

Alma Mater Studiorum - Università di Bologna

DEPARTMENT OF PHYSICS AND ASTRONOMY

Master Degree in Astrophysics and Cosmology

Gravitational microscope in MACSJ1206: unveiling galaxy cluster member mass distribution

Master Degree Thesis

Candidate:

Massimiliano Zago

Supervisor:

Robert Benton Metcalf

Abstract

Gravitational lensing stands as a powerful tool for probing the mass distribution properties of both baryonic and dark matter components. Taking advantage of Strong Gravitational Lensing, this study aims to infer the characteristics of a galaxy-scale substructure within the galaxy cluster *MACSJ1206*.

While modeling the visible components of a galaxy is relatively straightforward, the morphology of the dark matter distribution remains elusive. Strong lensing, with its sensitivity to the entire mass content, offers a unique way for investigating dark matter distribution, making it the predominant tool for this kind of substructure studies.

The primary objective of this research is to conduct an exhaustive analysis of the previously mentioned substructure, an early-type galaxy (ETG). The approach involves analyzing the highly magnified Einstein ring surrounding the ETG, due to the galaxy-galaxy strong lensing effects, to impose stringent constraints on its matter distribution, typically well-described by an isothermal profile for ETGs. This investigation holds significance for three distinct reasons. Firstly, galaxies in clusters are subject to various interaction processes, both dynamic and hydrodynamic, capable of significantly altering their matter distribution. Thus, discovering a deviation from the expected profile may indicate structural changes induced by such processes. Secondly, probing the mass distribution allows for an exploration of the dark matter component characteristics, a field still affected by many uncertainties nowadays. Thirdly, the distribution of dark matter in a clustered environment may deviate from that in an isolated galaxy.

Typically, an Early-Type Galaxy (ETG) is well-described by an isothermal profile. In this thesis, to address the aforementioned objectives and verifying potential alterations in the mass distribution profile, two distinct model profiles are employed to describe the lens. The first model, the elliptical power-law (EPL), detailed by [Tessore and R Benton Metcalf 2015], is nested within the second. The second model, a more complex representation known as the broken power-law (BPL), is outlined in the study by [O’Riordan, Stephen J Warren, and Daniel J Mortlock 2021]. For both models, the isothermal profile serves as a specific case, allowing to determine if the standard ETG trend is maintained. Notably, the BPL introduces the possibility to examine the presence of a change point in the mass distribution profile, indicating potential truncation in the mass distribution, including the dark matter component.

Since the galaxy under analysis is not isolated, the model that describes it must incorporate the effects of the cluster deflection field. An existing lens model for MACSJ1206 is documented in the research by [Bergamini et al. 2019]. This study provided a comprehensive model for the entire cluster, employing the `lenstool` software package [Jullo, Kneib, et al. 2007, Jullo and Kneib 2009]. To manage this external field, the results presented in this article are used to generate a map of the deflection field produced by the entire cluster at the specific position of the galaxy we are taking under examination.

The work by [Bergamini et al. 2019] focused on reconstructing the overall mass distribution of the cluster using multiple images of lensed background objects across the entire cluster. While providing extensive information on the Mpc scale, this approach sacrificed resolution, especially on the Kpc scale. Furthermore, the use of `lenstool` in their work assumed an analytical or parameterized relationship between observed galaxy properties and mass distribution. Although beneficial for fitting the entire cluster, this introduced a form of averaging that might lack accuracy for individual galaxies.

This thesis extends the analysis by considering multiple images not previously addressed in [Bergamini et al. 2019], with a specific focus on the Einstein ring surrounding the target galaxy. Additionally, the study fully exploits the capabilities of the recently released `PyAutoLens` software [J. Nightingale et al. 2021]. Unlike `lenstool`, `PyAutoLens` has the ability to reconstruct the lensed source adaptively, making use of the adaptive Delaunay triangular mesh grid without relying on any parametric models. This non-parametric approach provides greater degrees of freedom, resulting in a significantly higher precision level in the system’s reconstruction compared to scenarios where parametric models are assumed.

The advantages of this approach are manifold. The data supplied to `PyAutoLens` and extracted from [Bergamini et al. 2019] are comprehensive, drawing from the knowledge of the entire cluster. This eliminates the need to reconstruct the external deflection field solely relying on information from the region surrounding the galaxy on the Kpc scale, but rather on the Mpc scale. Furthermore, considering that the model by [Bergamini et al. 2019] in the region surrounding the galaxy at a kpc scale may not be optimal, the Bayesian approach, along with the capacity of `PyAutoLens` to establish priors on the parameters, allows the outcomes from [Bergamini et al. 2019] to serve as a prior for the investigation. This introduces a certain level of uncertainty to test the [Bergamini et al. 2019] model on the kpc scale.

The exhaustive analysis undertaken in this thesis reveals a noteworthy discovery: the total mass density

distribution of the observed early-type galaxy exhibits a slope significantly steeper than the isothermal profile $\rho \propto r^{-2}$ typically expected for such galaxies - approximately one factor greater $\rho \approx r^{-3}$. Furthermore, the data, alongside with the Bayesian model selection criterion, does not support the introduction of a break radius. This implies that the distribution is not isothermal up to a certain point with an abrupt change, rather, it maintains a steeper profile than the isothermal throughout the entire mass distribution, extending to the outer radius where dark matter dominates.

Introduzione

Il lensing gravitazionale è uno strumento ideale per esaminare le proprietà della distribuzione di massa, sia della materia barionica che di quella oscura. Sfruttando la lensing gravitazionale forte, questo studio mira a inferire le caratteristiche di una sotto-struttura su scala galattica all'interno di *MACSJ1206*. Mentre modellare i componenti visibili di una galassia è relativamente semplice, la morfologia della distribuzione di materia oscura rimane sfuggente. Il lensing gravitazionale forte, con la sua sensibilità all'intera massa, offre un modo unico per indagare la distribuzione di materia oscura, facendolo diventare lo strumento predominante per studi riguardanti sotto-strutture di questo tipo.

L'obiettivo principale di questa ricerca è condurre un'analisi approfondita della sotto-struttura precedentemente menzionata, una galassia di tipo early-type (ETG). L'approccio utilizzato in questo lavoro di tesi coinvolge l'analisi dell'anello di Einstein l'ETG, dovuto agli effetti di lensing gravitazionale forte di tipo galassia-galassia, per imporre vincoli più stringenti sulla sua distribuzione di materia, comunemente ben descritta da un profilo isotermico per le ETG. Questa indagine si dimostra estremamente interessante per tre ragioni distinte.

In primo luogo, le galassie nei gruppi sono soggette a vari processi di interazione, sia dinamici che idrodinamici, capaci di alterare significativamente la loro distribuzione di materia. Scoprire una deviazione dal profilo atteso potrebbe indicare cambiamenti strutturali indotti da tali processi. In secondo luogo, esaminare la distribuzione di massa consente di esplorare le caratteristiche della componente di materia oscura, un campo ancora afflitto da molte incertezze al giorno d'oggi. In terzo luogo, la distribuzione di materia oscura in una galassia facente parte del cluster può essere differente da quella in una galassia isolata.

Tipicamente, una galassia di tipo Early-Type (ETG) è ben descritta da un profilo isotermo. In questa tesi, per affrontare gli obiettivi sopra menzionati e verificare potenziali alterazioni nel profilo di distribuzione di massa, vengono impiegati due distinti modelli. Il primo modello, la elliptical power-law (EPL), studiato in Tessore and R Benton Metcalf 2015, è una versione semplificata del secondo. Il secondo modello, una rappresentazione più complessa nota come broken power-law (BPL), è definito nello studio di O'Riordan, Stephen J Warren, and Daniel J Mortlock 2021. Per entrambi i modelli, il profilo isotermo funge da caso specifico, consentendo di determinare se il trend standard di ETG è mantenuto. In particolare, il BPL introduce la possibilità di esaminare la presenza di un punto di troncamento nel profilo di distribuzione di massa, indicando, per l'appunto, un potenziale troncamento nella distribuzione di massa, compresa la componente di materia oscura.

Dal momento che la galassia in analisi non è isolata, il modello che la descrive deve incorporare gli effetti del campo di deflessione del cluster. Un modello di lente esistente per MACSJ1206 è documentato nella ricerca di Bergamini et al. 2019. Questo studio ha fornito un modello completo per l'intero cluster, utilizzando il pacchetto software `lenstool` Jullo, Kneib, et al. 2007; Jullo and Kneib 2009. Per gestire questo campo esterno, i risultati presentati in questo articolo vengono utilizzati per generare una mappa del campo di deflessione prodotto dall'intero cluster nella posizione specifica della galassia che stiamo esaminando.

Il lavoro di Bergamini et al. 2019 si è concentrato sulla ricostruzione della distribuzione di massa complessiva del cluster utilizzando immagini multiple distribuite nell'intero cluster, di oggetti retrostanti lo stesso. Pur fornendo informazioni approfondite sulla scala dei Mpc, questo approccio ha sacrificato la risoluzione, in particolare sulla scala dei Kpc. Inoltre, l'uso di `lenstool` nel loro lavoro assume una relazione analitica o parametrizzata tra le proprietà osservate delle galassie e la distribuzione di massa. Sebbene utile per ricostruire l'intero cluster, questo introduce una sorta di media che potrebbe mancare di precisione per le singole galassie.

Questa tesi estende l'analisi considerando immagini multiple non affrontate in Bergamini et al. 2019, con un focus specifico sull'anello di Einstein che circonda la galassia in esame. Inoltre, lo studio sfrutta appieno le capacità del software `PyAutolens` recentemente rilasciato J. Nightingale et al. 2021. A differenza di `lenstool`, `PyAutolens` ha la capacità di ricostruire in modo adattivo la sorgente lensata, sfruttando la griglia adattiva triangolare di Delaunay, senza fare affidamento su modelli parametrici. Questo approccio non parametrico fornisce maggiori gradi di libertà, risultando in un livello di precisione significativamente più elevato nella ricostruzione del sistema rispetto a scenari in cui si assumono modelli parametrici.

I vantaggi di questo approccio sono molteplici. I dati forniti a `PyAutolens` e estratti da Bergamini et al. 2019 sono esaustivi, attingendo alla conoscenza dell'intero cluster. Ciò elimina la necessità di ricostruire il campo di deflessione esterno facendo affidamento esclusivamente sulle informazioni fornite dalla regione circostante la galassia sulla scala di Kpc, ma piuttosto sulla scala di Mpc. Inoltre, considerando che il modello di Bergamini et al. 2019 nella regione circostante la galassia su scala Kpc potrebbe non essere ottimale, l'approccio bayesiano, insieme alla capacità di `PyAutolens` di stabilire prior sui parametri, consente ai risultati di Bergamini et al. 2019 di servire come base per l'indagine. Ciò introduce un certo livello di incertezza per testare il modello di Bergamini et al. 2019 sulla scala di Kpc.

L'analisi approfondita condotta in questa tesi rivela una scoperta significativa: la distribuzione della densità di massa totale dell'osservata galassia di tipo early-type presenta una pendenza significativamente maggiore rispetto al profilo isotermo $\rho \propto r^{-2}$ generalmente atteso per tali galassie - approssimativamente un fattore maggiore $\rho \approx r^{-3}$. Inoltre, i dati, insieme al criterio di selezione del modello bayesiano, non supportano l'introduzione di un raggio di troncamento. Ciò implica che la distribuzione non è isoterna fino a un certo punto con un cambiamento repentino, ma mantiene un profilo più ripido rispetto all'isoterna per l'intera distribuzione di massa, estendendosi fino al raggio esterno in cui la materia oscura domina.

Contents

1	Gravitational Lensing theory	1
1.1	Lensing Equation	1
1.1.1	Deflection Angle	1
1.1.2	Lens Equation	4
1.1.3	Convergence, Lensing Jacobian and Magnification	5
2	Galaxy mass and light profile	13
2.1	Baryonic, dark and total mass distribution	13
2.2	Surface Mass Density Distribution	16
2.2.1	The elliptical power law profile	16
2.2.2	The elliptical broken power law profile	22
2.3	Surface Brightness profile	28
2.3.1	The Elliptical Sersic Profile	28
3	The Bayesian Approach	31
3.1	Model selection and parameters estimation through the bayesian approach	31
3.1.1	Bayes' Theorem and Bayesian approach	31
3.1.2	Parameters estimation	32
3.1.3	Model selection	32
3.2	Numerical sampling methods	33
3.2.1	Numerical Sampling Method: Introduction	33
3.2.2	Nested Sampling	33
4	PyAutolens Features	39
4.1	The log-likelihood for parametric models	39
4.2	Multiple image position setting	41
4.3	Search chaining and prior passing	42
4.4	Adaptive grid	44
4.4.1	The Inversion procedure	44
4.4.2	The Adaptive Pixelization: the Delaunay magnification pixelization	46
4.4.3	Adaptive features improvement: The Delaunay brigtness pixelization and the adaptive regularization	46
4.4.4	The log-likelihood for adaptive models	51
5	Data	53
5.1	The cluster MACS J1206	53
5.1.1	The HST data	53
5.1.2	The data	54
5.1.3	The point spread function	55
5.1.4	The cluster deflection field	57

6	Data analysis: the Analysis Pipeline	63
6.1	1 Fit: Lens light fitting	65
6.2	2 Fit: Fitting the system for the first time	69
6.3	3 Fit: Last parametric fit of the source	74
6.4	4 Fit: first adaptive reconstrucion of the source	78
6.5	5 fit: fitting the lens plane with adaptive source reconstruction for the first time	82
6.6	6 Fit: Delaunay Brightness Image adaptive grid	86
6.7	7 Fit: Last fit	91
6.7.1	Elliptical power-law	91
6.7.2	Elliptical broken power-law	95
7	Results analysis	99
7.1	Bergamini et al. 2019 mass model	99
7.2	Mass and Einstein radius of the power-law	101
7.3	Shear results and uncertainties	105
7.4	Comparison between Epl and dPIE models	106
7.5	The slope of the lens: power-law or broken power-law model	106
7.5.1	Elliptical power-law slope	106
7.5.2	Do the data provide support for the introduction of a broken radius?	107
7.6	Conclusions and future perspectives	109

Chapter 1

Gravitational Lensing theory

1.1 Lensing Equation

1.1.1 Deflection Angle

The theory of General Relativity (GR) states that any object with mass bends the surrounding space-time, also causing the deviation of the trajectory of any light beam that was to pass through the gravitational field of that mass, this phenomenon is called *gravitational lensing* and the mass distributions are the *gravitational lenses*. This is illustrated in figure 1.1.

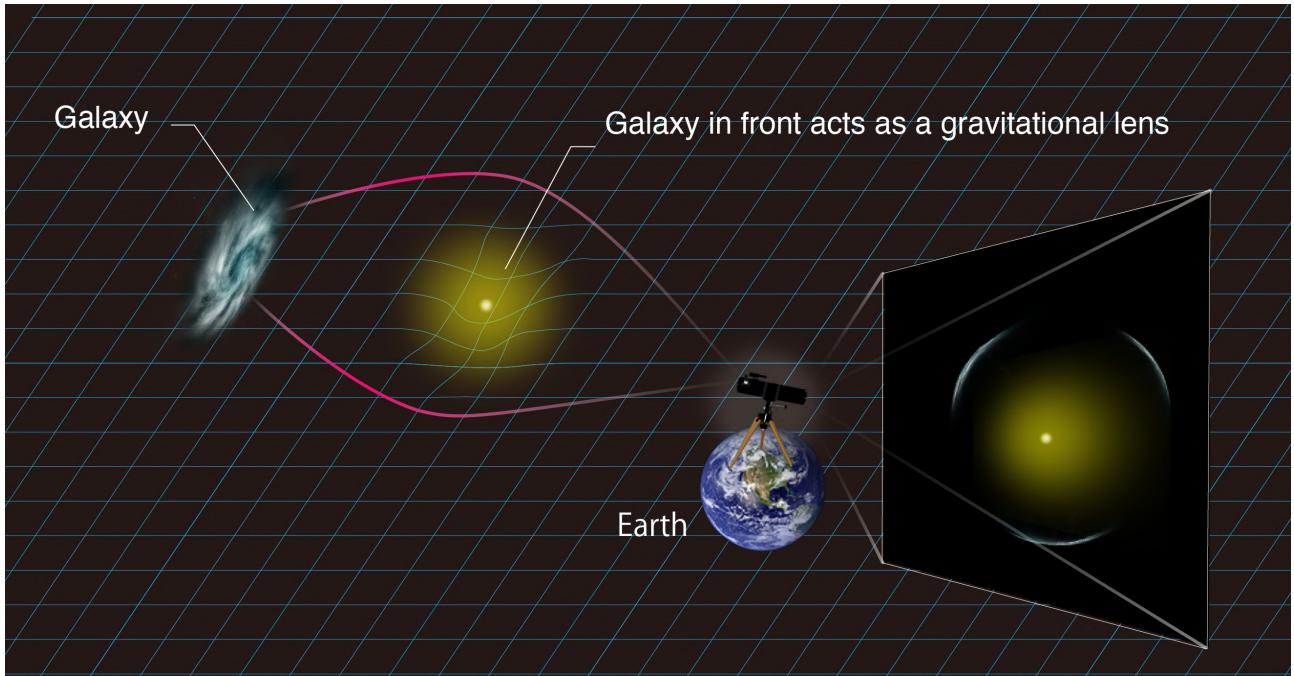


Figure 1.1: In the image is possible to observe how the foreground galaxy bends the space-time, causing the light rays of the background source galaxy to be bended too. Image from: Kavli IPMU, University of Tokyo.

The gravitational lensing theory here below follows the book '*Introduction to gravitational lensing*' [Meneghetti 2021a].

Light path in any gravitational field follows the so called space-time geodesics, that are the trajectories of any given object within a gravitational field and that can be calculated starting from the Einstein gravitational field equations.

The equations that describe the gravitational lensing phenomena can also be derived starting from the Fermat's

principle. This states that light will follow a trajectory along which the quantity

$$\int \frac{n}{c} dl \quad (1.1)$$

is an extremal, where n is the refraction index and the integral is performed along the light path. In fact the perturbed region, where the gravitational field exists, can be described in terms of an effective refractive index. Thus, thanks to *least action principle*, the light path $\vec{x}(l)$, through the gravitational field is given by

$$\delta \int_A^B n(\vec{x}(l)) dl = 0 \quad (\text{Fermat's principle}) \quad (1.2)$$

where A and B are respectively the starting and the ending point of the path, and $\vec{x}(l)$ is the position along the path l .

In order to find the refraction index is necessary to take into account the GR. The starting point is the space-time line element ds^2

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (1.3)$$

where $g_{\mu\nu}$ is the space-time metric and dx^μ and dx^ν are the space-time infinitesimal displacement. It is now necessary to make an approximation: the weak-field approximation. In other words, the Newtonian gravitational potential ϕ is much smaller than c^2 , so that $\phi/c^2 \ll 1$. Most of the lensing phenomena, such as those covered in this thesis, can be described by this weak field approximation. In this case the metric in equation 1.3 can be described as a perturbation of the flat space-time Minkowsky metric:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.4)$$

so, in the weak-field approximation the metric is

$$g_{\mu,\nu} = \eta_{\mu,\nu} + h_{\mu,\nu} \quad (1.5)$$

where $|h_{\mu,\nu}| \ll 1$.

Another common assumption, in addition to the previous weak-field approximation ($\phi/c^2 \ll 1$), consists in assuming a static gravitational field, that is independent on time. In this case the perturbed metric becomes:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} = \eta_{\mu\nu} + \begin{pmatrix} -2\phi/c^2 & 0 & 0 & 0 \\ 0 & -2\phi/c^2 & 0 & 0 \\ 0 & 0 & -2\phi/c^2 & 0 \\ 0 & 0 & 0 & -2\phi/c^2 \end{pmatrix} \quad (1.6)$$

The space-time interval, using equation (1.3), can be written as:

$$ds^2 = -\left(1 + \frac{2\phi}{c^2}\right)c^2 dt^2 + \left(1 - \frac{2\phi}{c^2}\right)d\vec{x}^2 \quad (1.7)$$

where $d\vec{x}$ is the three dimensional space interval.

Photons propagate along null line elements ($ds^2 = 0$, null proper time), covering a distance $|d\vec{x}|$ in a coordinate time dt . This, from equation (1.7), leads to:

$$ds^2 = 0 \rightarrow \left(1 - \frac{2\phi}{c^2}\right)d\vec{x}^2 = \left(1 + \frac{2\phi}{c^2}\right)c^2 dt^2 \quad (1.8)$$

and the speed of light, c' , measured by an external observer, in a region where a potential ϕ exists is:

$$c' = \frac{|d\vec{x}|}{dt} = c \cdot \sqrt{\frac{1 + \frac{2\phi}{c^2}}{1 - \frac{2\phi}{c^2}}} \stackrel{2\phi/c^2 \ll 1}{\approx} c \cdot \left(1 + \frac{2\phi}{c^2}\right). \quad (1.9)$$

The refraction index $n(\vec{x})$ is given by:

$$n = \frac{c}{c'(\vec{x})} = \frac{1}{1 + \frac{2\phi(\vec{x})}{c^2}} \stackrel{\phi(\vec{x})/c^2 \ll 1}{\approx} 1 - \frac{2\phi(\vec{x})}{c^2} \quad (1.10)$$

Given that $\phi(\vec{x}) < 0$, $n(\vec{x}) > 1$ and the speed of light $c' < c$, the vacuum value. Now it is possible to reparametrize the path l :

$$dl = \left| \frac{d\vec{x}}{d\lambda} \right| d\lambda \quad (1.11)$$

where λ is an arbitrary curve parameter. The equation (1.2) becomes

$$\delta \int_{\lambda_A}^{\lambda_B} n(\vec{x}(\lambda)) \left| \frac{d\vec{x}}{d\lambda} \right| d\lambda = 0 \quad (1.12)$$

where λ_A and λ_B are respectively the starting and the ending point of the path.

Now,

$$n(\vec{x}(\lambda)) \left| \frac{d\vec{x}}{d\lambda} \right| \equiv L(\vec{x}, \dot{\vec{x}}, \lambda) \quad (1.13)$$

resembles the Lagrangian, and

$$\dot{\vec{x}} \equiv \frac{d\vec{x}}{d\lambda} \quad (1.14)$$

is the tangent to the curve that defines the light path, and so

$$\delta \int_{\lambda_A}^{\lambda_B} L(\vec{x}, \dot{\vec{x}}, \lambda) d\lambda = 0 \quad (1.15)$$

This leads to Euler-Lagrange equations:

$$\frac{d}{d\lambda} \frac{\partial L}{\partial \dot{\vec{x}}} - \frac{\partial L}{\partial \vec{x}} = 0 \quad (1.16)$$

From equation (1.13):

$$\frac{\partial L}{\partial \vec{x}} = |\dot{\vec{x}}| \cdot \text{grad}(n) \quad (1.17)$$

and, from equation (1.14):

$$\frac{\partial L}{\partial \dot{\vec{x}}} = n \frac{\dot{\vec{x}}}{|\dot{\vec{x}}|} \quad (1.18)$$

It is possible to assume that $|\dot{\vec{x}}| = 1$ and call $\vec{e} \equiv \dot{\vec{x}}$ the tangent unit vector along the light path. Equation (1.16) can now be written as :

$$\frac{d}{d\lambda} (n\vec{e}) - \text{grad}(n) = 0 \quad (1.19)$$

$$\begin{aligned} n\dot{\vec{e}} + \vec{e}(\text{grad}(n)\dot{\vec{x})) &= \text{grad}(n) \\ n\dot{\vec{e}} &= \text{grad}(n) - \vec{e}(\text{grad}(n)\dot{\vec{x}}) \end{aligned} \quad (1.20)$$

The term $\vec{e}(\text{grad}(n)\dot{\vec{x}})$ is the gradient component along the light path, so the right hand side of equation (1.20) is the gradient of n perpendicular to the light path

$$\dot{\vec{e}} = \frac{1}{n} \nabla_{\perp} n = \nabla_{\perp} \ln(n) \quad (1.21)$$

From equation (1.10) $n = 1 - \frac{2\phi(\vec{x})}{c^2}$ and with the approx $\phi(\vec{x})/c^2 \ll 1$, $\ln(n) \approx \frac{2\phi(\vec{x})}{c^2}$ and thus

$$\dot{\vec{e}} = -\frac{2}{c^2} \nabla_{\perp} \phi(\vec{x}) \quad (1.22)$$

It follows that

$$\vec{e}_{out} = \vec{e}_{in} + \int_{\lambda_A}^{\lambda_B} \dot{\vec{e}} d\lambda \quad (1.23)$$

where \vec{e}_{in} is the direction of the light at the moment it "enters" the gravitational well and \vec{e}_{out} is the direction of the light when it "leaves" the gravitational well.

Thus the total deflection angle is

$$\hat{\alpha} = \vec{e}_{in} - \vec{e}_{out} = \int_{\lambda_A}^{\lambda_B} \dot{\vec{e}} d\lambda = \frac{2}{c^2} \int_{\lambda_A}^{\lambda_B} \nabla_{\perp} \phi(x(\vec{\lambda})) d\lambda \quad (1.24)$$

Suppose the line of sight to be aligned with the z axis; since $\phi(\vec{x})/c^2 \ll 1$, the deflection angle is small, so is possible to use the *Born approximation*, integrating along the unperturbed light path by substituting the variable λ in equation (1.24) with z . Consider the distance D_{OS} between the observer and the source to be much greater than the region where the gravitational potential generated by the lens exists, set the position of the lens to be at $z = 0$, and consider an impact parameter b , then, with the above *Born approximation*, the deflection angle can be written as

$$\hat{\alpha} = \frac{2}{c^2} \int_{-\infty}^{\infty} \nabla_{\perp} \phi(\vec{x}) dz \quad (1.25)$$

It can be demonstrated [Meneghetti 2021a] that an alternative form for the equation (1.25) is

$$\hat{\alpha}(\vec{\xi}) = \frac{4G}{c^2} \int \frac{(\vec{\xi} - \vec{\xi}') \Sigma(\vec{\xi}')}{|\vec{\xi} - \vec{\xi}'|^2} d^2 \vec{\xi}' \quad (1.26)$$

where

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z) dz \quad (1.27)$$

is the surface mass density, and $\vec{\xi}$ is the planar coordinate on the *lens plane*. In fact, as said before, the above equations assume that the distance between observer, lens and source are much greater than the lens physical size. This justifies the subsequent assumption that the deflection occurs in a very small section of the light path. Eventually, this even justifies the *thin screen approximation*. The lens is approximated by a planar distribution of matter, the so called *lens plane*. Even the source is assumed to lie on a plane, the so called *source plane*.

1.1.2 Lens Equation

Thanks to geometrical considerations, showed in figure (1.2) and described in [Meneghetti 2021a], where, for simplicity, $D_{OS} \equiv D_S$, $D_{OL} \equiv D_L$, it is possible to obtain the lens equation

$$\vec{\theta} D_S = \vec{\beta} D_S + \hat{\alpha}(\vec{\theta}) D_{SL} \quad (1.28)$$

where $\vec{\theta}$ is the 2D adimensional angular position vector on the lens plane and $\vec{\beta}$ is the 2D adimensional angular position vector on the source plane.

Defining the *reduced deflection angle*

$$\vec{\alpha}(\vec{\theta}) \equiv \frac{D_{LS}}{D_S} \hat{\alpha}(\vec{\theta}) \quad (1.29)$$

the equation (1.28) becomes

$$\vec{\theta} = \vec{\beta} + \vec{\alpha}(\vec{\theta}) \quad (1.30)$$

where $\vec{\theta}$ and $\vec{\beta}$ can be expressed in radians or arcseconds. Hereafter, for this thesis, the arcseconds will be assumed for all the physical quantities.

Thanks to *lens equation* (1.30) it is possible to map the lensed source image on the lens plane back to the source plane and reconstruct the true source light distribution, obviously keeping in mind the limitations imposed by the instrumentation and the precision of the reconstruction algorithms.

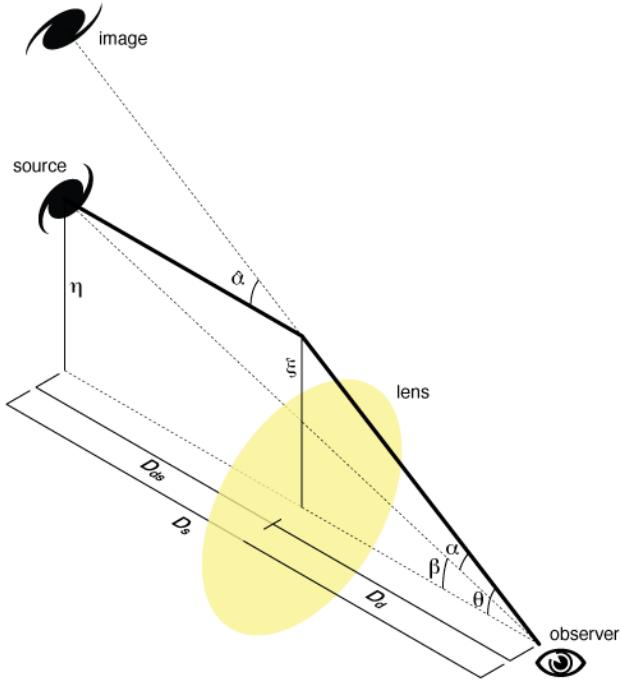


Figure 1.2: Scheme of a lensing system. In yellow the lens plane, at the distance D_{OL} , where the light deflection occurs. D_{OS} is the distance between the observer and the source. D_{LS} is the distance between the source and the lens. ξ and $\vec{\eta}$ are respectively the physical distances on the lens and source plane. Image from: https://en.wikipedia.org/wiki/Gravitational_lensing_formalism

1.1.3 Convergence, Lensing Jacobian and Magnification

Convergence

In lensing formalism, it is common to use the so-called *convergence*, that consists in the *adimensional surface mass density distribution*. Hereafter, in this thesis, its spatial dependencies will be expressed in terms of angular position on the lens plane. From equation (1.27):

$$k(\vec{\theta}) = \frac{\Sigma(D_L \vec{\theta})}{\Sigma_{crit}} \quad (1.31)$$

where Σ_{crit} is the *critical surface density on the lensing system*

$$\Sigma_{crit} = \frac{c^2 D_S}{4\pi G D_L D_{LS}} \quad (1.32)$$

It is important to note that convergence is a quantity strictly dependent on the redshift, through the three different distances that define it that depend in turn on the redshift.

Lensing Jacobian and Shear

Suppose that the deflection angle does not vary significantly over the scale $d\vec{\theta}$ (fig. 1.3), in this case, the lens equation (1.30) can be expressed in a differential form at the first order as

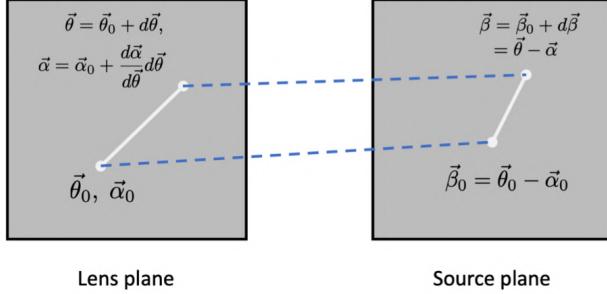


Figure 1.3: Scheme of a lensing system and lens mapping at the first order. [Meneghetti 2021b]

$$d\vec{\beta} = \left(I - \frac{d\vec{\alpha}(\vec{\theta})}{d\vec{\theta}} \right) d\vec{\theta} \rightarrow A \equiv \frac{d\vec{\beta}}{d\vec{\theta}} = \left(I - \frac{d\vec{\alpha}(\vec{\theta})}{d\vec{\theta}} \right) \quad (1.33)$$

$\frac{d\vec{\beta}}{d\vec{\theta}}$ is the *Lensing Jacobian*, that is a symmetric second rank tensor describing the mapping between the source and the lens plane at the first order. In fact, at the first order the positions of the multiple images on the lens plane can be traced back to the source plane through the relation

$$\vec{\beta} = A \cdot \vec{\theta} \quad (1.34)$$

The tensor can be decomposed in a sum of an isotropic diagonal tensor and a trace-less anisotropic symmetric tensor

$$A_{isotropic} = (1 - k(\vec{\theta})) \quad (1.35)$$

where $k(\vec{\theta})$ is the convergence described above.

This component of the tensor consists in a rescaling of the the source dimensions of a factor $1/(1-k\vec{\theta})$, because the tensor is isotropic, or, in other words, the source retains its original shape, just rescales its dimensions. Instead, as far as the anisotropic part is concerned, it is called *Shear* and it is usually indicated as $A_{anisotropic} = \Gamma$

$$A_{anisotropic} = \Gamma = \begin{pmatrix} \frac{1}{2}(\alpha_{1,1}(\vec{\theta}) - \alpha_{2,2}(\vec{\theta})) & \alpha_{1,2}(\vec{\theta}) \\ \alpha_{2,1}(\vec{\theta}) & -\frac{1}{2}(\alpha_{1,1}(\vec{\theta}) - \alpha_{2,2}(\vec{\theta})) \end{pmatrix} = \begin{pmatrix} \gamma_1(\vec{\theta}) & \gamma_2(\vec{\theta}) \\ \gamma_2(\vec{\theta}) & -\gamma_1(\vec{\theta}) \end{pmatrix} \quad (1.36)$$

where $\alpha_{1,2}(\vec{\theta}) = \alpha_{2,1}(\vec{\theta})$. Furthermore, $\alpha_{1,1}(\vec{\theta})$ means the derivative of the first component of the deflection angle with respect to the component θ_1 of the position vector $\vec{\theta}$, $\alpha_{2,2}(\vec{\theta})$ means the derivative of the second component of the deflection angle with respect to the component θ_2 of the position vector $\vec{\theta}$ and $\alpha_{1,2}(\vec{\theta})$ means the derivative of the first component of the deflection angle with respect to the component θ_2 of the position vector $\vec{\theta}$. The shear represents an anisotropic distortion along a privileged direction, so the source does not retain its shape anymore, but it is deformed along the shear eigenvector directions.

The shear is a symmetric tensor, and therefore, there is a rotation $R(\phi(\vec{\theta}))$ that allows to change the reference frame and pass into one where the shear is in diagonal form

$$\Gamma \rightarrow \Gamma' = R(\phi(\vec{\theta}))^T \Gamma R(\phi(\vec{\theta})) \quad (1.37)$$

Thus, the shear components transform as:

$$\begin{aligned} \gamma_1(\vec{\theta}) \rightarrow \gamma'_1(\vec{\theta}) &= \gamma_1(\vec{\theta})\cos(2\phi(\vec{\theta})) + \gamma_2(\vec{\theta})\sin(2\phi(\vec{\theta})) \\ \gamma_2(\vec{\theta}) \rightarrow \gamma'_2(\vec{\theta}) &= -\gamma_1(\vec{\theta})\sin(2\phi(\vec{\theta})) + \gamma_2(\vec{\theta})\cos(2\phi(\vec{\theta})) \end{aligned} \quad (1.38)$$

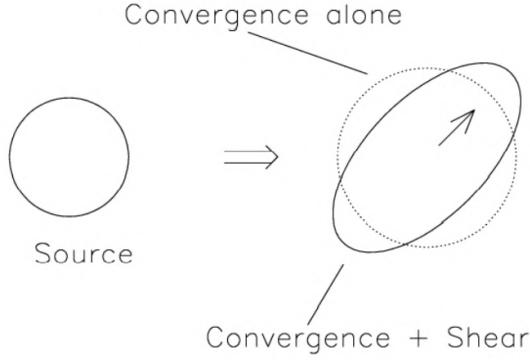


Figure 1.4: Combined effects of isotropic (convergence) and anisotropic (shear) tensors [Meneghetti 2021b]. This representation of the variation of the shape of the source is correct only in the case when the source is completely symmetrical (as in the figure), otherwise the induced deformation is more complex than an ellipticization. This is intended to be a representative image of a simple case but that allows us to understand the idea behind the shear

The eigenvalue of the tensor are

$$\pm \sqrt{\gamma_1^2 + \gamma_2^2} = \pm \gamma \quad (1.39)$$

In its diagonal form the shear is

$$\Gamma = \begin{pmatrix} \gamma & 0 \\ 0 & -\gamma \end{pmatrix} = R(\phi(\vec{\theta}))^T \Gamma R(\phi(\vec{\theta})) \quad (1.40)$$

Thus

$$\begin{aligned} \gamma &= \gamma'_1(\vec{\theta}) = \gamma_1(\vec{\theta})\cos(2\phi(\vec{\theta})) + \gamma_2(\vec{\theta})\sin(2\phi(\vec{\theta})) \\ 0 &= \gamma'_2(\vec{\theta}) = -\gamma_1(\vec{\theta})\sin(2\phi(\vec{\theta})) + \gamma_2(\vec{\theta})\cos(2\phi(\vec{\theta})) \end{aligned} \quad (1.41)$$

and so it is possible to rewrite the two components γ_1 and γ_2 of the shear as

$$\begin{aligned} \gamma_1(\vec{\theta}) &= \gamma(\vec{\theta})\cos(2\phi(\vec{\theta})) \\ \gamma_2(\vec{\theta}) &= \gamma(\vec{\theta})\sin(2\phi(\vec{\theta})) \end{aligned} \quad (1.42)$$

As said above, the shear stretches the source image along a particular directions, the one identified by the angle ϕ and shrinks along the one identified by $\phi + \pi/2$.

The lensing Jacobian can be expressed as

$$A = \begin{pmatrix} 1 - k(\vec{\theta}) - \gamma_1(\vec{\theta}) & -\gamma_2(\vec{\theta}) \\ -\gamma_2(\vec{\theta}) & 1 - k(\vec{\theta}) + \gamma_1(\vec{\theta}) \end{pmatrix} = (1 - k(\vec{\theta})) - \gamma \begin{pmatrix} \cos(2\phi(\vec{\theta})) & \sin(2\phi(\vec{\theta})) \\ \sin(2\phi(\vec{\theta})) & -\cos(2\phi(\vec{\theta})) \end{pmatrix} \quad (1.43)$$

Its determinant $\det A$ is

$$\det A = (1 - k(\vec{\theta}) - \gamma(\vec{\theta}))(1 - k(\vec{\theta}) + \gamma(\vec{\theta})) \quad (1.44)$$

where $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$.

Magnification, critical lines and caustics

During a gravitational lensing event the surface element $d\beta^2$ on the source plane is mapped into the surface element $d\theta^2$ on the lens plane through the lens equation. Furthermore, during gravitational lensing event there is no absorption or creation of photons and this, in addition to the Liouville's theorem, ensure the surface brightness conservation. The magnification or de-magnification effect during the lensing event depends exclusively on the solid angle under which the source is seen.

The magnification μ is defined as the ratio between the flux received from a lensed image and the those coming from the unlensed source, so it is an adimensional quantity. It can be demonstrated that the convergence can be expressed as

$$\mu = \frac{1}{detA} = \frac{1}{(1 - k^2(\vec{\theta})) - \gamma^2(\vec{\theta})} \quad (1.45)$$

It is also possible to quantify the magnification along the tangential μ_t and radial μ_r direction

$$\mu_t = \frac{1}{\lambda_t} = \frac{1}{(1 - k(\vec{\theta}) - \gamma(\vec{\theta}))} \quad \mu_r = \frac{1}{\lambda_r} = \frac{1}{(1 - k(\vec{\theta}) + \gamma(\vec{\theta}))} \quad (1.46)$$

For $\lambda_t = (1 - k(\vec{\theta}) - \gamma(\vec{\theta})) = 0$ and $\lambda_r = (1 - k(\vec{\theta}) + \gamma(\vec{\theta})) = 0$ the magnification goes to infinity along the tangential and radial direction respectively. These two conditions define two curves on the lens plane, $\lambda_t(\vec{\theta})$ and $\lambda_r(\vec{\theta})$, called respectively *tangential and radial critical lines*. Images close to radial critical lines are distorted perpendicularly respect to these lines while those near to the tangential critical lines are distorted tangentially. The mapping of $\lambda_t(\vec{\theta})$ on the source plane through the lens equation defines respectively the so called *tangential and radial caustics*.

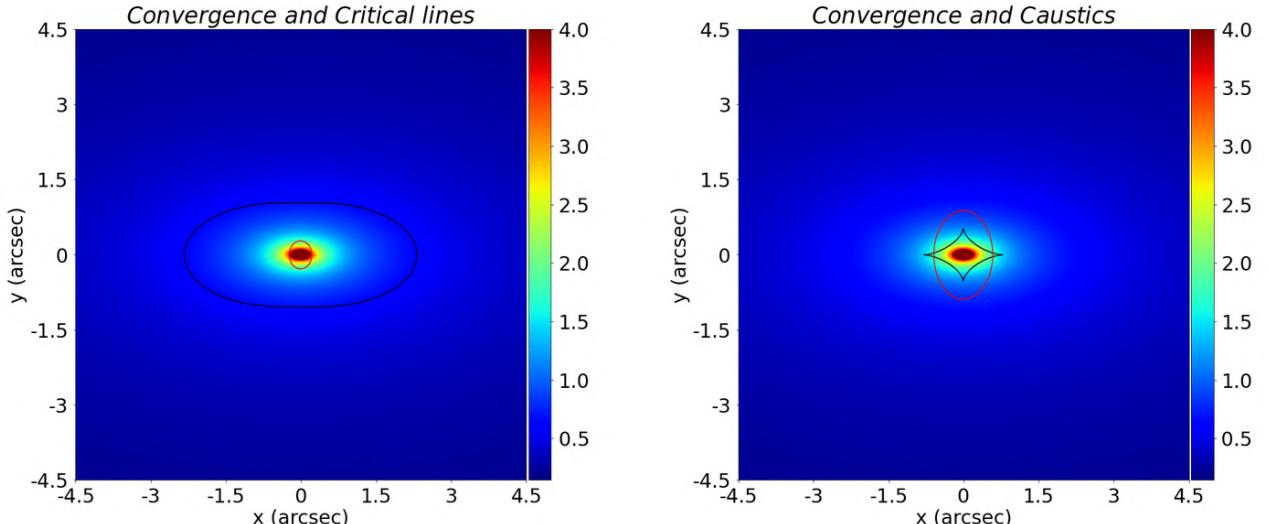


Figure 1.5: Elliptical Power-Law convergence (section (2.2.1): Einstein radius=1.60'', axis-ratio=0.6, slope=0.75). In the image on the left the tangential critical lines are in black and the radial critical lines (slope < 2) are in red. On the left image the tangential caustic are in black and the radial caustics are in red.

External Perturbation

Lens are often not isolated, as the one treated in this thesis, that is a cluster member. For this reason it is necessary to consider the surrounding environment effects. One possibility is to use an approximate form of the first order expansion covered in section (1.1.3). This approximation supposes

$$\begin{aligned}\gamma_1(\vec{\theta}) &= \text{constant} \\ \gamma_2(\vec{\theta}) &= \text{constant} \\ k(\vec{\theta}) &= \text{constant}\end{aligned}\tag{1.47}$$

all over the plane. This means, on the base of equation (1.42), that

$$\begin{aligned}\gamma(\vec{\theta}) &= \text{constant} \\ \phi(\vec{\theta}) &= \text{constant}\end{aligned}\tag{1.48}$$

all over the plane.

One possibility to describe the external perturbation is to use the first order expansion of the deflection angle and take advantage of the above constant terms.

The first order expansion of the deflection angle is

$$\vec{\alpha}_{\text{external}}(\vec{\theta}) \approx \vec{\alpha}_0 + \frac{\partial \vec{\alpha}}{\partial \vec{\theta}}(\vec{\theta} - \vec{\theta}_0)\tag{1.49}$$

Now, from equation (1.33) we have that

$$\frac{\partial \vec{\alpha}}{\partial \vec{\theta}} = \left(I - \frac{\partial \vec{\beta}}{\partial \vec{\theta}} \right) \stackrel{\text{eq.1.43}}{=} \begin{pmatrix} k + \gamma_1 & \gamma_2 \\ \gamma_2 & k - \gamma_1 \end{pmatrix}\tag{1.50}$$

And so, from equation (1.49) we have hat:

$$\begin{aligned}\alpha_1 &\approx \alpha_{0,1} + k(\theta_1 - \theta_{0,1}) + \gamma(\cos(2\phi)(\theta_1 - \theta_{0,1}) + \sin(2\phi)(\theta_2 - \theta_{0,2})) \\ \alpha_2 &\approx \alpha_{0,2} + k(\theta_2 - \theta_{0,2}) + \gamma(\sin(2\phi)(\theta_1 - \theta_{0,1}) - \cos(2\phi)(\theta_2 - \theta_{0,2}))\end{aligned}\tag{1.51}$$

where the terms k, γ, ϕ are the constants from equation (1.47). From now on, for this section, $\gamma = \gamma_{\text{ext}}$. Often, however, we tend to insert in the models of external shear only the part concerning the actual shear tensor, or in other words

$$\vec{\alpha}_{\text{external}}(\vec{\theta}) \approx \Gamma \cdot \vec{\theta} = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \cdot \vec{\theta} = \gamma_{\text{ext}} \begin{pmatrix} \cos(2\phi) & \sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{pmatrix} \cdot \vec{\theta}\tag{1.52}$$

In this way, thanks to the superposition principle, the equation (1.30) becomes

$$\vec{\theta} = \vec{\beta} + \vec{\alpha}_{\text{lens}}(\vec{\theta}) + \vec{\alpha}_{\text{external}}(\vec{\theta}) = \vec{\beta} + \vec{\alpha}_{\text{lens}}(\vec{\theta}) + \Gamma \cdot \vec{\theta}\tag{1.53}$$

Insert external perturbations, modify the critical lines, let's see what terms are the cause of the change. To do this we will use the full first order expansion in the equation (1.50).

Firstly, the two components of the total deflection angle are given by

$$\begin{aligned}\alpha_{1,tot} &\approx \alpha_{1,lens} + \alpha_{0,1} + k(\theta_1 - \theta_{0,1}) + \gamma_{ext}(\cos(2\phi)(\theta_1 - \theta_{0,1}) + \sin(2\phi)(\theta_2 - \theta_{0,2})) \\ \alpha_{2,tot} &\approx \alpha_{2,lens} + \alpha_{0,2} + k(\theta_2 - \theta_{0,2}) + \gamma_{ext}(\sin(2\phi)(\theta_1 - \theta_{0,1}) - \cos(2\phi)(\theta_2 - \theta_{0,2}))\end{aligned}\quad (1.54)$$

Then, we need to calculate the two components of the shear, γ_1 and γ_2 , using the definition given in equation (1.36):

$$\gamma_1 = \frac{1}{2} \left(\frac{\partial \alpha_{1,tot}}{\partial \theta_1} - \frac{\partial \alpha_{2,tot}}{\partial \theta_2} \right) \quad \gamma_2 = \frac{\partial \alpha_{1,tot}}{\partial \theta_2} \quad (1.55)$$

Thus, we have that

$$\begin{aligned}\frac{\partial \alpha_{1,tot}}{\partial \theta_1} &= \frac{\partial \alpha_{1,lens}}{\partial \theta_1} + k + \gamma_{ext} \cos(2\phi) \\ \frac{\partial \alpha_{2,tot}}{\partial \theta_2} &= \frac{\partial \alpha_{2,lens}}{\partial \theta_1} + k - \gamma_{ext} \cos(2\phi) \\ \frac{\partial \alpha_{1,tot}}{\partial \theta_2} &= \frac{\partial \alpha_{1,lens}}{\partial \theta_2} + \gamma_{ext} \sin(2\phi)\end{aligned}\quad (1.56)$$

Now, we have that

$$\begin{aligned}\gamma_1 &= \frac{1}{2} \left(\frac{\partial \alpha_{1,tot}}{\partial \theta_1} - \frac{\partial \alpha_{2,tot}}{\partial \theta_2} \right) \\ &= \frac{1}{2} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} + k + \gamma_{ext} \cos(2\phi) - \left(\frac{\partial \alpha_{2,lens}}{\partial \theta_1} + k - \gamma_{ext} \cos(2\phi) \right) \right) \\ &= \frac{1}{2} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right) + \gamma_{ext} \cos(2\phi)\end{aligned}\quad (1.57)$$

With these terms we can compute the quantities that make up the critical line equation. Because the tangential critical lines' equation and the radial one differ just for a minus sign, only the tangential will be demonstrated.

The equation that must be calculated is

$$\lambda_t = (1 - (k_{lens}(\vec{\theta}) + k_{external})) - \gamma(\vec{\theta}) = 0 \quad (1.58)$$

where $k_{external}$ is the one from equation (1.47) and the θ dependence is in the deflection angle derivative. The γ term is evaluated using

$$\begin{aligned}\gamma(\vec{\theta}) &= \sqrt{\gamma_1^2 + \gamma_2^2} \\ &= \sqrt{\left(\frac{1}{2} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right) + \gamma_{ext} \cos(2\phi) \right)^2 + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_2} + \gamma_{ext} \sin(2\phi) \right)^2} \\ &= \left(\frac{1}{4} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right)^2 + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right) \cdot \gamma_{ext} \cos(2\phi) + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_2} \right)^2 \right. \\ &\quad \left. + 2 \cdot \frac{\partial \alpha_{1,lens}}{\partial \theta_2} \cdot \gamma_{ext} \sin(2\phi) + \gamma_{ext}^2 \right)^{\frac{1}{2}}\end{aligned}\quad (1.59)$$

From equation (1.58) we have

$$\begin{aligned}\lambda_t &= 1 - (k_{lens}(\vec{\theta}) + k_{external}) - \gamma(\vec{\theta}) = \\ &= 1 - (k_{lens}(\vec{\theta}) + k_{external}) - \left(\frac{1}{4} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right)^2 + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right) \cdot \gamma_{ext} \cos(2\phi) \right. \\ &\quad \left. + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_2} \right)^2 + 2 \cdot \frac{\partial \alpha_{1,lens}}{\partial \theta_2} \cdot \gamma_{ext} \sin(2\phi) + \gamma_{ext}^2 \right)^{\frac{1}{2}} = 0\end{aligned}\quad (1.60)$$

Where in the case of no external perturbation we would have

$$\lambda_t = 1 - k_{lens}(\vec{\theta}) - \gamma(\vec{\theta}) = 1 - k_{lens}(\vec{\theta}) - \left(\left(\frac{1}{4} \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_1} - \frac{\partial \alpha_{2,lens}}{\partial \theta_2} \right)^2 + \left(\frac{\partial \alpha_{1,lens}}{\partial \theta_2} \right)^2 \right)^{\frac{1}{2}} \right) \quad (1.61)$$

Before proceeding, it is also worth mentioning how the lens equation (1.30) is modified by the external perturbation of the complete first order expansion. In fact, using the vectorial form of equation (1.51), it becomes

$$\vec{\theta} = \vec{\beta} + \vec{\alpha}_{lens}(\vec{\theta}) + \vec{\alpha}_{external}(\vec{\theta}) = \vec{\beta} + \vec{\alpha}_{lens}(\vec{\theta}) + \vec{\alpha}_{0,ext} + k_{ext} \vec{\theta} + \Gamma \cdot \vec{\theta} \quad (1.62)$$

It can be rearranged into

$$\begin{aligned} \vec{\beta} + \vec{\alpha}_{0,ext} &= \vec{\beta}' = (1 - k_{ext}) \vec{\theta} - \vec{\alpha}_{lens}(\vec{\theta}) - \Gamma \cdot \vec{\theta} \\ \frac{\vec{\beta}'}{(1 - k_{ext})} &= \vec{\theta} - \frac{\vec{\alpha}_{lens}(\vec{\theta})}{(1 - k_{ext})} - \frac{\Gamma}{(1 - k_{ext})} \cdot \vec{\theta} \end{aligned} \quad (1.63)$$

and eventually into

$$\vec{\beta}'' = \vec{\theta} - \vec{\alpha}'_{lens}(\vec{\theta}) - \Gamma' \cdot \vec{\theta} \quad (1.64)$$

where

$$\begin{aligned} \vec{\beta}'' &= \frac{\vec{\beta}'}{(1 - k_{ext})} \\ \vec{\alpha}'_{lens}(\vec{\theta}) &= \frac{\vec{\alpha}_{lens}(\vec{\theta})}{(1 - k_{ext})} \\ \Gamma' &= \frac{\Gamma}{(1 - k_{ext})} \end{aligned} \quad (1.65)$$

This is the same form as (1.53), and therefore is solved in the same way. This means that in the fitting procedure the deflection angle that is found is $\vec{\alpha}'_{lens}(\vec{\theta})$ and not $\vec{\alpha}_{lens}(\vec{\theta})$. What has just been pointed out it is very important, because the convergence profile is built from deflection angle fitting, this means that even the convergence is rescaled by a factor $(1 - k_{ext})$ and thus the mass of the lens, that is the 2D integral of the convergence.

In fact, equation (1.29), in combination with equation (1.26) and dividing by Σ_{crit} , can be written in a rescaled form as

$$\vec{\alpha}_{lens}(\vec{\theta}) = \frac{1}{\pi} \int \frac{\vec{\theta} - \vec{\theta}'}{\|\vec{\theta} - \vec{\theta}'\|^2} k(\vec{\theta}') d^2 \vec{\theta}' \quad (1.66)$$

In the case where equation (1.65) is solved we have

$$\begin{aligned} \vec{\alpha}'_{lens}(\vec{\theta}) &= \frac{1}{\pi(1 - k_{ext})} \int \frac{\vec{\theta} - \vec{\theta}'}{\|\vec{\theta} - \vec{\theta}'\|^2} k(\vec{\theta}') d^2 \vec{\theta}' \\ &= \frac{1}{\pi} \int \frac{\vec{\theta} - \vec{\theta}'}{\|\vec{\theta} - \vec{\theta}'\|^2} k'(\vec{\theta}') d^2 \vec{\theta}' \end{aligned} \quad (1.67)$$

where

$$k'(\vec{\theta}) = \frac{k(\vec{\theta})}{(1 - k_{ext})} \quad (1.68)$$

Therefore, because the mass is the 2D integral of the convergence, and in this case we have $k'(\vec{\theta})$, the mass is not the correct one. To obtain the real lens mass, it is necessary to multiply the mass value just obtained by the factor $(1 - k_{ext})$.

Obviously, if we consider just the external shear term, and thus the equation (1.53), the external convergence is fixed to be zero $k_{ext} = 0$ and the rescaling procedure it is not needed.

In the scenario where the equation that is solved is the (1.64), we need to rescale the mass, the deflection angle and the source positions by multiplying the factor $(1 - k_{ext})$.

Eventually, the first constant term of the expansion can be excluded during the optimization of the lens plane and can be added in the end thanks to the superposition principle. In fact, as it is showed above it does not modify any quantity, the critical lines still remain the same.

Chapter 2

Galaxy mass and light profile

2.1 Baryonic, dark and total mass distribution

In this section we briefly describe the reason of the usage of the two different surface mass density profile, the elliptical power-law profile and the elliptical broken power-law profile.

In general, the density distribution for an ETG is dominated by stars for $r \lesssim r_{\text{effective}}$, where $r_{\text{effective}}$ is the radius that contains half of the galaxy light, and by dark matter for $r \gtrsim r_{\text{effective}}$. Thus, the total three dimensional density $\rho_{\text{tot}}(\vec{r})$ of early type galaxies (ETGs) is, in general, well approximated by a power-law $\rho_{\text{tot}}(\vec{r}) \propto ||\vec{r}||^{-\gamma}$, with $\gamma \approx 2$, i.e the density profile of a *singular isothermal ellipsoid (SIE)*, [Cimatti, Fraternali, and Nipoti 2019, (§5.3.5)].

This trend can be easily explained taking into consideration two of the main profiles that are usually used to describe ETGs: the Hernquist profile for the stellar component and the NFW profile for the dark matter component.

In order to show this we take advantage of what is written in [Van De Ven, Mandelbaum, and Keeton 2009]. Consider an ellipsoidal mass density distribution

$$\begin{aligned} \rho(x, y, z) &= \rho(m) \\ m^2 &= \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \\ a &\geq b \geq c \end{aligned} \tag{2.1}$$

The major semi-axis length a is a scale parameter, m is a dimensionless ellipsoidal radius and in the spherical case we have $a = b = c$ with $m = r/a$.

The ellipsoidal density profiles is the *cusped-density profile*, that have a smoother power-law at small radii and a steeper power-law at large radii. Both the two profiles are particular cases of this.

$$\rho(m) = \frac{\rho_0}{m^\gamma(1 + m)^{n-\gamma}} \tag{2.2}$$

It includes:

- Spherical case $m = r/a$, where a is the scale radius.
- The Hernquist profile: $\gamma = 1$, $n = 4$.
- The NFW profile: $\gamma = 1$, $n = 3$.

What we are interested to know about these two profiles are their logarithmic slope

$$\gamma_{\log} = \frac{d \ln(\rho(m))}{d \ln(m)} = \frac{m}{\rho(m)} \cdot \frac{d \rho(m)}{d m} = -\frac{\gamma + n \cdot m}{1 + m} \tag{2.3}$$

Take into account the spherical case, just to simplify calculations ($m = r/a$), and assume the scale radius a , that is different for each profile.

For the Hernquist profile we have that:

$$\begin{aligned}\gamma_{\log} &\xrightarrow{m \rightarrow 0} -1 \\ \gamma_{\log} &\xrightarrow{m \rightarrow 1} -2.5 \\ \gamma_{\log} &\xrightarrow{m \rightarrow \infty} -4\end{aligned}\tag{2.4}$$

For the NFW profile we have that:

$$\begin{aligned}\gamma_{\log} &\xrightarrow{m \rightarrow 0} -1 \\ \gamma_{\log} &\xrightarrow{m \rightarrow 1} -2 \\ \gamma_{\log} &\xrightarrow{m \rightarrow \infty} -3\end{aligned}\tag{2.5}$$

Superposing the two profiles it is possible to reproduce figure (2.1).

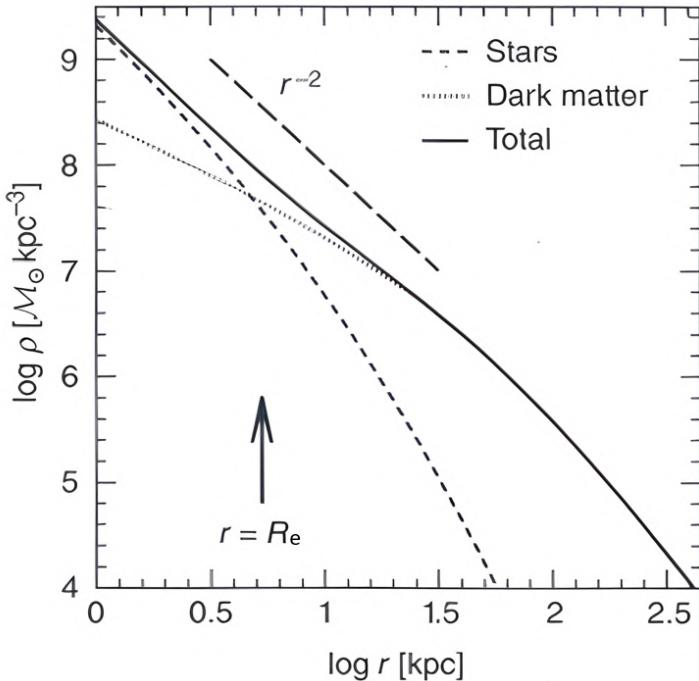


Figure 2.1: Total density radial distribution of an ETG galaxy. Superposition of baryonic and dark matter component. Image from [Cimatti, Fraternali, and Nipoti 2019].

We use both power-law and broken power-law as a mass model, to test the hypothesis that the galaxy we are observing has a usual mass distribution, as the one in the figure (2.1), or if it is an anomalous galaxy, which, therefore, does not present a trend like the one in the figure (2.1).

We initially use power-law for two reasons:

- Fewer parameters to fit, and therefore less problems concerning degeneration.
- The slope of the power-law is one of the free parameters, therefore, if this does not coincide, more or less, with the typical trend of the SIE, it means that the mass distribution has anomalies.

Eventually, we use broken power-law. Through the use of this profile we verify two aspects:

- Through bayesian statistics, especially through the comparison of bayesian evidence (section (3.1.3)), we check whether the observed data support the introduction of an additional parameter in the model, namely the brake radius.

- In case the introduction of brake radius is justified, we use this mass model to investigate at what distance from the center of the lens the change of slope occurs. This allows to acquire more precise information about the actual distribution of matter.
- Moreover, assuming that the introduction of brake radius is justified and that the slopes are not those of a SIE, this model allows us to see what are the trends that determine the mass distribution of the lens.

Eventually, in general early type galaxies show a Sérsic index n that is greater than $n_{ETGs} \geq 2.5$ and usually it has a value of about $n_{ETGs} \geq 4$, typical of the profile of De Vaucouleurs, that well describes ETGs. Thus, we even test the Sérsic index of this galaxy by fitting its light profile with an elliptical Sérsic Profile.

2.2 Surface Mass Density Distribution

To treat the elliptical case, the so called pseudo-elliptical model is often taken, because in this way quantities such as the shear, the deflection angle and the convergence can be expressed directly in terms of derivatives of the potential, without the need to solve integrals. Although, this way of proceeding leads to an unphysical surface mass densities with a peanut-shaped isodensity contours or negative values, and this problem becomes more evident as the ellipticity increases. Therefore, to avoid these problems it is necessary to use an elliptical mass distribution. In the following, two mass distributions will be presented, or rather the relative convergence, the *Elliptical power-law profile* [Tessore and R Benton Metcalf 2015] and the *two-dimensional broken power-law profile* [O’Riordan, Stephen J Warren, and Daniel J Mortlock 2021].

2.2.1 The elliptical power law profile

The general ellipse and the elliptical profile

The dimensionless axis-symmetrical profile for a power-law lens is

$$k(\theta) = \frac{2-t}{2} \left(\frac{b}{\theta} \right)^t \quad (2.6)$$

where $0 < t < 2$ (otherwise if $t > 2$ there are not radial critical lines) is the slope of the profile, $b > 0$ is the scale radius, also known as the *Einstein Radius* and $\theta = \sqrt{\theta_1^2 + \theta_2^2} > 0$ is the two-dimensional distance from the lens centre.

The more general elliptical form of the profile in equation (2.6) can be obtained by stretching one of the axes, in this case the θ_1 -axis, by a factor of q^{-1} , where $0 < q \leq 1$ is the so called *axes-ratio*. Performing this operation, the circular isodensity contour defined by $\theta = \text{const}$ become now ellipses with the semi-major axis θ/q and a semi-minor axis θ . The ellipse equation is

$$\frac{\theta_1^2}{a^2} + \frac{\theta_2^2}{b^2} = 1 \quad (2.7)$$

that, in the specific case of a circle $a = b = r$ of given radius $r = \theta$, it becomes

$$\frac{\theta_1^2}{r^2} + \frac{\theta_2^2}{r^2} = 1 \quad (2.8)$$

So, taking $\theta_1 \rightarrow q\theta_1$

$$\theta_1^2 + \theta_2^2 = r^2 \rightarrow q^2\theta_1^2 + \theta_2^2 = r^2 \quad (2.9)$$

$$\frac{q^2\theta_1^2}{r^2} + \frac{\theta_2^2}{r^2} = 1 \quad \text{ellipse equation} \quad (2.10)$$

$$\frac{\theta_1^2}{\frac{r^2}{q^2}} + \frac{\theta_2^2}{r^2} = 1 \quad \text{ellipse equation} \quad (2.11)$$

where $\frac{\theta^2}{q^2}$ is the semi-major axis and θ^2 is the semi-minor axis.

Therefore, the most general case, that is the equation implemented in the software PyAutoLens, used for the data analysis, is

$$\frac{q^2((\theta_1 - \theta_{1,0})\cos(\epsilon) + (\theta_2 - \theta_{2,0})\sin(\epsilon))^2}{r^2} + \frac{((\theta_1 - \theta_{1,0})\sin(\epsilon) - (\theta_2 - \theta_{2,0})\cos(\epsilon))^2}{r^2} = 1 \quad (2.12)$$

where $(\theta_{1,0}, \theta_{2,0})$ is the center of the ellipse and $\epsilon \in [0, 2\pi]$ is the rotation angle on the ellipse’s plane, defined in the counterclockwise from the θ_1 -axis to the semi-major axis of the ellipse.

The general elliptical power-law profile is

$$k(\theta) = \frac{2-t}{2} \left(\frac{b}{\theta} \right)^t \quad (2.13)$$

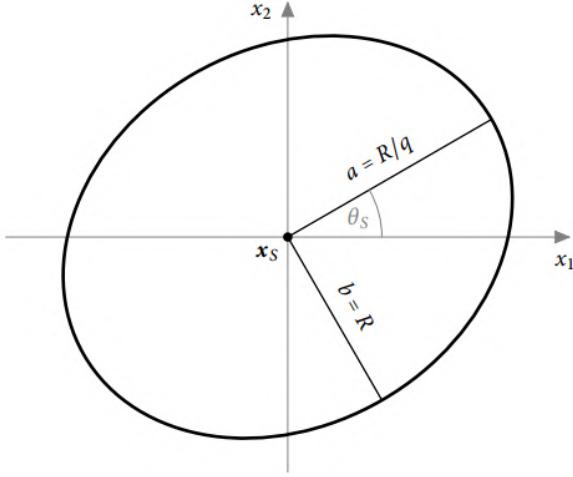


Figure 2.2: General ellipse from equation (2.12). \vec{x}_s is the center of the ellipse, whose notation in equation (2.12) is $\vec{\theta}_0$. θ_s is the orientation angle of the ellipse, whose notation in equation (2.12) is ϵ . Image from Tessore 2015.

where now θ is the *elliptical radius* or the semi-minor axis of the ellipse

$$r^2 = \theta^2 = q^2((\theta_1 - \theta_{1,0})\cos(\epsilon) + (\theta_2 - \theta_{2,0})\sin(\epsilon))^2 + ((\theta_1 - \theta_{1,0})\sin(\epsilon) - (\theta_2 - \theta_{2,0})\cos(\epsilon))^2 \quad (2.14)$$

where the elliptical angle is defined as

$$\epsilon = \arctan(\theta'_2/(q\theta'_1)) \quad (2.15)$$

and the relation between cartesian and polar coordinates (parametric representation) is

$$\begin{aligned} \theta'_1 &= a \cdot \cos(\phi)\cos(\epsilon) - b \cdot \sin(\phi)\sin(\epsilon) + \theta_{1,0} = \frac{r}{q} \cdot \cos(\phi)\cos(\epsilon) - r \cdot \sin(\phi)\sin(\epsilon) + \theta_{1,0} \\ \theta'_2 &= a \cdot \cos(\phi)\sin(\epsilon) + b \cdot \sin(\phi)\cos(\epsilon) + \theta_{2,0} = \frac{r}{q} \cdot \cos(\phi)\sin(\epsilon) + r \cdot \sin(\phi)\cos(\epsilon) + \theta_{2,0} \end{aligned} \quad (2.16)$$

where $\phi \in [0, 2\pi]$ is the polar angle.

To treat such lenses it is useful to use the complex notation.

The elliptical power-law profile deflection angle

The heart of this discussion is the complex deflection angle, thanks to which it is possible to calculate all the others important lensing quantities.

The starting point is the complex lens equation

$$z' = z - \alpha(z) \quad (2.17)$$

where z' is the complex source-plane position, z is the complex lens plane position and $\alpha(z)$ is the complex deflection angle. The above $z = \theta_1 + i\theta_2$ and $\alpha(z) = \alpha_1 + i\alpha_2$ replace two-dimensional vector quantities. Now, the complex conjugated deflection angle of an elliptical adimensional surface mass density is

$$a^*(z) = 2 \frac{\sqrt{z^2}}{z} \int_0^{R(z)} dr \frac{k(r)r}{\sqrt{q^2 z^2 - (1 - q^2)r^2}} \quad (2.18)$$

where $R(z)$ is the semi-minor axis of the ellipse from equation (2.14) (i.e. $R(z) = r(z)$, but $R(z)$ is used because r is the variable of the integral and we do not want to confuse) passing through the point z .

Inserting the power-law profile of equation (2.13) into the above

$$\begin{aligned}
a^*(z) &= \frac{2-t}{q} \frac{b^t}{z} \int_0^{r(z)} r^{1-t} \left(1 - \frac{1-q^2}{q^2} \frac{r^2}{z^2} \right) dr \\
&= \frac{2-t}{2q} \frac{b^2}{z} \left(\frac{b}{R(z)} \right)^{t-2} \int_0^1 x^{-t/2} \left(1 - \frac{1-q^2}{q^2} \frac{R^2(z)}{z^2} x \right)^{-1/2} dx \quad r \rightarrow x = r^2/R^2(z) \\
&= \frac{1}{q} \frac{b^2}{z} \left(\frac{b}{R(z)} \right)^{t-2} {}_2F_1 \left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; \frac{1-q^2}{q} \frac{R^2(z)}{z^2} \right)
\end{aligned} \tag{2.19}$$

where in the last step the integral representation of the Gaussian hypergeometric function ${}_2F_1(a, b, c, x)$ was used (2.2.1). Thanks to the relation between the parameters $c = a + b + 1/2$, the hypergeometric function can be transformed through a quadratic transformation, and the complex conjugated angle can be expressed through factorisation into radial and angular part as:

$$\alpha^*(r, \phi) = \frac{2b}{1+q} \left(\frac{b}{r(z)} \right)^{t-1} e^{-i\phi} {}_2F_1 \left(1, \frac{t}{2}; 2 - \frac{t}{2}; -\frac{1-q}{1+q} e^{-i2\phi} \right) \tag{2.20}$$

and the complex deflection angle is

$$\alpha(r, \phi) = \frac{2b}{1+q} \left(\frac{b}{r(z)} \right)^{t-1} e^{i\phi} {}_2F_1 \left(1, \frac{t}{2}; 2 - \frac{t}{2}; -\frac{1-q}{1+q} e^{i2\phi} \right) \tag{2.21}$$

where $\phi \in [0, 2\pi]$ is the polar angle in the complex plane.

The Function F is the *generalized hypergeometric function*.

The generalized hypergeometric function

A generalized hypergeometric function ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$ is a function that can be expressed in terms of an *hypergeometric series*, [[Generalized Hypergeometric Function n.d.](#)].

An *hypergeometric series* is a series $\sum_k c_k$ for which $c_0 = 1$ and the ratio of the consecutive terms can be expressed as

$$\frac{c_{k+1}}{c_k} = \frac{P(k)}{Q(k)} \tag{2.22}$$

where $P(k)$ and $Q(k)$ are two polynomials

$$\frac{c_{k+1}}{c_k} = \frac{P(k)}{Q(k)} = \frac{(k+a_1)(k+a_2)\dots(k+a_p)}{(k+b_1)(k+b_2)\dots(k+b_q)(k+1)} \tag{2.23}$$

The *generealized hypergeometric function* is

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = {}_pF_q \left[\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; x \right] = \sum_k^\infty c_k x^k = \sum_k^\infty \frac{(a_1)_k (a_2)_k \dots (a_p)_k}{(b_1)_k (b_2)_k \dots (b_p)_k} \frac{x^k}{k!} \tag{2.24}$$

where $(a)_k$ is the Pochhammer symbol

$$(a)_k \equiv \frac{\Gamma(a+k)}{\Gamma(a)} = a(a+1)\dots(a+k-1) \tag{2.25}$$

If $p = 2$ and $q = 1$ the above becomes the *Gaussian hypergeometric function*

$${}_2F_1(a, b, c, x) = \sum_k^\infty \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!} = \sum_k^\infty \frac{\Gamma(a+k) \Gamma(b+k) \Gamma(c)}{\Gamma(c+k) \Gamma(a) \Gamma(b)} \frac{x^k}{k!} \tag{2.26}$$

Where the parameters a , b and c are independent of x and, in general, they may be complex. The Gaussian hypergeometric function also has an integral form

$${}_2F_1(a, b, c, x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tz)^{-a} dt \tag{2.27}$$

For example, in equation (2.19), the factor $\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}$, with $a = 1/2$, $b = 1 - t/2$, $c = 2 - t/2$, is equal to

$$\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} = \frac{\Gamma(2-t/2)}{\Gamma(1-t/2)\Gamma(1)} = 1 - \frac{t}{2} \quad (2.28)$$

and so, thanks to equation (2.27), it is possible to express the (2.19) in terms of the Gaussian hypergeometric function.

The elliptical power-law lensing quantities

The others lensing quantities, such as the complex shear and the convergence, can be expressed as

$$\gamma = \gamma_1 + i\gamma_2 \rightarrow \gamma^*(z) = \frac{\partial \alpha^*}{\partial z} \quad (2.29)$$

and the convergence

$$k(z) = \frac{\partial \alpha^*}{\partial z^*} \quad (2.30)$$

From equation (2.29) the shear is

$$\gamma^*(z) = -k(z) \frac{z^*}{z} + (1-t) \frac{\alpha^*(z)}{z} \quad (2.31)$$

and, in terms of polar coordinates,

$$\gamma(r, \phi) = -e^{i2\phi} k(r) + (1-t)e^{i\phi} \frac{\alpha(r, \phi)}{r} \quad (2.32)$$

Now, with the expression for the shear and the convergence, from equation (1.45) it is possible to get μ^{-1}

$$\mu^{-1} = (1-k)^2 - |\gamma|^2 = 1 - 2k \left(1 - (1-t) \frac{\theta_1 \alpha_1 + \theta_2 \alpha_2}{r^2} \right) - (1-t)^2 \frac{|\alpha|^2}{r^2} \quad (2.33)$$

Slope and radial critical lines

Before ending the description of the elliptical power-law profile it is useful to point out some particular slope values and when the radial critical lines are present or not:

- For $t = 1$ we have the singular isothermal elliptical profile.
- For $t = 2$ we have the point mass.
- For $t \geq 1$ we do not have radial critical lines.

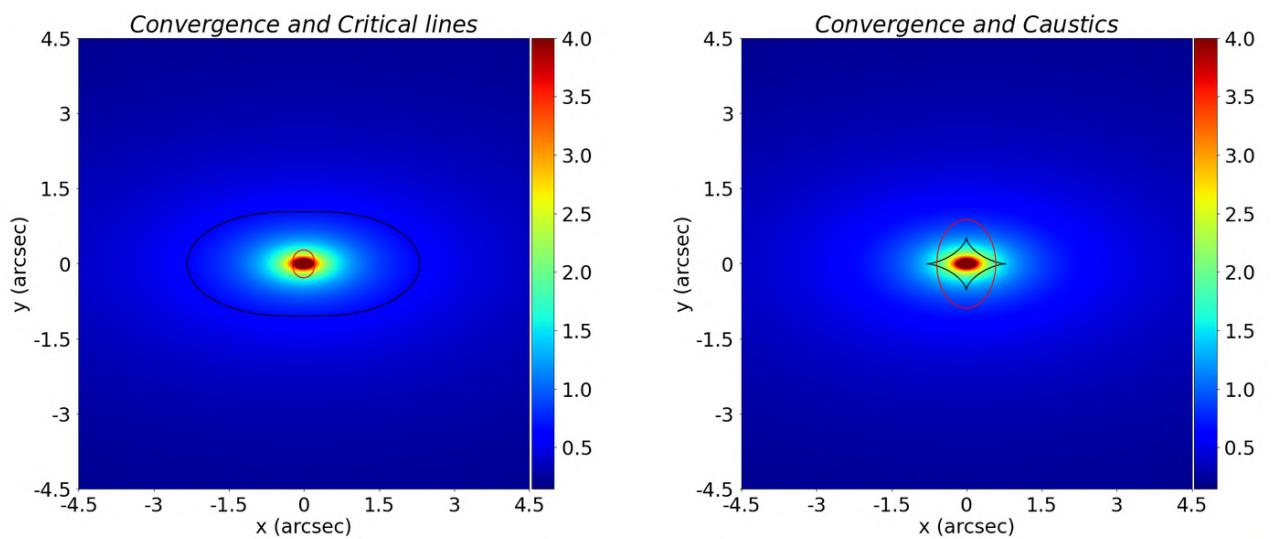


Figure 2.3: Elliptical Power-Law convergence (section (2.2.1)): Einstein radius=1.60'', axis-ratio=0.6, slope=0.75 . In the image on the left the angular critical lines are in black and the radial critical lines (slope < 2) are in red. On the left image the tangential caustic are in black and the radial caustics are in red.

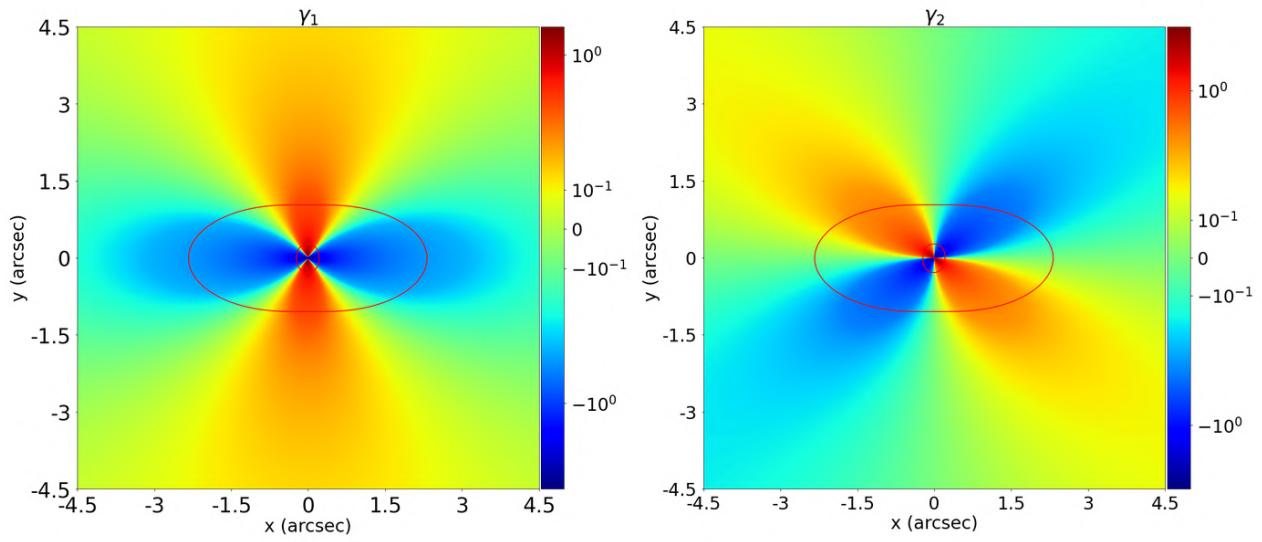


Figure 2.4: Elliptical Power-Law shear components (section (2.2.1)): on the left the γ_1 and on the right the γ_2 component.

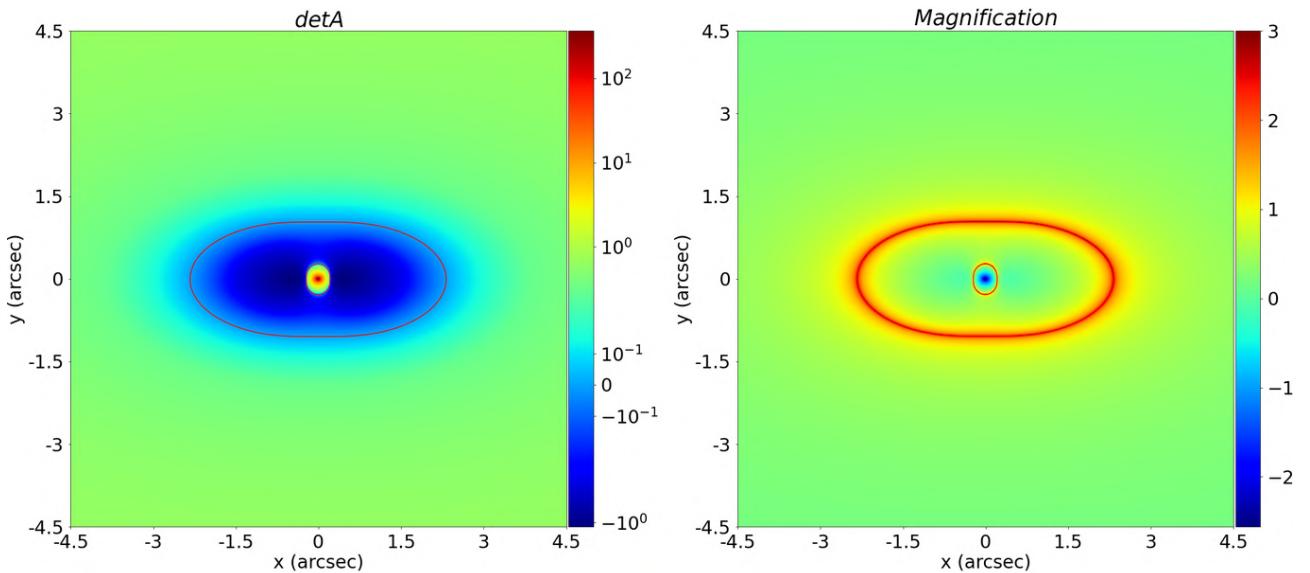


Figure 2.5: Elliptical Power-Law Lensing Jacobian determinant and magnification (section (2.2.1)): on the left the Jacobian determinant and on the right the Magnification.

2.2.2 The elliptical broken power law profile

The description will follow the article[O’Riordan, Stephen J Warren, and Daniel J Mortlock 2021].

In order to describe the broken power-law profile it is first necessary to briefly introduce the *truncated power-law profile*.

The elliptical truncated power-law profile

The convergence of the *elliptical truncated power-law profile* is described by a power-law within the elliptical truncation radius and it is zero outside. It is described as follow

$$k(\theta) = \begin{cases} k_T \left(\frac{\theta_T}{\theta} \right)^t & \theta \leq \theta_T \\ 0 & \theta > \theta_T \end{cases} \quad (2.34)$$

where θ_T is the elliptical truncation radius and k_T is the convergence normalization calculated at the truncation radius θ_T and t is the slope. The elliptical radius is the one defined in equation (2.14).

The slope $t = 1$ corresponds to the isothermal profile.

To give a definition for k_T it is first necessary to introduce the scale length b . It is a particular value of the elliptical radius whithin which the average density is equal to the critical density described in equation (1.32). The mass $M(\theta)$ is the mass enclosed by the elliptical radius θ , thus, if $\theta = b$ we have that

$$M(b) = \Sigma_c \pi \frac{b^2 D_L^2}{q} \quad (2.35)$$

where D_L is the lens distance from the observer.

In general the mass enclosed within a given elliptical radius θ is

$$M(\theta) = \frac{\Sigma_c D_L^2}{q} \int_0^\theta k(\theta') 2\pi \theta' d\theta' \quad (2.36)$$

Equation (2.34) establishes that beyond the truncation radius the convergence is zero, so it does not contribute to the mass, because it means adding a zero (the integral is nothing but a sum). Thus, the total mass is

$$M(\theta > \theta_t) = \frac{2\pi \Sigma_{crit} D_L^2 k_T \theta_T^2}{q(2-t)} \quad (2.37)$$

In order to find the K_T it is necessary to combine equation (2.35) and equation (2.36). There are two different cases:

- In the case which the scale length b is lower than the truncation radius θ_T we have that:

$$M(\theta = b) = \frac{\Sigma_c D_L^2}{q} \int_0^b k(\theta') 2\pi \theta' d\theta' = \frac{\Sigma_c D_L^2}{q} 2\pi k_T \theta_T \int_0^b \theta^{1-t} d\theta = \frac{\Sigma_c D_L^2}{q} 2\pi k_T \theta_T^t \frac{b^{2-t}}{2-t} \quad (2.38)$$

thus, equating the (2.35) and the (2.38) we have that

$$\Sigma_c \pi \frac{b^2 D_L^2}{q} = \frac{\Sigma_c D_L^2}{q} 2\pi k_T \theta_T \frac{b^{2-t}}{2-t} \rightarrow k_T = \frac{2-t}{2\theta_T^t/b^t} \quad (2.39)$$

- In the case which the scale length b is greater than the truncation radius θ_T , equation (2.35) and (2.37) must be used

$$\Sigma_c \pi \frac{b^2 D_L^2}{q} = \frac{2\pi \Sigma_{crit} D_L^2 k_T \theta_T^2}{q(2-t)} \rightarrow k_T = \frac{2-t}{2\theta_T^2/b^2} \quad (2.40)$$

Thus, we have that

$$k_T = \begin{cases} \frac{2-t}{2\nu^2} & \nu \leq 1 \\ \frac{2-t}{2\nu^t} & \nu > 1 \end{cases} \quad (2.41)$$

where $\nu = \theta_T/b$

It is now necessary to introduce the elliptical truncated power-law deflection angle, because it will be useful in the next section to derive the broken power-law deflection angle.

Using the equation (2.18) and inserting here the equation (2.34) we have

$$\alpha^*(z) = \frac{2k_T \theta_T^t}{qz} \int_0^{\theta(z)} \theta'^{1-t} \left(1 - \frac{1-q^2}{q} \frac{\theta'^2}{z} \right)^{-1/2} d\theta' \quad (2.42)$$

where

$$\theta(z) = \begin{cases} \theta(z) & \theta(z) \leq \theta_T \\ \theta_T & \theta(z) > \theta_T \end{cases} \quad (2.43)$$

and $\theta(z)$ is the radius in equation (2.14).

From equation (2.19), it follows that the integral can be expressed in terms of the Gaussian hypergeometric function (it is the same integral of equation (2.19)), and thus

$$I(\theta, t, z) = \frac{\theta^{2-t}}{2-t} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2}\right) \quad (2.44)$$

Thus, we have that

$$\alpha^*(z) = \frac{b^2}{qz} \cdot \begin{cases} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2}\right) \left(\frac{\theta_T}{\theta}\right)^{t-2} & \theta(z) \leq \theta_T \\ {}_2F_1\left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2}\right) & \theta(z) > \theta_T \end{cases} \quad (2.45)$$

The broken power-law profile

The convergence of the broken power-law profile is defined as

$$k(\vec{\theta}) = \begin{cases} k_B \left(\frac{\theta_B}{||\vec{\theta}||} \right)^{t_1} & \theta \leq \theta_B \\ k_B \left(\frac{\theta_B}{||\vec{\theta}||} \right)^{t_2} & \theta > \theta_B \end{cases} \quad (2.46)$$

where θ_B is the *brake radius*, k_B is the convergence at the brake radius, and t_1 and t_2 are respectively the slope within and outside the brake radius.

The isothermal case is given when $t_1 = t_2 = 1$ and when the break radius is near to 0.

To obtain the convergence k_B at the brake radius it is possible to follow the same reasonment of the previous section:

- If b (that retains the same meaning as before) is lower than θ_B , using equation (2.46) for $\theta \leq \theta_B$, we have that

$$M(\theta = b \leq \theta_B) = \frac{\Sigma_c D_L^2}{q} \int_0^b k(\theta') 2\pi \theta' d\theta' = \frac{\Sigma_c D_L^2}{q} 2\pi k_B \theta_T \int_0^b \theta^{1-t_1} d\theta = \frac{\Sigma_c D_L^2}{q} 2\pi k_B \theta_B^{t_1} \frac{b^{2-t_1}}{2-t_1} \quad (2.47)$$

and thus, equating (2.35) with (2.47) we have that

$$k_B = \frac{2-t_1}{2(\theta_B/b)^{t_1}} \quad (2.48)$$

- If the scale length b is greater than the brake radiu

$$\begin{aligned}
M(\theta = b > \theta_B) &= \frac{\Sigma_c D_L^2}{q} \int_0^b k(\theta') 2\pi \theta' d\theta' \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \theta_B^{t_1} \int_0^{\theta_B} \theta^{1-t_1} d\theta + \frac{2\pi \Sigma_c D_L^2}{q} k_B \theta_B^{t_2} \int_{\theta_B}^b \theta^{1-t_2} d\theta \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} k_B \theta_B^{t_2} \left[\frac{b^{2-t_2}}{2-t_2} - \frac{\theta_B^{2-t_2}}{2-t_2} \right] \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} \frac{k_B}{2-t_2} \left[b^2 \frac{\theta_B^{t_2}}{b^{t_2}} - \theta_B^2 \right] \\
&\text{with } \nu = \frac{\theta_B}{b} \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} \frac{k_B}{2-t_2} \left[b^2 \nu^{t_2} - \theta_B^2 \right] \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} \frac{k_B b^2}{2-t_2} \left[\nu^{t_2} - \nu^2 \right] \\
&= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} \frac{k_B b^2 \nu^2}{2-t_2} \left[\nu^{t_2-2} - 1 \right]
\end{aligned} \tag{2.49}$$

Now, equating the above with (2.36), we have that

$$\begin{aligned}
\frac{\Sigma_c \pi b^2 D_L^2}{q} &= \frac{2\pi \Sigma_c D_L^2}{q} k_B \frac{\theta_B^2}{2-t_1} + \frac{2\pi \Sigma_c D_L^2}{q} \frac{k_B b^2 \nu^2}{2-t_2} \left[\nu^{t_2-2} - 1 \right] \\
1 &= \frac{2\nu^2}{2-t_1} k_B + \frac{2k_B \nu^2}{2-t_2} \left[\nu^{t_2-2} - 1 \right] \\
1 &= \frac{2\nu^2}{2-t_1} k_B \left[1 + \frac{2-t_1}{2-t_2} \left[\nu^{t_2-2} - 1 \right] \right]
\end{aligned} \tag{2.50}$$

thus

$$k_B = \frac{2-t_1}{2\nu^2 \left[1 + \frac{2-t_1}{2-t_2} \left[\nu^{t_2-2} - 1 \right] \right]} \tag{2.51}$$

The broken power-law profile deflection angle

To obtain the deflection angle for this profile it necessary to use equation (2.42). The solution has the form of equation (2.44).

There are two cases:

- Inside the break-radius

$$\alpha^*(z) = \frac{2k_B}{2-t_1} \frac{\theta_B^2}{qz} \left(\frac{\theta_B}{\theta} \right)^{t_1-2} {}_2F_1 \left(\frac{1}{2}, 1 - \frac{t_1}{2}; 2 - \frac{t_1}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2} \right) \tag{2.52}$$

The reason why inside the break-radius the deflection angle is due only to the contribution of the internal power-law depends on the third Newton's shell theoreme.

Consider a family of concentric, coaxial, and homothetic ellipsoidal surfaces

$$\begin{cases} \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = m^2 & m \in \mathbb{R} \\ a \geq b \geq c \end{cases} \tag{2.53}$$

Taking a finite homeoid, that is, the region of space between two concentric, coaxial and homothetic ellipsoids as in equation (2.53), if this has a homogeneus density, we have that the gravitational field

inside the homeoid itself is 0. This means that if we have a heterogeneous homeoid, or in other words overlapped homeoids of different density $\rho = \rho(m)$, thanks to the superposition principle, the field in one specific point inside the heterogeneous homeoid depends only on the matter below that specific point, external homeoids do not contribute.

- Outside the break-radius:

$$\begin{aligned} \alpha^*(z) = & \frac{2k_B}{2-t_1} \left[\frac{\theta_B^2}{qz} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_1}{2}; 2 - \frac{t_1}{2}; \frac{1-q^2}{q} \frac{\theta_B^2}{z^2}\right) + \right. \\ & \frac{2-t_1}{2-t_2} \left[\frac{\theta_B^2}{qz} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_1}{2}; 2 - \frac{t_1}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2}\right) \left(\frac{\theta_B}{\theta}\right)^{t_2-2} - \right. \\ & \left. \left. \frac{\theta_B^2}{qz} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_2}{2}; 2 - \frac{t_2}{2}; \frac{1-q^2}{q} \frac{\theta_B^2}{z^2}\right) \right] \right] \end{aligned} \quad (2.54)$$

The deflection angle is given by the contribute of three components (superposition principle): the α for the inner power-law with slope t_1 truncated at θ_B from equation (2.45) with a slightly differnt normalization (first term of the sum), the α for a single power-law (section (2.2.1) with the slope t_2 and a slighly different normalization (second term) and then the contribution of the power-law with slope t_2 within the brake-radius must be removed (Newton's third shell theoreme), thus the deflection angle of a truncated power-law with slope t_2 is subtracted (third term).

It is interesting to analize the case in which $t_1 = t_2 = t$. Thwe normalizaion k_B in this case is

$$k_b = \frac{2-t}{2\nu^t} \quad (2.55)$$

and even the first and third term in equation (2.54) cancel out, so both the equations (2.52) and (2.54) become

$$\alpha^*(z) = \frac{b^2}{qz} \left(\frac{b}{\theta}\right)^{t-2} {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_1}{2}; 2 - \frac{t_1}{2}; \frac{1-q^2}{q} \frac{\theta^2(z)}{z^2}\right) \quad (2.56)$$

that is the deflection angle due to a single elliptical power-law profile and it is the same as in equation (2.19).

The broken power-law profile shear

The shear can be calculated starting from equations (2.52) and (2.54) using equation (2.29).

- Inside the broken-radius:

$$\gamma^*(z) = \frac{2K_B\nu^{t_1}}{2-t_1} \gamma_{t_1}^*(z) \quad (2.57)$$

where γ_{t_n} with $n = (1, 2)$ is

$$\gamma_{t_n}^*(z) = (1-t_n) \frac{\alpha_t^*(z)}{z} - k(z) \frac{z^*}{z} \quad (2.58)$$

and α^* is given by equation (2.56).

- To obtain the shear outside the break-radius it is necessary to combine equation (2.54) and (2.29).

$$\begin{aligned} \gamma^*(z) = & \frac{2k_B}{2-t_1} \left[\frac{\theta_B^2}{qz^2} \left[\frac{t_1-t_2}{\sqrt{1-\frac{1-q^2}{q^2}\theta_B^2/z^2}} \right. \right. \\ & + (1-t_1) {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_1}{2}; 2 - \frac{t_1}{2}; \frac{1-q^2}{q} \frac{\theta_B^2}{z^2}\right) \\ & \left. \left. - (1-t_2) {}_2F_1\left(\frac{1}{2}, 1 - \frac{t_2}{2}; 2 - \frac{t_2}{2}; \frac{1-q^2}{q} \frac{\theta_B^2}{z^2}\right) \right] \right] \\ & + \nu^{t_2} \frac{2-t_1}{2-t_2} \gamma_{t_2}^*(z) \end{aligned} \quad (2.59)$$

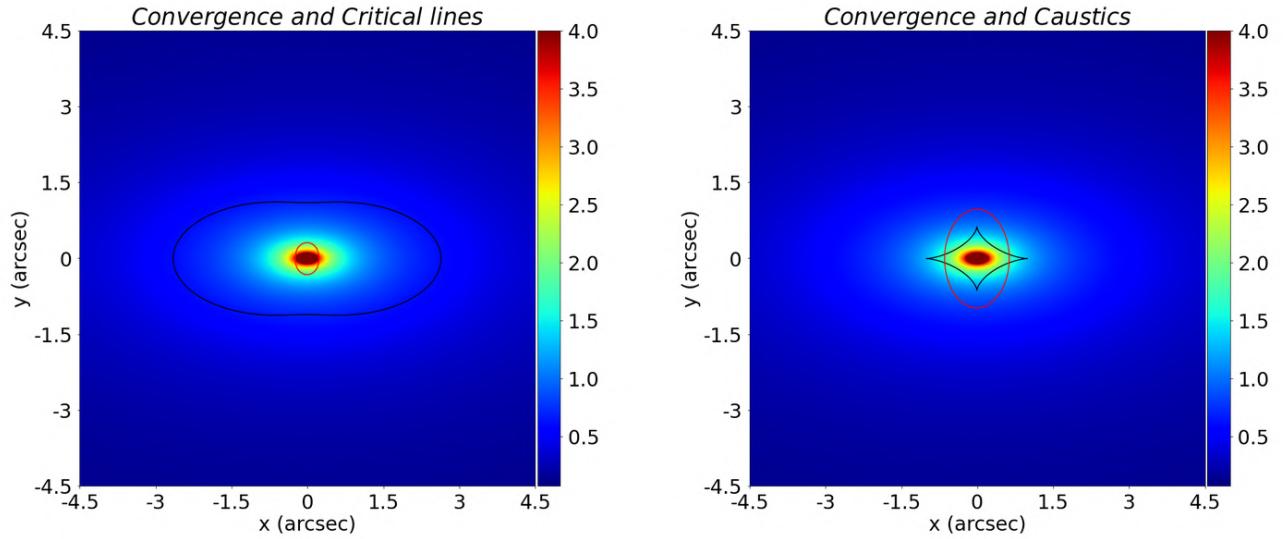


Figure 2.6: Elliptical Broken Power-Law convergence: Einstein radius=2.0'', axis-ratio=0.5, inner-slope=0.75, outer-slope=1.2, brake-radius=1.5 . In the image on the left the tangential critical lines are in black and the radial critical lines (slope < 2) are in red. On the left image the tangential caustic are in black and the radial caustics are in red.

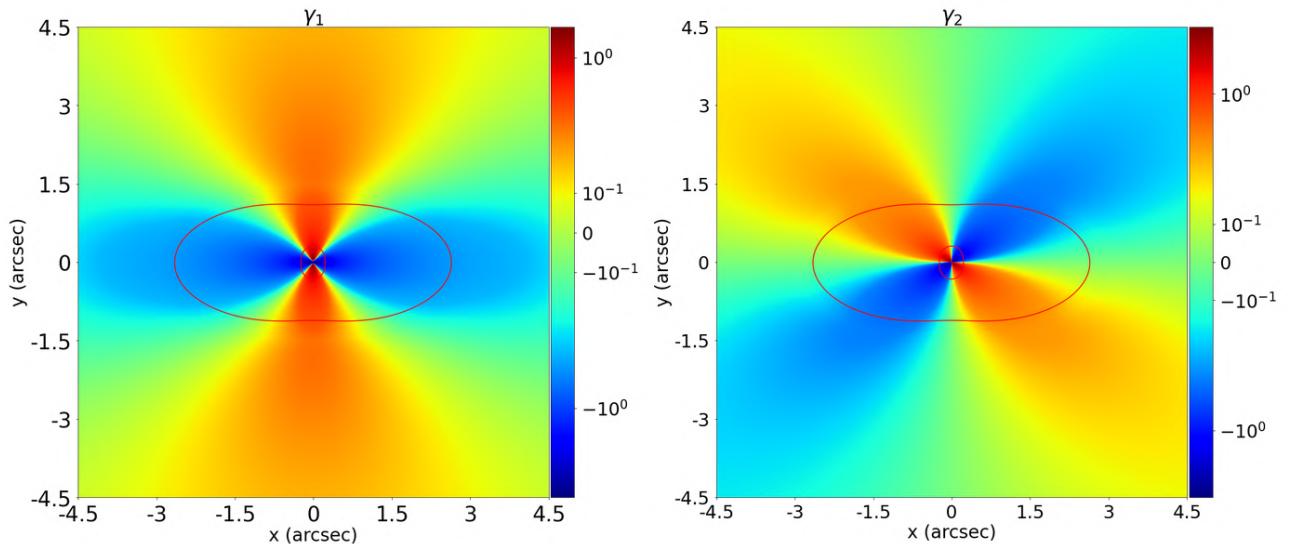


Figure 2.7: Elliptical Broken Power-Law shear components: on the left the γ_1 and on the right the γ_2 component.

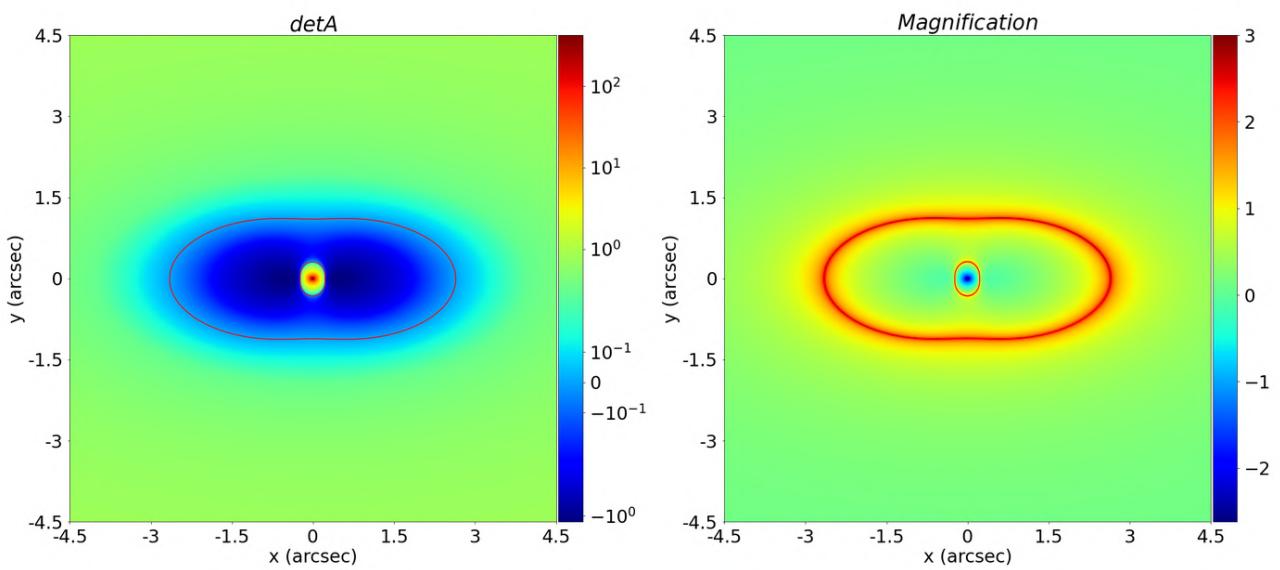


Figure 2.8: Elliptical Broken Power-Law Lensing Jacobian determinant and magnification: on the left the Jacobian determinant and on the right the Magnification.

2.3 Surface Brightness profile

In this section we will briefly analyze the surface brightness profile used in the first part of data analysis data analysis. In fact, as will be shown in the next chapters, a method will be used that makes use of parametric light sources, that is, described by numerical parameters. Both the lens and the source will be described using the so called textit{elliptical Sèrsic profile}.

2.3.1 The Elliptical Sersic Profile

The Sèrsic surface brightness profile is a parametric profile given by

$$S(\theta) = I_{eff} \cdot \exp \left\{ -b(n) \left(\frac{\theta}{R_{eff}} \right)^{\frac{1}{n}} - 1 \right\} \quad (2.60)$$

This profile is described by 7 parameters:

- The distance from the profile center is given by θ , that is the elliptical radius described by equation (2.14), thus from this we have three parameters
 - The two coordinates of the center $\theta_{1,0}$ and $\theta_{2,0}$
 - The ellipse orientation angle ϵ
 - The axis ratio q
- The effective radius R_{eff} , also known as the half-light radius and defined as the radius such that it contains half of the total brightness of the light source.
- The intensity at the effective radius, I_{eff} .
- The Sèrsic index n .

The parameter b is a function of n . To obtain its expression it is first necessary to describe the *luminosity* related to this profile.

The *luminosity* of any given elliptical surface brightness profile is given by

$$L(\theta) = \frac{2\pi}{q} \int_0^\theta \theta' S(\theta') d\theta' \quad (2.61)$$

where *theta* is the radius given by equation (2.14).

Two values of the luminosity are particularly interesting, the luminosity within the effective radius R_{eff} and the total luminosity.

The first, using the (2.61), is given by

$$L_{eff} = \frac{2\pi}{q} \int_0^{R_{eff}} \theta' S(\theta') d\theta' = \frac{2\pi}{q} R_{eff}^2 I_{eff} \frac{n}{b^{2n}} (\Gamma(2n) - \Gamma(2n, b)) \quad (2.62)$$

where $\Gamma(z)$ (in this case we have that z is an integer) is the gamma function and $\Gamma(z, a)$ (we have that z is an integer even in this case) is the incomplete gamma function.

The total luminosity is

$$L_{tot} = \frac{2\pi}{q} \int_0^\infty \theta' S(\theta') d\theta' = \frac{2\pi}{q} R_{eff}^2 I_{eff} \frac{n}{b^{2n}} \Gamma(2n) \quad (2.63)$$

As previously said R_{eff} is the half-light radius, thus, by definition, we have

$$2L_{eff} = L_{tot} \quad (2.64)$$

and therefore it is possible to derive the parameter b from the relation

$$\Gamma(2n, b) = \frac{\Gamma(2n)}{2} \quad (2.65)$$

though, this is a transcendental equation that must be solved numerically/grafigally or with the Poincarè asymptotic series technique

$$b(n) \approx 2n - \frac{1}{3} + \frac{4}{405n} \quad (2.66)$$

In addition, the total brightness of the galaxy can be derived from the magnitude of the object through the relation:

$$\frac{L_{object}}{L_{zp}} = 10^{-0.4(M_{object}-M_{zp})} \quad (2.67)$$

where L_{zp} is the zero-point luminosity, or in other words the reference luminosity for which the absolute bolometric magnitude $M_{zp} = 0$. L_{object} is the total luminosity of the object under analysis and M_{object} is the absolute bolometric magnitude of the object under analysis. Thus, we have

$$L_{object} = L_{zp} 10^{-0.4M_{object}} \quad (2.68)$$

Thanks to this, if the absolute bolometric magnitude of the object the numbers of parameters that describe the Sérsic profile reduce to 6, because for example the intensity at the effective radius is univocally determined by equation (2.63)

$$I_{eff} = \frac{L_{tot}}{\pi R_{eff}^2/q} \frac{b^{2n}}{\Gamma(2n+1)} \quad (2.69)$$

where the relation $\Gamma(x+1) = x\Gamma(x) \forall x \in \mathbb{R}$ was used with $x = 2n$.

Chapter 3

The Bayesian Approach

3.1 Model selection and parameters estimation through the bayesian approach

3.1.1 Bayes' Theorem and Bayesian approach

To select an appropriate model for a more accurate description of the data and to obtain the values of the model parameters, a Bayesian approach is employed.

The latter relies on the *Bayes Theorem*:

Theorem 1.

$$P(A|B) = \frac{P(B)P(B|A)}{P(B)} .$$

Now, supposing \mathbf{D} is a set of Data, and let M_i be one of a set of models $M_1, M_2, \dots, M_i, \dots, M_N$ that try to describe the data, where each of the previous M_i model is composed by a set of parameters $\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i$. By applying the the Bayes theorem (1):

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} \quad (3.1)$$

where:

- $P(M_i|D)$ is the *posterior probability* and it represents the probability the model M_i is effective in describing the data once they are known.
- $P(D|M_i)$ is the *likelihood probability* and it represents the probability of getting the observed data given the model M_i . It is denoted as $\mathcal{L}(D|M_i)$.
- $P(M_i)$ is the *prior probability* and it represents the probability associated to the model itself, before the data are known. It's denoted as $\pi(M_i)$.
- $P(D)$ is the *evidence* and it represents the probability of the data, regardless of the model. Since the data do not change, and the evidence is independent on the model, fixed the latter, it does not change. It is denoted as $\mathcal{E}(D)$.

Each of the M_i model is a function of a certain number of parameters

$$M_i = M_i(\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i) . \quad (3.2)$$

Thus, the (3.1) can be written:

$$\begin{aligned} P(\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i|D) &= \frac{\mathcal{L}(D|\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)P(\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)}{P(D)} \\ &= \frac{\mathcal{L}(D|\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)P(\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)}{\int_{\Omega_{\theta_i}} d^n \boldsymbol{\theta}^i \mathcal{L}(D|\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)P(\theta_1^i, \theta_2^i, \dots, \theta_j^i, \dots, \theta_k^i)} \end{aligned} \quad (3.3)$$

where the last passage is justified by :

$$P(D) = \sum_i P(D, M_i) = \sum_i P(D|M_i)P(M_i) . \quad (3.4)$$

considering all the parameters assume continuos range of value and integrating over the whole $i-th$ parameters set's domain Ω_{θ^i} .

3.1.2 Parameters estimation

One of the most common use of the Bayesian inference is the parameter estimation. In this case all the M_i models are the same model, with the same set of parameters $\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k$, though the parameters have different values.

Whit this in mind, the (3.3) becomes:

$$P(\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k | D) = \frac{\mathcal{L}(D|\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k)P(\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k)}{\int_{\Omega_\theta} d^n \theta \mathcal{L}(D|\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k)P(\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_k)} \quad (3.5)$$

The value of each parameters, can theoretically be obtained by an optimization procedure, in this case maximizing the *posterior* in (3.5).

Though, in the case of a multivariant posterior with high dimensionality, i.e. with a great numbers of parameters, the optimization procedure may become to hard to be computed analitically or even impossible to be computed analitically, making it compulsory to use a numerical approach. The latter is the case that concerns this thesis and the numerical approach to adress this problem will be explained in section (3.2.1).

3.1.3 Model selection

Suppose the method to find parameters works and provides the values for the whole set of parameters, what gives the certainty that the values found are those that are most supported by the data? In fact, Bayesian analysis is somewhat relative, suffering from a kind of 'systematic error', in the sense that it is always possible to obtain an answer for the problem, even though the model is a wrong one. This affects even the method used to deal with parameters estimation.

In order to deal with this problem it is necessary to compare the models and define a method that allows to exclude the one that is not correct.

Suppose the set of different models M_i in section (3.1.1), in general they do not just differ from each other by having different values of parameters, they might have completely different set of parameters or some of the models might be nested one within the other, gradually becoming more complex by adding parametersto the previous.

Suppose two different models M_1 and M_2 are provided in order to describe the same set of data D , write down the ratio of the respective posterior using the equation (3.1) :

$$O_{1,2} = \frac{P(M_1|D)}{P(M_2|D)} = \frac{\frac{P(D|M_1)P(M_1)}{P(D)}}{\frac{P(D|M_2)P(M_2)}{P(D)}} = \frac{P(D|M_1)P(M_1)}{P(D|M_2)P(M_2)} = \frac{P(M_1)}{P(M_2)} \cdot B_{1,2} \quad (3.6)$$

where $O_{1,2} = \frac{P(M_1|D)}{P(M_2|D)}$, the ratio of the two posteriors, is known as the *odds*; $B_{1,2} = \frac{P(D|M_1)}{P(D|M_2)}$, the ratio of the two *likelihood* $\mathcal{L}(D|M_i)$, is called the *Bayes' factor*. The $P(D)$, the probability of having the data given all the possible models, cancels out becuse the data are always the same, and this also avoids the need of knowing a priori the probabilities of the data (the *evidence* $\mathcal{E}(D)$).

It is possible to express the Bayes' theorem with the model conditionality (all the different parameters belonging to one model) explicitly shown, or in other words the Bayes' theorem for a single model with the explicit dependence on its parameters. Simply speaking, in equation (3.6) the probabilities related to a specific model take into account the informations of all set of parameters belonging to a specific model, because in the model comparison all the informations about a specific model are needed. Whereas, when a specific model is selected

and its parameters (the best values for these) are to be found, it is necessary to make the dependency on the parameters explicit:

$$P(\boldsymbol{\theta}|D, M) = \frac{P(D|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)}{P(D|M)} \quad (3.7)$$

Where the term $P(D|M_i)$ can be written as:

$$P(D|M_i) = \int_{\Omega_{\theta^i}} d^n \boldsymbol{\theta}^i P(D|\boldsymbol{\theta}, M_i)P(\boldsymbol{\theta}|M_i) = \mathcal{E}(D|M_i) \quad (3.8)$$

So, the term $P(D|M)$ in equation (3.6) is the evidence for each model, and the Bayes' factor is the ratio of evidences for two different models.

It is not strange that the evidence in the case of parameters selection (eq. (3.8)) represents the textit{likelihood} in case of model selection (eq. (3.6)), because, as said just above, the probability of having the data given a specific model requests all the possible informations about that model, or, in other words, the knowledge of all the possible values for its parameters, hence the evidence in (eq. (3.8)). The problem is how to evaluate the $\mathcal{E}(D|M_i)$. In fact, as previously said, the *likelihood* is in general a complex fuction of many different parameters, and the integral above can be very difficult to evaluate. The solution to this problem is given in section (3.2.1). The above is a general approach to model selection that allows to discriminate between two different models (they do not just differ from each other by having different values of parameters), though, in this work the interest concerns a particular case, that is finding parameters value. Hence, here the case of different models is the same as the one in section (3.1.2), the discussion is about the same model but different parameters values, i.e. the purpose is to check which of the parameters' set best describes the data. This procedure is of fundamental importance for the numerical algorithm described in the following section.

3.2 Numerical sampling methods

3.2.1 Numerical Sampling Method: Introduction

In the previous section was highlightend the problem that arises when one try to optimize the posterior or even to integrate the evidence $\mathcal{E}(D|M_i)$ in equation (3.8). This is due to the fact that in general the dimension of parameter space is high or there is not an analytical solution.

In order to adress this problem it is possible to use the *Numerical sampling method*.

The idea consists in the use of the *Monte carlo Method*, a numerical algorithm that allows to perform random repeated sampling in order to estimate a possible solution for a specific problem, such as the optimization or the integration above.

The Monte Carlo method works as follow:

- Set up the model (for example the posterior in equation (3.5) or the integral in equation (3.8)), define the dependent variables that must be predicted (or evaluated in the two cases above) and the independent variables that will drive the process.
- Specify the probability distribution for the independent variables. For the problem that is being analyzed the probability distributions that are exploited are the *posteriors* $\pi(\boldsymbol{\theta})$.
- Run multiple simulations, sampling as many random values as needed for the independent variables from the probability distribution in point two, and for each iteration perform the operations of the model under analysis (such as optimization or integration).
- Stop the iteration procedure when the ending threshold is reached.

3.2.2 Nested Sampling

Nested Sampling is a numerical sampling Monte Carlo integration based method, first developed by [Skilling 2004], used to solve Bayesian inference problems, primarily applied for estimating the *evidence* $\mathcal{E}(D|M_i)$ in equation (3.8), that can be rewrite as:

$$Z = \int_{\Omega_{\theta}} d^n \boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (3.9)$$

The basic idea is to approximate the evidence by integrating the prior in nested shells of constant likelihood (Figure (3.1)). Moreover, nested sampling allows even to simultaneously estimate the both the evidence and the posterior.

Before exploring the algorithm itself, it is necessary to describe a series of mathematical steps that allow to rewrite the equation (3.9).

The integration domain \mathcal{D} can be written in a manner that is normal to λ or normal to $\boldsymbol{\theta}$. Respectively:

$$\mathcal{D} = \mathcal{D}_\lambda = \{(\boldsymbol{\theta}, \lambda) | \lambda \in [0, \infty[, \mathcal{L}(\boldsymbol{\theta}) \geq \lambda\} \quad (3.10)$$

$$\mathcal{D} = \mathcal{D}_{\boldsymbol{\theta}} = \{(\boldsymbol{\theta}, \lambda) | \boldsymbol{\theta} \in \Omega, 0 \leq \lambda \leq \mathcal{L}(\boldsymbol{\theta})\}. \quad (3.11)$$

Suppose now to have the two functions, $f(\boldsymbol{\theta})$ and $g(\lambda)$, and their integral over the above domain:

$$\int_{\mathcal{D}} f(\boldsymbol{\theta}) g(\lambda) d\boldsymbol{\theta} d\lambda. \quad (3.12)$$

Thanks to the Fubini's theorem (then order of integration can be exchanged) it's possible to exchange the integration domain in equations (3.10) and (3.11):

$$\begin{aligned} \int_{\mathcal{D}} \{(\boldsymbol{\theta}) g(\lambda) d\boldsymbol{\theta} d\lambda &= \\ \int_0^\infty \left(g(\lambda) \int_{\{\boldsymbol{\theta} \in \Omega | \mathcal{L}(\boldsymbol{\theta}) \geq \lambda\}} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) d\lambda &\stackrel{\text{F. theo.}}{=} \\ \int_\Omega \left(f(\boldsymbol{\theta}) \int_0^{\mathcal{L}(\boldsymbol{\theta})} g(\lambda) d\lambda \right) d\boldsymbol{\theta} \end{aligned} \quad (3.13)$$

The first step consists in defining the so called *cumulative prior mass function*:

$$X(\lambda) = \int_{\Omega_{\boldsymbol{\theta}: \mathcal{L}(\boldsymbol{\theta}) > \lambda}} \pi(\boldsymbol{\theta}) d^n \boldsymbol{\theta} \quad (3.14)$$

$X(\lambda)$ is the fraction of the prior with likelihood greater than a certain likelihood value $\lambda = \mathcal{L}(\boldsymbol{\theta})$, or, in other words, the fraction contained within the iso-likelihood contour $\lambda = \mathcal{L}(\boldsymbol{\theta})$, and, given that $X(\lambda)$ is a cumulative function of the prior, it takes value in the range $[0, 1]$, where it is 1 over the entire parameters' space. Moreover, as λ increases $X(\lambda)$ decreases from 1 to 0, because the domain $\Omega_{\boldsymbol{\theta}}$ is getting more and more shrunked.

Hence, using the equation (3.13) and considering the case where $f(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})$ and $g(\lambda) = 1$, the integration of the evidence Z , that demands a multi-dimensional integral, can be rearranged into one dimensional integration, writing the (3.9) using the new variables:

$$\begin{aligned} Z &= \int_{\Omega_{\boldsymbol{\theta}}} d^n \boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) = \\ \int_\Omega \left(\pi(\boldsymbol{\theta}) \int_0^{\mathcal{L}(\boldsymbol{\theta})} 1 d\lambda \right) d\boldsymbol{\theta} &\stackrel{\text{F. theo.}}{=} \\ \int_0^\infty \left(\int_{\{\boldsymbol{\theta} \in \Omega | \mathcal{L}(\boldsymbol{\theta}) \geq \lambda\}} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) d\lambda &\stackrel{(3.14)}{=} \\ \int_0^\infty X(\lambda) d\lambda \end{aligned} \quad (3.15)$$

where the range of integration extends from 0 to ∞ considering that the likelihood (when it is not normalized) has an upper limit of ∞ . Furthermore, as far as $\mathcal{L}(X)$, the inverse of $X(\lambda)$, exists, the above integral can be transformed into

$$Z = \int_0^1 \mathcal{L}(X) dX, \quad (3.16)$$

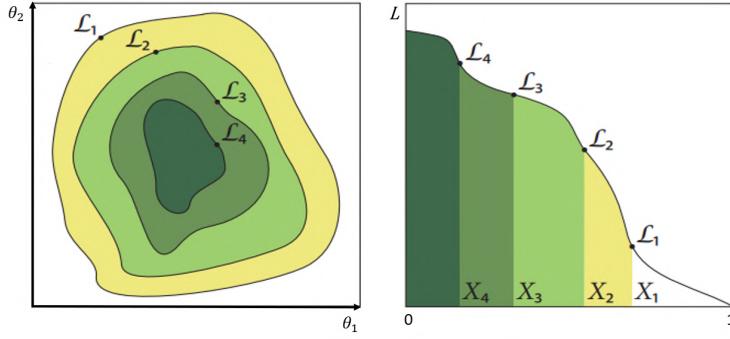


Figure 3.1: In the left image is illustrated the two dimensional integration of $X(\lambda)$ whithin different iso-likelihood contours. Between every iso-likelihood contour X remains constant (approximation). In the right image is illustrated the transformation from the two-dimensional problem to the one-dimensional one.[Feroz et al. 2019].

given that $X(\lambda)$ ranges from 0 to 1. A two-dimensional $\boldsymbol{\theta} = (\theta_1, \theta_2)$ example of this procedure, taken from [Feroz et al. [2013]], is illustrated in figure (3.1).

The algorithm begins by sampling N points from a the

Evaluating the iso-likelihod contours $\mathcal{L}_i = \mathcal{L}(X_i)$ on a sample of cumulative prior mass function

$$0 < X_N < X_{N-1} < \dots < X_2 < X_1 < 1 , \quad (3.17)$$

it is possible to evaluate the (3.16) using a numerical integration algorithm.

The evidence in equation (3.16) can be estimated through:

$$\mathcal{E} \approx \sum_{n=1}^M \mathcal{L}_n v_n \quad (3.18)$$

where M is the total number of cycles used and

$$v_n = \frac{1}{2}(X_{n-1} - X_{n+1}) \quad (3.19)$$

where the above difference is justified by the relation (3.17) ($X_{n-1} > X_{n+1}$) .

The process starts by sampling N random samples from the prior $\pi(\boldsymbol{\theta})$, the so called *Live point*

$$(\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_N) , \quad (3.20)$$

using a random generator or a *Markov Montecarlo Chain (MCMC)*. Afterwards, the likelihood is evaluated at each of these N points and they are sorted following the next one:

$$\mathcal{L}(\boldsymbol{\theta}_1), \mathcal{L}(\boldsymbol{\theta}_2), \dots, \mathcal{L}(\boldsymbol{\theta}_k), \dots, \mathcal{L}(\boldsymbol{\theta}_N) \quad (3.21)$$

where the index k does not follow the same order of the index j in relation (3.20). Hence, the question is how to evaluate the X_i . The idea is to model every X_i in the following way:

$$X_i = t_i X_{i-1} \quad (3.22)$$

setting $X_0 = 1$, and supposing that, for a given number n of *live points*, each shrinkage ratio it is independently distributed as the largest of n random variables from the interval $[0,1]$. The distribution of t is then:

$$p(t|N) = Nt^{N-1} \quad (3.23)$$

The next step involves a rearranging of the PDF above, performing a change of variables, more in particular it is necessary to find the distribution of $y = f(t) = \ln(t)$.

$$\begin{aligned} p(y|N) &= p(f^{-1}(y)) \left| \frac{d(f^{-1}(y))}{dy} \right| \\ &= N(e^y)^{N-1} (e^y) \\ &= Ne^{Ny} \end{aligned} \quad (3.24)$$

where now $0 < e^y = t < 1 \rightarrow \ln(0) < y < \ln(1) \rightarrow -\infty < y < 0$. The expectation value of this distribution $E[y = \ln(t)]$ is given by:

$$E[y = \ln(t)] = \int_{-\infty}^0 y \cdot p(y|N) dy = \int_{-\infty}^0 y \cdot N e^{Ny} dy = -\frac{1}{N} \quad (3.25)$$

The variance of this distribution is given by:

$$E[y^2 = (\ln(t))^2] = \int_{-\infty}^0 y^2 \cdot p(y|N) dy = \int_{-\infty}^0 y^2 \cdot N e^{Ny} dy = \frac{2}{N^2} \quad (3.26)$$

and so

$$\sigma = \sqrt{Var[y = \ln(t)]} = \sqrt{E[y^2 = (\ln(t))^2] - E[y = \ln(t)]^2} = \frac{1}{N} \quad (3.27)$$

It is now possible to approximate $\ln(t)$ in the following way:

$$\ln(t) \approx E[y = \ln(t)] \pm \sigma \quad (3.28)$$

and so

$$t \approx e^{(E[y = \ln(t)] \pm \sigma)} = e^{(-\frac{1}{N} \pm \frac{1}{N})} \approx e^{-\frac{1}{N}} \quad (3.29)$$

Supposing the number of live points remains constant all over the sampling procedure and that the shrinkage ratio t_i is the same and given by the (3.29) for each nested shell, and so:

$$t_i = t = e^{-\frac{1}{N}} \quad \forall i \quad (3.30)$$

$$\begin{aligned} X_i &= t_i X_{i-1} = t_i t_{i-1} X_{i-2} = \dots = t_i t_{i-1} \dots t_1 X_0 \\ &= \left(\prod_{j=1}^i t_j \right) \cdot X_0 = t^i \cdot X_0 = e^{-\frac{i}{N}} \cdot X_0 = e^{-\frac{i}{N}} \end{aligned} \quad (3.31)$$

recalling that $X_0 = 1$.

Since all the elements are defined, it is time to integrate (3.18) through the Monte Carlo technique in section (3.2.1):

- Extract another sample from the prior $\pi(\boldsymbol{\theta})$, compare its likelihood $\mathcal{L}(\boldsymbol{\theta}_{extracted})$ with the previous values in relation (3.21) and accept it only if its likelihood is larger than the minimum found in relation (3.21). Repeat this procedure until a good one that satisfy this constrain is found.
- Once a good point is found, discard the $\boldsymbol{\theta}_{lowest}$ in (3.21), add the new good point to the list and resort the whole sample.
 - This point is one of a series of M discarded points that are known as *dead point*. Each of these *dead points* in each iteration correspond to a specific X_i , in this case $\boldsymbol{\theta}_l$ corresponds to X_m , where m is the current step of the Monte Carlo iteration: at each iteration the volume X , relying on the (3.31), is shrunked to $e^{-\frac{m}{N}}$, because m live points were substituted following the constrain on the likelihood. In more simple terms it is the volume interior with respect to the dead point $\boldsymbol{\theta}_l$. The m -th term of the likelihood \mathcal{L}_m that appears in the summation (3.18), corresponding to the X_m volume, is evaluated at the $\boldsymbol{\theta}_l$ just discarded.
 - It is not necessary to sample N new points at each iteration, because $N - 1$ points, except the point with lowest \mathcal{L} (highest X), of the previous sample still satisfy the restriction condition $\mathcal{L}(\boldsymbol{\theta}_k) > \mathcal{L}(\boldsymbol{\theta}_1)$ and can be used in current step.
- Evaluate (3.19).
- Update (3.18)

- Repeat the whole procedure M times until the threshold is reached. The simplest way to find the threshold consists in examining the current addition term (at the iteration m) to the evidence at the iteration m (the sum from the previous $m - 1$ iterations). It begins by setting the follow

$$\Delta\hat{\mathcal{E}}_m = \mathcal{L}_{max}X_m \quad (3.32)$$

where $\Delta\hat{\mathcal{E}}_m$ is the last element that should be added to the summation in (3.18) at the iteration m , \mathcal{L}_{max} is the maximum likelihood contained within the remaining set of N live points at the $m - th$ iteration and X_m is the volume element at the $m - th$ iteration. The threshold is found using the above:

$$\Delta log(\hat{\mathcal{E}}_m) = ln(\hat{\mathcal{E}}_{m-1} + \Delta\hat{\mathcal{E}}_M) - ln(\hat{\mathcal{E}}_{m-1}) \quad (3.33)$$

where the value of $\Delta log(\hat{\mathcal{E}}_m)$ is set up arbitrarily. If the term $\Delta\hat{\mathcal{E}}_m$ is very small compared to the current evidence \mathcal{E}_{m-1} at the iteration m , which means the following iterations are not likely to contribute significantly to the growth of \mathcal{E} , then the loop is going to be terminated.

Eventually, the algorithm produces a set of M dead points

$$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k, \dots, \boldsymbol{\theta}_N), \quad (3.34)$$

ordered by the likelihood constrain previously established, where $\boldsymbol{\theta}_1$ corresponds to the lowest likelihood value and $\boldsymbol{\theta}_N$ to the highest one. Thanks to these, it is possible to evaluate the posterior in (3.5) and find the maximum one. In fact, the posterior is evaluated at each one of these dead points, and, since the evidence is the same for each of the m dead points (the integral over the parameters domain makes the evidence independent on the parameters, i.e. a marginalization procedure), the posterior depends only on the likelihood \mathcal{L} and the prior π .

$$P_k(\boldsymbol{\theta}_k) \approx \frac{\mathcal{L}(\boldsymbol{\theta}_k)\pi(\boldsymbol{\theta}_k)}{\mathcal{E}} \quad (3.35)$$

The maximum of the posterior can be found iterating the relation (3.6) over the set of M dead points.

Chapter 4

PyAutolens Features

PyAutolens is an open-source Python-based software, aimed at analyzing strong lensing events [J. Nightingale et al. 2021]. In this chapter we will analyze some features of this software, necessary for data analysis. The description is based on what is taught in the PyAutolens tutorials [[PyAutolens Tutorials 2022](#)].

4.1 The log-likelihood for parametric models

The purpose of this section is to illustrate how the log-likelihood function (the logarithm of the function described in the previous chapter) is calculated in the various fitting procedures.

The first important step is to load the data we fit and then to define the so-called fitting mask within which the fit will be performed.

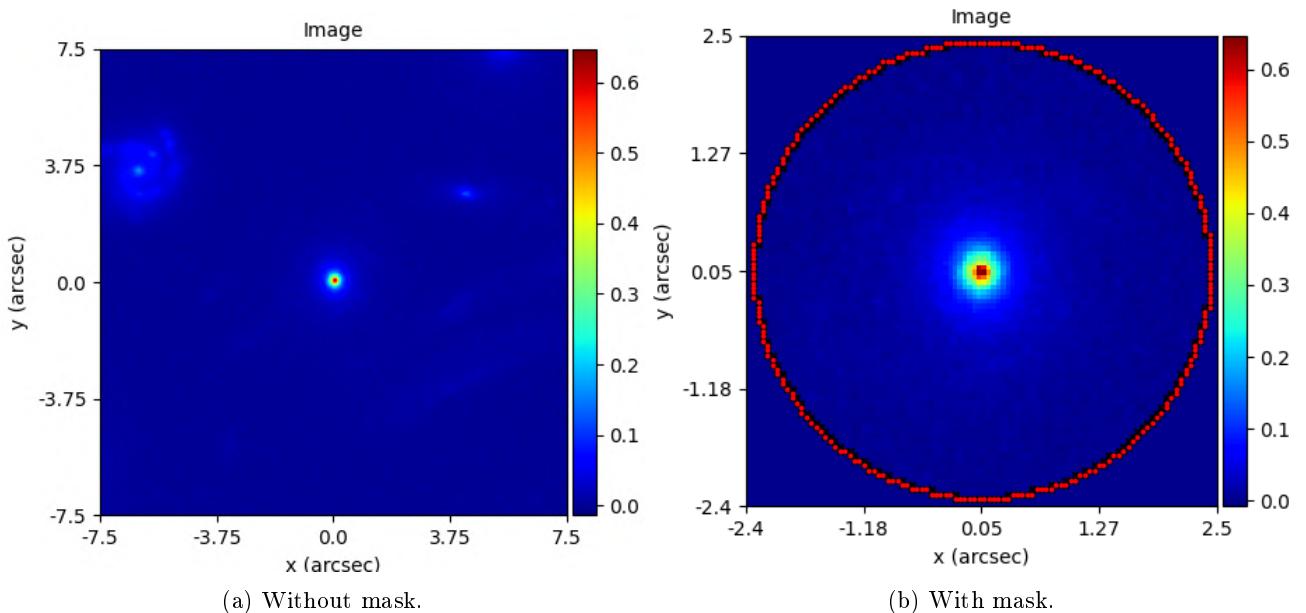


Figure 4.1: Masking procedure. The fit is performed using only the pixel within the mask. On the left, the initial image without mask, on the right, the galaxy, that is the one in the center of the left image, enclosed in the mask.

In fact, the log-likelihood function is calculated exclusively using the image pixels within this circular region, which is defined before each fit, providing the software with the dimensions of the radius of this region. Now that we have the data, in order to calculate the log-likelihood function it is necessary to assume a model

and apply it to the data. Thus, the log-likelihood function for a parametric model is defined as

$$-2\log\mathcal{L} = \chi^2 + \sum_{j=1}^J \ln[2\pi(\sigma_j)^2] \quad (4.1)$$

So, what are the terms that make up the log-likelihood?

The first term in the r.h.s is the χ^2 . The χ^2 is made up by different terms:

- The first is the *model data*. Once we have the parametric model, the one we think best describes the data, we need to be able to compare it with the real data we get from the telescope. To do this, we generate the theoretical profile is generated and once we have this, it is convolved with the *Point Spread Function* (PSF) profile due to the telescope's optics. As what is observed passes through the optics of the telescope, that distort the intrinsic profile.
- The second element is the *residual map*. This consists in the difference between the true data and the model data described just above.
- The third term is the *normalized residual map*, that consists in the above residual map divided by the *noise map*. (riquadro ... in figura).
- The chi-square map consists in the square of the normalized residual map, that represents what regions are fitted accurately and those that are not.
- Eventually, the χ^2 is the sum of each pixel-value of the chi-square map above.

The second term on the r.h.s of the equation (4.1) is the *noise normalization term*.

The log-likelihood function implemented in the software assumes the imaging data consists of independent Gaussian noise in every image pixel, thus, the *noise normalization term* consists in the sum of the log of every noise-map pixel value squared. Given the noise map is fixed, this term does not change during the lens modeling process and has no impact on the model we infer.

The complete description with some example of this log-likelihood function is in the [log-likelihood description](#), [PyAutolens](#).

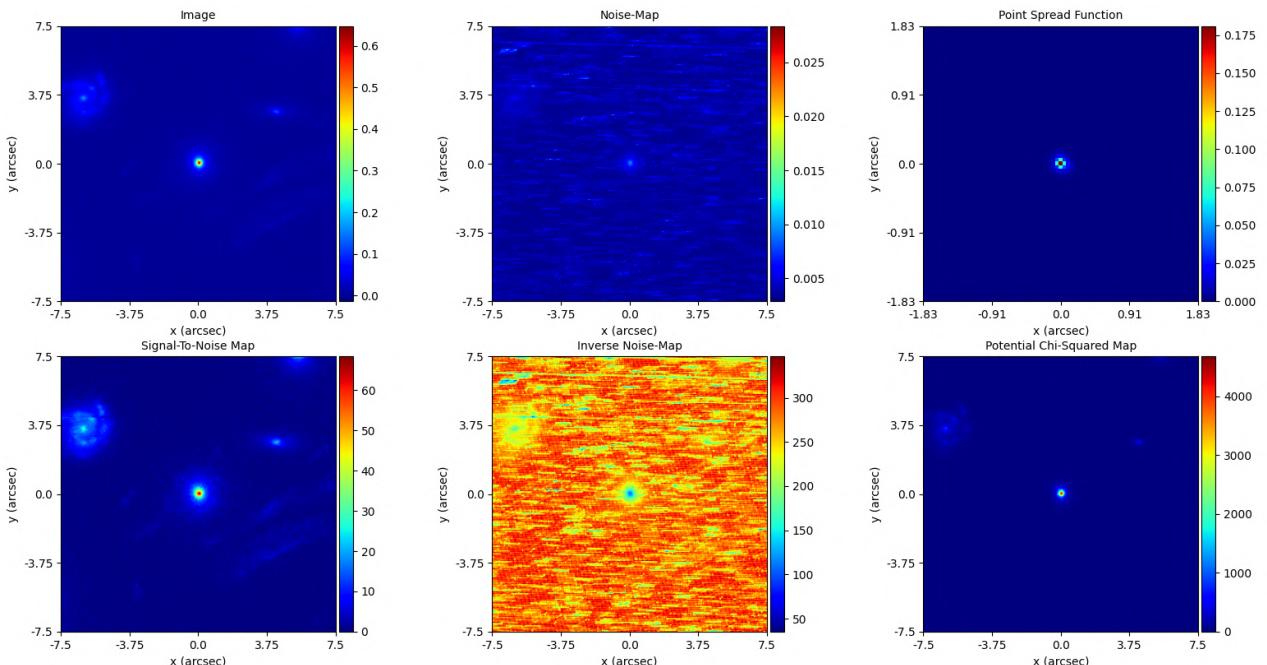


Figure 4.2: The various elements that make up the likelihood

4.2 Multiple image position setting

Before the fitting procedure begins, we can provide the software with additional information to help it to move through the parameter space during its non-linear search of the values of the parameters that can best describe the mass-profiles that constitute the system under examination.

In fact, it is possible to manually specify a set of image-plane coordinates, that correspond to the multiple images of a lensed source-galaxy. If we supply these positions, every time a lens model is fitted, the software will first check that these pixels trace within a specified arc-second threshold of one another in the source-plane (which is controlled by an input position threshold parameter). If they do not trace within this threshold of one another, the log-likelihood of this lens model is penalized and the model is discarded and the software samples a new lens model. In other words: the software looks for the source's multiple images supplied by hand in the lens plane (these are the first pixels it looks to), than it perform ray tracing from these images back to the source plane. The rays of the same multiple images must fall within a certain threshold (the same limited region) in the source plane, because they are the same point in the source plane

If they do not do, the model must be discarded.

This has two consequences:

- The model-fit is faster as the non-linear search avoids searching regions of parameter space where the mass-model is clearly not accurate (given that models which do not trace the source's brightest image-plane pixels close to one another in the source plane can not fit the data accurately).
- Removing these solutions means removing possible local maxima in the parameter space and giving the software the possibility to reach absolute maxima. This is because removing the incorrect mass models makes the non-linear parameter space less complex, i.e. it removes some regions of the latter and the software will not look there anymore.

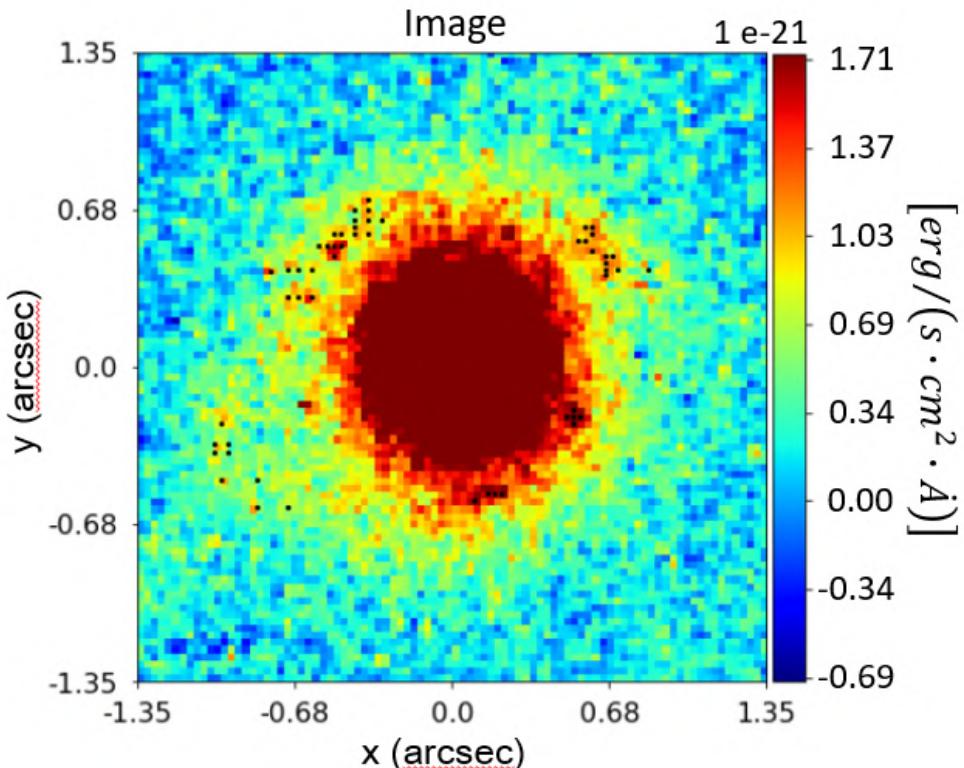


Figure 4.3: The black dots represent the multiple images that are given as input to the software.

4.3 Search chaining and prior passing

One of the great potentiality of PyAutoLens consists in the possibility of chaining subsequent fitting procedures, where each new non-linear search begins to look for the parameters in the parameter space starting from the best solution of the previous search (the refining procedure). Thanks to this chaining procedure, starting from very simple solutions, it is possible to build complex models that can more accurately describe the data. In fact, starting from simple solutions (that can be an over-simplification of what is really observed), search chaining eases the non-linear search in parameters space. At the beginning, the model will not be the best one, however, in the last fits, degrees of complexity will be added to the latter, creating the so-called nested-model, where the simplest model is enclosed in the most complex one, which should be able to accurately describe the observed data. The advantage of using nested-model and search chaining together consists in the creation of the so-called hyper-images, or models that, thanks to the previous fits, contain information on both the lens and the source. For example, the coordinates of the centers in their respective planes, the location of multiple images, and so on. These are of great help, because they allow the software to move in the new parameter space, which now has additional dimensions. However, when data analysis will be presented, these hyper-images will be clearer. Usually, at the second to last fit of the chain, during which the same model is always used, we get the most refined version of the latter and therefore the information used to build the hyper-image, before adding parameters and creating the most complex final model. As mentioned above, in search-chaining the best results of the previous search are the starting point of the next search. In order to transmit this information, the prior function, always described in the previous chapter, is used. The prior function can be seen as the set function for a given parameter, that is, it establishes what should be the range of values within which to look for the value of a given parameter. In this case the results of previous non-linear search are passed to the new fit through the *gaussian prior*, this means that the function that constitutes the prior is a gaussian, with a given mean, sigma, an upper and lower limit.

Now, how does exactly the gaussian prior passing work?

- The mean, where the gaussian is centered, corresponds to the best fit values of the previous search, in particular the median PDF model (e.g. the one at 3 sigma limits, but also 1 sigma limits) parameter values, that is not the max loglikelihood value. These values are in the *model.result file*, that is produced each time a fit is performed and is in the same folder of the fit. This ensures that the initial sampling of the new non-linear search starts by searching the region of non-linear parameter space that correspond to highest log likelihood solutions in the previous search.
- The sigma value is chosen between the value in the *config file* (that comes along with the PyAutoLens software), and in this case is the maximum of the following two:
 - The error on the parameter computed at 3σ from the previous search.
 - The value specified in the the **.json(autolens_workspace/config/priors/* .json* file under "width_modifier".

If on *type* there is written *absolute*, sigma assumes the value written in value, otherwise, if it is written *relative*, sigma assumes a value equal to the value on which the Gaussian is centered multiplied by the value written in value (i.e. sigma is a percentage of the average value).

The reason for this is due to the fact that lens modeling is prone to an effect called '*over-fitting*' where the errors on the lens model parameters is underestimated. This happens when fast non-linear search settings or simplified lens models are used.

Then, the sigma obtained from the previous step is compared with the width of the result at 3 sigma limits (*model.result file*, this file is produced each time a fit is performed and is in the same folder of the fit). The higher between the two is taken as the sigma.

- Eventually, there are the *lower and upper limits*. These are the values at which the gaussian's wings are truncated. They are written in the same **.json(autolens_workspace/config/priors/* .json* file under the keyword "gaussian_limits" or they can be set manually.

If the prior on one parameter is manually set to be gaussian and it is not manually changed in the code in a following fit, it will remain gaussian. This means that if nothing is specified, the software will set a gaussian prior according to the above procedure.

It also possible that in the file `*.json`(`autolens_workspace/config/priors/*.json`) there is written that on some specific parameter the prior that must be used is the *uniform prior*. In this case the previous fit best likelihood information is not passed to the next fit and the parameter is searched in the range established by the uniform prior. Though, if the software detects the over-fitting problem, it switches to the gaussian prior and builds the gaussian as described before. An example of this is reported in (6.1).

4.4 Adaptive grid

In the previous chapter, we used light profiles to model the light of a strong lens's source galaxy, where the light profile was an analytic description of how the luminosity varies as a function of radius. However, PyAutoLens allows to reconstruct the light distribution of the source even without assuming an analytical profile. In this section, we will briefly explain the procedures to be able to do what has just been said.

4.4.1 The Inversion procedure

First of all, it is necessary to explain how the fundamental Inversion procedure works.

The starting point is the inversion procedure is mappings between the source-pixels and the image. Suppose to some source-pixels, how are they mapped onto the image plane? Consider the example in figure (4.4)

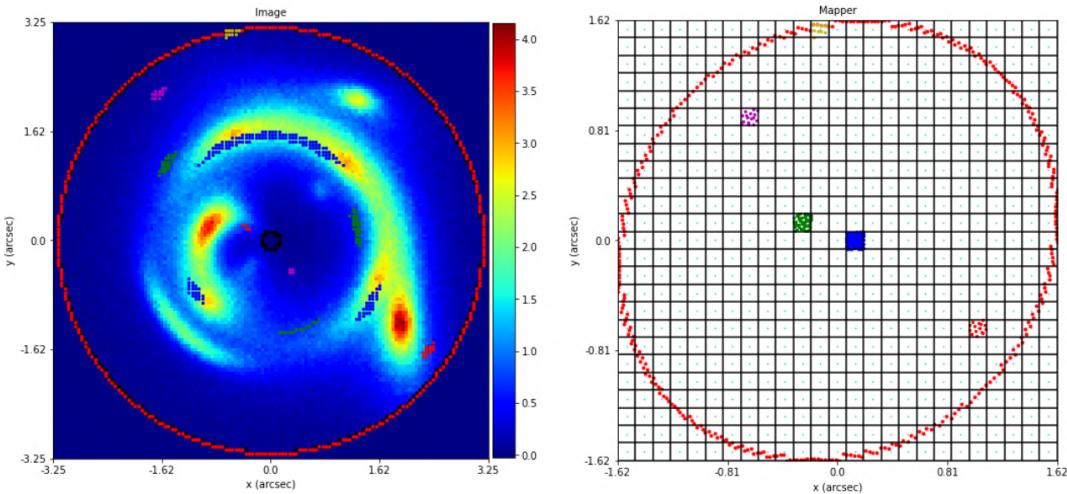


Figure 4.4: Mapping between the image and the source plane. The linear mapping process of some pixels from the source-plane (right) to the lens-plane (left). Image from [[PyAutoLens Tutorials 2022](#)]

These mappings are known before the inversion procedure, that is the one that reconstructs the source, begins. Therefore, the initial information that the software has available are:

- The mapping between each image-pixel and their corresponding source-pixels.
- The flux value in each image-pixel

The software uses this information it has available to perform the *linear inversion* process. In fact, with these two pieces of information the software can linearly solve for the set of source-pixel fluxes that best-fit (i.e. maximize the log-likelihood) the observed data.

The problem is solved in the same way as a set of linear equations is usually solved. This means that a matrix containing the information for the mapping between the image and source-plane is built, and then the multiple equations are solved to find the flux for each source-plane pixel. To have the computational details look into [Adaptive models log-likelihood function](#), from step 6 to 11 included. Obviously, during the fitting process, when it is said that the mapping is known a priori, it means that the software supposes a given mass model and associates a mapping with this model. It solves equations with the linear inversion process and compares the results with the data, i.e. verifies the log-likelihood. The linear part of the inversion is the mapping from the lens-plane to the source-plane and vice versa (figure (4.4)), however, to determine the lens parameters that allow the correct mapping we still have to use a non-linear search.

The software proceeds until a good model is found. When the threshold set in the algorithm, for example that of the equation (3.33) in chapter 3, is reached, the algorithm stops. An example of *linear inversion*, with the [pixelized source reconstruction](#) (i.e. without using parametric models), with an uniform grid is represented in figure (4.5).

Before proceeding, it is worth mentioning two other procedures performed by the software during the linear inversion process.

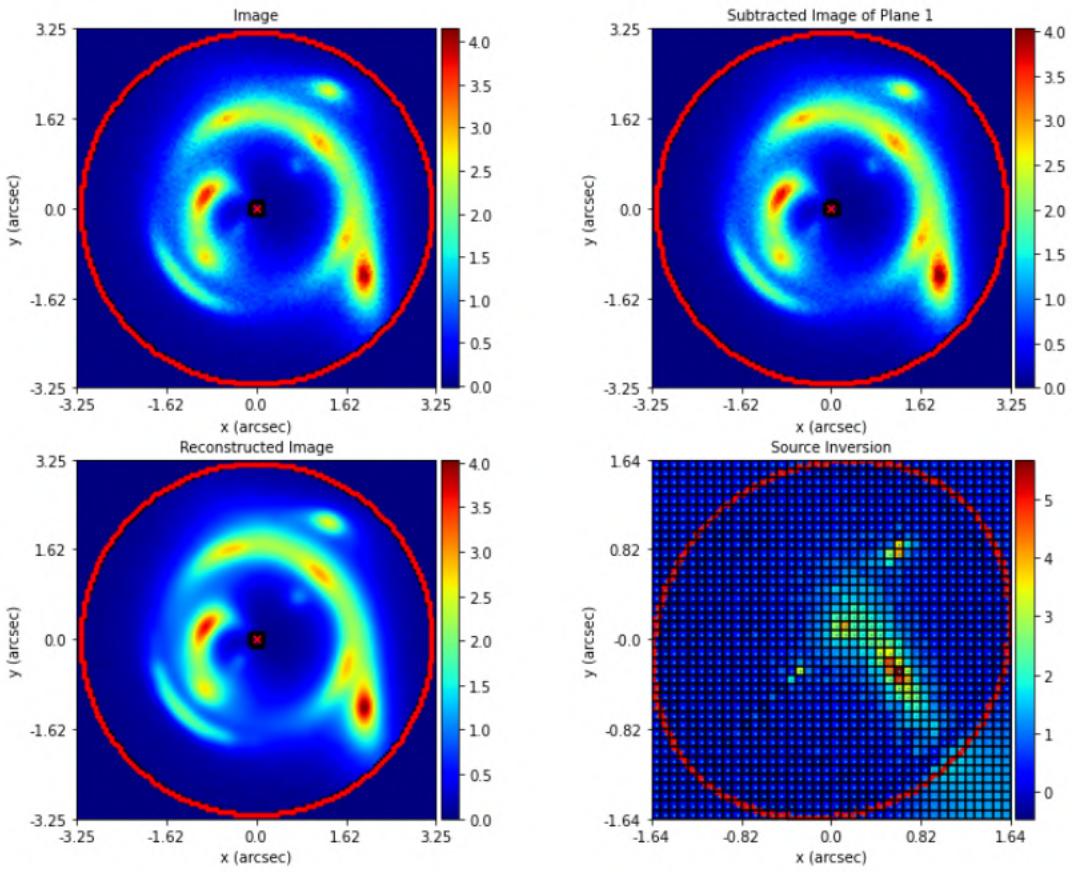


Figure 4.5: Mapping between the image and the source plane. Top-left: true data image. top-right: The lensed source image without the lens light (in this case because the data are simulated, lens light was not included in the simulation, thus we only see the source light). Bottom-left: reconstructed image of the source on the lens-plane with the best mass model. Bottom-right: reconstructed profile of the source on the source-plane thanks to the *inversion process*, the so-called [pixelized source reconstruction](#). Image from [\[PyAutolens Tutorials 2022\]](#)

- Even in this case, as in the approach that made use of parametric models (4.1), the solutions that the software produces, in order to be compared with the real data, are convolved with the PSF that is provided to the software.
- The solutions produced by the software undergo a *regularization process*. During the inversion process the software takes every pixel on the rectangular source-plane grid and computes the difference between the reconstructed flux values of every source pixel with its 4 neighboring pixels. If the difference in flux is large, then the solution is penalized, reducing its log-likelihood. In more mathematical terms, what happens is that it adds a 'penalty term' to the log-likelihood of any inversion process, that consists in the summed difference between the reconstructed fluxes of every source-pixel pair multiplied by the regularization coefficient

$$\text{regularization_term} = \sum_i^N (\text{flux}_i - \text{flux}_{i-1}) \cdot \text{reg_coeff} \quad (4.2)$$

where this subtraction must be taken for each pixel on the grid (in the summation it is taken into account that the pixels are repeated).

If the regularization coefficient were to be set to 0, it would mean completely neglecting regularization. This term of regularization must be added because otherwise it would meet the so-called 'over-fitting' of the image, reconstructing a source with a lot of noise and with strange features, often not physical. This

is due to the fact that the inversion goal consists in maximizing the log-likelihood, and in order to do this, it tries to fit everything as accurately as possible, adding noise and strange features. The solution maximizes log-likelihood, but it is not physical because extra elements have been added (over-fitting). Thus, decreasing the regularization coefficient increases the log-likelihood, because less smoothing allows the source reconstruction to fit the data better, contrary, the higher the regularization coefficient is, the lower the log-likelihood is.

Before proceeding, it is also interesting to understand the reason why we do not use the inversion from the first fit in the search-chaining procedure. In fact, in the version of the software used in the data analysis of this thesis, there is no possibility to jump directly to the inversion phase avoiding the use of parametric models. This is due to the fact that in this software's version it is very likely to incur the so-called *Demagnified solutions*.

This is also the reason why the software is provided with multiple positions. In fact, these require that a mass model traces the multiple image positions specified within a given threshold value (set manually) of one another in the source-plane. If this criteria is not met, a large penalty term is applied to the likelihood and this is massively reduces. This penalty is larger if the positions trace further from one another. This ensures the unphysical solutions that produce demagnified solutions have a much lower likelihood than the physical solutions we desire, because these multiple position are not traced near of one another.

4.4.2 The Adaptive Pixelization: the Delaunay magnification pixelization

The rectangular pixelization introduced in the previous section is not actually the best way to reconstruct the source with inversion process. In this section the *Delaunay Magnification Pixelization* will be introduced. This pixelization does not use a uniform grid of rectangular pixels, but instead uses an irregular grid of Delaunay pixels.

The reasons why the Delaunay adaptive grid is used, instead of the uniform rectangular one, are mainly two:

- Using different size triangular mesh grid elements (see figure (4.6)) and concentrating a greater number of them in the areas of greatest interest also allows to reduce the number of information element (here they're not properly pixels); in fact, reached a sufficient number of triangular mesh grid elements in the most significant regions on the source plane, to describe the areas that are of less interest because the source is not located there (i.e. they do not have sensitive information if not noise) a smaller number of larger size triangular mesh grid elements are sufficient. Thus, the computational time is reduced.
- Through the adaptive grid it is also possible to concentrate several triangular mesh grid elements where it is needed, where the most of informations come from, in the case of Delaunay adaptive pixelization around the areas of strong magnification (caustic) or where the source is located on the source-plane. This allows to increase the resolution where it is more important and give a better description of the source profile.

In figures (4.7) and (4.8) we can see the comparison between the regular uniform grid and the Delaunay adaptive grid.

Before proceeding, it is worth observing that even in the case the Delaunay adaptive grid is used, the regularization is performed. For a Delaunay grid, each source-pixel is compared with all other source-pixels with which it shares a direct vertex.

This means that different source-pixels may be regularized with different numbers of source-pixels, depending on how many neighbors are formed.

4.4.3 Adaptive features improvement: The Delaunay brigtness pixelization and the adaptive regularization

In the previous sections, both the regularization and Delaunay adaptive grid were introduced. However, they can both be further improved, in order to improve the result of the fit itself. There are different reasons for both. Let's start with the Delaunay adaptive pixelization. Concentrating a greater number of triangular mesh grid elements where the great abundance of information is located is very advantageous, both from a computational point of view, as previously mentioned, and from a resolute point of view. However, if the source were to be far from the regions of strong magnification, Delaunay adaptive pixelization would be ineffective, because, as previously mentioned, pixelization uses fewer and fewer pixels as we move away from these regions. This means low resolution of the source and therefore a worse and worse fit.

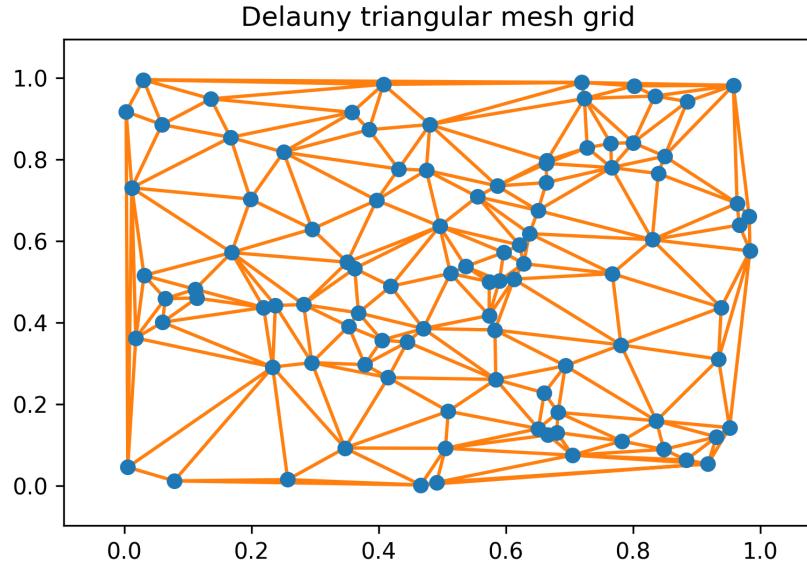


Figure 4.6: Delaunay triangular mesh grid.

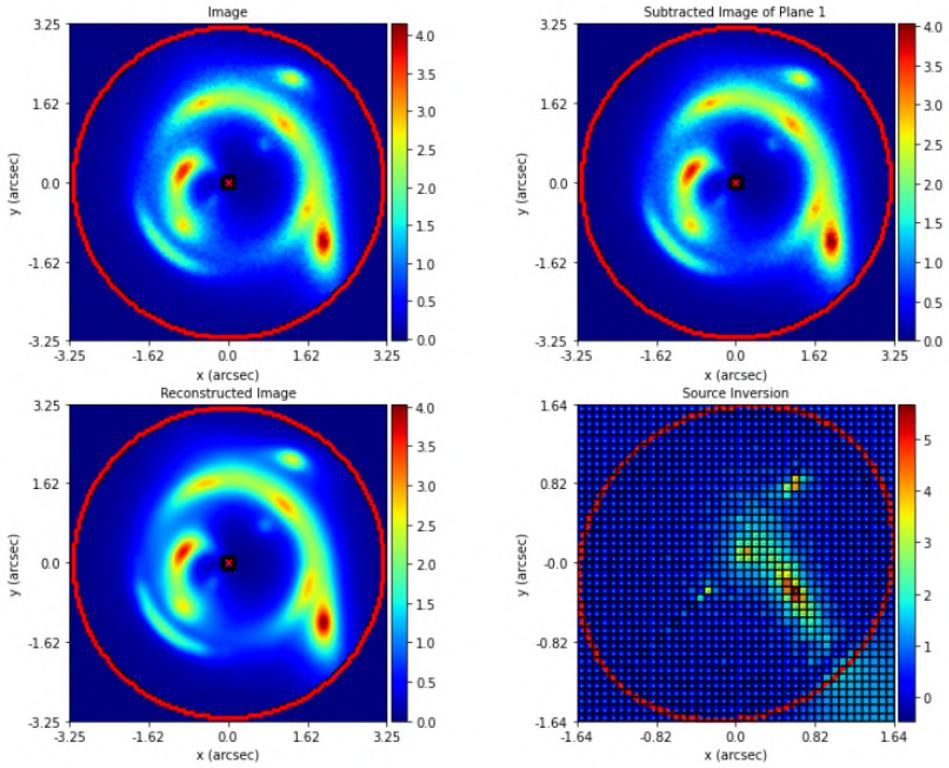


Figure 4.7: Inversion process using an uniform rectangular grid. In the bottom-right panel there is the reconstructed source using the rectangular grid. Image from [\[PyAutolens Tutorials 2022\]](#)

As far as regularization is concerned, which was introduced in the previous sections, it also presents a problem: the regularisation coefficient is constant. Regularize the source means adding up the difference in fluxes between all source-pixels multiplied by one single value, the regularization coefficient, to the log-likelihood. In this

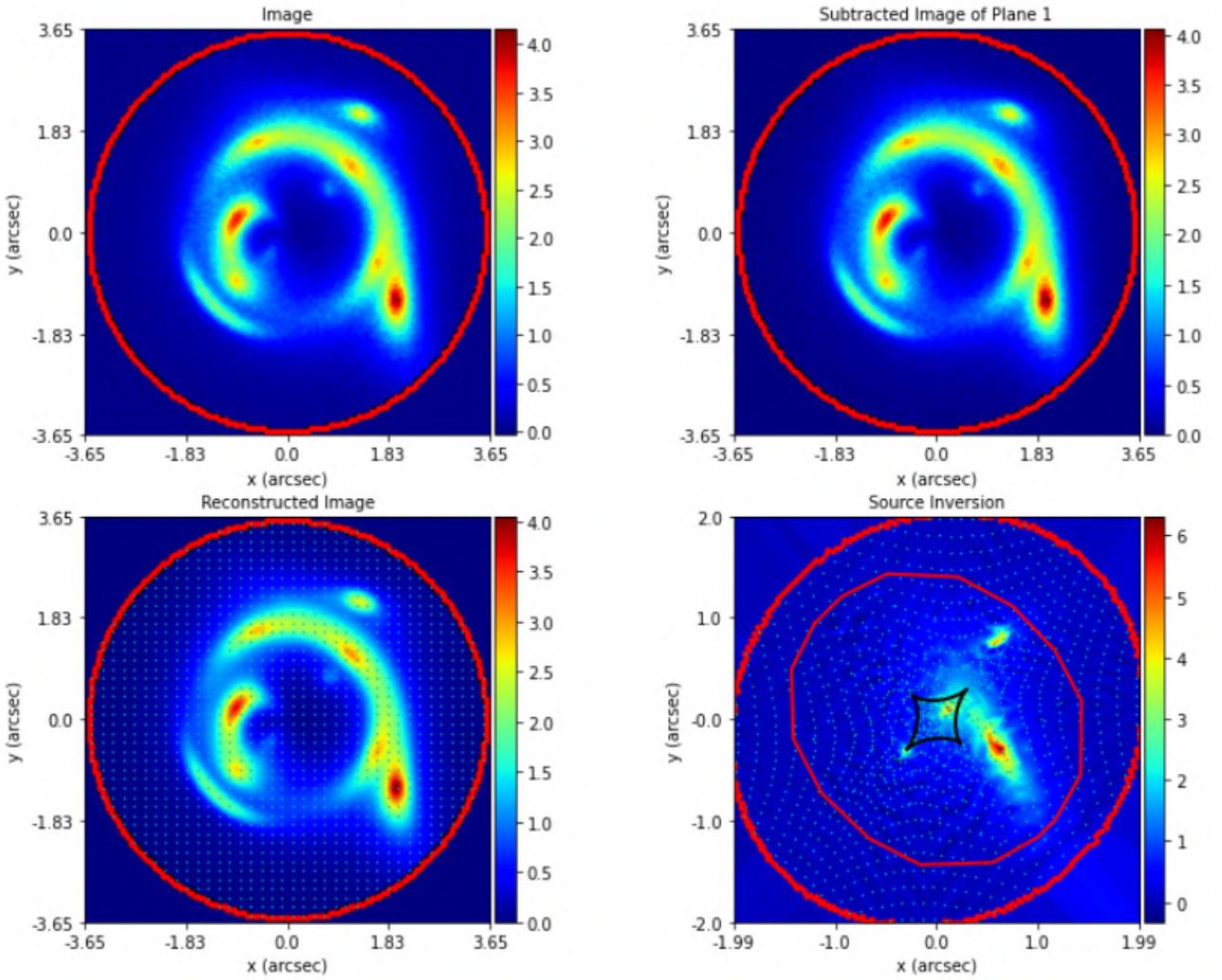


Figure 4.8: Inversion process using a Delaunay adaptive grid. In the bottom-right panel there is the reconstructed source using the Delaunay adaptive grid. The blue dots in the bottom-left panel represent the center of each pixel of the grid. The blue dots in the bottom-right panel represent the vertices of the Delaunay triangles in figure (4.6). As it is possible to see from these, in the right panel they are not uniformly distributed and moreover they are concentrated around the region where there is more information, in this case the caustics (black cusped lines). The triangular mesh grid elements around the caustics are much more than those located where the source is brighter. Image from [[PyAutolens Tutorials 2022](#)]

procedure every single source triangular mesh grid elements receives the same level of regularization, regardless of whether it is reconstructing the bright central regions of the source or its faint exterior regions. In fact, in general different regions of the source demand different levels of regularization. In the source's central regions the flux gradient is steep, as the surface brightness of the source increases naturally the closer we get to the central regions; the change in flux between two source pixels is much larger than in the exterior regions where the gradient is flatter (or there is no source flux at all).

Therefore, different regions cannot be weighted, or rather regularized, in the same way. So how can we improve both situations?

- Let's start again from the adaptive grid. Here above have been reported the problems of the magnification based Delaunay adaptive grid; in order to solve them, we need to adapt the pixelization to the source's morphology, such that source pixels congregates in the source's brightest regions, regardless of where the

source is located in the source-plane.

To do this, we build the so called 'hyper-images' of the lensed source galaxy. A hyper-image is a model image of the source computed using a previous lens model that has been fitted to the image (e.g. in the earlier searches of the chain by the use of the parametric models). This image tells us where in the source-plane our source is located, thus informing us of where we need to adapt our source pixelization.

- In order to solve the problem related to adaptive regularization, the above mentioned hyper-image is used again.

For every source-triangular mesh grid elements, we have a mapping between that pixel and a set of pixels in the hyper-image. Therefore, for every source-pixel, if we sum the values of all hyper-image pixels that map to it, it is possible to obtain an estimate of how much of the lensed source's signal it is expected to be reconstructed in that given pixel of the source-plane. The sum of all signals from each pixel that is mapped to a given pixel in the source-plane is called textit{pixel}-signal.

If a source-pixel has a higher pixel-signal, that will have to reconstruct more signal/flux, and therefore must be characterized by a lower regularization coefficient. Conversely, if the pixel-signal is close to zero, the source-pixel will reconstruct near-zero signal/flux and the regularization will smooth over these pixels by using a high regularization coefficient.

So, now that we have understood how to improve the adaptive capabilities of the software, how to find the parameters that characterize the adaptive grid and adaptive regularization? It is about making fits again. In fact, using hyper-images, non-linear searches are carried out, where in this case all the parameters concerning the lens, mass and light profiles, are fixed and the only parameters that are searched in the parameter space are those concerning the source, in other words the parameters of the adaptive grid and adaptive regularization. In the context of software adaptive capabilities, however, non-linear search in parameter space is no longer about log-likelihood of section (4.1) , but about a new modified log-likelihood.

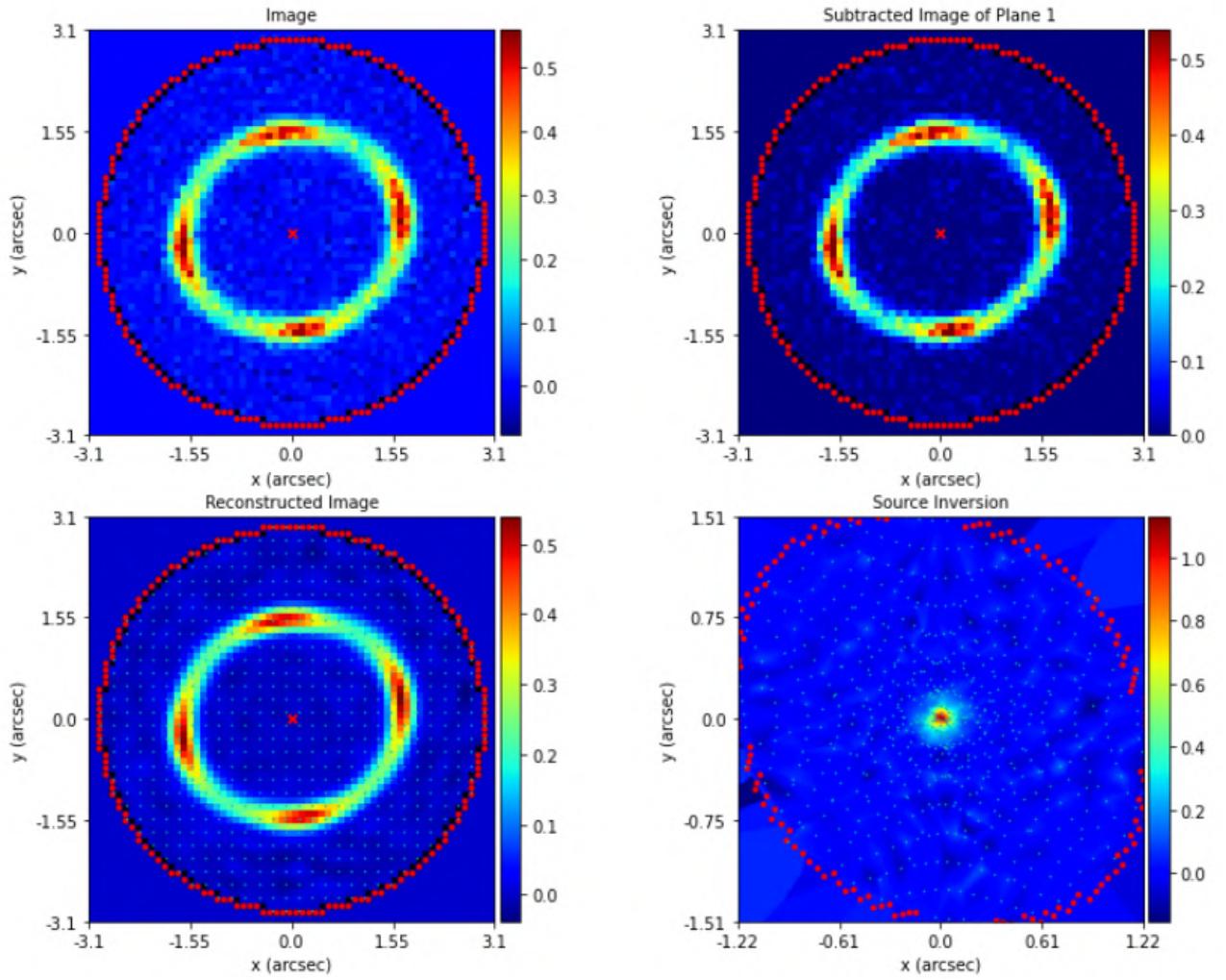


Figure 4.9: Inversion process using a Delaunay brightness adaptive grid. In the bottom-right panel there is the reconstructed source using the Delaunay brightness adaptive grid. The blue dots in the bottom-left panel represent the center of each pixel of the grid. The blue dots in the bottom-right panel represent the vertices of the Delaunay triangles in figure (4.6) As it is possible to see from these, in the right panel they are not uniformly distributed and moreover they are concentrated around the region where the source is located, they follow the source morphology. Image from [\[PyAutolens Tutorials 2022\]](#)

4.4.4 The log-likelihood for adaptive models

The log-likelihod for adaptive features (all subsection of section (4.4)) is no longer the simple one explained in section (4.1), but it is more complex. Below, the log-likelihood for adaptive models will be briefly introduced, without going into too much details about how the various terms are constructed, but only explaining what they are for. A more exhaustive and complete explanation of how the various terms are constructed can be found in [Adaptive models log-likelihood function](#) and in [J. W. Nightingale, Dye, and Massey 2018]. In particular, in the first link is described the log-likelihood for adaptive grid and constant regularization term and int the article (equation 10 and appendix B) is described the extension to adaptive regularization.

The log-likelihood function

$$-2\log\mathcal{L} = \chi^2 + s^T \mathbf{H} s + \ln[\det(F + H)] - \ln[\det(H)] + \sum_{j=1}^J \ln[2\pi(\sigma_j)^2] \quad (4.3)$$

The first term in the r.h.s is the χ^2 . The χ^2 is made up by different terms:

- The first is the *model data*. This time they're made up by the sum of the *reconstructed 2D image* and by the lens light, both convolved with the PSF. What is the reconstructed image? Once a model for the mass and for the source morphology (i.e. the parameters of the adaptive grid into the source-plane) is considered, through the ray-tracing the source light is mapped into the lens-plane, the mapped light is the reconstructed 2D image (figure (4.10)).

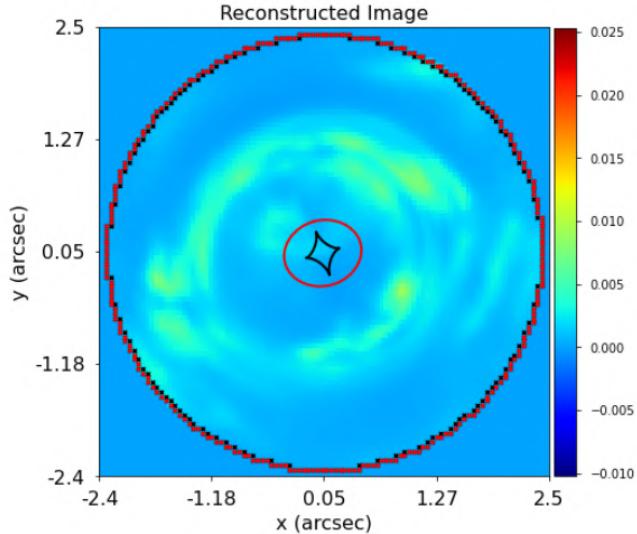


Figure 4.10: Reconstructed 2D image

- The second element is the *residual map*. This consists in the difference between the true data and the model data described just above.
- The third term is the *normalized residual map*, that consists in the above residual map divided by the *noise map*. (riquadro ... in figura).
- The chi-square map consists in the square of the normalized residual map, that represents what regions are fitted accurately and those that are not.
- Eventually, the χ^2 is the sum of each pixel-value of the chi-square map above.

The second term, $s^T \mathbf{H} s$, is the *regularization term*. This is the term which sums up the difference in flux of all reconstructed source pixels, and reduces the likelihood of solutions where there are large differences in flux (e.g. the source is less smooth and more likely to be overfitting noise). The *regularization coefficient* we are looking for is within the regularization matrix \mathbf{H} . In general \mathbf{H} is written to deal with the most general adaptive

regularization, and can be reduced to the special case of constant regularization. Here s is a matrix containing the reconstructed source-pixels' fluxes for each pixel that is used in the source-plane.

The third term is the *complexity term*. Now, before explaining it, it is necessary to point out the fact that if we decrease the regularization coefficient, this will always increase the log-likelihood, because less smoothing allows the source reconstruction to fit the data better (overfitting problem explained in section (4.4.1) under equation (4.2)). The higher the regularization coefficient, the lower the log-likelihood is. If, instead, we increase the regularization coefficient, this has two consequences:

- In contrast to what was said just above, it decreases chi-squared by fitting the data worse and so it produces a lower log-likelihood.
- Increases the *regularization term*, $s^T \mathbf{H} s$, by penalizing the differences between source pixel fluxes more, and this reduces the log-likelihood.

Finding the regularization coefficient relying only on the first two term would not give the correct result. Thus, adding the term $\ln[\det(F + H)] - \ln[\det(H)]$ solves this problem. Here F is the curvature matrix and its mathematical description is at the step 11 in [Adaptive models log-likelihood function](#). In essence, they measure the complexity of the source reconstruction and punish solutions when it is higher. In reality, as was previously stated, decreasing the regularization coefficient increases the complexity of source reconstruction since a source with less smoothing makes use of more flexibility to more closely match the data (over-fitting). Therefore, these two components work in opposition to the chi-squared and regularization terms to provide solutions that fit the data with a more rounded and straightforward source a higher log-likelihood (e.g. one with a higher regularization-coefficient).

The fifth term is the same as in equation (4.1).

Chapter 5

Data

In this chapter we will present the data that will be analyzed in the next chapter.

5.1 The cluster MACS J1206

MACSJ1206 is a cluster located at redshift $z=0.44$, whose data were obtained with the Cluster Lensing And Supernova-VLT (CLASH-VLT) survey. Its goal was to take a census of the cluster population, obtaining large numbers of multiple lensed image and identifying their redshift. In particular, the imaging data were obtained with the Hubble Space Telescope (HST) during the survey CLASH [Postman et al. 2012], while the spectroscopic data come from several Very Large Telescope (VLT) programs, where these aimed to add spectroscopic confirmations to the CLASH survey observations [Rosati et al. 2014].



Figure 5.1: MACSJ1206 RGB image.

5.1.1 The HST data

The imaging and photometric data analyzed in the next chapter were obtained with the Wide Field Channel (WFC) of the Advanced Camera for Surveys (ACS) of HST. The ACS camera works between 3700 Å and 11000 Å, and, in particular, the data analyzed were obtained using the broad near infrared I filter $f814W$, that works approximately between 7000 Å and 9500 Å (figure (5.1)).

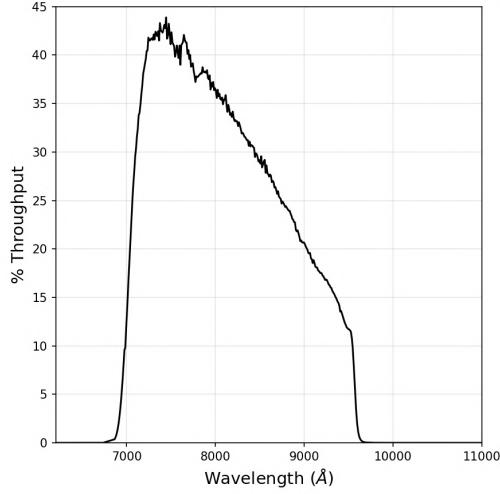


Figure 5.2: F814W filter transmission curve response at each different wavelength. Image from Space Telescope Science Institute (STScI).

5.1.2 The data

The data, both the scientific and weight data, were taken from [MACSJ1206_F814W_30mas](#). The units of the images are in [electrons/s] and can be converted in physical flux units using the [ACS Zeropoints](#).

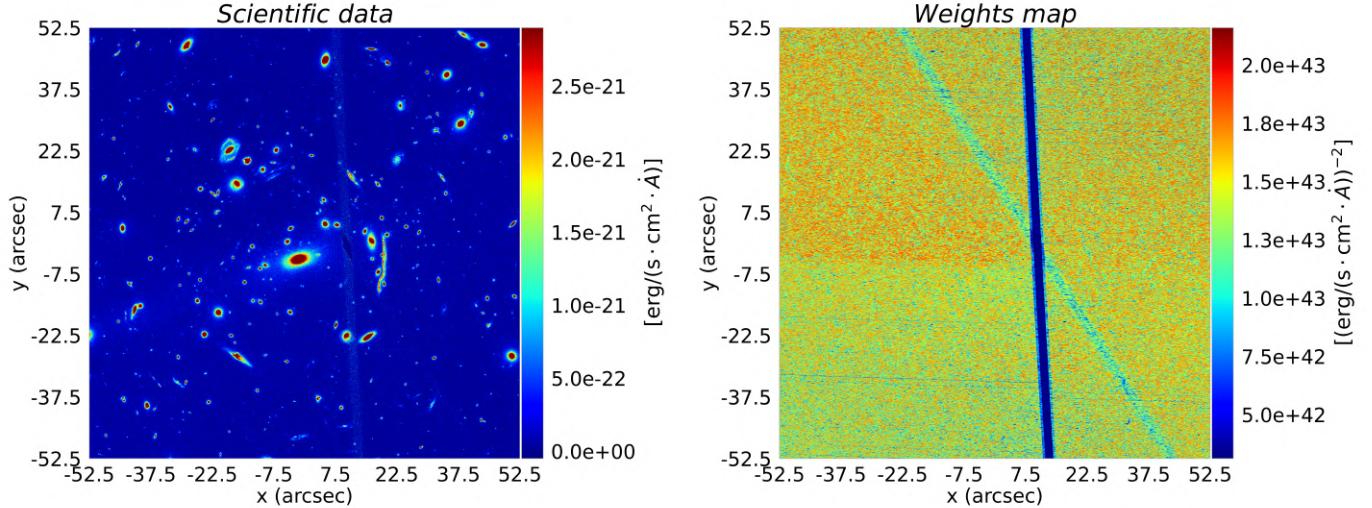


Figure 5.3: F814W band scientific image of the cluster (_drz.fits) and of the weights map (_wht.fits) in physical units

In addition, in order to analyze the data, the software requires not only the scientific image but also the noise map. To obtain this, it is necessary to combine the two previous images:

$$\text{noise map} = \sqrt{\frac{\text{(exposure time)}^2}{\text{weight map}} + \text{scientific image} \cdot \text{exposure time}} = \sqrt{\frac{1}{\text{weight map}} + \frac{\text{scientific image}}{\text{exposure time}}} \quad (5.1)$$

The weight map constitutes the inverse variance per pixel, thus $1/rms^2$, where the variance is given by the superposition of the sky background (modulated by the flat field), the readout noise and the dark current. The effects of Poisson shot noise due to signal from objects in the image are not included in this noise model. This explains the presence of the second term in equation (5.1), we need to add the Poisson noise. The multiplication of *exposure time* is needed in order to have the signal (electrons) and not the counts rate (electron/s) under the square root, which would lead to a noise greater of the signal. Then, it is necessary to normalize another time to the exposure time in order to have the same units of the image which is [electrons/s]. The result is in image (5.5).

Eventually, before proceeding, it is important to point out the pixel-scale of the data image, that can be read from the two keywords in the header files of the scientific data, one for each axis:

- x-axis pixel-scale: $CD1_1 = 8 \cdot 10^{-6} \text{ degree/pixel} = 0.03''/\text{pixel}$
- y-axis pixel-scale: $CD2_2 = 8 \cdot 10^{-6} \text{ degree/pixel} = 0.03''/\text{pixel}$

Thus, the same pixel-scale is assumed on each axis. The location of the lens galaxy is in image (5.4).

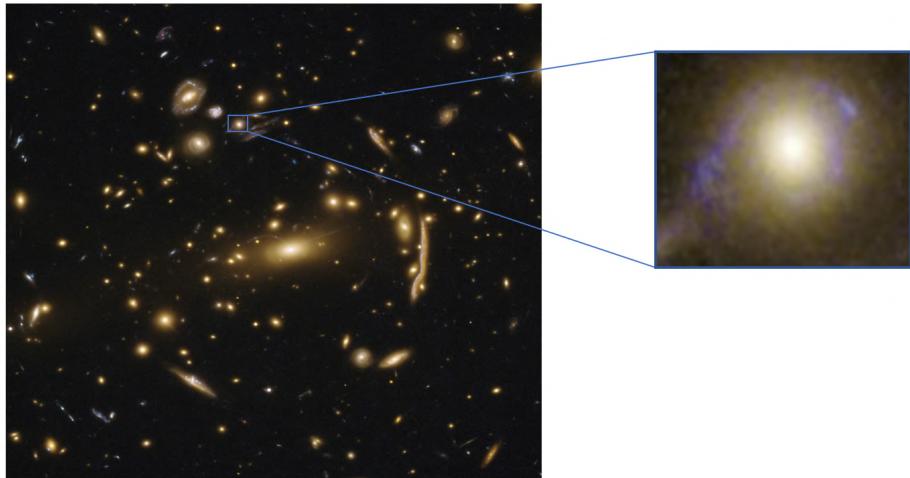


Figure 5.4: RGB image of the lens galaxy

5.1.3 The point spread function

As pointed out in the previous chapter, in order to compare our model with the real data, it is necessary to convolve model output with the PSF. To obtain the PSF the software *Tiny Tim* was used.

In order to generate the PSF the following parameters were setted:

- ACS camera -WFC
- Detector: WFC1
- f814w band
- Spectrum type: Blackbody, 6000 K
- PSF diameter: 3.550 arcsec
- Position where to generate the PSF : (2047, 1023), the exact center, because the PSF varies little inside the detector and for our purposes this variation could be considered negligible.

The result is figure (5.5).

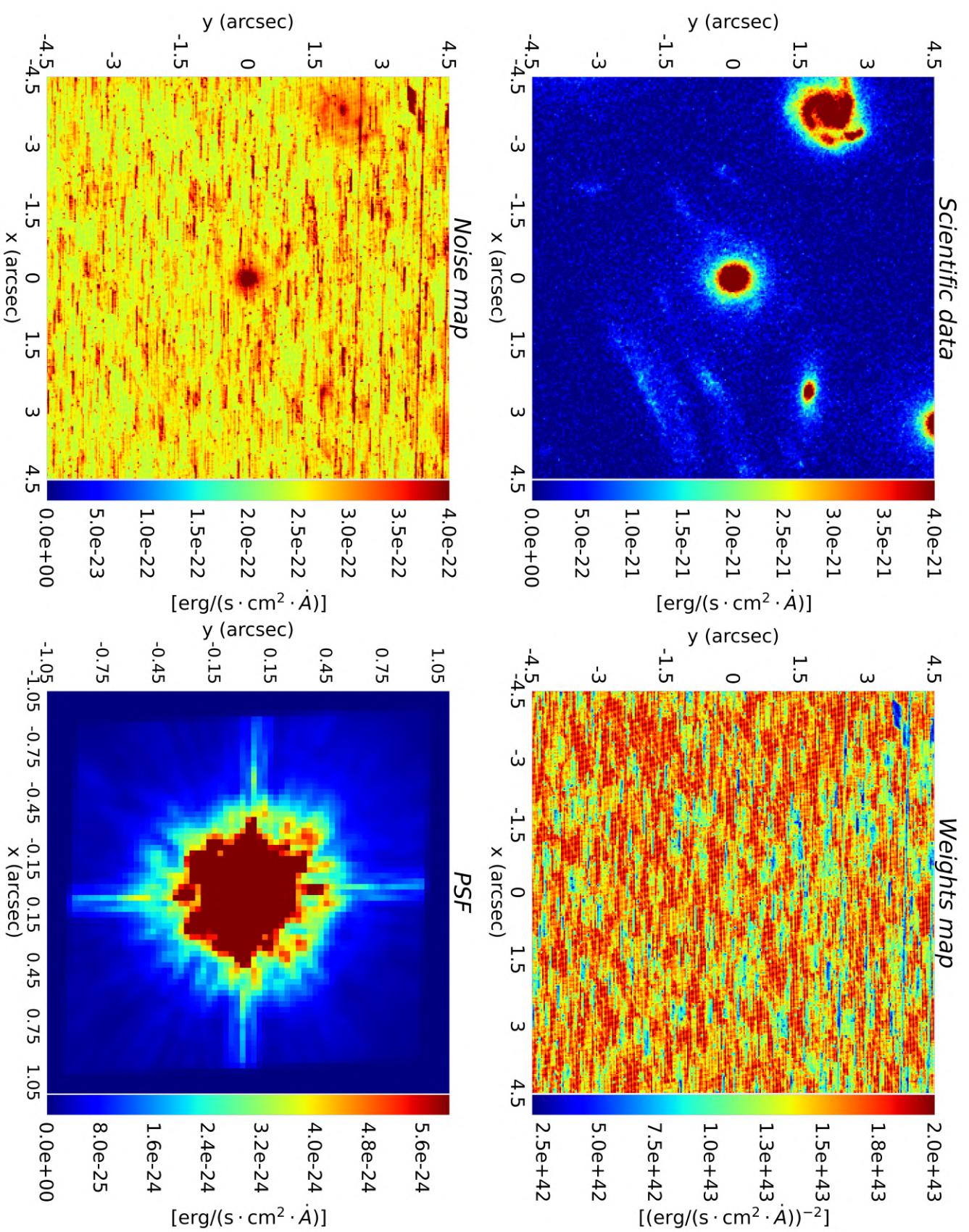


Figure 5.5: **TOP LEFT:** F814W band scientific image of the galaxy under analysis (the object in the center). **TOP RIGHT:** F814W weights map in physical units. **BOTTOM LEFT:** F814W band noise map in physical units. **BOTTOM RIGHT:** F814W band PSF profile in physical units.

5.1.4 The cluster deflection field

The galaxy under analysis is not isolated, it a cluster member galaxy. This means that we must deal with the deflection field generated by the cluster.

The software that is used for the analysis, PyAutolens, allows the possibility to insert into the model an external shear, like the case in equation (1.53). Though, the two components of this external shear are fitted together with the other system parameters (we could have parameters degeneration) and, moreover, it would try to fit the whole cluster deflection, even if at the first order, using just one single galaxy.

In order to overcome this issue we have chosen to use the previous results obtained in the article [Bergamini et al. 2019]. In this article they analyzed several clusters, including MACSJ1206, using the software *Lenstool*, that is a software that was developed to parametrically constrain, thus using parametric models, the total mass distribution of a galaxy cluster using the observed numerous multiple images positions all over the cluster. By doing so, in [Bergamini et al. 2019], using the information obtained in the Hubble filter F160W (in contrast to our analysis performed in the filter F814W), they were able to build a model using as much information as possible from each part of the cluster.

So, to insert the effects of the cluster in the model that must describe the galaxy under analysis, we started from their results. This has some advantages:

- The software is provided with a starting point to start searching in the parameter space and this allows us to remove some of the degeneration.
- The information provided to the software is derived from the knowledge of the entire cluster, eliminating the need to reconstruct the external deflection field using only the information provided by the galaxy in question.
- Furthermore, because PyAutolens allows us to set priors on the parameters, we can use the results of [Bergamini et al. 2019] model to do so, as described in section (4.3). Because of the prior, we can avoid having to fix the external deflection field around the galaxy under study using the Bergamini parameters, which may not be the best and assume a certain degree of uncertainty.

As previously stated, it was decided to include the external deflection field as External Shear in the model that describes the system. We will demonstrate below how the first-order approximation can reproduce the trend of the true deflection angle quite well.

First of all, the deflection angle generated by the model proposed in Bergamini et al. 2019 is illustrated in figure (5.6).

In order to reproduce the deflection field a first order approximation is used.

Initially, the components γ_1 and γ_2 are extracted, and they are illustrated in figure (5.7).

The two values of γ_1 and γ_2 are those at the exact center:

- $\gamma_1 = 0.0276$
- $\gamma_2 = 0.0878$

Now it is possible to calculate γ and ϕ , in order to have the constant components to write the external perturbation as written in section (1.1.3).

The value of gamma is obtained from the relation:

$$\gamma = \sqrt{\gamma_1^2 + \gamma_2^2} = 0.0906 \quad (5.2)$$

whereas the value of the angle ϕ (radians) is obtained from

$$\phi = 0.5 \cdot \text{arctg}\left(\frac{\gamma_2}{\gamma_1}\right) = 0.6015 \quad (5.3)$$

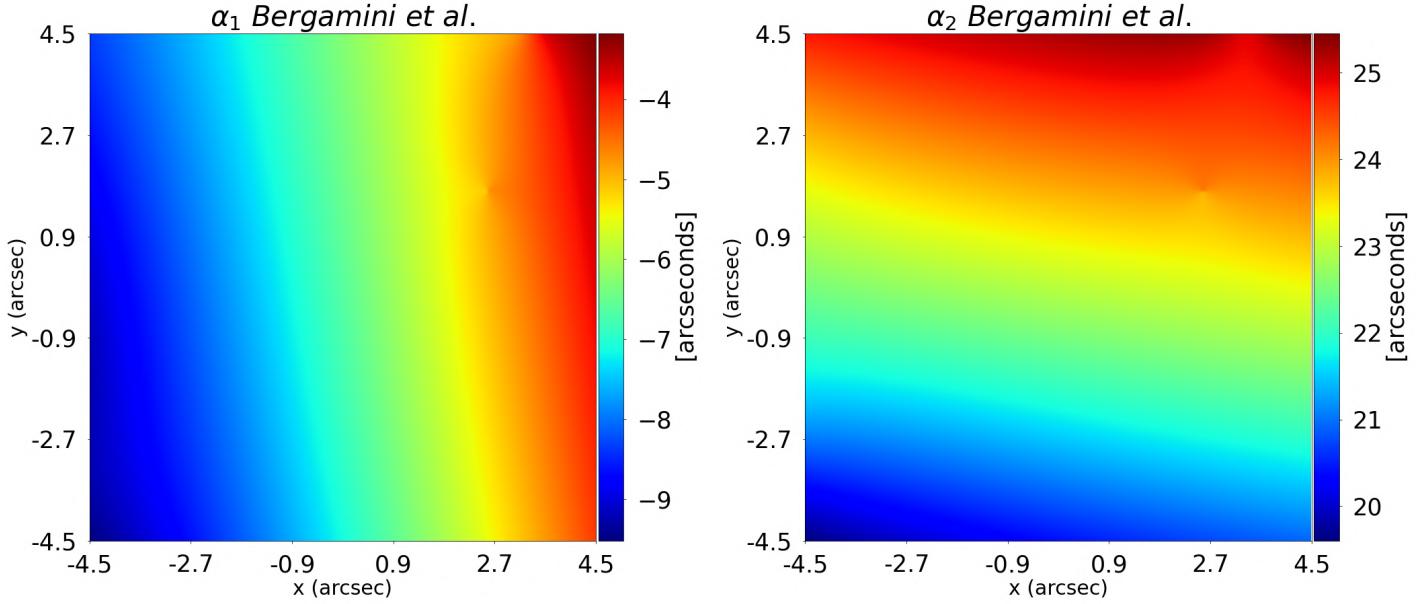


Figure 5.6: True cluster deflection in the region around the galaxy under analysis without the contribution of this. It is generated thanks to the model in [Bergamini et al. 2019].

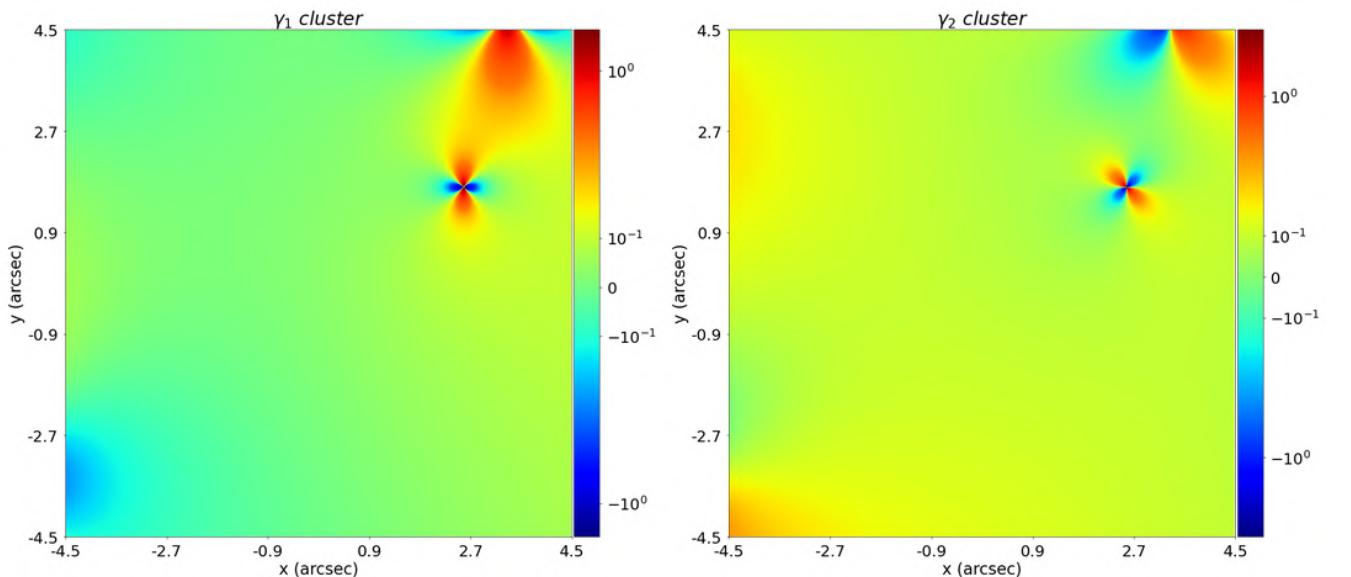
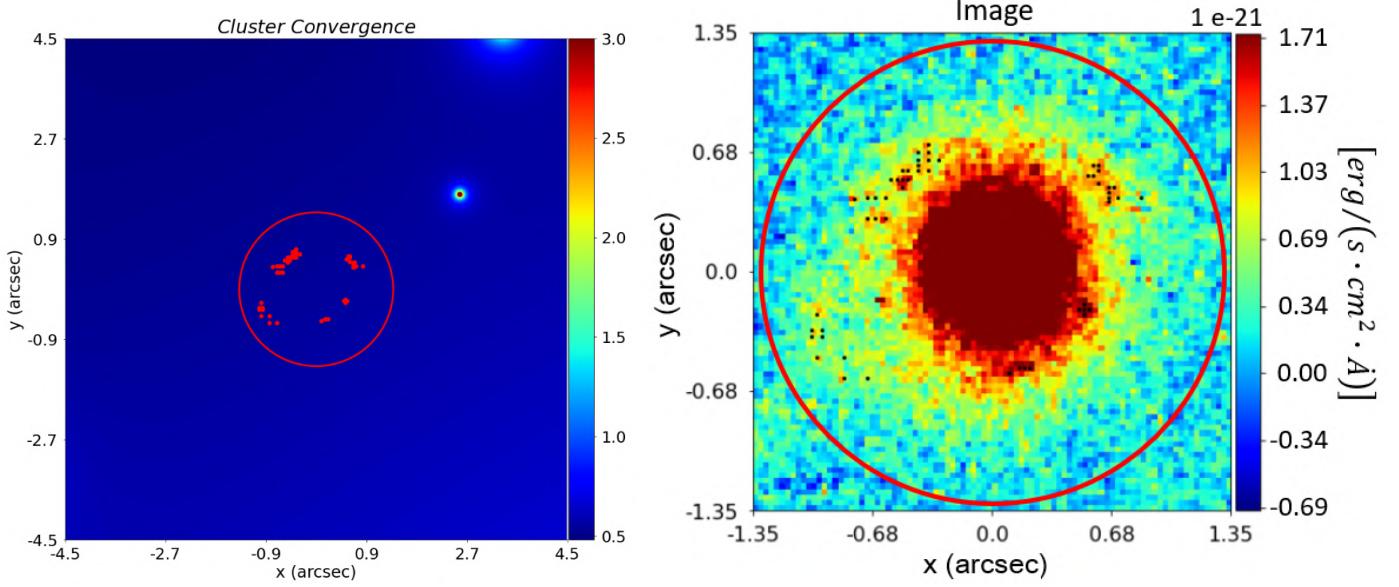


Figure 5.7: γ_1 and γ_2 components in the region around the galaxy under analysis without the contribution of this. It is generated thanks to the model in [Bergamini et al. 2019].

In conclusion, the convergence map of the cluster around the galaxy is presented in figure (5.8a, 5.8b), with a central value of $k_{ext} = 0.5446$.



(a) Convergence alongside the mask encompassing the lens (red circle) and multiple images locations (red dots).

(b) Light of the source alongside the mask encompassing the lens (red circle) and multiple images locations (black dots).

Figure 5.8: The figure displays the cluster convergence analyzed in the region surrounding the examined galaxy, excluding its contribution. The red circle and the red dots, represent respectively the mask used in the fitting procedure around the galaxy and the positions of multiple images. It is generated thanks to the model in [Bergamini et al. 2019].

As previously said, PyAutoLens allows the possibility to use external shear into the model. The model that PyAutoLens uses is the one in the equation (1.53), so no k_{ext} is considered.

Before proceeding, it is worth demonstrating that the first-order approximation can reproduce quite well the true deflection.

To do this we use equation (1.51), where $\alpha_{0,1}, \alpha_{0,2}$ are the central value of the deflection field model proposed in Bergamini et al. 2019 and illustrated in figure (5.6), and $\theta_{0,1} = \theta_{0,2} = 0$ are the grid center coordinates (the center of the grid is located in the exact center of the array).

The first term is the one in equation (1.51), so, giving the values of γ and ϕ just calculated, we have (5.9).

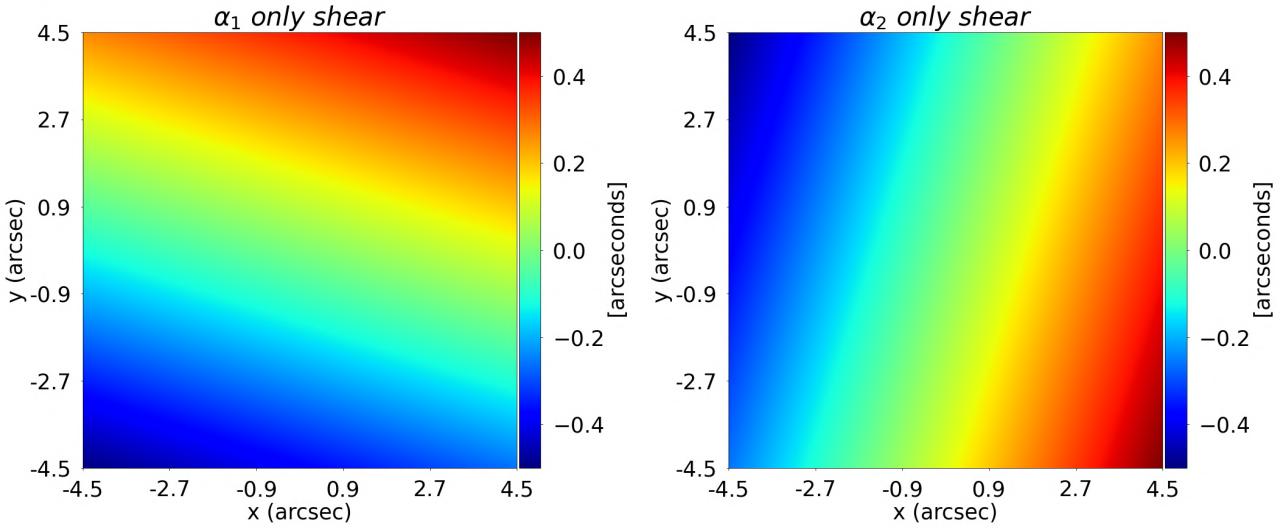


Figure 5.9: External shear deflection angle in the first order approximation

If we now add the convergence term $k_{ext}\vec{\theta}$ we obtain (5.10).

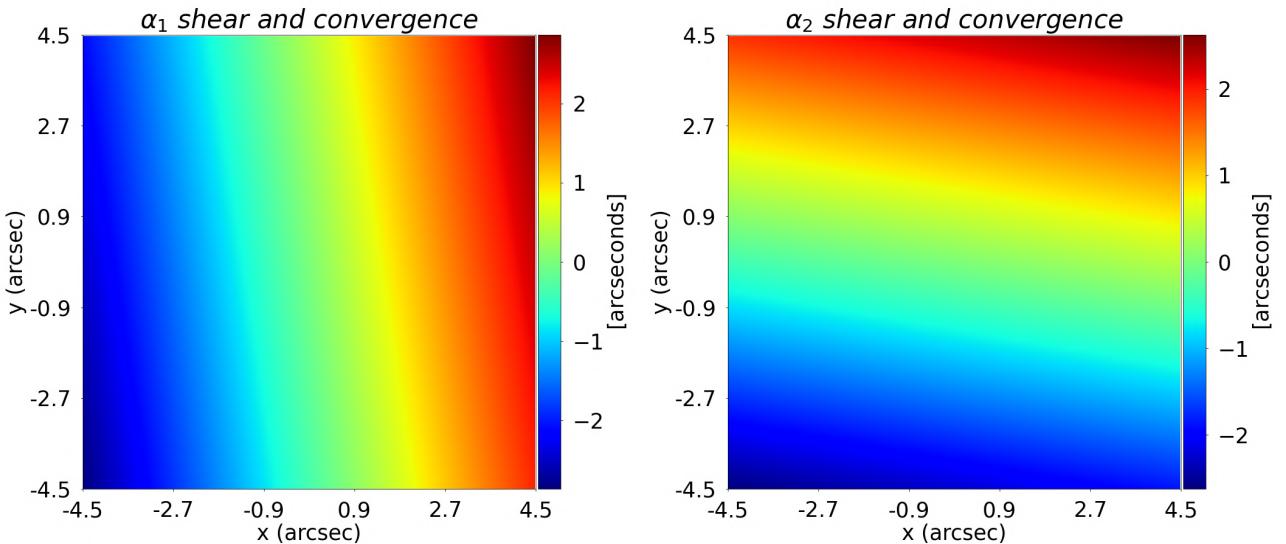


Figure 5.10: External shear + convergence term deflection angle in the first order approximation

Eventually, adding the constant central terms $\alpha_{0,1}$ and $\alpha_{0,2}$, we obtain (5.11).

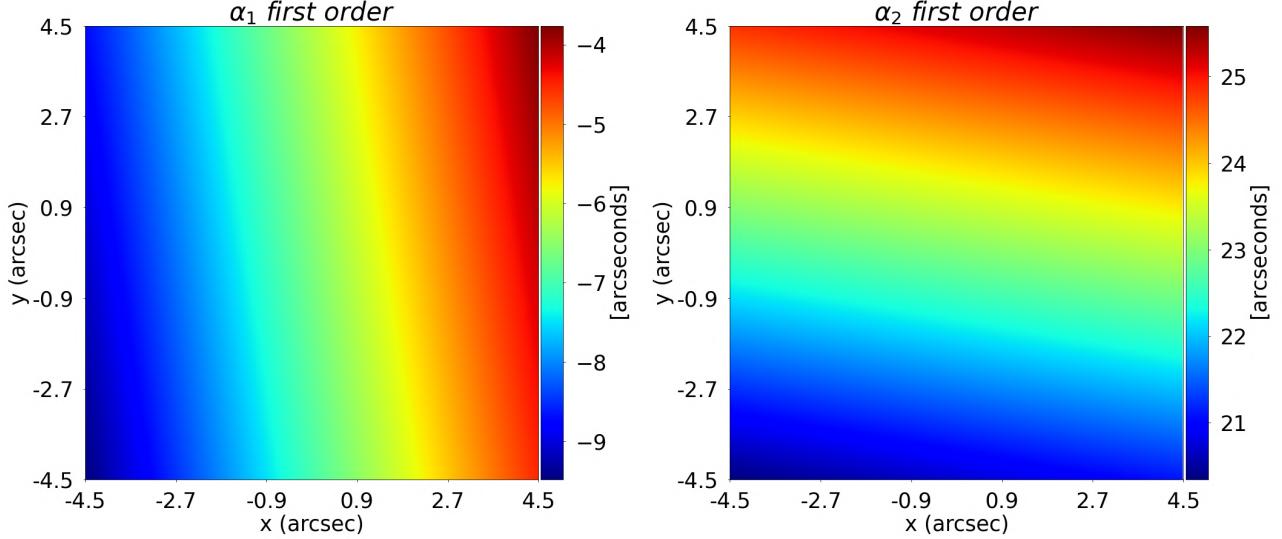


Figure 5.11: External deflection generated using equation (1.51).

As it is possible to see the gradient structure is respected as well as the magnitude. Furthermore, in the region exactly surrounding the lens galaxy under examination, i.e. the box that goes from $-1.35''$ to $1.35''$, the first order deflection reproduces very accurately the true deflection field. In fact if we take the difference between the first order approximation and the real deflection we have (5.12).

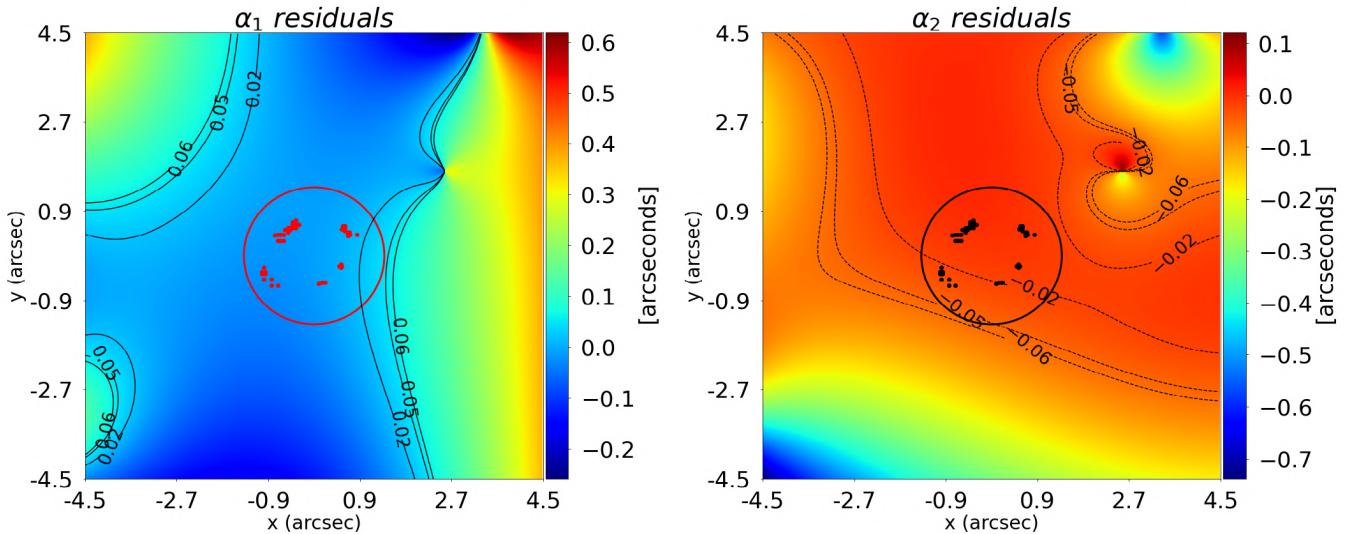


Figure 5.12: Difference between the true deflection and the first order approximation. Left figure contours: the innermost is at 0.02, in the middle 0.05 and the outermost is at 0.06. Right image contours: the innermost is at -0.02, in the middle -0.05 and the outermost is at -0.06. In black the mask encompassing the lens (black circle) and multiple images locations (black dots).

Chapter 6

Data analysis: the Analysis Pipeline

In the upcoming subsections, a systematic and structured series of data processing steps are applied to the dataset introduced earlier, in order to extract meaningful insights and explore in depth the physics of the system. This approach can be best described as a workflow composed of interconnected data processing modules or stages, where the output of one stage seamlessly becomes the input for the next, enabling a smooth and efficient advancement in the data analysis procedure. In particular, the procedure involves eight distinct fitting processes, each aimed at enhancing the system's description with every subsequent fit and the search-chaining procedure described in section (4.3).

The system being examined comprises the central galaxy at the image's center and the multiple images surrounding it (6.1).

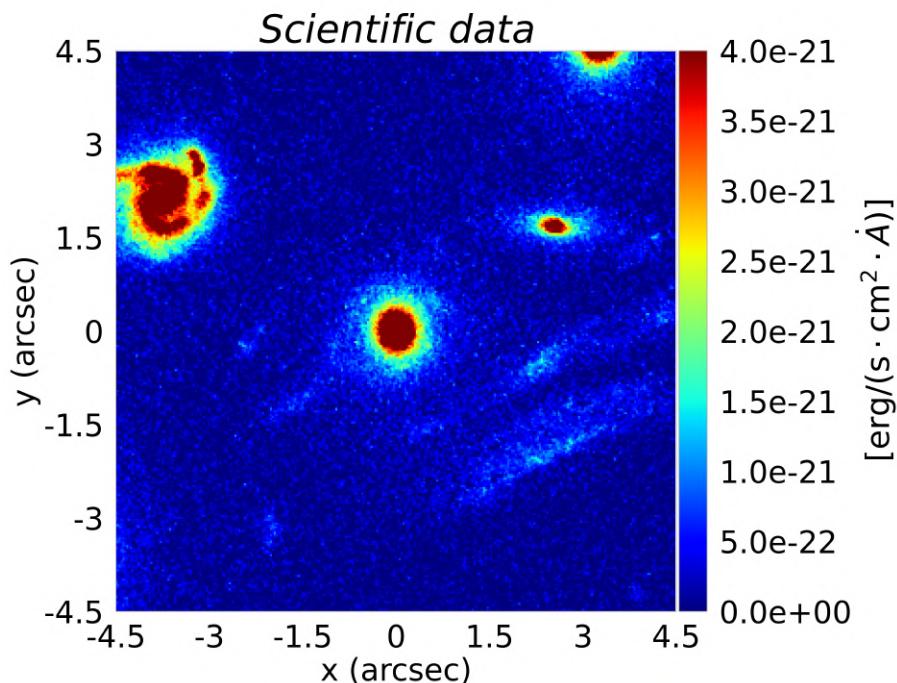


Figure 6.1: Image of the lens galaxy (very center) and multiple image around it.

Before moving forward, it is essential to outline some precautions that will be adopted throughout the analysis procedure:

1. The mass model being considered consists of the elliptical power-law up to the seventh fit. Fit number 7, where the most suitable source model is adopted, also integrates the elliptical broken power-law, allowing to test the presence of a changing point in the trend of the mass profile.

2. Within the PyAutoLens framework, the equation (1.64) is effectively solved, requiring a rescaling operation for the mass, deflection angles, and source positions by means of a multiplication factor of $(1 - k_{ext})$. However, this last procedure will only be done after the last fit and for the quantities physically relevant to the description of the lens.
3. The magnitude of shear, along with its corresponding orientation angle and associated errors, will be provided exclusively within the final fit results. In all preceding fits, solely the elliptical components and their respective uncertainties will be provided.
4. The masses of the lens, both the one calculated by the software and the one scaled by the factor $(1 - k_{ext})$, will be provided. Furthermore, only in the last fit, the Einstein radius, properly adjusted for the correct factor, will be presented.
5. Each of the forthcoming images, pertaining to the fitting process and the analyzed data, is subject to dimensionless rescaling by a factor $6.87 \cdot 10^{-20}$ [erg/s · cm⁻²·Å], obtained during the *ACS Zeropoint procedure*, that allows the transformation from the instrumental units to the physical ones. This rescaling serves the dual purpose of expediting computational procedures and maintaining the result's consistency, as both the scientific image depicting the system under examination and its corresponding noise image undergo the identical scaling transformation, ensuring uniformity in units.
6. In the context of successive fits, the term *Einstein radius* refers to the scale-radius denoted as b within the elliptical power-law model. The actual Einstein radius, which is derived from the value of b , will be presented in the concluding section, specifically within the discussion of the final fit's results (referenced as (7)).
7. Measurements along the planes of the lens and source will be expressed in arcseconds. To convert these measurements into physical units, one must simply employ the source's and lens' redshifts, provided within the analysis, and select the appropriate cosmological framework to compute the respective distances.
8. The adopted cosmology is Planck 2015 cosmology ([Planck Collaboration et al. 2016](#)): $H_0 = 67.7$, $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$.

6.1 1 Fit: Lens light fitting

The first fitting procedure consists in analyzing only the surface brightness profile of the lens galaxy. The objective of this procedure is to meticulously reconstruct the surface brightness profile of the lens with the utmost precision to facilitate the software in recognizing the multiple images with a high degree of accuracy. To describe the brightness profile of the source we use the Sersic profile in equation (2.60).

In order to reconstruct this profile, it is necessary to shrink the fitting mask so that it encloses only the light of the lens galaxy and excludes the outermost part, where the multiple images of the source galaxy are located. The following image (6.2a, 6.2b) shows the lens galaxy and highlight the location of multiple images (black dots) of the source galaxy. The black dots do not represent the multi-image instances themselves; rather, as it will be explained in the following section, they serve as a tool employed by the software to enhance the reconstruction of the source as it moves through the parameter space.

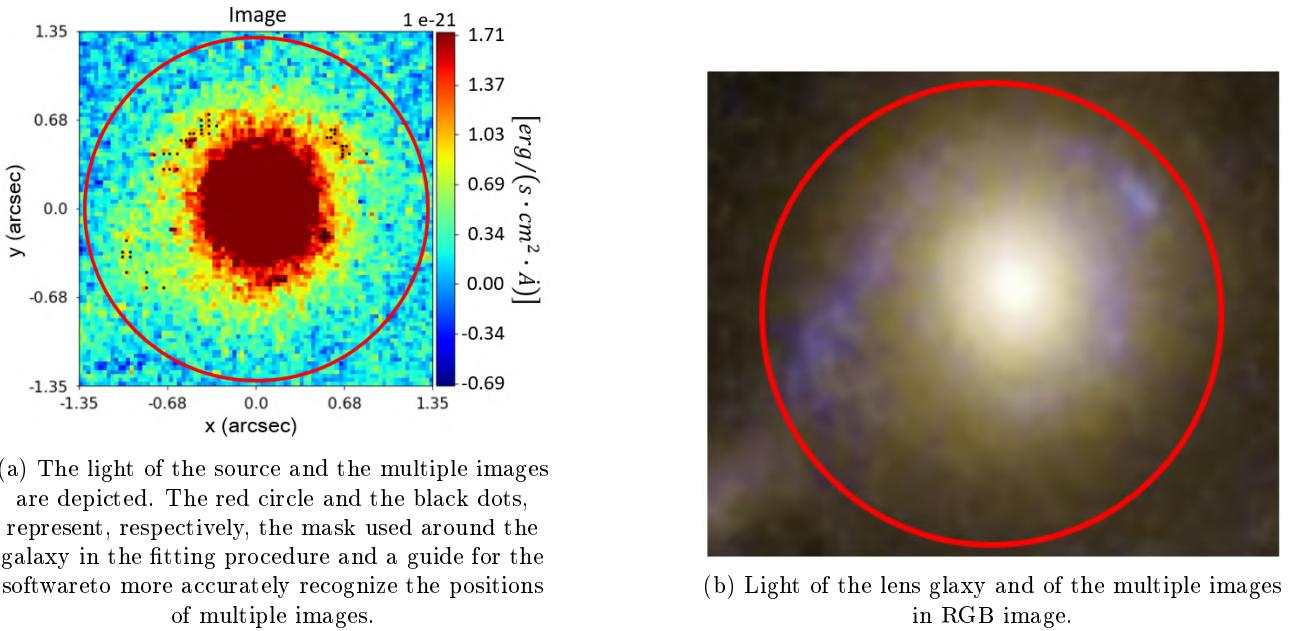


Figure 6.2: The figure depicts the light of the lens galaxy and the one of the multiple images.

Thus, to fit the lens light we must shrink the mask around the galaxy. We choose a mask radius of 0.6 arcseconds, figure (6.3).

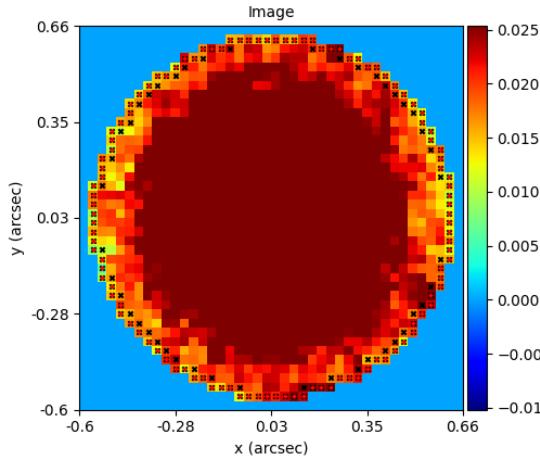


Figure 6.3: The mask encircles only the lens light.

The fitting procedure, which uses 7 parameters, gives the following maximum log-likelihood model parameters (6.1). Visual results are in figures (6.4,6.5, 6.6).

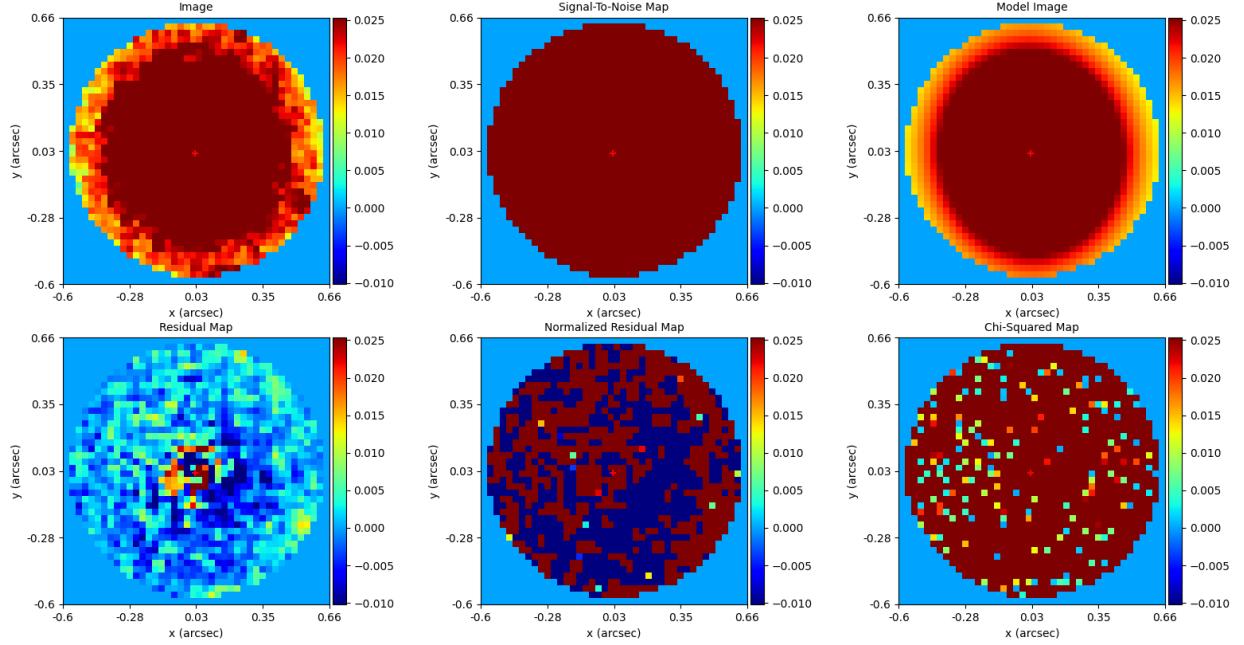


Figure 6.4: The outcome derived from the initial fitting procedure.

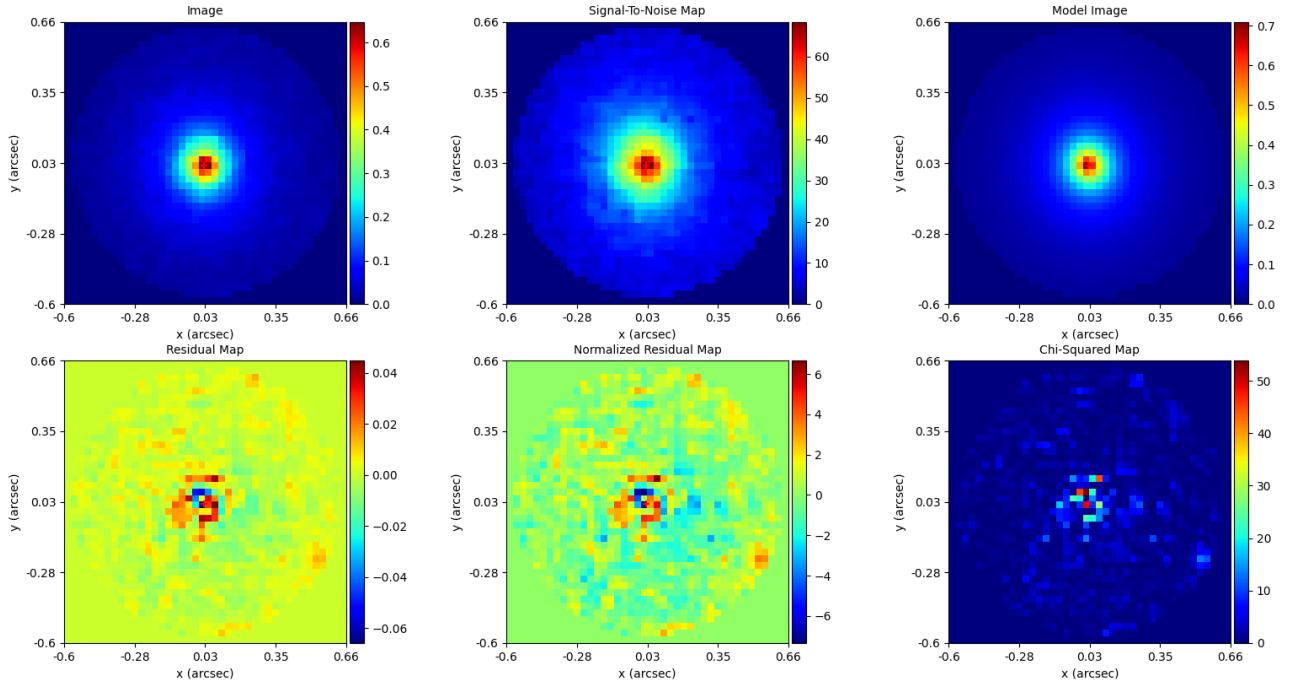


Figure 6.5: The outcome of the initial fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

The fact that in the center the elliptical profile of Sérsic is not able to so accurately fit the surface brightness is not a big problem, as the multiple images are located in the outer regions where the fit is good. In fact, this

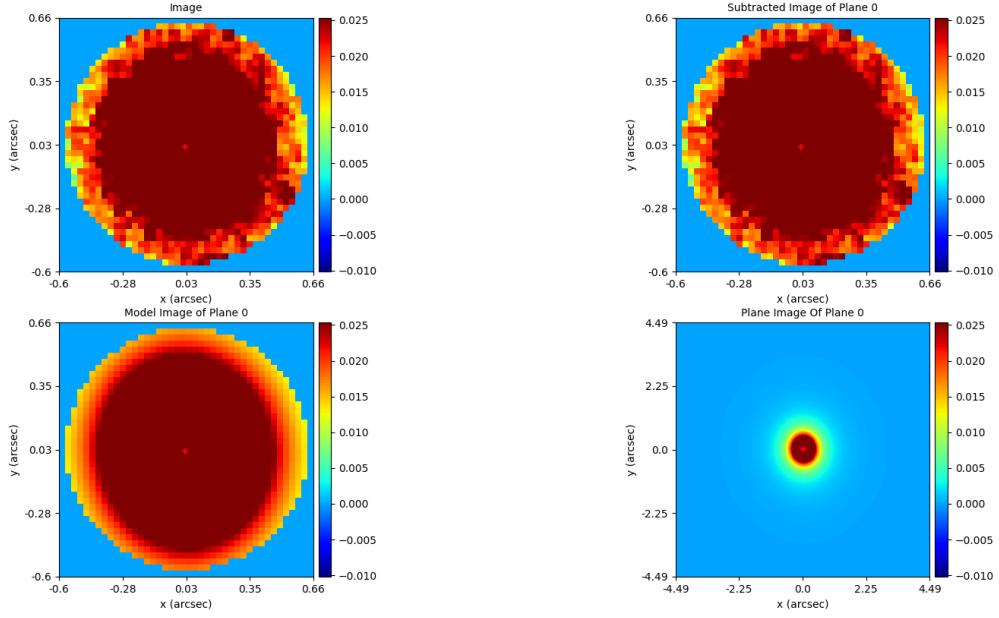


Figure 6.6: Outcome of the initial fitting procedure. The subtracted image of plane 0 is the image from which what has been recognized as the light of the source has been taken away. In this case, in the image there is only the light of the lens, so real image and subtracted image coincide.

fit aims to facilitate the identification of multiple images by the software, identifying which component of light belongs to the lens galaxy and which to the multiple images, in the position where the latter are located. For this purpose the result of the fit are suitable.

The maximum log-likelihood model parameters are:

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens bulge centre y-component ["]	2.09×10^{-2}	1.02×10^{-3}	9.88×10^{-4}
galaxies lens bulge centre x-component ["]	2.30×10^{-2}	8.30×10^{-4}	6.93×10^{-4}
galaxies lens bulge elliptical comps 0 [/]	-7.88×10^{-3}	6.52×10^{-3}	6.30×10^{-3}
galaxies lens bulge elliptical comps 1 [/]	-6.12×10^{-2}	6.18×10^{-3}	7.57×10^{-3}
galaxies lens bulge intensity [/]	3.56×10^{-2}	2.88×10^{-3}	2.86×10^{-3}
galaxies lens bulge effective radius ["]	3.76×10^{-1}	1.66×10^{-2}	1.83×10^{-2}
galaxies lens bulge sersic index [/]	2.80	9.26×10^{-2}	9.99×10^{-2}

Table 6.1: Outcomes of the first optimization process for the parameters.

The intensity, is adimensional because of the previous adimensional rescaling.
The elliptical components are a combination of the axis-ratio and of the orientation angle of the profile:

$$fac = \frac{1-q}{1+q}$$

$$elliptical\ component\ zero = \epsilon_0 = fac \cdot \sin(2\phi) \quad (6.1)$$

$$elliptical\ component\ one = \epsilon_1 = fac \cdot \cos(2\phi)$$

and thus, the axis-ratio and the angle are given by

$$\begin{aligned}\phi &= 0.5 \cdot \arctan\left(\frac{\epsilon_0}{\epsilon_1}\right) \\ fac &= \frac{\epsilon_0}{\sin(2\phi)} \\ q &= \frac{1 - fac}{1 + fac}\end{aligned}\tag{6.2}$$

Although, to provide a comprehensive overview of the brightness profile, the parameters of greater significance are reported in physical units in the table (6.2) and the results of the fit is in figure (6.7).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
Orientation angle from positive x-axis [°]	93.66	2.87	2.31
Axis ratio [/]	0.88	0.06	0.2
Effective Intensity [$\text{erg}/(\text{s} \cdot \text{cm}^{-2} \cdot \text{\AA})$]	2.62×10^{-21}	2.10×10^{-22}	2.80×10^{-22}
Effective radius [kpc]	2.20	0.09	0.10

Table 6.2: Physical parameters of greater significance

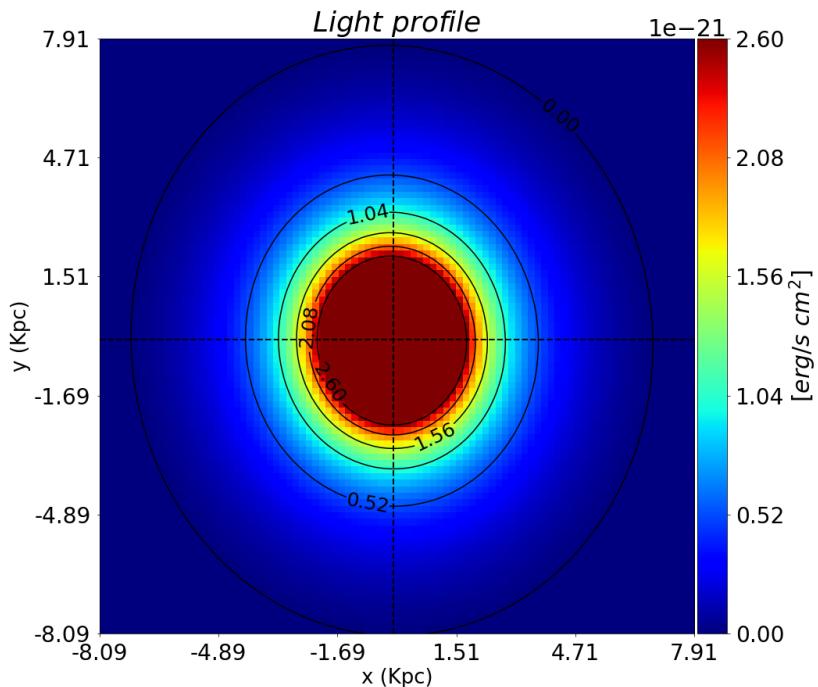


Figure 6.7: The result of the initial fit displays the lens light in physical units, with black ellipses representing the isophotes.

6.2 2 Fit: Fitting the system for the first time

The second fit is the one in which for the first time the source lens system is reconstructed, therefore the first where the characteristics of the lens and the source are reconstructed. The second fit involves the use of parametric models to describe both the lens and the source. In particular, the lens is described using elliptical power-law (2.13), while the source is described using an elliptical Sérsic profile (2.60). All parameters describing the source have been left free. Instead, with regard to the lens plane, assumptions have been made and prior have been set:

- The center of the mass distribution of the lens has been set equal to the center of the light distribution obtained in the first fit:

$$\begin{aligned} x_{\text{center}}^{\text{lens}} &= x_{\text{center}}^{\text{lens bulge}} \\ y_{\text{center}}^{\text{lens}} &= y_{\text{center}}^{\text{lens bulge}} \end{aligned} \quad (6.3)$$

- A gaussian prior was imposed on the elliptical components of the lens galaxy, using the results concerning the elliptical components of fit 1:

```

1 af.GaussianPrior(mean=max_log_likelihood_value, # max likelihood parameters from previous
# search
2 sigma=sigma, # sigma of the gaussian
3 lower_limit=lower_limit, # lower truncation limit for the gaussian prior
4 upper_limit=upper_limit # upper truncation limit for the gaussian prior
5 )
```

Listing 6.1: Gaussian prior applied to parameter space by PyAutolens software

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{lens mass}} [/]$	max loglikelihood value $\epsilon_0^{\text{lens bulge}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$
$\epsilon_1^{\text{lens mass}} [/]$	max loglikelihood value $\epsilon_1^{\text{lens bulge}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$

The idea is that the axis-ratio and the orientation of the mass have the possibility to vary, but not too much with respect to the distribution of light.

- The prior on the other lens' parameters is described in section (4.3).
- Eventually, the last prior was imposed on the external shear, used to describe the external deflection of the cluster. In PyAutolens one is allowed to pass just the external shear magnitude and orientation angle (γ and ϕ). PyAutolens does not work with the equations (1.53) and (1.62), because it is not possible to give the external convergence k_{ext} and the first constant term of the expansion $\vec{\alpha}_0$. In order to have the same effect in PyAutolens, without considering the term $\vec{\alpha}_0$, that can be added thanks to the superposition principle in a second moment, it is necessary to scale the shear component. PyAutolens solves the equation of the form (1.53). In order to take into account the effects of $k_{external}$ also in Autolens, as is done in Lenstool, it is necessary to rescale equation (1.62) for the term $(1 - k_{external})$ and obtain equation (1.64), which is of the same form used by PyAutolens. Obviously, the solutions found by PyAutolens for $\vec{\beta}$ and $\vec{\alpha}$ are the primed one in this way and therefore it will be necessary to rescale the quantities obtained for the factor $(1 - k_{external})$.

In order to deal with the uncertainty concerning the [Bergamini et al. 2019] model, a gaussian prior is imposed on the external shear. The external shear is defined by a magnitude γ and an orientation angle ϕ , PyAutolens works with the elliptical components even in this case, except that the *fac* term is now substituted by the magnitude γ :

$$\begin{aligned} \text{elliptical component zero shear} &= \epsilon_0^{\text{shear}} = \gamma \cdot \sin(2\phi) \\ \text{elliptical component one shear} &= \epsilon_1^{\text{shear}} = \gamma \cdot \cos(2\phi) \end{aligned} \quad (6.4)$$

and thus, the magnitude and the angle are given by

$$\begin{aligned}\phi &= 0.5 \cdot \arctan\left(\frac{\epsilon_0^{\text{shear}}}{\epsilon_1^{\text{shear}}}\right) \\ \gamma &= \frac{\epsilon_0^{\text{shear}}}{\sin(2\phi)}\end{aligned}\quad (6.5)$$

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{ext shear}}$ [/]	[Bergamini et al. 2019] model $\epsilon_0^{\text{ext shear}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$
$\epsilon_1^{\text{ext shear}}$ [/]	[Bergamini et al. 2019] model $\epsilon_1^{\text{ext shear}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$

The source is located at redshift $z = 1.425$ and the filter $F814W$ works in the range [7000, 9500] Å, this means that the light of the lensed source we are observing is emitted in the range [2886.6, 3917.5] Å, UV range ([100, 4000] Å).

The lens-source system is rebuilt in the second fit. To accomplish this, the fitting procedure must take into account the multiple images of the lensed source, which necessitates enlarging the fitting mask to include multiple images (??). We choose a mask radius of 1.44 arcseconds. From here on the fitting mask will not change anymore.

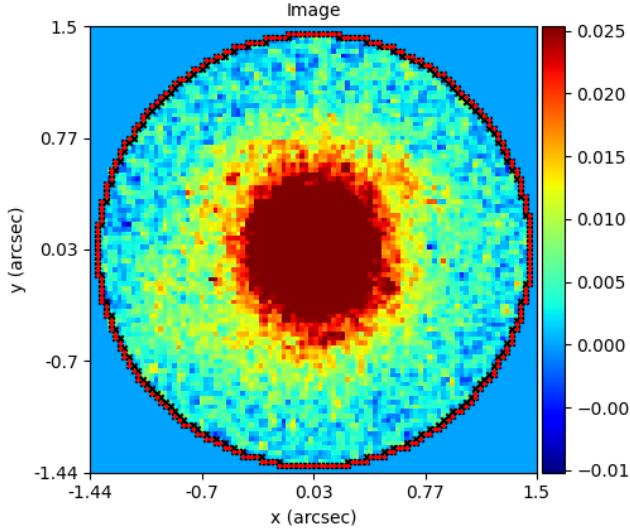


Figure 6.8: This time the mask encircles even the lensed source light. The mask is not centered at 0.0, but at $(0.0264, 0.0336)$ arcseconds.

The model describing the system is:

- Lens mass distribution: 6 parameters in total, the two center coordinates, the axis-ratio, the orientation angle, the Einstein radius (b) and the slope. Four actual parameters, because the center is fixed.
- Source light distribution: 7 parameters in total, section (2.3).
- External shear components: 2 parameters

The fitting procedure, which uses 15 parameters (13 effective), gives the following maximum log-likelihood model parameters (6.3). Visual results are in figures (6.9, 6.10, 6.11, 6.11, 6.12, 6.13).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens mass einstein radius ["]	7.7028×10^{-1}	3.5358×10^{-2}	3.6699×10^{-2}
galaxies lens mass slope [/]	2.9929	2.0364×10^{-1}	4.5311×10^{-2}
galaxies lens mass elliptical comps 0 [/]	-1.1048×10^{-2}	1.1421×10^{-2}	8.4794×10^{-3}
galaxies lens mass elliptical comps 1 [/]	-5.3884×10^{-2}	7.4571×10^{-2}	7.6310×10^{-2}
galaxies source disk centre 0 ["]	1.0818×10^{-2}	3.1671×10^{-2}	3.3401×10^{-2}
galaxies source disk centre 1 ["]	-3.0880×10^{-1}	2.9275×10^{-2}	4.3462×10^{-2}
galaxies source disk elliptical comps 0 [/]	-2.3164×10^{-2}	1.4681×10^{-1}	1.2477×10^{-1}
galaxies source disk elliptical comps 1 [/]	2.7373×10^{-1}	1.5241×10^{-1}	1.6373×10^{-1}
galaxies source disk intensity [/]	5.1856×10^{-5}	2.1384×10^{-5}	4.1462×10^{-5}
galaxies source disk effective radius ["]	4.1213	2.0188	7.3794
galaxies source disk sersic index [/]	3.8350	1.0766	6.7706×10^{-1}
galaxies lens shear elliptical comps 0 [/]	2.5966×10^{-1}	3.4116×10^{-2}	3.7290×10^{-2}
galaxies lens shear elliptical comps 1 [/]	1.0767×10^{-1}	3.5916×10^{-2}	4.0434×10^{-2}

Table 6.3: Outcomes of the second optimization process for the parameters

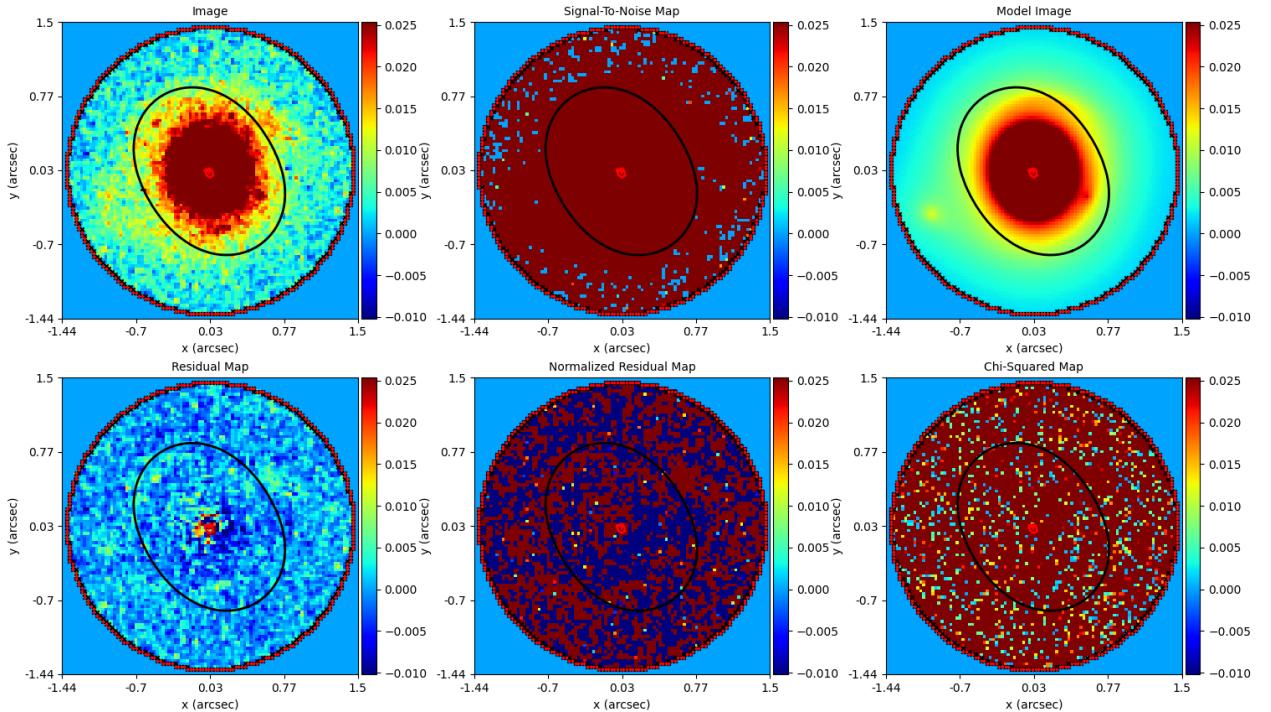


Figure 6.9: The outcome derived from the second fitting procedure.

The critical lines that appears into the images are those modified by the presence of the shear (1.60).

NOTE: The red piecewise constant curve at the center should not be there, because it is the radial critical

lines. The slope is too steep for the radial critical lines to exist, they should be an infinitesimal zero dimension small point in the center, but the software in this version and in this case has not the resolution to deal with this and the radial critical lines appear.

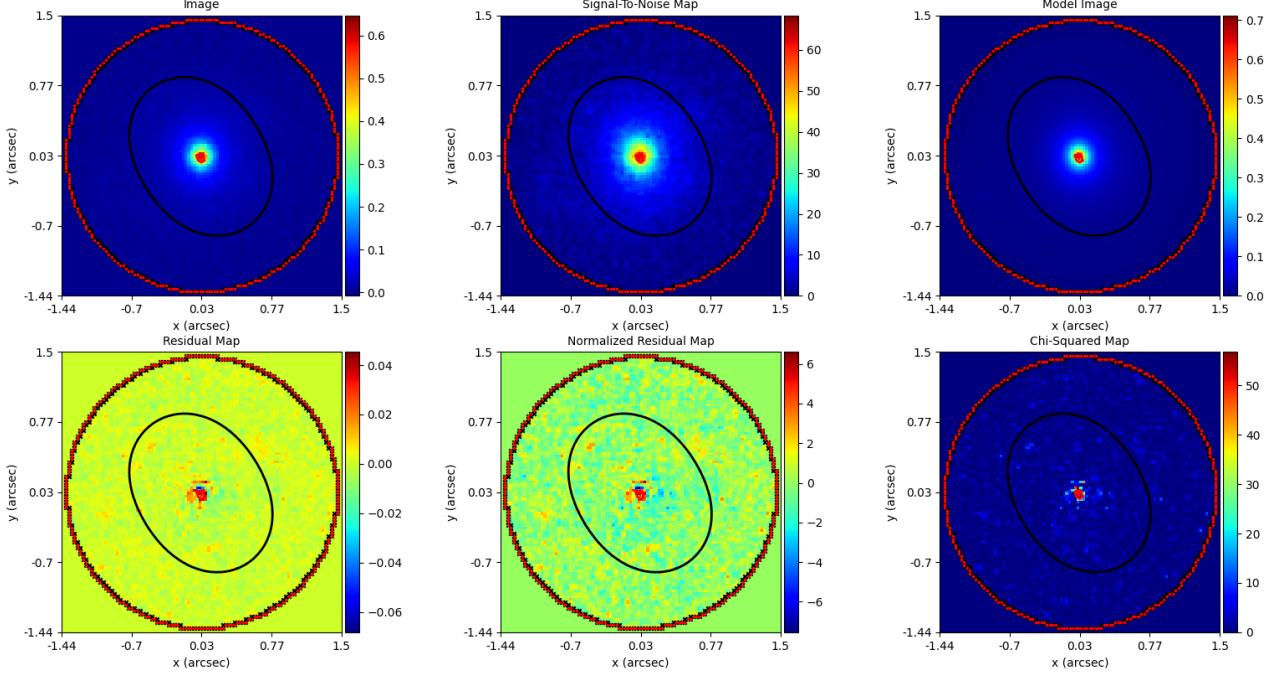


Figure 6.10: The outcome of the second fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

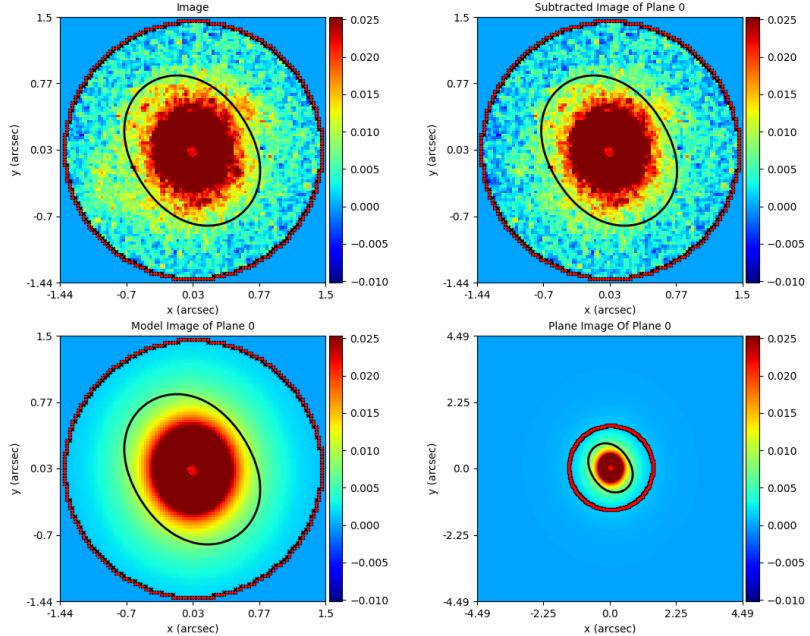


Figure 6.11: Outcome of the second fitting procedure. The subtracted image of plane 0 is the image from which what has been recognized as the light of the source has been taken away.

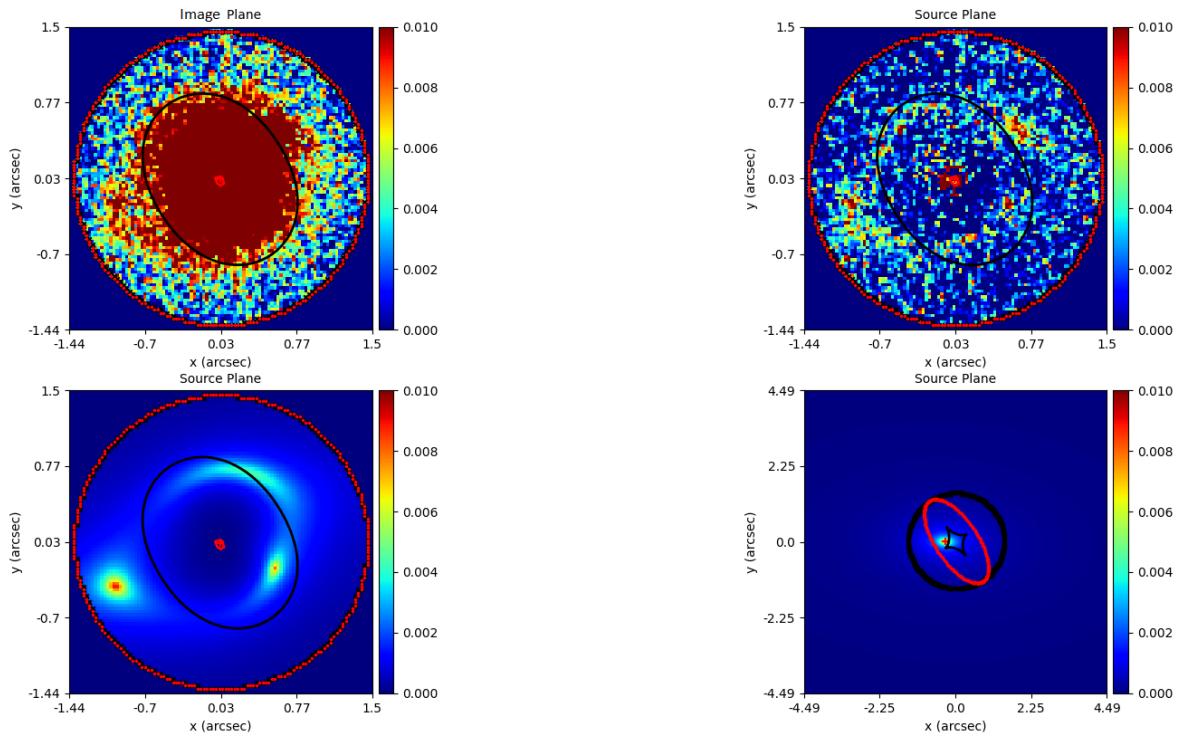


Figure 6.12: The source resulting from the second fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

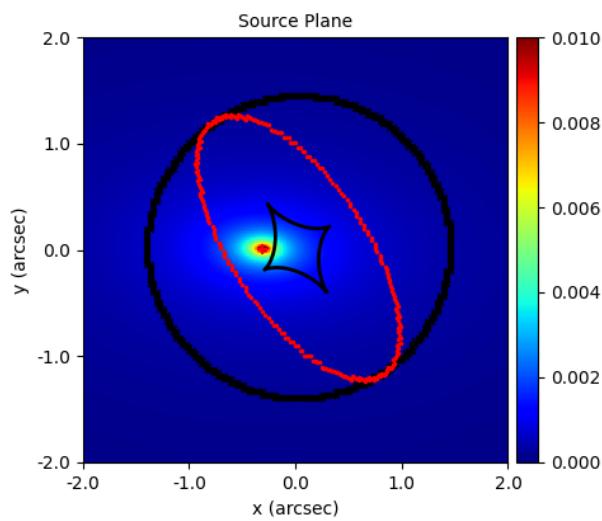


Figure 6.13: The source resulting from the second fit. The black circle is the mask used for the fitting procedure. The red ellipse is the mask traced to the source plane.

6.3 3 Fit: Last parametric fit of the source

The third fit aims to clean up and redefine the results of the second fit more precisely, so we can create a hyper image as precise as possible to be used in the fourth fit, where adaptive reconstruction will be introduced (4.4.2). In the third fit parametric model both for the source and the lens are still used. The model for the system is the following:

- The lens' bulge is described by an elliptical Sérsic profile. The prior on the parameters is described in section (4.3).
- The center of the mass distribution of the lens has been set equal to the center of the light distribution:

$$\begin{aligned} x_{\text{lens}}^{\text{lens}} &= x_{\text{center}}^{\text{lens bulge}} \\ y_{\text{lens}}^{\text{lens}} &= y_{\text{center}}^{\text{lens bulge}} \end{aligned} \quad (6.6)$$

- A gaussian prior was imposed on the elliptical components of the lens galaxy, using the results concerning the elliptical components of fit 2:

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{lens mass}} [/]$	max loglikelihood value $\epsilon_0^{\text{lens mass}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$
$\epsilon_1^{\text{lens mass}} [/]$	max loglikelihood value $\epsilon_1^{\text{lens bulge}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$

This time the gaussian was truncated at 2σ instead of 3σ , because we need to redefine orientation and ellipticity, but their values were already found in the previous fit. However, a certain degree of freedom is left.

- The prior on the other lens' parameters is described in section (4.3).
- A gaussian prior was imposed on the elliptical components of the lens shear, using the results concerning the elliptical components of fit 2:

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{shear}} [/]$	max loglikelihood value $\epsilon_0^{\text{shear}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$
$\epsilon_1^{\text{shear}} [/]$	max loglikelihood value $\epsilon_1^{\text{shear}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 3\sigma$	$\text{mean} + 3\sigma$

- The source parameters are set equal to the results of the previous fit.

The fitting procedure, which uses 15 parameters (13 effective, because 2 are fixed), gives the following maximum log-likelihood model parameters (6.4). Visual results are in figures(6.14, 6.15, 6.16, 6.17, 6.18).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens bulge centre 0 ["]	2.0776×10^{-2}	8.6640×10^{-4}	8.5123×10^{-4}
galaxies lens bulge centre 1 ["]	2.3063×10^{-2}	7.1170×10^{-4}	7.7538×10^{-4}
galaxies lens bulge elliptical comps 0 [/]	-4.7162×10^{-3}	5.6496×10^{-3}	5.4355×10^{-3}

Table 6.4: Outcomes of the third optimization process for the parameters.

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens bulge elliptical comps 1 [/]	-6.3244×10^{-2}	6.0398×10^{-3}	5.9641×10^{-3}
galaxies lens bulge intensity [/]	3.7347×10^{-2}	1.7189×10^{-3}	2.1230×10^{-3}
galaxies lens bulge effective radius ["]	3.6328×10^{-1}	1.1398×10^{-2}	9.5424×10^{-3}
galaxies lens bulge sersic index [/]	2.7606	8.2949×10^{-2}	7.2763×10^{-2}
galaxies lens mass slope [/]	2.8758	2.2206×10^{-1}	1.1478×10^{-1}
galaxies lens mass elliptical comps 0 [/]	-1.3075×10^{-2}	1.0414×10^{-2}	1.1267×10^{-2}
galaxies lens mass elliptical comps 1 [/]	-1.0210×10^{-1}	4.5845×10^{-2}	5.9566×10^{-2}
galaxies lens mass einstein radius ["]	7.9827×10^{-1}	3.0688×10^{-2}	2.7014×10^{-2}
galaxies lens shear elliptical comps 0 [/]	2.4127×10^{-1}	4.1997×10^{-2}	3.8396×10^{-2}
galaxies lens shear elliptical comps 1 [/]	9.1947×10^{-2}	4.0119×10^{-2}	3.0082×10^{-2}

Table 6.5: Outcomes of the third optimization process for the parameters.

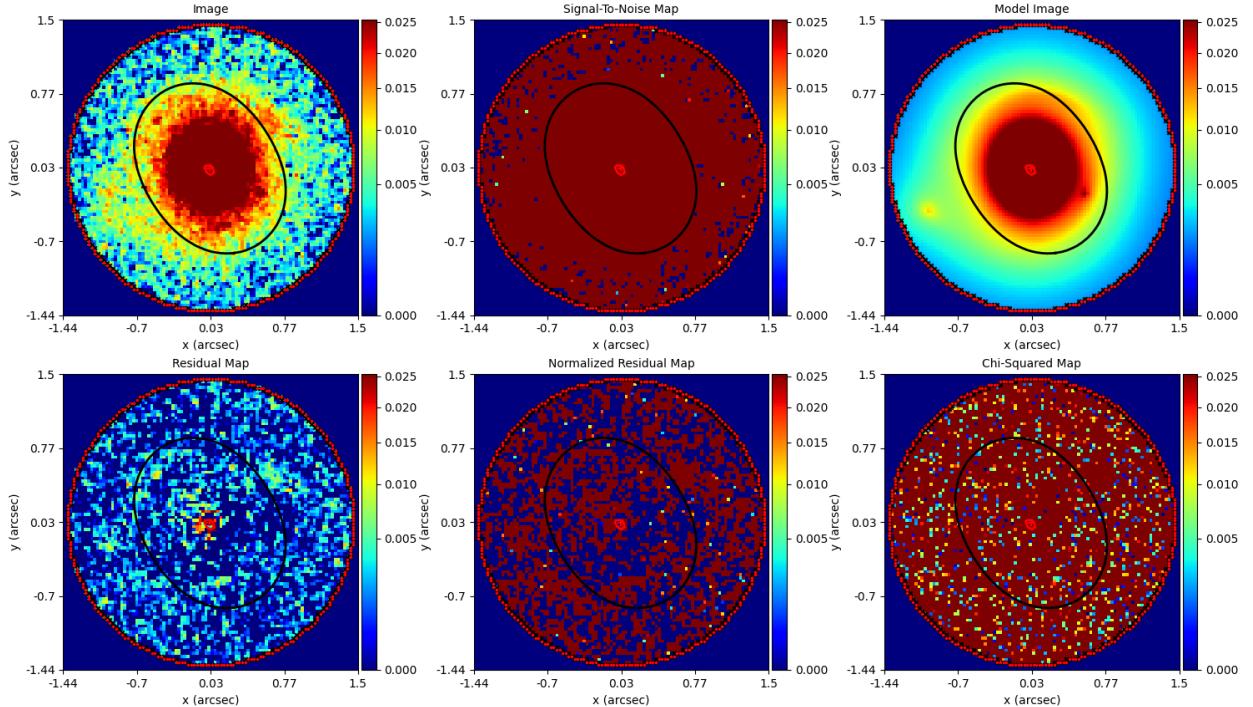


Figure 6.14: The outcome derived from the third fitting procedure.

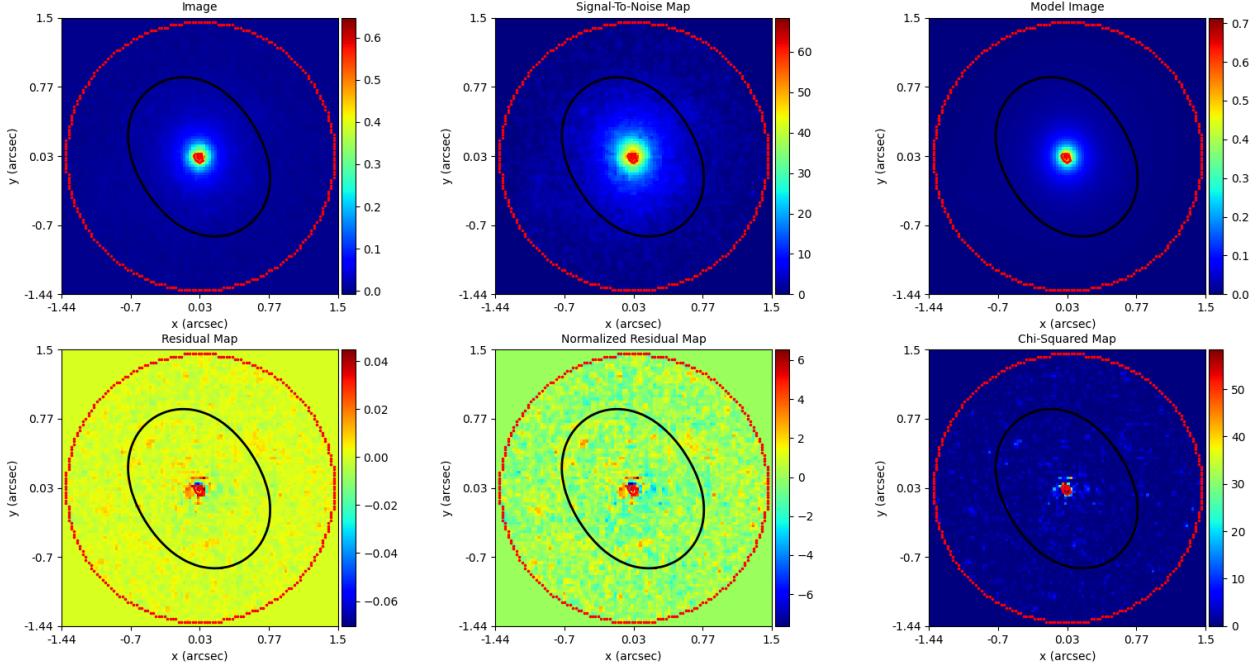


Figure 6.15: The outcome of the third fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

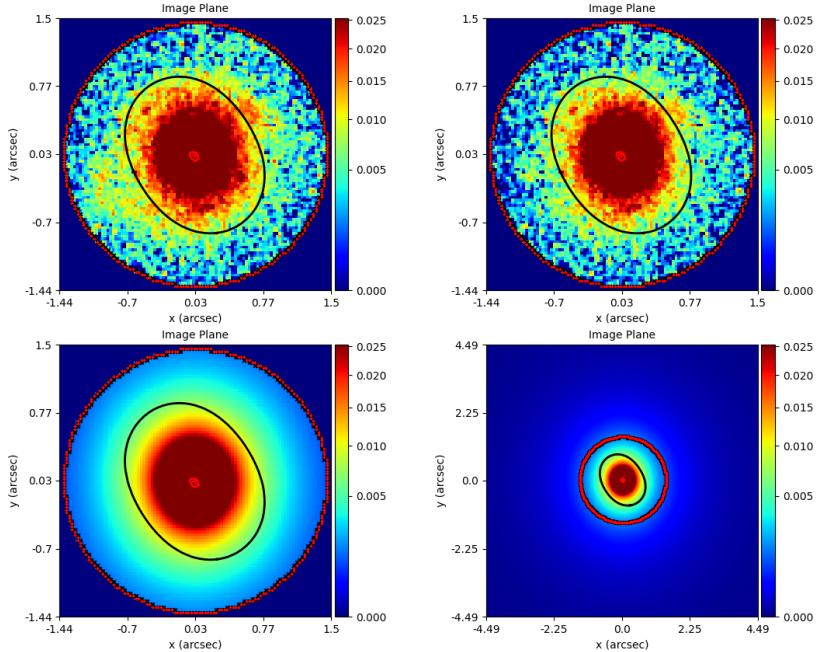


Figure 6.16: Result of the third fit. The subtracted image of plane 0 is the image from which what has been recognized as the light of the source has been taken away.

The critical lines that appears into the images are those modified by the presence of the shear (1.60). The slope is to steep for the radial critical lines to exist, they should be an infinitesimal zero dimension small point in the center, but the software in this version and in this case has not the resolution to deal with this and the radial critical lines appear.

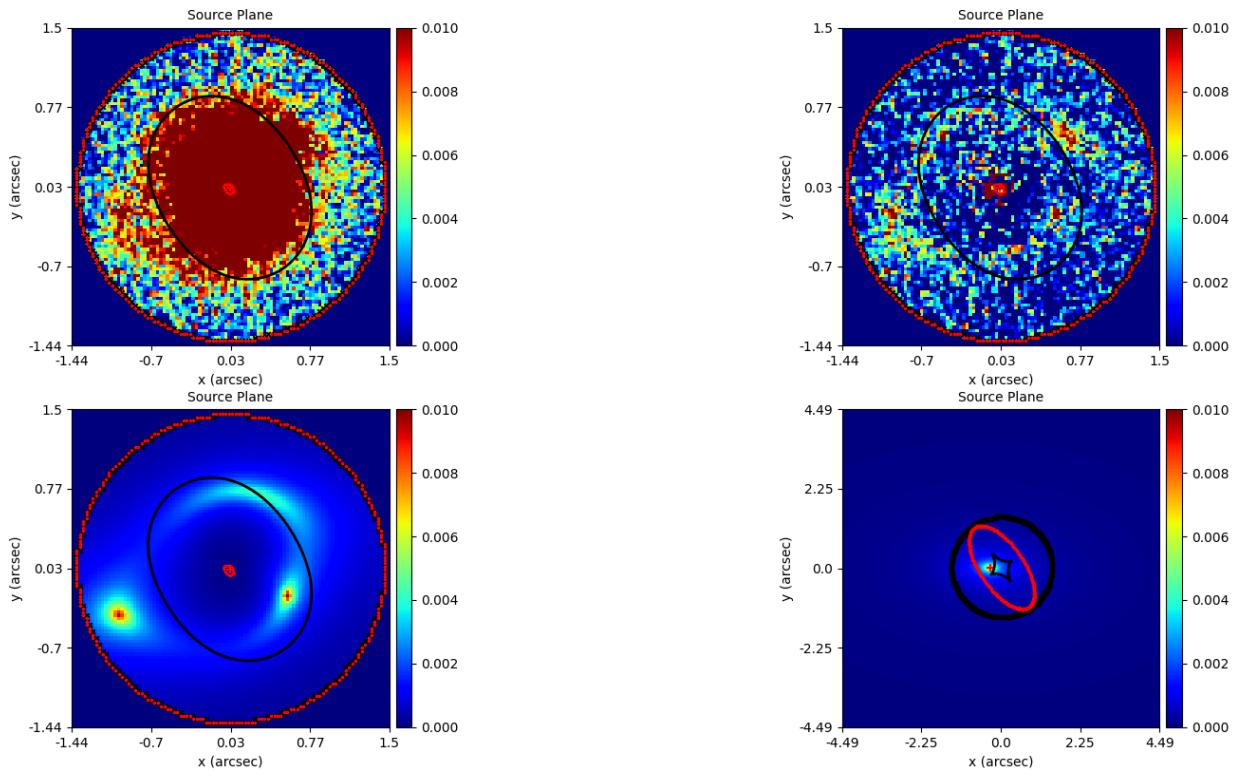


Figure 6.17: The source resulting from the second fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

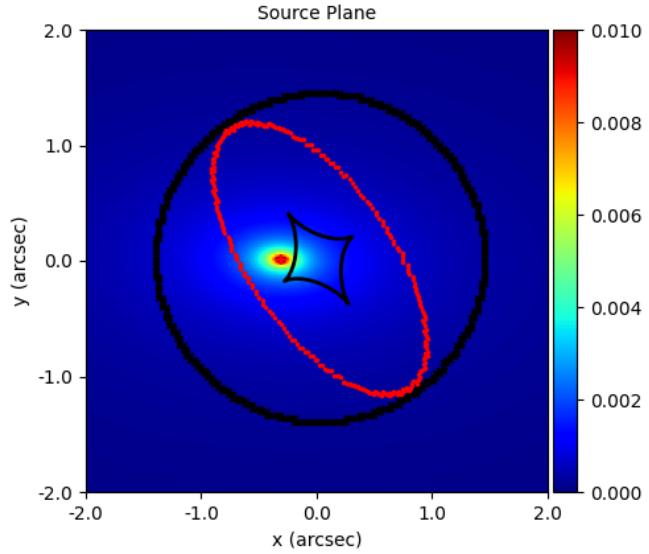


Figure 6.18: The source resulting from the second fit. The red ellipse is the mask traced to the source plane.

6.4 4 Fit: first adaptive reconstruction of the source

With the third fit, we now have the hyper-image of the source and we can therefore proceed with the adaptive reconstruction of this. The first use of adaptive reconstruction involves that pixels are concentrated more around the regions of maximum magnification in the plane of the source, i.e. the caustics (4.4.2). The fit number four is used to find the parameters that best describe the adaptive grid on the source plane. In this case all the parameters that concerns the lens plane, thus lens mass, light and shear are kept fixed, and the only free parameters are those of the adaptive grid: the regularization, constant in this model, and the distribution of the pixels around the regions of the caustics in the source plane.

Specifically, the code is configured as presented in (6.2):

```

1 model = af.Collection(
2     galaxies=af.Collection(
3         lens=af.Model(
4             al.Galaxy,
5             redshift=0.44,
6             bulge=result_3.instance.galaxies.lens.bulge, # fixed
7             mass=result_3.instance.galaxies.lens.mass, # fixed
8             shear=result_3.instance.galaxies.lens.shear, # fixed
9             hyper_galaxy=setup_hyper.hyper_galaxy_lens_from(result=result_3), # galaxy hyper-image
10        ),
11        source=af.Model(
12            al.Galaxy,
13            redshift=1.425,
14            pixelization=al.pix.DelaunayMagnification, #distribution of pixels around the caustics
15            regularization=al.reg.Constant,# regularization
16        ),
17    )
)

```

Listing 6.2: Input commands for the PyAutoLens model representing the adaptive image

The fitting procedure, which uses 3 parameters gives the following maximum log-likelihood model parameters (6.6). Visual results are in figures (6.19, 6.20, 6.21, 6.22, 6.23, 6.24a, 6.24b).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies source pixelization shape 0 [/]	2.4414×10^1	1.3510×10^1	1.1024×10^1
galaxies source pixelization shape 1 [/]	3.8765×10^1	1.8127×10^1	6.2069
galaxies source regularization coefficient [/]	2.4605×10^2	5.1760×10^1	9.5189×10^1

Table 6.6: Outcomes of the fourth optimization process for the parameters.

Among these results, however, the primary outcome of the fitting process relates to the determination of the elements within the triangular mesh grid. In figure (6.19) it is possible to observe the corner of the triangles that constitute the Delaunay triangular mesh grid (4.6).

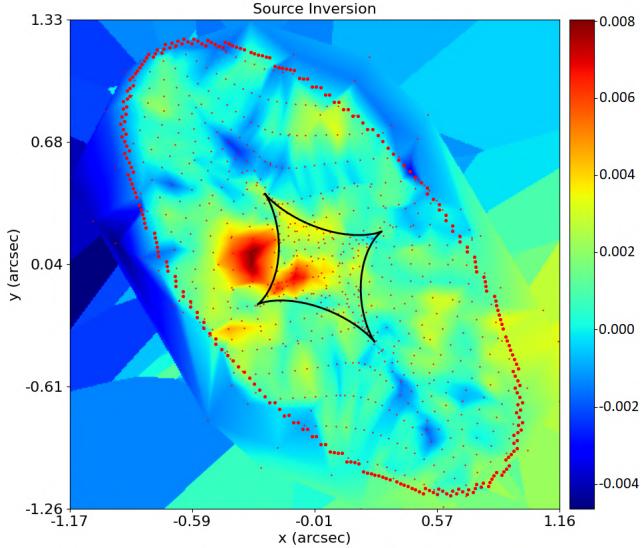


Figure 6.19: Source plane and vertices (red dots) of Delaunay triangular mesh grid.

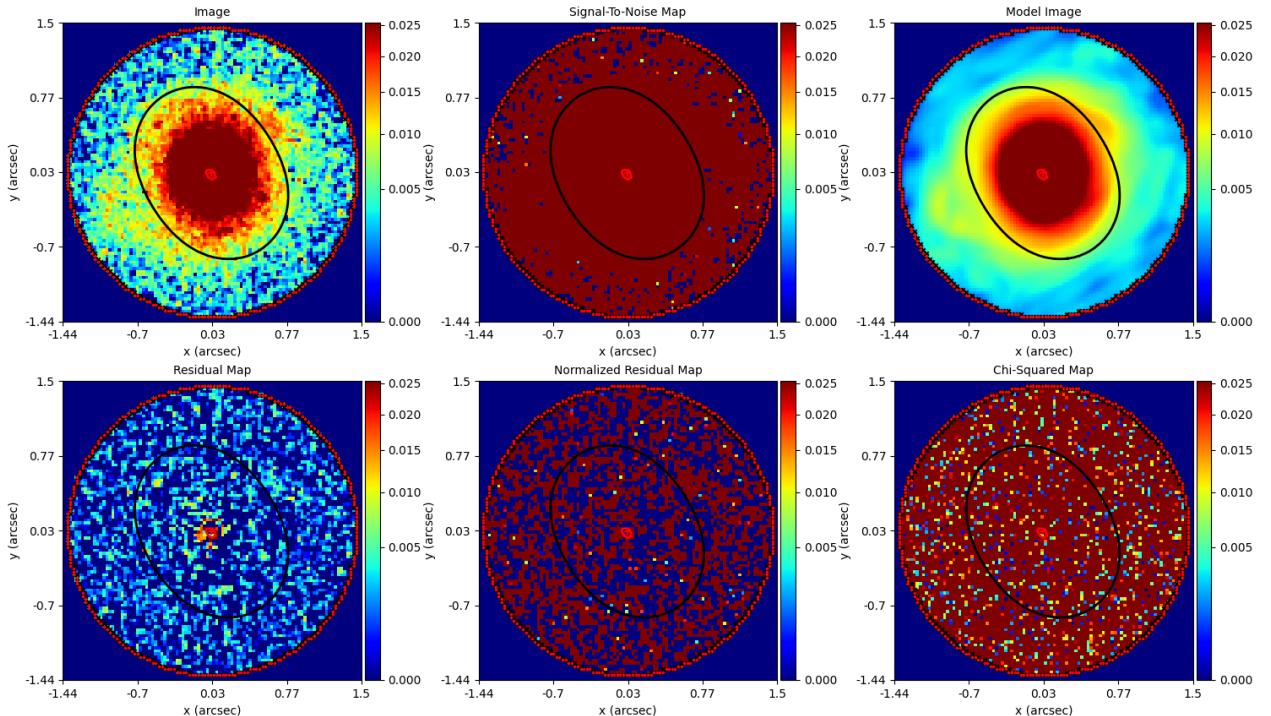


Figure 6.20: The outcome derived from the fourth fitting procedure.

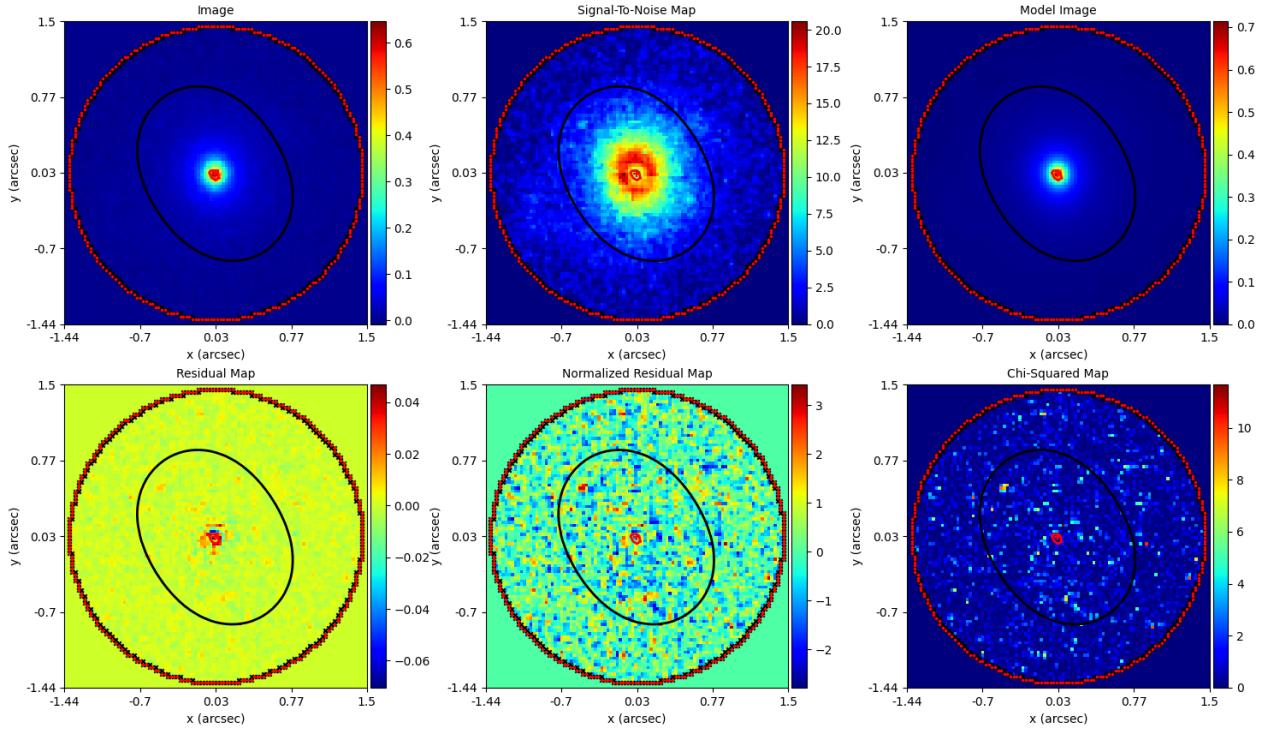


Figure 6.21: The outcome of the fourth fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

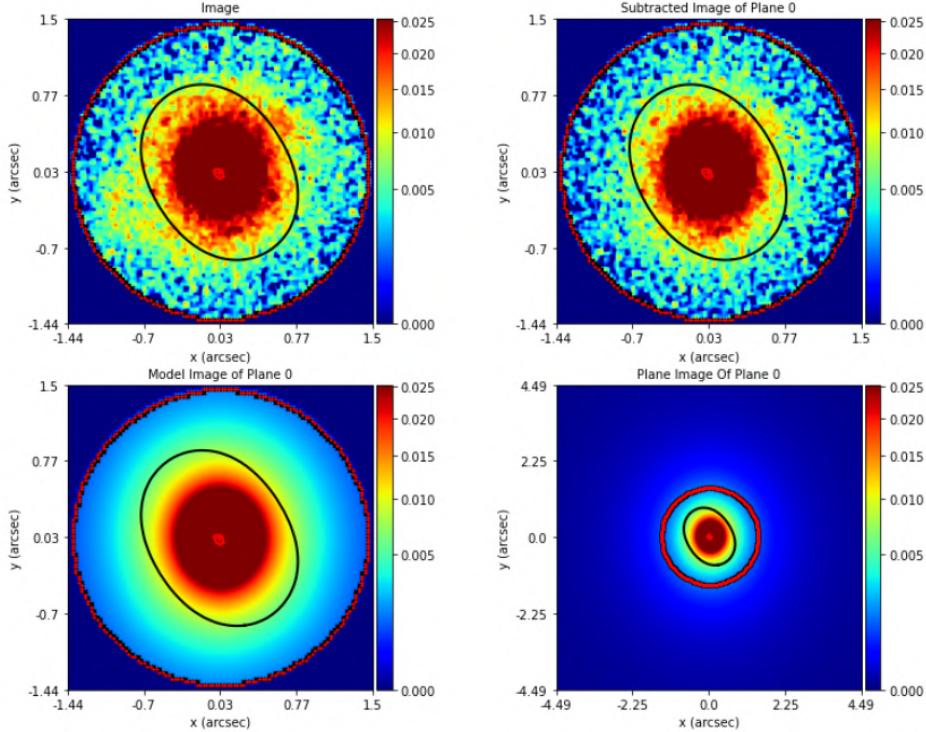


Figure 6.22: The lens plane resulting from the fourth fit. The subtracted image of plane 0 is the the lens light without the lensed light of the source.

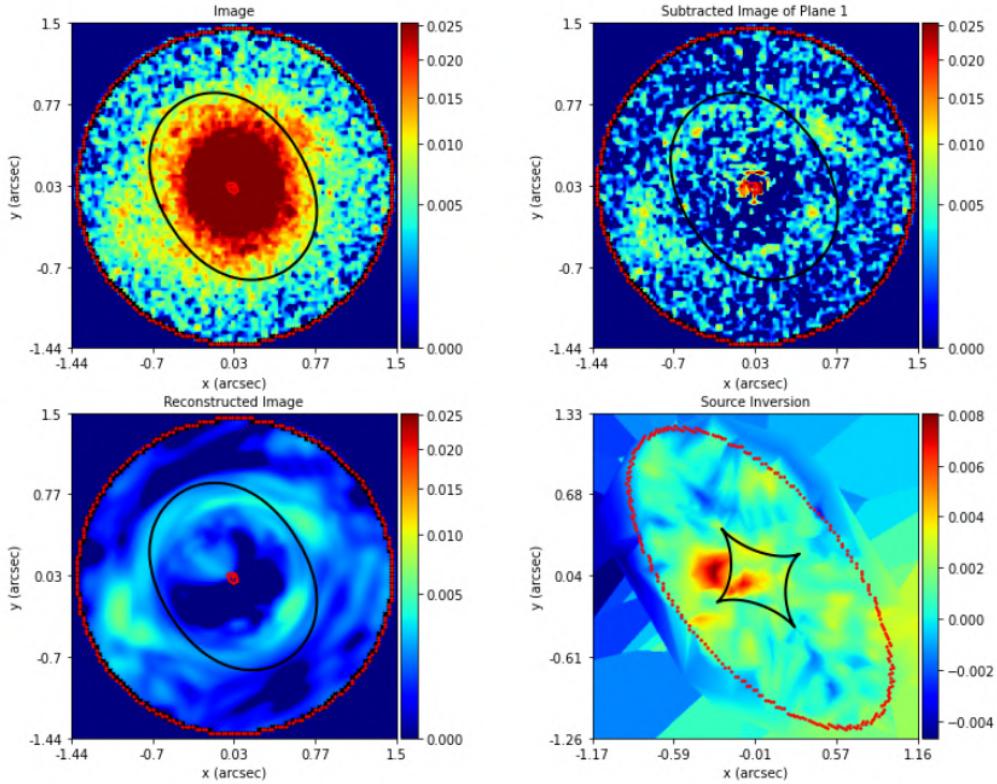


Figure 6.23: The source resulting from the fourth fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

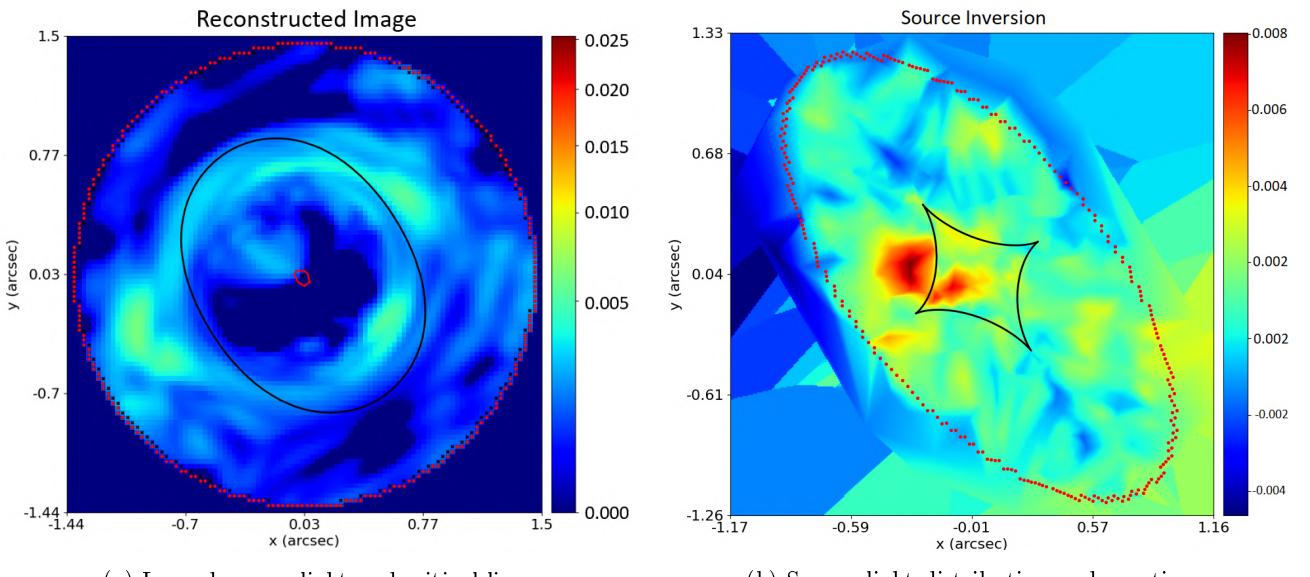


Figure 6.24: Dimensionless source light, normalized to $6.87 \cdot 10^{-20}$ [erg/s · cm $^{-2}$ · Å].

6.5 5 fit: fitting the lens plane with adaptive source reconstruction for the first time

In the fit number five we look for the lens plane quantities using for the first time the source reconstructed with an adaptive grid (4.4).

In this case the model is the following one

- The center of the mass distribution of the lens has been set equal to the center of the light distribution from the third fit:

$$\begin{aligned} x_{\text{center}}^{\text{lens}} &= x_{\text{center}}^{\text{lens bulge}} \\ y_{\text{center}}^{\text{lens}} &= y_{\text{center}}^{\text{lens bulge}} \end{aligned} \quad (6.7)$$

- The lens light distribution parameters are fixed using the results from fit number three.
- The shear elliptical components are two of the free parameters of the model, and on them a prior is imposed. The software uses the information of the file `*.json(autolens_workspace/config/priors/mass_sheet.json)` and of the fit 3 results (`model.results`) and deals with them following the gaussian prior theory in section (4.3).

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{shear}}$	PDF median value fit 3 $\epsilon_0^{\text{shear}}$	0.05	$-\infty$	∞
$\epsilon_1^{\text{shear}}$	PDF median value fit 3 $\epsilon_1^{\text{shear}}$	0.05	$-\infty$	∞

- A gaussian prior was imposed on the elliptical components of the lens galaxy, using the results concerning the elliptical components of fit 3:

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{lens mass}}$	max loglikelihood value $\epsilon_0^{\text{lens mass}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$
$\epsilon_1^{\text{lens mass}}$	max loglikelihood value $\epsilon_1^{\text{lens mass}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$

This time the gaussian was truncated at 2σ instead of 3σ , because we need to redefine orientation and ellipticity, but their values were already found in the previous fit. However, a certain degree of freedom is left.

- The prior on the other lens' parameters is described in section (4.3).

The fitting procedure, which uses 8 parameters (6 effective, because two are fixed), gives the following maximum log-likelihood model parameters (6.7, 6.8). Visual results are in figures(6.25, 6.26, 6.27, 6.28, 6.29a, 6.29b).

Parameter	Value	Lower error (3 σ)	Upper error (3 σ)
galaxies lens mass slope [/]	2.9993	1.9821×10^{-1}	3.3649×10^{-2}
galaxies lens mass einstein radius ["]	7.8847×10^{-1}	4.9098×10^{-2}	5.3276×10^{-2}

Table 6.7: Outcomes of the fifth optimization process for the parameters.

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens shear elliptical comps 0 [/]	2.4192×10^{-1}	5.2177×10^{-2}	5.1744×10^{-2}
galaxies lens shear elliptical comps 1 [/]	5.6034×10^{-2}	4.0016×10^{-2}	5.2507×10^{-2}
galaxies lens mass elliptical comps 0 [/]	-1.1115×10^{-2}	1.1839×10^{-2}	1.3271×10^{-2}
galaxies lens mass elliptical comps 1 [/]	-7.4779×10^{-2}	9.1256×10^{-2}	1.0646×10^{-1}

Table 6.8: Outcomes of the fifth optimization process for the parameters.

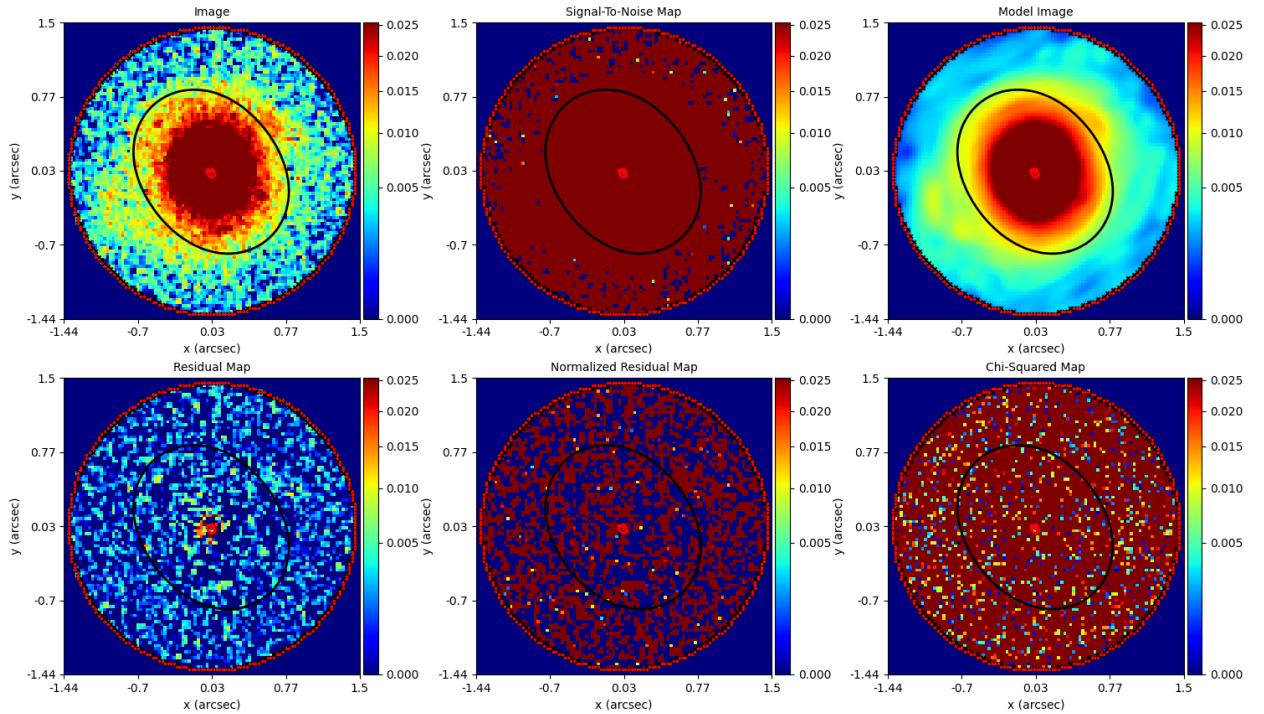


Figure 6.25: The outcome derived from the fifth fitting procedure.

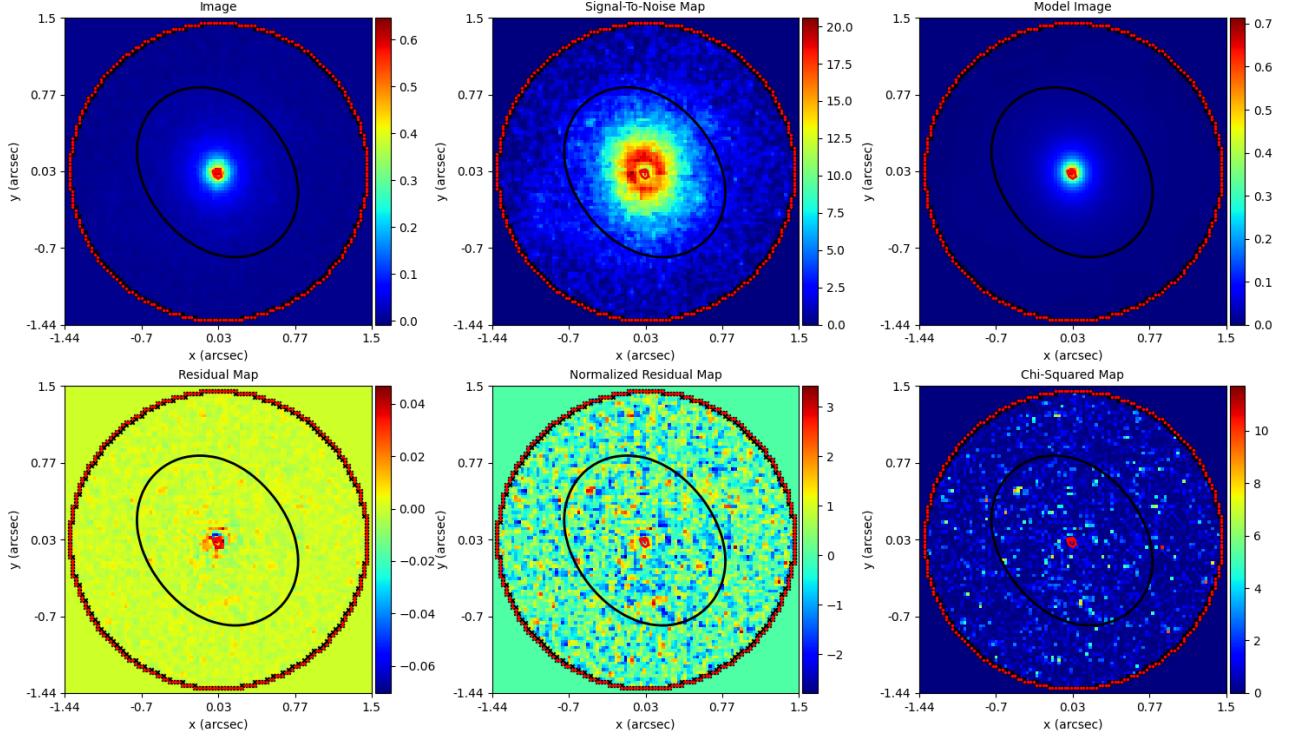


Figure 6.26: The outcome of the initial fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

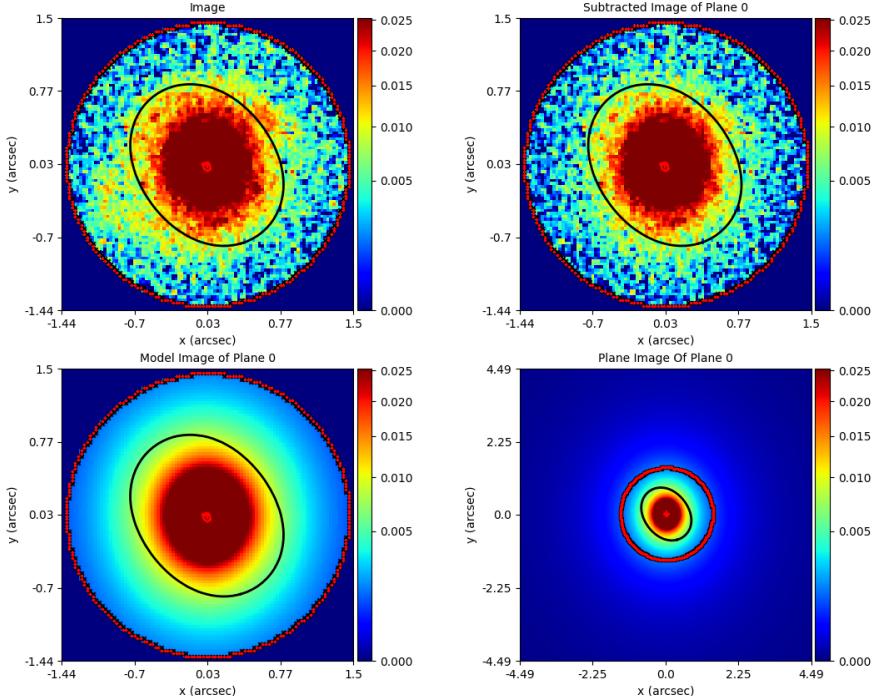


Figure 6.27: The lens plane resulting from the fifth fit. The subtracted image of plane 0 is the the lens light without the lensed light of the source.

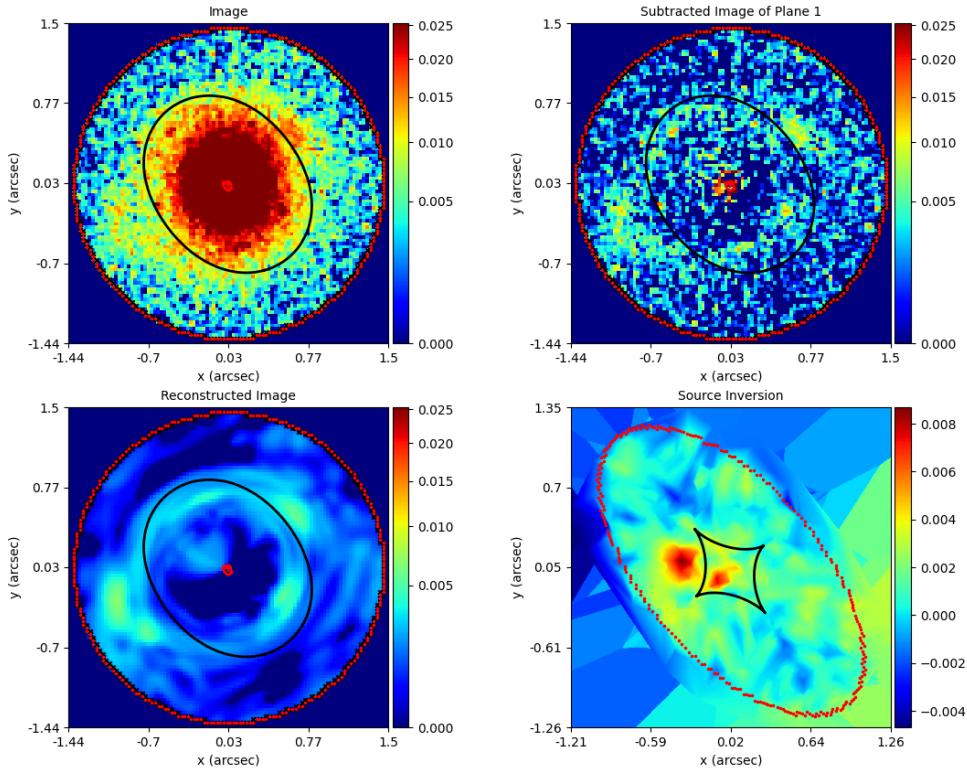


Figure 6.28: The source resulting from the fifth fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

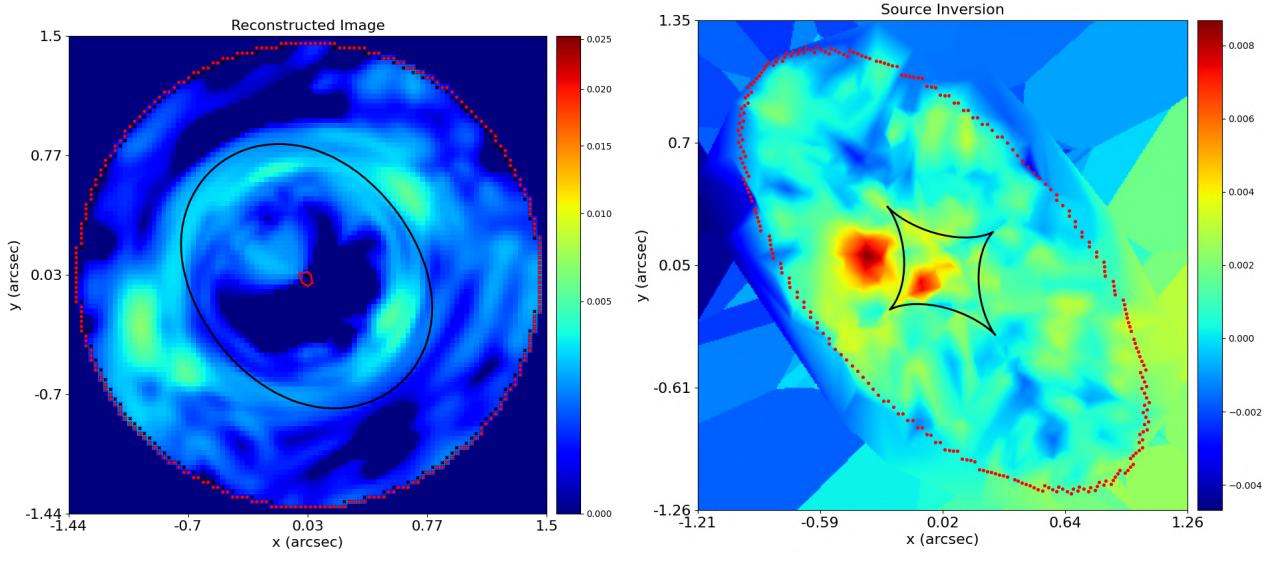


Figure 6.29: Dimensionless source light, normalized to $6.87 \cdot 10^{-20}$ [erg/s · cm $^{-2}$ · Å].

6.6 6 Fit: Delaunay Brightness Image adaptive grid

In section (4.4.3) is pointed out the reason why we need to improve the previous results. The main goal of this fit are mainly two:

- Use the adaptive regularization.
- Congregate the triangular mesh grid elements around the position of the source.

Similar to fit number four, all parameters pertaining to the lens plane, specifically those characterizing the lens light, lens mass, and shear components, remain fixed. The only degrees of freedom are associated with the source, encompassing the adaptive regularization and the Delaunay triangular mesh grid. These parameters, along with their corresponding maximum log-likelihood values, are presented in Table 6.9. Additionally, Figures (6.31, 6.32, 6.33, 6.34, 6.35a, and 6.35b) visually depict the outcomes of the fitting process.

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies source pixelization pixels [/]	1.3061×10^3	5.6232×10^2	8.2117×10^2
galaxies source pixelization weight floor [/]	7.2311×10^{-1}	1.9211×10^{-1}	3.1308×10^{-2}
galaxies source pixelization weight power [/]	5.7274	5.6298	2.7119
galaxies source regularization inner coefficient [/]	8.8355	7.7833	3.7283
galaxies source regularization outer coefficient [/]	2.1204×10^1	5.4715	7.6893
galaxies source regularization signal scale [/]	3.3118×10^{-1}	2.2832×10^{-1}	3.9618×10^{-1}

Table 6.9: Outcome of the sixth fitting procedure.

Among these results, however, the primary outcome of the fitting process relates to the determination of the elements within the triangular mesh grid. In figure (6.30) it is possible to observe the corner of the triangles that constitute the Delaunay triangular mesh grid (4.6).

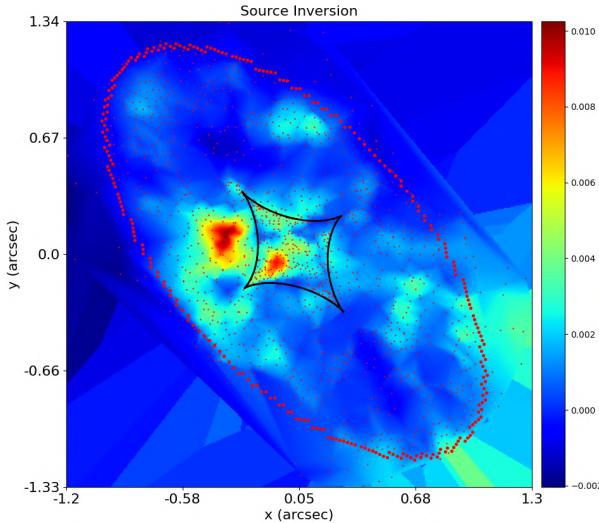


Figure 6.30: Source plane and vertices (red dots) of Delaunay triangular mesh grid.

Even in this case there are a lot of red dots (triangles vertices) near to the caustics, but this is due to the fact that the source of light, where the triangular mesh grid elements are congregated, are near to the caustics too. Though, in this case we can see that the triangles are congregated at the positions of the light sources and there are voids (there still are the triangles but in a lower number and they are bigger) where there is no source.

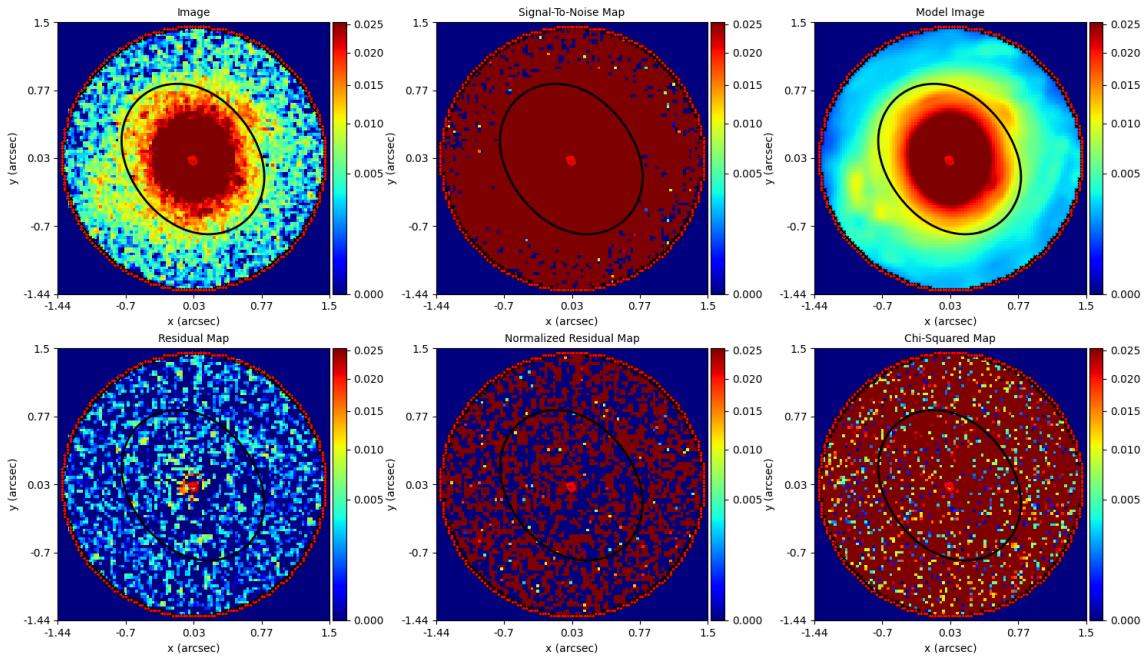


Figure 6.31: The outcome derived from the sixth fitting procedure.

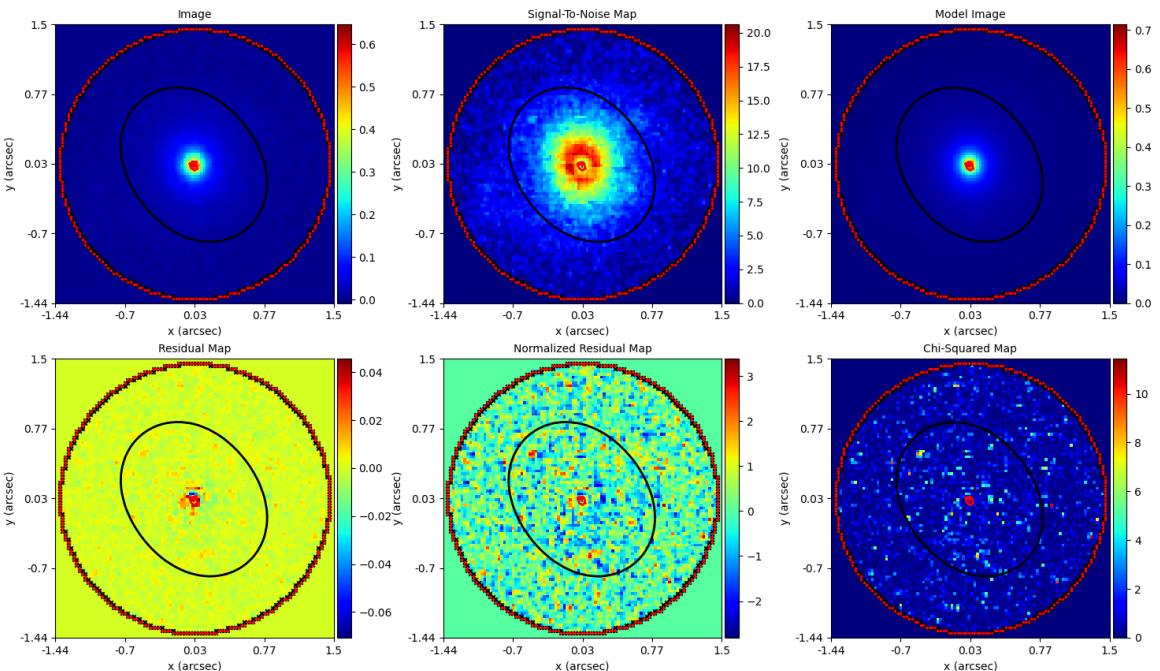


Figure 6.32: The outcome of the sixth fitting is presented with an alternative normalization aimed at enhancing the visibility of elements not perceptible in the initial image. This, in particular, includes the Signal-to-Noise map and the Chi-Square map.

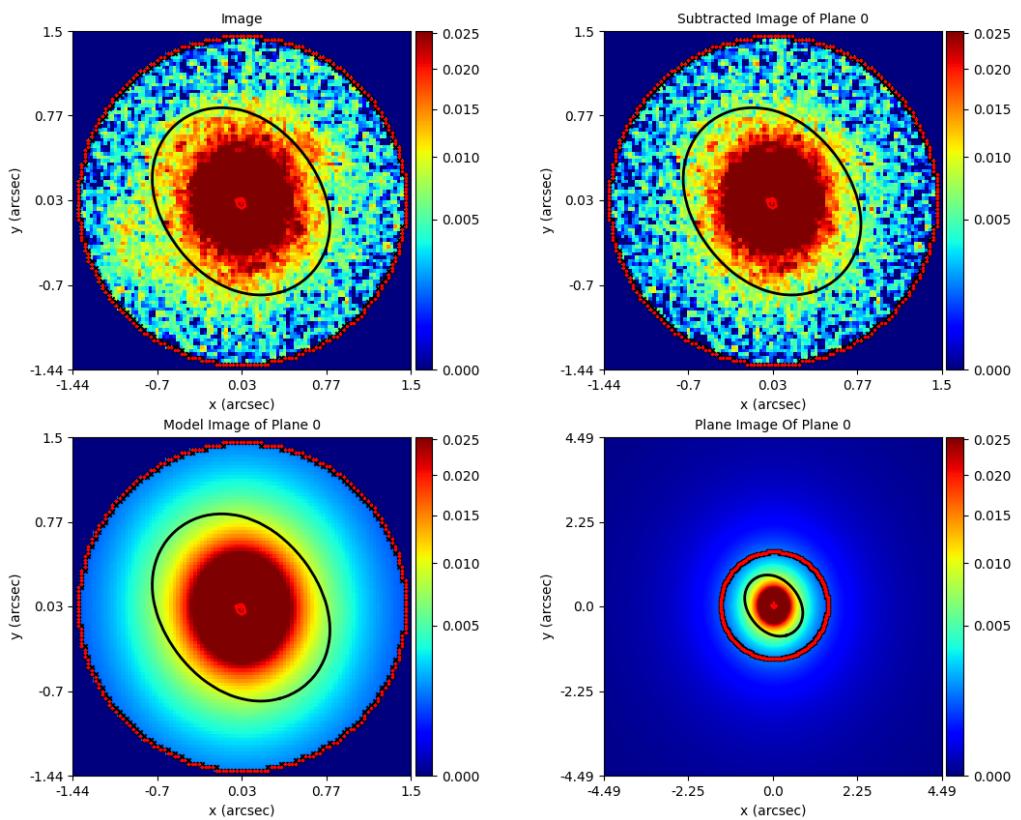


Figure 6.33: The lens plane resulting from the sixth fit. The subtracted image of plane 0 is the the lens light without the lensed light of the source.

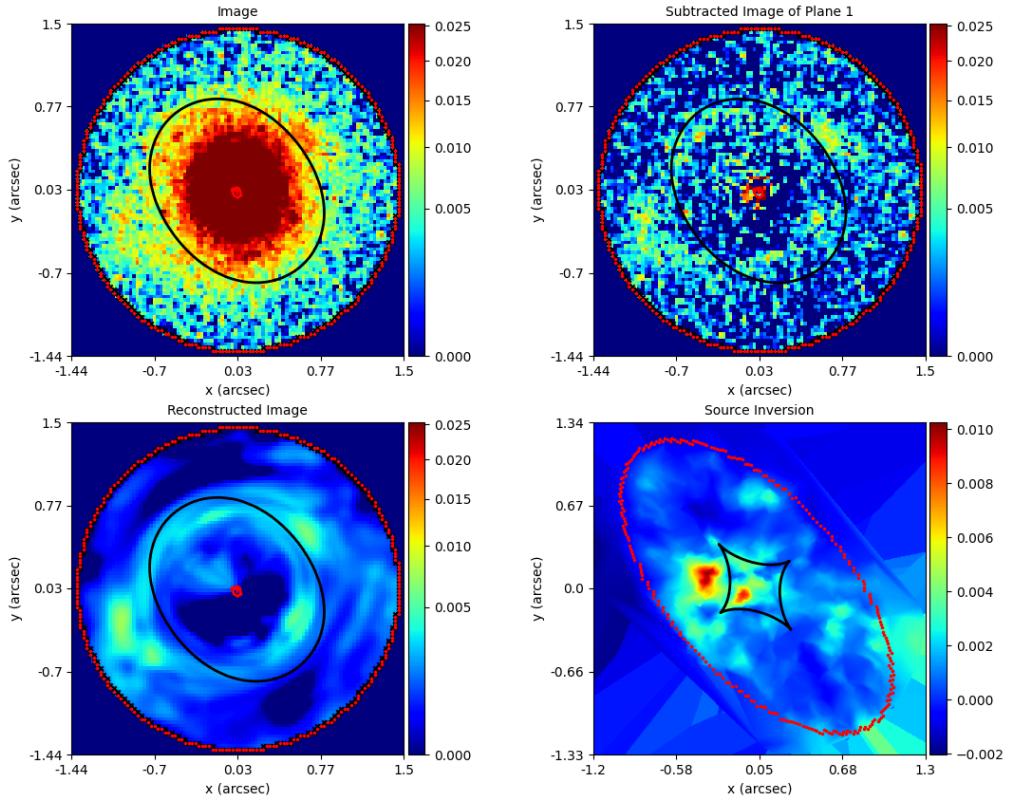
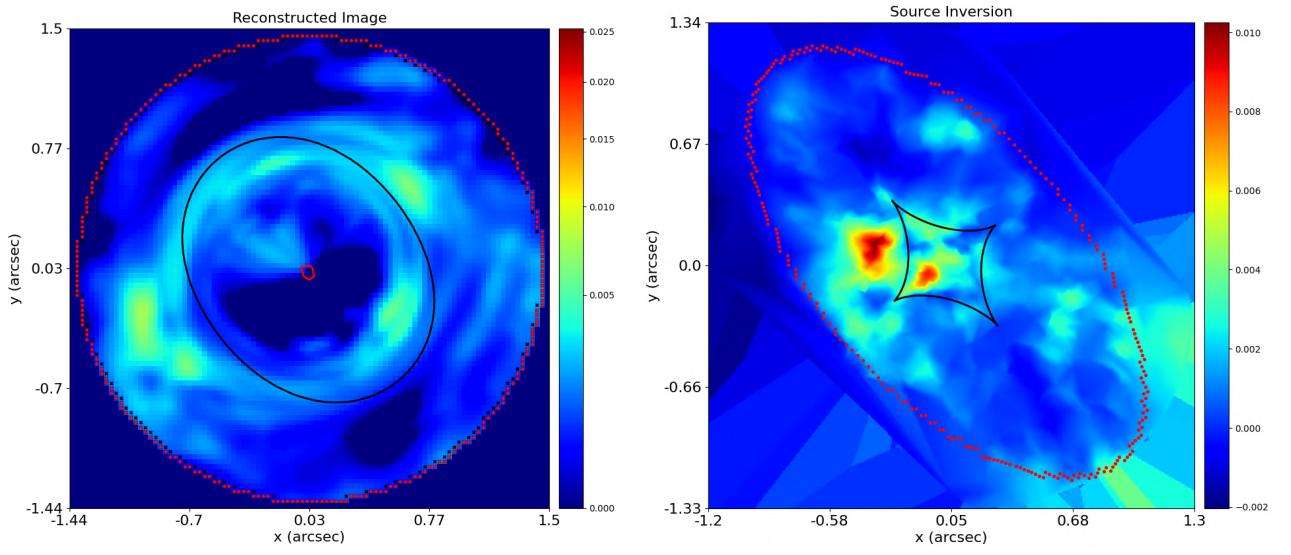


Figure 6.34: The source resulting from the sixth fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.



(a) Lensed source light and critical lines

(b) Source light distribution and caustics

Figure 6.35: Dimensionless source light, normalized to $6.87 \cdot 10^{-20}$ [erg/s · cm $^{-2}$ · Å].

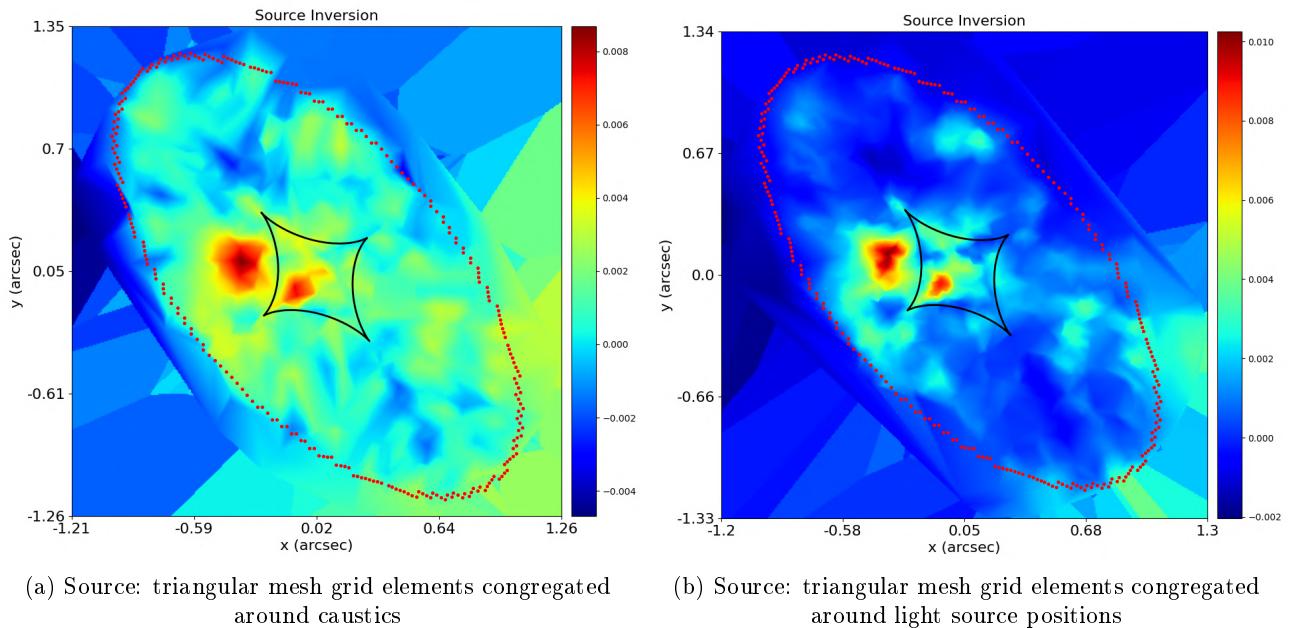


Figure 6.36: Dimensionless source light, normalized to $6.87 \cdot 10^{-20}$ [erg/s · cm $^{-2}$ · Å].

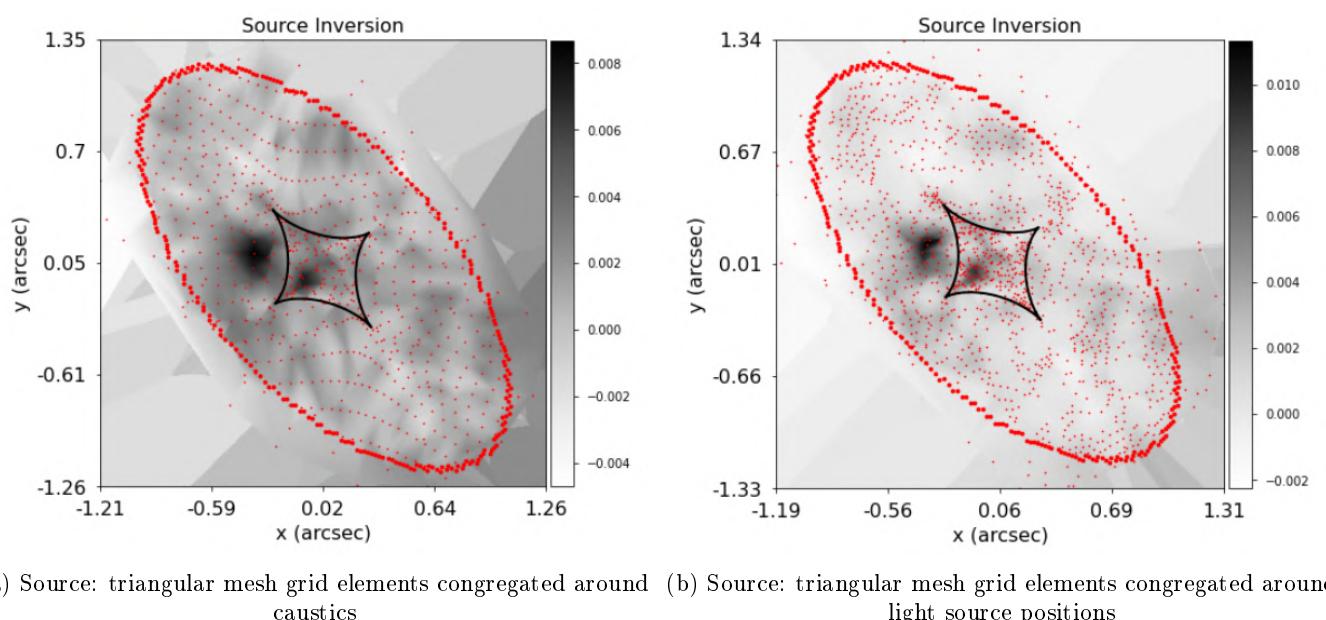


Figure 6.37: Source and triangular mesh grid elements

6.7 7 Fit: Last fit

6.7.1 Elliptical power-law

The fit number 7 is the last fit. The best model of the source is completed (fit 6), now we only have to use this to properly fit the lens plane.

In this case the model is the following one:

- The center of the mass distribution of the lens has been set equal to the center of the light distribution from the third fit (after the third fit the lens light never changed):

$$\begin{aligned} x_{\text{center}}^{\text{lens}} &= x_{\text{center}}^{\text{lens bulge}} \\ y_{\text{center}}^{\text{lens}} &= y_{\text{center}}^{\text{lens bulge}} \end{aligned} \quad (6.8)$$

- The lens light distribution parameters are fixed using the results from fit number three (fit 3 was the last time the light profile was fitted).
- The shear elliptical components are two of the free parameters of the model, and on them a prior is imposed. The software uses the information of the file `*.json(autolens_workspace/config/priors/mass_sheet.json)` and of the fit 5 results (`model.results`) and deals with them following the gaussian prior theory in section (4.3).

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{shear}}$	PDF median value fit 5 $\epsilon_0^{\text{shear}}$	3σ error fit 5	$-\infty$	∞
$\epsilon_1^{\text{shear}}$	PDF median value fit 5 $\epsilon_1^{\text{shear}}$	3σ error fit 5	$-\infty$	∞

- A gaussian prior was imposed on the elliptical components of the lens galaxy, using the results concerning the elliptical components of fit 5:

Parameter	Mean	σ	Lower limit	Upper limit
$\epsilon_0^{\text{lens mass}}$	max loglikelihood value $\epsilon_0^{\text{lens bulge}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$
$\epsilon_1^{\text{lens mass}}$	max loglikelihood value $\epsilon_1^{\text{lens bulge}}$	$0.5 \cdot \text{mean}$	$\text{mean} - 2\sigma$	$\text{mean} + 2\sigma$

This time the gaussian was truncated at 2σ instead of 3σ , because we need to redefine orientation and ellipticity, but their values were already found in the previous fit. However, a certain degree of freedom is left.

- The prior on the other lens' parameters is described in section (4.3).

The fitting procedure, which uses 8 parameters (6 effective, because two are fixed), gives the following maximum log-likelihood model parameters (6.10). Visual results are in figures(6.38, 6.39, 6.40, 6.41a, 6.41b).

The **logarithmic Bayesian evidence** for this model, computed at the conclusion of the fitting procedure, is 31263.958 .

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens mass slope [/]	2.9778	2.6245×10^{-1}	4.8978×10^{-2}
galaxies lens mass einstein radius ["]	7.8510×10^{-1}	5.1287×10^{-2}	3.5296×10^{-2}
galaxies lens shear elliptical comps 0 [/]	2.3031×10^{-1}	5.6093×10^{-2}	7.6532×10^{-2}
galaxies lens shear elliptical comps 1 [/]	5.7037×10^{-2}	4.6268×10^{-2}	5.1062×10^{-2}
galaxies lens mass elliptical comps 0 [/]	-1.0213×10^{-2}	1.0696×10^{-2}	1.0076×10^{-2}
galaxies lens mass elliptical comps 1 [/]	-9.9563×10^{-2}	5.9894×10^{-2}	7.8501×10^{-2}

Table 6.10: The outcome derived from the seventh fitting procedure.

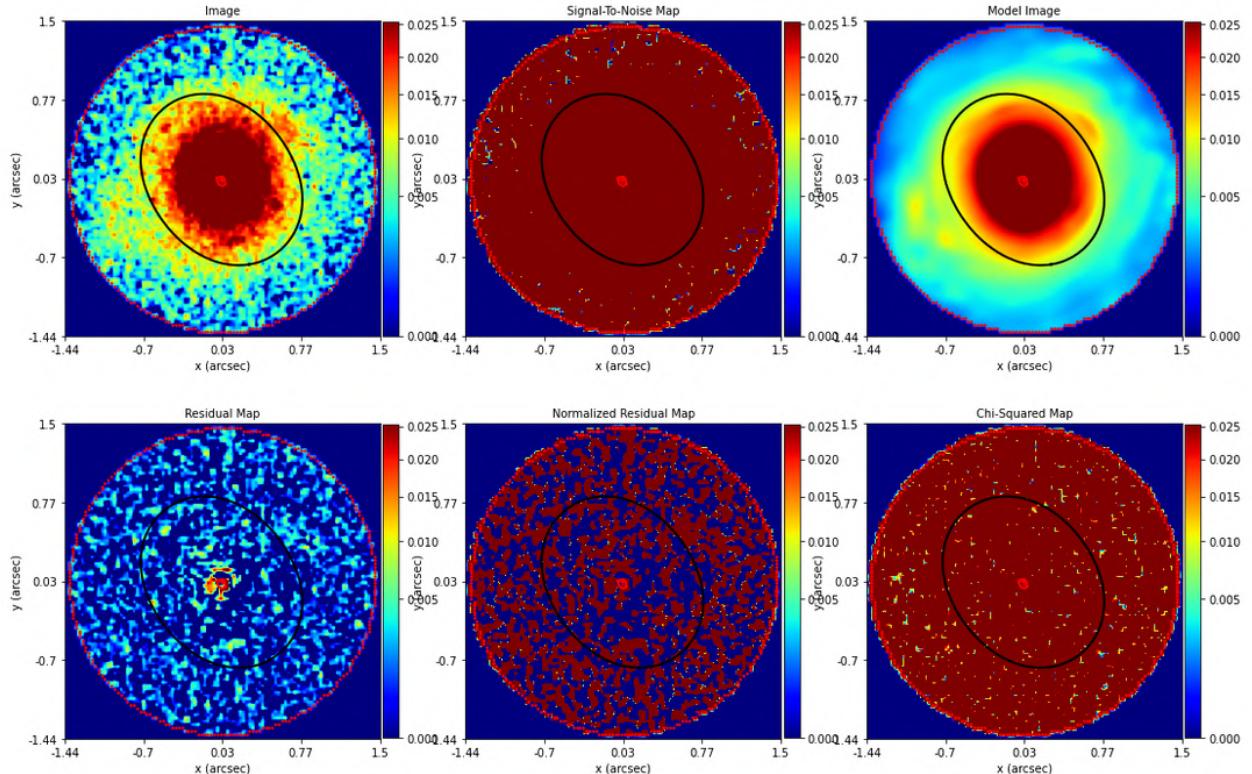


Figure 6.38: Outcome of the seventh fitting procedure.

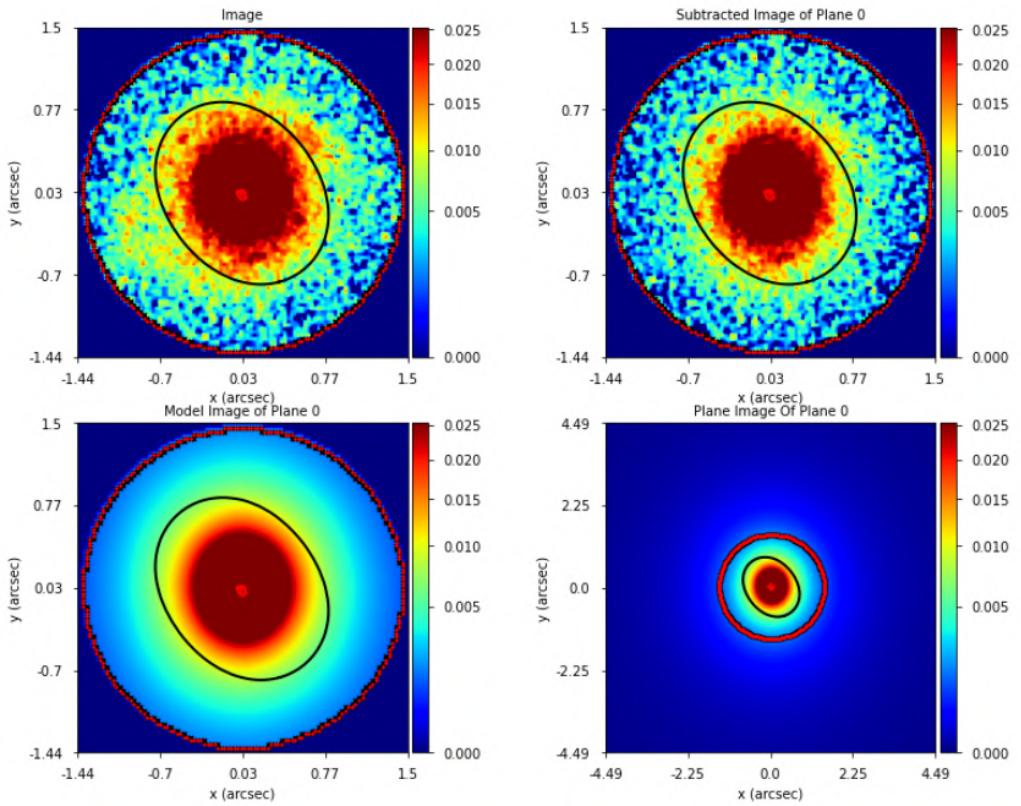


Figure 6.39: The lens plane resulting from the seventh fit. The subtracted image of plane 0 is the the lens light without the lensed light of the source.

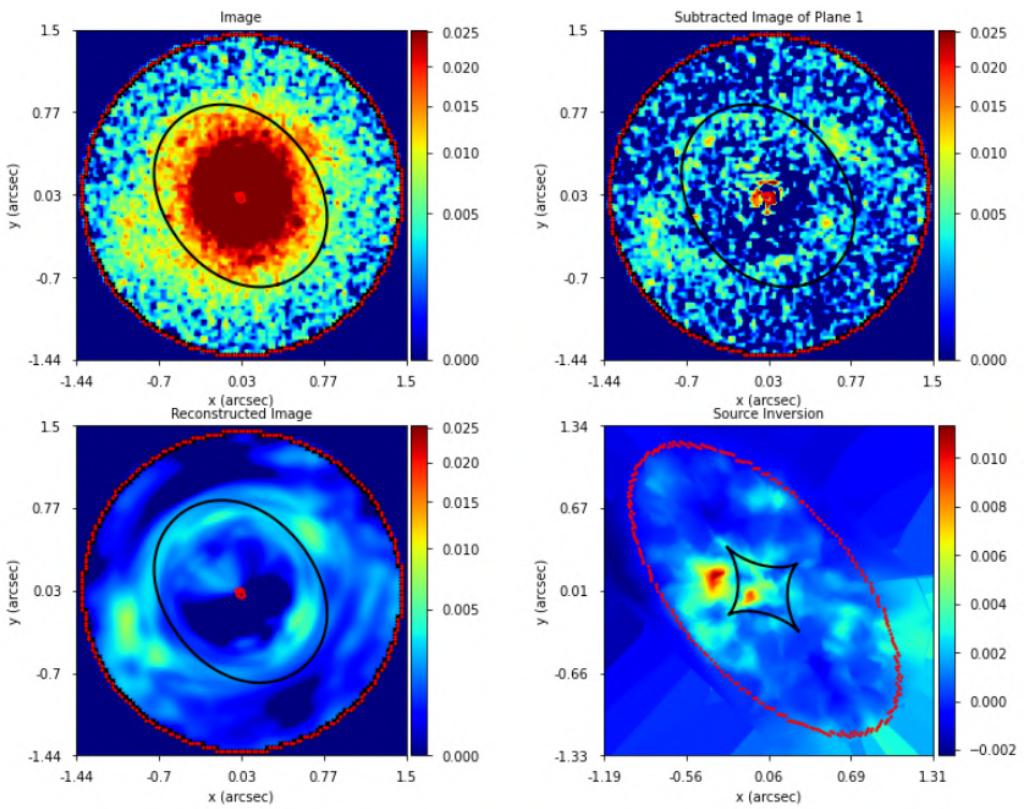


Figure 6.40: The source resulting from the seventh fit. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

In Figure 6.41a and Figure 6.41b, the *reconstructed image* and the *lensed image* of the source are presented in physical units.

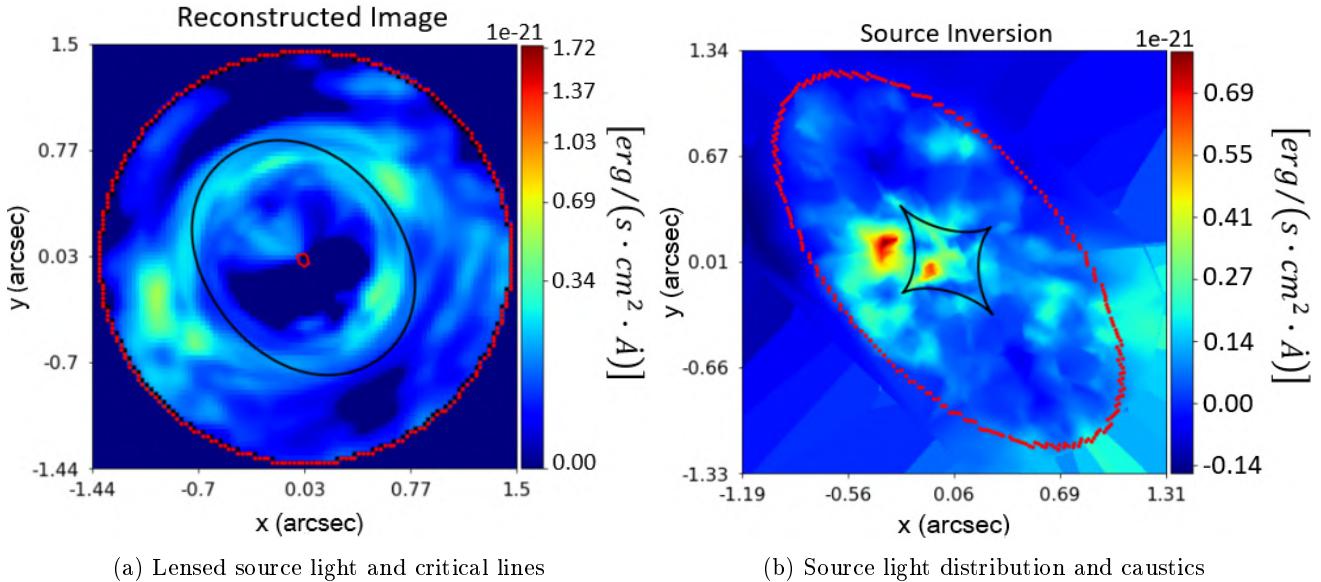


Figure 6.41: Light of the source in physical units.

6.7.2 Elliptical broken power-law

In this section the results of the fit that uses the elliptical broken power-law as model are reported.

Two different fits are performed using the elliptical broken power-law.

In both the fits the results of the fit number 7 that uses the elliptical power-law are used.

1° elliptical broken power-law model

The model is:

- The center's coordinates were fixed exactly as in the case of the power-law of fit 7.
- The elliptical components of the lens are fixed using the results of the fit 7 that uses the elliptical power-law. This is done because the fit 7 has achieved a high degree of accuracy, using adaptive grids on the source plane. Thus, it is assumed that the orientation and ellipticity of the lens have already been correctly fitted, regardless of the lens model.
- The scale radius b is fixed to the result of the fit 7 with elliptical power-law. The same scale radius appears in the definition of the broken power-law convergence and, as in the case of the elliptical components, we have that fit 7 has achieved a high degree of accuracy, using adaptive grids on the source plane. Thus, it is assumed that the orientation and ellipticity of the lens have already been correctly fitted, regardless of the lens model.
- On the *break radius* is imposed an uniform prior, where the lower limit is equal to 0.0 and the upper limit is set to be two times the scale radius b ($2 \cdot b$).
- On the *inner slope* t_1 is imposed an uniform prior, where the lower limit is 0.0 and the upper limit is set to be 3.0 .
- On the *outer slope* t_2 is imposed an uniform prior, where the lower limit is 0.0 and the upper limit is set to be 3.0 .

The fitting procedure gives the following maximum log-likelihood model parameters (6.11). Visual results are in figures(6.42, 6.43).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens mass break radius ["]	1.2006	4.8543×10^{-1}	3.2045×10^{-1}
galaxies lens mass inner slope [/]	1.8500	2.0144×10^{-1}	1.0176×10^{-1}
galaxies lens mass outer slope [/]	1.8422	1.3901	1.5533

Table 6.11: The outcome derived from the first *broken power-law* fitting procedure.

The *logarithmic Bayesian evidence* for this model, computed at the conclusion of the fitting procedure, is 31263.258 .

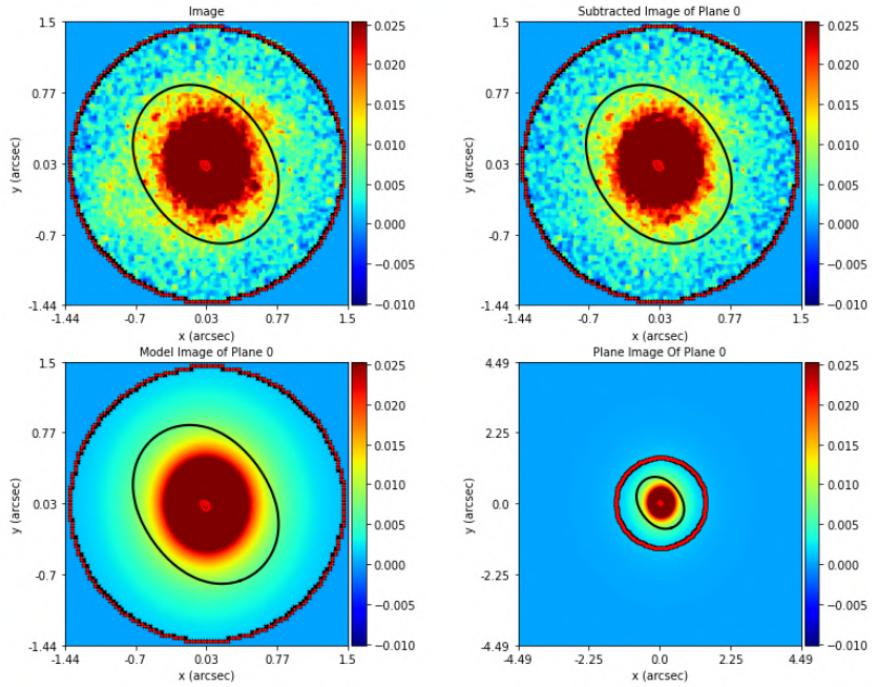


Figure 6.42: The lens plane resulting from the seventh fit with the elliptical broken power-law. The subtracted image of plane 0 is the the lens light without the lensed light of the source. The red piecewise line in the center is the radial critical line and it should be an infinitesimal point in the center, but the resolution is not high enough.

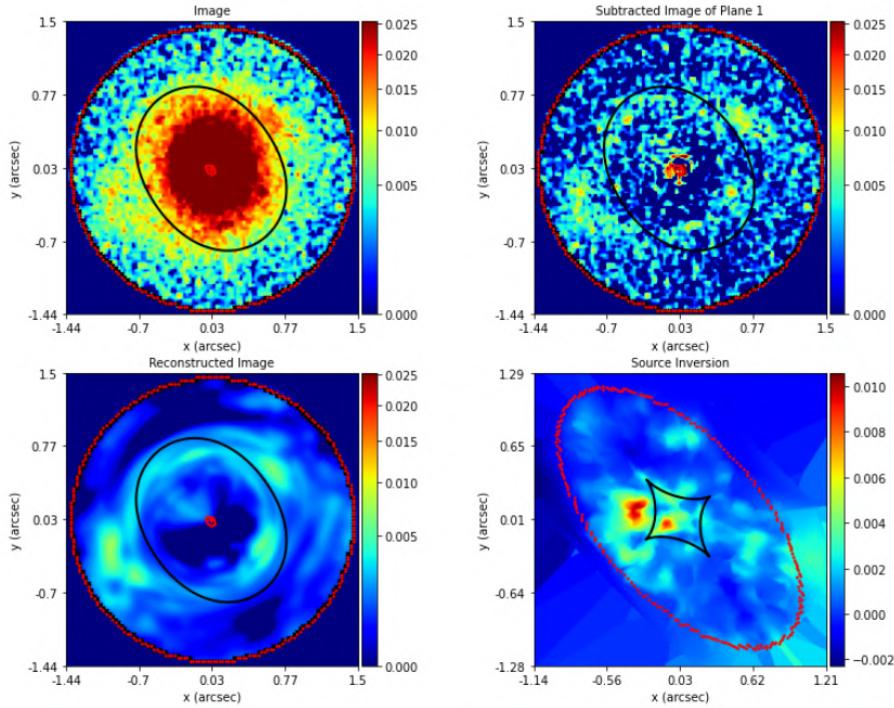


Figure 6.43: The source resulting from the seventh fit with the elliptical broken power-law. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

2° elliptical broken power-law model

The model is:

- The center's coordinates were fixed exactly as in the case of the power-law of fit 7.
- The elliptical components of the lens are fixed using the results of the fit 7 that uses the elliptical power-law. This is done because the fit 7 has achieved a high degree of accuracy, using adaptive grids on the source plane. Thus, it is assumed that the orientation and ellipticity of the lens have already been correctly fitted, regardless of the lens model.
- The scale radius b is fixed to the result of the fit 7 with elliptical power-law. The same scale radius appears in the definition of the broken power-law convergence and, as in the case of the elliptical components, we have that fit 7 has achieved a high degree of accuracy, using adaptive grids on the source plane. Thus, it is assumed that the orientation and ellipticity of the lens have already been correctly fitted, regardless of the lens model.
- The *inner slope* t_1 is fixed to a value of 1.00, representing the slope of a *singular isothermal ellipsoid (SIE)*. This constraint is applied to investigate the preservation of the SIE profile in elliptical galaxies up to a specified transition point. .
- On the *break radius* is imposed an uniform prior, where the lower limit is equal to 0.0 and the upper limit is set to be two times the scale radius b ($2 \cdot b$).
- On the *outer slope* t_2 is imposed an uniform prior, where the lower limit is 0.0 and the upper limit is set to be 3.0 .

The fitting procedure gives the following maximum log-likelihood model parameters (6.12). Visual results are in figures(6.44, 6.45).

Parameter	Value	Lower error (3σ)	Upper error (3σ)
galaxies lens mass break radius ["]	2.0239×10^{-1}	1.9056×10^{-1}	1.1465×10^{-1}
galaxies lens mass outer slope [/]	1.5859	7.2951×10^{-1}	3.6248×10^{-1}

Table 6.12: The outcome derived from the second *broken power-law* fitting procedure.

The *logarithmic Bayesian evidence* for this model, computed at the conclusion of the fitting procedure, is 31254.689 .

The results of this series of fit are in the *Elliptical_Broken_Power_Law_Analysis* folder.

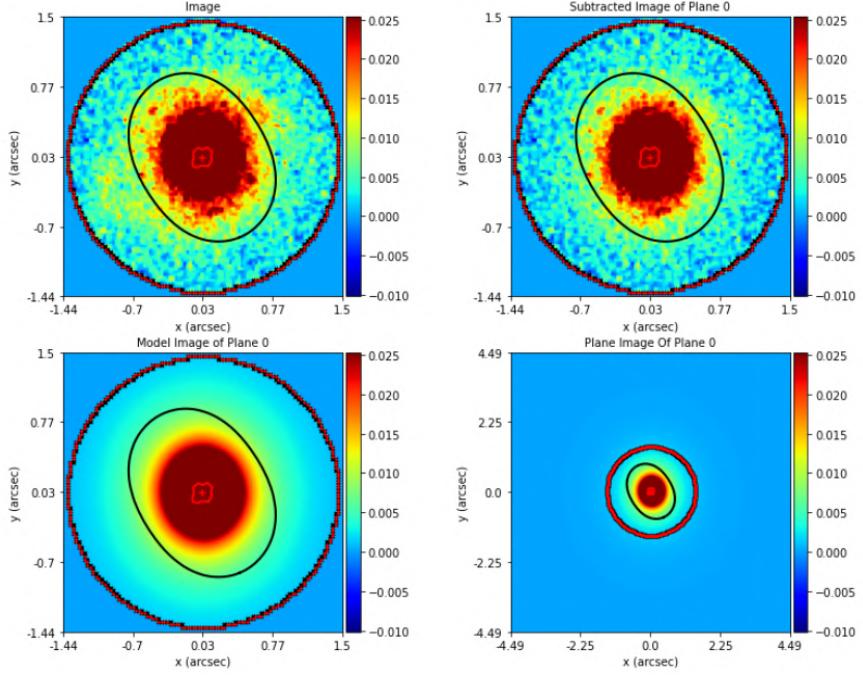


Figure 6.44: The lens plane resulting from the seventh fit with the elliptical broken power-law. The subtracted image of plane 0 is the the lens light without the lensed light of the source. The red piecewise line in the center is the radial critical line and it should be an infinitesimal point in the center, but the resolution is not high enough.

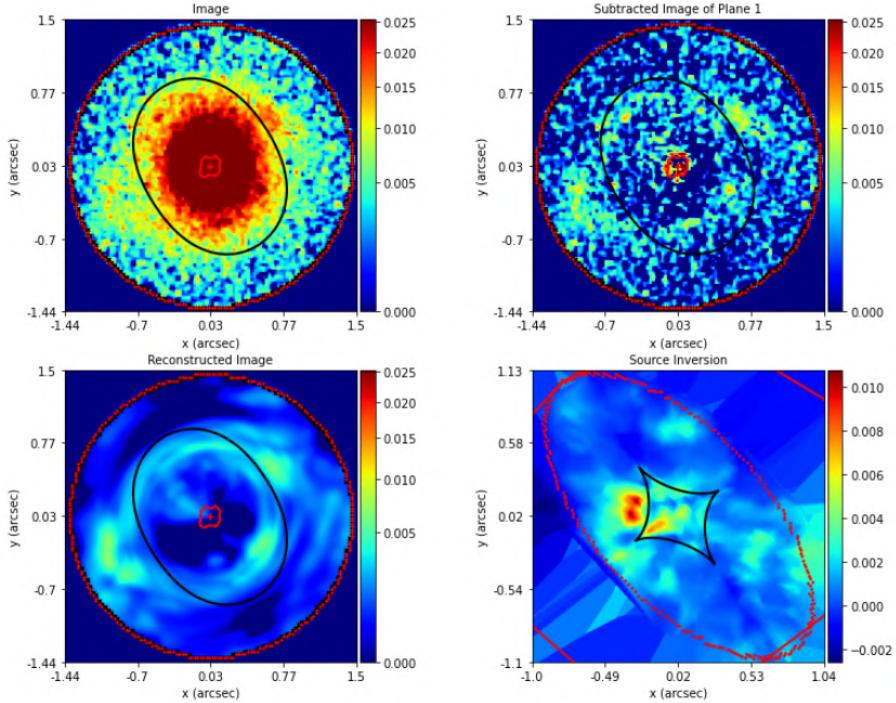


Figure 6.45: The source resulting from the seventh fit with the elliptical broken power-law. The subtracted image of plane 1 is the the lensed light of the source without the light of the lens.

Chapter 7

Results analysis

In this section, the outcomes of the previous fits are analyzed. The primary objective of this thesis, as pointed out in the introduction, is to determine the distribution of matter, both baryonic and dark, in the substructure we take under analysis.

First of all, we compare the results on the mass obtained in [Bergamini et al. 2019] with those obtained in fit number 7, the last fit, that is the most precise one.

7.1 Bergamini et al. 2019 mass model

The authors model the mass distribution using the *dual Pseudo-Isothermal Elliptical* surface mass density distribution (*dPIE*):

$$k_{dPIE}(\theta) = \frac{\sigma_0^2}{2G\Sigma_{crit}D_L} \frac{r_{cut}}{r_{cut} - r_{core}} \left(\frac{1}{\sqrt{\theta^2 + r_{core}^2}} - \frac{1}{\sqrt{\theta^2 + r_{cut}^2}} \right) \quad (7.1)$$

where:

- σ_0 is the normalization of the convergence, and has the units of the velocity [*length/time*]. It is not the velocity dispersion of this mass distribution.
- r_{core} is the *core-radius*. This allows to avoid the singularity for $\vec{\theta} \rightarrow 0$. Same units as θ .
- r_{cut} is the *truncation-radius*. $r_{core} < r_{cut}$. This allows to have a non-divergent mass value for the lens (the surface integral of the (7.1)). Same units as θ .
- θ is the elliptical radius of equation (2.14).

In the central region ($\theta \rightarrow 0$) and we have that the (7.1) becomes

$$k_{dPIE}(\theta = 0) = \frac{\sigma_0^2}{2G\Sigma_{crit}} \cdot \frac{1}{r_{cut} - r_{core}} \cdot \frac{1}{r_{core}} \quad (7.2)$$

In the region $r_{core} < \theta < r_{cut}$ the (7.1) behaves as the elliptical isothermal profile:

$$k_{dPIE}(r_{core} < \theta < r_{cut}) \propto \frac{1}{\theta} \quad (7.3)$$

Eventually, in the region $\theta \gg r_{cut}$ the (7.1) decreases as:

$$k_{dPIE}(r_{cut} \ll \theta) \propto \frac{1}{\theta^2} \quad (7.4)$$

The mass of this surface mass density distribution is given by equation (2.36):

$$\begin{aligned} M_{dPIE}(\theta) &= \frac{2\pi\Sigma_{crit}D_L^2}{q} \int_0^\theta k(\theta')\theta' d\theta' \\ &= \frac{\pi\sigma_0^2 D_L}{qG} \cdot \frac{r_{cut}}{r_{cut} - r_{core}} \cdot (\sqrt{r_{core}^2 + \theta^2} - r_{core} - \sqrt{r_{cut}^2 + \theta^2} + r_{cut}) \end{aligned} \quad (7.5)$$

The parameters value are those reported in the the *best.par* file at [Results \[Bergamini et al.\]](#). The lens under investigation is the *potential 9152*. The parameters are:

- $q = 0.00$
- $r_{core} = 0.002886$ arcseconds
- $r_{cut} = 0.763913$ arcseconds
- $\sigma_{0, lenstool} = 135.845530$ km/s
- $z_{lens} = 0.44$
- ΛCDM model: $\Omega_\Lambda = 0.7, \Omega_M = 0.3, \Omega_K = 0, H_0 = 70 \frac{km}{s \cdot Mpc}$

The theoretical σ_0 that appears in (7.1) is related to the one of lenstool by the relation:

$$\sigma_0 = \sqrt{\frac{3}{2}} \sigma_{0, lenstool} \quad (7.6)$$

In order to compare the mass of different surface mass density distribution it is necessary to choose the same radius within which calculate the mass. A radius $\theta = 1.5$ arcseconds is choosen.

The mass of the dPIE is $M_{dPIE}(1.5'') = 6.66 \cdot 10^{10} M_\odot$.

The *Einstein radius* of the dPIE of the Bergamini et al. 2019 is calculated using the more general relation for asymmetric, irregular and realistic lenses in the article [Etherington et al. 2022]:

$$Einstein\ radius = \sqrt{\frac{A}{\pi}} \quad (7.7)$$

where A is the area enclosed within the tangential critical line, that can be calculated thanks to the Green's theorem:

Theorem 2. *The circuitation of \vec{f} , that lies on the plane of the curve γ , is*

$$C(\vec{f}) = \int_{\gamma(t)} \langle \vec{f}, \frac{d\vec{x}}{dt} \rangle dt = \int_{\Omega} \langle \text{rot} \vec{f}, \vec{n} \rangle d^2\vec{x} = \int_{\Omega} \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) d^2\vec{x}$$

where $\gamma(t)$ is the parametric curve along which the circuitation is performed, Ω is the region bounded by this curve and \vec{n} is the normal to the infinitesimal surface element. Lying on the same plane of the curve, the $\text{rot} \vec{f}$ is parallel to \vec{n} and thus the last passage is justified.

Thus, in our case

$$\begin{aligned}
A &= \int_{\Omega} 1 \, dx \, dy \\
\vec{f}(x, y) &= (f_1(x, y), f_2(x, y)) \rightarrow \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) = 1 \\
&= \int_{\Omega} \langle \text{rot } \vec{f}, \vec{n} \rangle \, d^2 \vec{x} = \int_{\Omega} \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) d^2 \vec{x} \\
&= \int_{\gamma(t)} f_1 dx + f_2 dy \\
\vec{f}(x, y) &= (0, x) \\
&= \int_{\gamma(t)} x \, dy \\
&= \int_{\gamma(t)} x(t) \frac{dy(t)}{dt} \, dt
\end{aligned} \tag{7.8}$$

$\gamma(t)$ are the tangential critical lines.

In the case of the dPIE of [Bergamini et al. 2019] the Einstein radius, numerically evaluated, is

$$\theta_E = 0.36 \text{ arcseconds} \tag{7.9}$$

The mass within the critical lines is

$$M(\theta = \theta_E) = 3.19 \cdot 10^{10} M_{\odot} \tag{7.10}$$

7.2 Mass and Einstein radius of the power-law

In fit number 7 both the elliptical power-law and the elliptical broken power-law were used. The mass of the lens and the Einstein radius are calculated using the elliptical power-law mass model. The matter distribution properties are studied using the elliptical broken power-law.

The mass of the lens, using (2.36) for the elliptical power-law model is :

$$M_{epl}(\theta) = \frac{\pi \Sigma_c D_L^2}{q} \cdot b^t \cdot \theta^{2-t} \cdot \text{arcs_to_rad} \tag{7.11}$$

where the radius θ is given in arcseconds and the `arcs_to_rad` is the conversion factor between arcseconds and radians, because the θ in the formula must be in radians.

Though, as previously said, PyAutoLens solves the equation (1.64), thus we need to rescale by the factor $(1 - k_{ext})$ the different quantities.

The slope t and the axis-ratio q do not need to be rescaled. The first quantity that must be rescaled is the scale-radius b . In order to rescale it, we use the definition of the mass. Using equation (7.11) and considering the fact that it must be rescaled by the factor $(1 - k_{ext})$, we have that:

$$\begin{aligned}
\frac{M(\theta)}{(1 - k_{ext})} &= M'(\theta) = \frac{\pi \Sigma_c D_L^2}{q} \cdot b^t \cdot \theta^{2-t} \\
\frac{\pi \Sigma_c D_L^2}{q(1 - k_{ext})} \cdot b^t \cdot \theta^{2-t} &= \frac{\pi \Sigma_c D_L^2}{q} \cdot b'^t \cdot \theta^{2-t} \\
\frac{b^t}{(1 - k_{ext})} &= b'^t \\
\frac{b}{\sqrt[1.97]{(1 - k_{ext})}} &= b' \\
b &= \sqrt[1.97]{(1 - k_{ext})} \cdot b'
\end{aligned} \tag{7.12}$$

NOTE: the deflection implemented in PyAutolens is the one in the article [Tessore and R Benton Metcalf 2015], but the convention on the slope is changed. The SIE is 2 in PyAutolens for this mass model. We need to subtract 1 to the result of PyAutolens in order to have the same convention defined in the elliptical power-law convergence.

Thus, the scale radius b is equal to

$$b = \sqrt[1.97]{(1 - k_{ext})} \cdot b' = \sqrt[1.97]{0.46} \cdot 0.79 = 0.53 \text{ arcseconds} \tag{7.13}$$

The lens' orientation angle and axis-ratio are given by equation (6.2) using the results in table (6.10):

$$\begin{aligned}
\phi &= 0.5 \cdot \arctan \left(\frac{\epsilon_{lens,0}}{\epsilon_{lens,0}} \right) = 1.62 \text{ (with respect to } \theta_1 \text{ axis)} \\
fac &= \frac{\epsilon_{lens,0}}{\sin(2\phi)} = -0.10 \\
q &= \frac{1 + fac}{1 - fac} = 0.82
\end{aligned} \tag{7.14}$$

PyAutolens makes a profile elliptical by multiplying the axis-ratio q to θ_2 -axis, thus it gives the orientation angle with respect to this and even the elliptical components rely on this. Thus, in order to use the orientation angle ϕ , make sure to use it in radians and with respect to θ_2 -axis.

The mass of the elliptical power-law lens, using equation (7.11), is $M_{epl} = 8.58 \cdot 10^{10} M_\odot$.

The error on the parameters are computed (taking a mean between the upper and lower 3σ error on the elliptical components) using the standard error propagation formula. Given a function f of N real variables, we have that the error on this function is given by:

$$\sigma_f = \sqrt{\sum_i^N \left(\frac{\partial f}{\partial x_i} \cdot \sigma_i \right)^2} \tag{7.15}$$

The error on the previous parameters are:

Error on ϕ :

$$\begin{aligned}
\sigma_\phi &= \sqrt{\left(\frac{\partial \phi}{\partial \epsilon_{lens,0}} \cdot \sigma_{\epsilon_{lens,0}}\right)^2 + \left(\frac{\partial \phi}{\partial \epsilon_{lens,1}} \cdot \sigma_{\epsilon_{lens,1}}\right)^2} \\
\frac{\partial \phi}{\partial \epsilon_{lens,0}} &= 0.5 \cdot \frac{1}{1 + \left(\frac{\epsilon_{lens,0}}{\epsilon_{lens,1}}\right)^2} \cdot \frac{1}{\epsilon_{lens,1}} = -4.54 \\
\sigma_{\epsilon_{lens,0}} &= 0.01 \\
\left(\frac{\partial \phi}{\partial \epsilon_{lens,0}} \cdot \sigma_{\epsilon_{lens,0}}\right)^2 &= 0.002 \\
\frac{\partial \phi}{\partial \epsilon_{lens,1}} &= 0.5 \cdot \frac{1}{1 + \left(\frac{\epsilon_{lens,0}}{\epsilon_{lens,1}}\right)^2} \cdot \epsilon_{lens,0} \cdot (-\epsilon_{lens,1}^{-2}) = -0.45 \\
\sigma_{\epsilon_{lens,1}} &= 0.08 \\
\left(\frac{\partial \phi}{\partial \epsilon_{lens,1}} \cdot \sigma_{\epsilon_{lens,1}}\right)^2 &= 0.001 \\
\sigma_\phi &= \sqrt{0.002 + 0.001} = 0.054
\end{aligned} \tag{7.16}$$

Error on fac :

$$\begin{aligned}
\sigma_{fac} &= \sqrt{\left(\frac{\partial fac}{\partial \epsilon_{lens,0}} \cdot \sigma_{\epsilon_{lens,0}}\right)^2 + \left(\frac{\partial fac}{\partial \phi} \cdot \sigma_\phi\right)^2} \\
\frac{\partial fac}{\partial \epsilon_{lens,0}} &= \frac{1}{\sin(2\phi)} = 10.1 \\
\sigma_{\epsilon_{lens,0}} &= 0.01 \\
\left(\frac{\partial fac}{\partial \epsilon_{lens,0}} \cdot \sigma_{\epsilon_{lens,0}}\right)^2 &= 0.01 \\
\frac{\partial fac}{\partial \phi} &= -2\epsilon_{lens,0} \cdot \frac{\cos(2\phi)}{\sin(2\phi)} \cdot \frac{1}{\sin(2\phi)} = 1.996 \\
\sigma_\phi &= 0.054 \\
\left(\frac{\partial fac}{\partial \phi} \cdot \sigma_\phi\right)^2 &= 0.01 \\
\sigma_{fac} &= \sqrt{0.01 + 0.01} = 0.08
\end{aligned} \tag{7.17}$$

Error on q :

$$\begin{aligned}
\sigma_q &= \frac{\partial q}{\partial fac} \cdot \sigma_{fac} \\
\frac{\partial q}{\partial fac} &= \frac{-2}{(1+fac)^2} = 2.46 \\
\sigma_q &= 2.46 \cdot 0.14 = 0.34
\end{aligned} \tag{7.18}$$

The error on the mass is computed using the equations (7.11) and (7.15).

$$\sigma_M = \sqrt{\left(\frac{\partial M_{epl}}{\partial q} \cdot \sigma_q\right)^2 + \left(\frac{\partial M_{epl}}{\partial b} \cdot \sigma_q\right)^2 + \left(\frac{\partial M_{epl}}{\partial t} \cdot \sigma_t\right)^2} \tag{7.19}$$

The first term, calculated at $\theta = 1.5''$, is:

$$\begin{aligned} \frac{\partial M_{epl}}{\partial q} &= \pi \Sigma_c D_L^2 \cdot b^t \cdot \theta^{2-t} \cdot (-1) \cdot q^{-2} \cdot \text{arcs_to_rad} = -5.63 \cdot 10^{10} M_\odot \\ \left(\frac{\partial M_{epl}}{\partial q} \cdot \sigma_q \right)^2 &= 3.67 \cdot 10^{20} M_\odot \end{aligned} \quad (7.20)$$

The error on b , using equation (7.13) and the mean error in table (6.10), is just

$$\sigma_b = \sqrt[1.97]{1 - k_{ext}} \cdot \sigma_{b'} = 0.67 \cdot 0.04 = 0.027 \quad (7.21)$$

The second term, calculated at $\theta = 1.5''$, is:

$$\begin{aligned} \frac{\partial M_{epl}}{\partial b} &= \frac{\pi \Sigma_c D_L^2}{q} \cdot \theta^{2-t} \cdot t \cdot b^{t-1} \cdot \text{arcs_to_rad} = 3.20 \cdot 10^{11} M_\odot \\ \left(\frac{\partial M_{epl}}{\partial b} \cdot \sigma_b \right)^2 &= 7.40 \cdot 10^{19} M_\odot \end{aligned} \quad (7.22)$$

The third term, calculated at $\theta = 1.5''$ and using $\sigma_t = 0.15$, is:

$$\begin{aligned} \frac{\partial M_{epl}}{\partial t} &= \frac{\pi \Sigma_c D_L^2}{q} \cdot \theta^2 \cdot \frac{\partial}{\partial t} \left(e^{\ln\left(\frac{b}{\theta}\right)^t} \right) \cdot \text{arcs_to_rad} = \\ &= \frac{\pi \Sigma_c D_L^2}{q} \cdot \theta^2 \cdot e^{t \cdot \ln\left(\frac{b}{\theta}\right)} \cdot \ln\left(\frac{b}{\theta}\right) = -8,93 \cdot 10^{10} M_\odot \\ \left(\frac{\partial M_{epl}}{\partial t} \cdot \sigma_t \right)^2 &= 1.80 \cdot 10^{20} M_\odot \end{aligned} \quad (7.23)$$

Thus, the error on the mass within $1.5''$ is:

$$\begin{aligned} \sigma_M &= \sqrt{\left(\frac{\partial M_{epl}}{\partial q} \cdot \sigma_q \right)^2 + \left(\frac{\partial M_{epl}}{\partial b} \cdot \sigma_b \right)^2 + \left(\frac{\partial M_{epl}}{\partial t} \cdot \sigma_t \right)^2} \\ &= \sqrt{3.67 \cdot 10^{20} + 7.40 \cdot 10^{19} + 1.80 \cdot 10^{20}} \\ &= 2.5 \cdot 10^{10} M_\odot \end{aligned} \quad (7.24)$$

The Einstein radius for this model is the one defined using equation (4) in [O'Riordan, S J Warren, and D J Mortlock 2020]:

$$\theta_E = \frac{b \pm \sigma_b}{\sqrt{q} \pm \sigma_q} = \frac{b}{\sqrt{q}} \cdot \sqrt{b^2 \sigma_b^2 + q \sigma_q^2} = \frac{0.53}{\sqrt{0.81}} \pm 0.3 = 0.59 \pm 0.3 \text{ arcseconds} \quad (7.25)$$

The mass of the elliptical power-law lens within the tangential critical lines, thus the Einstein radius, using equation (7.11) and (7.19), is $M_{epl} = 8.35 \cdot 10^{10} \pm 2.05 \cdot 10^{10} M_\odot$.

7.3 Shear results and uncertainties

The results concerning the shear (even the rescaled one), using equations (6.5), (7.16) and (7.17), are:
Error on ϕ_{shear} :

$$\begin{aligned}
\sigma_{\phi_s} &= \sqrt{\left(\frac{\partial \phi_s}{\partial \epsilon_{s,0}} \cdot \sigma_{\epsilon_{s,0}}\right)^2 + \left(\frac{\partial \phi_s}{\partial \epsilon_{s,1}} \cdot \sigma_{\epsilon_{s,1}}\right)^2} \\
\frac{\partial \phi_s}{\partial \epsilon_{s,0}} &= 0.5 \cdot \frac{1}{1 + \left(\frac{\epsilon_{s,0}}{\epsilon_{s,1}}\right)^2} \cdot \frac{1}{\epsilon_{s,1}} = 0.036 \\
\sigma_{\epsilon_{s,0}} &= 0.07 \\
\left(\frac{\partial \phi_s}{\partial \epsilon_{s,0}} \cdot \sigma_{\epsilon_{s,0}}\right)^2 &= 0.002 \\
\frac{\partial \phi_s}{\partial \epsilon_{s,1}} &= 0.5 \cdot \frac{1}{1 + \left(\frac{\epsilon_{s,0}}{\epsilon_{s,1}}\right)^2} \cdot \epsilon_{s,0} \cdot (-\epsilon_{s,1}^{-2}) = -0.020 \\
\sigma_{\epsilon_{s,1}} &= 0.01 \\
\left(\frac{\partial \phi_s}{\partial \epsilon_{s,1}} \cdot \sigma_{\epsilon_{s,1}}\right)^2 &= 0.0004 \\
\sigma_{\phi_s} &= \sqrt{0.002 + 0.0004} = 0.04
\end{aligned} \tag{7.26}$$

Error on γ :

$$\begin{aligned}
\sigma_\gamma &= \sqrt{\left(\frac{\partial \gamma}{\partial \epsilon_{s,0}} \cdot \sigma_{\epsilon_{s,0}}\right)^2 + \left(\frac{\partial \gamma}{\partial \phi_s} \cdot \sigma_{\phi_s}\right)^2} \\
\frac{\partial \gamma}{\partial \epsilon_{s,0}} &= \frac{1}{\sin(2\phi_s)} = 1.05 \\
\sigma_{\epsilon_{s,0}} &= 0.07 \\
\left(\frac{\partial \gamma}{\partial \epsilon_{s,0}} \cdot \sigma_{\epsilon_{s,0}}\right)^2 &= 0.005 \\
\frac{\partial \gamma}{\partial \phi_s} &= -2\epsilon_{s,0} \cdot \frac{\cos(2\phi_s)}{\sin(2\phi_s)} \cdot \frac{1}{\sin(2\phi_s)} = -0.16 \\
\sigma_{\phi_s} &= 0.04 \\
\left(\frac{\partial \gamma}{\partial \phi_s} \cdot \sigma_{\phi_s}\right)^2 &= 0.003 \\
\sigma_\gamma &= \sqrt{0.01 + 0.01} = 0.09
\end{aligned} \tag{7.27}$$

Thus, the value of γ and ϕ_s are:

$$\begin{aligned}
\gamma &= 0.237 \pm 0.090 \xrightarrow{\cdot(1-k_{ext})} 0.109 \pm 0.041 \\
\phi &= 0.66 \pm 0.04 \text{ rad (with respect to } \theta_1 \text{ axis)} \\
\phi &= 38.04^\circ \text{ pm } 2.3 \text{ deg (with respect to } \theta_1 \text{ axis)}
\end{aligned} \tag{7.28}$$

In the [Bergamini et al. 2019] model are:

$$\begin{aligned}\gamma &= 0.202 \xrightarrow{\cdot(1-k_{ext})} 0.092 \\ \phi &= 0.63 \text{ rad (with respect to } \theta_1 \text{ axis)} \\ \phi &= 36.25^\circ \text{ deg (with respect to } \theta_1 \text{ axis)}\end{aligned}\quad (7.29)$$

7.4 Comparison between Epl and dPIE models

The table (7.1) encompasses the most important results obtained during the analysis.

Model	Epl (this work, PyAutolens) $\pm 3\sigma$	[Bergamini et al. 2019] (Lenstool) $\pm 3\sigma$
θ_E [arcseconds]	0.59 ± 0.03	0.36 ± 0.07
θ_E [kpc]	3.46 ± 0.18	2.17 ± 0.41
$M(1.5'')$ [M_\odot]	$8.58 \cdot 10^{10} \pm 2.5 \cdot 10^{10}$	$6.66 \cdot 10^{10} \pm 2.08 \cdot 10^{10}$
M_{θ_E} [M_\odot]	$8.35 \cdot 10^{10} \pm 2.05 \cdot 10^{10}$	$3.19 \cdot 10^{10} \pm 1.00 \cdot 10^{10}$
Shear amplitude γ [/]	0.109 ± 0.041	0.092
Shear ϕ angle[/]	0.66 ± 0.04	0.63

Table 7.1: The outcome derived from the analysis procedure.

The discrepancy can be attributed to the difference in resolution between the two analyses. Bergamini's analysis, which does not use the Kpc scale Einstein ring and relies on average relations between parameters, loses information at the Kpc scale, likely accounting for the observed difference. Another discrepancy between the models that may have led to these results is the different cosmology employed. In this study, ΛCDM model with Planck 2015 cosmology $H_0 = 67.7 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$, $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$ is used, while in [Bergamini et al. 2019] model, the ΛCDM model utilized is: $\Omega_\Lambda = 0.7$, $\Omega_M = 0.3$, $\Omega_K = 0$, $H_0 = 70 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$. Conversely, the parameters used to describe the external deflection, the amplitude γ and the orientation angle ϕ , are in agreement within their respective error margins (3σ): 0.109 ± 0.041 and 0.66 ± 0.04 found in this thesis, compared to 0.092 and 0.63 in [Bergamini et al. 2019]. This suggests that the Mpc scale properties, as presented in [Bergamini et al. 2019], effectively describe external perturbations originating from the entire cluster, impacting the Kpc scale.

7.5 The slope of the lens: power-law or broken power-law model

In this section we compare the results of the last fit with what is reported in section (2.1). To compare the slope we obtain from the fit number 7 to the one of the 3D SIE, we need to convert the 2D slope to the 3D one. The relation between the 2D slope t , for both the elliptical power-law and the elliptical broken power-law, is:

$$\gamma_{3D} = t_{2D} + 1 \quad (7.30)$$

In this way it is possible to relate $\rho \propto |\vec{r}|^{-\gamma_{3D}}$.

7.5.1 Elliptical power-law slope

What distinguishes this research is the observed trend in mass distribution. This detailed examination reveals a noteworthy discovery: the total mass distribution of the observed early-type galaxy exhibits a slope significantly steeper than the isothermal profile typically expected for such galaxies - approximately one factor greater.

The relation between PyAutolens slope and the one in [Tessore and R Benton Metcalf 2015] is:

$$t = t_{PyA} - 1 \quad (7.31)$$

thus, the 3D slope resulting from our analysis, using the result of the elliptical power-law model, is:

$$\gamma_{3D} = t + 1 = t_{PyA} = 2.978 \pm 0.156 \quad (7.32)$$

where the error is the mean of the upper and lower 3σ error in table (6.10).

It is steeper than a SIE, that typically describes ETGs as pointed out in section (2.1) and in figure (7.1).

7.5.2 Do the data provide support for the introduction of a broken radius?

The slope convention in PyAutolens is the same as in [O'Riordan, Stephen J Warren, and Daniel J Mortlock 2021]. Here we have that the SIE is given by $t_1 = t_2 = 1$.

Is the introduction of another parameter justified? In both the test the answer is no. The bayesian evidence does not support it in both the two different fits with the broken power-law. The justifications are outlined in Table 7.2, where *Model A* represents the Elliptical broken power-law with both t_1 and t_2 as free parameters, and *Model B* corresponds to the Elliptical broken power-law with the inner slope t_1 set to be isothermal and the outermost slope t_2 as a free parameter.

Model	Elliptical power-law $\pm 3\sigma$	Model A $\pm 3\sigma$	Model B $\pm 3\sigma$
break radius ["]	/	1.2 ± 0.4	0.2 ± 0.2
inner slope	2.97 ± 0.15	2.85 ± 0.15	2
outer slope	2.97 ± 0.15	2.84 ± 0.15	2.58 ± 0.56
log Bayesian evidence	31263.958	31263.258	31254.689

Table 7.2: The Elliptical broken power-law radius and respective values of the bayesian log-evidence.

According to section (3.1.3), the introduction of the new parameter must be justified using the Bayes's factor, that is the ratio of the evidence of two different models. The ratio is approximately 1 for the elliptical power-law model and for the elliptical broken power-law (6.7.2) and lower than 1 for the elliptical power-law and the elliptical broken power-law (6.7.2). Therefore, the data do not support the introduction of the break radius.

And not only that, as can be seen from the results, the slope is steeper than an SIE, already from the innermost regions. In fact, even imposing the inner slope equal to the SIE ($t_1 = 1$), the break radius is very small and the external slope is much steeper than a SIE, as reported in figure (7.1).

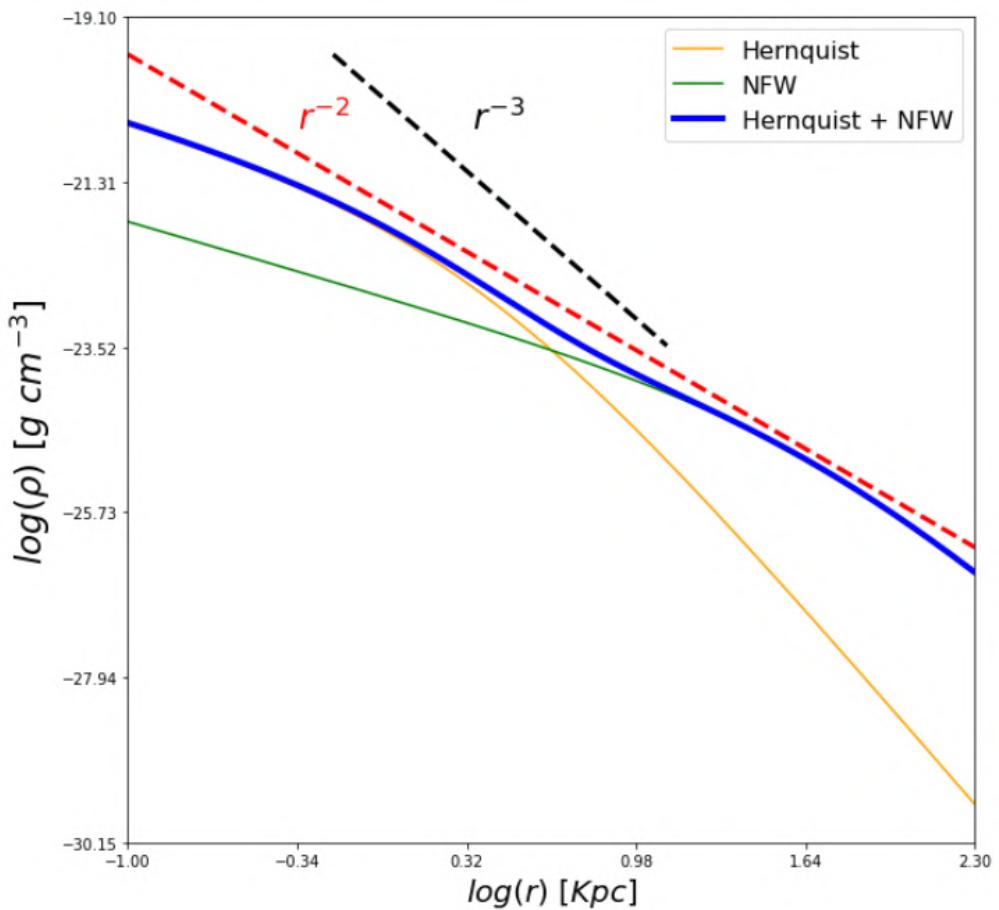


Figure 7.1: Density profile of an ETG galaxy as the one in section (2.1). The reference density of the Hernquist profile is $14.81e-23 [g \cdot cm^{-3}]$, the reference density of the NFW is $2.93e-25 [g \cdot cm^{-3}]$. The scale radius for the Hernquist profile is 2.24, the one for the NFW is 49.7.

7.6 Conclusions and future perspectives

The mass profile is steeper than expected (section (2.1)) for an ETG galaxy. This finding aligns with the scenario proposed by [(Dobke, King, and Fellhauer 2007), (Auger et al. 2008) and (Auger 2008)], suggesting that early-type satellite gravitational lenses within groups or clusters of galaxies are better modeled with mass density profiles steeper than isothermal. However, it is essential to note that the slope identified in this research is noticeably steeper than those reported in their works, particularly those of [(Auger et al. 2008) and (Auger 2008)]. Furthermore, it is worth highlighting that in their 2007 work, [Dobke, King, and Fellhauer 2007] conducted simulations within the mass range among group members spanning masses from $10^{11} M_{\odot}$ to $10^{13} M_{\odot}$. In contrast, the mass range identified in our research is on the order of $10^{10} M_{\odot}$. Eventually, [Dobke, King, and Fellhauer 2007] observed isothermal profiles ($\gamma_{3D} = 2$) steepening to $\gamma_{3D} = 3$, but specifically for a higher mass range. In our study, we observed a similar steepening, but for a lower mass range $10^{10} M_{\odot}$.

While it is fundamental to conduct additional tests before making an unequivocal claim about the nature of the physical system based on the results, it is undeniably a noteworthy outcome.

A plausible source of discrepancies, for instance, could be identified in the deflection field of the cluster:

- The results found in [Bergamini et al. 2019] may not be the most correct to accurately describe the region of the cluster where the galaxy under examination is located, and therefore, the values that were used to generate the first-order deflection may not be the most accurate ones.
- The effects of the cluster are used only up to the first order, and the approximation may not be enough to describe the external deflection field. Therefore, using orders higher than the first to describe the deflection field of the cluster at the coordinates of the lens may change the value of the galaxy's slope. Unfortunately, up to now, PyAutoLens does not incorporate higher order terms than the external shear to describe the external deflection field.

An additional test involves using the image acquired in another filter with an equivalent Signal-to-Noise ratio. Given that gravitational lensing is not wavelength-dependent, it is expected that the results will remain consistent across different filters.

In conclusion, conducting numerical simulations emerges as a preferred method for a comprehensive assessment, generating a simulated system with comparable quality to the original data. This involves maintaining the same signal-to-noise ratio while introducing varying slope scenarios and fitting them with the Elliptical Power Law. Using PyAutoLens on simulated results would provide an opportunity to validate the software's capability to recover a slope similar to that obtained from the analysis of real data. Specifically, simulating images with equivalent Signal-to-Noise ratios facilitates an examination of whether the software can genuinely recover the correct parameters within the errors or if such notable values are merely a result of random fluctuation in the data.

Bibliography

- Auger, M. W. (Jan. 2008). "The environments of SLACS gravitational lenses". In: *Monthly Notices of the Royal Astronomical Society: Letters* 383.1, pp. L40–L44. ISSN: 1745-3925. DOI: [10.1111/j.1745-3933.2007.00408.x](https://doi.org/10.1111/j.1745-3933.2007.00408.x). eprint: https://academic.oup.com/mnrasl/article-pdf/383/1/L40/54685588/mnrasl_383_1_140.pdf. URL: <https://doi.org/10.1111/j.1745-3933.2007.00408.x>.
- Auger, M. W. et al. (Feb. 2008). "Lens Galaxy Properties of SBS 1520+530: Insights from Keck Spectroscopy and AO Imaging". In: *The Astrophysical Journal* 673.2, p. 778. DOI: [10.1086/524351](https://doi.org/10.1086/524351). URL: <https://doi.org/10.1086/524351>.
- Bergamini, P. et al. (Nov. 2019). "Enhanced cluster lensing models with measured galaxy kinematics". In: *Astronomy & Astrophysics* 631, A130. DOI: [10.1051/0004-6361/201935974](https://doi.org/10.1051/0004-6361/201935974). URL: <https://doi.org/10.1051%2F0004-6361%2F201935974>.
- Buchner, Johannes (Jan. 2021). *Nested Sampling Methods*.
- Cimatti, Andrea, Filippo Fraternali, and Carlo Nipoti (2019). *Introduction to Galaxy Formation and Evolution: From Primordial Gas to Present-Day Galaxies*. Cambridge University Press.
- Dobke, Benjamin M., Lindsay J. King, and Michael Fellhauer (June 2007). "Steepened inner density profiles of group galaxies via interactions: an N-body analysis". In: *Monthly Notice of the Royal Astronomical Society* 377.4, pp. 1503–1510. DOI: [10.1111/j.1365-2966.2007.11683.x](https://doi.org/10.1111/j.1365-2966.2007.11683.x). arXiv: [astro-ph/0702741 \[astro-ph\]](https://arxiv.org/abs/astro-ph/0702741).
- Dynamic Nested Sampling Background* (2022). URL: <https://dynesty.readthedocs.io/en/latest/overview.html>.
- Etherington, Amy et al. (Sept. 2022). "Automated galaxy–galaxy strong lens modelling: No lens left behind". In: *Monthly Notices of the Royal Astronomical Society* 517.3, pp. 3275–3302. DOI: [10.1093/mnras/stac2639](https://doi.org/10.1093/mnras/stac2639). URL: <https://doi.org/10.1093%2Fmnras%2Fstac2639>.
- Feroz, Farhan et al. (Nov. 2019). "Importance Nested Sampling and the MultiNest Algorithm". In: *The Open Journal of Astrophysics* 2.1, 10, p. 10. DOI: [10.21105/astro.1306.2144](https://doi.org/10.21105/astro.1306.2144). arXiv: [1306.2144 \[astro-ph.IM\]](https://arxiv.org/abs/1306.2144).
- Generalized Hypergeometric Function* (n.d.). URL: <https://mathworld.wolfram.com/%20HypergeometricFunction.html>.
- Higson, Edward et al. (Sept. 2019). "Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation". In: *Statistics and Computing* 29.5, pp. 891–913. DOI: [10.1007/s11222-018-9844-0](https://doi.org/10.1007/s11222-018-9844-0). arXiv: [1704.03459 \[stat.CO\]](https://arxiv.org/abs/1704.03459).
- Jing, Liang (2019). *Nested Sampling: Introduction and Implementation*.
- Jullo, E. and J. -P. Kneib (May 2009). "Multiscale cluster lens mass mapping - I. Strong lensing modelling". In: *Monthly Notice of the Royal Astronomical Society* 395.3, pp. 1319–1332. DOI: [10.1111/j.1365-2966.2009.14654.x](https://doi.org/10.1111/j.1365-2966.2009.14654.x). arXiv: [0901.3792 \[astro-ph.CO\]](https://arxiv.org/abs/0901.3792).
- Jullo, E., J. -P. Kneib, et al. (Dec. 2007). "A Bayesian approach to strong lensing modelling of galaxy clusters". In: *New Journal of Physics* 9.12, p. 447. DOI: [10.1088/1367-2630/9/12/447](https://doi.org/10.1088/1367-2630/9/12/447). arXiv: [0706.0048 \[astro-ph\]](https://arxiv.org/abs/0706.0048).
- Kneib, J. -P. et al. (Nov. 1996). "Hubble Space Telescope Observations of the Lensing Cluster Abell 2218". In: *Astrophysical Journal* 471, p. 643. DOI: [10.1086/177995](https://doi.org/10.1086/177995). arXiv: [astro-ph/9511015 \[astro-ph\]](https://arxiv.org/abs/astro-ph/9511015).
- Meneghetti, Massimo (2021a). *Introduction to Gravitational Lensing: With Python Examples*. Vol. 956. Springer Nature.
- (2021b). *Lensing Lectures*.
- Metcalf, Robert Benton (2022). *Notes: Practical Statistics for Physics and Astronomy*. URL: <https://rbmetcalf.github.io/Practical-Statistics/lecturenotes/%20notes.pdf>.
- Nightingale, J W, S Dye, and Richard J Massey (May 2018). "AutoLens: automated modeling of a strong lens's light, mass, and source". In: *Monthly Notices of the Royal Astronomical Society* 478.4, pp. 4738–4784. DOI: [10.1093/mnras/sty1264](https://doi.org/10.1093/mnras/sty1264). URL: <https://doi.org/10.1093%2Fmnras%2Fsty1264>.

- Nightingale, James. et al. (Feb. 2021). “PyAutoLens: Open-Source Strong Gravitational Lensing”. In: *Journal of Open Source Software* 6.58, p. 2825. DOI: [10.21105/joss.02825](https://doi.org/10.21105/joss.02825). URL: <https://doi.org/10.21105/2Fjoss.02825>.
- O’Riordan, C M, S J Warren, and D J Mortlock (June 2020). “Galaxy mass profiles from strong lensing II: The elliptical power-law model”. In: *Monthly Notices of the Royal Astronomical Society* 496.3, pp. 3424–3435. DOI: [10.1093/mnras/staa1697](https://doi.org/10.1093/mnras/staa1697). URL: <https://doi.org/10.1093%2Fmnras%2Fstaa1697>.
- O’Riordan, Conor M, Stephen J Warren, and Daniel J Mortlock (2021). “Galaxy mass profiles from strong lensing—III. The two-dimensional broken power-law model”. In: *Monthly Notices of the Royal Astronomical Society* 501.3, pp. 3687–3694.
- Postman, Marc et al. (Apr. 2012). “The Cluster Lensing and Supernova Survey with Hubble: An Overview”. In: *Astrophysical Journal* 199.2, 25, p. 25. DOI: [10.1088/0067-0049/199/2/25](https://doi.org/10.1088/0067-0049/199/2/25). arXiv: [1106.3328 \[astro-ph.CO\]](https://arxiv.org/abs/1106.3328).
- PyAutolens Tutorials* (2022). URL: https://github.com/Jammy2211/autolens_workspace.
- Rosati, P. et al. (Dec. 2014). “CLASH-VLT: A VIMOS Large Programme to Map the Dark Matter Mass Distribution in Galaxy Clusters and Probe Distant Lensed Galaxies”. In: *The Messenger* 158, pp. 48–53.
- Skilling, John (Nov. 2004). “Nested Sampling”. In: *AIP Conf. Proc.* 735. DOI: [10.1063/1.1835238](https://doi.org/10.1063/1.1835238).
- Tessore, Nicolas (Dicembre 2015). “Gravitational Lensing as a Tool on Galactic and Cosmological Scales”. PhD thesis. alma. URL: <http://amsdottorato.unibo.it/7238/>.
- Tessore, Nicolas and R Benton Metcalf (2015). “The elliptical power law profile lens”. In: *Astronomy & Astrophysics* 580, A79.
- Van De Ven, Glenn, Rachel Mandelbaum, and Charles R. Keeton (Sept. 2009). “Galaxy density profiles and shapes – I. Simulation pipeline for lensing by realistic galaxy models”. In: *Monthly Notices of the Royal Astronomical Society* 398.2, pp. 607–634. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2009.15167.x](https://doi.org/10.1111/j.1365-2966.2009.15167.x). eprint: <https://academic.oup.com/mnras/article-pdf/398/2/607/4009862/mnras0398-0607.pdf>. URL: <https://doi.org/10.1111/j.1365-2966.2009.15167.x>.