prof. Lauro Moscardini**

Co-supervisor:  
**Dott. Massimo Meneghetti**

---

Academic year 2022-2023  
Graduation date V



*Nature uses only the longest threads to  
weave her patterns, so that each small  
piece of her fabric reveals the  
organization of the entire tapestry*

---

Richard P. Feynman



# Abstract

Gravitational lensing, a remarkable consequence of Einstein's theory of general relativity, provides a unique opportunity to explore the fundamental properties of the universe by studying the distortions caused by massive objects on the path of light rays. However, analyzing and modeling gravitational lenses poses many challenges, due to the complex nature of the lensing effects and the vast amount of observational data and computational resources required.

The primary objective of this thesis is to address these challenges by developing advanced Python algorithms based on differentiable programming paradigm, leveraging the capabilities of PyTorch and TensorFlow frameworks to enable precise modeling and analysis of gravitational lenses. By employing parametric models, these algorithms exploit automatic differentiation to backpropagate errors and compute gradients of a loss function, facilitating the optimization of high-dimensional parameter spaces. Through the training of these parametric models, relevant features can be extracted and key parameters of the lensing system can be estimated. The resulting models can then be applied to real observational data, improving the characterization and classification of strong lenses with enhanced accuracy and efficiency.

The use of PyTorch and TensorFlow in the implementation of these algorithms allows efficient utilization of modern computational resources, such as GPUs, to handle the inherent computational complexity involved in strong lens analysis. Furthermore, the flexibility and extensibility of these frameworks enable seamless integration with other astrophysical and computational tools, facilitating a comprehensive and robust analysis of strong lenses.

The structure of this thesis is the following: an introduction to the main concepts of cosmology and gravitational lensing theory is presented in Chapters 1 and 2, followed by an extensive description of the most important lens models in Chapter 3. Chapter 4 describes the differentiable programming paradigm, its characteristics, and some examples of algorithms to model and analyze gravitational lenses and light sources. Finally, Chapter 5 specifically focus on the application of differentiable programming and automatic differentiation methods to microlensing and strong lensing optimization problems. Finally, an example of surface brightness fitting is presented as a means to derive the shape and ellipticity of a galaxy, fundamental information for performing weak lensing measurements.



# Sommario

Il lensing gravitazionale, una straordinaria conseguenza della teoria della relatività generale di Einstein, offre un'opportunità unica di esplorare le proprietà fondamentali dell'universo, studiando le distorsioni causate da oggetti massicci sul percorso dei raggi luminosi. Il lensing gravitazionale forte, in particolare, che si manifesta come immagini multiple altamente amplificate e distorte di sorgenti di background, offre preziose indicazioni sulla distribuzione della materia oscura e sulla formazione delle strutture cosmiche. Tuttavia, analizzare e modellare con precisione le lenti gravitazionali forti presenta varie difficoltà, a causa della natura complessa degli effetti di lensing e della vasta quantità di dati osservativi richiesti, oltre alle grandi risorse computazionali necessarie. L'obiettivo principale di questo lavoro di tesi è quello di affrontare tali questioni sviluppando algoritmi Python avanzati per la precisa modellizzazione e analisi di lenti gravitazionali forti, fondati su tecniche di programmazione differenziabile, implementate utilizzando i framework PyTorch e TensorFlow. Utilizzando modelli parametrici, questi algoritmi sfruttano la differenziazione automatica per retropropagare gli errori e calcolare i gradienti di una funzione di costo, facilitando l'ottimizzazione nello spazio dei parametri. Attraverso l'implementazione in tali modelli parametrici, è possibile estrapolare le caratteristiche rilevanti e stimare i parametri chiave del sistema in esame. I modelli risultanti possono essere applicati a dati osservativi reali, migliorando la caratterizzazione e la classificazione delle lenti con maggiore precisione ed efficienza. L'uso di PyTorch e TensorFlow nell'implementazione di questi algoritmi consente di utilizzare in modo efficiente le moderne risorse di calcolo, come le GPU, per gestire la complessità computazionale intrinseca all'analisi delle lenti forti. Inoltre, la flessibilità e l'estensibilità di questi framework consentono una perfetta integrazione con altri strumenti astrofisici e di calcolo, facilitando un'analisi completa e robusta.

La struttura di questa tesi è la seguente: nei Capitoli 1 e 2 viene presentata un'introduzione ai concetti principali della cosmologia e della teoria delle lenti gravitazionali, seguita da un'ampia descrizione dei più importanti modelli di lente nel Capitolo 3. Il Capitolo 4 descrive il paradigma della programmazione differenziabile, le sue caratteristiche e alcuni esempi di algoritmi per modellare e analizzare le lenti gravitazionali. Il Capitolo 5 si concentra specificamente sull'applicazione dei metodi di differenziazione automatica a problemi di ottimizzazione di microlensing e di lensing forte. Infine, viene presentato un esempio di analisi della brillanza superficiale per derivare la forma e l'ellitticità di una galassia, informazioni fondamentali per effettuare misurazioni di weak lensing.



# Contents

<b>Abstract</b>	i
<b>Sommario</b>	ii
<b>List of Figures</b>	iv
<b>List of Tables</b>	vi
<b>1 Cosmological background</b>	1
1.1 History . . . . .	1
1.2 Friedmann Equations . . . . .	2
1.2.1 Expansion rate . . . . .	2
1.2.2 Dynamics of the expansion . . . . .	3
1.2.3 Cosmological distances . . . . .	5
<b>2 Gravitational lensing</b>	7
2.1 Light deflection . . . . .	7
2.2 Lens equation . . . . .	10
2.2.1 Lensing potential and convergence . . . . .	12
2.2.2 First-order lens mapping . . . . .	13
2.2.3 Magnification . . . . .	17
2.2.4 Time-delay and multiple images . . . . .	18
<b>3 Lens models</b>	22
3.1 Microlenses . . . . .	22
3.1.1 Deflection angle and lensing potential . . . . .	22
3.1.2 Lens equation and multiple images . . . . .	23
3.1.3 Critical lines, caustics and magnification . . . . .	24
3.1.4 Microlensing light-curve . . . . .	25

3.2	Extended lenses . . . . .	28
3.2.1	Axially symmetric profiles . . . . .	28
3.2.2	Power-law profiles . . . . .	29
3.2.3	Singular Isothermal Sphere . . . . .	31
3.2.4	Non-singular Isothermal Sphere . . . . .	33
3.2.5	Singular Isothermal Ellipsoid . . . . .	35
3.2.6	Non-singular Isothermal Ellipsoid . . . . .	40
3.2.7	External shear . . . . .	41
<b>4</b>	<b>Algorithms</b>	<b>43</b>
4.1	Differentiable programming . . . . .	43
4.1.1	Automatic differentiation . . . . .	45
4.1.2	Backpropagation algorithm . . . . .	49
4.2	Light sources modeling . . . . .	53
4.2.1	Sérsic profile . . . . .	54
4.2.2	Core-Sérsic profile . . . . .	54
4.2.3	Sky component . . . . .	55
4.2.4	Point Spread Function . . . . .	57
4.3	Lens reconstruction . . . . .	58
4.3.1	Parametric reconstruction . . . . .	60
<b>5</b>	<b>Applications</b>	<b>65</b>
5.1	Microlensing light-curve fit . . . . .	65
5.2	Strong lens parametric reconstruction . . . . .	73
5.2.1	Lens modelling . . . . .	74
5.2.2	Solving the lens equation . . . . .	75
5.2.3	Lens optimization . . . . .	76
5.3	Surface brightness and galaxy shapes . . . . .	83
5.3.1	Fit of Sérsic profile of surface brightness . . . . .	84
<b>6</b>	<b>Summary and conclusions</b>	<b>93</b>
<b>Bibliography</b>		<b>viii</b>
<b>Ringraziamenti</b>		<b>xvi</b>

# List of Figures

1.1	Hubble's Law: radial velocities of Extra-Galactic Nebulae . . . . .	2
1.2	Temporal evolution of the cosmic scale factor . . . . .	5
2.1	Born approximation schematics . . . . .	9
2.2	Lensing system geometry . . . . .	11
2.3	Lensing distance variation with redshifts . . . . .	14
2.4	Convergence and shear distortion on circular source . . . . .	15
2.5	Images orientations according to different combinations of $\gamma_1$ and $\gamma_2$ . . . . .	16
2.6	Caustics and critical lines shapes . . . . .	18
2.7	The lensed pair S1–S2 in AC114 . . . . .	19
2.8	Time delay functions and surfaces . . . . .	21
2.9	Einstein radius . . . . .	21
3.1	Lens equation solutions for point-mass lens . . . . .	23
3.2	Lens position and source trajectory in a microlensing event . . . . .	26
3.3	Microlensing light-curves for different impact parameters . . . . .	27
3.4	Image diagrams for power-law lenses with $1 < n < 2$ . . . . .	31
3.5	Image diagrams for power-law lenses with $n > 2$ . . . . .	31
3.6	Imaging of an extended source by a SIS lens . . . . .	33
3.7	NIS caustic critical lines radii as a function of $x_c$ . . . . .	34
3.8	SIE caustics and critical lines . . . . .	37
3.9	Image multiplicity for a SIE lens . . . . .	38
3.10	Lensing of circular sources by a SIE . . . . .	39
3.11	NIE caustics topologies . . . . .	40
4.1	Summary of techniques to calculate derivatives . . . . .	45
4.2	Example of computational graph . . . . .	47
4.3	Forward mode AD example . . . . .	48

4.4	Reverse mode AD example . . . . .	49
4.5	One/two-dimensional gradient descents . . . . .	51
4.6	Examples of learning rates too small or too high . . . . .	51
4.7	Optimization algorithm diagram . . . . .	53
4.8	Sérsic profiles for different values of $n$ . . . . .	55
4.9	Core-Sérsic profiles for different values of $\alpha$ . . . . .	56
4.10	Gaussian PSF and FWHM . . . . .	58
4.11	Strong gravitational lensing observations . . . . .	59
4.12	Simulated image of a strong lensing event with multiple images . . . . .	62
5.1	Microlensing initial predictions and residuals . . . . .	67
5.2	Microlensing corner plot showing degeneracy . . . . .	70
5.3	Microlensing initial predictions and residuals . . . . .	71
5.4	Microlensing corner plot non degenerate parameters . . . . .	72
5.5	Microlensing loss value as a function of epochs . . . . .	73
5.6	Simulated strong lensing event . . . . .	77
5.7	Initial sources predictions source plane optimization . . . . .	78
5.8	Best-fit model loss history and image positions for source plane optimization	79
5.9	Best-fit model source and image plane . . . . .	80
5.10	Multiple images prediction for image plane optimization . . . . .	81
5.11	Best-fit model loss history and image positions for image plane optimization	82
5.12	Simulated Sérsic surface brightness: theoretical + noisy . . . . .	87
5.13	Simulated Sérsic surface brightness vs radius . . . . .	88
5.14	Initial prediction Sérsic profile + residuals . . . . .	89
5.15	Best-fit prediction Sérsic profile + residuals . . . . .	89
5.16	Loss function vs epochs for Sérsic profile fit . . . . .	90
5.17	Corner plot Sérsic profile . . . . .	92

# List of Tables

5.1	Summary of the parameters before and after fitting the model with the set of degenerate parameters. . . . .	68
5.2	Summary of the parameters before and after fitting the model a second time to avoid degeneracy. . . . .	69
5.3	Summary of lens and source parameters. . . . .	74
5.4	Summary of the parameters before and after fitting the model using both source plane or image plane optimization. . . . .	76
5.5	Summary of true and initial parameters before fitting. . . . .	88
5.6	Summary of true, initial and best-fit parameters. . . . .	91



# 1

## Cosmological background

### 1.1 History

Throughout history, questions surrounding the origin, age, and size of the universe have fascinated humans. Plato believed in a Universe that remains constant, envisioning it as an entity that was created perfect, unaging, and free from decay ([Plato, 2010](#)). This view of a static universe prevailed for more than two millennia, and it was so entrenched in cosmological thinking that even Albert Einstein initially subscribed to it. To reconcile his field equations of general relativity with the notion of a static universe, Einstein introduced the cosmological constant, a term that provided a mathematical means to allow for static solutions to his equations ([Einstein, 1917](#)).

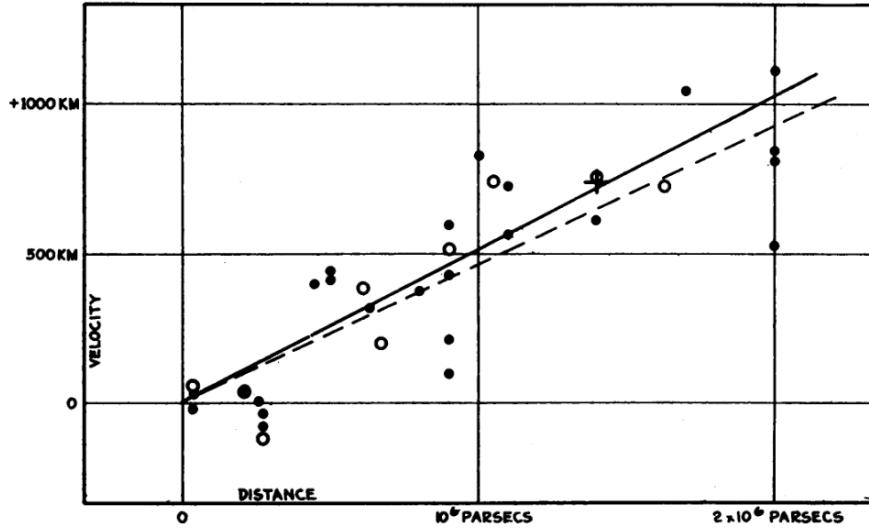
However, the early 20th century brought discoveries that challenged the long-standing paradigm of a static universe. Vesto Slipher made pivotal observations, noting that most galaxies recede from the Milky Way at high velocities. Building on this foundation, Edwin Hubble, in the 1920s, conducted an analysis of the escape velocities of distant galaxies, leading to a groundbreaking discovery. Hubble observed that the farther away galaxies are, the faster they appear to be moving away. He plotted the radial velocities of these galaxies against their distances and found that the data could be best described by a straight line, indicating a linear relationship (Fig. 1.1).

Hubble's choice of a linear fit was influenced by his belief in the predictions of Friedmann's equations, which are solutions to Einstein's field equations of general relativity and suggest an expanding universe. Hubble's analysis and the linear relationship he identified, the basis for what is now known as *Hubble's Law*, marked a fundamental shift in the understanding of the universe. It provided strong evidence for an expanding Universe, a concept that fundamentally contradicted the long-held belief in a static cosmos. This discovery not only revolutionized cosmology, but also led Einstein to reconsider the cosmological constant he had introduced.

The receding velocity  $v$  of a galaxy leads to a Doppler shift in the galaxy's spectrum, observable as a redshift  $z$ :

$$v(r) = c \frac{\Delta\lambda}{\lambda} = cz = H_0 r, \quad (1.1)$$

where  $H_0$  is the proportionality constant, subsequently named after Hubble, for which,



**Figure 1.1.** Radial velocities, corrected for solar motion, plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster.  
Credits: [Hubble \(1929\)](#).

in 1929 he published ([Hubble, 1929](#)) a value of

$$H_0 = 530 \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad (1.2)$$

much different from the current value ([Tully, 2023](#)) of

$$H_0 = (74.6 \pm 0.8) \text{ km s}^{-1} \text{ Mpc}^{-1}. \quad (1.3)$$

The unit of the Hubble constant is chosen according to the relation between escape velocity and distance, but actually it has the unit of an inverse time, representing the age of the universe in a uniformly expanding model:

$$\tau_0 = H_0^{-1} \approx 13.9 \cdot 10^9 \text{ yr}. \quad (1.4)$$

## 1.2 Friedmann Equations

### 1.2.1 Expansion rate

According to the *Cosmological Principle*, on sufficiently large scales, the universe is homogeneous and isotropic. Considering a sphere with homogeneous density and a test particle at location  $\vec{x}$ , and introducing a spherical coordinate system that is

allowed to expand with time, due to the cosmological principle, the expansion and the time-dependent position  $\vec{r}(t)$  can be expressed by

$$\vec{r}(t) = a(t)\vec{x}, \quad (1.5)$$

where  $a(t)$  is the *cosmic scale factor*, which does not depend only on time. For  $t_0 = \text{today}$ , the scale factor is conventionally set to  $a(t_0) = 1$ . Scalar quantities  $r, x$  can be used instead of  $\vec{r}, \vec{x}$  due to isotropy.

The velocity of a test particle caused by cosmic expansion is

$$v(r, t) = \frac{dr(t)}{dt} = \frac{da(t)}{dt}x = \frac{\dot{a}(t)}{a(t)}r = H(t)r, \quad (1.6)$$

which leads to the definition of the *Hubble parameter*, used to quantify the relative expansion rate of the universe

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}, \quad (1.7)$$

whose value in the present epoch,  $H(t_0) = H_0$ , is the *Hubble constant*, expressed in Eq. (1.3).

### 1.2.2 Dynamics of the expansion

To derive the Friedmann Equations and to study the evolution of the scale factor  $a(t)$ , to better understand the development of the universe, one has to start from Einstein field's equation, that describes the geometry of space-time:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (1.8)$$

where  $R_{\mu\nu}$  and  $R$  are the Ricci tensor and Ricci scalar, respectively,  $\Lambda$  is the *cosmological constant* and  $T_{\mu\nu}$  is the *energy-momentum tensor*, which includes all contributions of energy and acts as the source of gravity.

The energy-momentum tensor of the universe is that of a homogeneous perfect fluid, characterized by its density  $\rho(t)$  and pressure  $p(t)$ . Using the Robertson-Walker metric to describe homogeneity and isotropy, the Einstein's equations simplify to the

*Friedmann equations* (Friedman, 1922; Friedmann, 1924):

$$H(t)^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda c^2}{3} - \frac{Kc^2}{a^2}, \quad (1.9a)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \quad (1.9b)$$

where  $K$  is a constant parameter that defines the curvature of spatial surfaces:

- $K = -1$  means open, hyperbolic space (i.e. infinite) with negative curvature;
- $K = 0$  means flat, Euclidean space;
- $K = 1$  means closed, spherical space with positive curvature.

It is then possible to split up the density  $\rho$  into

$$\rho_m(t) = \rho_{m,0}a(t)^{-3} \Rightarrow \text{non-relativistic matter}, \quad (1.10a)$$

$$\rho_r(t) = \rho_{r,0}a(t)^{-4} \Rightarrow \text{relativistic matter}. \quad (1.10b)$$

Furthermore, by defining the constant vacuum energy density  $\rho_\Lambda = \frac{\Lambda}{8\pi G}$  and the critical density  $\rho_{cr} = \frac{3H_0^2}{8\pi G}$ , it is convenient to introduce dimensionless density parameters for matter, radiation and vacuum energy:

$$\Omega_m = \frac{\rho_m}{\rho_{cr}}, \quad \Omega_r = \frac{\rho_r}{\rho_{cr}}, \quad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{cr}} = \frac{\Lambda}{3H_0^2}. \quad (1.11)$$

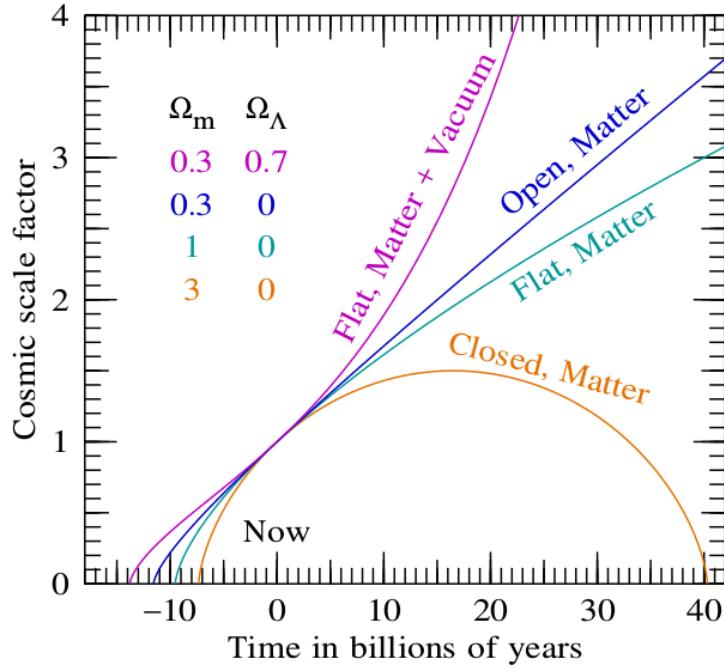
Finally, introducing the curvature parameter

$$\Omega_K = -\frac{Kc^2}{H_0^2} = 1 - (\Omega_m + \Omega_r + \Omega_\Lambda), \quad (1.12)$$

and subtracting Eq. (1.9a) from Eq. (1.9b):

$$H(t)^2 = \left[\frac{\dot{a}(t)}{a(t)}\right]^2 = H_0^2 \left[ \Omega_r a(t)^{-4} + \Omega_m a(t)^{-3} + \Omega_K a(t)^{-2} + \Omega_\Lambda \right]. \quad (1.13)$$

Equation (1.13) is a fundamental equation that completely describes the expansion of the universe and contains all the information about geometry, matter, and energy content.



**Figure 1.2.** Numerical solutions for Eq. (1.13) for different cosmological models.  
Credits: [Hamilton \(2019\)](#).

### 1.2.3 Cosmological distances

In a curved space-time, there are various ways to define the distance between two points. Setting  $P_0$  as the origin of a set of polar coordinates  $(r, \theta, \phi)$  and assuming  $dt = d\phi = d\theta = 0$ , integrating the Friedmann-Robertson-Walker metric

$$ds^2 = c^2 dt^2 - a(t)^2 \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (1.14)$$

in this coordinate system, it is possible to determine the distance measured by a “chain” of observers in every point between  $P_0$  and a generic point  $P$  at time  $t$ . This is the so-called *proper distance*:

$$d_P = \int_0^r \frac{a(t) dr'}{\sqrt{1 - Kr'^2}} = a(t)F(r), \quad (1.15)$$

with

$$F(r) = \begin{cases} \operatorname{arcsinh}(r) & \text{if } K = -1; \\ r & \text{if } K = 0; \\ \operatorname{arcsin}(r) & \text{if } K = 1. \end{cases} \quad (1.16)$$

By computing the proper distance at present time,  $t = t_0$ , it is possible to obtain the *comoving distance*:

$$d_C \equiv d_P(t_0) = a(t_0)F(r), \quad (1.17)$$

and since the proper distance of a source can change in time as a consequence of the time dependence of the scale factor, the source at  $P$  has a radial velocity with respect to  $P_0$ :

$$v_r = \dot{a}(t)F(r) = \frac{\dot{a}(t)}{a(t)}d_P = H(t)d_P, \quad (1.18)$$

which is the Hubble Law mentioned in Section 1.1.

There is no unique way to define the distance of an astronomical object in cosmology, and, in addition to the proper and comoving distances, which are not directly measurable, other kinds of distance can be defined that are, in principle, measurable.

Firstly, one can define the *luminosity distance* of a source at a distance  $r$ , at time  $t$  with emitted power  $L$  and flux  $f$ :

$$d_L \equiv \sqrt{\frac{L}{4\pi f}}. \quad (1.19)$$

Due to the expansion of the Universe it is necessary to take into account time-dilation effect, a stretch of the spherical surface centered on the source, and a cosmological redshift on the photons:

$$d_L = a(t_0)r(1+z). \quad (1.20)$$

Finally, the *angular diameter distance*, very useful for gravitational lensing purposes, is defined as the ratio of the physical diameter  $d_P$  of a source and the angle  $\Delta\theta$  that it subtends:

$$d_A = \frac{d_P}{\Delta\theta} = a(t)r, \quad (1.21)$$

and consequently

$$d_A = d_L \frac{a(t)^2}{a(t_0)^2} = \frac{d_L}{(1+z)^2}, \quad (1.22)$$

given that  $1+z = \frac{a(t_0)}{a(t)}$ .

# 2

## Gravitational Lensing

In the realm of astrophysics and gravitational theory, the concept that gravity affects not only matter but also light has a long history, dating back to Newton's *Opticks* ([Newton, 1704](#)). The initial calculation of the Newtonian deflection of light rays passing near massive bodies was undertaken by Cavendish in 1784, and von Soldner subsequently published these findings in 1804 ([Will, 1988, 2014](#)). It was Albert Einstein himself who, in 1915, as he completed his General Theory of Relativity ([Einstein, 1915](#)), recognized that the Newtonian prediction of light-ray deflection near the Sun was only half the value foreseen by his revolutionary theory. The confirmation of this twice as large deflection during the solar eclipse of May 29, 1919, observed by Eddington ([Dyson & Eddington, 1920; Will, 2015](#)), marked a momentous confirmation of this nascent theory and captivated global attention. Not only did this observation represent the first in a series of triumphs for general relativity, but it also inaugurated the practical application of *gravitational lensing*, a method that would later become a cornerstone of observational astrophysics.

More than a century has passed since that pivotal moment, and gravitational lensing has evolved into a well-established and respected tool within the fields of astronomy and astrophysics. Presently, the study of lensing theory and its applications can be broadly categorized into three distinct components ([Kochanek, 2004](#)): *strong lensing*, characterized by non-linear deflection at the scale of galaxies and galaxy clusters, produces distinct phenomena such as multiple images, arcs, and rings, *weak lensing*, observed on both cluster and cosmological scales, is a subtle linear effect that gently aligns background galaxies with intervening matter. Statistical analysis of the observed distribution of light allows for the extraction of information regarding the distribution of intervening matter. Finally, *microlensing* involves the dynamic fluctuation of light when compact objects pass in front of background sources at scales too minute to be resolved.

### 2.1 Light deflection

According to Einstein's theory of relativity, objects with gravitational pull have the ability to alter the fabric of space-time, resulting in the bending of light rays ([Narayan &](#)

[Bartelmann, 1997](#)) . Gravitational lensing occurs when a substantial mass distribution can effectively curve and amplify the light emitted from a source positioned behind it.

The calculation of light deflection involves the examination of geodesic curves originating from the field equations of general relativity. Light deflection can also be understood through Fermat's principle<sup>1</sup>, similar to how it is described in geometrical optics. The main approach is to consider light deflection within the framework of general relativity as a refraction problem, for which a refractive index  $n$  can be introduced.

To investigate the bending of light and to determine the refractive index, an initial approximation is made by assuming that the lens is “weak” and significantly smaller than the source-lens-observer optical system, an assumption true for nearly all astrophysical scenarios. This “weak field” approximation refers to a lens with a relatively small Newtonian gravitational potential, which means  $\phi \ll c^2$ , where  $c$  is the speed of light. Additionally, it is plausible to assume that light deflection occurs within a region small enough that the expansion of the universe can be disregarded. Leveraging the principle of equivalence, one can select a locally inertial frame in which space-time is flat and described by Minkowski's metric. In this context, the line element of the local metric tensor can be written as a small perturbation of the metric, such as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = \left(1 + \frac{2\phi}{c^2}\right) c^2 dt^2 - \left(1 - \frac{2\phi}{c^2}\right) d\vec{x}^2 . \quad (2.1)$$

Since light travels on null geodesics, for which  $ds^2 = 0$ , the light speed in the gravitational perturbation is thus

$$c' = \frac{d\vec{x}}{dt} \approx c \left(1 + \frac{2\phi}{c^2}\right) , \quad (2.2)$$

and given that  $\phi \leq 0$ ,  $c'$  is smaller than in the absence of a gravitational potential.

This leads to the definition of the *effective refraction index* as

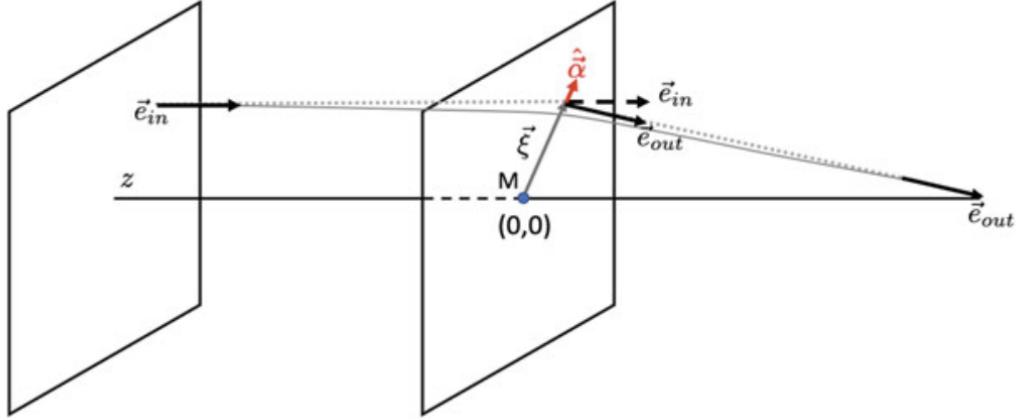
$$n = \frac{c}{c'} \approx 1 - \frac{2\phi}{c^2} \geq 1 . \quad (2.3)$$

Applying Fermat's principle, the total deflection angle of a photon is the integral over the gradient of the potential perpendicular to the light path along the proper light path. Thanks to the *Born approximation*<sup>2</sup>, it can be shown ([Schneider et al., 1992](#)) that

---

<sup>1</sup>Light travels between two points along the path that requires the least time.

<sup>2</sup>Simplification valid when the gravitational potential is small: deflection of light is treated like a linear process, neglecting higher-order corrections.



**Figure 2.1.** Schematics for the Born approximation.  
Credits: [Meneghetti \(2021\)](#).

the deflection angle can be obtained integrating over the unperturbed light path:

$$\hat{\alpha}(\vec{\xi}) = \frac{2}{c^2} \int_{-\infty}^{+\infty} \vec{\nabla}_\perp \phi(\vec{\xi}, z) dz , \quad (2.4)$$

where  $\xi$  is the impact parameter of the photon traveling along the  $\vec{e}_z$  direction that passes through the lens at  $z = 0$  (Fig. 2.1).

If the lens can be approximated by a point mass, the potential is  $\phi = -\frac{GM}{\sqrt{\xi^2 + z^2}}$ , where  $G$  is the gravitational constant, and thus Eq. (2.4) simplifies to the following:

$$\hat{\alpha}(\vec{\xi}) = \frac{4GM}{c^2 \xi} \vec{e}_\xi = \frac{4GM}{c^2 \xi^2} \vec{\xi} . \quad (2.5)$$

Under the assumption of weak field, the superposition principle can be applied to extend the previous definition to calculate the deflection angle of an ensemble of point-like lenses. Having a sparse distribution of  $N$  point masses on a plane, the deflection angle of a light ray crossing the plane at  $\vec{\xi}$  will be

$$\hat{\alpha}(\vec{\xi}) = \sum_i^N \hat{\alpha}_i(\vec{\xi} - \vec{\xi}_i) = \frac{4G}{c^2} \sum_i^N M_i \frac{\vec{\xi} - \vec{\xi}_i}{|\vec{\xi} - \vec{\xi}_i|^2} , \quad (2.6)$$

where  $\vec{\xi}_i$  and  $M_i$  are the positions and masses of each lens.

When discussing more realistic situations of lensing, such as those involving three-dimensional distributions of matter, it is important to note that the physical extent of the lens is typically much smaller than the distances between the observer, the lens, and

the source. Consequently, the bending of light occurs primarily over a brief segment of its path. This observation allows for the application of the *thin screen approximation*, where the lens is represented as a planar distribution of matter, termed the lens plane. In this simplified model, the distribution of matter responsible for lensing is completely characterized by its surface density

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z) dz , \quad (2.7)$$

where  $\vec{\xi}$  is a two-dimensional vector on the lens plane and  $\rho$  is the three-dimensional density.

In this context, the total deflection angle can be obtained by summing the contributions of all the mass elements  $\Sigma(\vec{\xi})d^2\xi$ :

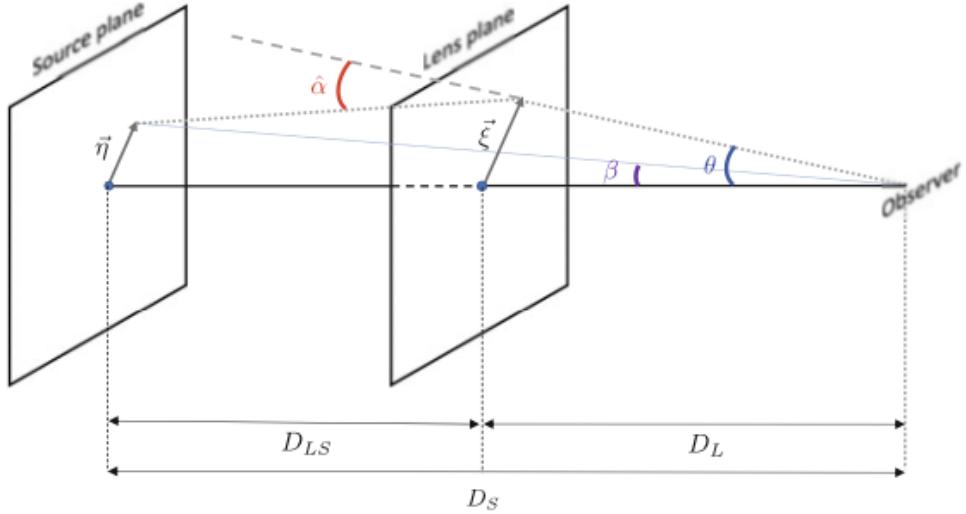
$$\hat{\alpha}(\vec{\xi}) = \frac{4G}{c^2} \int \frac{(\vec{\xi} - \vec{\xi}')\Sigma(\vec{\xi}')}{|\vec{\xi} - \vec{\xi}'|^2} d^2\xi' . \quad (2.8)$$

## 2.2 Lens equation

In order to define the observable light path, the connection between the observed and true positions of a source during a gravitational lensing event must be explored. Without the presence of the lens, light from a distant source would travel directly to an observer who would perceive the source at a specific location in the sky, denoted by  $\vec{\beta}$  (measured in angular units), representing the source's *intrinsic* position. However, when the gravitational lens causes a deflection of the photons, the observer detects them coming from an altered direction  $\vec{\theta}$ , which is known as the *apparent* (or *observed*) *image* position of the source.

Following the typical gravitational lensing geometry, depicted in Fig. 2.2, a mass is placed at redshift  $z_L$ , which equates to an angular diameter distance  $D_L$ . This mass acts as a lens, deflecting light rays coming from a source located at redshift  $z_S$  (corresponding to an angular distance  $D_S$ ). An observer at  $z_O = 0$  gathers the photons originating from this distant source. The angular diameter distance from the lens to the source is denoted as  $D_{LS}$ .

A source located at the intrinsic angular position  $\vec{\beta}$ , which lies on the source plane at a distance  $\vec{\eta} = \vec{\beta}D_S$  from the optical axis, emits photons that pass through the lens plane at a point  $\vec{\xi} = \vec{\theta}D_L$ . These photons are then deflected by an angle  $\hat{\alpha}$ , ultimately reaching the observer. The magnitude of this deflection is specified by Eq. (2.4). As a result of this deflection, the observer perceives the light as originating from the apparent



**Figure 2.2.** Typical lensing system geometry.  
Credits: [Meneghetti \(2021\)](#).

angular position  $\vec{\theta}$ .

When the angles  $\vec{\theta}$ ,  $\vec{\beta}$ ,  $\hat{\vec{\alpha}}$  are small, the actual position of the source and where it appears to be located in the sky are connected by a straightforward relationship, the so-called *lens equation*:

$$\vec{\theta}D_S = \vec{\beta}D_S + \hat{\vec{\alpha}}D_{LS}. \quad (2.9)$$

Defining then the reduced deflection angle as

$$\vec{\alpha}(\vec{\theta}) \equiv \frac{D_{LS}}{D_S} \hat{\vec{\alpha}}(\vec{\theta}), \quad (2.10)$$

Eq. (2.9) can be rewritten as

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}), \quad (2.11)$$

which allows to determine the intrinsic source position if the image position and the deflection angle are known. In contrast, by solving the lens equation for the unknown  $\vec{\theta}$ , it is possible to compute the image position(s) of a source placed at  $\vec{\beta}$ , lensed by a lens with a deflection field  $\vec{\alpha}(\vec{\theta})$ .

A common choice is to write the lens equation in its dimensionless form by defining a length scale  $\xi_0$  on the lens plane and a corresponding length scale  $\eta_0 = \xi_0 D_S / D_L$  on

the source plane and using them to define the dimensionless (scaled) vectors

$$\vec{x} \equiv \frac{\vec{\xi}}{\xi_0}, \quad \vec{y} \equiv \frac{\vec{\eta}}{\eta_0}, \quad (2.12)$$

as well as the dimensionless deflection angle

$$\vec{\alpha}(\vec{x}) = \frac{D_L D_{LS}}{\xi_0 D_S} \hat{\vec{\alpha}}(\xi_0 \vec{x}). \quad (2.13)$$

By substituting these dimensionless quantities into Eq. (2.11), the dimensionless lens equation can be obtained

$$\vec{y} = \vec{x} - \vec{\alpha}(\vec{x}). \quad (2.14)$$

### 2.2.1 Lensing potential and convergence

The deflection properties of an extended distribution of matter are defined by its *effective lensing potential*, namely the properly re-scaled projection on the lens plane of the three-dimensional Newtonian potential  $\Phi$ :

$$\hat{\Psi}(\vec{\theta}) = \frac{2}{c^2} \frac{D_{LS}}{D_L D_S} \int \Phi(D_L \vec{\theta}, z) dz. \quad (2.15)$$

It can be shown that the effective lensing potential satisfies two important properties:

1. the gradient of  $\hat{\Psi}$  is the reduced deflection angle:

$$\vec{\nabla}_{\theta} \hat{\Psi}(\vec{\theta}) = \vec{\alpha}(\vec{\theta}); \quad (2.16)$$

2. the Laplacian of  $\hat{\Psi}$  is twice the *convergence*  $\kappa$ :

$$\nabla_{\theta}^2 \hat{\Psi}(\vec{\theta}) = 2\kappa(\vec{\theta}). \quad (2.17)$$

The convergence is defined as a dimensionless surface density

$$\kappa(\vec{\theta}) \equiv \frac{\Sigma(\vec{\theta})}{\Sigma_{cr}} \quad \text{with} \quad \Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}}, \quad (2.18)$$

where  $\Sigma_{cr}$  is the so-called *critical surface density*, a pivotal characteristic of the lens system, a function of the angular diameter distances of the lens and source, which allows to discriminate between *strong* and *weak* gravitational lensing regimes.

In particular, strong lensing is usually observed in regions of high mass concentration, such as the cores of galaxy clusters or around massive galaxies, when the surface mass

density of the lens is comparable to or exceeds the critical surface density ( $\Sigma \geq \Sigma_{cr}$ ) and the gravitational field of the lens is strong enough to produce multiple images of the source, form Einstein rings, or create arc-like structures. On the contrary, weak lensing occurs when the surface mass density of the lens is below the critical surface density ( $\Sigma < \Sigma_{cr}$ ). The gravitational effect is subtler, leading to slight distortions in the shapes of background galaxies, which can only be detected statistically over large populations of galaxies. Weak lensing does not produce multiple images or highly visible arcs but instead slightly stretches the images of background galaxies, causing shear and magnification effects.

From the previous definitions in Eqs. (2.15) and (2.18), lensing quantities such as lensing potential (and therefore deflection angle) and critical surface density strongly depend on distances between observer, lens and source, which in turn depend on the source and lens redshifts. The distance ratio

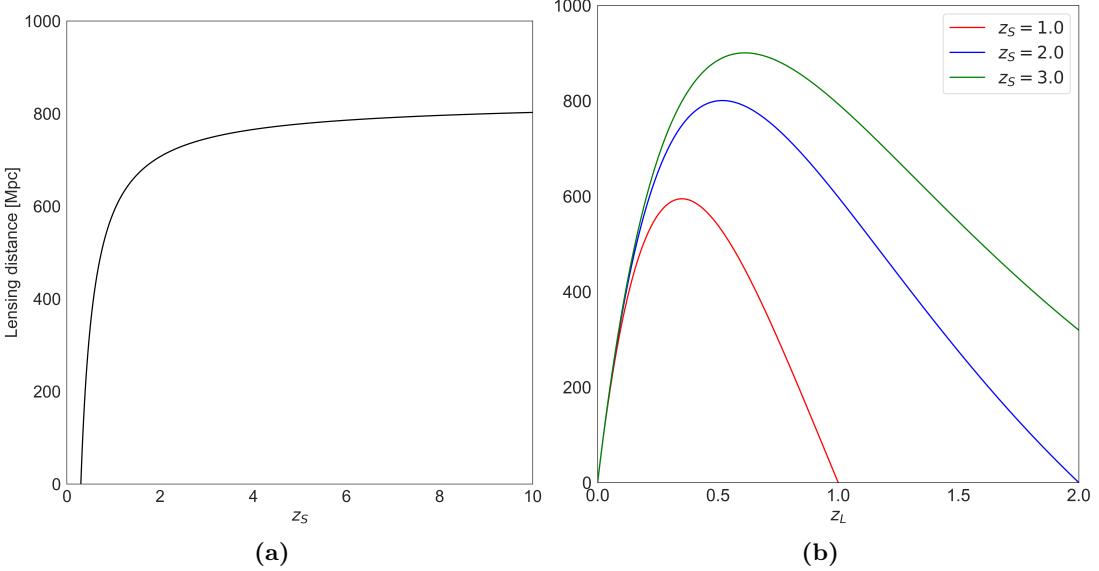
$$\frac{D_L D_{LS}}{D_S} \quad (2.19)$$

is called *lensing distance* and Fig. 2.3 shows how it varies with the redshifts of source and lens. As can be seen, the lensing distance increases with the source redshift and peaks when the lens is at an intermediate distance between the source and the observer. Clearly, the larger the lensing distance (and so the convergence), the stronger the effects generated by the lensing event.

### 2.2.2 First-order lens mapping

A significant effect of gravitational lensing is image distortion. This alteration becomes markedly noticeable with sources of extended size. For instance, galaxies in the background can be observed as elongated arcs when they are lensed by clusters of galaxies or individual galaxies. The cause of this distortion is the varying deflection of light rays. Ideally, by solving the lens equation for every point of the extended source, the configuration of the images can be determined. Specifically, when the size of the source is significantly less than the angular scale over which the lens deflection angle field varies, the correspondence between the position of the source and the images can be approximated as linear on a local scale, allowing for a first-order approximation.

By computing the distance between two points  $\vec{\beta}$  and  $\vec{\beta}' = \vec{\beta} + d\vec{\beta}$  on the source plane, it is possible to define a linear mapping between the source and the lens plane,



**Figure 2.3.** (a) Lensing distance variation as a function of the source redshift and (b) as a function of the lens redshift.

described by the Jacobian matrix

$$A \equiv \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \left( \delta_{ij} - \frac{\partial \alpha_i(\vec{\theta})}{\partial \theta_j} \right) = \left( \delta_{ij} - \frac{\partial^2 \hat{\Psi}(\vec{\theta})}{\partial \theta_i \partial \theta_j} \right). \quad (2.20)$$

This metric is a second-rank symmetric tensor called the *lensing Jacobian* and can be divided into an isotropic and an anisotropic part:

$$A_{iso,ij} = \frac{1}{2} \text{Tr } A \delta_{ij} = \left( 1 - \frac{1}{2} \nabla^2 \hat{\Psi} \right) \delta_{ij} = (1 - \kappa) \delta_{ij}, \quad (2.21a)$$

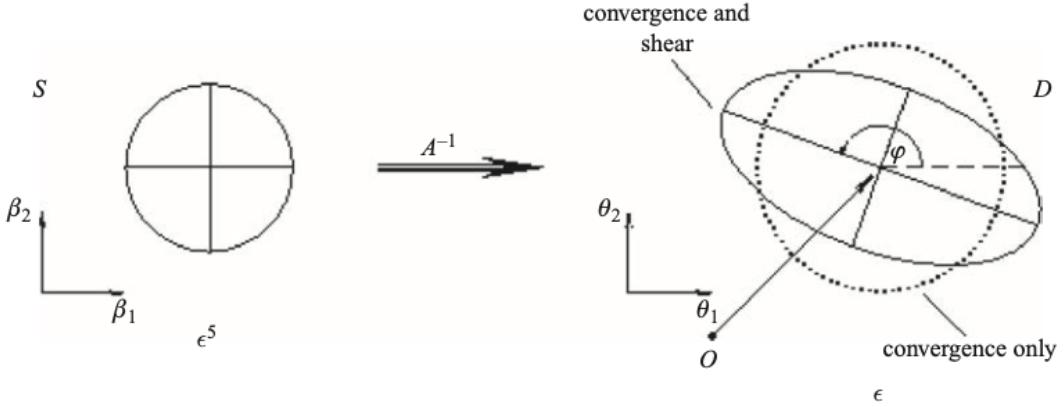
$$A_{aniso,ij} = A_{ij} - \frac{1}{2} \text{Tr } A \delta_{ij} = \begin{pmatrix} -\frac{1}{2} (\hat{\Psi}_{11} - \hat{\Psi}_{22}) & -\hat{\Psi}_{12} \\ -\hat{\Psi}_{12} & \frac{1}{2} (\hat{\Psi}_{11} - \hat{\Psi}_{22}) \end{pmatrix}. \quad (2.21b)$$

It is possible to describe the anisotropic component of the lensing Jacobian by defining the *shear tensor*  $\Gamma$ , a 2x2 symmetric, traceless tensor, usually written in the form of a pseudo-vector  $\vec{\gamma} = (\gamma_1, \gamma_2)$ , whose components are:

$$\gamma_1 = \frac{1}{2} (\hat{\Psi}_{11} - \hat{\Psi}_{22}), \quad (2.22a)$$

$$\gamma_2 = \hat{\Psi}_{12} = \hat{\Psi}_{21}. \quad (2.22b)$$

The shear tensor has eigenvalues  $\pm \sqrt{\gamma_1^2 + \gamma_2^2} = \pm \gamma$  and, representing the direction



**Figure 2.4.** Distortion effects due to convergence and shear on a circular source.  
Credits: [Suyu \(2016\)](#).

of the eigenvector corresponding to the positive eigenvalue with  $\phi$ ,  $\Gamma$  can be written as:

$$\Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} = \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}. \quad (2.23)$$

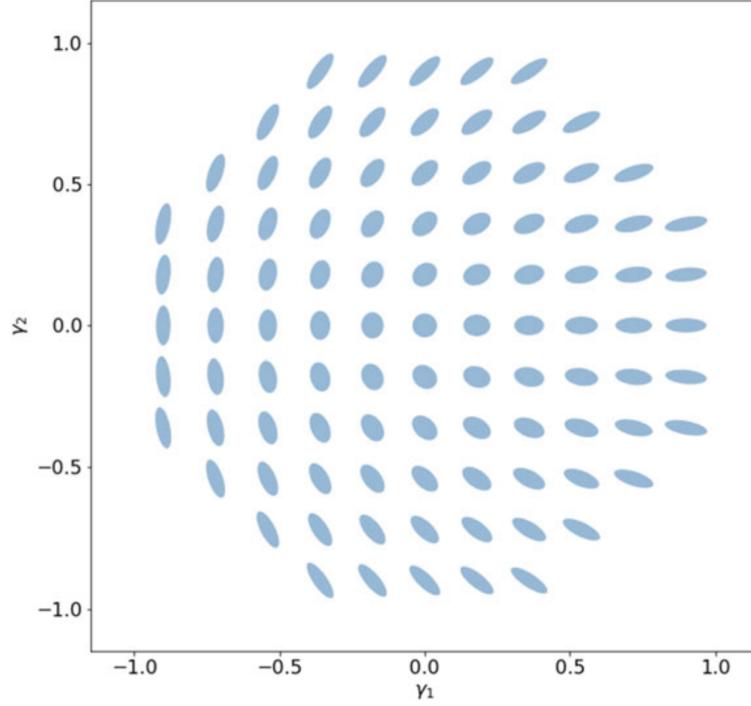
Summarizing, from Eqs. (2.21a) and (2.21b) and considering this last definition of  $\Gamma$ , the lensing Jacobian becomes

$$A = (1 - \kappa) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \quad (2.24)$$

From the last expression of the lensing Jacobian, it can be seen that the first-order mapping is characterized by two parts. The first term, depending on the convergence, describes an isotropic transformation of the source image, which is therefore expanded or contracted by the same factor in all directions. Instead, the second term describes an anisotropic transformation dependent on the shear, and thus causes distortions of the source image along a specific direction, given by the angle  $\phi$ . More precisely, the image size is increased compared to the source in the direction of the eigenvectors of  $A$  with eigenvalue  $\gamma$ , and decreased in the perpendicular direction. The amounts of these magnifications and de-magnifications are given by the inverse of the tangential and radial eigenvalues, namely  $\lambda_t = 1 - \kappa - \gamma$  and  $\lambda_r = 1 - \kappa + \gamma$ .

As an example, Fig. 2.4 shows how a circular source of radius  $r$  would be distorted by convergence and shear effects. Due to the convergence term, the source is mapped to a larger (or smaller) circle, whose radius is  $r/(1 - \kappa)$ .

Due to the shear term, the circle is further elongated in the direction given by the angle



**Figure 2.5.** Images orientations according to different combinations of  $\gamma_1$  and  $\gamma_2$ .  
Credits: [Meneghetti \(2021\)](#).

$\phi$  and contracted in the perpendicular direction to form an ellipse.

This effect can be understood considering the circular source isophotes, described by

$$\beta_1^2 + \beta_2^2 = r^2. \quad (2.25)$$

It is always possible to choose a reference frame where the Jacobian matrix is diagonal, so that the lens equation for a circular source is

$$\beta_1 = (1 - \kappa - \gamma)\theta_1, \quad (2.26a)$$

$$\beta_2 = (1 - \kappa + \gamma)\theta_2. \quad (2.26b)$$

By substituting these values into Eq. (2.25) the equation of the isophotes becomes

$$(1 - \kappa - \gamma)^2\theta_1^2 + \beta_2 = (1 - \kappa + \gamma)^2\theta_2^2 = r^2, \quad (2.27)$$

which represents the equation for an ellipse, with major and minor axes  $a = r/\lambda_t$  and  $b = r/\lambda_r$ , respectively.

Moreover, the fact that the shear tensor has spin 2 translates in different distortions according to its components:

- $\gamma_1 > 0, \gamma_2 = 0$ : major axis of the ellipse is along  $\theta_1$ ;
- $\gamma_1 = 0, \gamma_2 > 0$ : major axis of the ellipse forms an angle of  $\pi/4$  with  $\theta_1$ ;
- $\gamma_1 < 0, \gamma_2 = 0$ : major axis of the ellipse is perpendicular to  $\theta_1$ ;
- $\gamma_1 = 0, \gamma_2 < 0$ : major axis of the ellipse forms an angle of  $3\pi/4$  with  $\theta_1$ .

Figure 2.5 shows how images orientations are affected by these combinations of  $\gamma_1$  and  $\gamma_2$  and many more.

### 2.2.3 Magnification

One of the most important features introduced by gravitational lensing is the *magnification* effect: through the lens equation, the solid angle  $\delta\beta^2$  is mapped to the solid angle  $\delta\theta^2$ . In the absence of photon emission or absorption, the Liouville theorem states that the surface brightness of a source remains constant as it passes through the lensing field. Thus, the change in the solid angle under which the source is observed implies that the flux received is magnified (or demagnified). From Eq. (2.20), the magnification introduced by the lensing is given by the inverse of the determinant of the Jacobian matrix. For this reason, the matrix  $M = A^{-1}$  is called the *magnification tensor* and therefore:

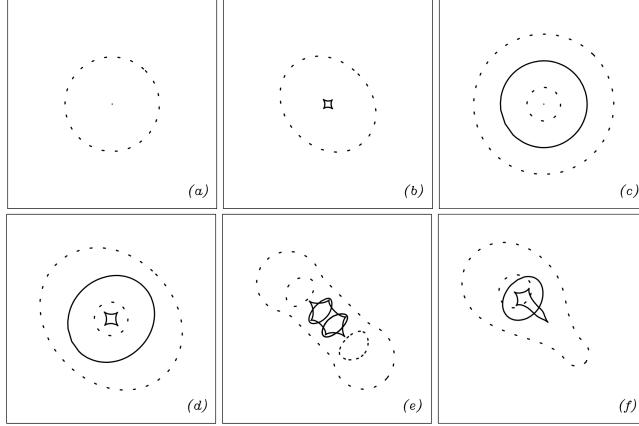
$$\mu \equiv \det M = \frac{1}{\det A} = \frac{1}{(1 - \kappa)^2 - \gamma^2}, \quad (2.28)$$

while the relation between intrinsic and observed flux can be written

$$F_\nu = \int_I I_\nu(\vec{\theta}) d^2\theta = \int_S I_\nu^S(\vec{\beta}(\vec{\theta})) \mu(\vec{\theta}) d^2\beta, \quad (2.29)$$

where the first integral is over the image plane and the second over the source plane. From Eq. (2.28), it is possible to define  $\mu_t = \lambda_t^{-1}$  and  $\mu_r = \lambda_r^{-1}$ , the so-called *tangential* and *radial* magnification factors, so that  $\mu = \mu_t \mu_r$ .

Since both convergence and shear (and therefore magnification) are functions of the position on the lens plane  $\vec{\theta}$ , there exists a set of critical points where  $\det A = 0$ , which means either  $\lambda_t = 0$  or  $\lambda_r = 0$ . The ensemble of these points defines the *critical lines*, along which the magnification diverges. By mapping the critical lines to the source plane through the lens equation, new sets of points are obtained, which are called *caustics*. The shape of these critical lines and caustics varies with the mass distribution of the lens. Some examples for different mass distributions are shown in Fig. 2.6.



**Figure 2.6.** Caustics (solid) and critical lines (dashed) for different mass models: (a) a singular isothermal circular mass distribution generates only the critical lines, in particular the radial critical line is the central point and the tangential critical line is the circle; (b) a singular isothermal elliptical lens produces a tangential caustic (astroid) and the corresponding elliptical tangential critical line; (c) a circular and (d) elliptical mass distribution with shallower inner slope than the isothermal mass distribution generates both the critical lines and caustics; a bimodal mass distribution with two clumps of equal (e) or unequal (d) mass produces more complex critical lines and caustics.

Credits: [Kneib & Natarajan \(2011\)](#).

#### 2.2.4 Time-delay and multiple images

The deflection of light by a gravitational potential causes a delay in the travel time of light between the source and the observer, resulting in the appearance of multiple images of the source on the lens plane at different times. This time delay has two components: one is geometrical, due to the different path length of the deflected light rays compared to the unperturbed ones, and one is gravitational, due to the effective speed of light in the presence of a gravitational potential.

The total time delay as a function of the position of the image, for a given source and lens, can be described by a surface known as *time delay surface*:

$$t(\vec{\theta}) = t_{geom} + t_{grav} = \frac{(1 + z_L)}{c} \frac{D_L D_S}{D_{LS}} \left[ \frac{1}{2} (\vec{\theta} - \vec{\beta})^2 - \hat{\Psi}(\vec{\theta}) \right] = \frac{D_{\Delta t}}{c} \tau(\vec{\theta}), \quad (2.30)$$

where  $z_L$  is the lens redshift and the quantities

$$D_{\Delta t} = (1 + z_L) \frac{D_L D_S}{D_{LS}}, \quad (2.31a)$$

$$\tau(\vec{\theta}) = \frac{1}{2} (\vec{\theta} - \vec{\beta})^2 - \hat{\Psi}(\vec{\theta}) \quad (2.31b)$$

are called *time delay distance* and *Fermat potential*, respectively.

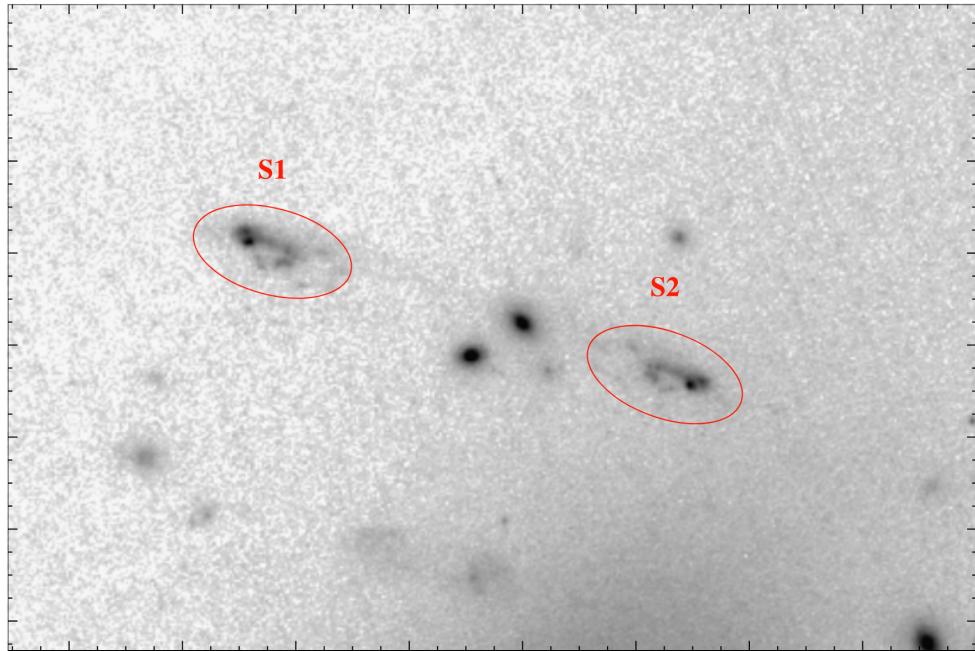
The lens equation can be obtained taking the gradient of Eq. (2.30):

$$\vec{\nabla} \left[ \frac{1}{2}(\vec{\theta} - \vec{\beta})^2 - \hat{\Psi}(\vec{\theta}) \right] = 0. \quad (2.32)$$

Therefore, the images are formed at the stationary points of the time delay surface, whose curvature is given by its Hessian matrix:

$$T_{ij} = \frac{\partial^2 t(\vec{\theta})}{\partial \theta_i \partial \theta_j} \propto (\delta_{ij} - \hat{\Psi}_{ij}) = A_{ij}. \quad (2.33)$$

The multiplicity of images in gravitational lensing is closely related to the characteristics of the time delay surface, and the configuration of this surface near stationary points offers insights into the shapes of the images and their *parity*. Image parity is determined by the magnification sign: magnifications  $> 1$ , whether positive or negative, both amplify the image. However, a positive magnification retains the original orientation of the unlensed source in the image, whereas a negative magnification results in an image with reversed parity. Crossing a critical line results in a sign change in one of the



**Figure 2.7.** The lensed pair S1–S2 in AC114. This galaxy at  $z = 1.867$  displays the surprising morphology of a hook, with an obvious change in parity.  
Credits: [Kneib & Natarajan \(2011\)](#).

Jacobian matrix's eigenvalues, leading to a shift in the image's parity. As a consequence, the parity of images located on opposite sides of a critical line is opposite to each other (Fig. 2.7). Furthermore, it can be shown that the number of images changes by two when the source crosses a caustic (Schneider et al., 1992) and that the total number of images of a generic lens is odd (Burke, 1981).

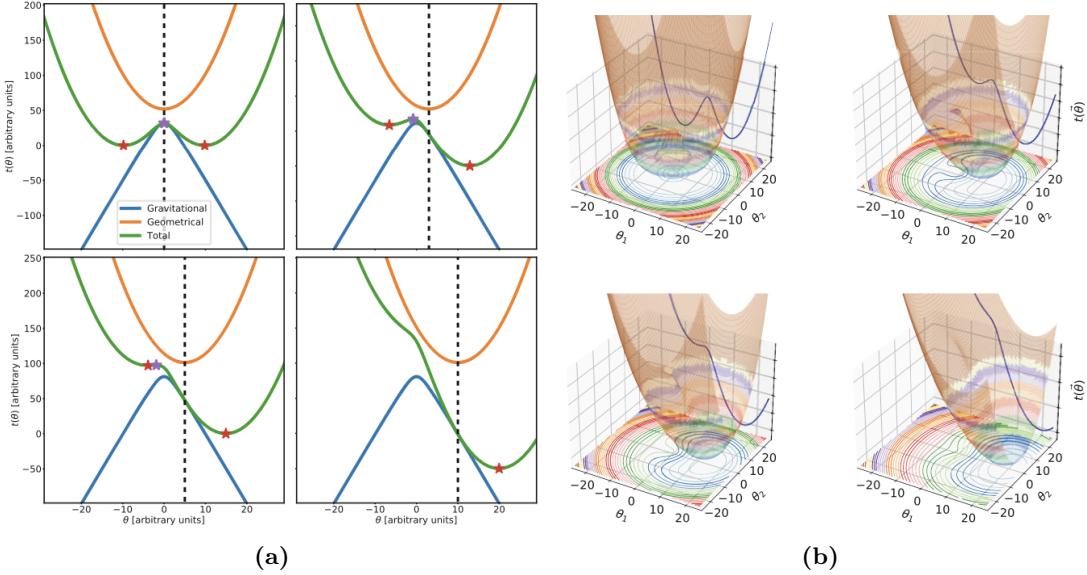
Different stationary points result in different types of image:

- **type I images** arise at the minima of surface where both Eq. (2.33) eigenvalues are positive, so  $\det A > 0$ ,  $\text{Tr } A > 0 \rightarrow$  positive magnification;
- **type II images** appears at the saddle points of the surface where Eq. (2.33) eigenvalues have opposite signs, so  $\det A < 0 \rightarrow$  negative magnification (i.e. reversed image parity, not de-magnified);
- **type III images** arise at the maxima of the surface where both Eq. (2.33) eigenvalues are negative, so  $\det A > 0$ ,  $\text{Tr } A < 0 \rightarrow$  positive magnification.

Figure 2.8a shows the geometrical and gravitational components of the time delay and their combination for a few positions of the source relative to the lens, considered axially symmetric. The geometric time delay is described by a parabola, and the gravitational function will vary depending on the potential considered. In particular, for a source perfectly aligned with the center of the lens (top left of Fig. 2.8b) there exist three stationary points, two of which (the minima) merge onto a single ring. This configuration is the so-called *Einstein ring*.

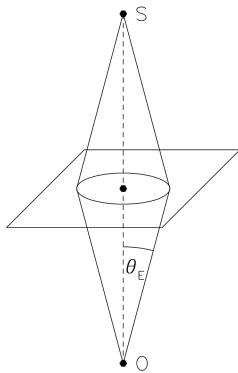
Thus, it is possible to introduce the so-called Einstein radius (Fig. 2.9), which is properly an angle, useful to define the radius of the Einstein ring:

$$\theta_E = \sqrt{\frac{4GM(\theta)}{c^2} \frac{D_{LS}}{D_L D_S}}. \quad (2.34)$$



**Figure 2.8.** (a) One-dimensional time delay functions for a non-singular isothermal potential. The stars indicate the position of the images. Each panel corresponds to a different position of the source relative to the lens (dashed line). (b) Time delay surfaces for the same lens. Each panel corresponds to a different position of the source relative to the lens. Clearly visible Einstein ring in the top-left plot.

Credits: [Meneghetti \(2021\)](#).



**Figure 2.9.** A source S exactly behind the center of an axially symmetric lens is mapped into the Einstein ring, with angular radius given by  $\theta_E$ .

Credits: [Narayan & Bartelmann \(1997\)](#).

# 3

## Lens models

This chapter delves into the intricate process of lens modeling, a pivotal technique for interpreting gravitational lensing observations. Gravitational lens models can be separated into two main categories: point-mass lenses (i.e. *microlenses*) and extended lenses, each possessing distinctive characteristics. Microlensing refers to the lensing effect caused by objects with relatively small masses, such as stars or planets, acting as lenses. Unlike their more massive counterparts, microlenses do not produce multiple discernible images of the background source. Instead, they induce magnification variations over time as the lens moves relative to the observer and the source. Transitioning to a grander scale, extended lenses involve massive structures like galaxies and galaxy clusters, capable of producing multiple, resolvable images of background sources.

### 3.1 Microlenses

This section is devoted to exploring the phenomenon of microlensing, which refers to the lensing effects caused by objects of relatively small mass in the universe, such as planets, stars, star clusters, and other compact objects located within the Milky Way or distant galaxies. Typically, these microlenses are considered, in a first-order approximation, to be point-masses or collections of point-masses.

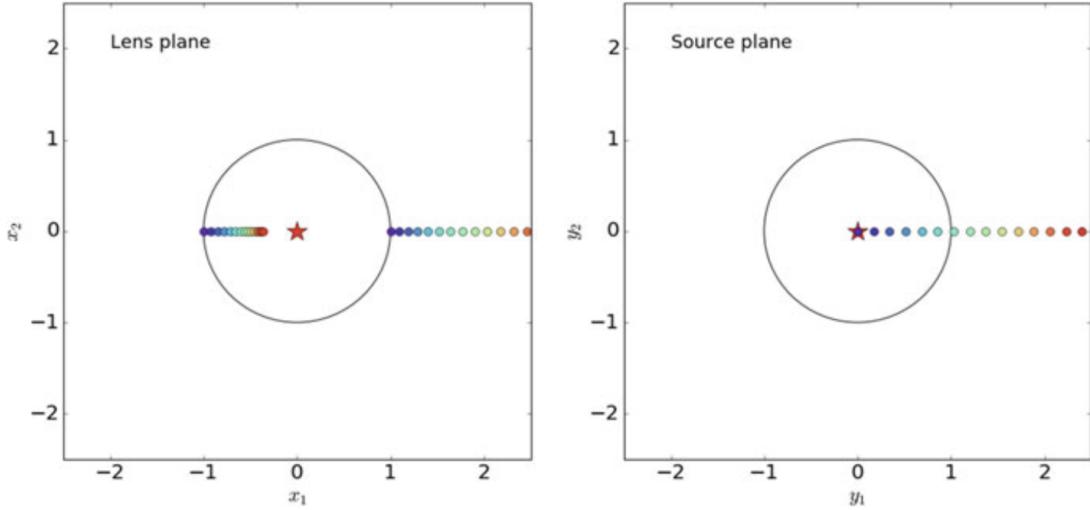
#### 3.1.1 Deflection angle and lensing potential

As already derived with Eq. (2.5), by setting the lens position as the center of the reference frame and using the relation  $\xi = D_L \theta$ , the deflection angle for a point mass lens can be written as

$$\vec{\alpha}(\vec{\theta}) = \frac{D_{LS}}{D_S} \hat{\vec{\alpha}}(\vec{\theta}) = \frac{D_{LS}}{D_S} \frac{4GM}{c^2 D_L} \frac{\vec{\theta}}{|\vec{\theta}|^2}. \quad (3.1)$$

Given that  $\vec{\alpha}(\vec{\theta}) = \vec{\nabla} \hat{\Psi}(\vec{\theta})$ , the lensing potential of the point mass lens is

$$\hat{\Psi}(\vec{\theta}) = \frac{4GM}{c^2} \frac{D_{LS}}{D_L D_S} \ln |\vec{\theta}|. \quad (3.2)$$



**Figure 3.1.** Solutions of the lens equation for a point-mass, with the lens represented by the star at the center. The Einstein ring is highlighted in black. In the right diagram, the locations of various sources are marked with colored circles. The images produced, as calculated using Eq. (3.5), are displayed in the left diagram.  
Credits: [Meneghetti \(2021\)](#).

### 3.1.2 Lens equation and multiple images

Given the deflection angle of Eq. (3.1), the lens equation becomes

$$\beta = \theta - \frac{4GM}{c^2\theta} \frac{D_{LS}}{D_L D_S}, \quad (3.3)$$

where the vector signs can be omitted due to the fact that the vector  $\hat{\vec{\alpha}}$  always points away from the lens.

As already anticipated in Section 2.2, the lens equation can be written in a more concise way by introducing a scale radius  $\theta_E$ , i.e. the *Einstein radius* defined in Eq. (2.34), and setting  $y = \beta/\theta_E$  and  $x = \theta/\theta_E$ , results:

$$\beta = \theta - \frac{\theta_E^2}{\theta} \quad \Rightarrow \quad y = x - \frac{1}{x}. \quad (3.4)$$

This equation is quadratic in  $\theta$  (or  $x$ ) and has two solutions:

$$x_{\pm} = \frac{y \pm \sqrt{y^2 + 4}}{2}, \quad (3.5)$$

which means that there always exist two images for a given source position.

In the right section of Fig. 3.1, some sources are arranged at varying angular distances

from the lens, which is marked by a red star. Each source is represented by a unique color to facilitate the identification of its corresponding images in the left section. Each source generates two images: one positioned at  $x_+ > 0$  and the other in the range of  $-1 < x_- < 0$ . These images appear on either side of the lens, with the  $x_-$  image always lying within a circle of radius  $x = 1$ . This circle is equivalent to the image produced by a source directly behind the point lens at  $y = 0$ , resulting in a ring-shaped image with radius  $\theta_E$ , the Einstein ring. The size of the Einstein radius is typically

$$\theta_E \approx 1'' \left( \frac{M}{10^{12} M_\odot} \right)^{1/2} \left( \frac{D}{\text{Gpc}} \right)^{-1/2}, \quad (3.6)$$

where

$$D \equiv \frac{D_L D_S}{D_{LS}} \quad (3.7)$$

is the *effective lensing distance*.

As the angular separation  $y \rightarrow 0$ , it is observed that  $x_- \rightarrow 0$ , whereas  $x_+ \rightarrow y$ . This indicates that when the angular distance between the lens and the source increases significantly, the source is unlensed. In theory, an image still exists at  $x_- = 0$ , but this central image has zero magnification.

### 3.1.3 Critical lines, caustics and magnification

The Jacobian determinant for a point-mass lens can be written as

$$\det A(x) = \frac{y}{x} \frac{dy}{dx}, \quad (3.8)$$

which means that the eigenvalues of the Jacobian matrix are

$$\lambda_t(x) = \frac{y}{x} = \left( 1 - \frac{1}{x^2} \right), \quad (3.9a)$$

$$\lambda_r(x) = \frac{dy}{dx} = \left( 1 + \frac{1}{x^2} \right). \quad (3.9b)$$

The second eigenvalue is never zero, and therefore the point-mass lens only has a tangential critical line, a circle with equation  $x^2 = 1$ , which represents the Einstein ring. This line can be mapped onto the source plane to find the relative tangential caustic, resulting in a single point at  $y = 0$ .

Given that the magnification is the inverse of the Jacobian determinant

$$\mu(x) = \left( 1 - \frac{1}{x^4} \right)^{-1}, \quad (3.10)$$

which can also be written as a function of the source position

$$\mu_{\pm}(y) = \frac{x \, dx}{y \, dy} = \frac{1}{2} \left( 1 \pm \frac{y^2 + 2}{y\sqrt{y^2 + 4}} \right), \quad (3.11)$$

the magnifications of the two images always have the same signs of  $x_{\pm}$  and therefore different parity.

The total source magnification will then be:

$$\mu(y) = \mu_+(y) + |\mu_-(y)| = \frac{y^2 + 2}{y\sqrt{y^2 + 4}}, \quad (3.12)$$

while the sum of the signed magnifications is always  $\mu = 1$ .

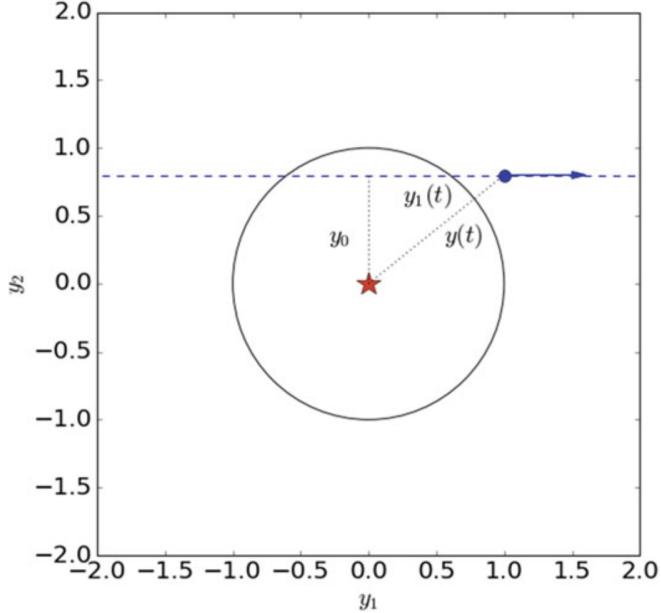
Furthermore, it can be shown through series expansion that, for large  $y$ ,

$$\lim_{y \rightarrow \infty} \left| \frac{\mu_+}{\mu_-} \right| \propto y^4. \quad (3.13)$$

This means that the magnification rapidly becomes negligible outside the Einstein ring and has a very simple form well inside it. Hence, deviations from simple point-lens microlensing can usually be easily spotted.

### 3.1.4 Microlensing light-curve

The Einstein radius of a typical lens indicates the scale of image separation observed in microlensing events. For a star with the mass of the Sun located within the Milky Way, the separation is approximately of the order of milliarcseconds, a measurement that is beyond the detection capabilities of current instruments. However, stars move around the Galactic center (and have an additional random velocity component with respect to one another). The relative velocities are such that the time scale of the relative change of lens and source positions is of order of weeks or shorter. Therefore, this motion introduces a temporal component in the lensing geometry, causing the distance between the lens and the source, and thus the magnification, to vary measurably as a function of time, accordingly to Eq. (3.12). In general, a source with intrinsic flux  $F_s$  will appear to have flux  $F(t) = \mu(t)F_s$ .



**Figure 3.2.** Illustration of the lens position and source trajectory. The dimensionless impact parameter is  $y_0$ ,  $y_1(t)$  indicates the dimensionless distance of the source from its closest point to the lens, and  $y(t)$  is the dimensionless distance between the lens and source.

Credits: [Meneghetti \(2021\)](#).

The *microlensing light-curve* describes the temporal behavior of the magnification, dictated by the relative motion of the lens across the observer-source line of sight.

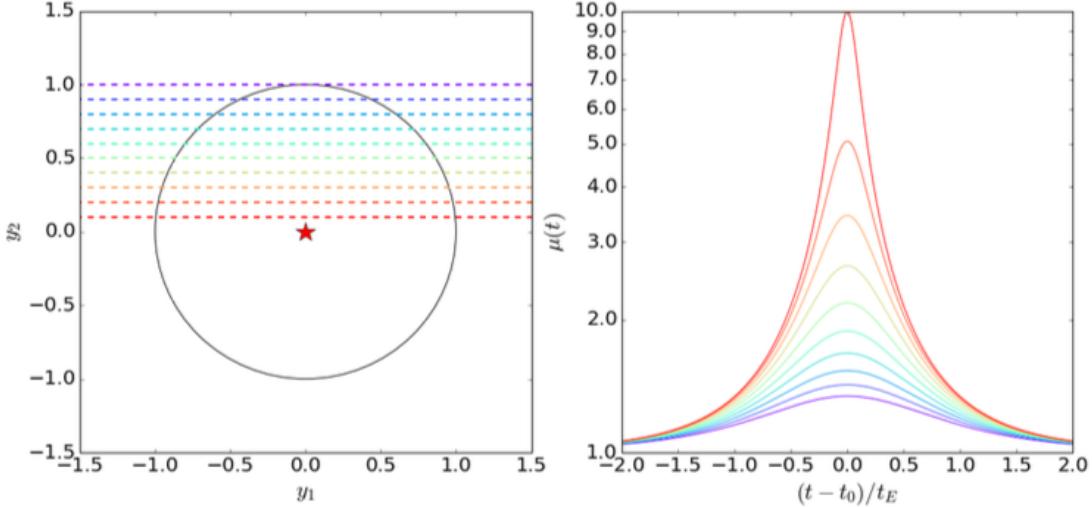
As shown in Fig. 3.2, assuming a straight line can approximate the path of the source relative to the lens, the former moving with transverse velocity  $v$  and reaching the minimum dimensionless distance  $y_0$  (i.e. the *impact parameter* of the source), from the lens at time  $t_0$ , the dimensionless distance of the source from  $y_0$  can be written

$$y_1(t) = \frac{v(t - t_0)}{D_L \theta_E}. \quad (3.14)$$

Since magnification significantly deviates from unity only for sources with  $|y| \lesssim 1$ , the characteristic timescale of the microlensing event is given by

$$t_E = \frac{D_L \theta_E}{v}, \quad (3.15)$$

which is known as the *Einstein crossing time*. Stellar lenses in the Milky Way are associated with typical  $t_E$  on the order of a month, and thus the changes in the observed brightness of the source they induce are referred to as microlensing “events”.



**Figure 3.3.** Left panel: some examples of source trajectory across the Einstein ring with different impact parameters  $y_0$ . Right panel: corresponding light-curves. Credits: [Meneghetti \(2021\)](#).

Inserting Eq. (3.15) into Eq. (3.14), the trajectory of the source can be represented by

$$y(t) = \sqrt{y_0^2 + y_1^2(t)} = \sqrt{y_0^2 + \frac{(t - t_0)^2}{t_E^2}}. \quad (3.16)$$

The corresponding light-curve of the source, examples of which are shown in Fig. 3.3, is then obtained by combining Eqs. (3.12) and (3.16) and multiplying by the unlensed source flux

$$F(t) = \mu(t)F_s = \frac{y(t)^2 + 2}{y(t)\sqrt{y(t)^2 + 4}}F_s. \quad (3.17)$$

The *standard* microlensing light-curve of a simple point-source, point-lens system is thus described by four parameters: unlensed flux  $F_s$ ,  $t_0$ ,  $y_0$  and  $t_E$ . Of these,  $F_s$  can be measured in the absence of microlensing,  $t_0$  and  $y_0$  from the position and height of the light-curve peak, respectively. Only  $t_E = D_L\theta_E/v$  contains physical information about the lens system and determines the peak width, i.e. the duration of the event. Assuming that the source distance can be determined from its properties (membership in a stellar system, spectral type and apparent magnitude), there are three physical parameters, lens mass  $M$ , lens distance  $D_L$ , and lens relative transverse velocity  $v$ , to determine from one observable: this is the so-called *microlensing degeneracy*. Although the standard light-curve model is successful in numerous instances, there exist circumstances where some foundational assumptions of this model no longer hold. In such situations, it might be feasible to derive additional constraints that can partially lift the degeneracy.

Non-standard light-curves, for instance, may occur when either the source or the lens are not point-like, or when the path of the source moving relative to the lens is not linear. Such instances often occur when either the lens or the source, or both, are part of binary systems.

## 3.2 Extended lenses

This section aims to introduce the main models used to represent extended gravitational lenses, such as massive galaxies and galaxy clusters. These cosmic structures, characterized by their complex, gravitationally bound mass distributions, are formidable gravitational lenses capable of producing striking lensing phenomena, including multiple images and gravitational arcs.

The analysis starts with circular, axially symmetric models and describes how different mass profiles affect lensing properties. Subsequently, deviations from circular symmetry are introduced through properties such as ellipticity and substructures. Finally, the impact of the environment surrounding the lenses is taken into account. The thin screen approximation is used throughout this description of analytical lens models.

### 3.2.1 Axially symmetric profiles

The most simple description of an extended lens is an axially symmetric profile. For such lenses, the potential is constant on circles centered on the lens center. This symmetry allows most equations to be simplified to a one-dimensional form.

The deflection angle is radially directed and its amplitude depends only on the distance from the lens center. Its expression is the same as that of the point-mass lens, with the only difference that now  $M = M(\theta)$  is the mass enclosed in a circle of radius  $\theta$ . For this reason, introducing an arbitrary reference scale  $\xi_0$ , the *dimensionless mass* profile can be written as

$$m(x) = 2 \int_0^x x' \kappa(x') dx' = \frac{M(\xi_0 x)}{\pi \xi_0^2 \Sigma_{cr}}, \quad (3.18)$$

which leads to a deflection angle

$$\alpha(x) = \frac{2}{x} \int_0^x x' \kappa(x') dx' = \frac{m(x)}{x}. \quad (3.19)$$

Therefore, the lens equation is

$$y = x - \frac{m(x)}{x}. \quad (3.20)$$

Given that  $\Psi'(x) = \alpha(x)$ , from Eq. (2.17) the convergence profile can be obtained:

$$\kappa(x) = \frac{1}{2} \left[ \alpha'(x) + \frac{\alpha(x)}{x} \right] = \frac{m'(x)}{2x}. \quad (3.21)$$

Finally, introducing  $\bar{\kappa}(x) = m(x)/x^2$  as the mean convergence within a circle of radius  $x$ , the shear vector results

$$\gamma(x) = \frac{1}{2} \left| \alpha'(x) - \frac{\alpha(x)}{x} \right| = |\kappa(x) - \bar{\kappa}(x)|. \quad (3.22)$$

Using the precedent results and Eq. (2.28), it is possible to compute the determinant of the lensing Jacobian

$$\det A(x) = \frac{y}{x} \frac{dy}{dx} = [1 - \kappa(x) - \gamma(x)][1 - \kappa(x) + \gamma(x)], \quad (3.23)$$

and the correspondent magnification profile  $\mu(x) = [\det A(x)]^{-1}$ .

As described in Section 2.2, the set of points that satisfies  $\det A(x) = 0$  represents the critical lines of the lens. For an axially symmetric lens these are circles, whose radii can be found by solving the equations

$$\lambda_t(x) = 0 \Rightarrow \kappa(x) + \gamma(x) = 1, \quad (3.24a)$$

$$\lambda_r(x) = 0 \Rightarrow \kappa(x) - \gamma(x) = 1, \quad (3.24b)$$

which define the *tangential critical line* and the *radial critical line*, respectively.

Through the lens equation, it can be seen that all the points along the tangential critical line are mapped to the point  $y = 0$  on the source plane: this type of lenses have point-like tangential caustics. Instead, the radial critical points are mapped onto a circular radial caustic on the source plane.

### 3.2.2 Power-law profiles

This class of lenses is characterized by a mass profile with a power-law form of the kind

$$m(x) = x^{3-n}, \quad (3.25)$$

where  $n$  is a parameter that can assume any real value  $> 1$ .

From the previous mass profile definition, all the other lensing properties can be derived:

$$\begin{aligned}\kappa(x) &= \frac{m'(x)}{2x} = \frac{3-n}{2}x^{1-n}, \\ \alpha(x) &= \frac{m(x)}{x} = x^{2-n}, \\ \gamma(x) &= \left| \frac{m'}{2x} - \frac{m(x)}{x^2} \right| = \frac{n-1}{2}x^{1-n}, \\ y &= x - x^{2-n}.\end{aligned}\tag{3.26}$$

In particular, given the value of the parameter  $n$ , it is possible to distinguish different scenarios.

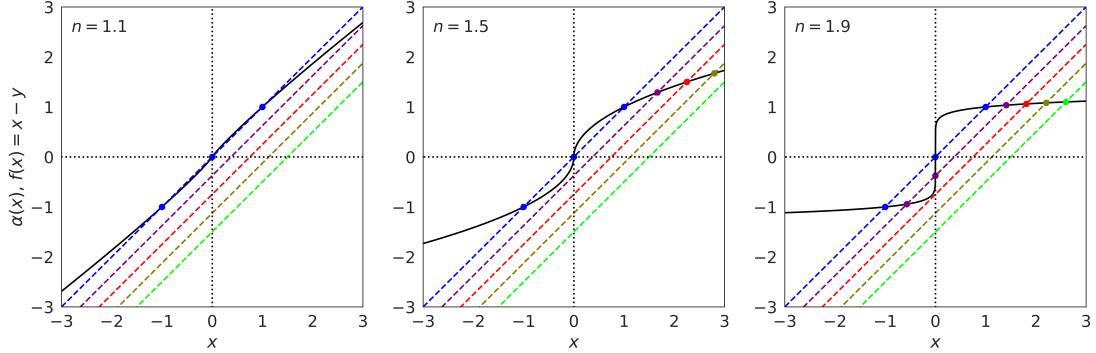
- **$n = 1$** : perfectly convergent lens with constant convergence and  $y(x) = 0, \forall x$ ;
- **$1 < n < 2$** : mass and deflection angle profiles increase with  $x$ , while the convergence and shear profiles show a decrease. The tangential critical line forms a circle of radius  $x_t = 1, \forall n$ , and its corresponding caustic line is a single point at  $y_t = 0$ . On the other hand, the size of the radial critical line varies according to the value of  $n$ . As it increases, the radial critical line becomes smaller, whereas the size of the caustic line follows an increasing trend. In particular, for  $n = 2$ , the radial critical line is absent. For lenses that are axially symmetric, there are multiple methods to solve the lens equation, but the most effective approach is often referred to as the *image diagram*, by which multiple image positions can be identified where the profile of the deflection angle intersects the lines  $f(x) = x - y$ .

Figure 3.4 demonstrates that the separation between images is influenced by the value of  $n$ : a higher value results in a more curved profile of the deflection angle, leading to a greater number of images that are closer together, and conversely, a lower  $n$  leads to fewer images with larger separations. More precisely, these lenses produce three images if the source is inside the radial critical line (i.e.  $y < y_r$ ), otherwise they form a single image.

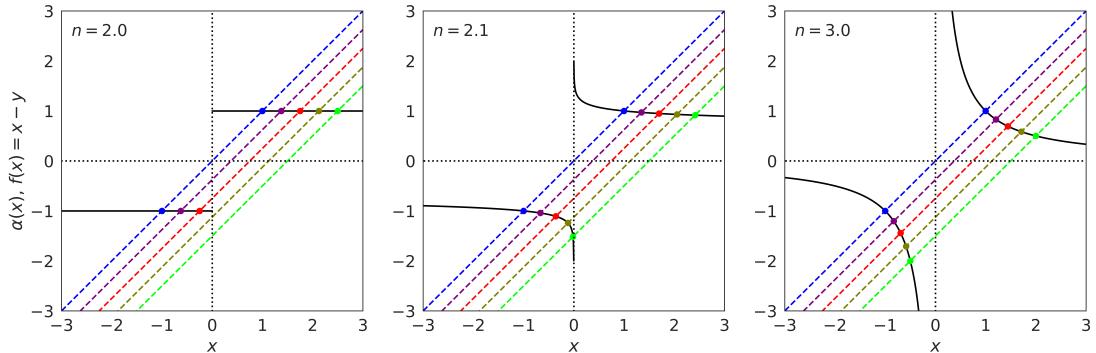
- **$n = 2$** : flat deflection angle profile, Singular Isothermal Sphere model (see Section 3.2.3);
- **$n > 2$** : deflection angle profile decreases with  $x$  and has a singularity in  $x = 0$ . These lenses produce two images, one inside and one outside the Einstein radius (see Fig. 3.5). The radial eigenvalue is always non-zero, causing the absence of

a radial critical line and ensuring that the radial magnification factor,  $\mu_r < 1$ , always. This means that the images are consistently demagnified in the radial direction.

It should be noted that for  $n = 3$  the scenario corresponds to that of a point-mass lens.



**Figure 3.4.** Image diagrams for different values of  $1 < n < 2$ . The solid curves show the function  $\alpha(x)$ , while the dashed lines represent the function  $f(x) = x - y$ , with varying  $y$  value.



**Figure 3.5.** Image diagrams for different values of  $n > 2$ . The solid curves show the function  $\alpha(x)$ , while the dashed lines represent the function  $f(x) = x - y$ , with varying  $y$  value.

### 3.2.3 Singular Isothermal Sphere

One of the simplest and most widely used models for axially symmetric lenses is the Singular Isothermal Sphere (SIS). The density profile is obtained by assuming that the matter constituting the lens acts like an ideal gas, contained within a spherically

symmetric gravitational potential. This gas is presumed to be in states of thermal and hydrostatic equilibrium.

By projecting the three-dimensional density onto the lens plane, the surface mass density can be calculated

$$\Sigma(\xi) = \int_0^\infty \rho(\xi, z) dz = \frac{\sigma_v^2}{2G\xi}, \quad (3.27)$$

where  $\xi$  is the distance from the center of the lens and  $\sigma_v$  represents the gas particle velocity dispersion (i.e. the stars in a galaxy or the galaxies in a cluster). Although this profile is not physical, having a singularity for  $\xi = 0$  and an infinite total mass, it reproduces very well the observed flat rotation curves of spiral galaxies for  $0 \ll \xi < \infty$ .

Setting the length scale at  $\xi_0 = 4\pi (\frac{\sigma_v}{c})^2 \frac{D_L D_{LS}}{D_S}$ , it is possible to rewrite Eq. (3.27) in a dimensionless form and derive the convergence profile

$$\Sigma(x) = \frac{\Sigma_{cr}}{2x} \Rightarrow \kappa(x) = \gamma(x) = \frac{1}{2x}, \quad (3.28)$$

which shows that the SIS lens corresponds to a power-law lens with  $n = 2$ .

From the convergence definition, the deflection angle and the lens equation can be obtained:

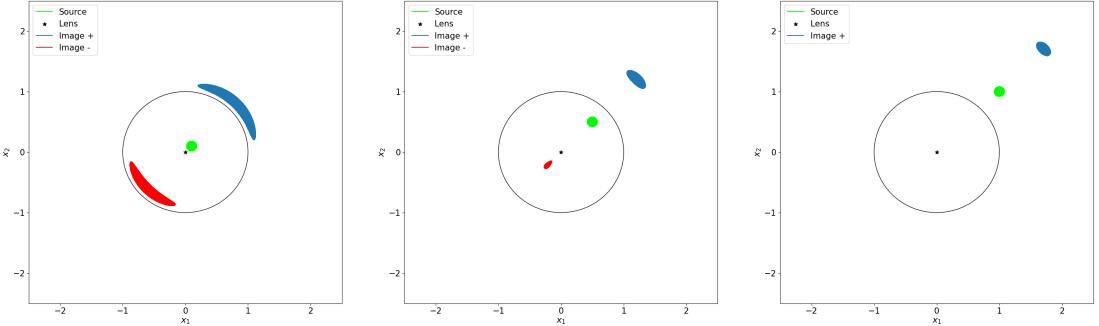
$$\alpha(x) = \frac{x}{|x|}, \quad (3.29a)$$

$$y = x - \frac{x}{|x|}. \quad (3.29b)$$

The solutions of the lens equation are dependent on the source position  $y$ , as it can be seen in the left panel of Fig. 3.5:

- when  $0 < y < 1$  there exist two solutions on opposite sides of the lens center; one is at  $x_- = y - 1$  and one at  $x_+ = y + 1$ . Their angular separation is always  $\Delta(\theta_E) = 2\theta_E$ ;
- when  $y > 1$  there is only one solution at  $x_+ = y + 1$ .

Therefore, the circle of radius  $y = 1$  serves a similar function to that of the radial caustic for power-law lenses with  $1 < n < 2$ , delineating areas on the source plane that are associated with different image multiplicities. However, this circle does not qualify as caustic because  $\alpha'(x) = 0, \forall x$ , indicating that  $\lambda_r = 1$ . Instead, the circle of radius  $y_{cut} = 1$  is referred to as pseudo-caustic, known as the *cut*. In addition, unlike "normal" caustics, the number of images changes only by one when crossing the cut, instead of changing by two.



**Figure 3.6.** Imaging of an extended source by a SIS lens. Each panel shows the images produced for different values of  $y$ . The black circle represents the tangential critical line of the lens, the Einstein ring.

Given that  $\lambda_r = 1$ , the images are only magnified in the tangential direction, while the radial size of all the images remains unchanged. In particular, when the source is placed inside the cut, the magnifications can be expressed as:

$$\mu_-(y) = 1 - \frac{1}{y}, \quad (3.30a)$$

$$\mu_+(y) = 1 + \frac{1}{y}, \quad (3.30b)$$

for the image inside and outside the Einstein ring respectively.

From Eqs. (3.30a) and (3.30b) can be seen that for  $y \rightarrow 1$  the image  $x_-$  becomes weaker and weaker until it disappears for  $y = 1$ . On the other hand, when  $y \rightarrow \infty$  the source magnification tends to unity: sources at a large distance from the lens experience little to no magnification.

Figure 3.6 shows some examples of images produced by a SIS lens of an extended circular source, placed at different distances from the lens.

### 3.2.4 Non-singular Isothermal Sphere

One way to solve the central singularity issue of the SIS lens is to introduce an additional parameter  $\xi_c$ , describing a flat central core of the surface density profile (Kormann et al., 1994):

$$\Sigma(\xi) = \frac{\Sigma_0}{\sqrt{1 + (\xi/\xi_c)^2}} \quad \text{with} \quad \Sigma_0 = \frac{\sigma_v^2}{2G\xi_c}, \quad (3.31)$$

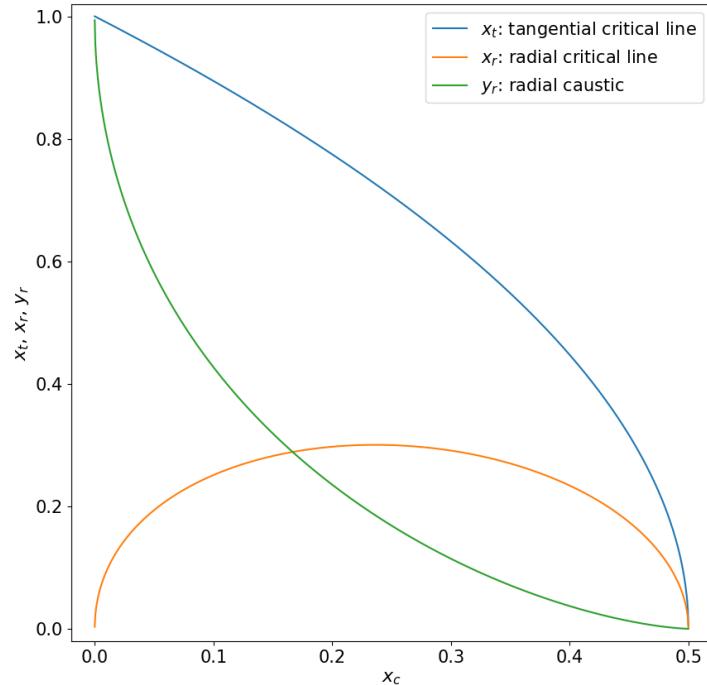
where  $\Sigma_0$  represents the constant surface density for  $\xi \ll \xi_c$ .

By introducing the same scale length  $\xi_0$  defined in Section 3.2.3 and re-scaling both  $\xi$  and  $\xi_c$ , the dimensionless relevant quantities for the Non-singular Isothermal Sphere

(NIS) can be derived:

$$\begin{aligned}\kappa(x) &= \frac{1}{2\sqrt{x^2 + x_c^2}}, \\ \mu(x) &= \sqrt{x^2 + x_c^2} - x_c, \\ \alpha(x) &= \sqrt{1 + \left(\frac{x_c}{x}\right)^2} - \frac{x_c}{x}, \\ y &= x - \sqrt{1 + \left(\frac{x_c}{x}\right)^2} - \frac{x_c}{x}.\end{aligned}\tag{3.32}$$

The lens equation above can be reduced to a third-order polynomial: the NIS lens can produce up to three images of a given source, with multiplicity depending on the value of  $x_c$ . It can be shown (Meneghetti, 2021) that tangential and critical lines exist only for values of  $x_c < 1/2$ , i.e. for  $x_c > 1/2$  the lens is “weak” and cannot produce multiple images. In particular, the tangential caustic is a point at  $y_t = 0$ , while the radii of the tangential critical line, the radial caustic and the radial critical line vary with  $x_c$  as shown in Fig. 3.7.



**Figure 3.7.** NIS caustic critical lines radii as a function of  $x_c$ . For  $x_c = 0$ ,  $y_r = 1$ , i.e. the SIS cut.

### 3.2.5 Singular Isothermal Ellipsoid

After investigating the impact of altering the slope of the density profile and the incorporation of a central core, it is possible to assess the influence of ellipticity on lens characteristics. Introducing ellipticity surely improves the representation of galaxies and their mass distributions. Kormann et al. (1994) developed the Singular Isothermal Ellipsoid (SIE) model, which can be derived from the SIS model by substituting

$$\xi \Rightarrow \sqrt{\xi_1^2 + f^2 \xi_2^2}, \quad (3.33)$$

where  $0 < f \leq 1$  is the axis ratio of the ellipses that, in this scenario, define the iso-density contours of the mass profile.

Through this substitution, the surface density profile becomes

$$\Sigma(\vec{\xi}) = \frac{\sigma_v^2}{2G} \frac{\sqrt{f}}{\sqrt{\xi_1^2 + f^2 \xi_2^2}}, \quad (3.34)$$

which, as already said, is constant on ellipses with minor axis  $\xi$  and major axis  $\xi/f$ .

By choosing, once again, the same scale length  $\xi_0$  as for the SIS model, the convergence can be written as

$$\kappa(\vec{x}) = \frac{\sqrt{f}}{2\sqrt{x_1^2 + f^2 x_2^2}}, \quad (3.35)$$

or, using polar coordinates,

$$\kappa(x, \varphi) = \frac{\sqrt{f}}{2x\Delta(\varphi)}, \quad (3.36)$$

where

$$\Delta(\varphi) = \sqrt{\cos(\varphi)^2 + f^2 \sin(\varphi)^2}. \quad (3.37)$$

By solving the Poisson equation, the lensing potential in polar coordinates can be obtained, and then, taking its gradient, it is possible to derive the deflection angle, which has two components, none of which depends on  $x$ :

$$\begin{aligned} \alpha_1(\vec{x}) &= \sqrt{\frac{f}{1-f^2}} \operatorname{arcsinh} \left( \frac{\sqrt{1-f^2}}{f} \cos(\varphi) \right), \\ \alpha_2(\vec{x}) &= \sqrt{\frac{f}{1-f^2}} \operatorname{arcsin} \left( \sqrt{1-f^2} \sin(\varphi) \right). \end{aligned} \quad (3.38)$$

The shear components can be obtained from the derivatives of the deflection angle

$$\begin{aligned}\gamma_1(\vec{x}) &= -\kappa(\vec{x}) \cos(2\varphi), \\ \gamma_2(\vec{x}) &= -\kappa(\vec{x}) \sin(2\varphi),\end{aligned}\tag{3.39}$$

and, as for the SIS model,  $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2} = \kappa$ .

Finally, from the lensing Jacobian determinant, the two eigenvalue are

$$\lambda_t(\vec{x}) = 1 - 2\kappa(\vec{x}),\tag{3.40a}$$

$$\lambda_r(\vec{x}) = 1.\tag{3.40b}$$

Similarly to the SIS lens, the SIE does not have a radial critical line. Instead, the tangential critical line is the ellipse defined by

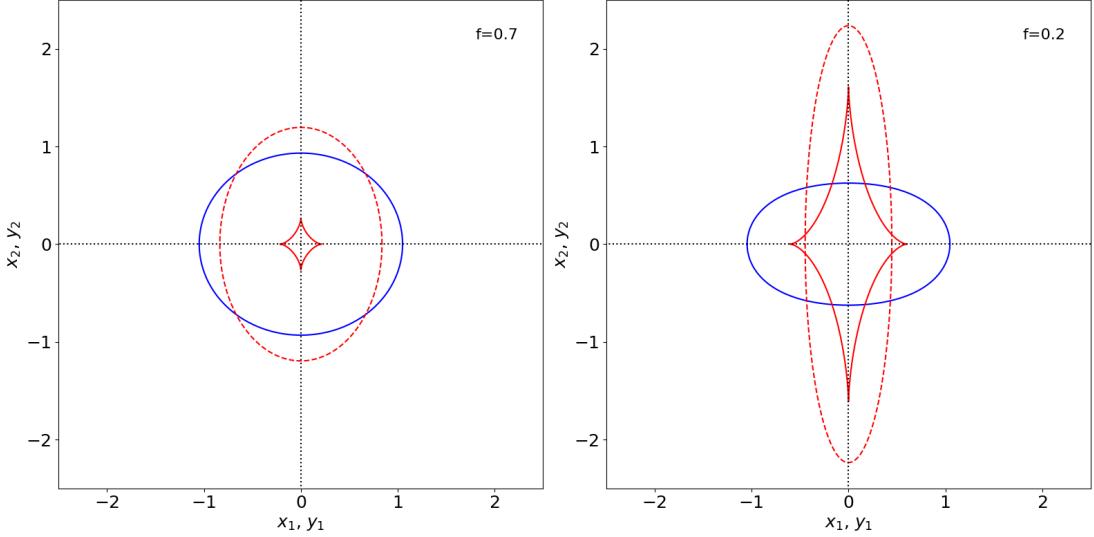
$$\kappa(\vec{x}) = \frac{1}{2} \Rightarrow \vec{x}_t(\varphi) = \frac{\sqrt{f}}{\Delta(\varphi)} [\cos(\varphi), \sin(\varphi)].\tag{3.41}$$

As always, the points on the critical line can be mapped onto the source plane using the lens equation: introducing ellipticity to the lens disrupts its axial symmetry, leading to the transformation of the tangential caustic from a central point into an astroid-shaped caustic, featuring four cusps and four folds. As for the SIS, due to the singularity at the center of the lens, also for the SIE profile does not exist a radial critical line and, as a result, the multiple images region is not enclosed by the radial caustic, but by the cut, which is an ellipse.

The number of multiple images formed by this lens is influenced by the positioning of the cut and caustic lines. In particular, for large ellipticities (small values of  $f$ , in particular  $f < f_0 = 0.3942$ ), the tangential caustic extends outside the cut, as shown in Fig. 3.8. The cusps that are not contained within the cut are called *naked*.

Given that crossing the cut alters the number of images by one and traversing the caustic alters it by two, the scenarios with  $f > f_0$  are capable of generating one, two, or four images. In contrast, a smaller value of  $f$  can lead to the existence of another possible scenario with the formation of three multiple images (see Fig. 3.9).

Figure 3.10 illustrate the lensing of circular sources by a SIE lens. The upper panels illustrate the change in image geometry as the source approaches the lens center, traversing the cut and the caustic fold. Sources positioned outside the cut generate a single image, which is located in the same quadrant of the image plane as the source's projected position. Upon crossing the cut, an additional image emerges near the lens center, a consequence directly related to the cut definition. As the source moves nearer

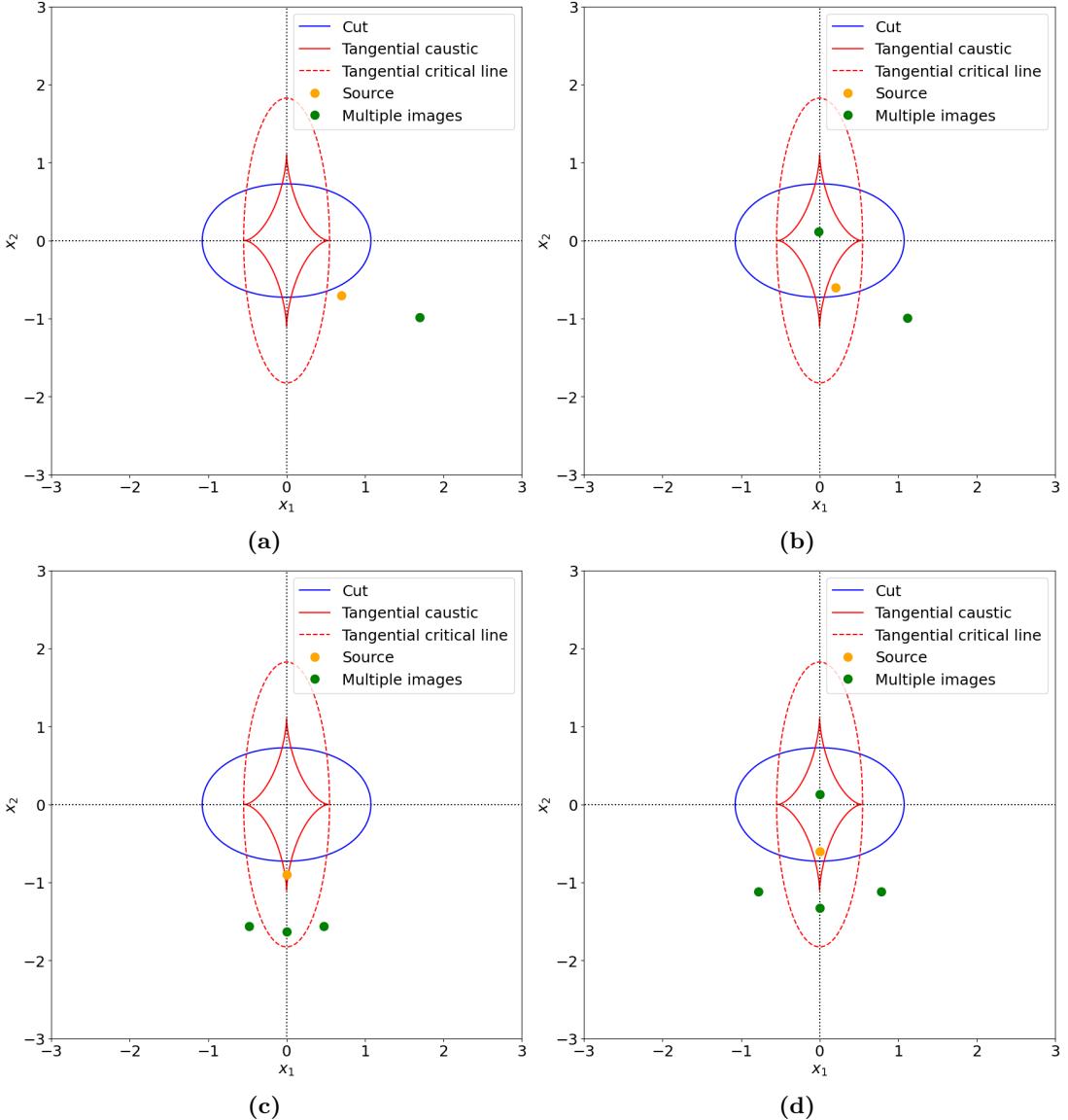


**Figure 3.8.** Tangential critical line (dashed red), tangential caustic (solid red) and cut (solid blue) for two SIE lenses with  $f = 0.7$  and  $f = 0.2$ .

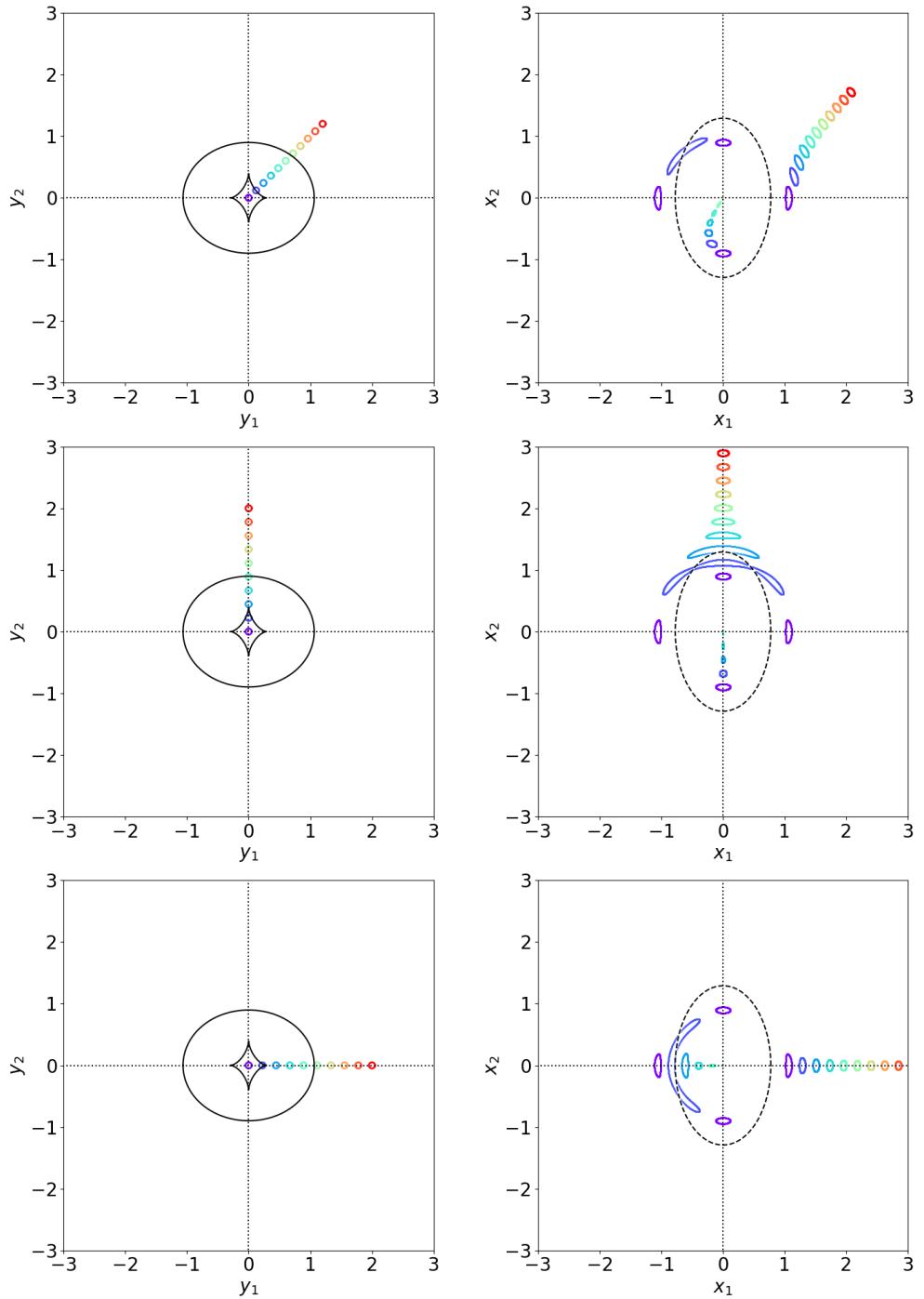
to the caustic, the central image shifts away from the lens center, appearing in the quadrant of the lens plane that is opposite to the source's projected quadrant. When the source crosses the caustic, a pair of images appear on either side of the critical line. Specifically, if the source is close to the fold in the source plane's first quadrant, the resultant images are found in the lens plane's second quadrant, adhering to the left-right mapping rule relevant to tangential critical points. As the source approaches the lens center, the resulting images form a symmetrical arrangement known as the *Einstein cross*.

In the middle and bottom panels of Fig. 3.9, which focus on the same lens, the source moves from beyond the cut towards the lens's center, crossing through the tangential caustic cusps. At the point of caustic crossing, three images converge at the critical line, with the left-right rule once again coming into effect.

Images near the critical line are distorted tangentially, leading to the creation of *gravitational arcs*. These images become elongated and tend to converge at the critical line, with the most significant gravitational arcs forming from the convergence of three images of sources located near the caustic cusps. The extent of observed distortions is influenced by the source's size in relation to the caustic. When the source is significantly larger than the caustic, the impact of the ellipticity is barely noticeable.



**Figure 3.9.** Illustration of the different image multiplicity scenarios: (a) one image if the source is outside the cut, (b) two images if the source is inside the cut, (c) three images if the source is inside the caustic and (d) four images if the source is inside both caustic and cut.



**Figure 3.10.** Lensing of circular sources by a SIE lens with  $f = 0.6$ . In the left panels, sources at different angular separations from the lens center. In the right panels, the correspondent multiple images.

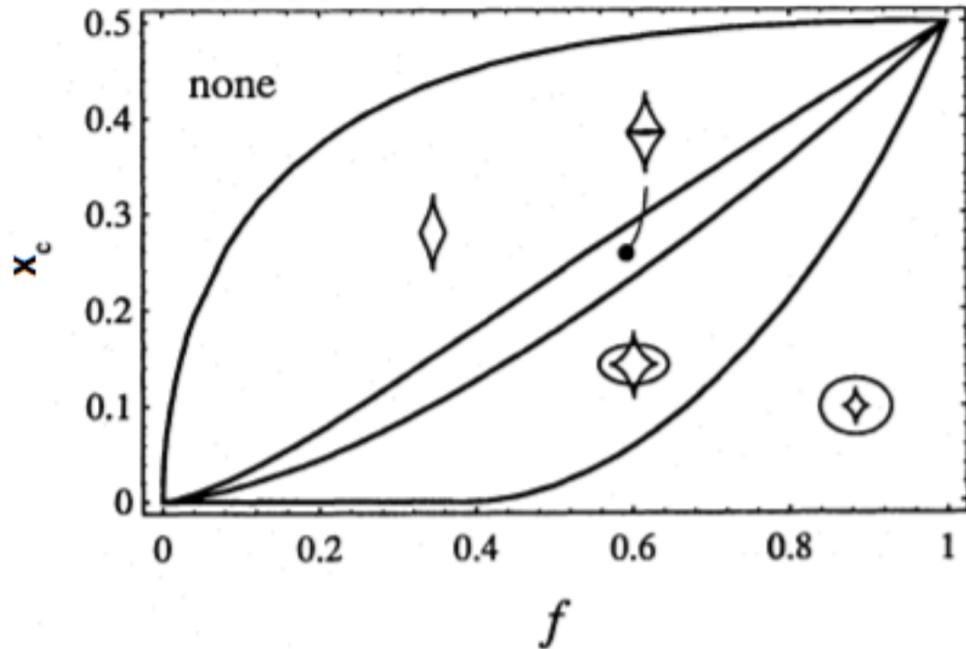
### 3.2.6 Non-singular Isothermal Ellipsoid

As previously done with the SIS model, to remove the central singularity of the surface mass density, it is also possible to introduce a core for the SIE model. The resulting model, the Non-singular Isothermal Ellipsoid (NIE), has been thoroughly described by [Kormann et al. \(1994\)](#); [Tessore & Benton Metcalf \(2015\)](#). By introducing a core radius  $\xi_c$  and with the usual choice of  $\xi_0 = \xi_{0,SIS}$ , the surface mass density and the convergence can be written

$$\Sigma(\vec{\xi}) = \frac{\sigma^2}{2G} \frac{\sqrt{f}}{\sqrt{\xi_1^2 + f^2\xi_2^2 + \xi_c^2}}, \quad (3.42a)$$

$$\kappa(\vec{x}) = \frac{\sqrt{f}}{2\sqrt{x_1^2 + f^2x_2^2 + x_c^2}}. \quad (3.42b)$$

Figure 3.11 shows the different multiplicities that a NIE lens can produce of a source, depending on the value of  $f$  and  $x_c$ . It can have two separate critical lines and caustics, only one tangential critical line and caustic, or no critical lines and caustics at all. Depending on the caustics structure, this lens can produce one, three, or five images of a source.



**Figure 3.11.** NIE caustics topologies for varying  $f$  and  $x_c$ .  
Credits: [Kormann et al. \(1994\)](#).

### 3.2.7 External shear

When analyzing a gravitational lens located in a densely populated environment, it becomes crucial to consider the gravitational influence of the surrounding mass distributions. A usual approach to encapsulate the effects of this environment is to employ the concept of an external shear field. This field is characterized through a potential that allows the quantification of the environmental impact on the lensing effects observed. The external shear field effectively models the tidal forces exerted by nearby mass distributions that are not directly part of the lens, but nevertheless influence the path of light rays passing near the lens.

The presence of the external shear field can be modeled by means of a potential  $\Psi_\gamma$ , such that

$$\begin{aligned}\gamma_1 &= \frac{1}{2}(\Psi_{11} - \Psi_{22}) = \text{const.} \\ \gamma_2 &= \Psi_{12} = \text{const.} \\ \kappa &= \frac{1}{2}(\Psi_{11} + \Psi_{22}) = \text{const.}\end{aligned}\tag{3.43}$$

This means that  $\Psi_{11}$  and  $\Psi_{22}$  must be both constants and the potential is quadratic:

$$\Psi_\gamma(\vec{x}) = Cx_1^2 + C'x_2^2 + Dx_1x_2 + E.\tag{3.44}$$

Differentiating Eq. (3.44) and substituting into Eq. (3.43):

$$\begin{aligned}\gamma_1 &= \frac{1}{2}(\Psi_{11} - \Psi_{22}) = C - C', \\ \gamma_2 &= \Psi_{12} = D, \\ \kappa &= \frac{1}{2}(\Psi_{11} + \Psi_{22}) = C + C'.\end{aligned}\tag{3.45}$$

At this point, it is possible to distinguish two cases:

- $\kappa = 0$  if the external perturbation does not contribute to the convergence, which means  $C = C' = \gamma_1/2$  and

$$\Psi_\gamma = \frac{\gamma_1}{2}(x_1^2 - x_2^2) + \gamma_2 x_1 x_2.\tag{3.46}$$

- $\gamma_1, \gamma_2 = 0$  if the lens is embedded in a sheet of constant surface mass density producing no shear (no privileged directions); in this case the external convergence potential is

$$\Psi_\kappa = \frac{\kappa}{2}x^2,\tag{3.47}$$

and can be inserted into the lens equation

$$\vec{\alpha}(\vec{x}) = \vec{\nabla}\Psi_\kappa(\vec{x}) = \kappa\vec{x}, \quad (3.48)$$

to obtain

$$\vec{y} = \vec{x} - \vec{\alpha}(\vec{x}) = \vec{x}(1 - \kappa). \quad (3.49)$$

If  $\kappa = 1$ , this sheet acts as a perfectly convergent lens, mapping every position  $\vec{x}$  on the lens plane to the same point  $y = 0$ .

# 4

## Algorithms

Building on the foundational understanding of gravitational lensing and its computational challenges, a critical aspect of modern astrophysical research involves the optimization of parametric functions to model the complex phenomena of the Universe accurately. This optimization is essential in gravitational lensing studies, where precise models of the mass distribution within lensing objects are paramount for interpreting the observed lensing effects. The optimization process often requires navigating a high-dimensional parameter space to find the best-fit parameters that reconcile theoretical models with observational data, a task that can be computationally intensive and algorithmically complex.

In this context, the advent of frameworks such as PyTorch<sup>1</sup> (Paszke et al., 2019) and TensorFlow<sup>2</sup> (Abadi et al., 2016) represents a significant leap forward for astrophysical research. These open-source libraries, primarily developed for deep learning applications, offer powerful tools for automatic differentiation, a technique that facilitates the calculation of gradients automatically, a cornerstone for any optimization algorithm. The ability of PyTorch and TensorFlow to efficiently compute derivatives of highly complex, nested functions makes them exceptionally well suited for optimizing parametric models in gravitational lensing studies.

Moreover, PyTorch and TensorFlow are designed to exploit the capabilities of modern computing hardware, including GPUs and TPUs, enabling the parallel processing of large datasets and the acceleration of computational tasks. This feature is particularly beneficial for gravitational lensing research, where handling large amounts of observational data and running complex simulations is commonplace. Using the computational power offered by these frameworks, one can significantly reduce the time required for model optimization and data analysis, thus enhancing the efficiency of the investigations.

### 4.1 Differentiable programming

Differentiable programming is an advanced computational paradigm that unites traditional programming concepts with the principles of differentiation, a fundamental

---

<sup>1</sup><https://pytorch.org>

<sup>2</sup><https://www.tensorflow.org>

concept in calculus. This approach extends the idea of computing derivatives to entire programs, enabling the automatic calculation of gradients of program outputs with respect to inputs. Differentiable programming is particularly powerful in the context of optimization, machine learning, and artificial intelligence, where it facilitates efficient parameter tuning to minimize or maximize some objective function. Methods for computing derivatives in computer programs can be classified into four categories (see Fig. 4.1):

1. manually working out derivatives and coding them;
2. **numerical differentiation**, involves approximating the derivative of a function using values of the original function evaluated at some sample points ([Burden et al., 2016](#)). In its simplest form, it is based on the limit definition of a derivative. It is quite simple to implement and apply to a wide range of problems, especially when dealing with data-driven models or functions that lack a clear analytical representation. Its effectiveness diminishes in high-dimensional settings, where the complexity and computational demands increase exponentially, highlighting the method's limitations in handling complex, multi-variable functions efficiently;
3. **symbolic differentiation**, which refers to the automatic process of finding derivatives using the rules of differentiation to obtain an exact symbolic expression for the derivative of a given function ([Grabmeier et al., 2003](#)). This method works similarly to how humans perform differentiation “by hand”, manipulating symbols according to mathematical laws;
4. **automatic differentiation (AD)**, also called algorithmic differentiation, is a method that computes the derivative of a function efficiently and accurately by systematically applying the chain rule of calculus to the sequence of elementary operations (additions, multiplications, trigonometric functions, etc.) that constitute a computer program. All numerical computations are ultimately compositions of a finite set of elementary operations for which derivatives are known ([Verma, 2000](#); [Griewank & Walther, 2008](#)), and combining the derivatives of the constituent operations through the chain rule of calculus<sup>3</sup> gives the derivative of the overall composition. AD is not an approximation like numerical differentiation, but rather computes derivatives to machine precision.

---

<sup>3</sup>The derivative of the composition of two (or more) differentiable functions  $f$  and  $g$  can be expressed in terms of the derivatives of the single functions  $f$  and  $g$ . If  $h(x) = f(g(x))$ , then  $h'(x) = f'(g(x))g'(x)$ .

Technique	Advantage(s)	Drawback(s)
<b>Hand-coded analytical derivative</b>	Exact and often fastest method.	Time consuming to code, error prone, and not applicable to problems with implicit solutions. Not automated.
<b>Finite differentiation</b>	Easy to code.	Subject to floating point precision errors and slow, especially in high dimensions, as the method requires at least $D$ evaluations, where $D$ is the number of partial derivatives required.
<b>Symbolic differentiation</b>	Exact, generates symbolic expressions.	Memory intensive and slow. Cannot handle statements such as unbounded loops.
<b>Automatic differentiation</b>	Exact, speed is comparable to hand-coding derivatives, highly applicable.	Needs to be carefully implemented, although this is already done in several packages.

**Figure 4.1.** Summary of techniques to calculate derivatives.

Credits: [Margossian \(2019\)](#).

In particular, differentiable programming refers to using automatic differentiation in some way that allows a program to optimize its parameters to improve on some task. It only has three requirements:

1. a parameterized function/model to be optimized;
2. (automatic) differentiability of the object to be optimized;
3. a function suitable to measure the performance of the model.

The process of optimizing a model in the field of differentiable programming is often called *training*. In the subsequent portion of this section, a comprehensive overview of automatic differentiation and its role in the training process will be provided. This will be followed by an in-depth exposition of the process itself.

#### 4.1.1 Automatic differentiation

##### Computational graph

As already stated, automatic differentiation operates on the fundamental principle that complex functions can be decomposed into a series of elementary arithmetic operations. Given a target composite function  $f(x) = h \circ g(x) = h(g(x))$ , with  $x \in \mathbb{R}^n$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , and  $h : \mathbb{R}^k \rightarrow \mathbb{R}^m$ , applying the chain rule and elementary matrix multiplication, the

corresponding Jacobian matrix<sup>4</sup>  $J$  is thus:

$$J = J_{h \circ g} = J_h(g(x)) \cdot J_g(x), \quad (4.1)$$

with  $(i, j)^{th}$  element:

$$J_{ij} = \frac{\partial f_i}{\partial x_j} = \frac{\partial h_i}{\partial g_1} \frac{\partial g_1}{\partial x_j} + \frac{\partial h_i}{\partial g_2} \frac{\partial g_2}{\partial x_j} + \dots + \frac{\partial h_i}{\partial g_k} \frac{\partial g_k}{\partial x_j}. \quad (4.2)$$

More generally, if  $f$  is the composite expression of  $L$  functions

$$f = f^L \circ f^{L-1} \circ \dots \circ f^1, \quad (4.3)$$

the corresponding Jacobian matrix will be

$$J = J_L \cdot J_{L-1} \cdot \dots \cdot J_1. \quad (4.4)$$

Hence, given a complex function  $f$ , it is possible to break down the action of the Jacobian matrix on a vector into simple components. So, following [Griewank & Walther \(2008\)](#) notation, a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  can be constructed using intermediate variables  $v_i$  such that

- variables  $v_{i-n} = x_i$ ,  $i = 1, \dots, n$  are the input variables,
- variables  $v_i$ ,  $i = 1, \dots, l$  are the intermediate variables,
- variables  $y_{m-i} = v_{l-i}$ ,  $i = m - 1, \dots, 0$  are the output variables.

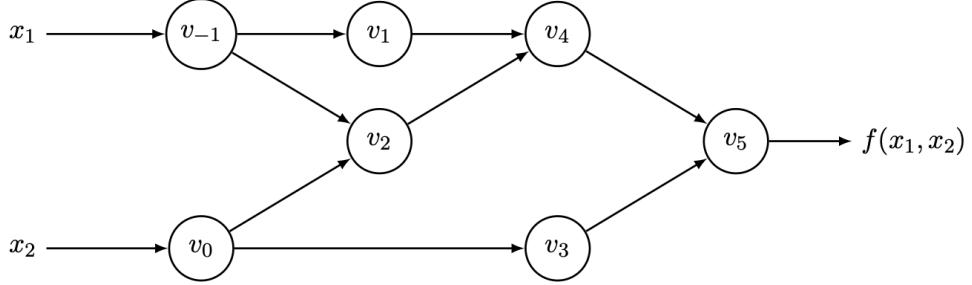
The representation of all the elementary operations that take place to construct a certain function  $f$  is called the *evaluation trace*, which can also be pictured as a *computational graph* ([Bauer, 1974](#)), useful for visualizing the dependency relations between intermediate variables. Figure 4.2 shows the computation graph for an example function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ :

$$f(x_1, x_2) = \ln(x_1) + x_1 x_2 - \sin(x_2). \quad (4.5)$$

Given the computational graph of all elementary operations, AD can be implemented in two main modes: *forward* accumulation mode and *reverse* accumulation mode (backward mode).

---

<sup>4</sup>The Jacobian matrix of a vector-valued function of several variables is the matrix of all its first-order partial derivatives.



**Figure 4.2.** Computational graph of the example  $f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$ . Credits: [Baydin et al. \(2018\)](#).

### Forward mode

Automatic differentiation in forward accumulation mode (or *tangent linear* mode) is the conceptually most simple type. The differentiation process is aligned with the function evaluation itself. Starting from the inputs, the program computes both the function's value and its derivative step by step through the computation graph. For each elementary operation, the forward mode calculates the derivative of the output with respect to the inputs, carrying these derivatives (or “tangents”) forward through the computation graph.

Considering as an example the evaluation trace of the function defined in Eq. (4.5) (left-hand side of Fig. 4.3), for computing the derivative of  $f$  with respect to  $x_1$ , the first step is to associate with each intermediate variable  $v_i$  a derivative

$$\dot{v}_i = \frac{\partial v_i}{\partial x_1}. \quad (4.6)$$

Applying then the chain rule to each elementary operation in the forward primal trace, the corresponding tangent (derivative) trace is generated (right-hand side of Fig. 4.3), until the required derivative in the final variable  $\dot{v}_5 = \frac{\partial y}{\partial x_1}$  is obtained.

Generalizing this procedure to a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , each forward pass of AD provides one column of the Jacobian matrix (i.e. the partial derivatives of all output variables with respect to one input variable). Thus, the complete Jacobian can be computed in  $n$  evaluations. For this reason, forward AD is efficient and straightforward especially for functions  $f : \mathbb{R} \rightarrow \mathbb{R}^m$ , for which all derivatives can be computed with just one forward pass.

In general, for functions with many inputs  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  where  $n \gg m$ , the reverse accumulation mode of AD is preferred.

Forward Primal Trace			Forward Tangent (Derivative) Trace		
$v_{-1} = x_1$	= 2		$\dot{v}_{-1} = \dot{x}_1$	= 1	
$v_0 = x_2$	= 5		$\dot{v}_0 = \dot{x}_2$	= 0	
$v_1 = \ln v_{-1}$	= $\ln 2$		$\dot{v}_1 = \dot{v}_{-1}/v_{-1}$	= $1/2$	
$v_2 = v_{-1} \times v_0$	= $2 \times 5$		$\dot{v}_2 = \dot{v}_{-1} \times v_0 + \dot{v}_0 \times v_{-1}$	= $1 \times 5 + 0 \times 2$	
$v_3 = \sin v_0$	= $\sin 5$		$\dot{v}_3 = \dot{v}_0 \times \cos v_0$	= $0 \times \cos 5$	
$v_4 = v_1 + v_2$	= $0.693 + 10$		$\dot{v}_4 = \dot{v}_1 + \dot{v}_2$	= $0.5 + 5$	
$v_5 = v_4 - v_3$	= $10.693 + 0.959$		$\dot{v}_5 = \dot{v}_4 - \dot{v}_3$	= $5.5 - 0$	
$y = v_5$	= 11.652		$\dot{y} = \dot{v}_5$	= 5.5	

**Figure 4.3.** Forward mode AD example for  $f(x_1, x_2) = \ln(x_1) + x_1 x_2 - \sin(x_2)$ , evaluated at  $(x_1, x_2) = (2, 5)$  by setting  $\dot{x}_1 = 1$ . The original forward evaluation of the primals on the left is augmented by the tangent operations on the right, where each line complements the original directly to its left.

Credits: [Baydin et al. \(2018\)](#).

## Reverse mode

AD in reverse accumulation mode (or *adjoint* or *cotangent linear* mode) ([Griewank et al., 2012](#)) corresponds to a generalized backpropagation algorithm, in that it propagates derivatives backward from a given output. This is done by complementing each variable  $v_i$  with an adjoint

$$\bar{v}_i = \frac{\partial y_j}{\partial v_i}, \quad (4.7)$$

which represents the sensitivity of a considered output  $y_j$  with respect to changes in  $v_i$ .

In reverse mode AD, derivatives are computed in the second of a two-phase process. In the first phase, the original function code is run forward, creating intermediate variables  $v_i$  and recording the dependencies in the computational graph. In the second phase, derivatives are calculated by propagating adjoints  $\bar{v}_i$  in reverse, from outputs to inputs. The reverse mode AD for the example function of Eq. (4.5) is shown in Fig. 4.4.

An important advantage of the reverse mode is that it is significantly less costly to evaluate (in terms of operation count) than the forward mode for functions with a large number of inputs. In the extreme case of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , only one application of the reverse mode is sufficient to compute the full gradient, compared with the  $n$  passes of the forward mode needed to populate the same. Because the optimization of astrophysical and gravitational lensing parametric functions usually involves the gradient of a scalar-valued objective with respect to a (large) number of parameters, this establishes the reverse mode, as opposed to the forward mode, as the mainstay technique in the form of the backpropagation algorithm ([Schmidhuber, 2015](#)).

Forward Primal Trace			Reverse Adjoint (Derivative) Trace		
$v_{-1} = x_1$	= 2		$\bar{x}_1 = \bar{v}_{-1}$	= 5.5	
$v_0 = x_2$	= 5		$\bar{x}_2 = \bar{v}_0$	= 1.716	
$v_1 = \ln v_{-1}$	= $\ln 2$		$\bar{v}_{-1} = \bar{v}_{-1} + \bar{v}_1 \frac{\partial v_1}{\partial v_{-1}} = \bar{v}_{-1} + \bar{v}_1/v_{-1} = 5.5$		
$v_2 = v_{-1} \times v_0$	= $2 \times 5$		$\bar{v}_0 = \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0} = \bar{v}_0 + \bar{v}_2 \times v_{-1} = 1.716$		
$v_3 = \sin v_0$	= $\sin 5$		$\bar{v}_{-1} = \bar{v}_2 \frac{\partial v_2}{\partial v_{-1}} = \bar{v}_2 \times v_0 = 5$		
$v_4 = v_1 + v_2$	= $0.693 + 10$		$\bar{v}_0 = \bar{v}_3 \frac{\partial v_3}{\partial v_0} = \bar{v}_3 \times \cos v_0 = -0.284$		
$v_5 = v_4 - v_3$	= $10.693 + 0.959$		$\bar{v}_2 = \bar{v}_4 \frac{\partial v_4}{\partial v_2} = \bar{v}_4 \times 1 = 1$		
$y = v_5$	= 11.652		$\bar{v}_1 = \bar{v}_4 \frac{\partial v_4}{\partial v_1} = \bar{v}_4 \times 1 = 1$		
			$\bar{v}_3 = \bar{v}_5 \frac{\partial v_5}{\partial v_3} = \bar{v}_5 \times (-1) = -1$		
			$\bar{v}_4 = \bar{v}_5 \frac{\partial v_5}{\partial v_4} = \bar{v}_5 \times 1 = 1$		
			$\bar{v}_5 = \bar{y}$	= 1	

**Figure 4.4.** Reverse mode AD example for  $f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$ , evaluated at  $(x_1, x_2) = (2, 5)$ . After forward evaluation of the primals on the left, adjoint operations on the right are evaluated in reverse.

Credits: [Baydin et al. \(2018\)](#).

#### 4.1.2 Backpropagation algorithm

The backpropagation algorithm is the cornerstone of neural network training ([Rumelhart et al., 1986](#); [LeCun et al., 2012](#)), mainly used to minimize the error by adjusting the weights of the connections in the network. However, when it comes to a model optimization problem outside the realm of neural networks, such as training parametric models, the concept of backpropagation refers to the method of computing gradients efficiently using reverse mode AD.

The training process (or better, the training loop) is characterized by a sequence of operations performed recursively. The optimization itself is performed by an *optimizer* which operates by tweaking the model parameters in a certain way, with the purpose of minimizing a *loss* function, which quantifies the difference between the predicted outputs of the model and the actual target/observed values. The main algorithm that allows for this optimization is the so-called *gradient descent*.

#### Loss function

As mentioned above, the loss function provides a measure of the performance of the model. The goal of optimization is to adjust the model parameters in a way that minimizes the loss function. The choice of loss function depends on the specific problem and the model. Common examples include the Mean Squared Error (MSE) for regression problems

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.8)$$

where  $y_i$  and  $\hat{y}_i$  are the observed and predicted values, respectively, and the Cross-Entropy Loss for classification problems.

A common loss function used in the optimization of the gravitational lensing models is the chi-squared ( $\chi^2$ ) statistic, which is the sum of squared differences between observed ( $y$ ) and model-predicted ( $\hat{y}$ ) values, normalized by uncertainties ( $\sigma$ ) in observations:

$$\chi^2 = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{\sigma_i} \right)^2. \quad (4.9)$$

The loss function is a crucial component because it guides the training process by indicating the direction in which the model parameters should be adjusted.

### Gradient descent

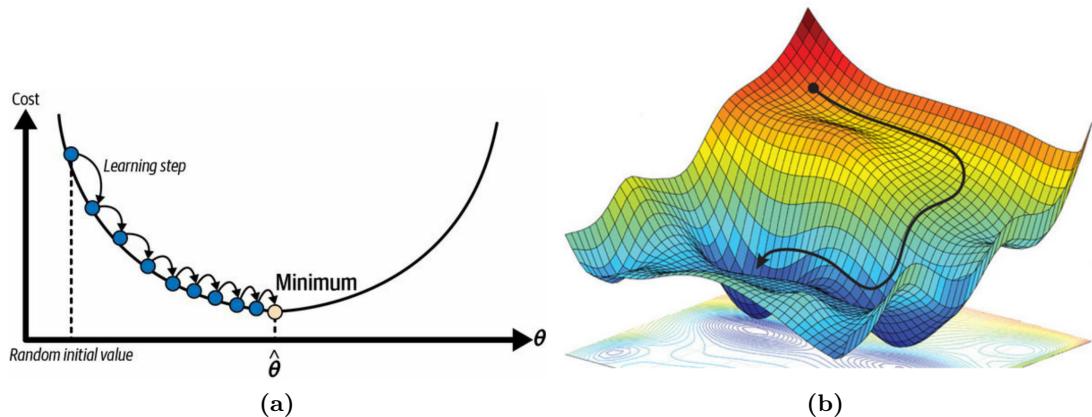
Gradient-based optimization is one of the pillars of machine learning (Bottou et al., 2018) and is one of the most common optimization algorithms for parametric functions. It is used to find the values of the parameters of a function that decrease the loss function as much as possible (Chandra et al., 2022; Ruder, 2016). Given a loss function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , and starting from random initial values of the parameters  $\vec{\theta}$ , classical gradient descent has the objective of finding (local) minima

$$\hat{\vec{\theta}} = \arg \min_{\vec{\theta}} f(\vec{\theta}), \quad (4.10)$$

which means finding the set of parameters  $\hat{\vec{\theta}} \in \mathbb{R}^n$  that minimizes the loss function, as shown in Fig. 4.5. In complex models,  $f$  might have multiple local minima, and the algorithm seeks to find at least one of these, via updates of the form

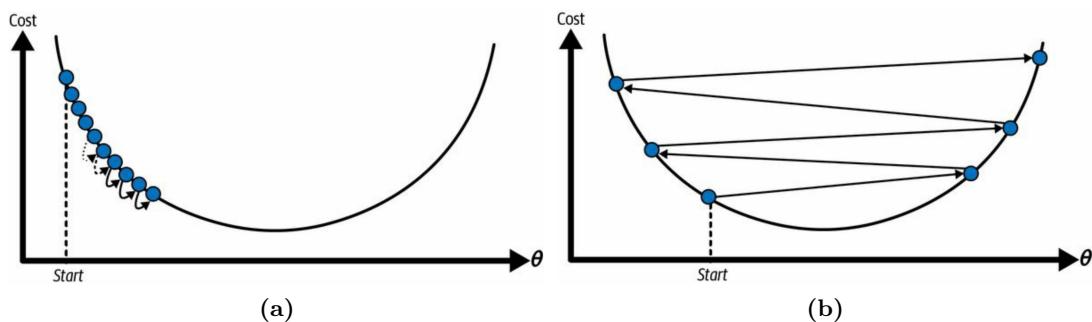
$$\Delta \vec{\theta} = -\eta \vec{\nabla}_{\vec{\theta}} f \quad (4.11)$$

where  $\eta > 0$  is the so-called step size, also known as *learning rate*, which determines how big a step is taken in the direction opposite to the gradient. Choosing the appropriate  $\eta$  is crucial; too small, and the algorithm converges slowly, too large, and it can overshoot the minimum or diverge (see Fig. 4.6). Gradient-based methods make use of the fact that  $f$  decreases the steepest moving in the direction of the negative gradient. The convergence rate of gradient-based methods is generally improved by adaptive step size techniques that adjust step size  $\eta$  on each iteration (Duchi et al., 2011; Schaul et al., 2013; Kingma & Ba, 2017).



**Figure 4.5.** Examples of gradient descents for a (a) one-dimensional loss function and for a (b) two-dimensional one.

Credits: [Géron \(2019\)](#); [Amini et al. \(2018\)](#).



**Figure 4.6.** Examples of gradient descent for a one-dimensional loss (cost) function with (a) learning rate too small and (b) learning rate too high.

Credits: [Géron \(2019\)](#).

All gradient-based optimization logic is encapsulated in the *optimizer* object: it determines the direction in which the model parameters should be adjusted and the magnitude of the parameter update (learning rate).

Advanced optimizers, such as Adam ([Kingma & Ba, 2017](#)) and AdaDelta ([Zeiler, 2012](#)), go beyond the basic gradient descent by adapting the learning rate during the training process based on the properties of the loss surface (e.g. , its curvature). These optimizers can accelerate convergence by dynamically adjusting the step size for each parameter individually, based on the history of gradients. This can be particularly beneficial for navigating complex loss surfaces with varying curvatures.

## Training loop

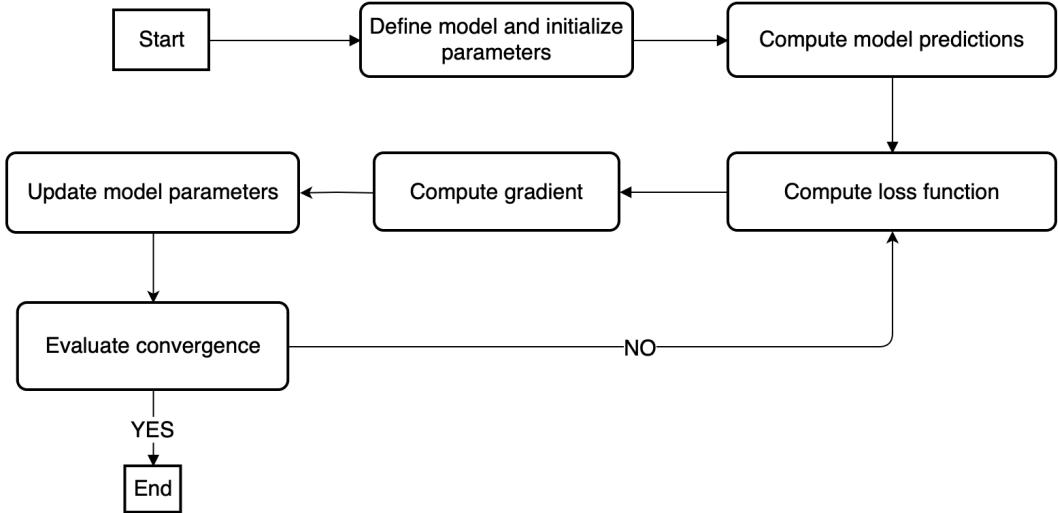
Having outlined all the essential components within the optimization framework, it is now possible to describe the training process in all its phases, as shown in Fig. 4.7.

1. **Initialization:** set up the model to be optimized and initialize its parameters to some (random) values.
2. **Forward pass:** obtain the model predictions based on the initial parameters and compute the loss to quantify how far the model's outputs are from the actual target/observed values.
3. **Backward pass (backpropagation):** calculate the gradient of the loss function with respect to each parameter in the model. This involves propagating the error backwards from the outputs to the inputs, as described in Section 4.1.1. This process systematically computes the partial derivatives using the chain rule to effectively determine how each parameter should be adjusted to minimize loss.
4. **Parameter update (gradient descent):** for each parameter, update its value by moving it in the direction that minimally decreases the loss

$$\vec{\theta}_{final} = \vec{\theta}_{initial} - \eta \vec{\nabla}_{\vec{\theta}_{initial}} f(\vec{\theta}_{initial}). \quad (4.12)$$

5. **Iteration or convergence:** repeat steps 2 through 4 for a predefined number of iterations (or *epochs*) or until the change in the loss function between iterations falls below a small threshold, indicating convergence.

The following sections will detail the methods used for the simulation and analysis of the objects presented in Chapter 5. Beginning with the modeling of light sources and the adopted surface brightness profiles, the discussion will advance to the techniques



**Figure 4.7.** Optimization algorithm diagram.

used for simulating and reconstructing strong gravitational lenses. Finally, methods will be described to assess the accuracy of the model fit to the observed/simulated data.

## 4.2 Light sources modeling

The prevalent approach to modeling light sources such as galaxies involves using a parametric profile ( $R$ ), where  $R$  represents a measure of distance from the center of the object. These models often presuppose an excessively idealized level of symmetry, yet offer the advantage of being straightforward to define and apply. When building galaxy components, it is generally assumed that the profiles exhibit elliptical symmetry (Peng et al., 2002, 2010), and therefore  $R$  represents the elliptical radius

$$R(\vec{x}') = \sqrt{{x'_1}^2 + {x'_2}/q^2} \quad (4.13)$$

of a source with an axis ratio  $q$ . The coordinate system  $(x'_1, x'_2)$  of the source can be rotated by the *position angle*  $\varphi$  with respect to the coordinate system  $(x_1, x_2)$  of the observation. The resulting isophotes<sup>5</sup> of the surface brightness distribution of such an elliptical source are ellipses with semi-major axis  $R/q$ , semi-minor axis  $R$  and orientation  $\varphi$  with respect to the  $x_1$  axis of observation.

---

<sup>5</sup>Curves of constant surface brightness.

### 4.2.1 Sérsic profile

The most common model to describe elliptical surface brightness distributions is the *Sérsic law* ([Sérsic, 1963, 1968](#)), which is given by the exponential

$$I(R) = I_e \exp \left\{ -b_n \left[ \left( \frac{R}{R_e} \right)^{\frac{1}{n}} - 1 \right] \right\}, \quad (4.14)$$

where  $I_e$  is the surface brightness at the effective radius<sup>6</sup>  $R_e$  and  $n > 0$  is called the *Sérsic index*, which characterizes the slope of the profile. The function  $b_n$  depends only on the Sérsic index and is defined by

$$\gamma(2n, b_n) = \frac{1}{2} \Gamma(2n), \quad (4.15)$$

where  $\Gamma, \gamma$  are, respectively, the Gamma function and the incomplete Gamma function. It can be shown ([Ciotti, 1991; Ciotti & Bertin, 1999](#)) that, for a Sérsic index in the range  $0.5 \leq n \lesssim 8$ ,  $b_n$  can be approximated by

$$b_n \approx 2n - \frac{1}{3} + \frac{4}{405n} \approx 1.9992n - 0.3271. \quad (4.16)$$

With this, the profile is now fully determined by the seven parameters for position  $x_1, x_2$ , effective radius  $R_e$ , Sérsic index  $n$ , intensity at the effective radius  $I_e$ , axis ratio  $q$  and position angle  $\varphi$ .

The Sérsic profile is a versatile model; by varying  $n$  it is possible to obtain many of the classical galaxy profiles as special cases, such as Gaussian profiles ( $n = 0.5$ ), exponential profiles ( $n = 1$ ) and de Vaucouleurs ([de Vaucouleurs, 1948](#)) profiles ( $n = 4$ ).

### 4.2.2 Core-Sérsic profile

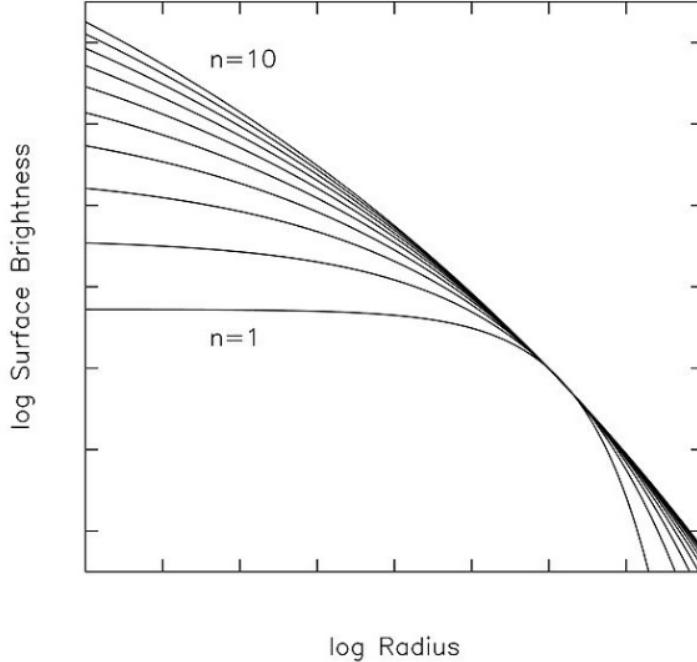
An extension of the Sérsic profile has been developed by [Graham et al. \(2003, 2004\); Trujillo et al. \(2004\)](#) to better reproduce the observed radial profiles of surface brightness. This model consists of a power law to model the inner radii and a Sérsic function to describe the outer stellar distribution. It is given by

$$I(R) = I' \left[ 1 + \left( \frac{R_b}{R} \right)^\alpha \right]^{\gamma/\alpha} \exp \left\{ -b_n \left[ \frac{R^\alpha + R_b^\alpha}{R_e^\alpha} \right]^{1/(\alpha n)} \right\}, \quad (4.17)$$

where  $R_b$  is the break-radius separating the inner power law with logarithmic slope  $\gamma$  from the outer Sérsic profile with slope  $n$ . The intensity  $I_b$  at the break-radius  $R_b$  can

---

<sup>6</sup>Also called the half-light radius, the radius within which half of the galaxy's luminosity is contained.



**Figure 4.8.** Sérsic profiles for different values of  $n$ . On average,  $n \approx 2 - 10$  for bulges and elliptical galaxies,  $n \approx 1$  for disk galaxies and  $n \leq 0.5$  for bars and stellar clumps. Credits: [Burke \(2013\)](#).

be evaluated from the expression

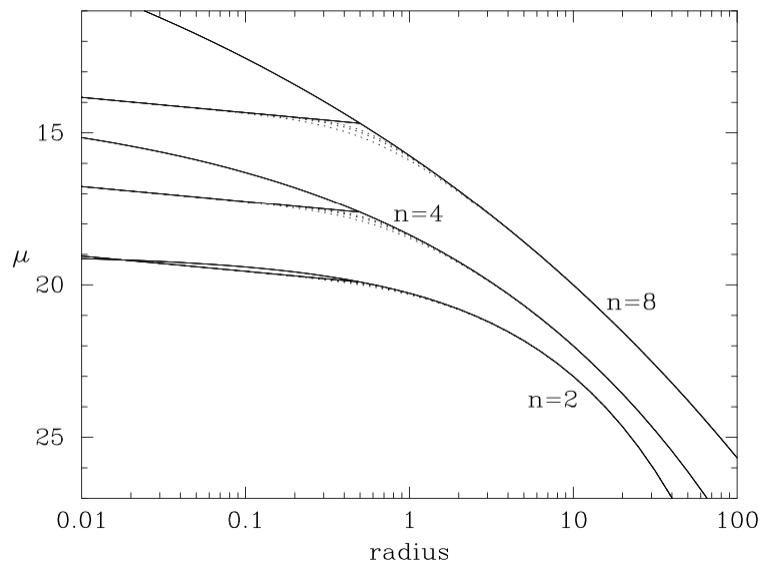
$$I' = I_b 2^{-\gamma/\alpha} \exp \left\{ b_n \left[ \frac{2^{1/\alpha} R_b}{R_e} \right]^{1/n} \right\}. \quad (4.18)$$

The parameter  $\alpha$  controls the sharpness of the transition between the inner (power law) and outer (Sérsic) regimes, with higher values indicating sharper transitions (Fig. 4.9).

In this case, the number of parameters becomes larger to include the inner-outer change of regime: the profile is fully determined by ten parameters: for position  $x_1$ ,  $x_2$ , effective radius  $R_e$ , Sérsic index  $n$ , intensity  $I_b$  at the break-radius  $R_b$ , slope of the inner power law  $\gamma$ , sharpness of the transition  $\alpha$ , axis ratio  $q$  and position angle  $\varphi$ .

#### 4.2.3 Sky component

Observations usually include a diffuse distribution of light known as the *sky background*, which must be considered in the reconstruction process. The most basic model for this sky component is a flat surface brightness distribution, characterized by a constant value or a fixed gradient along the  $x_1$  and  $x_2$  axes of the lens plane. Even if the



**Figure 4.9.** Different core-Sérsic profiles illustrated by the dotted curves for a range of structural parameters. Profiles with values of  $\alpha$  equal to 2, 3, and 4 are shown, the latter giving the sharpest transition. In all models  $R_e = 10$  arcsec,  $R_b = 0.5$  arcsec, and  $\gamma = 0.2$ . For comparison, an inner power law with slope equal to  $-0.2$  is shown (diagonal solid lines), as are Sérsic profiles (solid curves) having the same Sérsic shape index  $n$ .

Credits: [Graham et al. \(2003\)](#).

diffuse background has been removed during a pre-processing step, incorporating a sky component as a free element in the model is advisable. Distinguishing between sky and actual signal is challenging, and any over- or underestimation of the subtracted light can have a considerable impact on the reconstruction outcomes.

#### 4.2.4 Point Spread Function

The Point Spread Function (PSF) is a fundamental concept in optical physics and astronomy that describes how a system blurs or spreads a point light source in an image. The PSF encapsulates the response of an imaging system to a point source or point object, representing the diffraction pattern caused by the optics of the system, including aberrations (Bovik, 2005). Understanding the PSF is crucial for interpreting and processing observational data, as it affects the accuracy with which one can measure the properties of observed objects.

Mathematically, the PSF is often modeled as a convolution kernel that, when applied to an ideal image (the image that would be observed in the absence of any blurring effects), produces the observed image. The process of convolution mathematically represents spreading the light from point sources over a larger area of the detector, affecting the observed shapes, sizes, and brightness of these objects.

The effect of PSF on an observed image on a plane  $(x, y)$  can be described by the convolution of the true image  $f(x, y)$  with the Point Spread Function  $\text{PSF}(x, y)$ , resulting in the observed image  $g(x, y)$  (Howell, 2006):

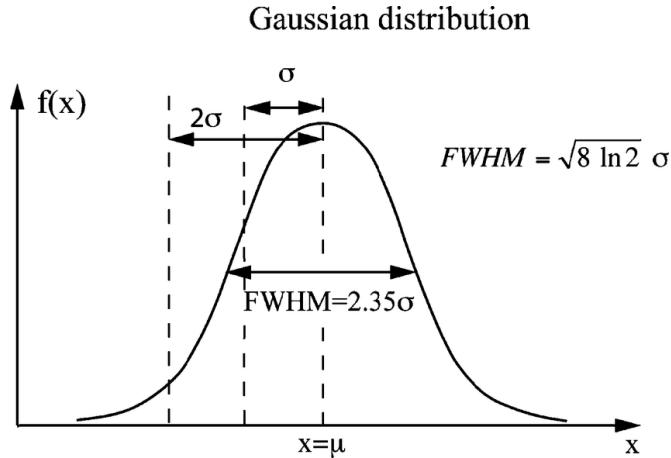
$$g(x, y) = (f * p)(x, y) = \iint f(x', y') \text{PSF}(x - x', y - y') dx' dy' . \quad (4.19)$$

The Point Spread Function (PSF) is usually expressed by means of the Full Width at Half Maximum (FWHM). Their link is rooted in the definition and characterization of the PSF itself. The PSF describes the response of an imaging system to a point source, represented as a distribution or function of intensity across the image plane. The FWHM is a derived characteristic of this distribution, specifically measuring the width of the PSF at half of its maximum intensity. In the usual case of a Gaussian PSF (Fig. 4.10):

$$\text{PSF}(x, y) = \text{PSF}_0 \exp \left\{ -\frac{4 \ln 2}{\text{FWHM}^2} (x^2 + y^2) \right\} , \quad (4.20)$$

where  $\text{PSF}_0$  is the peak intensity, the FWHM is directly related to the standard deviation  $\sigma$  of the Gaussian distribution by:

$$\text{FWHM} = 2\sqrt{2 \ln 2}\sigma . \quad (4.21)$$



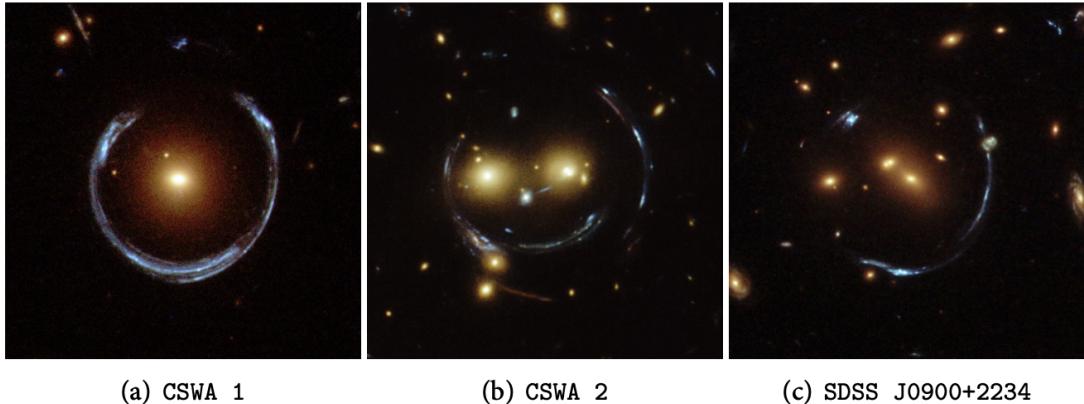
**Figure 4.10.** Gaussian or normal PSF and the link with the FWHM.  
Credits: [Tavernier \(2010\)](#).

### 4.3 Lens reconstruction

One of the primary applications of gravitational lensing is reconstructing the lens mass distribution. The modeling of gravitational lenses begins with a collection of observables (Fig. 4.11), such as the relative positions and fluxes of images, time delays between images, and other properties of the lens. This modeling can be approached in one of two different but related ways: as a *forward* problem of creating a model of the lens system (lenses and sources) that approximates the observed image the best, or as the *inverse* problem of finding a lens model that deconstructs the observation into a self-consistent and physically viable image of the source. Many successful applications have been published for both the forward ([Bandara et al., 2013](#); [Bolton et al., 2008](#); [Newton et al., 2011](#); [Peng et al., 2006](#)) and the inverse method ([Dye & Warren, 2005](#); [Kochanek & Narayan, 1992](#); [Nightingale & Dye, 2015](#); [Suyu et al., 2006](#); [Vegetti & Koopmans, 2009](#); [Wallington et al., 1996](#); [Warren & Dye, 2003](#)).

Three types of constraints can be used for this purpose:

1. the locations of the multiple images produced by a lensed source help to map out the deflection field of the lens, which corresponds to the first derivatives of the lensing potential ([Cardone et al., 2001](#)). Obtaining these constraints is relatively straightforward assuming the availability of high-resolution imaging data. This work will mainly focus on this type of constraint.
2. the magnifications (fluxes) and shapes of the multiple images and gravitational arcs explore the higher-order derivatives (mainly the second) of the lensing potential



(a) CSWA 1

(b) CSWA 2

(c) SDSS J0900+2234

**Figure 4.11.** Examples of observations of strong gravitational lensing. The large arcs contain information to constrain the gravitational lenses producing these images. Credits: ESA/Hubble & Nasa.

([Gilman et al., 2019](#)). Therefore, these constraints are particularly sensitive to the smaller-scale mass components of the lens;

3. the relative time delays between multiple images serve as probes of the lensing potential, as shown in Section 2.2.4. However, time delays can only be measured in a limited number of lenses (only a few tens) because they require the lensed sources to be intrinsically variable, such as quasars or supernovae ([Refsdal, 1964](#)). These types of source are rare. Additionally, measuring time delays is challenging; it demands dedicated telescope time for the continuous observation of these sources over extended periods, along with precise photometry.

The method of translating observed strong lensing constraints into distributions of matter is referred to as lens *inversion* or *reconstruction*.

Such reconstruction is often tackled using two main categories of inversion algorithms *parametric* and *non-parametric* algorithms, based on whether the calculation is “model-based” (parametric) or “model-free” (non-parametric) at the start of the process ([Coe et al., 2008](#)). This work is focused on the first type of approach, applying parametric optimization to retrieve the mass distribution of the lens.

Each approach has its own particular set of strengths and weaknesses, which are summarized here.

1. Parametric models employ a clear physical parameterization from the beginning, defining the mass distribution based on a set of specific parameters. Parametric simulation models are generally used to solve the forward problem, taking a source and lensing mass and then predicting the resulting image. Also known as

simply-parameterized models, they assume a physical model that fits the data with relatively few defined parameters ([Jullo et al., 2007](#)). In parametric models, the data is fitted to a physical object (e.g. Point-Mass, Singular Isothermal Sphere, Singular Isothermal Ellipsoid, De Vaucouleurs model, etc.) and a model of the lensing mass made using that physical object to predict the effect on the light from the source, often with the assumption that mass follows light. The exploration of these models parameter space aims to identify the optimal combination that accurately replicates the observed positions, shapes, magnitudes, and relative time delays of the multiple images and arcs;

2. non-parametric methods, which do not start with a predefined physical model but instead use a “grid-based” approach, among others, to model the mass distribution. The lens is divided into a mesh, which can be either structured or unstructured, and the lensing observables are projected onto this mesh. Subsequently, this mesh is converted into a pixelized mass distribution using the relationships between the observables and the surface density of the lens. This technique has been extensively applied on a variety of scales, from galaxies to clusters, and has been implemented in numerous ways ([Birrer et al., 2015](#); [Blandford et al., 2000](#); [Coles et al., 2014](#); [Diego et al., 2005, 2007](#); [Koopmans, 2005](#); [Liesenborgs et al., 2006](#); [Merten, 2016](#); [Saha & Williams, 2004](#); [Sebesta et al., 2016](#); [Suyu & Blandford, 2006](#); [Suyu et al., 2009](#)).

#### 4.3.1 Parametric reconstruction

Strong lensing parametric reconstruction, also known as inversion, involves the use of predefined models to describe both the mass distribution of the lensing object and the light distribution of the background source. The aim is to adjust the parameters of these models to best fit the observed lensing phenomena, such as arcs, rings, and multiple images of a background source. The process is detailed and involves several key steps and algorithms:

1. **model selection** for both the mass distribution and the light distribution;
2. **parameter initialization** for the parameters of the mass and light distribution models are made based on observational data or theoretical considerations;
3. **ray-tracing simulation** using the initial models: a ray-tracing algorithm computes the deflection angles at each point in the lens plane, which are used to trace the light rays back to the source plane. This step predicts the appearance of the lensed images based on the current model parameters;

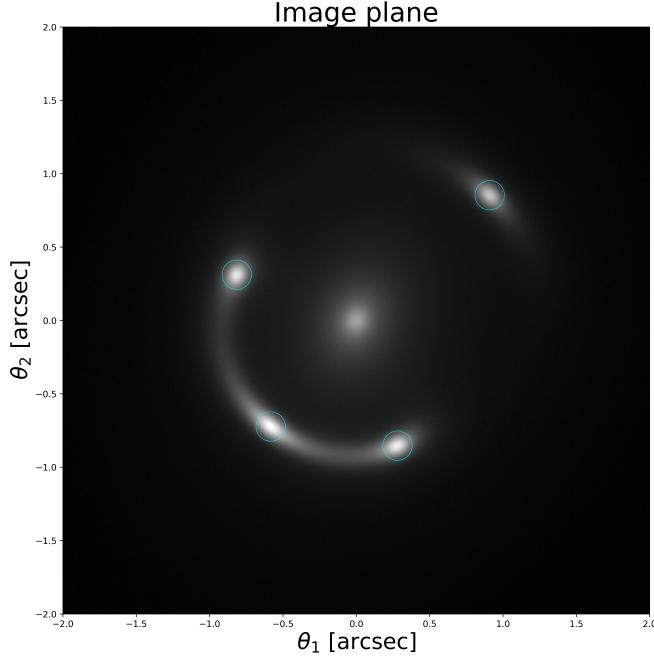
4. **optimization** via an objective (loss) function that quantifies the difference between the observed and predicted lensed images, as described in Section 4.1. This function often includes terms for the positions, shapes, and flux ratios of the observed and model-predicted images. Methods such as gradient descent (Ruder, 2016; Canu et al., 2016), Markov Chain Monte Carlo (MCMC) (Geyer, 1992; Speagle, 2019), or Nested-Sampling (Skilling, 2004; Buchner, 2023) are used to adjust model parameters to minimize the objective function. This iterative process refines the model to better fit the observations;
5. **model comparison and selection** if multiple models are being tested, and statistical criteria such as the Akaike Information Criterion (AIC) (Cavanaugh & Neath, 2019) or Bayesian Information Criterion (BIC) (Liddle, 2007; Chen & Chen, 2008) are used to compare their fit to the data, helping to select the best model;
6. **uncertainty quantification** for the model parameters, often through bootstrapping or by analyzing the posterior probability distributions in methods like MCMC, providing insights into the confidence levels of the mass and light distribution parameters.

As already stated, parametric reconstruction is highly dependent on the chosen models and their ability to accurately represent the complex mass distributions of astronomical objects. Its strength lies in its relatively straightforward interpretation and the physical significance of the model parameters, but it also faces limitations in flexibility compared to non-parametric methods.

There are several ways to implement parametric lens reconstruction and, as will be shown in Chapter 5, one of the most common approaches is to use the observed positions of multiple images (Fig. 4.12). The optimization can then be performed in mainly two ways: with the so-called *lens* or *image plane optimization* or with the *source plane optimization* technique.

### Lens plane and source plane optimization

In the context of lens inversion, the positions of multiple images  $\vec{\theta}_i$  are the main observable feature that can be used to constrain the lens parameters. It is assumed that the lens is described by a surface density model (e.g. SIS, SIE, etc.) depending on a set of  $n$  parameters  $\vec{p} = [p_1, \dots, p_n]$ , starting at some initial values (random or constrained by other observables) and that the redshifts of the source and lens are known.



**Figure 4.12.** Simulated image of a strong lensing event with multiple images, identified by the azure circles. The image is  $4'' \times 4''$ , with a pixel scale of  $0.03''$ .

- **Lens plane optimization**

Given a set of lens parameters, it is possible to calculate the lens deflection angle at the observed positions of the images  $\vec{\alpha}(\vec{\theta}_i|\vec{p})$  and, using the lens equation, map these positions on the source plane:

$$\vec{\beta}_i(\vec{p}) = \vec{\theta}_i - \vec{\alpha}(\vec{\theta}_i|\vec{p}). \quad (4.22)$$

The resulting points  $\vec{\beta}_i(\vec{p})$  will be spread over a certain region in the source plane since the model parameters are not correct. The predicted source position can be estimated by taking the mean position of these points:

$$\vec{\beta}(\vec{p}) = \frac{1}{N_{ima}} \sum_{i=1}^{N_{ima}} \vec{\beta}_i(\vec{p}), \quad (4.23)$$

where  $N_{ima}$  is the observed number of multiple images.

Subsequently, by solving the lens equation, the predicted source position can be

mapped back to the lens plane, obtaining a set of predicted image positions

$$\vec{\beta}(\vec{p}) \rightarrow \{\vec{\theta}_i(\vec{p}); i = 1, \dots, N_{ima,\vec{p}}\}, \quad (4.24)$$

where the number of predicted images,  $N_{ima,\vec{p}}$ , can clearly be different from  $N_{ima}$ . At this point a cost function can be defined to compare the predicted and observed positions of the images

$$\chi^2(\vec{p}) = \sum_{i=1}^{N_{ima}} \frac{[\vec{\theta}_i - \vec{\theta}_i(\vec{p})]^2}{\sigma_i^2}, \quad (4.25)$$

where  $\vec{\theta}_i(\vec{p})$  is the position of the predicted image closest to the image at  $\vec{\theta}_i$  and  $\sigma_i$  is the error on the position of the  $i$ -th image.

To identify the optimal lens model, one can adjust the parameters  $\vec{p}$  to minimize the loss function, or equivalently, to maximize the likelihood of the observed image positions given the parameters  $\vec{p}$

$$\mathcal{L}(\vec{p}) = \frac{1}{\prod_{i=1}^{N_{ima}} \sigma_i \sqrt{2\pi}} \exp \left[ -\frac{\chi^2(\vec{p})}{2} \right]. \quad (4.26)$$

When dealing with multiple families of multiple images, the same approach can be utilized. However, the likelihood function must be adjusted to become the product of the likelihoods corresponding to each image family

$$\mathcal{L}(\vec{p}) = \prod_{j=1}^{N_{fam}} \frac{1}{\prod_{i=1}^{N_{ima}} \sigma_{ji} \sqrt{2\pi}} \exp \left[ -\frac{\chi_j^2(\vec{p})}{2} \right], \quad (4.27)$$

with  $\chi_j^2$  as defined in Eq. (4.25), but referring to the  $j$ -th family of multiple images and  $\sigma_{ji}$  is the error on the position of the  $i$ -th image of the  $j$ -th family.

Maximizing the aforementioned likelihood is equivalent to minimizing the total chi-squared value  $\chi_{tot}^2$ , which is defined by

$$\chi_{tot}^2(\vec{p}) = \sum_{j=1}^{N_{fam}} \chi_j^2(\vec{p}). \quad (4.28)$$

- **Source plane optimization**

An alternative to optimizing in the lens plane is to perform optimization in the source plane. In an ideal scenario, if the model parameters precisely match the lens parameters, multiple images would map to a single position on the source plane. Consequently, a cost function can this time be defined on the source plane to facilitate this optimization process

$$\chi_s^2(\vec{p}) = \sum_{i=1}^{N_{ima}} \frac{[\vec{\beta}_i(\vec{p}) - \vec{\beta}(\vec{p})]^2}{\sigma_i^2} \mu_i(\vec{p})^2, \quad (4.29)$$

where  $\mu_i(\vec{p})$  is the model estimated magnification at the  $i$ -th image position.

The primary benefit of source plane optimization is the elimination of the most complex step in the previously described lens plane optimization procedure. This step, which involves solving the lens equation, can be computationally intensive and is typically addressed by numerical methods. When leveraging multiple families of images as constraints, optimization on the lens plane becomes considerably more time consuming compared to source plane optimization. Furthermore, this approach does not require a one-to-one match between the observed and predicted images by the model when calculating the chi-squared  $\chi^2$ . However, a potential downside is that the error for each predicted source position  $\beta_i(\vec{p})$  must be estimated using the model-predicted magnification  $\mu_i(\vec{p}) = \mu(\vec{\theta}_i|\vec{p})$  at the image positions. Additionally, optimization on the source plane may produce biased solutions, often favoring models with flatter density profiles and higher ellipticity ([Kneib & Natarajan, 2011](#)).

# 5

## Applications

In the following sections, we discuss specific lensing applications on three different scales. Our discussion initiates with a detailed examination of microlensing events, phenomena where the gravitational field of a relatively small object, such as a star or planet, magnifies the light of a background star. This section will focus on the technique of light-curve fitting, a method that analyzes the temporal change in brightness of the lensed star to extract critical information about the lensing object. Microlensing offers a unique method for detecting elusive objects, including exoplanets ([Sumi et al., 2011](#); [Gaudi, 2012](#)) and dark compact objects, providing insights into the mass, motion, and distribution of such entities within the Milky Way ([Alcock et al., 1993](#); [Udalski et al., 1993](#)).

Transitioning from the stellar to the galactic scale, we next explore the realm of galaxy-scale lensing ([Treu, 2010](#)) by optimizing a parametric strong lensing model. This segment emphasizes the development and optimization of parametric models for strong lensing, where entire galaxies act as lenses, distorting and magnifying the light of more distant galaxies into arcs and rings. Through the optimization of these models, we aim to reconstruct the mass distribution of lensing galaxies ([Auger et al., 2010](#)), including the elusive dark matter component ([Vegetti et al., 2010](#)). This approach not only sheds light on the structure and composition of galaxies, but also contributes to our understanding of the evolution of the large-scale structure of the Universe.

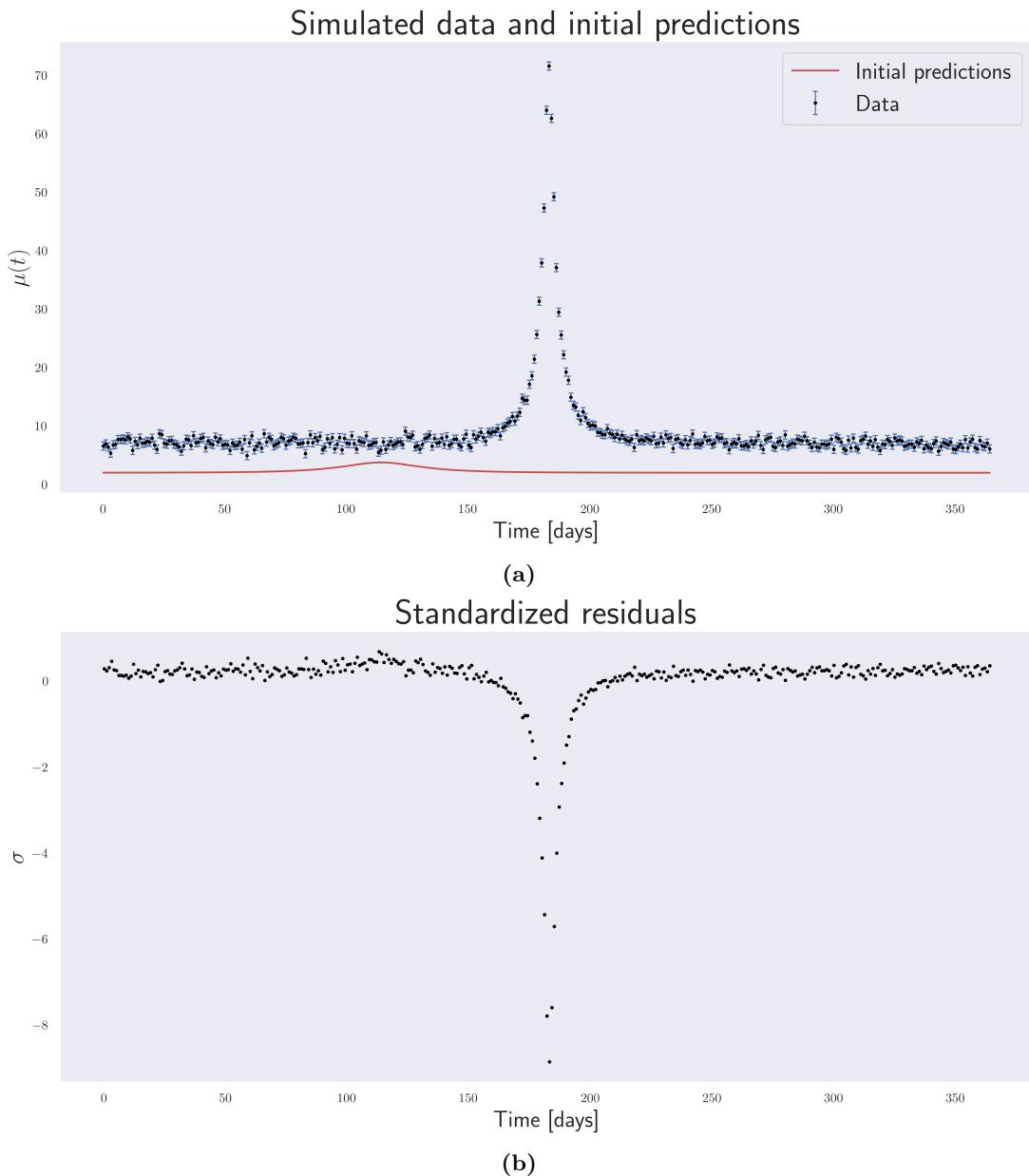
Finally, we will address weak lensing by fitting surface brightness profiles and measuring ellipticities.

### 5.1 Microlensing light-curve fit

The standard microlensing light-curve, also known as the Paczynski light-curve, ([Paczynski, 1986](#)), as introduced in Section 3.1, is defined by three parameters: the impact parameter  $y_0$ , the maximum magnification time  $t_0$ , and the Einstein radius crossing time  $t_E$ , which represents the time scale of the microlensing event. The latter implicitly depends on other relevant quantities of the lensing system, the lens mass  $M$ , the relative velocity between lens and source  $v_{rel}$  and the distances between lens and source from the observer  $D_L$  and  $D_S$ . The combination of these parameters defines the amplification

to which a source with intrinsic flux  $f_S$  is subjected.

The first step consists of simulating a microlensing event to obtain synthetic observational data, to which noise is then added in such a way as to mimic observational errors. We start by considering a lens with mass  $M = 0.3 M_\odot$  placed at a distance  $D_L = 4.0$  kpc. We then define a source behind the lens, at a distance  $D_S = 8.0$  kpc, characterized by an intrinsic flux of  $f_S = 7.0$  (in arbitrary units), which moves with relative velocity  $v_{rel} = 200.0 \text{ km s}^{-1}$  with respect to the lens. Assuming that one could monitor the source for a long period (one year in this case), a microlensing event happens when the source moves across the Einstein radius of the lens, causing a peak in the magnification. In this simulation, this occurs at  $t_0 = 183.0$  d. After setting up the model parameters, shown in Table 5.1, the simulated light curve can be computed and the result is shown in Fig. 5.1 (black dots). The errors in the photometric measurement are assumed to be 10% of the baseline flux. Subsequently, the model was initialized with random parameters. Predictions of the model, given the set of initial parameters, are shown in Fig. 5.1 (red line), along with the standardized residuals, which are expressed in units of standard deviation from the mean.



**Figure 5.1.** (a) Initial predictions of the model and (b) residuals expressed in units of standard deviation from their mean.

**Table 5.1.** Summary of the parameters before and after fitting the model with the set of degenerate parameters.

Parameter	True value	Initial value	Best-fit value (degeneracy)	Best-fit value (MCMC)
$f_S$	7.0	2.0	6.993	$(7.017 \pm 0.040)$
$M$	$0.3 M_\odot$	$0.8 M_\odot$	$0.556 M_\odot$	$(0.620 \pm 0.466) M_\odot$
$y_0$	0.1	0.6	0.100	$(0.101 \pm 0.001)$
$v_{rel}$	$200.0 \text{ km s}^{-1}$	$180.0 \text{ km s}^{-1}$	$182.881 \text{ km s}^{-1}$	$(231.722 \pm 49.037) \text{ km s}^{-1}$
$t_0$	183.0 d	114.0 d	183.002 d	$(183.018 \pm 0.023)$ d
$D_L$	4.0 kpc	2.0 kpc	1.548 kpc	$(3.166 \pm 1.698)$ kpc
$D_S$	8.0 kpc	7.0 kpc	3.786 kpc	$(8.658 \pm 3.457)$ kpc

After initializing the model, the cost function was defined to quantify the discrepancy between the model and the observed data. In this case, a reduced chi-square  $\chi^2_\nu$  was specified as the metric of choice. It is defined as

$$\chi^2_\nu = \frac{\chi^2}{\nu} = \frac{1}{\nu} \sum_i \left[ \frac{(O_i - E_i)^2}{\sigma_i^2} \right], \quad (5.1)$$

where  $O_i$  and  $E_i$  are the  $i$ -th observed and expected values respectively, while  $\sigma_i$  is the data uncertainty on the  $i$ -th observed value. The term  $\nu$  represents the degrees of freedom, calculated as the number of observations minus the number of fitted parameters. A value of  $\chi^2_\nu$  close to 1 suggests that the model fits the data well, with deviations between the observed and expected values largely attributable to random noise. A value significantly greater than 1 indicates that the model does not fit the data well, possibly due to systematic errors, an incorrect model, or an underestimation of the variability in the data. On the contrary, a value significantly less than 1 may suggest that the model fits the data too well, potentially due to overfitting or overestimation of the variability in the data.

At this stage, Adam (Kingma & Ba, 2017) is set as the optimizer to perform a gradient descent optimization on the cost function of choice and minimize its value by updating the model parameters at each iteration. Training starts with a loss value of  $\chi^2_\nu = 192.939$  and, after 1000 iterations, the algorithm stops after reaching a  $\chi^2_\nu = 1.028$ .

As can be seen Table 5.1, even though the input parameters  $f_S$ ,  $y_0$  and  $t_0$  are accurately estimated, almost matching their true values precisely, the other parameters show considerable deviation from their actual values. This discrepancy was expected

**Table 5.2.** Summary of the parameters before and after fitting the model a second time to avoid degeneracy.

Parameter	True value	Initial value	Best-fit value (training)	Best-fit value (MCMC)
$f_S$	7.0	2.0	7.065	$(7.052 \pm 0.041)$
$M^*$	$0.3 M_\odot$	$0.8 M_\odot$		
$y_0$	0.1	0.6	0.103	$(0.100 \pm 0.001)$
$v_{rel}^*$	$200.0 \text{ km s}^{-1}$	$180.0 \text{ km s}^{-1}$		
$t_0$	183.0 d	114.0 d	182.989 d	$(183.010 \pm 0.022)$ d
$D_L^*$	4.0 kpc	2.0 kpc		
$D_S^*$	8.0 kpc	7.0 kpc		
$t_E$	19.1369 d	29.3461 d	19.025 d	$(19.054 \pm 0.207)$ d

\*: Parameters not fitted to avoid degeneracy.

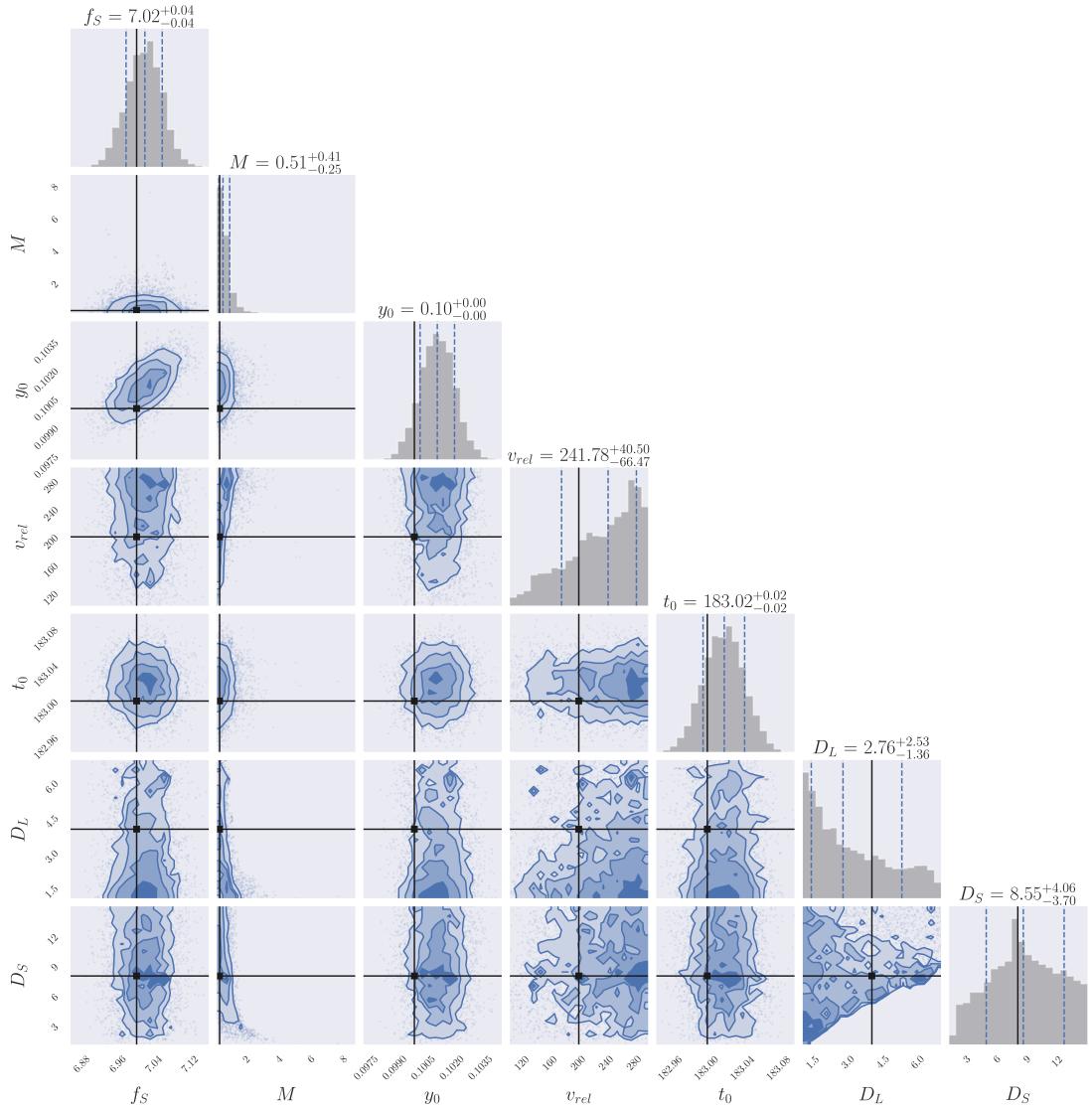
and, as introduced in Section 3.1, is due to the high microlensing degeneracy, as these parameters collectively influence the determination of the Einstein crossing time  $t_E$ . In fact, it is the Einstein crossing time alone that shapes the light curve’s profile.

This degeneracy can also be seen after performing a Bayesian sampling of the posterior probability distribution of the parameters using the Monte Carlo Markov Chain (MCMC) sampler implemented in the Pyro<sup>1</sup> framework (Bingham et al., 2018), which is a probabilistic programming language built on top of PyTorch, designed to create and run complex probabilistic models. The resulting posterior probability distributions of the parameters are shown in Fig. 5.2. In the literature, numerous methods have been known to attempt to break this degeneracy (Lee, 2017). In this case, the choice was made to redefine the model so that it directly depends on  $t_E$ , a quantity that, as mentioned, implicitly contains all the degenerate parameters within it.

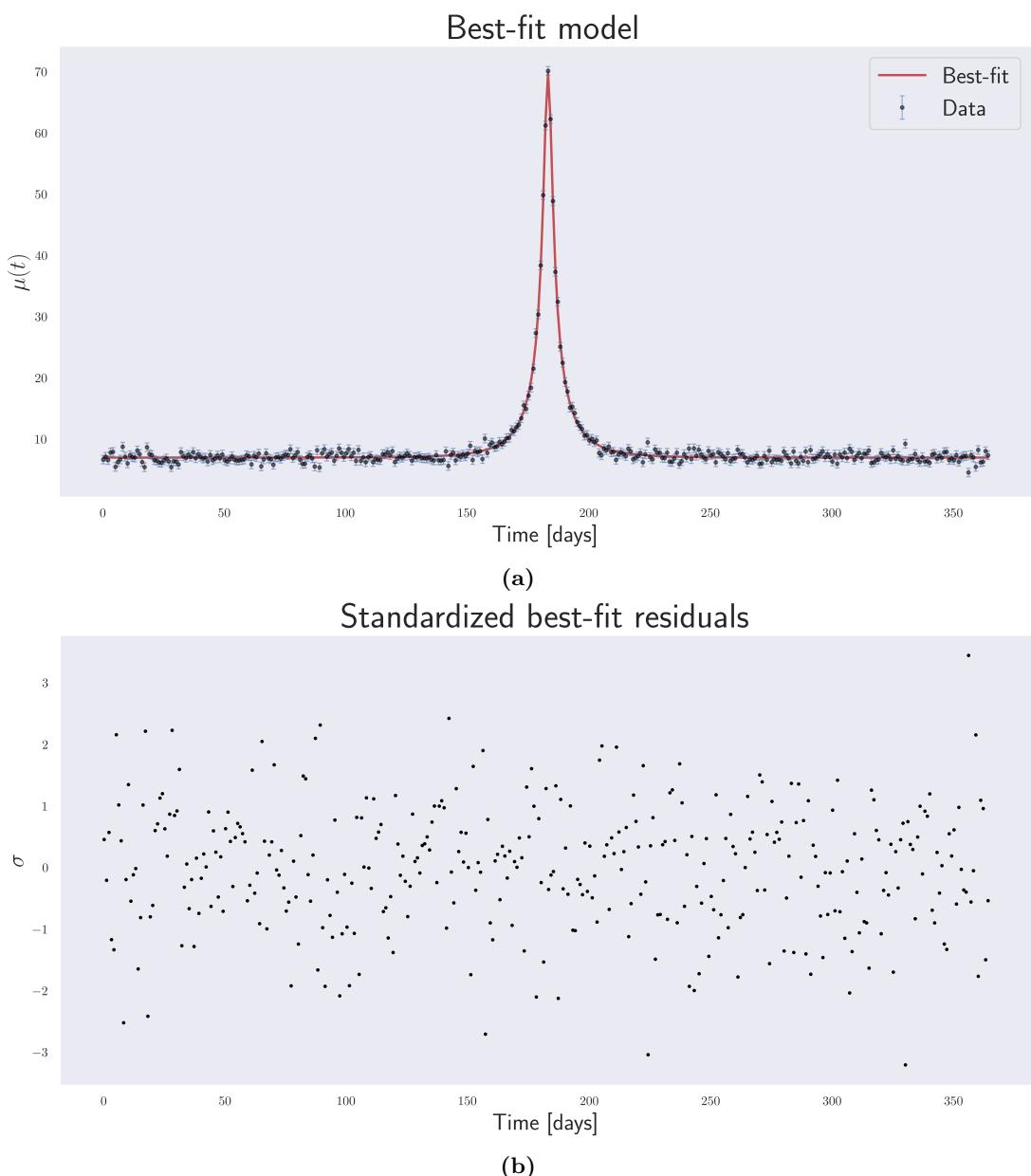
The fitting procedure for the simulated observations was therefore repeated, defining a model with the same initial parameters but fitting only the parameters  $f_S$ ,  $t_0$ ,  $y_0$  and  $t_E$ . In this case, the initial reduced chi-square value is  $\chi_\nu^2 = 193.270$ , and after training the model for the same number of epochs as before, it drops to a value of  $\chi_\nu^2 = 1.009$ . The best-fit parameters are indicated in Table 5.2, while Figs. 5.3 and 5.5 show, respectively, the best-fit model with its standardized residuals, and the trend of the cost function as a function of the number of iterations.

---

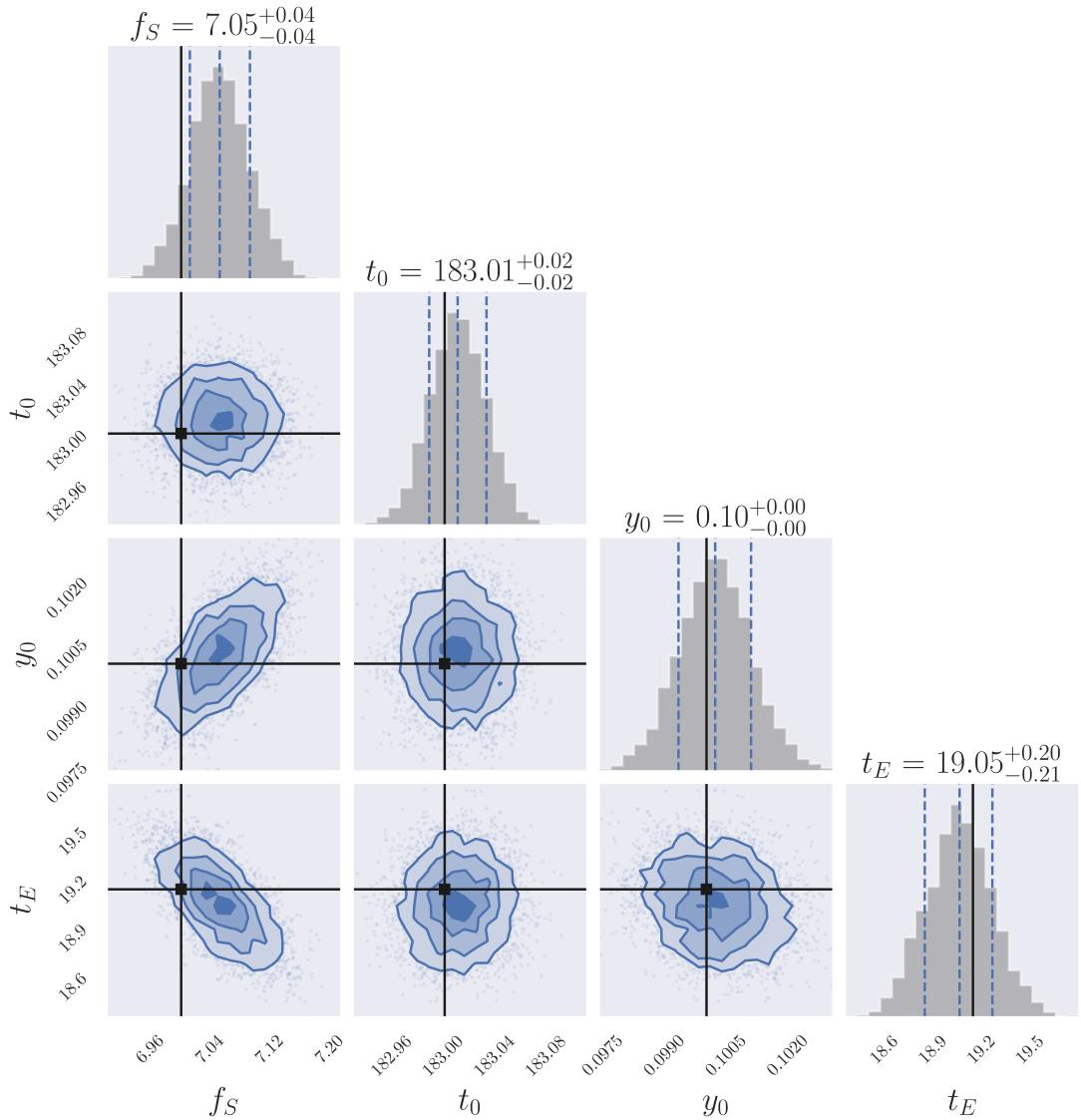
<sup>1</sup><https://pyro.ai/>



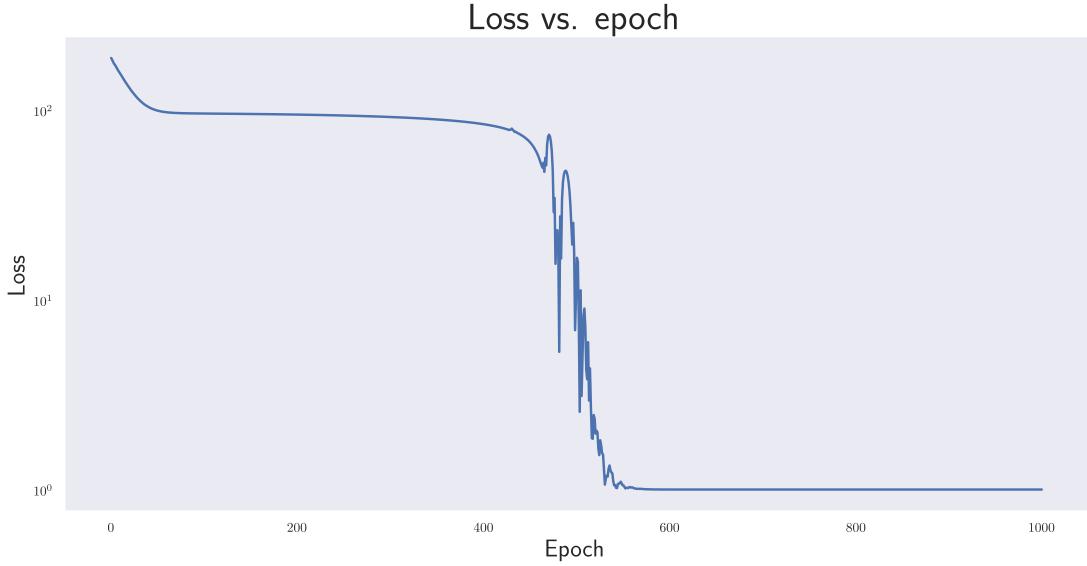
**Figure 5.2.** Corner plot showing the posterior probability distributions of the parameters used to fit the light-curve in Fig. 5.1a. The projections of the probability density on planes defined by each couple of parameters, as well as one-dimensional marginalized distributions are shown. The degeneracy among the parameters  $M$ ,  $v_{rel}$ ,  $D_L$  and  $D_S$  is clearly visible.



**Figure 5.3.** (a) Predictions of the best-fit model and (b) residuals expressed in units of standard deviation from their mean.



**Figure 5.4.** As Fig. 5.2 but for a model whose free parameters are  $f_S$ ,  $t_0$ ,  $y_0$ ,  $t_E$ .



**Figure 5.5.** Loss value as a function of epochs.

## 5.2 Strong lens parametric reconstruction

This section of the thesis delves into the intricacies of parametric strong lensing, focusing on the Singular Isothermal Ellipsoid (SIE) model, a widely adopted representation for the mass distribution of lensing galaxies. The SIE model provides a robust framework for simulating gravitational lenses, offering a simplified, yet effective, portrayal of their mass profiles.

Taking advantage of the positions of multiple images generated by the lensing effect, one can derive the parameters of the lensing system that produced them. The objective is to optimize the lens parameters, which dictate the lensing behavior and, consequently, the appearance and positions of the lensed images.

This optimization is pursued through two distinct methodologies, as discussed in Section 4.3. Starting in both techniques from the positions of the multiple images, the first method focuses on optimizing the lens parameters within the source plane, by mapping the positions of the images back in the source plane and minimizing the distance between the theoretical and inferred position of the source.

The second method shifts the focus to the lens plane by directly minimizing the distances between each observed and predicted image to optimize the lens parameters.

**Table 5.3.** Summary of lens and source parameters.

LENS		SOURCE	
Parameter	Value	Parameter	Value
$z_L$	0.3	$z_S$	1.5
$s_v$	$200.0 \text{ km s}^{-1}$	$q$	0.8
$q$	0.3	$\varphi$	1.0472 rad
$\varphi$	0.7854 rad	$f$	$4.5 \times 10^3$
$f$	$1 \times 10^3$	$R_e$	$0.3''$
$R_e$	$5.0''$	$n$	4.0
$n$	3.5	$y_{1c}$	$-0.1676''$
$x_{1c}$	$0.0''$	$y_{2c}$	$0.0838''$
$x_{2c}$	$0.0''$		

### 5.2.1 Lens modelling

The first step, as always, is to produce a simulation of the lensing system and the main observables of this topic, the multiple images of the background source.

The process begins by defining the observational parameters: a grid of  $100 \times 100$  pix is established, with a resolution of  $0.03'' \text{ pix}^{-1}$ . A background noise per pixel of  $rms = 0.5$  mimics real observational noise. Additionally, the final image is also convolved with a Gaussian PSF with FWHM of  $0.1''$ . Subsequently, the definition of the lens and source hyperparameters follows. Regarding the mass distribution of the lens, an SIE model is assumed in the center of the grid, at position  $(0, 0)$ , characterized by a velocity dispersion of  $200.0 \text{ km s}^{-1}$ , axial ratio of 0.3 and position angle of 45 deg. Concerning the surface brightness distribution, a Sérsic profile is assumed for both objects: for the lens, an effective radius of  $5.0''$  and a Sérsic index  $n = 3.5$ , while for the source these parameters are set to  $0.3''$  and 4.0. Moreover, the source is placed at position  $(-0.2\theta_E, 0.1\theta_E)$  (with  $\theta_E$  the lens Einstein radius<sup>2</sup>), and is characterized by an axial ratio and a position angle of 0.8 and 60 deg, respectively. We then assume that the source galaxy hosts at its center a point-like source, e.g. a quasar. Finally, it is assumed that the redshifts are  $z_L = 0.3$  for the lens and  $z_S = 1.5$  for the source. A summary of the lens and source parameters is presented in Table 5.3.

By inserting these parameters into the model, it is possible to obtain all characteristics of the lens, but the first step required to simulate the strong lensing event is to solve the lens equation for the assigned position of the source, treated at this stage as point-like.

---

<sup>2</sup> $\theta_E = 25.92 \times 10^{11} \left( \frac{\sigma_v^2 D_{LS}}{c^2 D_S} \right) ''$ .

This allows the positions of the multiple images produced to be assigned to the lens plane.

### 5.2.2 Solving the lens equation

Given the complexity of the model, the best solution is to solve the lens equation numerically. To do this, the method proposed by [Kormann et al. \(1994\)](#) has been adopted to solve the lens equation in the case of SIE models. Starting with the two coordinates of the source (and considering  $(x, \phi)$  as the radial and angular polar coordinates.),

$$y_1 = x_1 - \alpha_1(\vec{x}), \quad (5.2a)$$

$$y_2 = x_2 - \alpha_2(\vec{x}), \quad (5.2b)$$

and multiplying Eq. (5.2a) by  $\cos \phi$  and Eq. (5.2b) by  $\sin \phi$  we obtain

$$y_1 \cos \phi = x_1 \cos \phi - \alpha_1(\vec{x}) \cos \phi = x \cos^2 \phi - \alpha(x, \phi) \cos^2 \phi, \quad (5.3a)$$

$$y_2 \sin \phi = x_2 \sin \phi - \alpha_2(\vec{x}) \sin \phi = x \sin^2 \phi - \alpha(x, \phi) \sin^2 \phi. \quad (5.3b)$$

Recalling that  $\alpha(x, \phi) = \tilde{\Psi}(\phi)$ , the angular part of the lensing potential, Eqs. (5.3a) and (5.3b) can be combined to calculate the distance from the lens center as a function of  $\phi$ :

$$x(\phi) = y_1 \cos \phi + y_2 \sin \phi + \tilde{\Psi}(\phi). \quad (5.4)$$

By reinserting Eq. (5.4) into the lens equation, and setting  $q' = \sqrt{1 - q^2}$ , we obtain

$$F(\phi) = \left[ y_1 + \frac{\sqrt{q}}{q'} \operatorname{arcsinh} \left( \frac{q'}{q} \cos \phi \right) \right] \sin \phi - \left[ y_2 + \frac{\sqrt{q}}{q'} \operatorname{arcsin} (q' \sin \phi) \right] \cos \phi = 0. \quad (5.5)$$

The task of determining the images of a source located at the coordinates  $(y_1, y_2)$  simplifies the identification of the roots of the function  $F(\phi)$ . After determining  $\phi$ , it can be applied to Eq. (5.4) to calculate  $x$ . Analytical solutions are unattainable; thus, a numerical root-finding technique is required.

The algorithm adopted to find the roots of the function  $F$  shares some similarities with Brent's method ([Brent, 1971](#)). In particular, it uses a numerical technique based on the bisection method, which is a type of bracketing method. This method is used to find the roots of a continuous function. The approach is iterative and relies on the intermediate-value theorem. The bisection method provides a reliable and straightforward numerical approach to find roots with a guarantee of convergence,

**Table 5.4.** Summary of the parameters before and after fitting the model using both source plane or image plane optimization.

Parameter	True value	Initial value	Best-fit value	Best-fit value
			(Source plane opt.)	(Image plane opt.)
$\sigma_v$	200.0 km s $^{-1}$	220.0 km s $^{-1}$	200.1128 km s $^{-1}$	200.9709 km s $^{-1}$
$q$	0.3	0.7	0.2990	0.2952
$\varphi$	0.7854 rad	1.0472 rad	0.7855 rad	0.7853 rad

assuming that the initial interval indeed contains a root and the function does not change sign multiple times within the interval.

Thanks to this method, it is therefore possible to solve the lens equation to derive the positions of the multiple images and produce the simulated image of the strong lensing event.

The simulated image of the lens-source system is shown in Fig. 5.6a.

### 5.2.3 Lens optimization

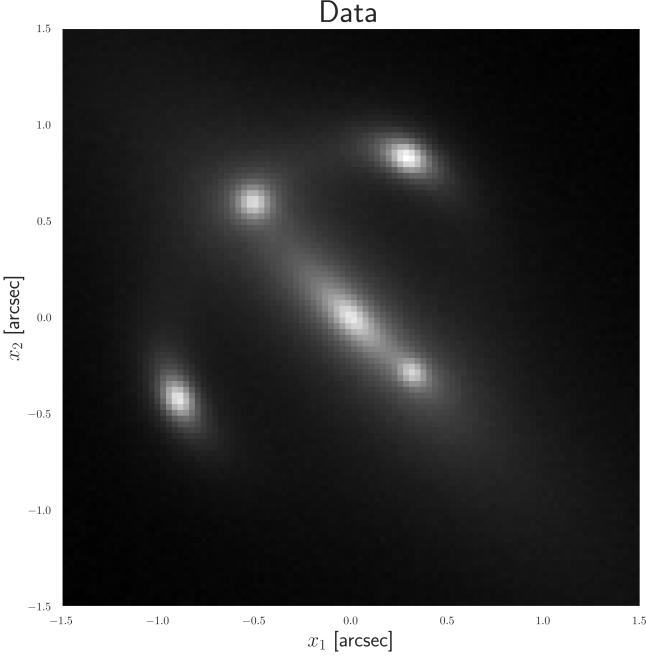
Initially, the model of the lens galaxy is constrained solely by the positions of the quasar’s multiple images. In practical terms, these positions, as determined from astronomical observations, are subject to a degree of imprecision. To simulate positional uncertainty, a minor scatter of 0.015” is introduced to the actual positions of the images.

To start the fitting process, we initialize the parameters of the SIE model to some values (see Table 5.4) and start by optimizing the lens parameter in the source plane.

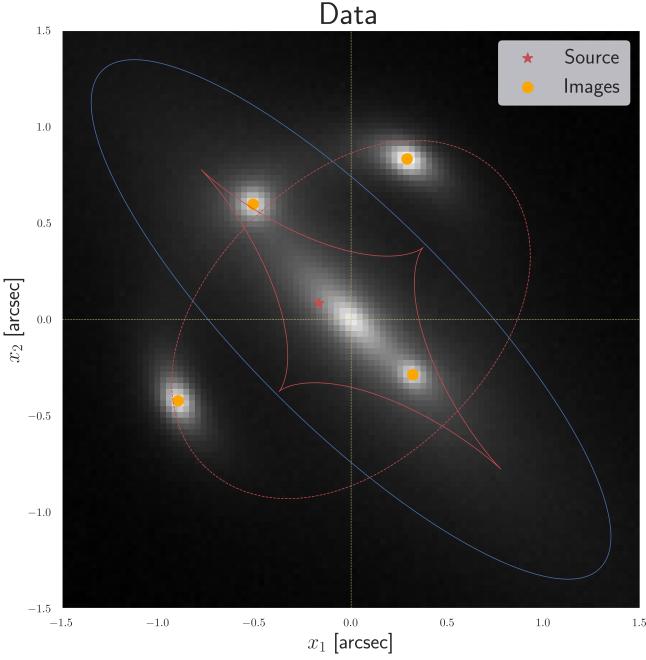
#### Source plane optimization

The first step in the optimization process involves calculating the deflection angle at the positions of the multiple images derived by solving the lens equation. This allows them to be mapped onto the source plane. Since the initial model is generally not accurate, a “source” will be obtained for each of the multiple images of the quasar. It is then assumed that the best estimate for the unlensed position of the quasar is the average position of these sources. The multiple sources obtained and their average position are shown in Fig. 5.7.

Subsequently, it is possible to define a cost function on the source plane that needs to be minimized. This is defined as the sum of the squared differences of the distances of the individual sources predicted from their average position. By minimizing this cost



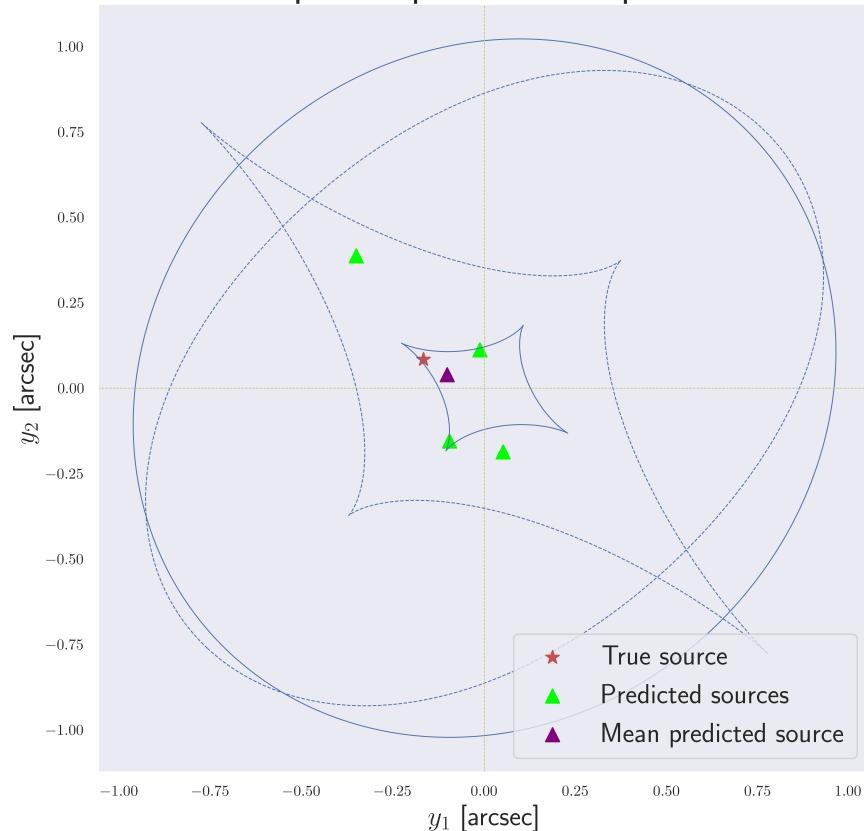
(a)



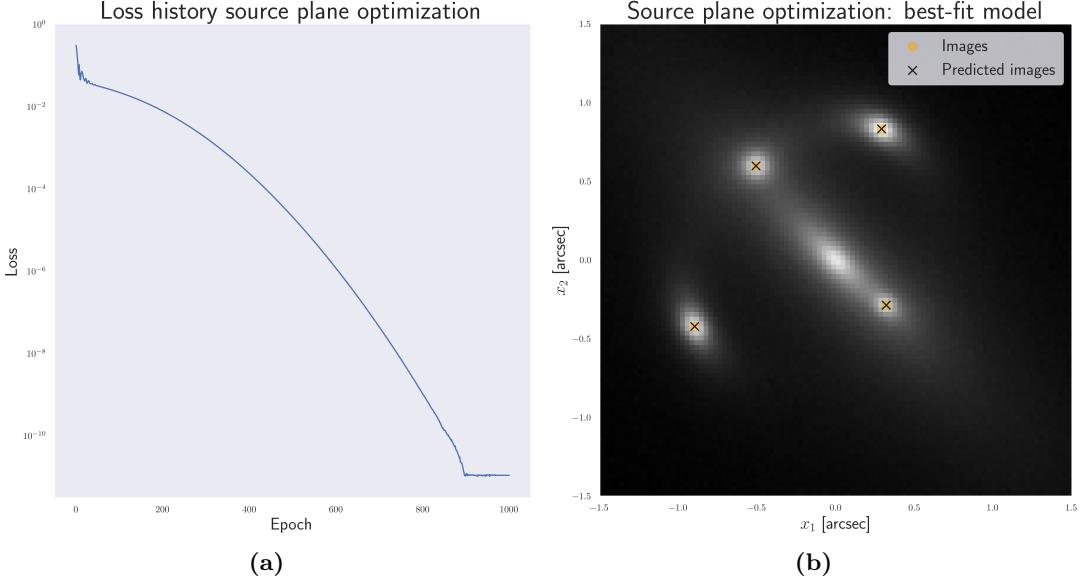
(b)

**Figure 5.6.** (a) Simulated image of a strong lensing event. The pixel scale is  $0.03'' \text{ pix}^{-1}$  and the image side length is  $3''$ . (b) The same simulated image, but with the tangential critical line (solid blue line), the cut (dashed red line) and the tangential caustic (solid red line) of the lens considered shown. The yellow dashed lines identify the lens center. The orange circles represent the positions of the multiple images obtained by solving the lens equation with the method discussed in Section 5.2.2.

### Source plane optimization: predictions



**Figure 5.7.** The multiple images of the source are mapped back to the source plane to multiple predicted sources (green triangles). Their mean position (purple triangle) is indicated. The solid blue curves represent the caustic and the cut of the lens model, respectively. The dashed blue curves give the true caustic and cut.



**Figure 5.8.** (a) Loss history for source plane optimization and (b) best-fit image positions for the source plane optimization.

function, it is possible to optimize the model parameters so that the multiple sources converge to a single position.

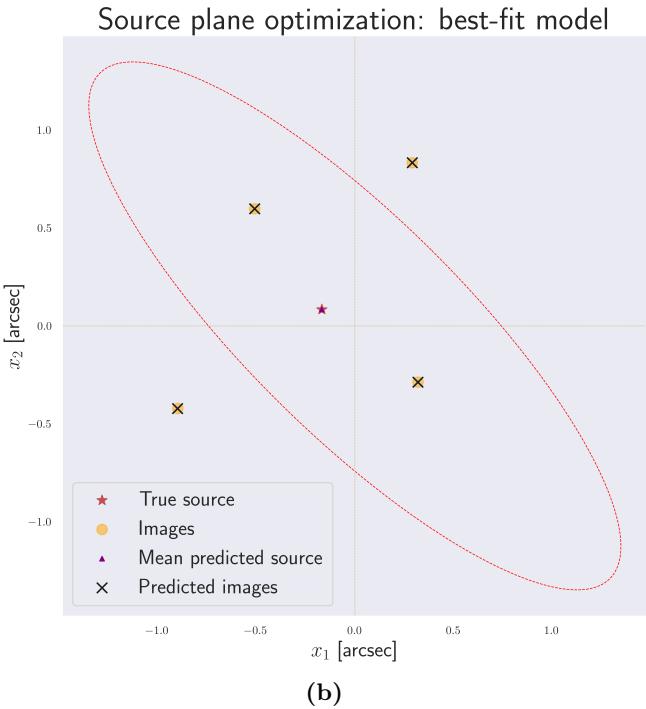
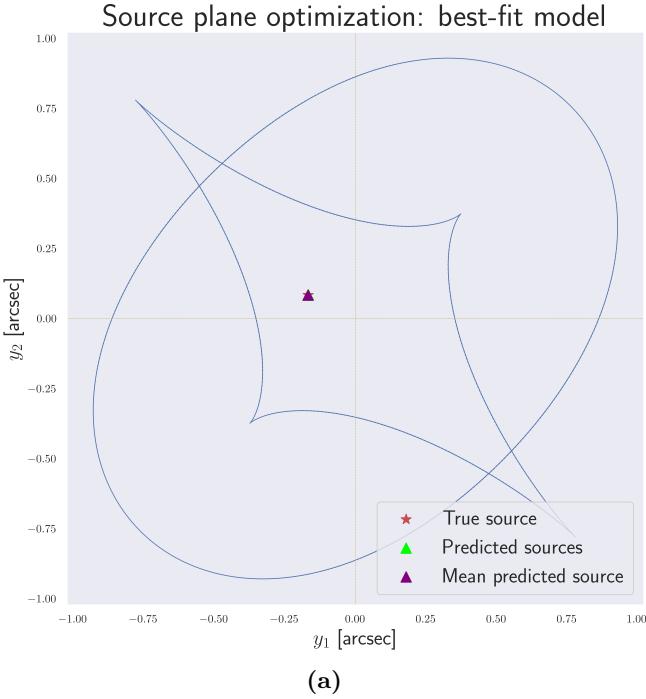
This is feasible, as always, through a training loop. After defining Adam as optimizer, training is started to minimize the cost function, which decreases from a starting value of  $3.1 \cdot 10^{-1}$  to a final value of  $1.1 \cdot 10^{-11}$  (see Fig. 5.8a).

In Table 5.4 the best-fit results are reported. Furthermore, as can be seen in Figs. 5.8b and 5.9, in both the source plane and the lens plane, the positions of the source and multiple images overlap. Additionally, from Figs. 5.9a and 5.9b a clear coincidence of the caustic, cut, and critical line is visible.

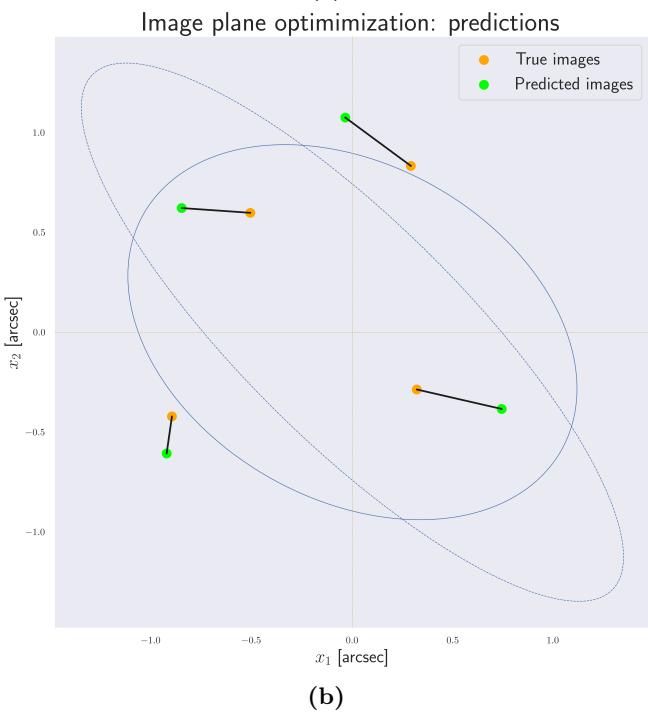
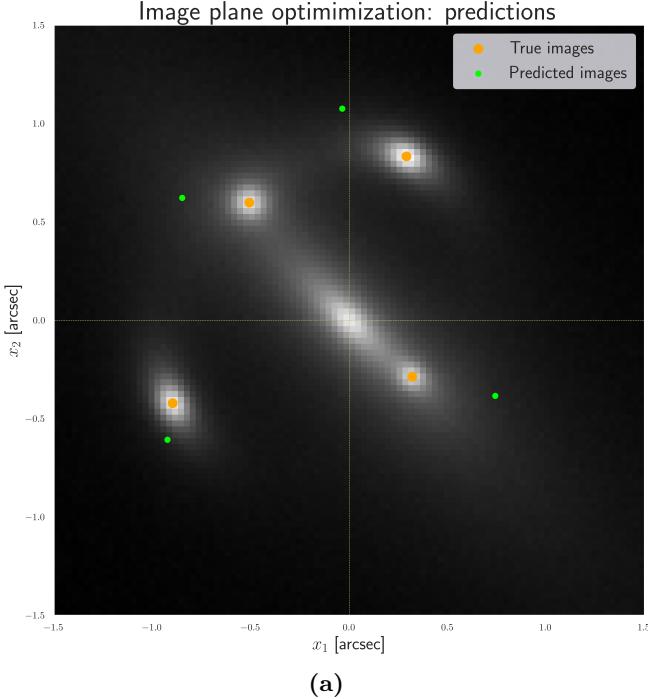
### Image plane optimization

Regarding the optimization on the lens plane, the first part of the process is similar: the positions of the multiple images are mapped onto the source plane and their average position is calculated. However, unlike the process discussed previously, it is now necessary to remap the average position of the sources onto the lens plane by solving the lens equation. This allows the model-predicted positions of the multiple images to be obtained. These are shown in Fig. 5.10a.

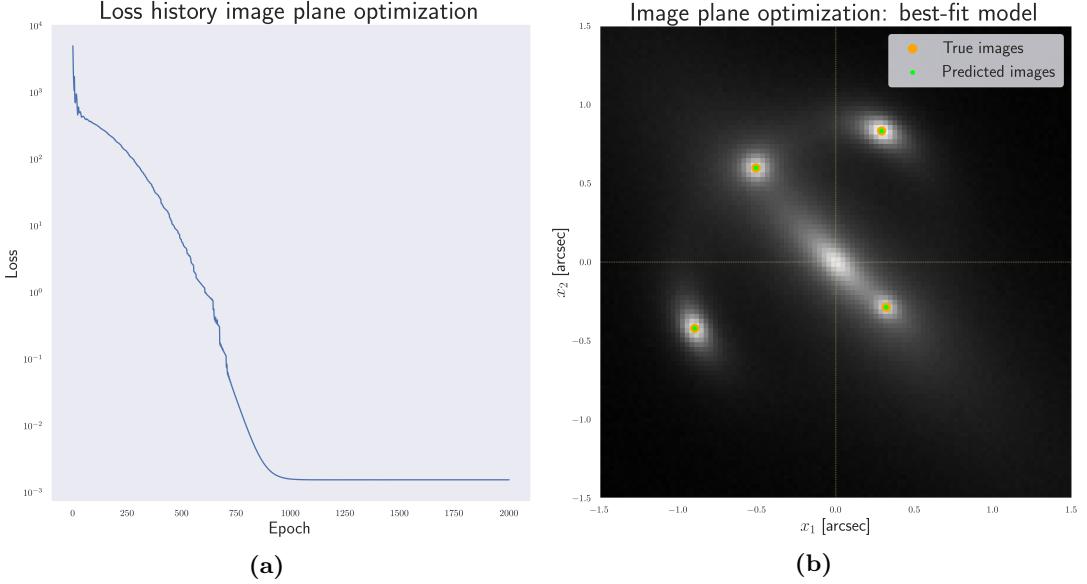
The process then proceeds as follows: for each observed image of the quasar, the closest predicted image from the model is identified. Then, the distance between each



**Figure 5.9.** Best-fit model for the source plane optimization, visualized both (a) in the source plane and (b) in the lens plane. On the source and lens planes are indicated the caustic, cut and critical line of the best-fit model and the true ones (which overlap).



**Figure 5.10.** (a) Multiple images predicted by the model and true images superimposed on the light distribution of the image plane. (b) Illustration in the lens plane of the multiple images predicted and observed, where each closest pair has been identified. Each black stick shows the distances between each quasar image and the closest model-predicted one. The tangential critical line of the model is represented by a solid blue curve, while the true one is shown as a dashed blue curve.



**Figure 5.11.** (a) Loss history for image plane optimization and (b) best-fit image positions for the image plane optimization.

closest pair of predicted and observed images is measured. Such distances are shown as black sticks in Fig. 5.10b.

Once the closest pairs of images have been identified and their distance calculated, all that remains is to explore the parameter space in search of the combination of  $\sigma_v$ ,  $q$ , and  $\varphi$  that minimizes this distance, in order to make every image predicted by the model converge to the real image produced by the observed system. The cost function implemented to quantify this discrepancy between images is a  $\chi^2$  in the lens plane that measures the sum of the absolute distances between each pair of images, divided by the observational error on each image position, as described in Section 4.3.1. With these settings, the minimization procedure starts with an initial loss value  $\chi^2 = 5.05 \cdot 10^3$  and ends with a value of  $\chi^2 = 1.56 \cdot 10^{-3}$ , as reported in Fig. 5.11a. The values for the best-fit parameters are reported in Table 5.4. With these parameters, the model-predicted image positions match the observed image positions very well, as can be seen in Fig. 5.11b.

Once the model has been fitted to the observational data and the best-fit parameters have been derived, these can be used to obtain information about the mass distribution of the lens.

In our particular case, the best-fit velocity dispersion results in an Einstein radius

$$\theta_E = 25.92 \times 10^{11} \left( \frac{\sigma_v^2 D_{LS}}{c^2 D_S} \right)'' = 0.8379'' = 4.06 \times 10^{-6} \text{ rad}, \quad (5.6)$$

which, knowing the redshift  $z_L = 0.3$  of the lens and so its angular diameter distance  $D_L = 9.5 \times 10^5$  kpc, can be converted into physical units:

$$R_E = \theta_E D_L = 3.86 \text{ kpc}. \quad (5.7)$$

Subsequently, we can compute the critical surface density  $\Sigma_{cr}$  from Eq. (2.18) as

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}} = 2.41 \times 10^9 M_\odot \text{ kpc}^{-2}, \quad (5.8)$$

allowing us to estimate the total mass enclosed inside the Einstein radius of the lens

$$M(< R_E) = \pi R_E^2 \Sigma_{cr} = 1.13 \times 10^{11} M_\odot. \quad (5.9)$$

### 5.3 Surface brightness and galaxy shapes

This section focuses on the significance of fitting surface brightness models to simulated/observational data within the context of weak lensing. Accurate modeling of surface brightness profiles is of utmost importance for the analysis of weak lensing effects.

Fitting surface brightness profiles is a crucial step in weak lensing studies for several reasons. First, it allows astronomers to precisely measure the shapes and orientations of galaxies, which can be subtly distorted by the gravitational lensing effect of foreground mass distributions. These distortions, although weak and difficult to detect, carry invaluable information about the mass distribution of the lensing objects, including dark matter, which is otherwise invisible. By analyzing the statistical alignment and shape changes of background galaxies, researchers can map the mass distribution of the lensing structures, offering insights into the nature of dark matter and the large-scale structure of the Universe.

Moreover, accurate fitting of surface brightness profiles helps in separating the lensing signal from intrinsic alignments and other observational effects, enhancing the reliability of weak lensing measurements. This process is fundamental in cosmology as it contributes to the precision modeling of the Universe's expansion and the distribution of matter on cosmic scales. Thus, the methodological advancements in fitting these profiles not only refine our understanding of individual galaxy structures but also bolster the broader applications of weak lensing as a tool for probing the fundamental components and dynamics of the cosmos.

By fitting observed surface brightness profiles with theoretical models, the ellipticity and orientation of galaxy images can be precisely determined, allowing for the recon-

struction of the lensing field and the mapping of the mass distribution along the line of sight.

### 5.3.1 Fit of Sérsic profile of surface brightness

The simulated galaxy is generated from a theoretical model, incorporating noise and applying a Point Spread Function (PSF), assuming that the sky background has already been subtracted from the image.

Incorporating noise into these simulations mimics the real observational conditions faced in astronomy. Sources of noise include detection equipment, atmospheric interference in ground-based observations, and intrinsic astronomical sources of variability. By integrating noise into the simulations, the fitting algorithms are tested for robustness, ensuring they can navigate the real-world data complexities.

Additionally, applying a Point Spread Function (PSF) is crucial. The PSF models the imaging system's response, capturing the effects of the telescope, detectors, and atmosphere that may blur and distort observed images. This is particularly relevant in the context of weak lensing because the PSF anisotropies can mimic lensing-induced ellipticities, biasing the shear measurements. Simulating these effects allows the fitting algorithms to consider them when analyzing objects' surface brightness.

We assume that the sky background has been subtracted from the image. This step isolates the light from the astronomical object from ambient environmental light, including atmospheric light, scattered light from nearby stars, and the Milky Way's diffuse glow. Precise background subtraction is essential for accurate surface brightness measurements.

The most common type of parametric surface brightness model is the Sérsic profile, which, as described in Section 4.2.1, is characterized by an exponential form. This is the usual profile adopted to represent the luminosity distribution of various astronomical objects, such as galaxies, by providing a flexible mathematical framework that describes a wide range of shapes from disk-like to elliptical structures.

The Sérsic profile is expressed mathematically as follows:

$$I(R) = I_e \exp \left\{ -b_n \left[ \left( \frac{R}{R_e} \right)^{\frac{1}{n}} - 1 \right] \right\}. \quad (5.10)$$

Assuming  $q$  is the axis ratio of the elliptical distribution, the elliptical radius  $R$  is thus

$$R = \sqrt{x_r^2 + (y_r/q)^2}, \quad (5.11)$$

where  $x_r$  and  $y_r$  the rotated coordinates obtained by aligning the reference frame  $(x, y)$  of the observer to the axes of the ellipse

$$\begin{aligned} x_r &= (x - x_0) \cos \varphi + (y - y_0) \sin \varphi, \\ y_r &= (y - y_0) \cos \varphi - (x - x_0) \sin \varphi, \end{aligned} \quad (5.12)$$

where  $x_0$  and  $y_0$  represent the coordinates of the center of the ellipse and  $\varphi$  the position angle.

Optimizing the Sérsic profile means fitting its seven parameters ( $I_e$ ,  $R_e$ ,  $n$ ,  $x_0$ ,  $y_0$ ,  $q$ ,  $\varphi$ ) to observational data, which requires careful consideration of the effects of instrumental and observational factors, such as the Point Spread Function (PSF).

### Simulation of observational data

The first step consists of simulating the observational data. It is done by constructing a  $(x, y)$  grid on the sky plane with dimensions  $100'' \times 100''$ , with a pixel scale of  $0.03'' \text{ pix}^{-1}$ .

The model for simulating the galaxy light profile and the parameters needed to create the mock data are then defined. The intensity at the effective radius,  $I_e$ , is expressed in arbitrary units, while the effective radius  $R_e$  is expressed in arcseconds, together with the center coordinates. Furthermore, the axis ratio  $q$  and position angle  $\varphi$  of the galaxy are defined. They are linked to the ellipticity components  $e_1$  and  $e_2$  by the following relations

$$\begin{aligned} e_1 &= \frac{1-q}{1+q} \cos(2\varphi), \\ e_2 &= \frac{1-q}{1+q} \sin(2\varphi), \end{aligned} \quad (5.13)$$

and their combination gives the ellipticity of the object

$$e = \sqrt{e_1^2 + e_2^2} = \frac{1-q}{1+q}. \quad (5.14)$$

The ellipticity is related to the semi-major and semi-minor axes, which is reflected in terms of shear. In particular, it can be demonstrated that the ellipticity coincides with the reduced shear  $g$ :

$$e \equiv g = \frac{\gamma}{1-\kappa}. \quad (5.15)$$

This is the concept at the basis of weak lensing: by measuring the ellipticity of the lensed images of background sources, we can infer the lenses' mass distribution. Given that ellipticity measures shear and convergence, it is then related to the second derivatives of the lensing potential.

We can then proceed with inserting the parameters inside the model. Table 5.5 report a summary of the parameters to which the model is initialized.

When the parameters are added to the model, the theoretical observational data is produced and is shown in Fig. 5.12a. Additionally, in Fig. 5.13, the theoretical Sérsic profile is shown: the surface brightness as a function of distance from the galaxy center, in units of  $R_e$ . Subsequently, to enhance the realism of the observational model, a Gaussian noise component with mean  $\mu = 0.0$  and standard deviation  $\sigma = 0.05$  was added to the image. Furthermore, the resulting surface brightness was convolved with a Point Spread Function (PSF) characterized by a Full Width at Half Maximum (FWHM) of  $0.1''$ . The final “noisy” surface brightness profile is illustrated in Fig. 5.12b.

### Fitting the model to data

The next step involves fitting the model to the observational data. In this context, we establish a cost function and an optimizer, and we proceed by utilizing automatic differentiation frameworks such as PyTorch and TensorFlow. As discussed in Chapter 4, these components are tasked with quantifying the discrepancy between the theoretical model and the observed data and updating the model parameters by gradient descent, respectively. For this task, one of the most commonly used cost functions in regression problems, the Mean Squared Error<sup>3,4</sup> (MSE), was selected. The MSE is advantageous for its simplicity and effectiveness in quantifying the average squared discrepancy between the predicted and actual values. This makes it particularly suitable for regression tasks that aim to minimize the difference between predicted results and observed data (Berger, 2006).

Regarding the optimizer, Adam<sup>5,6</sup> (Kingma & Ba, 2017) was chosen for its ability to combine the advantages of two other popular optimizers: AdaGrad and RMSProp (Tieleman & Hinton, 2014). Adam stands out for its adaptive learning rate, which adjusts as training progresses, making it efficient for large-scale and complex problems. This adaptability is particularly beneficial for navigating complex high-dimensional data landscapes, often leading to faster convergence than other gradient descent methods.

In addition, a learning rate scheduler has been defined to improve the efficiency of model training by automatically adjusting the step  $\eta$  taken by the optimizer in parameter updates<sup>7,8</sup>. The initial learning rate for the optimizer is set to  $\eta = 0.1$  and

---

<sup>3</sup><https://pytorch.org/docs/stable/generated/torch.nn.MSELoss.html>.

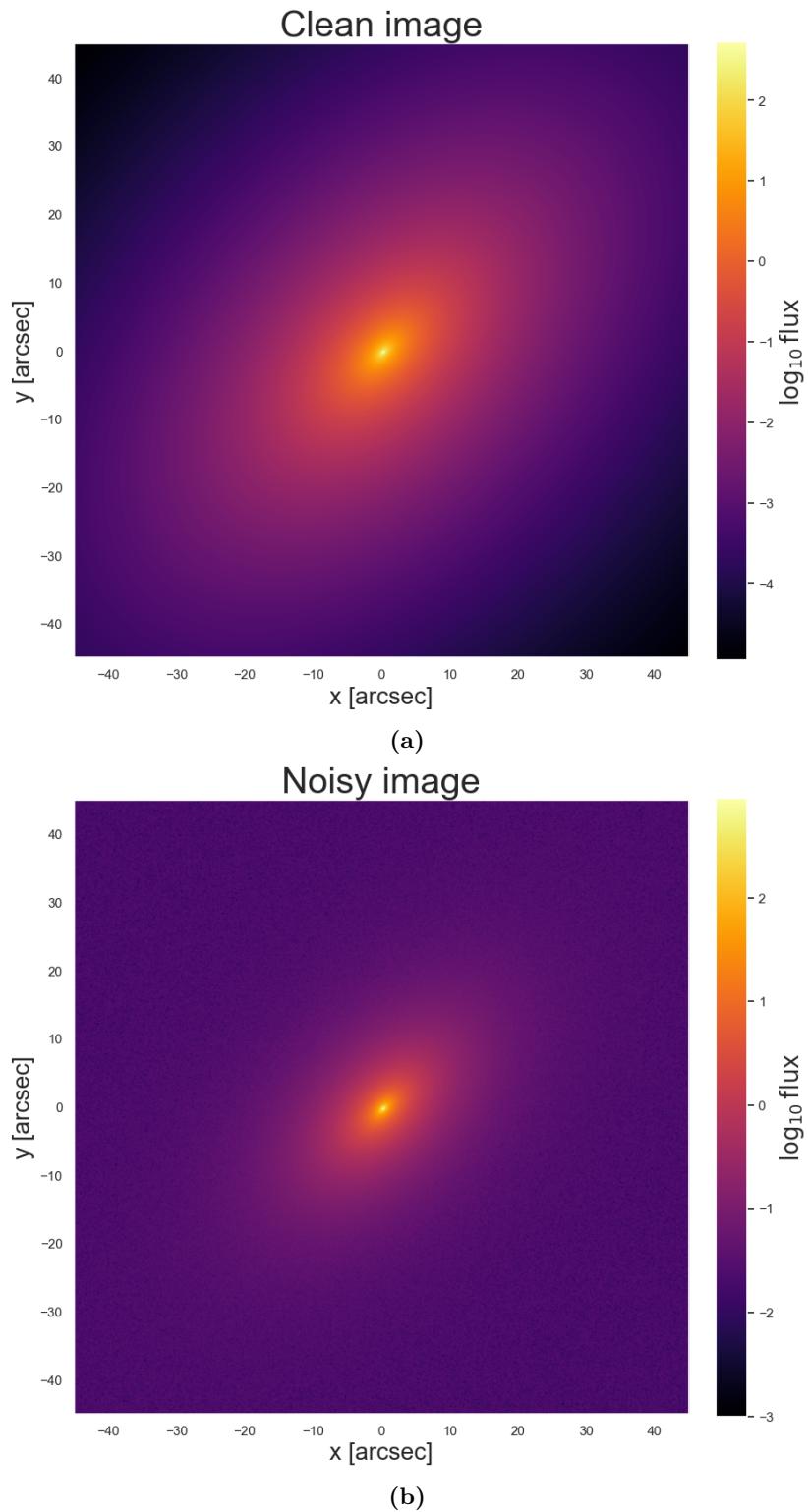
<sup>4</sup>[https://www.tensorflow.org/api\\_docs/python/tf/keras/losses/MeanSquaredError](https://www.tensorflow.org/api_docs/python/tf/keras/losses/MeanSquaredError).

<sup>5</sup><https://pytorch.org/docs/stable/generated/torch.optim.Adam.html>.

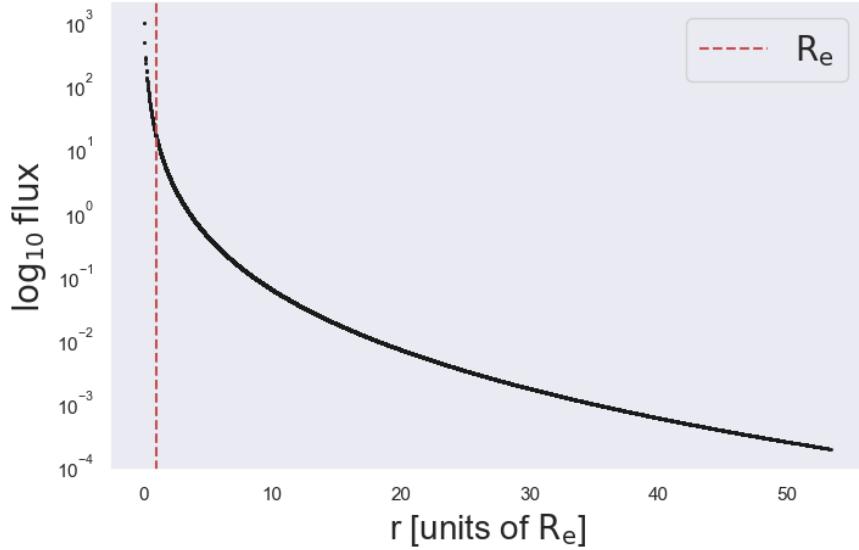
<sup>6</sup>[https://www.tensorflow.org/api\\_docs/python/tf/keras/optimizers/Adam](https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam).

<sup>7</sup>[https://pytorch.org/docs/stable/generated/torch.optim.lr\\_scheduler.ReduceLROnPlateau.html](https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html).

<sup>8</sup>[https://www.tensorflow.org/api\\_docs/python/tf/keras/callbacks/ReduceLROnPlateau](https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ReduceLROnPlateau).



**Figure 5.12.** Simulation of a galaxy with Sérsic surface brightness distribution: (a) theoretical model and (b) image with Gaussian noise+PSF.



**Figure 5.13.** Simulated Sérsic surface brightness distribution as a function of distance from the center of the galaxy, in units of the effective radius.

gradually decreases to improve the speed of finding the minimum loss function.

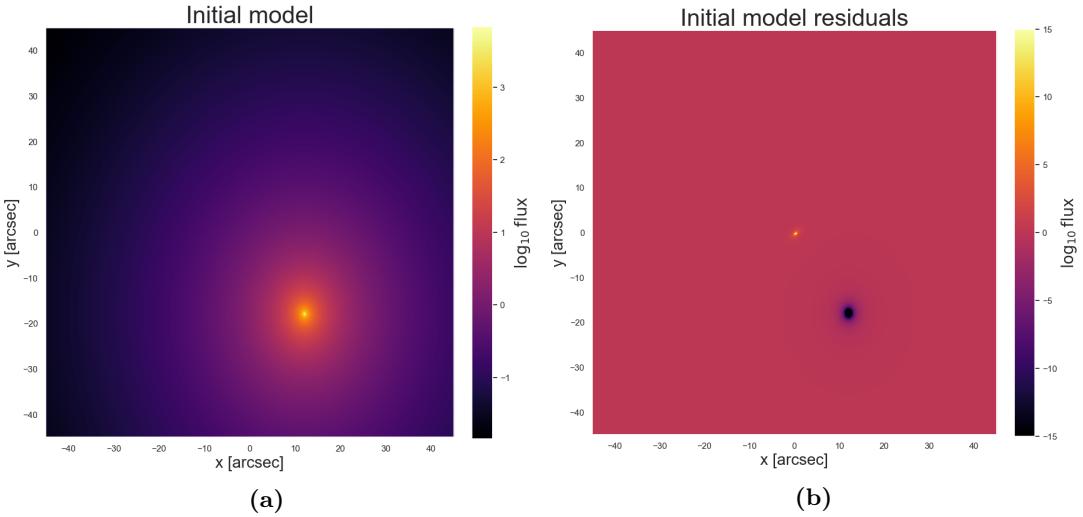
At this point, the model must be initialized with some initial (random) parameters that are appropriately constrained, shown in Table 5.5.

The initial prediction of the model, accompanied by residuals for the actual data, is presented in Fig. 5.14. Subsequently, the model training can start. To speed up the fitting process, it is possible to minimize the cost function computed on the logarithms of the model and data as a way to normalize the data between a smaller range of values (Huang et al., 2020).

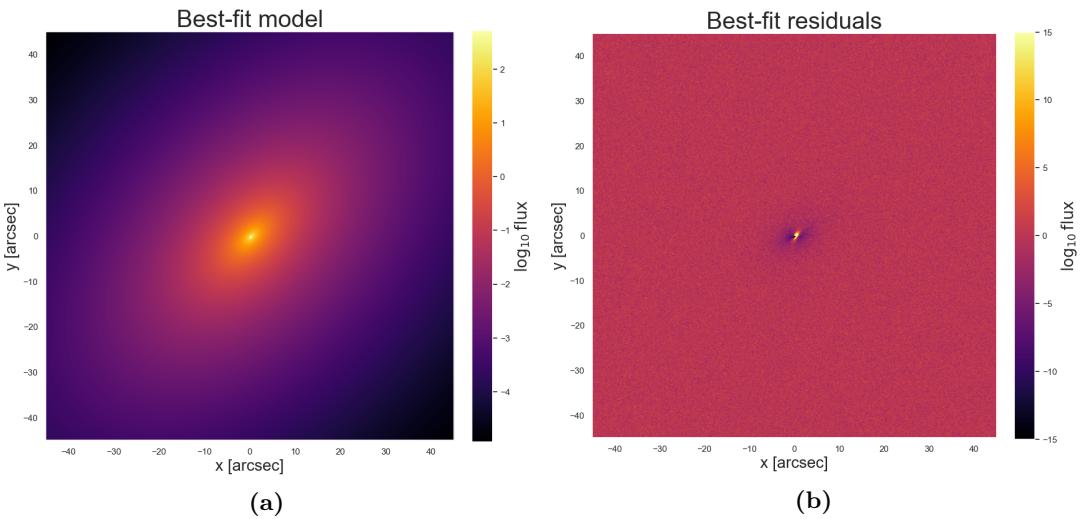
Before training the cost function, i.e. the MSE between the data and the model has

**Table 5.5.** Summary of true and initial parameters before fitting.

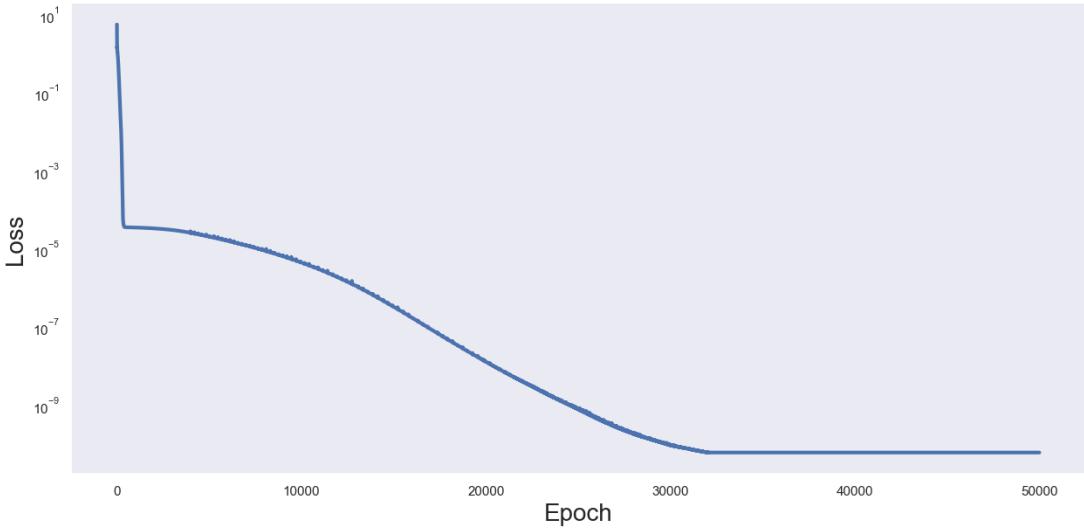
Parameter	True value	Initial value
$I_e$	15.0	25.0
$R_e$	2.0''	4.5''
$n$	5.5	7.0
$x_0$	0.2''	12.2''
$y_0$	-0.3''	-18.0''
$q$	0.6	0.8
$\varphi$	0.7853 rad	1.5708 rad



**Figure 5.14.** (a) Initial prediction of Sérsic surface brightness model, computed using the initial values of Table 5.5 and (b) residuals, normalized by their standard deviation, calculated as subtraction between true data and initial prediction.



**Figure 5.15.** (a) Best-fit prediction of Sérsic surface brightness model, computed using the best-fit values of Table 5.6 obtained from the training process and (b) residuals normalized by their standard deviation, calculated as subtraction between true data and best-fit prediction.



**Figure 5.16.** Loss function vs epochs for the training of the Sérsic model.

a value of  $\text{loss}_{\text{initial}} = 5.4$  and, after 50000 iterations (or epochs) of the training loop, it drops to  $\text{loss}_{\text{final}} = 3.1 \cdot 10^{-10}$ , reaching its absolute minimum. The best-fit parameters are reported in Table 5.6 and Fig. 5.16 shows the trend of the MSE as a function of iterations of gradient descent.

After training the model, an estimation of parameter errors was performed throughout MCMC sampling, using the Pyroframework. Pyro is a probabilistic programming language built on top of PyTorch, designed to create and run complex probabilistic models. It offers tools for defining flexible and expressive Bayesian models and algorithms for variational inference, making it well-suited for tasks in machine learning and statistics that involve uncertainty. Pyro’s integration with PyTorch allows for automatic differentiation and GPU acceleration, enabling efficient and scalable model fitting and evaluation.

The probabilistic model was defined with a uniform distribution for the parameters, and then 5000 samples from the posterior distributions were extracted. Their mean values and standard deviations are described in Table 5.6.

Figure 5.17 shows the corner plot to illustrate the multidimensional distributions of and between the parameters. Diagonal elements consist of histograms representing the marginal distributions of individual parameters. They allow one to discern the central tendencies (such as mean or median), the spread (variance or standard deviation), and the shape (indications of skewness or multiple peaks) of each parameter’s distribution. Off-diagonal elements display the joint distributions between pairs of parameters through

**Table 5.6.** Summary of true, initial and best-fit parameters.

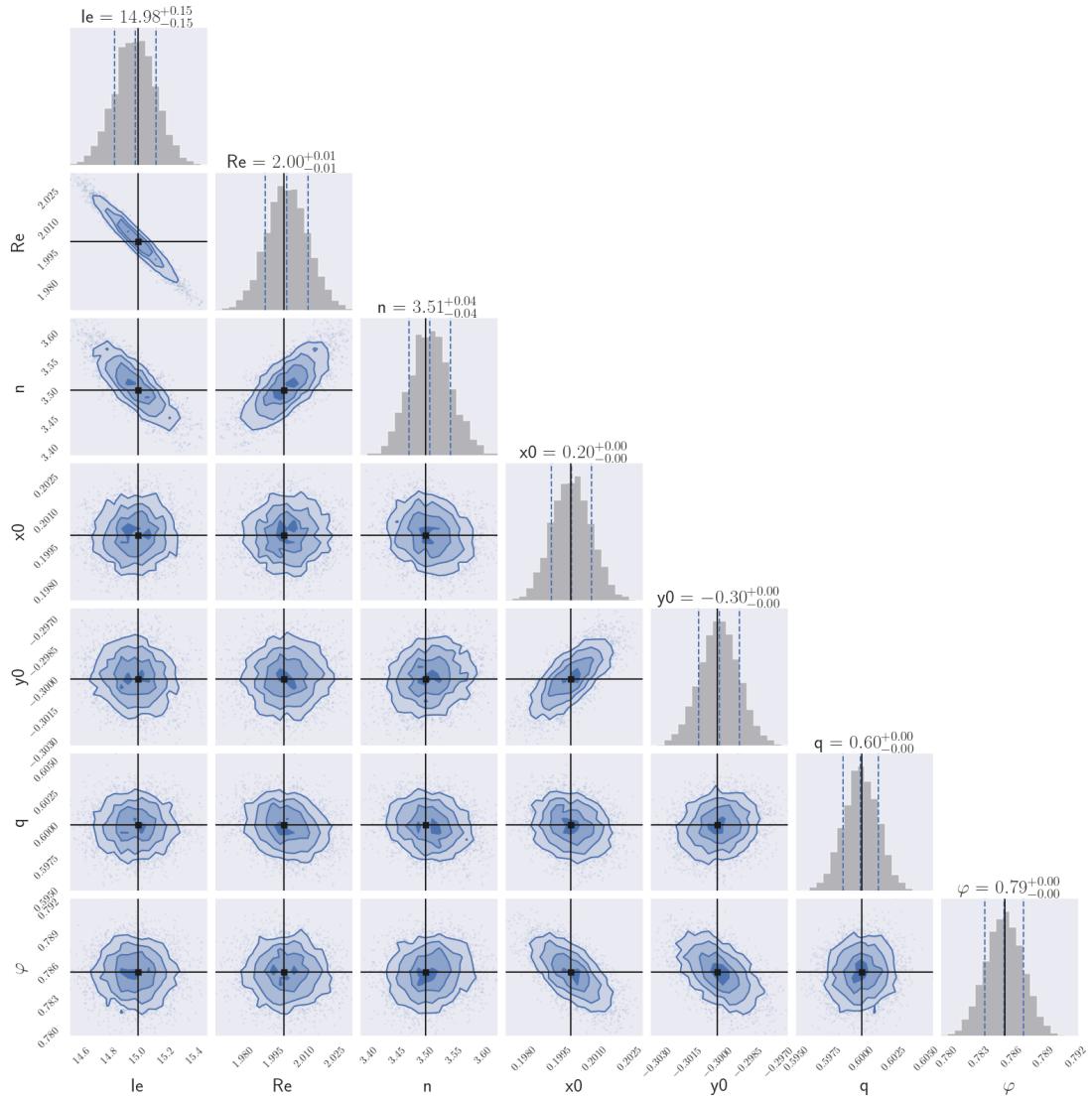
Parameter	True value	Initial value	Best-fit value	Best-fit value (MCMC)
$I_e$	15.0	25.0	15.0077	(14.9777 ± 0.1499)''
$R_e$	2.0''	4.5''	1.9997''	(2.0014 ± 0.0106)''
$n$	3.5	7.0	3.5002	(3.5080 ± 0.0376)
$x_0$	0.2''	12.2''	0.2000''	(0.2000 ± 0.0009)''
$y_0$	-0.3''	-18.0''	-0.3000''	(-0.2999 ± 0.0010)''
$q$	0.6	0.8	0.5999	(0.5999 ± 0.0014)
$\varphi$	0.7853 rad	1.5708 rad	0.7854 rad	(0.7854 ± 0.0018) rad

contour plots. They are instrumental in understanding the correlations or dependencies between parameters, with circular patterns indicating minimal correlation and elongated ellipsoids suggesting strong correlations. The orientation of these ellipsoids further clarifies the nature of the relationship, be it positive or negative.

In particular, the corner plot shows the correlation between some parameters involved. As is known from the literature (Graham & Driver, 2005; Ciotti & Bertin, 1999), the Sérsic profile presents degeneracies between some of its parameters:

- intensity  $I_e$  and effective radius  $R_e$ : larger radius with lower intensity can produce a similar overall brightness as a smaller radius with higher intensity;
- Sérsic index  $n$  and effective radius  $R_e$ : galaxies with higher Sérsic indices tend to also have larger effective radii, as both parameters are related to how light is distributed in a galaxy.

Additionally, Trujillo et al. (2001a,b) showed that in addition to the intrinsic degeneracies among the Sérsic parameters, there are some further degeneracies between the latter and the PSF. One can think of the intrinsic galaxy axis ratio as a very intuitive example. The PSF is, by definition, rounder than a flattened galaxy; hence, its effect on the two-dimensional light distribution produces a rounder galaxy. The higher the FWHM in arcsecs, the rounder the observed galaxy is (Peng et al., 2002; Li et al., 2022).



**Figure 5.17.** Corner plot showing the posterior probability distributions of the parameters used to fit the Sérsic surface brightness profile. The projections of the probability density on planes defined by each couple of parameters are shown, as well as one-dimensional marginalized distributions. The black and the blue dashed vertical lines in one-dimensional histograms indicate the true values and the 16th, 50th and 84th percentiles of the distributions.

# 6

## Summary and conclusions

In this thesis, we have systematically explored the application of automatic differentiation (AD) and differentiable programming methods, utilizing the computational capabilities of PyTorch and TensorFlow, to address various aspects of gravitational lensing. This exploration was aimed at enhancing the optimization of parametric functions and models, which are critical for accurately modeling gravitational lensing phenomena. This concluding chapter aims to concisely summarize this work, emphasizing the integration of AD in gravitational lensing and detailing the specific topics addressed.

Automatic differentiation has emerged as a pivotal tool in our investigation, offering significant advantages over traditional numerical differentiation methods in terms of both accuracy and efficiency. AD's capability to compute derivatives of complex functions accurately and efficiently is particularly valuable in gravitational lensing, where precise parameter estimation is crucial. By leveraging the computational frameworks of PyTorch and TensorFlow, we have demonstrated the potential of AD to advance the analysis and interpretation of gravitational lensing phenomena, facilitating the development of more detailed and nuanced models.

In the first application, we tackled the challenge of microlensing light curve fitting by leveraging the standard Paczynski curve, focusing on the impact of key parameters such as the impact parameter, the maximum magnification time, and the Einstein radius crossing time on the microlensing event profile. Our approach began with simulating a microlensing event and introducing observational noise to mimic real data. Initial model fitting, using randomly assigned parameters, demonstrated the critical issue of degeneracy in microlensing analysis, where multiple parameter sets produce similar light curves, complicating parameter estimation.

By adjusting our model to explicitly include the Einstein crossing time, we improved the fit, illustrating the effectiveness of our method in mitigating degeneracy. In addition, Bayesian sampling provided deeper insight into the parameter space, confirming the degeneracy.

Secondly, we focused on parametric reconstruction of strong gravitational lensing using the Singular Isothermal Ellipsoid (SIE) model to analyze the mass distribution of lensing galaxies. Our methodology utilized the SIE model's parameters to simulate lensing effects and optimize these parameters based on the positions of multiple lensed

images observed in astronomical data. We employed two optimization approaches: one that adjusted parameters based on their effects in the source plane and another that did so in the image plane, each aiming to reduce discrepancies between observed and predicted lensed images.

Through simulations that incorporate realistic observational conditions, we generated and optimized synthetic lensing scenarios. This process involved a detailed parameterization of both the lens and the source, which solved the lens equation numerically to predict the lensed image positions accurately. The optimization demonstrated the effectiveness of our methodology in aligning model predictions with observed data, as evidenced by significant improvements in model accuracy.

The results showcased the precision of the SIE model in replicating observed lensing phenomena, with optimized model parameters closely matching their intended values.

The final field of application focused on surface brightness and galaxy shapes: we have explored the critical role of accurately modeling surface brightness profiles for weak lensing analysis. This process is vital for measuring the shapes and orientations of galaxies, which are subtly altered by the gravitational lensing effects of foreground mass distributions. These alterations provide essential clues about the mass distribution of lensing objects, including elusive dark matter, and allow researchers to map the mass distribution of lensing structures, offering insights into dark matter and the universe's large-scale structure.

Our analysis focused on fitting a surface brightness profile using the Sérsic model, a versatile representation for various galaxy types. The fitting process involved simulating observational data with realistic noise and applying a Point Spread Function (PSF) to account for observational distortions. This step is crucial for separating the lensing signal from intrinsic galaxy alignments and observational effects, thereby enhancing the reliability of weak lensing measurements.

Through the simulation of a galaxy light profile, incorporating noise and PSF effects, we aimed to mimic the challenges encountered in real astronomical observations. The Sérsic profile was parameterized and fitted to this simulated data, optimizing its parameters to match the observed galaxy images. This process, facilitated by frameworks such as PyTorch and TensorFlow, involved the definition of a cost function and the employing of an optimizer for parameter updates, demonstrating practical applications of automatic differentiation.

This process highlights the importance of methodological advancements in fitting surface brightness profiles, not only for understanding individual galaxy structures, but also for leveraging weak lensing as a tool for probing the fundamental components and

dynamics of the cosmos.

Looking ahead, AD promises to revolutionize gravitational lensing studies further by facilitating the development of sophisticated models for complex phenomena and enabling efficient analysis of vast astronomical datasets. Its integration into time-delay cosmography and non-parametric mass reconstruction could yield deeper insights into the universe's expansion and structure. Moreover, AD's role in processing data from upcoming large-scale surveys will be pivotal in uncovering new discoveries and enhancing our cosmological understanding.

In summary, AD's contribution to gravitational lensing fields is profound, offering a pathway to novel discoveries and a deeper understanding of the cosmos. Its continued development and application stand to unlock even greater potentials in astrophysical research.



# Bibliography

- Abadi, M., Barham, P., Chen, J., et al. 2016, TensorFlow: A system for large-scale machine learning, arXiv. <http://arxiv.org/abs/1605.08695>
- Alcock, C., Akerlof, C. W., Allsman, R. A., et al. 1993, Nature, 365, 621, doi: [10.1038/365621a0](https://doi.org/10.1038/365621a0)
- Amini, A., Amini, A., Karaman, S., & Rus, D. 2018, Spatial Uncertainty Sampling for End-to-End Control
- Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, *\apj*, 724, 511, doi: [10.1088/0004-637X/724/1/511](https://doi.org/10.1088/0004-637X/724/1/511)
- Bandara, K., Crampton, D., Peng, C., & Simard, L. 2013, *The Astrophysical Journal*, 777, 1, doi: [10.1088/0004-637X/777/1/1](https://doi.org/10.1088/0004-637X/777/1/1)
- Bauer, F. L. 1974, *SIAM Journal on Numerical Analysis*, 11, 87, doi: [10.1137/0711010](https://doi.org/10.1137/0711010)
- Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. 2018, Automatic differentiation in machine learning: a survey, arXiv. <http://arxiv.org/abs/1502.05767>
- Berger, J. O. 2006, Statistical decision theory and Bayesian analysis, 2nd edn., Springer series in statistics (New York: Springer)
- Bingham, E., Chen, J. P., Jankowiak, M., et al. 2018, Pyro: Deep Universal Probabilistic Programming
- Birrer, S., Amara, A., & Refregier, A. 2015, *The Astrophysical Journal*, 813, 102, doi: [10.1088/0004-637X/813/2/102](https://doi.org/10.1088/0004-637X/813/2/102)
- Blandford, R., Surpi, G., & Kundic, T. 2000, Modeling Galaxy Lenses, arXiv. <http://arxiv.org/abs/astro-ph/0001496>

- Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, The Astrophysical Journal, 682, 964, doi: [10.1086/589327](https://doi.org/10.1086/589327)
- Bottou, L., Curtis, F. E., & Nocedal, J. 2018, SIAM Review, 60, 223, doi: [10.1137/16M1080173](https://doi.org/10.1137/16M1080173)
- Bovik, A. C., ed. 2005, Handbook of image and video processing, 2nd edn. (Amsterdam ; Boston, MA: Elsevier Academic Press)
- Brent, R. P. 1971, The Computer Journal, 14, 422, doi: [10.1093/comjnl/14.4.422](https://doi.org/10.1093/comjnl/14.4.422)
- Buchner, J. 2023, Statistics Surveys, 17, doi: [10.1214/23-SS144](https://doi.org/10.1214/23-SS144)
- Burden, R. L., Faires, J. D., & Burden, A. M. 2016, Numerical analysis, tenth edition edn. (Boston, MA: Cengage Learning)
- Burke, C. 2013, PhD Thesis
- Burke, W. L. 1981, The Astrophysical Journal, 244, L1, doi: [10.1086/183466](https://doi.org/10.1086/183466)
- Canu, S., Flamaury, R., & Mary, D. 2016, EAS Publications Series, 78-79, 127, doi: [10.1051/eas/1678007](https://doi.org/10.1051/eas/1678007)
- Cardone, V. F., Capozziello, S., Re, V., & Piedipalumbo, E. 2001, Astronomy & Astrophysics, 379, 72, doi: [10.1051/0004-6361:20011225](https://doi.org/10.1051/0004-6361:20011225)
- Cavanaugh, J. E., & Neath, A. A. 2019, WIREs Computational Statistics, 11, e1460, doi: [10.1002/wics.1460](https://doi.org/10.1002/wics.1460)
- Chandra, K., Xie, A., Ragan-Kelley, J., & Meijer, E. 2022, in Advances in Neural Information Processing Systems, ed. A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho. <https://openreview.net/forum?id=-Qp-3L-5ZdII>
- Chen, J., & Chen, Z. 2008, Biometrika, 95, 759, doi: [10.1093/biomet/asn034](https://doi.org/10.1093/biomet/asn034)
- Ciotti, L. 1991, Astronomy and Astrophysics, 249, 99. <https://ui.adsabs.harvard.edu/abs/1991A&A...249...99C>
- Ciotti, L., & Bertin, G. 1999, Analytical properties of the  $R^{\gamma}(1/m)$  luminosity law, arXiv. <http://arxiv.org/abs/astro-ph/9911078>
- Coe, D., Fuselier, E., Benítez, N., et al. 2008, *\apj*, 681, 814, doi: [10.1086/588250](https://doi.org/10.1086/588250)
- Coles, J. P., Read, J. I., & Saha, P. 2014, Monthly Notices of the Royal Astronomical Society, 445, 2181, doi: [10.1093/mnras/stu1781](https://doi.org/10.1093/mnras/stu1781)

- de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247. <https://ui.adsabs.harvard.edu/abs/1948AnAp...11..247D>
- Diego, J. M., Protopapas, P., Sandvik, H. B., & Tegmark, M. 2005, Monthly Notices of the Royal Astronomical Society, 360, 477, doi: [10.1111/j.1365-2966.2005.09021.x](https://doi.org/10.1111/j.1365-2966.2005.09021.x)
- Diego, J. M., Tegmark, M., Protopapas, P., & Sandvik, H. B. 2007, Monthly Notices of the Royal Astronomical Society, 375, 958, doi: [10.1111/j.1365-2966.2007.11380.x](https://doi.org/10.1111/j.1365-2966.2007.11380.x)
- Duchi, J., Hazan, E., & Singer, Y. 2011, Journal of Machine Learning Research, 12, 2121
- Dye, S., & Warren, S. J. 2005, The Astrophysical Journal, 623, 31, doi: [10.1086/428340](https://doi.org/10.1086/428340)
- Dyson, F. W., & Eddington, A. S. 1920, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 220, 291, doi: [10.1098/rsta.1920.0009](https://doi.org/10.1098/rsta.1920.0009)
- Einstein, A. 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 844. <https://ui.adsabs.harvard.edu/abs/1915SPAW.....844E>
- . 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 142. <https://ui.adsabs.harvard.edu/abs/1917SPAW.....142E>
- Friedman, A. 1922, Zeitschrift für Physik, 10, 377, doi: [10.1007/BF01332580](https://doi.org/10.1007/BF01332580)
- Friedmann, A. 1924, Zeitschrift für Physik, 21, 326, doi: [10.1007/BF01328280](https://doi.org/10.1007/BF01328280)
- Gaudi, B. S. 2012, \araa, 50, 411, doi: [10.1146/annurev-astro-081811-125518](https://doi.org/10.1146/annurev-astro-081811-125518)
- Geyer, C. J. 1992, Statistical Science, 7, doi: [10.1214/ss/1177011137](https://doi.org/10.1214/ss/1177011137)
- Gilman, D., Birrer, S., Treu, T., Nierenberg, A., & Benson, A. 2019, Monthly Notices of the Royal Astronomical Society, 487, 5721, doi: [10.1093/mnras/stz1593](https://doi.org/10.1093/mnras/stz1593)
- Grabmeier, J., Kaltofen, E., & Weispfenning, V., eds. 2003, Computer Algebra Handbook (Berlin, Heidelberg: Springer Berlin Heidelberg), doi: [10.1007/978-3-642-55826-9](https://doi.org/10.1007/978-3-642-55826-9)
- Graham, A. W., & Driver, S. P. 2005, \pasa, 22, 118, doi: [10.1071/AS05001](https://doi.org/10.1071/AS05001)
- Graham, A. W., Erwin, P., Trujillo, I., & Asensio Ramos, A. 2003, The Astronomical Journal, 125, 2951, doi: [10.1086/375320](https://doi.org/10.1086/375320)
- Graham, A. W., Erwin, P., Trujillo, I., & Asensio Ramos, A. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho, 21

Griewank, A., Kulshreshtha, K., & Walther, A. 2012, Computing, 94, 125, doi: [10.1007/s00607-011-0162-z](https://doi.org/10.1007/s00607-011-0162-z)

Griewank, A., & Walther, A. 2008, Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition, 2nd edn. (Society for Industrial and Applied Mathematics), doi: [10.1137/1.9780898717761](https://doi.org/10.1137/1.9780898717761)

Géron, A. 2019, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems, second edition edn. (Beijing [China] ; Sebastopol, CA: O'Reilly Media, Inc)

Hamilton, A. J. S. 2019. [https://jila.colorado.edu/~ajsh/courses/astr3740\\_20/evol.html](https://jila.colorado.edu/~ajsh/courses/astr3740_20/evol.html)

Howell, S. B. 2006, Handbook of CCD astronomy, 2nd edn., Cambridge observing handbooks for research astronomers No. 5 (Cambridge ; New York: Cambridge University Press)

Huang, L., Qin, J., Zhou, Y., et al. 2020, Normalization Techniques in Training DNNs: Methodology, Analysis and Application

Hubble, E. 1929, Proceedings of the National Academy of Sciences, 15, 168, doi: [10.1073/pnas.15.3.168](https://doi.org/10.1073/pnas.15.3.168)

Jullo, E., Kneib, J.-P., Limousin, M., et al. 2007, New Journal of Physics, 9, 447, doi: [10.1088/1367-2630/9/12/447](https://doi.org/10.1088/1367-2630/9/12/447)

Kingma, D. P., & Ba, J. 2017, Adam: A Method for Stochastic Optimization, arXiv. <http://arxiv.org/abs/1412.6980>

Kneib, J.-P., & Natarajan, P. 2011, The Astronomy and Astrophysics Review, 19, 47, doi: [10.1007/s00159-011-0047-3](https://doi.org/10.1007/s00159-011-0047-3)

Kochanek, C. S. 2004, The Saas Fee Lectures on Strong Gravitational Lensing, arXiv. <http://arxiv.org/abs/astro-ph/0407232>

Kochanek, C. S., & Narayan, R. 1992, The Astrophysical Journal, 401, 461, doi: [10.1086/172078](https://doi.org/10.1086/172078)

Koopmans, L. V. E. 2005, Monthly Notices of the Royal Astronomical Society, 363, 1136, doi: [10.1111/j.1365-2966.2005.09523.x](https://doi.org/10.1111/j.1365-2966.2005.09523.x)

Kormann, R., Schneider, P., & Bartelmann, M. 1994, Astronomy and Astrophysics, 284, 285. <https://ui.adsabs.harvard.edu/abs/1994A&A...284..285K>

- LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. 2012, in Neural Networks: Tricks of the Trade, ed. G. Montavon, G. B. Orr, & K.-R. Müller, Vol. 7700 (Berlin, Heidelberg: Springer Berlin Heidelberg), 9–48, doi: [10.1007/978-3-642-35289-8\\_3](https://doi.org/10.1007/978-3-642-35289-8_3)
- Lee, C.-H. 2017, Universe, 3, 53, doi: [10.3390/universe3030053](https://doi.org/10.3390/universe3030053)
- Li, R., Napolitano, N. R., Roy, N., et al. 2022, The Astrophysical Journal, 929, 152, doi: [10.3847/1538-4357/ac5ea0](https://doi.org/10.3847/1538-4357/ac5ea0)
- Liddle, A. R. 2007, Monthly Notices of the Royal Astronomical Society: Letters, 377, L74, doi: [10.1111/j.1745-3933.2007.00306.x](https://doi.org/10.1111/j.1745-3933.2007.00306.x)
- Liesenborgs, J., De Rijcke, S., & Dejonghe, H. 2006, Monthly Notices of the Royal Astronomical Society, 367, 1209, doi: [10.1111/j.1365-2966.2006.10040.x](https://doi.org/10.1111/j.1365-2966.2006.10040.x)
- Margossian, C. C. 2019, WIREs Data Mining and Knowledge Discovery, 9, e1305, doi: [10.1002/widm.1305](https://doi.org/10.1002/widm.1305)
- Meneghetti, M. 2021, Lecture Notes in Physics, Vol. 956, Introduction to Gravitational Lensing: With Python Examples (Cham: Springer International Publishing), doi: [10.1007/978-3-030-73582-1](https://doi.org/10.1007/978-3-030-73582-1)
- Merten, J. 2016, Monthly Notices of the Royal Astronomical Society, 461, 2328, doi: [10.1093/mnras/stw1413](https://doi.org/10.1093/mnras/stw1413)
- Narayan, R., & Bartelmann, M. 1997, Lectures on Gravitational Lensing, arXiv. <http://arxiv.org/abs/astro-ph/9606001>
- Newton, E. R., Marshall, P. J., Treu, T., et al. 2011, The Astrophysical Journal, 734, 104, doi: [10.1088/0004-637X/734/2/104](https://doi.org/10.1088/0004-637X/734/2/104)
- Newton, I. 1704, Opticks: or, A treatise of the reflections, refractions, inflexions and colours of light : also two treatises of the species and magnitude of curvilinear figures (London: Printed for Sam. Smith, and Benj. Walford), doi: [10.5479/sil.302475.39088000644674](https://doi.org/10.5479/sil.302475.39088000644674)
- Nightingale, J. W., & Dye, S. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2940, doi: [10.1093/mnras/stv1455](https://doi.org/10.1093/mnras/stv1455)
- Paczynski, B. 1986, \apj, 304, 1, doi: [10.1086/164140](https://doi.org/10.1086/164140)
- Paszke, A., Gross, S., Massa, F., et al. 2019, PyTorch: An Imperative Style, High-Performance Deep Learning Library, arXiv. <http://arxiv.org/abs/1912.01703>

- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *The Astronomical Journal*, 124, 266, doi: [10.1086/340952](https://doi.org/10.1086/340952)
- . 2010, *The Astronomical Journal*, 139, 2097, doi: [10.1088/0004-6256/139/6/2097](https://doi.org/10.1088/0004-6256/139/6/2097)
- Peng, C. Y., Impey, C. D., Rix, H., et al. 2006, *The Astrophysical Journal*, 649, 616, doi: [10.1086/506266](https://doi.org/10.1086/506266)
- Plato, ed. 2010, *Plato's cosmology: the Timaeus of Plato*, first issued in paperback edn., The international library of philosophy 4, Ancient philosophy : in 10 volumes No. 4 (Oxfordshire New York, NY: Routledge)
- Refsdal, S. 1964, *Monthly Notices of the Royal Astronomical Society*, 128, 307, doi: [10.1093/mnras/128.4.307](https://doi.org/10.1093/mnras/128.4.307)
- Ruder, S. 2016, doi: [10.48550/ARXIV.1609.04747](https://doi.org/10.48550/ARXIV.1609.04747)
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, *Nature*, 323, 533, doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0)
- Saha, P., & Williams, L. L. R. 2004, *The Astronomical Journal*, 127, 2604, doi: [10.1086/383544](https://doi.org/10.1086/383544)
- Schaul, T., Zhang, S., & LeCun, Y. 2013, in *Proceedings of Machine Learning Research*, Vol. 28, *Proceedings of the 30th International Conference on Machine Learning*, ed. S. Dasgupta & D. McAllester (Atlanta, Georgia, USA: PMLR), 343–351. <https://proceedings.mlr.press/v28/schaul13.html>
- Schmidhuber, J. 2015, *Neural Networks*, 61, 85, doi: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003)
- Schneider, P., Ehlers, J., & Falco, E. E. 1992, *Gravitational Lenses*, ed. I. Appenzeller, G. Börner, M. Harwit, R. Kippenhahn, J. Lequeux, P. A. Strittmatter, & V. Trimble, *Astronomy and Astrophysics Library* (Berlin, Heidelberg: Springer Berlin Heidelberg), doi: [10.1007/978-3-662-03758-4](https://doi.org/10.1007/978-3-662-03758-4)
- Sebesta, K., Williams, L. L. R., Mohammed, I., Saha, P., & Liesenborgs, J. 2016, *Monthly Notices of the Royal Astronomical Society*, 461, 2126, doi: [10.1093/mnras/stw1433](https://doi.org/10.1093/mnras/stw1433)
- Skilling, J. 2004, in *AIP Conference Proceedings*, Vol. 735 (Garching (Germany): AIP), 395–405, doi: [10.1063/1.1835238](https://doi.org/10.1063/1.1835238)
- Speagle, J. S. 2019, doi: [10.48550/ARXIV.1909.12313](https://doi.org/10.48550/ARXIV.1909.12313)

- Sumi, T., Kamiya, K., Udalski, A., et al. 2011, doi: [10.48550/ARXIV.1105.3544](https://arxiv.org/abs/1105.3544)
- Suyu, S. H. 2016, in *Astrophysical Applications of Gravitational Lensing*, 1st edn., ed. E. Mediavilla, J. A. Muñoz, F. Garzón, & T. J. Mahoney (Cambridge University Press), 1–30, doi: [10.1017/CBO9781139940306.002](https://doi.org/10.1017/CBO9781139940306.002)
- Suyu, S. H., & Blandford, R. D. 2006, *Monthly Notices of the Royal Astronomical Society*, 366, 39, doi: [10.1111/j.1365-2966.2005.09854.x](https://doi.org/10.1111/j.1365-2966.2005.09854.x)
- Suyu, S. H., Marshall, P. J., Blandford, R. D., et al. 2009, *The Astrophysical Journal*, 691, 277, doi: [10.1088/0004-637X/691/1/277](https://doi.org/10.1088/0004-637X/691/1/277)
- Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, *Monthly Notices of the Royal Astronomical Society*, 371, 983, doi: [10.1111/j.1365-2966.2006.10733.x](https://doi.org/10.1111/j.1365-2966.2006.10733.x)
- Sérsic, J. L. 1963, *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, 6, 41
- . 1968, *Atlas de Galaxias Australes*
- Tavernier, S. 2010, *Experimental techniques in nuclear and particle physics* (Berlin Heidelberg: Springer)
- Tessore, N., & Benton Metcalf, R. 2015, *Astronomy & Astrophysics*, 580, A79, doi: [10.1051/0004-6361/201526773](https://doi.org/10.1051/0004-6361/201526773)
- Tielemans, T., & Hinton, G. 2014, URL [http://www.cs.toronto.edu/tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](http://www.cs.toronto.edu/tijmen/csc321/slides/lecture_slides_lec6.pdf)
- Treu, T. 2010, *Annual Review of Astronomy and Astrophysics*, 48, 87, doi: [10.1146/annurev-astro-081309-130924](https://doi.org/10.1146/annurev-astro-081309-130924)
- Trujillo, I., Aguerri, J., Cepa, J., & Gutiérrez, C. 2001a, *Monthly Notices of the Royal Astronomical Society*, 328, 977, doi: [10.1046/j.1365-8711.2001.04937.x](https://doi.org/10.1046/j.1365-8711.2001.04937.x)
- Trujillo, I., Aguerri, J. A. L., Cepa, J., & Gutiérrez, C. M. 2001b, *Monthly Notices of the Royal Astronomical Society*, 321, 269, doi: [10.1046/j.1365-8711.2001.03987.x](https://doi.org/10.1046/j.1365-8711.2001.03987.x)
- Trujillo, I., Erwin, P., Asensio Ramos, A., & Graham, A. W. 2004, *\aj*, 127, 1917, doi: [10.1086/382712](https://doi.org/10.1086/382712)
- Tully, R. B. 2023, doi: [10.48550/ARXIV.2305.11950](https://arxiv.org/abs/2305.11950)
- Udalski, A., Szymanski, M., Kaluzny, J., et al. 1993, *\actaa*, 43, 289

- Vegetti, S., & Koopmans, L. V. E. 2009, Monthly Notices of the Royal Astronomical Society, 392, 945, doi: [10.1111/j.1365-2966.2008.14005.x](https://doi.org/10.1111/j.1365-2966.2008.14005.x)
- Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, Monthly Notices of the Royal Astronomical Society, 408, 1969, doi: [10.1111/j.1365-2966.2010.16865.x](https://doi.org/10.1111/j.1365-2966.2010.16865.x)
- Verma, A. 2000, Current Science, 78, 804. <https://www.jstor.org/stable/24103956>
- Wallington, S., Kochanek, C. S., & Narayan, R. 1996, The Astrophysical Journal, 465, 64, doi: [10.1086/177401](https://doi.org/10.1086/177401)
- Warren, S. J., & Dye, S. 2003, The Astrophysical Journal, 590, 673, doi: [10.1086/375132](https://doi.org/10.1086/375132)
- Will, C. M. 1988, American Journal of Physics, 56, 413, doi: [10.1119/1.15622](https://doi.org/10.1119/1.15622)
- . 2014, Living Reviews in Relativity, 17, 4, doi: [10.12942/lrr-2014-4](https://doi.org/10.12942/lrr-2014-4)
- . 2015, Classical and Quantum Gravity, 32, 124001, doi: [10.1088/0264-9381/32/12/124001](https://doi.org/10.1088/0264-9381/32/12/124001)
- Zeiler, M. D. 2012, ADADELTA: An Adaptive Learning Rate Method, arXiv. <http://arxiv.org/abs/1212.5701>



# Ringraziamenti

Nel momento in cui questo capitolo della mia vita si chiude e ne inizia uno nuovo, desidero esprimere la mia più profonda gratitudine a coloro che hanno reso possibile questo viaggio. La realizzazione di questa tesi magistrale non sarebbe stata possibile senza il sostegno, l'incoraggiamento e l'amore di molte persone straordinarie che mi hanno accompagnato lungo il percorso. È un privilegio avere l'opportunità di esprimere il mio sincero apprezzamento a queste persone speciali.

Non posso che iniziare senza dedicare il ringraziamento più speciale e profondo alla mia migliore amica, Federica. La nostra amicizia, iniziata per caso, si è trasformata in un legame indissolubile che ha arricchito e colorato ogni aspetto della mia vita, trasformando semplici momenti in ricordi preziosi che porterò sempre nel cuore. Federica, sei stata la mia roccia, la mia confidente, e la fonte di una forza incrollabile che mi ha permesso di navigare attraverso le difficoltà con determinazione e coraggio. Le tue parole di incoraggiamento, il tuo spirito indomito e la tua capacità di trovare la luce anche nei momenti più bui hanno acceso in me una scintilla di speranza e ottimismo, indispensabili per affrontare le sfide accademiche e personali.

Non meno importante è il ringraziamento che voglio estendere alla mia famiglia: ai miei genitori, Laura e Paolo, e a mia sorella, Giulia. Il vostro amore incondizionato, il vostro sostegno costante e la vostra fiducia in me hanno costituito il mio faro nella notte, guidandomi attraverso le incertezze e celebrando con me ogni traguardo raggiunto. La vostra presenza, anche se fisicamente distante, è stata una costante fonte di conforto e forza.

Un caloroso ringraziamento va esteso anche ai miei amici, pilastri della mia vita, fonte di gioia, ispirazione e sostegno. Edoardo, Mattia e Lorenzo, insieme abbiamo condiviso più di un decennio di risate, avventure e sfide. La distanza non ha fatto altro che rafforzare il legame che ci unisce, dimostrando che l'amicizia vera non conosce ostacoli.

Estendo la mia gratitudine ai miei compagni di università, Andrea, Leonardo e Chiara, con i quali ho condiviso non solo gli studi, ma anche un percorso di crescita

personale. La nostra collaborazione, il sostegno reciproco e i momenti condivisi hanno arricchito significativamente la mia esperienza universitaria.

Un ringraziamento speciale va anche a un altro Andrea, una presenza costante nelle nostre ormai passate serate in aula studio. La tua compagnia ha reso quelle lunghe ore notturne non solo più produttive, ma anche piacevoli e ricche di amicizia.

Non posso trascurare di ringraziare il mio caro coinquilino Michele, la cui compagnia è sempre stata una fonte di allegria (e caffé).

In questo momento di riflessione, desidero anche esprimere la mia gratitudine a tutte quelle persone che, magari non citate esplicitamente, hanno arricchito il mio percorso con la loro presenza, il loro incoraggiamento e il loro affetto. Ogni piccolo gesto, ogni parola di sostegno, ogni momento condiviso ha contribuito a rendere questo viaggio indimenticabile.

Grazie di cuore a tutti voi per aver camminato al mio fianco in questa avventura, lasciando un'impronta indelebile nel mio cuore e nella mia anima.

Grazie a tutti.

Simone





