Sparse Linear Inversion for Strong Gravitational Lenses Reconstruction, Component Separation from Morphological Component Analysis with Multiple Discrimination Criteria

PAR

Rémy Elie JOSEPH

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

C'est Francis Ford Coppola qui disait:
"je mets toujours une recette de cuisine dans mes films, comme ça même si on n'aime pas
mon film, au on a moins appris quelque chose".

It's Francis Ford Coppola who used to say:
"I always add a cooking recipe in my movies, so that if people don't like the movie, at least they
learned something."

— Alexandre Astier about science in a comedy show.

To my parents.

Acknowledgements

While I had been warned that taking on a thesis work would not be an easy feat, I have managed to enjoy this time of learning and hard work, which I could not have done without the help of a small number of people. The first group I would like to acknowledge is the double system formed by my two thesis directors, Frederic Courbin and Jean-Luc Starck. While you two have very different expertises and personalities, you share a similar vision of objective and progressive science. You seem to have built a trusting and lasting work and personal relationship that makes it a pleasure and true learning experience to work with you both. While each of you individually are great scientists, teachers and experts in your respective areas, combining your respective approaches has been a pleasure and a privilege.

The next mentor that I would like to thank is certainly Sandrine Pires. Being a taught engineer, starting a career in astrophysics was a long shot. Sandrine was brave enough to take me in as an intern and to guide me in my first painful steps in cosmology and coding. She taught me a lot and gave me the extraordinary opportunity to be part of the lensing community which I have learned to appreciate so much.

On the side of learning, many postdocs have had an influence on me and taught me a lot in terms of science but also personally. First I would like to thank my good friends Pierre Fleury and Franois Lanusse. I put them together, because to me these two have great qualities as scientists and humans beings. They have taught me a lot in their respective specialities, cosmology and signal processing and have grown to become friends that made these years much more enjoyable. Several postdocs: Anaïs Rassat, Mathilde Jauzac, David Harvey, Danuta Paraficz, Vivien Bonvin (yes, you are a postdoc now and one who stepped up and stood out), Hung-HSu Chan and John Blazec brought me a lot and helped me understand my subject and academia as a whole.

I have also been lucky enough to tutor a few students during these four years who enriched both my work and myself. I was lucky a second time for having students of great calibre who produced outstanding work and who now started promising academic careers themselves, so thanks Martin Millon, Damien Ringeisen, Nicolas Torcehboeuf and Minghan Chen.

My consecutive officemates through the years have made these long research hours much more entertaining and have also been of great support when answering my too frequent questions. You also had to endure my messiness, playtimes and toys in the office and for that you should be praised. Thanks for all you taught me Markus Rexroth, Thibault Kuntzer and Olga Tihhonova.

Thanks to my friends in Saclay as well who have always warmly greeted me and who taught

tout ce soutient et cette amour, pour tout cela vos noms ont toute leur place dans une thèse d'astrophysique Francoise Joseph, Jean-Marc Joseph et Lucas Joseph. Je vous aime.

Rémy

Abstract

Strong gravitational is a natural phenomenon that produces distorted images of distant galaxies due to massive galaxies or clusters, called deflectors or lenses, lying along the line of sight. Due to the mass of the deflector, and as a result of general relativity, the light coming from galaxies at the background follows a path that appears to us as bent, thus forming magnified and sheared images. The spectacular images of gravitational lens systems give us information about a wealth of cosmological and physical processes, from the expansion of the Universe to the nature of dark matter.

The study of strong gravitational lenses requires being able to extract information from images of such systems. When faced with this task, many problems arise. First, there is the problem of finding them. So far, only a few hundreds of strong gravitational lens systems are known, but future surveys are expected to bring hundreds of thousands new systems, to be found amongst billions of light sources. The second problem is to be able to identify lensed features and separate them from the light profile of the foreground deflector. The third problem is to be able to reconstruct the image of a lensed background source as if it had been unaffected by lensing. The last problem is the reconstruction of the deflector's mass distribution. In practice, the last two problems have to be solved as one. Recovering the light distribution of the source requires being able to invert the distortion by the gravitational lens, which depends on its mass distribution. In turn, the lens mass distribution is constrained by the distortions applied to the source's image.

In this thesis I propose to address these problems with the tools provided by sparse inverse problem solving with a particular emphasis on the problem of separating lens and source luminosity profiles. The problem of separating two overlapping light profiles is known as deblending.

I introduce three new techniques for the deblending of strong gravitational lenses that exploit different properties of lens systems and that each contribute to solving the problems of strong lens finding and modelling. The first technique allows the general subtraction of galaxies light profiles in large survey in order to facilitate the search for strong gravitational lenses. The second technique allows to separate lens and source light profiles based on their difference in colour and the sparsity of a galaxy's light profile in the wavelet. This technique allows, in particular, to remove most of the light from foreground galaxy clusters, which facilitates the identification of lensed galaxies in view of the modelling of the cluster's mass distribution.

Acknowledgements

Finally, I show that lensing itself can be used to separate the unperturbed light profile of a lens galaxy from the distorted profile of a lensed source. This technique involves the joint reconstruction of the source in its own referential based again on sparsity. At this stage, the joint reconstruction and separation is performed at fixed mass model, but this work paves the way for the development of a free form lens mass distribution reconstruction technique in the near future, based on the combination of these separation and reconstruction techniques.

Keywords: Strong gravitational lensing – linear inverse problems – sparsity – morphological component analysis – cluster lensing – galaxy surveys.

Résumé

L'effet de lentille gravitationnelle fort est un phénomène naturel qui produit la distortion d'images de galaxies lointaines par la présence galaxies et amas massifs, appelés déflecteurs, où lentilles, sur la ligne de visée. La relativité générale nous indique que la masse du déflecteur induit une apparente courbure de la trajectoire des photons émis par la galaxie d'arrière-plan, formant ainsi des images magnifiés et cisaillées. Les spectaculaires images de lentilles gravitationelles nous apportent une foule d'information sur la cosmologie et la physique de notre Univers, de l'expansion de l'Univers, à la nature de la matière noire.

L'étude des lentilles gravitationnelles fortes requiert la capacité d'interprêter les images de ces systèmes, ce qui pose plusieurs problèmes. Premièrement, le problème se pose de la recherche de ces images. Actuellement, nous connaissons quelques centaines de lentilles gravitationnelles fortes, mais les observations futures pourraient nous apporter des centaines de milliers de nouveaux systèmes qu'il nous faudra trouver parmis des milliards de cibles. Le second problème consiste à identifier les structures des galaxies lentillées afin de les séparer de la lumière du déflecteur situé à l'avant plan. Le troisième problème est la reconstruction du profil de lumière de la galaxie source d'arrière plan telle qu'elle apparaîtrait sans effet de lentillage. Le dernier problème est la reconstruction de la distribution de masse dans la lentille. En pratique, les deux derniers problèmes n'en forment qu'un. Reconstruire la distribution de lumière de la source requiert le connaissance de la distortion par la lentille, qu'elle a subi. à son, tour, la détermination du profil de masse de la lentille est contrainte par la compréhension des distorsions subies par l'image de la source.

Dans cette thèse, je propose d'adresser ces problèmes à l'aide des outils fournis par la résolution de problèmes parcimonieux inverse, avec un intérêt tout particulier pour les problèmes de séparation des profils de lumière des galaxies sources et lentilles. Les problèmes de séparation de sources lumineuses mélangées sont connus sous le nom de dé-mixage.

J'introduis ici trois nouvelles méthode de démixage des lentilles gravitationnelles fortes qui exploitent différentes propriétés des systèmes de lentilles et contribuent chacunes à la résolution d'un des problèmes de recherche et modélisation des lentilles. La première méthode permet la soustraction des profils de lumières des potentielles galaxies lentilles d'un relevé afin de faciliter la recherche automatique de systèmes de lentilles fortes. La seconde méthode permet la separation de galaxies sources et lentilles sur la base de leur différence de couleur et sur la parcimonie de leurs profils de lumière dans l'espace des ondelettes. Cette technique permet,

en particulier, de soustraire la lumière des galaxies d'un amas, ce qui facilite l'identification de sources lentillées à l'arrière plan, en vue de la modélisation de la distribution de masse de l'amas. Enfin, je montre que l'effet de lentille gravitationnelle lui-même permet de séparer la lumière non perturbée d'une galaxie lentille du profil de lumière distordu d'une galaxie source. Cette technique implique la reconstruction jointe du profil de lumière de la source, basé, encore une fois, sur la sparsité de la source dans son propre référentiel. À ce point de mes recherches, la reconstruction et séparation jointe sont faites pour un profil de masse supposé connu, cependant je montre que ce travail ouvre la voie au développement, dans un future proche, d'une méthode de reconstruction numérique de profils de masses basée sur la combinaison de ces techniques de séparation et reconstruction jointes.

Mots-clefs : Lentillage gravitationnel fort – problèmes linéaires inverse – parcimonie – Analyse en composante morphologique – lentillage d'amas – relevé de galaxies.

Contents

Contents

Contents

List of Figures

Notations

$\|\cdot\|_p$	P-norm of a vector defined as $\|X\|_p = \sqrt[p]{\sum_i |x_i|^p}$
$\langle \cdot, \cdot \rangle$	Inner product
$*$	Convolution product
$\odot$	Term by term product
$\oslash$	Term by term ratio
$\oplus$	Direct sum operator
$\cdot\|\cdot$	Matrix concatenation
$rk(\cdot)$	Operator that returns the rank of a matrix
$det(\cdot)$	Operator that returns the determinant of a matrix
$sign(\cdot)$	Operator that returns the sign of a variable
$D(\cdot)$	Operator that returns the diagonal elements of a square matrix as a vector
$\tilde{\cdot}$	Estimated value of a variable
$.^*$	Complex conjugate
$\downarrow_2$	decimation by a factor 2
$\uparrow_2$	up-sampling by a factor 2
$\hat{\cdot}$	Fourier transform of a vector
$\mathscr{F}(.)$	Fourier transform operator
$\mathscr{I}(.)$	Inverse Fourier transform operator
$.^T$	Transpose of a matrix
0_n	Square matrix of size $n \times n$ with all entries set to 0
I_n	Identity matrix of size $n \times n$
$\mathbb{R}$	The set of all real numbers
$\mathbb{N}$	The set of all natural integers
$\mathbb{Z}$	The set of all integers

Chapter 0. Notations

$L_2(\mathbb{R})$ The space of square integrable functions over $\mathbb{R}$

$M_\odot$ Solar mass

$M_\star$ Stellar mass

Acronyms

Most of these abbreviations can be found in the index that points to the pages where they are referenced

ACS	Advanced Camera for Surveys (HST instrument)
ADMM	Alternative Direction Method of Multipliers
AGN	Active Galactic Nuclei
ALMA	Atacama Large Millimeter/submilimeter Array
BELLS	BOSS Emission-Line Lens Survey
CCD	Coupled Charged Device
CFHTLS	Canada France Hawaï Telescope Legacy Survey
CNN	Convolutional Neural Network
COSMOGRAIL	COSmological MOnitoring of GRAvItational Lenses
CWT	Continuous Wavelet Transform
DES	Dark Energy Survey
DWT	Discrete Wavelet Transform
FoV	Field of View
GMCA	Generalised Morphological Component Analysis
GR	General relativity
HFF	Hubble Frontier Fields program
HST	Hubble Space Telescope
H0LiCOW	H_0 Lenses in COSMOGRAIL's Wellspring
ICA	Independent Component Analysis
ICL	Intra-Cluster Light
LSST	Large Synoptic Survey Telescope
MCA	Morphological Component Analysis

ML	Machine Learning
MuSCADeT	Multi-band morpho-Spectral Component Analysis Deblending Tool
MUSE	Multi-Unit Spectroscopic Explorer
NGVS	Next Generation Virgo cluster Survey
PCA	Principal Component Analysis
PSF	Point Spread Function
QSO	Quasi Stellar Object
RGB	Red Green Blue
SBF	Surface Brightness Fluctuation
SDR	Source Distortion Ratio
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
SIS	Singular Isothermal Sphere
SL	Strong gravitational Lensing
SLACS	Strong Lens ACS Survey
SLIT	Sparse Lens Inversion Technique
SNR	Signal to Noise Ratio
SVD	Sigular Value Decomposition
SWELLS	Sloan WFC Edge-on Late type Lens Survey
UWT	Undecimated Wavelet Transform
VLT	Very Large Telescope
WFC3	Wide Field Camera 3 (HST imager)

Introduction

"Strong gravitational lensing is the most beautiful phenomenon nature has given us to observe". This statement in itself is reason enough to start a thesis. What better motivation is there but to study what amazes and intrigues the eye? Not only is gravitational lensing a fascinating observational phenomenon, but it also constitutes a massive source of information for astrophysicists and cosmologists alike. A century ago, gravitational lensing provided one of the first tests of Einstein's theory of general relativity, and today, it remains one of the most promising probes to investigate dark matter, the expansion of the Universe, and the birth of the first galaxies.

The goal of my thesis is to apply modern techniques for image processing to analyse strong gravitational lens images. In particular, I want to address here the problem of separating images of different light sources that appear superimposed, or blended on the plane of the sky. All throughout this dissertation, we will refer to this problem as deblending.

Despite the very technical aspect of this thesis with regard to the study of strong gravitational lenses, I would like to give this dissertation a very observational-driven aspect. The goal is here to extract all the information that nature gives us in the least biased way possible. There are two concepts that I want to emphasize in this thesis and that resonate with this idea of unbiased and wonderstruck exploration of nature: accessibility and reproducibility. In the following pages I will attempt to provide a content that I want accessible to most future students willing to work on similar subjects by providing illustrated, but accurate descriptions of the concepts I use. On the side of reproducibility, I intend, when possible, to make public all the codes used to generate the illustrations I created for the purpose of this manuscript. To me, the ability to recreate the support of an argumentation, and the possibility to proof read and reuse the algorithms that are presented in any scientific work are paramount to the quality and the continuity of the scientific production.

The generic goal set for this thesis is the agnostic analysis and modelling of strong gravitational lenses. Many techniques already exist that provide satisfactory estimates of strong lenses mass and light distributions. So far these methods rely on the assumption that at either light or

mass is distributed according to a specific law, described by a mathematical model. While this assumption simplifies the problem to solve by reducing drastically the number of parameters to investigate when modelling lenses, it sometimes fail to represent the most complex objects and reduces our possibilities to explore new models. For this reason, several methods have started to emerge in the past decade that rely on semi-analytic models and decompositions with a much larger number of degrees of freedom. With the progress made in the fields of optimisation and computational astronomy, it becomes possible to consider full free-form modelling techniques that also have to be automated in order to be able to deal with the large amount of data of the upcoming years.

The first two chapters are dedicated to the introduction of the mathematical concepts used in this thesis: lensing formalism and linear inverse problem solving. While I could have extended these introductions to the broader picture in which they contribute, namely cosmology, I choose, on the contrary to give a more practical approach of my work by attempting to describe thoroughly the drivers of my work and the tools I have used. I designed these chapters so that they consist in a practical and illustrated introduction to the tools I have learned to use during these past four years. If I failed to make my introduction useful for students and scientists willing to reproduce my results or algorithms, interested readers should claim a complementary lecture to me.

Chapters 3, 4 and 5 detail my contribution to the strong gravitational lens search and modelling. These chapters heavily refer to the introductory chapters and contain technical details about the codes I developed and applied. In chapter 3, I show how the most rudimentary machine learning algorithm technique, Principal Component Analysis , can be used to facilitate the search for strong gravitational lenses. Chapter 4 gives the details of a colour-based deblending technique called `MuSCADeT`. The method was successfully applied to several datasets for which we show the results. The scope of applications of `MuSCADeT` extended beyond the problem of galaxy-scale strong lens deblending which it was designed to address and found significant applications in cluster lensing. In chapter 5 I set the first stone of a new lens modelling technique by detailing a source reconstruction method. The properties of sparsity that are used in the source reconstruction allowed to impose a separation criteria for the deblending of the lens and source light profiles in single-band images, which relies on the solving of the source inversion problem. Finally, I conclude by merging together all the concepts seen in the technical chapter to show how, the combination of these technique will allow to build a full, free form, lens modelling method based on linear optimisation.

1 Introduction to Strong Gravitational Lensing

> Lensing is not a field of study, it is a tool.
> — Paul Schechter

This chapter serves as an introduction to the broader context in which this thesis contributes. Although this work does not pretend solving any significant scientific questionings of modern Cosmology, it nonetheless intends to constitute one of the stones of a much larger structure. The following sections aim at giving the reader an overview of the main challenges in field and an outline of the few problems this work addresses, along with a clear description of the physical phenomenon studied here: strong gravitational lensing.

First, I would like to start by giving a shallow, conceptual view of what gravitational lensing is. The goal being to give the reader the necessary tools to appreciate the observational tour of gravitational lensing we are going to embark, and understand their relevance for cosmology. The next section is aimed at giving the reader a few examples of what gravitational lenses look like in a telescope, in order to fix a few ideas and concepts before delving in the more abstract sections that will follow. In the next section, we will see the details of lensing formalism, in particular, how images of gravitational systems form and what information they contain. Finally, we will review the latest discoveries and challenges in the field that this thesis intends to contribute addressing.

1.1 A history of gravitational lensing

1.1.1 The Newtonian case

Gravitational lensing refers to the bending of a light under the influence of a gravitational field. The first mention of such effect is attributed to Isaac Newton in his 1704 book Opticks (Newton, 1979, re-edition), where he considers the possibility that "bodies [could] act upon light at a distance and by their action bend its rays". This idea was used later by Johann Georg von Soldner who used the Newtonian law of gravity, under that assumption that light rays are made of low mass particles, to compute the deflection of a light ray emitted tangeantially

at the surface of the Earth (Soldner, 1804, accessed from Jaki (1978)). From this calculation, Soldner computed the deflection angle $\hat{\alpha}$ of a light ray emitted by a distant star and cruising near the surface of the Sun (see figure 1.1). This can be expressed as:

$$\hat{\alpha} = \frac{2GM}{bc^2},$$

(1.1)

where G is the Newtonian constant of gravitation, c is the speed of light, M is the mass of the Sun (or, more generally, the mass of the deflector), and b is the impact parameter of the incoming light ray, i.e. the smallest distance between the center of the Sun and the incoming light ray had it not been deflected (see figure 1.1). With an estimated value of $\hat{\alpha} = 0.84"$, Soldner concluded that if one was to observe a Star near the surface of the Sun, one would have to take this deflection into account to determine the actual position of the star in the sky. Figure 1.1, shows a scheme of the deflection of a light by the Sun. The trajectory of a light ray, in red, follows a hyperbola with the Sun at its focal point. Since an observer on Earth would only see the incidence of the light ray at the position of Earth, the star that emitted the light ray would appear to be at the position of the empty star in figure 1.1 instead of its actual position signalled by the filled star.
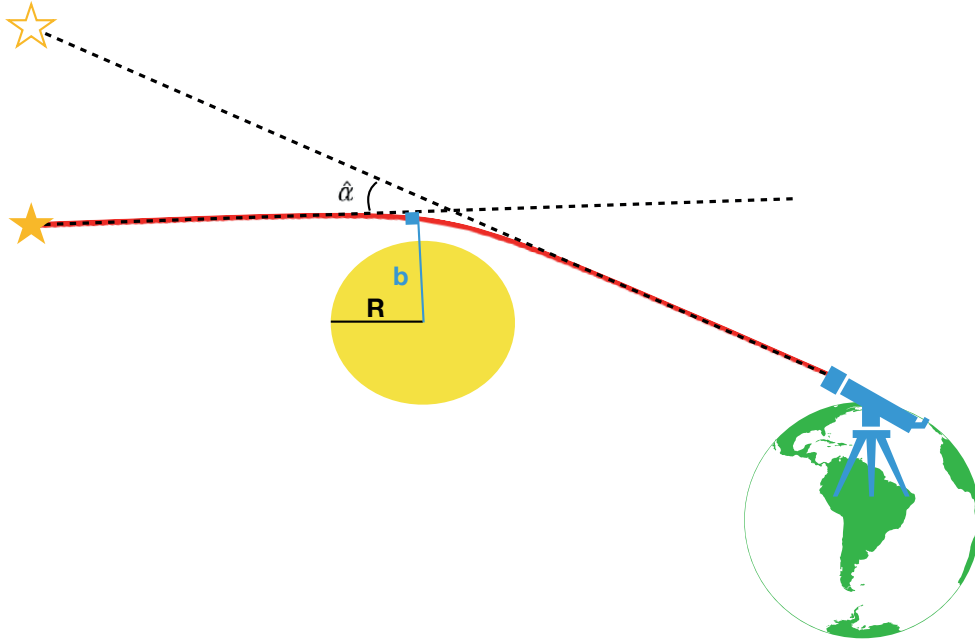
Figure 1.1 – Deflection of light by the Sun.

1.1.2 The bending of light in general relativity

Later, in 1911, Einstein, having already formulated the hypothesis of light energy quanta (Einstein, 1905, accessed in english from Arons & Peppard (1965)) that would later become photons, reproduced Soldner's calculation, which he was unaware of, and found the same expression and value as Soldner's for the deflection angle of light at the surface of the Sun (Einstein, 1911). Before anyone could verify this value through observation, gladly for Einstein, the latter had formulated his theory of general relativity (GR Einstein, 1915b). Using the framework of GR, Einstein recalculated the deflection angle of a massless photon cruising by the Sun and found the angle to be twice the angle he (and Soldner) had previously computed:

$$\hat{\alpha} = \frac{4GM}{bc^2},$$

(1.2)

The main difference from Newtonian physics introduced by GR, with regard to gravitation, stems from the realisation that space and time are intertwined and that gravity can be viewed as a local contraction of the metric in the vicinity of massive objects. These two concepts put together, we realise that the pull that massive objects exert on us results from a contraction of the space around us, but also that time is contracted in the same way as space is. As a result, a very intuitive way of explaining the factor two between the Newtonian and the relativistic values is to consider that a photon follows a path that is bent in space ($1 \times \hat{\alpha}$) but also in time ($2 \times \hat{\alpha}$). In the case of a massive, non-relativistic particle, cruising by the Sun at a speed $v << c$, however, the deflection angle computed by Soldner would be a valid approximation. A simplistic justification of this difference can be found when comparing the travel of a photon and a massive particle through space and time. Due to its high velocity, a photon explores as much as it explores space, while a massive particle with low velocity will explore much less space than a photon in the same duration. As a result, a photon is as sensitive to distortions of time as it is to distortions of space. This illustrates how the Newtonian law of gravity can be viewed as an approximation of Einstein's theory of gravity at small velocities and weak gravitational fields.

It is very important to notice here that the photon itself was never deflected, properly speaking. Instead, it followed a straight line, that, from the point of view of a distant observer, happens to be bent by the presence of a massive object, the Sun. From the point of view of the photon, its course was never altered and followed a straight line, due to the fact that its perception of space and time was altered.

1.1.3 Picturing lenses

With a value of a deflection angle larger than previously thought, the door was open to imagining observations of even more spectacular light bending phenomena. For instance, Chwolson (1924) argued that the perfect alignment between Earth, a background star and a massive foreground star would be observed as a ring around the foreground star and that a slight

misalignment would form two images of the background star form on either side of the foreground star. Einstein showed later that the angular separation between the multiple images of the background source and the deflector would be too small to provide good chances of ever observing such phenomenon (Einstein, 1936). Despite this pessimistic prediction, the Swiss astrophysicist Fritz Zwicky considered (Zwicky, 1937b,a) the alignment of two extragalactic nebulae (galaxies) and by doing so, opened the race for the search of gravitational lenses. As we will see it took many years before such phenomena could be observed, but it has now become a very active field of research, with more than 7000 papers published with the keyword "gravitational lensing"[1] to date.

The result of light deflection as predicted by general relativity, is therefore to create "mirage", or more accurately, multiple images, of background sources. Not only background images that are lensed may be seen multiple times around a deflector, but their images might also be distorted and magnified depending on the properties of the source and the geometry of the system. This will be expanded on in section 1.3.

1.1.4 Nomenclature

We use this opportunity of having described a gravitational lens to give the definition of the names and conventions we are going to use all throughout this thesis. As stated earlier, gravitational lensing refers to the bending of light by a massive object. The event of observing an alignment of objects such that the light from the background luminous object (the plain star in figure 1.1) is deflected by the foreground, massive object (the Sun in figure 1.1) before reaching us, is a **gravitational lens event**. By extension, we call **gravitational lens system** the observation, or the image of such an event . The background luminous object is referred to as the **background source** or simply, the **source** . The foreground object that causes the deflection is called a **deflector** or a **lens** . Very often in this thesis we will assimilate the light profile of the source or lens to the source or lens itself. The lensed image of a background source (e.g. an arc, an Einstein ring or a multiply imaged object) is referred to as the **lensed galaxy** or the **lensed source**. It is possible that when discussing separation of lens galaxy and lensed source, I assimilate the lensed source to the source.

Because lensing is caused by the alignment of objects in three dimensions, it is useful to define to planes with their respective coordinate systems. The **source plane**, will refer to the plane attached to the background source, while the **lens plane** will refer to the plane of the deflector, where photons are being deflected.

---

[1]https://ui.adsabs.harvard.edu/#search/q=keyword%3A%22Gravitational%20lensing%22&sort=date%20desc%2C%20bibcode%20desc&p_=0

1.2 First observations of gravitational lenses

1.2.1 Lensing by a Star

With Einstein's prediction of the possibility of observing the light from a star at grazing incidence with the surface of the Sun deflected by 1.74", it became possible to test whether this deflection was actually observed. With the light from the Sun making the observation of star in its vicinity impossible, this observation had to be conducted during a total solar eclipse. In 1919, two expeditions were commissioned to observe the solar eclipse from Sobral and from Principe, that reported the successful observation of the deflection of a background star (Dyson et al., 1920) with a deflection angle coherent with Einstein's prediction. This constitutes the first observation of the gravitational lens effect and provided yet another confirmation of Einstein's theory of relativity, after Einstein successfully explained the precession of Mercury's orbit (Einstein, 1915a).

1.2.2 First lensed quasar

Despite an early observation of the first gravitational lens event, four years after Einstein's prediction, it took decades before another gravitational lens event was observed: QSO 0957+561. It is only in 1979 that two quasi stellar objects (QSO, or quasars) were identified and reported in Walsh et al. (1979) as presenting similar properties that could qualify them as multiple images of a same object. The small angular separation on the plane of the sky (5.7"), combined with the similarity between their respective spectrum (see Walsh et al., 1979, figure 2) lead the authors to formulate the gravitational lens hypothesis. Later observations would finally reveal the presence of a galaxy at the foreground, between the two QSO images (Stockton, 1980) that could play the role of a massive deflector. This galaxy happens to be the brightest member of a galaxy cluster that is the actual deflector. The image of the lens galaxy could not have been observed before due to the relative brightness of the QSO images and their small separation in the sky. A recent composite image taken with the Hubble space telescope of the lens system QSO 0957+561 is shown in figure 1.2, that shows the apparent difference in magnitude between the two bright quasar images and the image of the lens galaxy in the middle. This single discovery confirmed the prediction made by Zwicky that lensing by extragalactic objects would produce gravitational lenses as well as the prediction made by Chwolson (1924) that two sources in an almost perfect alignment would produce two images of the same background source.

The discovery of such system would soon become a very active area of research, as recently demonstrated with the H0LiCOW publication Bonvin et al. (2017), where the authors derived a value of H_0 (More on this subject in section 1.4) at 3.5% accuracy from the monitoring of three lensed quasars.

Figure 1.2 – Optical image of the doubly imaged QSO 0957. Credit: Hubble NASA/ESA

1.2.3 First arc

The next leap in the field of gravitational lens observation comes from the discovery elongated arcs. While the previous observations of lensed objects had always consisted in the lensing of point sources (stars or QSOs), the detection of an extended, blue arc in galaxy cluster Abell 370 (Lynds & Petrosian, 1986; Soucail et al., 1987a), constituted the first observation of the lensing of an extended source. While the favoured hypothesis for explaining this arc-like structure at the time of the discovery was a star forming region, fuelled by the cooling of gas falling towards the center of the cluster (Soucail et al., 1987a), it was soon established that the arc was in fact at the background of the cluster (Soucail et al., 1987b) and therefore had to be a background galaxy, lensed by the massive cluster. A recent, high resolution composite image of cluster Abell 370 is shown in figure 1.3, where the stretching of the arc is clearly visible. Another characteristic of the gravitational arc in figure 1.3 is the brightness of the lensed galaxy. Contrary to the case of the lensed quasars, where the source had a high intrinsic luminosity, in this case, the high brightness of the lensed galaxy has to be imputed by the magnification of the lensing that makes the source galaxy appear 32 times brighter than it actually is (Richard et al., 2010). Due to their high concentration of matter, galaxy clusters are among the most powerful lenses in the Universe and as such, act as natural telescopes to observe distant galaxies by magnifying the images of objects at the background as recently illustrated by the discovery of a galaxy at redshift $z > 10$ by Oesch et al. (2016) and by the study of the detailed clumpy star forming structures in an extended lensed arc by Cava et al. (2018).

Due to their size, galaxy clusters also provide a formidable laboratory to study the distribution of mass at large scales. The large number of sources lying at the background of these clusters that are lensed provide us with as many constraints on the deflection angles their images were subject to, which tells us about the distribution of mass in these clusters at all scales
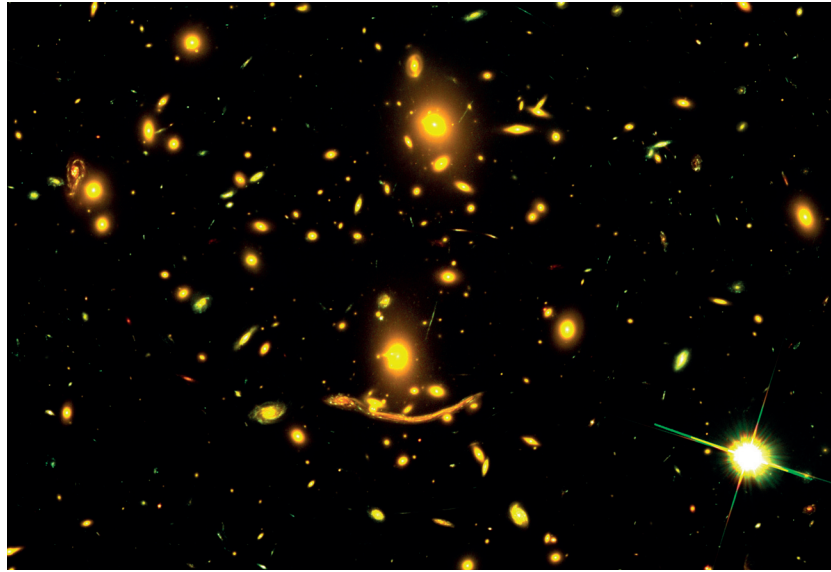
Figure 1.3 – Optical image of lensed arc in galaxy cluster Abell 370. Credit: Hubble NASA/ESA

(Harvey et al., 2016; Priewe et al., 2017; Diego et al., 2016; Sebesta et al., 2016; Jauzac et al., 2016; Meneghetti et al., 2017).

1.2.4   Einstein ring

The first Einstein ring was finally observed in the radio domain and reported in Hewitt et al. (1988). Its confirmation as a gravitational lens owes to the fact that the system could be fitted by a simple lens model (Kochanek et al., 1989). Lensing of extended sources by individual galaxies developed into yet another prolific field. With a current sample of a few hundreds galaxy-scale lens systems with a wide variety of lens galaxies: early-type galaxies (SLACS Bolton et al., 2008, see figure1.4), spiral galaxies (SWELLS Treu et al., 2011), emission lines galaxies (BELLS Bolton et al., 2012), groups of galaxies (More et al., 2012) and even lensing by AGNs (Courbin et al., 2012); future surveys such as DES (Dark Energy Survey), LSST (Large Synoptic Survey Telescope) or Euclid are expected to bring hundred thousands more of them (Collett, 2015).

With these predictions, it appears that astronomers will be faced with the problem of finding hundred thousands of objects amongst the billions of sources that these surveys are going to observe. Work in this direction has already started, and recent advances in the field of machine learning have given promising results as recently showed in Metcalf et al. (2018) and references therein.

The reason for finding strong gravitational lenses at galaxy scale stems from the amount of information these objects encode. A review on the subject can be found in Treu (2010), of which we give a brief account here. First and foremost, since the deflection in gravitational
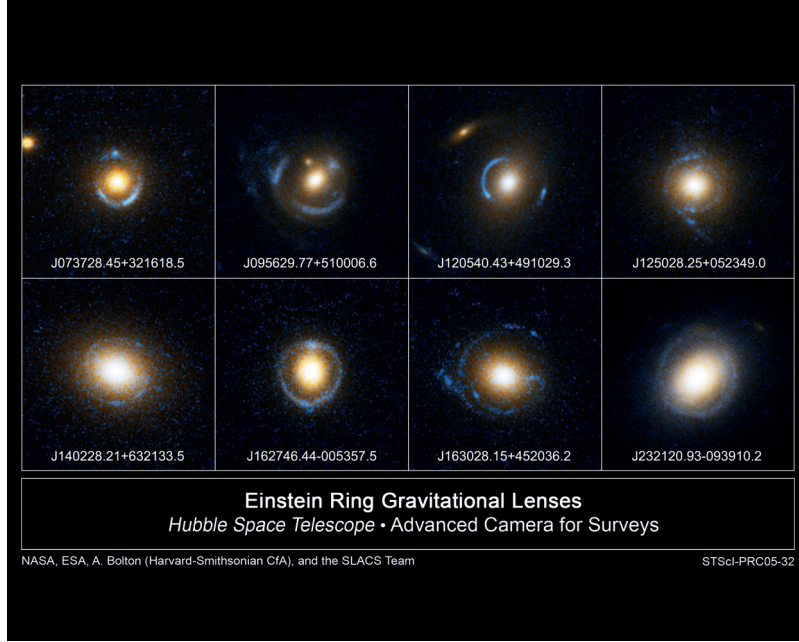
Figure 1.4 – Example of SLACS lenses. Credit: NASA, ESA, and the SLACS Survey team: A. Bolton (Harvard/ Smithsonian), S. Burles (MIT), L. Koopmans (Kapteyn), T. Treu (UCSB), and L. Moustakas (JPL/Caltech)

lensing depends on the mass of the lens galaxy (and on the relative positions of the lens and source galaxies) galaxy-scale strong gravitational lensing provides a way of weighting galaxies, and therefore learning about the structure of this objects beyond the information provided by light. Due to its sensitivity to mass, strong gravitational lensing also allows to probe the distribution of matter at small scales. Modelling the substructures in galaxy-scale lenses allows to find dwarf galaxies and sub-halos predicted by simulations (Vegetti et al., 2010; Vegetti & Vogelsberger, 2014). As stated earlier in the case of QSO0957, the monitoring of lensed quasars allows to explore the geometry of our Universe and to impose constraint on the cosmological constant H_0. Finally, galaxy-scale lenses, just as cluster-scale lenses, act as natural telescopes.

1.2.5 Lensed supernova

With many observations of lensed quasars, galaxy clusters and galaxy scale lenses across the years, the next significant first came recently with the detection of a multiply imaged supernova in galaxy cluster MACS J1149 (Kelly et al., 2015). The explosion of a distant supernova at the background of cluster MACS J1149 was observed in four images around one of the galaxy cluster members as shown in figure 1.5a. By looking at the mass distribution estimates for the cluster, an image of the supernova was expected to reappear in 2015 near the red star in the middle right panel of figure 1.5a. The predictions made by lens modellers with regard to the location, magnification and time of the reappearance provided an excellent test for our lens modelling capabilities, as shown in figure 1.5b.

(a) The lensed supernova "Refsdal" from Kelly et al. (2015)

(b) Prediction for the lensed supernova reappearance from Kelly et al. (2016)
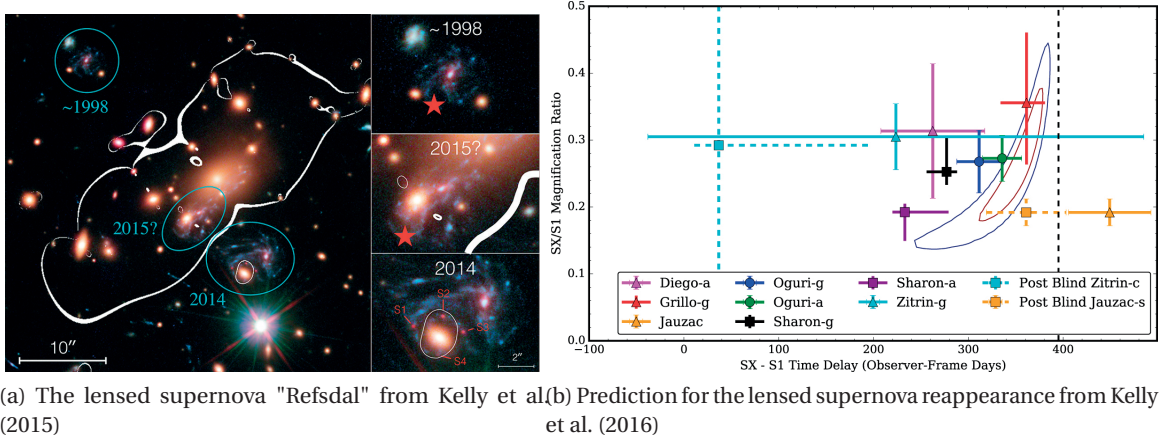
Figure 1.5 – The multiply imaged lensed supernova and its predicted reappearance in cluster MACS J1149.

The lensing of a supernova had been considered back in 1964 by Norwegian astrophysicist Sjur Refsdal in Refsdal (1964), where he predicted that the time delay between the successive appearances of the multiple images of a lensed supernova could be used to constrain the cosmological constant H_0. At the time this idea was dismissed by the community due to the low chances of ever observing such event. This was without accounting for the progress in the field of instrumentation, in particular with the development of coupled charged devices (CCD). In the mean time, the discovery of lensed QSOs as variable sources of light allowed to apply Refsdal's idea to lensed quasars (Courbin et al., 2005; Eigenbrod et al., 2005; Bonvin et al., 2017).

The supernova event observed in MACS J1149 is now known as the "Refsdal Supernova".

1.2.6 Lensed individual star

Another noticeable discovery in the field of strong gravitational lensing came, again, from observations of cluster MACS J1149. Kelly et al. (2018) reported the discovery of a transient object in a lensed arc belonging to a background spiral galaxy. With the event happening at the crossing of a line of extreme magnification (critical line) in the lens models and the spectral energy distribution (SED) of the object remaining unchanged throughout the event, it was deduced that the transient object was in fact, an individual star at the background, that was magnified by a factor $10^3 - 10^4$.

Not only magnification allowed to view the farthest individual star ever observed, but it also gives us an opportunity to investigate in details the mass distribution in the cluster at very small scales. As recently shown in Dai et al. (2018), the monitoring of the fluctuations in position and magnitude of the lensed star could tell us about the distribution of massive substructures in the lens cluster.

11

1.3 Gravitational Lensing formalism

We previously described gravitational lensing as a deflection of light by a massive object and gave the deflection angle of incoming light ray by a spherical mass. In this section, I give the deflection angle of an incoming light ray cruising by a lens with a generic mass distribution. The expression that relates this deflection angle to the observed and actual position of source in the plane of the sky is called the lens equation. To explicit this equation, we need to give a few definitions with regard to the geometry of the system and the lens mass distribution, with the help of figure 1.6. Most of the material presented here is adapted from the text books Schneider et al. (2006); Meneghetti (2006); Narayan & Bartelmann (1996); Bartelmann & Schneider (2001).

1.3.1 The thin lens approximation

I am going to use the approximation of thin lenses, which assumes that the size of the deflector is small compared to the distance between the source and the deflector and the distance between the lens and the observer, such that the curvature of the light ray that we drew on figure 1.1, can be approximated to a sharp deflection happening at the redshift of the lens. In practice, this approximation is well verified, even in the case of galaxy clusters. We define an arbitrary optical axis as the imaginary line that links the observer and the center of the lens, and that serves the as a referential to measure angular positions.
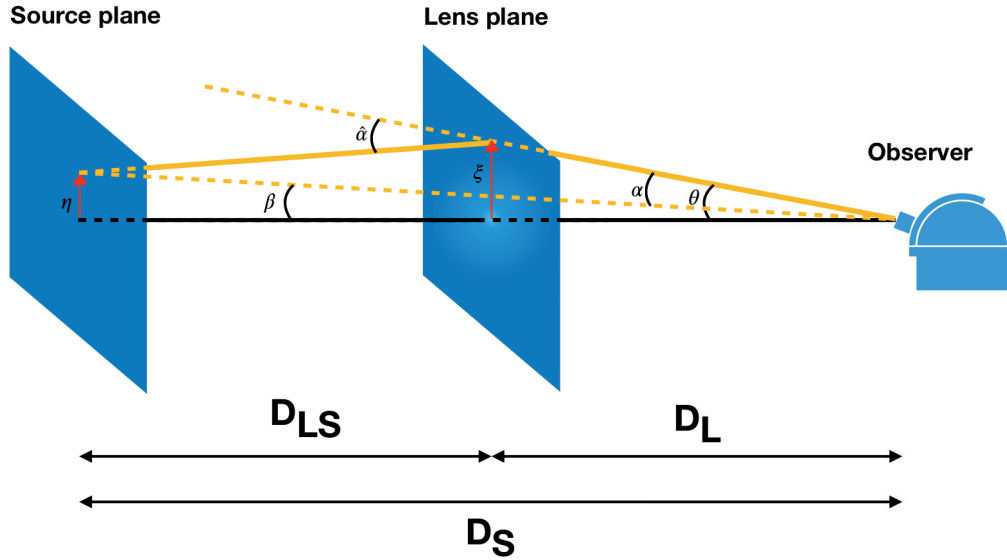
Figure 1.6 – Deflection of light by a thin lens

With this approximation, we can define two planes:

- The source plane, that lies at a distance D_S from the observer and that contains the source. the two dimensional position of a light source is given by the coordinates $\boldsymbol{\eta}$. We

write $\boldsymbol{\beta}$ the angular position of the source in the plane of the sky, relative to the optical axis, such that, in the approximation of small angles:

$$\boldsymbol{\eta} = D_S \boldsymbol{\beta} \tag{1.3}$$

- the lens plane, that lies at a distance D_L from the observer. The lens and the source are separated by a distance D_{LS}. An object in the lens plane is identified by its coordinates $\boldsymbol{\xi}$ and its angular position on the plane of the sky $\boldsymbol{\theta}$, such that:

$$\xi = D_L \theta \tag{1.4}$$

We notice here that each of the parameters $\boldsymbol{\eta}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$ are two dimensional vectors that encode a coordinate in their respective plane (lens or source), or referential (spatial or angular). We choose as a convention to write θ_x and θ_y the two components of a vector $\boldsymbol{\theta}$, such that $\boldsymbol{\theta} = (\theta_x, \quad \theta_y)$. The values D_S, D_L and D_{LS} that we coined as distance are actually angular diameter distances, that are defined by the ratio between the physical size of an object and its angular size at the considered distance.

1.3.2 Surface mass density

With the thin lens approximation, we consider that all the mass attributed to the lens is contained in the lens plane. As a result, we define the surface mass density of a lens as the projection of the lens mass on the lens plane along the radial axis r, such that the surface mass density at position $\boldsymbol{\xi}$ writes:

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, r) dr, \tag{1.5}$$

where ρ is the lens' volume density.

For the purpose of simplifying the coming equations, we also define the dimensionless surface mass density, or convergence, κ, as the ratio between the surface mass density and the critical surface mass density Σ_c, such that:

$$\kappa(\boldsymbol{\theta}) \quad = \frac{\Sigma(D_L\boldsymbol{\theta})}{\Sigma_c}, \quad and \tag{1.6}$$

$$\Sigma_c \quad = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}}. \tag{1.7}$$

As will be discussed later, if κ reaches values higher than 1 at some locations $\boldsymbol{\theta}$, the lens wil produce multiple images. The critical mass is therefore an indicator of the creation of multiple images.

1.3.3 The lens equation

The lens equation we are going to establish here, links the angular position of the observed image of a lensed object to its actual angular position in the sky. Of course, this relation depends on the deflection angle $\hat{\boldsymbol{\alpha}}$, which, in turn, depends on the mass distribution of the lens and the impact parameter of the deflected photon as illustrated with equation 1.2. Therefore, deflection angles are computed at a given angular location, such that a deflection angle writes $\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})$. We notice here that the information relative to the radial position of the source is encoded in the expression of the convergence, through the ratio $\frac{D_S}{D_{LS}D_L}$. With this definition it appears that the expression of the convergence depends on relative radial positions of the source and the lens.

Although $\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ is the deflection angle which was historically the value of interest in Soldner's and Einstein's work, we choose to use the scaled deflection angle $\boldsymbol{\alpha}(\boldsymbol{\theta})$[2], which directly expresses the difference between angular positions $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ as shown in figure 1.6, and relates to the deflection angle by:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \frac{D_{LS}}{D_S} \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \tag{1.8}$$

Using this angle, the lens equation is given by geometric construction as:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1.9}$$

This equation tells us that a photon emitted from angular position $\boldsymbol{\beta}$ will be seen by the observer at position $\boldsymbol{\theta}$ provided that the scaled deflection angle at position $\boldsymbol{\theta}$ verifies equation 1.9. It is possible that for a given $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ verifies equation 1.9 in several positions $\boldsymbol{\theta}$. In this case, the lens equation admits several solutions, and a light source at position $\boldsymbol{\beta}$ will be multiply imaged in the lens plane. With a generic lens mass distribution with a convergence $\kappa(\boldsymbol{\theta})$ at position $\boldsymbol{\theta}$, the scaled deflection angle is given by:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\boldsymbol{\theta} \in \mathbb{R}^2} \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} d\boldsymbol{\theta}', \tag{1.10}$$

which can be seen as a convolution operation of the convergence by a kernel: $\boldsymbol{\theta} \mapsto \frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2}$ (see Meneghetti, 2006, for the details on the derivation of $\boldsymbol{\alpha}$). By observing that $\forall x \in \mathbb{R}^+, \quad \nabla ln(|x|) = \frac{x}{|x|^2}$, it is possible to link the deflection angle and the convergence to the deflection potential Ψ defined as:

$$\Psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\boldsymbol{\theta} \in \mathbb{R}^2} \kappa(\boldsymbol{\theta}') ln(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) d\boldsymbol{\theta}', \tag{1.11}$$

---

[2]We notice here that angles $\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ and $\boldsymbol{\alpha}(\boldsymbol{\theta})$ are both two-dimensional vectors that account for an angular separation.

such that:

$$\boldsymbol{\alpha} = \boldsymbol{\nabla}\Psi, \quad and \tag{1.12}$$

$$\nabla^2\Psi = 2\kappa. \tag{1.13}$$

Due to the thin lens approximation and the normalisation of the convergence, the deflection potential can be seen as a rescaled projection of the gravitational potential Φ such that:

$$\Psi(\boldsymbol{\theta}) = \frac{D_{LS}}{D_L D_S}\frac{2}{c^2}\int_{\boldsymbol{\theta}\in\mathbb{R}^2}\Phi(\boldsymbol{\theta},r)dr. \tag{1.14}$$

Equations 1.11 and 1.10 tell us the deflection angle depends linearly on the convergence, which, in turn, gives access to the deflection potential. The analysis of strong gravitational lenses, through the modelling of the deflection angles therefore allows to derive the 2-dimensional deflection potential, which tells us about the distribution of mass of a lens, on the plane of the sky. In practice, the derivation of deflection angles in far from trivial. Indeed, when looking at equations 1.9 and 1.10, it appears that the only known value in these equations is θ at the position where a lensed image of a background source appears and that the relation between $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is non linear.

Multiple images

A solution to this problem of convergence estimation, comes from the identification of two (or more) multiple images of a same background source. By recording their positions $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, it is possible to use the lens equation 1.9 on their position vectors to derive an equation that has $\boldsymbol{\alpha}$ as its only unknown:

$$\boldsymbol{\theta}_1 - \boldsymbol{\alpha}(\boldsymbol{\theta}_1) = \boldsymbol{\theta}_2 - \boldsymbol{\alpha}(\boldsymbol{\theta}_2). \tag{1.15}$$

With $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ known, it becomes possible to solve equation 1.15 for κ as a linear inverse problem. This method is widely used in the field of cluster-scale lens modelling where the size and mass of galaxy clusters allow the formation of multiple images of several background sources (Jauzac et al., 2015; Hoag et al., 2016; Diego et al., 2016; Massey et al., 2015; Merten et al., 2011; Zitrin et al., 2013). In galaxy-scale strong lensing, the method is also used after careful identification of the multiple imaging of specific features in the lensed source (Eichner et al., 2012; Grillo et al., 2010; Suyu et al., 2012).

Extended sources

By extension, when the lensed source is an extended objects (e.g. a galaxy), it is possible to use the information provided by the whole image of the lensed object to derive deflection angles. To give an idea of how this works, let us consider the following example, where an extended

source that presents one particularly identifiable clump of intense star formation, is lensed and produces two images of the star forming clump. Because the rest of the galaxy does not present any particular feature, it is only possible to write equation 1.15 at the location of the clumps. However, the locations in the lens plane right next to the image of the clumps will very likely belong to regions right next to the clump in the source plane. And by transposing this idea to the neighbour of the neighbour and so on, we realise that the deflection angles at each location along the image of the lensed source has to be such that the (de-)projected image of the source back to source plane forms a coherent light profile that looks like a galaxy.

Using this concept, it is possible to design algorithms that account for the modelling of the full light profile of the source in source plane by finding the deflection angle that will reconstruct a source light profile consistent with a galaxy shape. Stating that a light profile has to be consistent with a galaxy shape is deliberately vague. In practice, many strategies can be devised that impose constraint on the source light profile's shape that encode the galaxy shape consistency. For instance, it is possible to impose that the source be made of smooth component that emulate a galaxy's radial profile (Kneib et al., 2011a; Bellagamba et al., 2017). Many modelling techniques perform numerical reconstruction of source light profiles and impose linear constraints that enforce the smoothness of the light profile (Vegetti et al., 2010; Birrer et al., 2015; Dye & Warren, 2005; Suyu et al., 2006; Wayth & Webster, 2006; Warren & Dye, 2003; Coles et al., 2014; Nightingale et al., 2018; Nightingale & Dye, 2015). This topic will be expanded on further in chapter 5.

Because the problem of finding κ at every location $\boldsymbol{\theta}$ on the plane of the sky, knowing only a small number of multiple image position, is highly ill-posed, most methods will rely on analytical profiles that describe κ using a small number of parameters. For instance, a simple lens mass model is the singular isothermal Sphere (SIS).

Singular Isothermal Sphere lens

The SIS mass model describes a system of self-gravitating particles with a given velocity dispersion σ_v, such that the density of the system at a distance r of its center is given by:

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2} \tag{1.16}$$

By integrating equation 1.16 along the line of sight, we obtain the surface mass density at position $\boldsymbol{\theta}$ on the plane of the sky, with the origin chosen to be at the center of the lens:

$$\Sigma(\boldsymbol{\theta}) = \frac{\sigma_v^2}{2GD_L\boldsymbol{\theta}}. \tag{1.17}$$

With this simple expression for the surface mass density, it is possible to compute analytically the convergence and deflection angle at location $\boldsymbol{\theta}$ as:

$$\kappa(\boldsymbol{\theta}) = \frac{R_E}{2|\boldsymbol{\theta}|}, \quad and \tag{1.18}$$

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{R_E \boldsymbol{\theta}}{|\boldsymbol{\theta}|} \tag{1.19}$$

With

$$R_E = 4\pi \frac{\sigma_v^2 D_{LS}}{c^2 D_S} \tag{1.20}$$

This simple expression for the deflection angle only depends on three parameters: R_E, the Einstein radii, or the radii of the critical curve (see section 1.3.5), and the choice for the center of the lens model $\boldsymbol{\theta_0}$ that was hidden in the choice of the origin for the coordinates system.

Under the assumption that a gravitational lens' mass density can be described by an SIS, it is possible to optimise for parameters R_E and $\boldsymbol{\theta}_0$ such that the corresponding deflection angles allow the reconstruction of the positions of a given lensed source image. More examples of lens mass density profiles are given in chapter 5.

1.3.4 Distortions

Another important aspect of strong gravitational lensing we have not dealt with yet, is the interpretation of the lens equation in terms of morphology of the lensed image. To do that, let us take a look at how a small, local perturbation of a source position is seen in the lens plane, by writing the Jacobian matrix, also called amplification matrix:

$$\mathscr{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(I_2 + \frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right)_{(i,j) \in (x,y)^2} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma 2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \tag{1.21}$$

where γ_1 and γ_2 are the shear components:

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_x \theta_x} - \frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_y \theta_y}\right), \quad and \tag{1.22}$$

$$\gamma_2 = \frac{1}{2}\left(\frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_x \theta_y}\right). \tag{1.23}$$

Matrix $\mathscr{A}$ therefore gives us a locally linearised version of the mapping of coordinate $\boldsymbol{\theta}$ into $\boldsymbol{\beta}$, to the first order. In order to get a better interpretation of the first order morphological transformation of an object in source plane into an image in lens plane, we need to take a look

at the inverse of the Jacobian, the magnification tensor $\mathcal{M}(\boldsymbol{\theta})$, which is obtained by simple matrix inversion and, if it exists, writes:

$$\mathcal{M}(\boldsymbol{\theta}) = \mu \begin{pmatrix} 1 - \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa - \gamma_1 \end{pmatrix} \tag{1.24}$$

$$= \mu \left(\begin{pmatrix} 1 - \kappa & 0 \\ 0 & 1 - \kappa \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_1 \end{pmatrix} \right). \tag{1.25}$$

From equation 1.24, it is clear that the existence of the magnification tensor is conditioned by the determinant of Jacobian matrix and requires $det(\mathcal{A}) \neq 0$. We define μ as the determinant of the magnification tensor and call it magnification for reasons that will soon become clear:

$$\mu = det(\mathcal{M}) = det(\mathcal{A})^{-1} = \frac{1}{(1 - \kappa)^2 - \gamma_1^2 - \gamma_2^2}. \tag{1.26}$$

The coordinate transform in equation 1.24 can be expanded as in equation 1.25. We see that the transformation matrix is the sum of the identity matrix rescaled by a factor $(1 - \kappa)\mu$, which corresponds to an isotropic stretch. The other component of equation 1.25 is a non diagonal matrix that depends only on γ_1 and γ_2. To have a clearer interpretation of this transformation, let us write the coordinates in lens plane (θ_x, θ_y) of a lensed source originally at position (β_x, β_y) in source plane, and decompose this coordinate transform:

$$\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} = \mu \begin{pmatrix} (1 - \kappa)\beta_x + \gamma_2 \beta_y + \gamma_1 \beta_x \\ (1 - \kappa)\beta_y + \gamma_2 \beta_x - \gamma_1 \beta_y \end{pmatrix}. \tag{1.27}$$

In equation 1.27, I highlight the different contributions of the lensing transform using colour. Now, let's assume that the coordinates (β_x, β_y) belong to the unit circle, and discuss the transformation of a circle in the source plane:

- The green component corresponds to an isotropic stretch of the coordinates, meaning that a circle in source plane will see its radius rescaled by a factor $\frac{1-\kappa}{\mu}$ in all directions in lens plane.

- The red component of equation 1.27 shows an equal contribution of β_x to θ_y and of β_y to θ_x, which corresponds to a stretch along the axis $\theta_x = \theta_y$.

- The blue component shows opposite contributions of β_x and β_y to θ_x and θ_y respectively, which indicates a stretch along one of the axes $\theta_x = 0$ or $\theta_y = 0$ and a contraction along the other. This axis along which the stretch will occur depends on the sign of γ_1, which, according to equation 1.22 depends on the slope of the gradient of the deflection potential along the horizontal and vertical axes.

From this interpretation of the first order of lensing, we get a better understanding of the naming of κ, γ_1 and γ_2. Indeed, the convergence κ, tends to make an image of a source bigger,
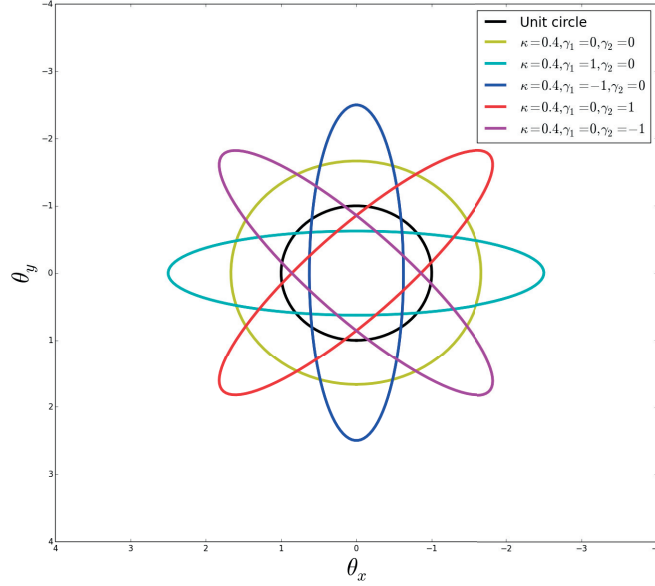
Figure 1.7 – Illustration of the magnification and shear effects. The black circle is the unit circle with coordinates (β_x, β_y) in the source plane. The coloured circles correspond to transformation of the unit circle through the first order approximation of the lens equation with different values of the shear and convergence components.

similarly to what a converging optical lens would do, while the γ_1 and γ_2 tend to shear a lensed source along the vertical or horizontal axis, and along the axis $\theta_x = \theta_y$. The effect of these transformations are illustrated on figure 1.7. By looking at the definitions for γ_1 and γ_2 in equations 1.22 and 1.23, we get the understanding that the lensing transformation produces a magnification combined with a shear in a direction orthogonal to the gradient of the deflection potential.

1.3.5 Magnification

An important consequence of the lensing distortion, is the magnification of lensed source due to the conservation of surface brightness. Surface brightness conservation is given by Liouville's theorem in the absence of absorption and emission of light, and tells us that the same surface brightness crossing a solid angle $d\Omega_\beta$ will cross its lensed counterpart $d\Omega_\theta$. From the previous linear development of lensing, we can express the area of a the solid angle in source plane, $||d\beta_x \wedge d\beta_y||_2$ as a function of the area of its counterpart in lens plane: $||d\theta_x \wedge d\theta_y||_2$. To do this, we use the definition of the amplification matrix to write:

$$d\beta_x = (1 - \kappa - \gamma_1)d\boldsymbol{\theta_x} - \gamma_2 d\boldsymbol{\theta_y} \tag{1.28}$$

$$d\beta_y = -\gamma_2 d\boldsymbol{\theta_x} + (1 - \kappa + \gamma_1)d\boldsymbol{\theta_y}. \tag{1.29}$$

Thus, the expression for an elementary solid angle in source plane:

$$d\Omega_{\boldsymbol{\beta}} \quad = ||d\boldsymbol{\beta_x} \wedge d\boldsymbol{\beta_y}||_2 = \begin{pmatrix} 1 - \kappa - \gamma_1 \\ -\gamma 2 \end{pmatrix} \wedge \begin{pmatrix} -\gamma_2 \\ 1 - \kappa + \gamma_1 \end{pmatrix} ||d\boldsymbol{\theta_x} \wedge d\boldsymbol{\theta_y}||_2 \tag{1.30}$$

$$= ((1 - \kappa)^2 - \gamma_1^2 - \gamma_2^2)||d\boldsymbol{\theta_x} \wedge d\boldsymbol{\theta_y}||_2 \tag{1.31}$$

$$= \frac{d\Omega_\theta}{\mu}. \tag{1.32}$$

We find that the ratio between the flux from an element of a lensed source and its unlensed counterpart, is equal to μ which we had thoughtfully called the magnification.

1.3.6 Consequences for lensed images

Image magnification

We have seen in the previous section that the light from a lensed source is magnified by a factor μ, which is the inverse of the determinant of the amplification matrix. One aspect of this that have kept hidden under the carpet, is the possibility for $det(\mathscr{A}(\boldsymbol{\theta}))$ to be null. Lenses admit a set of points $\boldsymbol{\theta}$ where $det(\mathscr{A}(\boldsymbol{\theta})) = 0$, that form a smooth closed curve called critical curve. The mapping of the critical curves to source plane forms caustic lines that are not necessarily smooth [3]. Formally, the magnification diverges to infinity and a background source crossing a region where $det(\mathscr{A}(\boldsymbol{\theta})) = 0$ should be magnified infinitely. In practice, the finite size of astronomical sources and the formal application of wave optics instead of its geometrical approximation, show that the magnification remains finite, even along critical curves (Ohanian, 1983). While this value remains finite, we have seen it to take extreme values as recently shown in the case of the lensed individual star in cluster MACS J1149 (Section 1.2.6).

Creation of multiple images

To better understand the creation of multiple image, it is interesting to come back to the lens equation and write using the deflection potential, such that:

$$\boldsymbol{\theta} - \boldsymbol{\beta} - \nabla\Psi(\boldsymbol{\theta}) = 0 \tag{1.33}$$

By noticing that

$$\boldsymbol{\theta} - \boldsymbol{\beta} = \frac{1}{2}\nabla(\boldsymbol{\theta} - \boldsymbol{\beta})^2, \tag{1.34}$$

---

[3]This is due to the tangent to the critical curve being parallel to the eigen-vector of $\mathscr{A}$ which eigenvalue is 0, which is always verified, given that the determinant of $\mathscr{A}$ is vanishing at any point of the critical curve. In this case, the derivative of the parametrisation of the caustic line vanishes, which can produce discontinuities.

we can rewrite the lens equation as:

$$\nabla(\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \Psi(\boldsymbol{\theta})) = \nabla\tau(\boldsymbol{\beta},\boldsymbol{\theta}) = 0, \tag{1.35}$$

where we introduced the Fermat potential, defined as:

$$\tau(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \Psi(\boldsymbol{\theta}). \tag{1.36}$$

The Fermat potential can be related to the difference in travel time delay between a photon emitted by a source at position $\boldsymbol{\beta}$ and redshift z_s, through a gravitational potential Ψ encountered at position $\boldsymbol{\theta}$ and hypothetical photon emitted from the same position, but that remains unperturbed by the gravitational potential. This time delay writes as:

$$t(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{1}{c}\frac{D_L D_S}{D_{LS}}(1 + z_s)\Big(\frac{(\boldsymbol{\theta} - \boldsymbol{\beta})^2}{2} - \Psi(\boldsymbol{\theta})\Big) = \frac{1}{c}\frac{D_L D_S}{D_{LS}}(1 + z_s)\tau(\boldsymbol{\beta},\boldsymbol{\theta}), \tag{1.37}$$

where the term in $(\boldsymbol{\theta} - \boldsymbol{\beta})^2$ relates to the geometrical time delay between light ray paths, and the term in $\Psi(\boldsymbol{\theta})$ is the so called Shapiro time delay introduced by the contraction of space-time due to the lensing potential.

With this formulation, we notice that the solutions for the lens equation coincide with the positions $\boldsymbol{\theta}$ where, for a given source position $\boldsymbol{\beta}$, the gradient of the time delay surface and the Fermat potential vanishes. This last point expresses Fermat's principle which states that light travels along the shortest path, hence the name for the potential. This means that images of lensed objects can only appear at minima, maxima or saddle points of the Fermat potential/time delay surface. Because the amplification matrix is also the Hessian of the Fermat potential, the curvature around a point of vanishing gradient of τ gives us the sign of the eigen-values of $\mathcal{A}$. Furthermore, the magnification is the determinant of the amplification matrix which is also the product of the matrix's eigen-values and the sum of $\mathcal{A}$'s eigen-values is the trace of $\mathcal{A}$, which is:

$$tr(\mathcal{A}) = 2(1 - \kappa). \tag{1.38}$$

As a result, the local curvature of the Fermat potential tells us about the local convergence and the magnification of the observed images:

- At a minimum, both eigen-values of the Hessian are positive, hence the positivity of the magnification and the trace, and a convergence $\kappa < 1$

- At a maximum, both eigen-values are negative, hence the positivity of the magnification. In this case, the trace of $\mathcal{A}$ is negative, meaning that $\kappa > 1$.

- At a saddle point, eigen-values have opposite signs, hence the negativity of the magnification. A negative magnification translates into a flipping of the source image. Nothing can be said about the convergence from this simple interpretation.

Looking at the time delay surface gives a slightly more intuitive idea about the production of multiple images and their magnification in the case of real lenses. Illustration 1.8 shows, in two dimensions instead of three, a background light source, which light we should, in the absence of a lens, receive at the minimum of the time delay surface. In this case, it would only be described by a geometrical component (Shapiro time delay is 0) and would admit 0 as its minimum. We find again Fermat's principle. In the case where a lens with an SIS mass distribution lies between the source and the observer aligned with the source, the time delay surface is perturbed as illustrated schematically by the plain black line at the top, in the case where lens and source are slightly misaligned. The dashed black line shows the unperturbed time delay surface. The light crossing through the center of the lens is slowed down by the intense metric distortion, thus forming a maximum in the time delay surface. With this perturbation at the center, the continuity of the time delay surface imposes that two minima exist on either side of the lens[4]. This first realisation gives us an intuition about the odd number theorem (Dyer & Roeder, 1980; Burke, 1981), which states that any lens with a smooth surface mass density decreasing faster than $|\boldsymbol{\theta}|^{-1}$ (the SIS model), will create an odd number of images. More precisely, the number extrema of the Fermat potential is equal to the number of saddle points plus one, provided that the source is not on a caustic line, thus enforcing oddness.

The odd number theorem and the characteristics of the extrema and saddle points as shown above tell us that multiple images occur for lenses that admit at least one position $\boldsymbol{\theta}$, where $det(\mathscr{A}(\boldsymbol{\theta})) < 0$. Similarly, if a lens forms an image at a position $\boldsymbol{\theta}$ where $\kappa(\boldsymbol{\theta}) > 1$, this image is either a saddle point, or a maximum, which imposes that other images exist. Another interpretation of this sketch can be made in terms of magnification of lensed images. Let us consider that the source is extended and fills the solid angle $d\Omega_\beta$. The representation of the solid angles as red intervals tangent to the time delay surface is technically not true, since each point of the extended source actually generates its own time delay surface. However, given that time delay surface are smooth, we approximate the area of the extremum of the time delay surfaces to the region around the extremum in the time delay surface of the central path of the source, where the surface can be considered flat. Due to the density of the mass profile around the center of the lensed, we see that the perturbed time delay surface produces an image in the center that with a smaller solid angle, while the images on either side of the lens appear with a larger solid angle. This means that the image seen through the center of the lens will appear demagnified. In practice, this central image is rarely seen due to the demagnification combined with the fact that the lens is usually a bright galaxy.

---

[4]In two dimensions, we would see that these two points are not strict maximum, given that their second eigen-value is zero. In this specific case of perfect alignment, the images, would actually form as a ring along the critical curve.
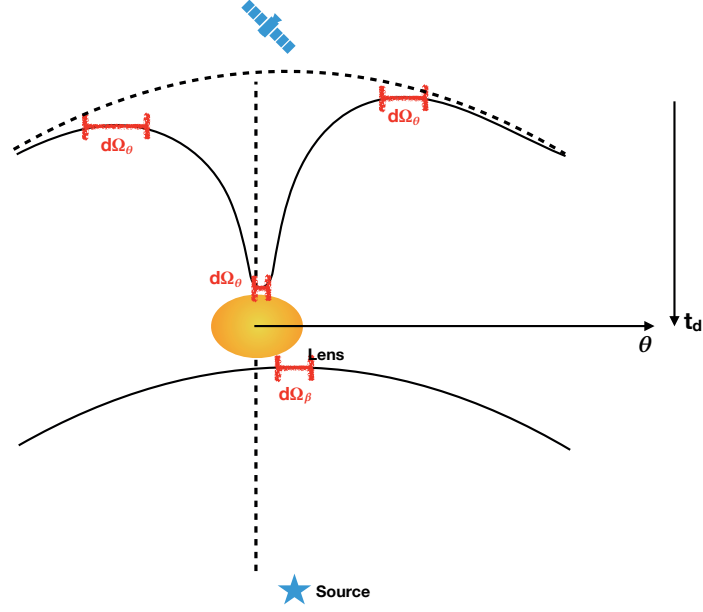
Figure 1.8 – Distortion of a time delay arrival surface.

1.3.7 The different regimes of lensing

As we have seen, gravitational lensing occurs when a lens and a source are aligned with an observer. In the case of a perfect alignment, an extended background source will be imaged as a ring around the lens. But as we deviate from perfect alignment, lensing can enter a different regime.

Strong gravitational lensing

Strong gravitational lensing is the regime where source and lens are sufficiently aligned so that multiple images and high order distortions are formed. As stated earlier, the fact that a surface mass density gets larger than the critical surface mass density leads to the possibility of multiple images to form. In general, gravitational lenses are called strong when a source is well aligned with a region where $\kappa > 1$, which tends to form multiple images and distortions. All the previous examples of lenses given in section 1.2: Multiple images of point sources, arcs, rings, and highly magnified point sources, are all strong gravitational lenses.

Weak gravitational lensing

Weak gravitational lensing occurs when the alignment between lens and source large enough so that the main effect of lensing on an extended spherical source is the imprinting of an ellipticity due to the shear. This effect is usually visible in the outer regions of massive galaxy clusters. Due to background source galaxies having their own intrinsic ellipticity, it is, in

practice, a challenging problem to decouple the weak lensing effect from intrinsic shapes and alignments.

Mililensing

Mililensing can be seen as a perturbation of an otherwise smooth lens macro-model. While so far we have only considered the formation of multiple and distorted images by a smooth, mono-modal lens mass distribution, it is possible that galaxies present massive substructures. These lenses of smaller angular size and mass will perturb the light profile of the lensed source by introducing variations of magnification, image source position, or small distortions.

Microlensing

Microlensing can be seen as an occurrence of strong gravitational lensing where the extent of the lens's critical curves is too small to be resolved. Observing a source behind such lens would only produce a higher magnification of the source. For instance, multiply imaged quasars are subject to microlensing by stars inside the lens galaxy producing an artificial luminosity variability in each of the quasar images.

In this thesis, we will only consider strong gravitational lenses. This type of lenses are, in practice, very complex to model for several reasons:

- The ill-posedness of the problem. Modelling lenses from an image of a gravitational lens requires to estimate both the surface mass density of the lens as well as the surface brightness of source that went through extreme distortions and magnification,

- The blending from the lens light that contaminates the surface brightness of the source,

- The non-linearity of the problem with regard to the mass surface density.

- Other problems may arise from the intrinsic complexity of observational constraints and of lens systems such as convolution by a telescope's impulse response, multi-plane lensing, contamination by other objects (e.g. stars in our galaxy)

1.4   Science with Strong Gravitational Lensing

Now that we have a clear overview of the type of lenses that we observed so far and that we have a better understanding of how gravitational lenses are made, let us take a look at what lenses tell us about our Universe.
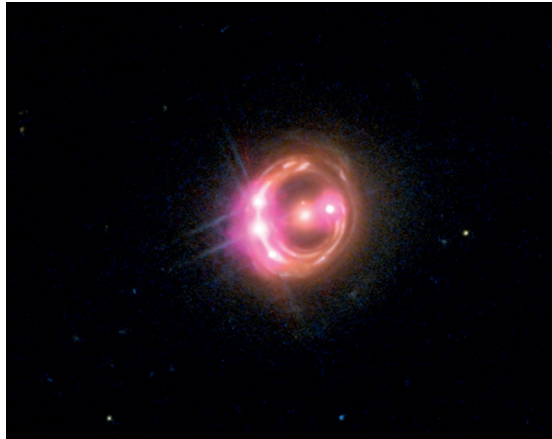
Figure 1.9 – Image of lens system RXJ1131.

1.4.1 Cosmic telescope

In gravitational lenses, magnification creates images of distant, background objects that appear bigger and brighter than if they were directly observed, without lensing. Each gravitational lens event therefore provides us with unique opportunities to probe the small scales of distant galaxies (Marshall et al., 2007). This was recently illustrated in Adamo et al. (2013); Johnson et al. (2017); Wuyts et al. (2014); Cava et al. (2018), where the authors used cluster lensing to study structures at sub-kiloparsec scale in lensed arcs down to redshift 2.5 (Johnson et al., 2017). Because lensing tends to enlarge images of background objects, it also makes the study of galaxies hosting active galactic nuclei (AGN) possible. Active galactic nuclei are supermassive blackholes at the center of a galaxy. In extreme cases, the inflow of matter in the blackhole releases energy as light and therefore produce extremely luminous objects such as QSOs. The properties of an AGN exhibit remarkably tight relations with the properties of the host galaxy (Hopkins et al., 2009; Gültekin et al., 2009). In cases where the AGN forms a QSO, the extreme brightness of the object combined with shape of the telescope's impulse response will see the luminosity of the host galaxy blend behind the diffraction figure of the telescope. When strongly lensed, the angular size of a host galaxy is increased while the size of the diffraction figure of the QSO remains the same and is only multiply imaged. This allows to study the luminosity of the host galaxy in great details. An example of the magnification of such object comes with lens system RXJ 1131-1231 (Sluse et al., 2003) shown in figure 1.9 [5] and recently reconstructed in source plane in Birrer et al. (2016), and by extension, with all the lenses from the COSMOGRAIL collaboration (Eigenbrod et al., 2005).

1.4.2 The mass content of galaxies

Since the ground breaking work of Vera Rubin on galaxy rotation curves (Rubin et al., 1978, 1980), we know that individual galaxies sit in a halo of matter that encompasses much more

---

[5]Image credit: X-ray: NASA, CXC, Univ of Michigan, R.C.Reis et al; Optical: NASA,STScI

mass than what luminous matter lets us think. This corroborated the previous observations made by Fritz Zwicky in Zwicky (1933) that the Coma cluster had a much higher gravitational potential that what could be inferred from its luminous mass. These observations led astronomers to form the hypothesis of large amounts of non luminous massive matter around galaxies and cluster, that could only be detected through its gravitational manifestations. This is what we now call dark matter (DM). In recent decades, the dark matter hypothesis found even more supporting ground with the observation of the bullet cluster, where it was revealed that the mass distribution of two merging clusters was not following the matter distribution traced by X-ray gas. This difference in distribution hints towards a dark matter component that does has low interactions with matter (at least X-ray gas), or with itself. Even more recently, Harvey et al. (2015) mapped the dark matter distribution in 72 merging clusters using weak gravitational lensing. They used the relative alignments with hot gas densities to impose an upper limit on the dark matter self interaction cross-section.

For now, the nature of dark matter remains a mystery. The fact that dark matter seems to manifest only through its gravitational interactions, makes gravitational lensing the perfect tool to investigate its nature. Indeed, by observing how dark matter behaves and interacts, we can access a better understanding of its composition as shown by Harvey et al. (2015).

1.4.3   The detection of Massive substructures

With advances in the field of numerical simulations, it is now possible to generate simulations of a Universe realisation based on our current understanding of cosmology, and compare the properties of the simulated Universe with our own (Somerville & Primack, 1999; Springel et al., 2001; Springel & Hernquist, 2003; Klypin et al., 2011; Genel et al., 2014; Schaye et al., 2015; Revaz & Jablonka, 2018). An interesting characteristic of these simulated Universes is the dark matter mass function, given by the number of dark matter halos per mass interval. While cosmological simulations predict a large abundance of small (dwarf galaxies with $M_\star \lesssim 10^9 M_\odot$) structures, observations do not seem to agree and tend towards a lower abundance of substructures (as illustrated in Bullock & Boylan-Kolchin, 2017, , figures 5 with the data from Sheth et al. (2001); Wright et al. (2017); Bernardi et al. (2013)). This discrepancy can be explained in two ways, which are best summed up by the names given to this problem: the excess subhalo problem on one side, and the missing satellite problem on the other side. The excess subhalo refers to the possibility that we should question our current understanding of cosmology and that the model we built so far to describe our Universe does not correspond to the reality. The missing satellite problem refers to a defect on the observational side. A likely possibility is that low mass substructures do not form enough stars to be detected. In this case subhalos would exist as dark, low mass substructures.

In the case that the latter be true, gravitational lensing would be, at the moment, the best probe to detect such substructures. Since the shape of a lensed source is sensitive to mass only, dark subhalos would imprint characteristic features in images of lensed galaxies. Such

features can only be identified through the accurate modelling of both source light profiles and lens mass distribution.

Strong gravitational lensing has already proven to be a promising way of detecting such substructures as shown in Mao & Schneider (1998); Bradač et al. (2002); Dalal & Kochanek (2002); Nierenberg et al. (2014); Gilman et al. (2017), where the authors of these studies showed anomalies in the flux ratios between multiple images of lensed quasars that could be best explained by massive substructures. In the works of Koopmans (2005); Vegetti et al. (2010, 2012), the authors successfully modelled anomalies in the lensed source surface brightness as massive, dark substructures. A recent and elegant approach was proposed in Hezaveh et al. (2016) and applied in Bayer et al. (2018) that, instead of modelling individual substructures, proposes to extract the power spectrum of the perturbations of the surface brightness of lensed galaxies.

1.4.4 Cosmography

Since the work of Lemaître (Lemaître, 1927), we know that galaxies in every directions are moving away from us. Even more, they are moving away faster as they are farther and this velocity distance relation scales linearly. This effect is known as the expansion of the Universe and expresses the fact that all galaxies in our Universe are moving away from one another. This expansion phenomenon can be seen as a conservation of the momentum given to all the particles in the Universe at the time of inflation. The expansion rate, that was first highlighted by Georges Lemaitre, was "naturally" called the Hubble constant, H_0 after Edwin Hubble who reached Lemaitre's conclusions in Hubble (1929). Due to galaxies moving away from us, the photons we receive from them are subject to a Doppler effect. The result is that the wavelength of the observed photons is shifted to longer wavelength making distant galaxies appear redder than they would in a Universe without expansion. Because galaxies move away from us faster as they are farther, the wavelength shift towards the red, henceforth called redshift and symbolised by the letter z, can be used as a proxy for their distance to us. In practice, the redshift of a galaxy is computed by identifying spectral features, such as emission or absorption lines and measuring by how much their measured frequency differs from their expected frequency in rest frame.

Coming back to the lensing formalism we previously established, we forgot (on purpose) to notice that the time delay written in equation 1.37 depends directly on the angular diameter distances of the system. In an expanding Universe, these angular diameter distances are scaled by the expansion rate. In other words, angular diameter distances are proportional to H_0^{-1}. In a curved space with matter and dark energy, the angular diameter distances also depend on the matter, dark energy and curvature densities but only to the power $-\frac{1}{2}$. While time delays themselves cannot be observed due to the fact that if the arrival time of a lensed photon is known, we cannot know the reference arrival time of a non-deflected photon emitted at the same time as the lensed photon, it is possible to access time delay differences through multiple

imaging. As proposed by Refsdal (1964), and recently illustrated in (Kelly et al., 2015; Kelly et al., 2016), the multiple imaging by a gravitational lens of a supernova exploding in a distant galaxy at redshift z_S would produce consecutive appearances of the supernova luminosity peak in each of the multiple images. By recording the time interval $\Delta t_{1,2}$ between the occurrence of the peak in two of the multiple images at position θ_1 and θ_2, one can link the angular diameter distances to an observational quantity:

$$\Delta t_{1,2} = (1 + z_S)\frac{D_L D_S}{c D_{LS}}(\tau(\boldsymbol{\beta}, \boldsymbol{\theta_1}) - \tau(\boldsymbol{\beta}, \boldsymbol{\theta_1})). \tag{1.39}$$

We note here that z_S and $\boldsymbol{\beta}$ are common to both images of lensed supernova. With the formulation of equation 1.39, we know have an observational quantity that is proportional to H_0^{-1}. Before we can conclude that H_0 is therefore easily measurable using multiple imaging, there are a few aspects that need to be emphasised:

- First, only a few lensed supernovae have been observed to date, with little monitoring of the multiple images light curves. However, Refsdal idea can easily be extended to general variable sources. In practice, QSOs present characteristic variations over timescales of a few days. However, since these variations are, in general, less peaked than supernovae explosion light curves, and contaminated by microlensing from the lens galaxy stars, it is necessary to build robust time delay measurement techniques. In the frame of the COSMOGRAIL survey, Tewes et al. (2013); Bonvin et al. (2016) developed the pyCS curve shifting package, which relies on simultaneous spline fitting, residual minimisation of the residuals between individually interpolated light curves, and dispersion minimisation between light curves.

- Second the time delay difference has a strong dependence on the Fermat potential which depends on the deflection angle and the deflection potential. Both of these values need to be carefully estimated before any computation of H_0 can be made, which requires accurate lens modelling. In the recent work of the H0LiCOW collaboration, the lens modelling was performed using a code optimised for galaxy scale modelling of lensed quasars, GLEE (Suyu et al., 2006; Halkola et al., 2008; Suyu & Halkola, 2010; Suyu et al., 2012), that reconstructs lens mass profiles based on multiple image positions and extended lensed source light profile reconstruction with an adaptive regularisation parameter.

Applying these techniques to decade long monitoring of the light curves of three gravitationally lensed quasars, Bonvin et al. (2017) achieved a competitive measurement of H_0, down to a 3.8% accuracy. With more lens systems to exploit (Suyu et al., 2017) and upcoming data with higher sampling, the future of time delay cosmography will be magnified.

1.5 Outline

Gravitational lensing is a now well understood phenomenon that results from the rare combination of factors. To be observable, a gravitational lens has to show a lensed image of the source which brightness and separation form the lens galaxy light profile allows detection by our Earthly telescopes. This does not only require that a light source be aligned on the plane of the sky with a massive lens. The system formed by aligned objects also has to verify several conditions with regard to the mass distribution of the lens, the relative redshifts of the lens and source, angular position and brightness of the source relative to the lens, and luminosity of the lens. Only under these very peculiar conditions can a lens be observed. Then, the fact that a lens be observable does not necessarily mean that it will be observed. Indeed, the most spectacular lensing regime: strong gravitational lensing, is found at galaxy scale, to produce systems of a few arcseconds in angular size on the plane of the sky. Knowing that the full sky covers 64800 square degrees and has billions of light sources, finding strong lens candidates is not an easy task. The first problem, when working on gravitational lenses is to detect them. In chapter 2, we are going to see how the community addresses the problem and how linear decompositions of images can help in this regard. With lensing being a large toolbox on its own, as shown in section 1.4, it is paramount that we develop robust tools to model them and extract useful and accurate information from these objects. Due to lenses being formed of aligned lens and source light profiles, the first challenge is to disentangle between both surface brightnesses. In the case of early type, mono-modal galaxies acting as lenses, the problem can be addressed by fitting parametric profiles to the lens luminosity, but for more complex lens shape, and in the case complete blending between lens and source light profiles, the separation requires higher orders of modelling.

The following chapters will contribute in large parts to solving this problem using different techniques that each exploit a different property of lens systems. In chapter 3, I exploit the fact that lensed objects have a widely different morphology from unlensed galaxies. This allows, in large surveys, to learn the morphological features of galaxies that could potentially be lens galaxies and decompose lens candidates on these features to separate lens form source light profiles. In chapter 4 I use the fact that lensed sources are, by construction, galaxies at higher redshift than lens galaxies which hints at source being more star forming and therefore have a different colour from lens galaxies. This difference in colour is observed in practice and allows to build separation techniques relying on the colour difference between the two. In chapter 5, I present a technique for lensed source light profile reconstruction in source at known surface mass density. Interestingly, the difference in morphology between lensed galaxy and lens galaxy light profiles, allows to perform a joint reconstruction of the peculiar features of each light profile in their respective referential and therefore, to solve the problem of lens and source deblending as I reconstruct a source light profile.

2 Introduction to inverse problem solving

Contents

In this second Chapter, we introduce the mathematical tools that will be referred to all along this dissertation. This introduction should not be seen as a toolbox, but instead, as an overview of the field of linear optimisation, which is central to this work. One of the novelties of the approach presented in this thesis lies in the effort made to express the many problems encountered in strong gravitational lensing, in particular the problems relative to blending, as linear inverse problems and solve them as such. The following sections detail the technical aspects of the linear framework we chose for our analysis.

2.1 Linear Inverse Problems

Across most of this dissertation, we are going to formulate various problems of denoising, reconstruction or deblending for astronomical images as, generally ill-posed, linear inverse problems. Linear inverse problems consist in solving:

$$Y = AX \tag{2.1}$$

in X, where Y is an $\mathbb{R}^n$ vector containing measurements of an experiment (images or cubes of images), X is the $\mathbb{R}^m$ vector containing the unknown model to estimate and A is an $\mathbb{R}^{n \times m}$ matrix that links the model to the measurements. In practical image processing applications, matrix A usually stands for a morphological transformation of a distribution on a 2D plane such as lensing [see chapters 1 & 5], a blurring operator in the case of convolution by a smooth kernel, or a masking operator that represents missing data. Equation 2.1 can be seen as a set of n equations with m unknowns.

2.1.1 Ill-conditioning

Equations such as equation 2.1 admit a unique solution that can be derived analytically if and only if the number of independent equations is equal to the number of unknowns, which translates by $rk(A|X) = m$. In cases where $rk(A|X) \leq m$, the solution is not unique and it is only possible to derive subspaces of solutions which then have to be constrained using physical assumptions on X in order to estimate a suitable solution for equation 2.1. If $rk(A|X) \geq m$ the equation is inconsistent and there is no solution that can satisfy all independent equations. All cases where $rk(A|X) \neq m$ are ill-posed problems in the sense of Hadamard, meaning that either:

- The problem admits no solution,

- There is no unique solution to the problem

- Or, there is continuous solution over the input parameter space.

A very simple illustration to this comes when facing a problem of the form :

$$y = a_1 x_1 + a_2 x_2, \tag{2.2}$$

where y, a_1 and a_2 are known real numbers, and x_1 and x_2 are unknown scalars. With one equation and two unknowns, it is only possible to define an ensemble of solutions of the form:

$$x_2 = \frac{y}{a_2} - \frac{a_1}{a_2} x_1. \tag{2.3}$$

Equation 2.3 defines a line in the (x_1, x_2) plane along which every point satisfies $Y - AX = 0$. Deriving a unique solution requires more constraints, either under the form of an other equation, or as a physical assumption on the possible solution. We will now use this example to illustrate a set of constraints that are commonly used in physics and that are essential to the rest of this manuscript, the ℓ_p-norm constraint, with $p \in \{0, 1, 2, \infty\}$. For a given vector $X = \{x_0, x_1, ..., x_{m-1}\}$ in $\mathbb{R}^m$ with $x_{i \in [[0, m-1]]} \in \mathbb{R}$, it ℓ_p-norms are defined as follows:

$$||X||_\infty \quad = \quad \max_{x_i}(|x_i|) \tag{2.4}$$

$$||X||_2 \quad = \quad \sqrt{\sum_{i=0}^{i=m-1} x_i^2} \tag{2.5}$$

$$||X||_1 \quad = \quad \sum_{i=0}^{i=m-1} |x_i| \tag{2.6}$$

$$||X||_0 \quad = \quad \sum_{x_i \neq 0} 1 \tag{2.7}$$

Strictly speaking, $||.||_0$ is not a norm as it does not follow the absolute homogeneity. Another downside of this function, as will be illustrated in section 2.2 is the non convexity of this application. Nonetheless, we find it very convenient constraint in practical applications.

The general idea of constraining a solution space with ℓ_p-norms is to favour the recovery solutions with low intensity, or at least, concentrated in a small number of coefficients, which is a common physical assumption. In order to apply these constraints we derive the solution $(\tilde{x}_1, \tilde{x}_2)$ that satisfies equation 2.3 while having the smallest ℓ_p-norm. For a more formal mathematical development of this aspect, we refer to section 2.2, the aim here being to give the reader a more empirical idea of how these constraints work. In figure 2.1 we show how we can compute a unique solution for equation 2.2, by inflating the ℓ_p-sphere (in colour in figure 2.1) until they intersect with the solution space. An ℓ_p-sphere is an ensemble of vectors X that have the same ℓ_p-norm. On figure 2.1, we see that these spheres intersect the solution space in a unique point, giving, for each of these constraints a unique solution that holds different physical properties. By closely looking at the shape of the ℓ_0 sphere that verifies $||X||_0 = 1$, we notice that it intersects the line described by $x_2 = y/a_2 - x_1 a_1/a_2$ in two
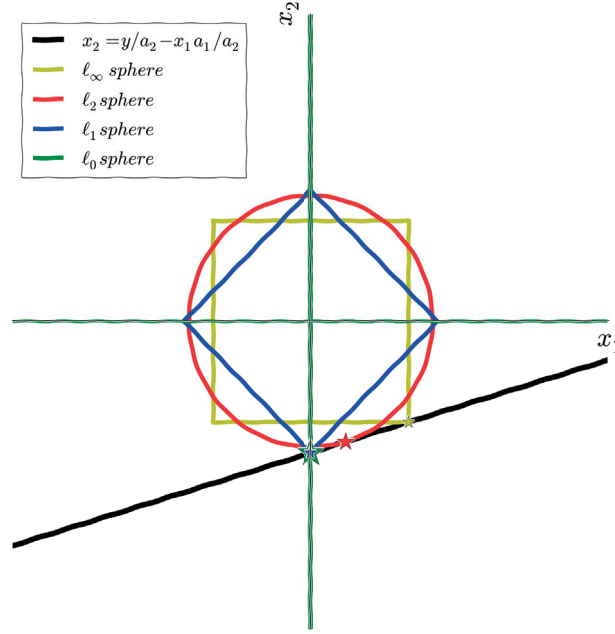
Figure 2.1 – The ℓ_p-norm constraints used to derive a unique solution for an ill-posed problem. The black line represents the sub-space of solutions for equation 2.2. Coloured curves show the ℓ_p spheres that intersect the solution space. Coloured stars show the solution derived with each corresponding norm.

positions: one highlighted by the green star, the other one, outside of the frame of figure 2.1. From this realisation, it is easy to understand that in higher dimensional and highly degenerate problems, the non unicity of the ℓ_0 solution becomes a combinatorial problem. This is due to the non-convexity of the ℓ_0 "norm", which we can understand graphically from figure 2.1, where we see that the green lines do not delimit a closed, convex set, while other coloured lines do.

2.1.2 Noise in linear inverse problems

In practice, experiment results are often contaminated by additive and/or multiplicative noise, that are represented by vectors M and Z respectively, in $\mathbb{R}^n$, such that equation 2.1 becomes:

$$Y = M \odot AX + Z. \tag{2.8}$$

In the following astrophysical image processing applications, we will exclusively consider additive noise, thus setting M to 1. Noise is usually a random degradation of the measurements introduced by the instrument and for which we can only access its statistics at best. Gaussian distribution, Poisson distribution or Gaussian and Poisson mixture are the most common

| (a) Target measurement | (b) Kernel | (c) Convolved image |
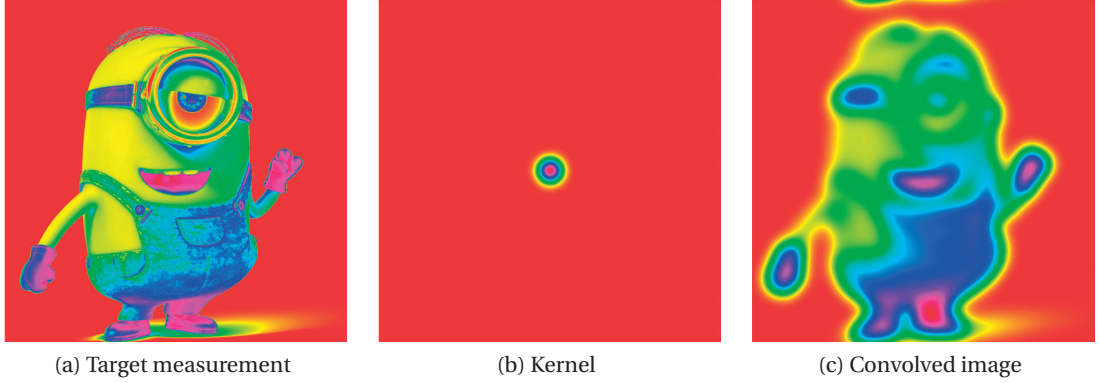
Figure 2.2 – Convolution of an image (*left panel)* by a kernel (*middle panel)* to obtain a blurred image (*right panel)*

noise statistics in astronomical imaging. Without noise, it is very often possible to recover the exact solution of equation 2.1 as seen in the previous section, but even at very high signal to noise ratio, the presence of noise changes the nature of the problem as illustrated in the following example.

Illustration on deconvolution

Deconvolution in Fourrier space Deconvolution is a very common problem encountered in astrophysics and in image processing in general, where we aim at recovering a signal that has been "blurred" by the response of the measurement instrument for instance. Blurring, refers to the operation of convolution of the targeted measurement by a smooth kernel. In this illustration, let us consider image $\underline{X}$ (fig. 2.2a) with N_p samples (or pixels), and blur it using convolution by a Gaussian kernel $\underline{k}$ (fig. 2.2b), with the same number of samples as in $\underline{X}$, such that the resulting image $\underline{Y}$ can be described by the relation:

$$\underline{Y}[m,n] = \underline{k} * \underline{X} = \sum_{i,j} \underline{k}[n-i, m-j]\underline{X}[j,i]. \tag{2.9}$$

The resulting image Y is showed in fig. 2.2c.

The problem of recovering image 2.2a from image 2.2c is a very common problem in astrophysics and in image processing in general. In this particular case, a simple solution comes from the formulation of the problem in Fourier space. When rewriting equation 2.9 with the Fourier transform of its terms, the convolution product becomes a term by term multiplication (see eq. 2.10).

$$\hat{\underline{Y}} = \hat{\underline{k}} \odot \hat{\underline{X}} \tag{2.10}$$

A very simple solution to recover $\underline{X}$, provided that we know $\underline{k}$ is therefore to divide $\hat{\underline{Y}}$ by $\hat{\underline{k}}$ and

| (a) Target measurement | (b) Convolved and noisy image | (c) Reconstructed image |
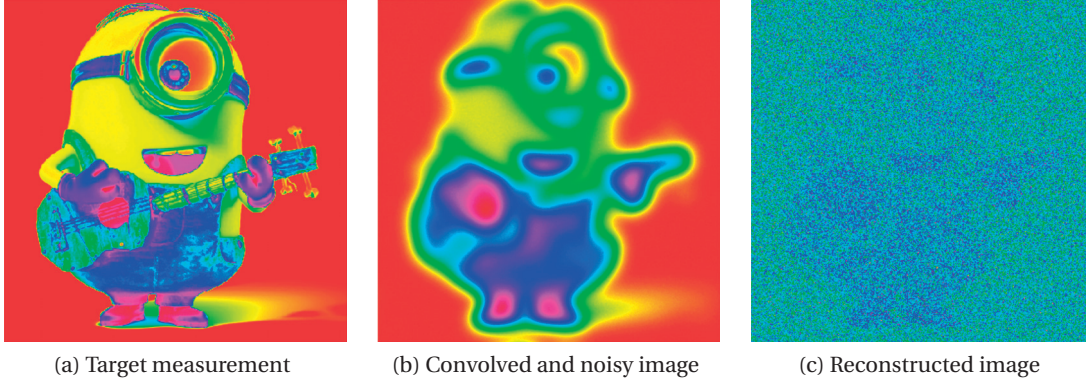
Figure 2.3 – Reconstruction of a blurred, noisy image from kernel division in the Fourrier domain. From left to right: the target measurement, the blurred and noisy image, and the reconstruction from Fourrier space division.

transform the result back into direct space. In this case, the result is an exact recovery of the original image, expressed as $\underline{X} = \mathscr{I}(\underline{\hat{Y}} \oslash \underline{\hat{k}})$.

Deconvolution of noisy images   As a more realistic study case, we will now perform the same inversion on a noisy image, meaning that white Gaussian noise $\underline{Z}$ with standard deviation σ is added on the right handside of equation 2.9. An illustration is given in fig. 2.3, where image 2.3a is the original image and image 2.3b is its convolved version with noise added. Dividing $\underline{\hat{Y}}$ by the operator $\underline{\hat{k}}$, yields a very different result shown in figure 2.3c.

This result is easily explained by writing explicitly the solution we computed here as equation 2.11.

$$\underline{X} = \mathscr{I}(\underline{\hat{Y}} \oslash \underline{\hat{k}} + \underline{\hat{Z}} \oslash \underline{\hat{k}}) \tag{2.11}$$

In this expression, the term $\underline{\hat{Z}} \oslash \underline{\hat{k}}$ contains frequencies from the noise divided by those of the convolution kernel. Since we used a Gaussian kernel, which is localised in frequency, while the white Gaussian noise is not by definition, we end up enhancing some of these frequencies compared to the frequencies in the signal. In figure 2.3c, we clearly see that high frequencies contaminate our reconstruction.

From equation 2.9, we observe that the convolution product is a linear operation that can be expressed as a matrix product:

$$Y = KX. \tag{2.12}$$

In equation 2.12, matrix K is a matrix of size $N_p \times N_p$[1] that accounts for the linear operation

---

[1]We considered X and Y have the same size here, but in the general case, they might of different sizes. If so, the size of matrix K is the product of the sizes of Y and X.

of convolution. One might also notice that I dropped the underlining of the matrices Y and X. This is to signify that X and Y are one dimensional vectors that are constructed by ordering the consecutive lines of their underlined counterparts as a single line. From this we see that the lines of the convolution operator K are circular variations of the transpose of $\underline{k}$ ordered as a one dimensional column. From now on, this is the formalism we are going to refer to, to describe the convolution operation, and, by extension, linear inverse problems in general.

Least squares   Because uncertainties introduced by noise make it impossible to derive a simple solution for X from the Fourier quotient, we need to invert the problem by accounting for the noise. The first thing in attempting to find an estimate for X is to choose a criteria that will help discriminate viable solutions. This is usually done by looking for solutions that minimise the distance between the observations and the model, which boils down to solving equation 2.13.

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \frac{1}{2} ||Y - KX||_2^2 \tag{2.13}$$

Not only the function $F : X \mapsto \frac{1}{2}||Y - KX||_2^2$ is a direct measurement of how close a model is from the observable, but it also holds two very useful properties, namely: differentiability and convexity. Because of these properties, we know that F reaches a minimum where

$$\frac{\partial F}{\partial X}(\tilde{X}) = 0 \tag{2.14}$$

By computing the derivative of F and solving equation 2.14, we are able to derive a closed form solution for equation 2.13 known as least squares:

$$\tilde{X} = (K^T K)^{-1} K^T Y \tag{2.15}$$

Such solution exists if and only if the columns of K are linearly independent (condition of existence for $(K^T K)^{-1}$). Since we cannot always assume that this condition is met in practical applications, we cannot rely on such solution unless by using a regularisation to lift the conditioning of $(K^T K)$.

Gradient descent   It is also possible to use the properties of differentiability and convexity of F to approximate a solution for equation 2.13 in an iterative scheme. Since F is convex, it is possible, from any point X_i, to find X_{i+1} such that $F(X_i) > F(X_{i+1})$, by following the direction of negative gradients around X_i. In other words, repeating iteration 2.16 will see variable X_i converge towards an estimate of X that minimises F. With this scheme, the gradient step μ gives a measure of how much we are moving in the direction of decreasing gradients. Therefore, if μ is too small, the algorithm will only converge after a very large number of iterations. If μ is too large, the algorithm will never find the minimum and diverge. In practice, μ will depend on the steepness of the gradient and must therefore satisfy $0 < \mu < 2/||K^T K||_S$, where $||K^T K||_S$

is the Lipschitz constant of the gradient.

$$X_{i+1} = X_i - \mu \nabla F(X) = X_i + \mu K^T (Y - KX_i) \tag{2.16}$$

While this is a perfectly acceptable solution, physicists in particular might need to impose constraints that have different motivation such as the ones illustrated in figure 2.1. These constraints are generally called regularisations. In the following section we are going to see how regularisations in the form of convex functions may be applied to general ill-posed problems.

The result of applying the gradient descent strategy to solve equation 2.9 is shown in figure 2.4b.

2.2   Regularisation

Building upon the previous notions of differentiability and convexity of a function, it is important to notice that a sum of convex functions over a given domain is a convex function on the same domain. This property allows us to impose further constraints, or regularisations, on a solution as long as this constraints can be expressed as convex functions. Regularisation is particularly useful in problems that are highly ill-posed ($rk(A|X) \ll m$) and therefore subject to degeneracies.

In this section, we will consider regularisation from using convex constraints, for which solving a regularised problem, comes down to minimising a sum of convex cost functions $f_k : \mathbb{R}^m \to \mathbb{R}$ (see equation 2.17) that each penalises solutions that are deemed physically implausible.

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \sum_k f_k \tag{2.17}$$

We have seen a very simple example of such minimisation problem in the previous section, where the only function to minimise was F and had the convenient property of being differentiable everywhere in $\mathbb{R}^m$. In cases where all function f_k are differentiable, the least square, or the gradient descent strategies are appropriate as shown in the following case of the ℓ_2 constraint. A common physical assumption is that the signal we aim to recover be of low amplitude. This is expressed, using the previous notations, in the following minimisation problem:

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \frac{1}{2} ||Y - KX||_2^2 + \lambda ||X||_2^2. \tag{2.18}$$

In this case, the minimum of the convex function $F_2 : X \mapsto \frac{1}{2}||Y - KX||_2^2 + \lambda ||X||_2^2$ can be computed by nullifying its derivative and solving in X, which gives the solution:

$$\tilde{X} = (K^T K + \lambda I)^{-1} K^T Y \tag{2.19}$$

This solution, called the ridge regression, has the advantage to lift the ill conditioning of matrix

$K^T K$ by λ, compared to the least square case and it is possible to find λ large enough such that $(K^T K + \lambda I)$ exists.

In a more general framework, many useful constraints are not differentiable. In these cases, we need to rely on proximal algorithms (Combettes & Pesquet, 2009).

2.2.1 Basics of proximal calculus

Notations and definitions

- Convex set: a non-empty set $\mathscr{C}$ is a convex set if:

$$\forall (x, y) \in \mathscr{C}^2, \forall \lambda \in [0, 1], \lambda x + (1 - \lambda y) \in \mathscr{C} \tag{2.20}$$

- Convex function: a function $f : \mathscr{C} \to \mathbb{R}$ is convex if $\mathscr{C}$ is a non-empty, convex set and if:

$$\forall (x, y) \in \mathscr{C}^2, \quad \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{2.21}$$

- Lower semi-continuity: a function $f : \mathbb{R}^m \to \mathbb{R}$ is lower semi-continuous if $\forall \alpha \in \mathbb{R}, \{x \in \mathbb{R}^m, f(x) < \alpha\}$ is a closed set. In the following, we call Γ_0, the class of lower semi-continuous convex functions from $\mathbb{R}^m$ to $\mathbb{R}$.

Proximal operators

Proximal calculus relies on a set of functions called proximal operators. The proximal operator of a function $f \in \Gamma_0$ is defined by:

$$prox_f : \mathbb{R}^m \to \mathbb{R}^m : x \mapsto \underset{y \in \mathbb{R}^m}{argmin} \quad f(y) + \frac{1}{2} ||x - y||^2, \tag{2.22}$$

and reaches an infimum at a unique point x_f (Moreau, 1962, 1965). Proximal operators can be seen as a local minimisation around a point x, of function f. This interpretation becomes particularly clear and useful when faced with the problem of constraining a solution to a convex set $\mathscr{C}$. This constraint is expressed by the indicator function, $i_{\mathscr{C}} : \mathbb{R}^m \to \mathbb{R}$, defined by:

$$i_{\mathscr{C}}(X) = \begin{cases} 0 & \text{if } X \in \mathscr{C} \\ +\infty & \text{otherwise} \end{cases}, \tag{2.23}$$

and its proximal operator is simply the orthogonal projection of x onto $\mathscr{C}$. This allows us to design an intuitive algorithm for solving the simple problem:

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{argmin} \quad \frac{1}{2} ||Y - KX||_2^2 + g(X), \tag{2.24}$$

with $g \in \Gamma_0$. We have seen previously how gradient steps converge towards a minimum of $||Y - KX||_2^2$. Because proximal operators are local minimisations of a function, we can devise the strategy of alternating between a step of gradient descent (forward) and a step of projection on $g(.)$ (backward) in order to minimise g locally. This strategy is called *forward-backward* algorithm which is formally introduced in (Combettes & Wajs, 2005) where the authors prove that iteration:

$$X_{i+1} = prox_{g(.)}(X_i - \mu \nabla F(X_i)) \tag{2.25}$$

converges towards a unique solution.

We applied the *forward-backward* algorithm to problem 2.9 with a positivity constraint: $g(\cdot) = i_{\mathbb{R}^{m+}}$, which asserts that all pixels in the unknown image X are positive. Because $\mathbb{R}^{m+}$ is a convex subset of $\mathbb{R}^m$, the proximity operator of $i_{\mathbb{R}^{m+}}$ is simply the orthogonal projection on $\mathbb{R}^{m+}$, which consists in setting to zero all negative coefficients in X. The result is shown in figure 2.4c. When comparing with the result of the gradient descent in figure 2.4b, we see that the positivity constraint sharpens the contour of the recovered image and removes the negative wriggles introduced by gradient descent. At this point it is important to notice that the specific positivity constraint we applied here was motivated by the fact that we knew the original image contained only positive pixels. In real applications, regularisations should always be motivated by a physical assumption we have about the signal to recover.

In many cases, the proximal operator of a convex function can be computed analytically as we have seen with the case of the positivity constraint and more generally, with projections onto convex sets. The explicit proximal operators of various functions can be found in Combettes & Pesquet (2009), (Table 10.2). When no closed form exists for the proximal operator of a function, we need to solve numerically the minimisation problem in equation 2.22, which can involve timely iterative steps, as we will se in section 2.2.3. Another approach consists in using the properties of proximal operators (see Combettes & Pesquet, 2009, table 10.1) to build algorithm that will efficiently compute the proximal operator of a function while solving the global minimisation problem (Chambolle & Pock, 2011; Vũ, 2013). In the following sections we will focus on a specific constraint coined as sparsity, which is well suited to many astrophysical applications and that will be widely used in the rest of this dissertation.

2.2.2 Sparsity

The concept of sparsity in signal processing stems from the realisation that a signal can be concentrated in a very small number of coefficients in an adequate transformed domain. A simple example is given by periodic signals, which, in the time domain, takes many non-zero values. In the Fourier domain, the signal is concentrated in as few coefficients as the number of frequencies they contain, under Parseval's theorem. The notion of sparsity in a transformed domain has given rise to the field of compressed sensing (Donoho, 2006; Candes et al., 2006; Candes & Tao, 2006), where it was shown that the knowledge of a sparse decomposition for a

(a) Target measurement

(b) Reconstruction from gradient descent

(c) Regularised solution
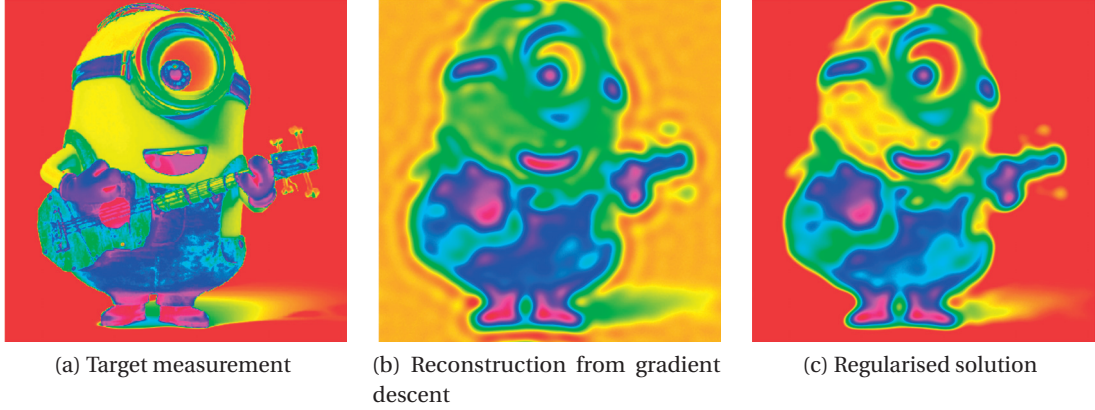
Figure 2.4 – Illustration of the convolution of an image (*left panel)* by a kernel (*middle panel)* to obtain a blurred image (*right panel)*

signal allowed for its recovery and reconstruction beyond Shanon's sampling theory, in cases where the acquisition matrix is a random matrix. Sparsity is now a widely used regularisation for solving ill-posed linear inverse problems with strong applications in various fields of astrophysics and cosmology (Starck et al., 2015; Lanusse et al., 2016; Farrens et al., 2017; Bobin et al., 2015b; Birdi et al., 2018; Leonard et al., 2014; Bobin et al., 2013; Pratley et al., 2018; Wallis et al., 2017).

Definition

A signal in $\mathbb{R}^m$ that can be fully represented by k non-zero coefficients in a given dictionary is called k-sparse. The notion of dictionary here, includes, but is not restricted to, orthonormal basis. A generative dictionary of $\mathbb{R}^m$ is a $m \times d$ matrix, where its d columns are vectors representing elements of $\mathbb{R}^m$, that we call atoms. Examples of atoms encompass sinusoids (Fourier dictionary), Diracs (direct space), gaussians or wavelets. In the context of sparsity, overcomplete ($d > m$) dictionaries spanning $\mathbb{R}^m$ allow for a larger degrees of freedom in representing a signal, which may lead to sparser decompositions. In practice, we favour overcomplete dictionaries that admit a fast transformation operator such as wavelets, in order to avoid the high computation cost of carrying large matrix multiplications.

Since k-sparse vectors of $\mathbb{R}^m$ are the vectors that have k non-zero elements we can quantitatively measure the sparsity of a vector $X \in \mathbb{R}^m$ by computing its ℓ_0 norm (see equation 2.7). From there, a sensible strategy to impose the k-sparsity of a vector is to set to zero all but k of its coefficients to zero. As we have seen with the graphical example of figure 2.1 this strategy leaves us with the problem of finding the combination of k non-zero coefficients that match our observables best, for which their is no simple solution. Because one of the key assumption of sparsity is that the energy of the signal is concentrated in a small number of coefficients, a fair solution to this combinatorial problem is to choose a value for k and favour the k highest

coefficients by setting the others to zero. This has the advantage of minimising the amount of signal lost in the process. This operation is carried out by applying the `hard-thresholding` operator ($HT_\lambda(.)$) described by:

$$HT_{\lambda \in \mathbb{R}^+} : \mathbb{R}^m \to \mathbb{R}^m : X \mapsto HT_\lambda(X) = \begin{cases} x_i & \text{if} |x_i| \geq \lambda \\ 0 & \text{if} |x_i| < \lambda \end{cases}. \tag{2.26}$$

The scalar λ in equation 2.26 is a threshold below which coefficients in X are set to zero and above which they are left untouched. In the case of a k-sparse vector, λ is set to the value of the $k^t h$ coefficient when considering all coefficients x_i ranked in order of decreasing magnitude.

In practical applications, and in astrophysics in particular, vectors are rarely strictly k-sparse, and if they are, the number k, of non-zero coefficients is not known a priori. In this case, we need to rely on weak sparsity, which assumes that the series formed by the coefficients of a weakly sparse vector X, sorted by decreasing magnitudes $\{x_{d \in [0..m]}\}$, decreases faster than a power law, such that $|x_d| < Ad^{-\alpha}$, with $A \in \mathbb{R}^+$ and $\alpha > 2$. With such signals, the error between the full vector X and its truncation to the k highest coefficients X_k is majored by:

$$||x - x_k||_2^2 \leq \sum_{i>k}^{i=m} Ai^{-\alpha} \leq \frac{A}{2\alpha - 1} k^{-2\alpha+1}, \tag{2.27}$$

meaning that the approximation made in setting to zero the $m - k$ smallest coefficient is small. In most noise-dominated problems encountered in astrophysics, it is common to chose k such that X is reconstructed down to noise levels, which is, provided that the sparsity assumption is verified, much larger than the error from equation 2.27.

Sparsity in the frame of convex optimisation

In the context of linear inverse problem solving, imposing a constraint of sparsity as formulated above, boils down to the following minimisation problem:

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \frac{1}{2} ||Y - AX||_2^2 + \lambda ||X||_0, \tag{2.28}$$

As we discussed before, the ℓ_0 "norm" is not a convex function, and solving such a minimisation problem is combinatorial. Instead, sparsity may be imposed as a convex constraint, using the ℓ_1 norm:

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \frac{1}{2} ||Y - AX||_2^2 + \lambda ||X||_1, \tag{2.29}$$

Instead of penalising solutions with a large number of non-zero coefficients, the ℓ_1 norm penalises those with a large sum of the coefficients' magnitudes. It was shown in Donoho & Huo (2001) that for highly sparse signals and without noise, the ℓ_1 minimisation problem leads to the same solution as the ℓ_0 problem. This is illustrated graphically by the superimposition

of the green and blue stars in our simplistic study case of figure 2.1.

Function $g_1 : \mathbb{R}^m \to \mathbb{R} : X \mapsto \lambda||X||_1$ is convex and admits a closed form proximal operator under the form of a projection of X onto the ℓ_1 ball of radius λ, coined as *soft thresholding* ($ST_\lambda(.)$) operator:

$$prox_{g_1} : \mathbb{R}^m \to \mathbb{R}^m : X \mapsto ST_\lambda(X) = \begin{cases} sign(x_i)(|x_i| - \lambda) & \text{if}|x_i| \geq \lambda \\ 0 & \text{if}|x_i| < \lambda \end{cases}. \tag{2.30}$$

Choice of the regularisation parameter

The regularisation parameter λ, which sets the sparsity of the solution is of paramount importance in the development of efficient algorithms to solve problem 2.29. Several strategies exist to choose this parameter wisely, which depend on the exact nature of the physical problem. One strategy, as discussed earlier is to have a prior knowledge of the sparsity of the solution, either in terms of number of non-zero coefficients, or in terms of power-law decrease of the ordered coefficients. With this knowledge, it is simple to choose λ such that only k coefficients are non-zero, or such that the error on the model remains below an acceptable limit $\left(\frac{A}{2\alpha - 1}k^{-2\alpha + 1}\right)$.

In noise dominated cases, it is wiser to choose a regularisation parameter based on noise levels and ensure that enough coefficients are used to reconstruct the signal without over fitting noise features. In Donoho & Johnstone (1994), the authors proposed a threshold for denoising problems ($A = I_m$ and $m = n$, in equation 2.8), with signals affected by white Gaussian noise with standard deviation σ and represented in an orthogonal basis, where λ is chosen as $\lambda = \sigma\sqrt{2log(m)}$. A more general approach for noise-dominated linear inverse problems consists in evaluating the noise standard deviation for each coefficient x_i from the noise covariance matrix. Assuming that the noise covariance in the observable, Σ, is known (in the case of white Gaussian noise with standard deviation σ, $\Sigma = \sigma^2 I_m$). It is possible to compute the noise standard deviation for each coefficient x_i of our model by computing the vector of $\mathbb{R}^m$ made of the square root of the diagonal elements of matrix $\Sigma_X = K^T\Sigma K$. Therefore, the corresponding noise standard deviation in the i-th coefficient of X is given by $\sigma_{x_i} = \sqrt{K^T\Sigma K_{i,i}}$. The threshold of the soft thresholding operator is simply given $\lambda = t \times \sqrt{D(\Sigma_X)}$, where t is a scalar, usually taken between 3 and 5 to ensure a good balance between completeness of the reconstruction and robustness to overfitting. We recall that $D(\cdot)$ is the operator that returns the diagonal elements of a square matrix as a vector.

2.2.3 Proximal operators of sparsity in a transformed domain

In practical astrophysical applications, the sparsity assumption is usually not verified in the direct (pixel) domain. In particular galaxy, light profiles show extended features that use large

numbers of pixels to be approximated properly. It is possible, however to go to a transformed domain where galaxies are actually sparse. A very common class of functions to decompose galaxies linearly, with a small number of coefficients are wavelets, which we will expand on in the next section, 2.3. Because a galaxy light profile is not sparse in the pixel domain, but is in the wavelet domain, we are now faced with the problem of reconstructing the wavelet coefficients of a galaxy's light profile rather than its pixel values.

Analysis versus synthesis formulation

Using the formalism of equation 2.8, let's consider that vector $X \in \mathbb{R}^m$ contains the coefficients of our target measurement (for instance pixel values of a galaxy light profile reshaped as vector). $\Phi \in \mathbb{R}^{n \times m}$ is a dictionary of vectors that can be an orthonormal basis or a generative dictionary of $\mathbb{R}^n$ over which X is known to be sparse, and α_X is a vector of $\mathbb{R}^n$ such that $\alpha_X = \Phi X$. Because we chose Φ to provide sparse representation of X, α_X is known to be a sparse vector. We can now reformulate problem 2.29 in one of two ways:

Synthesis formulation: Considering the problem in α_X, sparse, with X being our target measurement, we can simply solve the problem:

$$\tilde{X} = \tilde{\alpha}_X, \quad s.t. \quad \tilde{\alpha}_X = \underset{\alpha_X \in \mathbb{R}^n}{arg\,min} \quad \frac{1}{2} ||Y - K\Phi^T \alpha_X||_2^2 + \lambda ||\alpha_X||_1. \tag{2.31}$$

Equation 2.31 is formulated as a classical problem of minimisation of a sum of two convex, proximable functions. Such problem admits a simple solution, derived form the *forward-backward* algorithm of equation 2.25.

Analysis formulation Focusing on the recovery of the target measurement itself, rather than on an abstract of the variable in a transformed domain, we can choose to solve equation 2.8, assuming the sparsity of the transfom of X, by solving:

$$\tilde{X} = \underset{X \in \mathbb{R}^m}{arg\,min} \quad \frac{1}{2} ||Y - KX||_2^2 + ||\Phi X||_1. \tag{2.32}$$

With the formulation of equation 2.32, the emphasis is put on the recovery of X itself. Despite these two formulations looking very similar, we will see in the following discussion that in some cases they represent two different approaches, and that they may lead to different solutions.

Analysis versus synthesis Despite analysis and synthesis formulations seeming equivalent at first sight, their equivalence is actually conditional to $\Phi^T \Phi = I_m$. Meaning that in the cases

where Φ is an orthogonal basis, analysis and synthesis are exactly equivalent through a simple change of variable. In cases where Φ is an overcomplete dictionary, or where its rank is smaller than m, the two formulations are fundamentally different. In analysis the aim is to find the vector X that has a projection over Φ which is sparse, while in synthesis, we reconstruct the sparsest vector in the transformed domain. Because $\Phi^T \Phi \neq I_m$, these two solutions are not equal. In particular, synthesis formulation will lead only to solutions that can be represented by vectors in the transformed domain (which is limiting in the rank deficient case), while analysis can potentially explore solutions for X in the direct domain, but will not necessarily be able to reach some of the sparse representations α_X in the transformed domain. In other words, let X_a be the solution of the analysis problem 2.32 and $\alpha_{X,s}$ be the solution of the synthesis problem 2.31, then $\Phi X_a \neq \alpha_{X,a}$ and $\Phi^T \alpha_{X,s} \neq X_a$.

This has several implications when designing an algorithm for solving inverse problem under a sparse assumption. First, we notice that in the overcomplete case ($n > m$), the synthesis formulation contains more unknowns than the analysis formulation, making analysis problems simpler to solve. Another important feature to notice is the proximal operator of the regularisation term. In the general synthesis case, where the regularisation term is a proximable function $g(.)$ that admits a closed form operator, we can directly apply this operator in a *forward-backward* strategy for instance. In the analysis case, where the regularisation term in $g(\Phi.)$ does not generally admit a closed form proximal operator, even if g does. In cases where Φ is an orthonormal basis, the proximal operator $g(\Phi.)$ is known and writes as $W^T prox_g(.)$. In other cases, a closed form might not exist, and applying the *forward-backward* strategy, would require to numerically estimate $prox_{g(\Phi.)}(X)$ at every iteration.

The properties of proximal operators combined with primal-dual splitting make it possible to design efficient algorithms to solve analysis formulated problems without relying on subiterations. A few examples of such algorithms can be found in: *Chambolle-Pock* (Chambolle & Pock, 2011), *Condat-Vu* for problems with a third function to minimise (Condat, 2013; Vũ, 2013) or Alternating Direction Method of Multipliers (Gabay & Mercier, 1976). In Combettes & Pesquet (2009), the authors give a detailed overview of these algorithms.

2.3 Wavelet Transform

In the previous section, we have seen how the sparsity constraint could be imposed in transformed domains. This suggests that for a given signal, one might be able to find a domain such that the projection of the signal onto this domain be sparse. For instance, a sinusoid takes non-zero values over the whole domain $\mathbb{R}$, but in the Fourier domain, it is represented by only one non-zero coefficient. In astronomy and cosmology, galaxy light profiles are generally not sparse in the direct domain, and it takes many pixels to fully account for it surface brightness. However, galaxies usually present smooth variation of their luminosity across an image. The goal of this section is to detail a framework that allows sparse decomposition of piecewise smooth signals: wavelets. As we will see, this framework allows to perform decompositions at

various scales of a signal over a wide range of wavelet functions. The choice of the wavelet function allows to favour sparse representation of a specific features, making it a powerful tool for sparse optimisation.

Wavelet decomposition can be seen with many regards as an extension of Fourier analysis. Where Fourier analysis provides a global decomposition of a signal into sinusoids at various scales, the concept behind wavelets is to provide a local analysis of a signal at various scales with functions that are not restricted to sinusoids. The goal of such approach is many folds. In the analysis of time series, it allows to visualise the evolution of a signal's frequency as a function of time. This was recently illustrated with the first discovery of a gravitational wave event (Abbott et al., 2016), where the authors used a multiscale analysis (Chatterji et al., 2004) of the detected signal to characterise the event. In image analysis, wavelets provide tunable analysis functions that efficiently represent piece-wise smooth signals, thus providing adequate domains for sparse representation of images, as we will see across this dissertation, with strong implications with regard to image compression, as illustrated by the JPEG2000 format, which relies on discrete bi-orthogonal wavelets. Now wavelets find application beyond 1 and 2-D signals and provide sparse representations of signals on the sphere (Starck et al., 2006; McEwen et al., 2007; Chan et al., 2017) and in 3 dimensions (Starck & Murtagh, 2006; Woiselle et al., 2010; Lanusse et al., 2012; Leonard et al., 2014).

The notion of local scale-dependent analysis can be traced back to the work of Gabor (1946). The general idea of Gabor (1946) is to perform a Fourier analysis of a signal $s(t)$ at various locations τ of the time domain. This is done by performing the Fourier analysis of $s(t)$ multiplied by a Gaussian window centred in τ, which formally writes:

$$G_s(\tau, \nu) = \int_{-\infty}^{+\infty} s(t) e^{-\pi(t-\tau)^2} e^{-j\nu t} dt, \tag{2.33}$$

Where ν is the frequency. This corresponds to a decomposition of a signal into sinusoids at frequency ν attenuated by a Gaussian function centred in τ. The elements of this decomposition are represented graphically in the real and imaginary domain at several frequencies in figure 2.5.

From figure 2.5, we can see that the elementary function that is used to analyse a signal is shaped as a "small wave". With the works of (Grossmann & Morlet, 1984), a formal framework is introduced that defines generic transformations based on families of functions showing "small wave" patterns, henceforth called wavelets (actually extrapolated from the french for ondelettes). In Goupillaud et al. (1984), the authors even introduce a formal wavelet family called Morlet's wavelet, that builds upon Gabor's analysis. In the following lines we will see the formal definition of wavelets, and, in particular, the conditions under which a function can be used as a wavelet, as well as the techniques to efficiently compute wavelet transforms.

(a) Real part   (b) Imaginary part
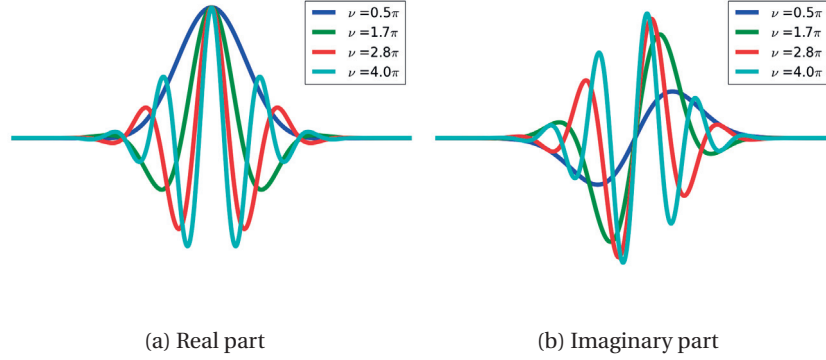
Figure 2.5 – Gabor elements in real (*left*) and imaginary (*right*) domains.

2.3.1 Continuous wavelet transform

Definition

Building upon the notions of Fourier and Gabor Analysis, the wavelet transform of a signal in time domain $s(t)$ can be seen as an inner product between s and an analysis function Ψ. To begin with, we will consider $s : \mathbb{R} \to \mathbb{R}$, a one-dimensional function of $L_2(\mathbb{R})$. The wavelet transform of s at location $\tau \in \mathbb{R}$, and scale $a \in \mathbb{R}^+$ is given in Grossmann & Morlet (1984), equations 1.6, and writes as:

$$W_s(a, \tau) = \langle s | \Psi_{a,\tau} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \Psi^* \left(\frac{t - \tau}{a} \right) dt, \tag{2.34}$$

and its inverse (Grossmann & Morlet, 1984, equation 1.9), provided that it exists, is defined as:

$$s(t) = \frac{1}{C_\Psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_s(a, \tau) \Psi_{a,\tau}(t) \frac{da}{a^2} d\tau. \tag{2.35}$$

where C_Ψ is defined by:

$$C_\Psi = \int_0^{+\infty} \frac{|\hat{\Psi}(\nu)|^2}{\nu} d\nu. \tag{2.36}$$

From equation 2.34, we see that each wavelet coefficient $W_s(a, \tau)$ is computed from a shifted and scaled version of the complex conjugate of Ψ. Function Ψ is called the mother wavelet and each wavelet coefficient is the inner product between s and a daughter wavelet: $\Psi_{a,\tau} = \frac{1}{\sqrt{a}} \Psi \left(\frac{t - \tau}{a} \right)$. As pointed out in Grossmann & Morlet (1984), the existence of the inverse of a wavelet transform is conditional to the admissibility condition of function Ψ, which imposes:

$$C_\Psi < +\infty. \tag{2.37}$$
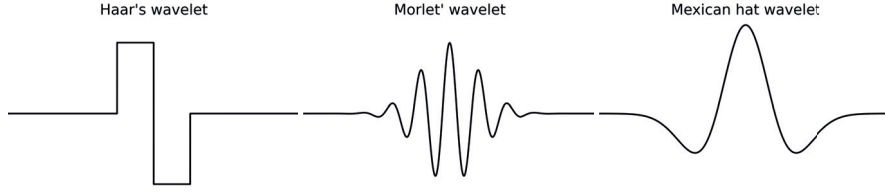
Figure 2.6 – Profile of mother wavelets. From left to right: Haar's, Morlet's (real part only) and the Mexican hat.

A necessary condition for equation 2.37 to be verified is that $|\hat{\Psi}(0)| = 0$, which means that wavelet functions must have a zero mean. Another observation we make from equation 2.34 is that at a given scale $a \in \mathbb{R}^+$, the operation:

$$\mathbb{R} \to \mathbb{R} : \tau \mapsto \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \Psi^* \left(\frac{t-\tau}{a} \right) dt \tag{2.38}$$

defines a convolution product. This means that all the wavelet coefficients at a given scale a can be computed from the convolution of s by a function $\bar{\Psi}_a$ defined as:

$$\forall a \in \mathbb{R}, \bar{\Psi}_a : \mathbb{R} \to \mathbb{R} : t \mapsto \frac{1}{\sqrt{a}} \Psi^* \left(\frac{-t}{a} \right). \tag{2.39}$$

This formulation provides an efficient way of computing wavelet transforms, by noticing that convolution is a product in the Fourier domain. The goal of wavelets being to represent signals with adequate analysis function, mainly for compression and sparse application purposes, many functions have been developed across the past century that emphasize various aspects of natural signals. For that reason, it is paramount that the wavelet we chose to analyse a given signal be well suited to extract the desired features of the signal. To give the reader a better feeling of what wavelets look like and of what features they can represent, we give, in the following paragraphs a few examples of common wavelet functions.

Examples of wavelet functions

In the following paragraphs, we show the explicit mother wavelet of three historical wavelet decompositions and illustrate their characteristic with the decomposition of a piece-wise smooth function in figures 2.7 and 2.8. The profile of their mother wavelets are shown in figure 2.6.

Haar wavelet   The first historical example of a wavelet function dates back from the early twentieth century with the work of Haar (1910), where the author describes an orthonormal

basis of $L_2(\mathbb{R})$ based on the function:

$$\Psi_h : \mathbb{R} \to \mathbb{R} : t \mapsto \begin{cases} 1 & \text{if} \quad 0 \le t < 0.5 \\ -1 & \text{if} \quad 0.5 \le t < 1 \, . \\ 0 & \text{otherwise} \end{cases} \tag{2.40}$$

By noticing that the set of discrete scaled and translated versions of Ψ_h were orthogonal from one-another, Haar constructed a basis of $L_2(\mathbb{R})$ that would later become the first example of a wavelet function. Because the Haar's wavelets is a combination of step functions, it is well suited to representing sharp edges and piece-wise constant signals, but is not adapted to the decomposition of smoothly varying signals.

Morlet wavelet   As stated earlier, the Morlet wavelet (Goupillaud et al., 1984) is inspired by the Gabor function, and the wavelet transform with Morlet's mother wavelet only differs from the Gabor transform by the scaling factor, such that the wavelet coefficients of the Morlet transform of a signal $s(t)$ are given by:

$$W_{a,\tau} = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{+\infty} s(t) e^{-\frac{(t-\tau)^2}{2a^2} + 2j\pi \frac{(t-\tau)}{a}} \, dt. \tag{2.41}$$

Unlike Haar's transform that has a compact support, Morlet's transform relies on integrating a function that has non-zero values over $\mathbb{R}$. In practice, for t sufficiently large, the mother wavelet reaches negligible values, making it possible to approximate the mother wavelet by its truncated version. Given the oscillatory nature of the mother wavelet, Morlet's transform performs very well at discriminating locally between modes of periodic signals.

Mexican hat wavelet   The Mexican hat wavelet was introduced in Murenzi (1989) and uses, as a mother wavelet, the second derivative of a gaussian, up to a minus sign:

$$\Psi(t) = (1 - t^2) e^{-\frac{t^2}{2}} \tag{2.42}$$

This wavelet family is best suited to represent blobs. Mexican hat is somewhat in between the previous wavelet families in the sense that they able to represent smooth signals well and do not require too many coefficients to represent discontinuities, as illustrated with the top panel of figure 2.8.

Figure 2.7 shows the scalograms of each wavelet decomposition of a signal shown in the top panel. Each line of the scalogram shows a different scale of the associated wavelet decomposition. The grayscale indicates the magnitude of the wavelet coefficient at the corresponding scale. The scalograms were generated using the previously discussed strategy of multiplying the signal by the rescaled mother wavelet in the Fourier domain and reverting the result back to direct domain. Figure 2.8 shows the magnitudes of specific scales (lines in the scalogram).

Signal

Haar scalogram

Morlet scalogram

Mexican hat scalogram
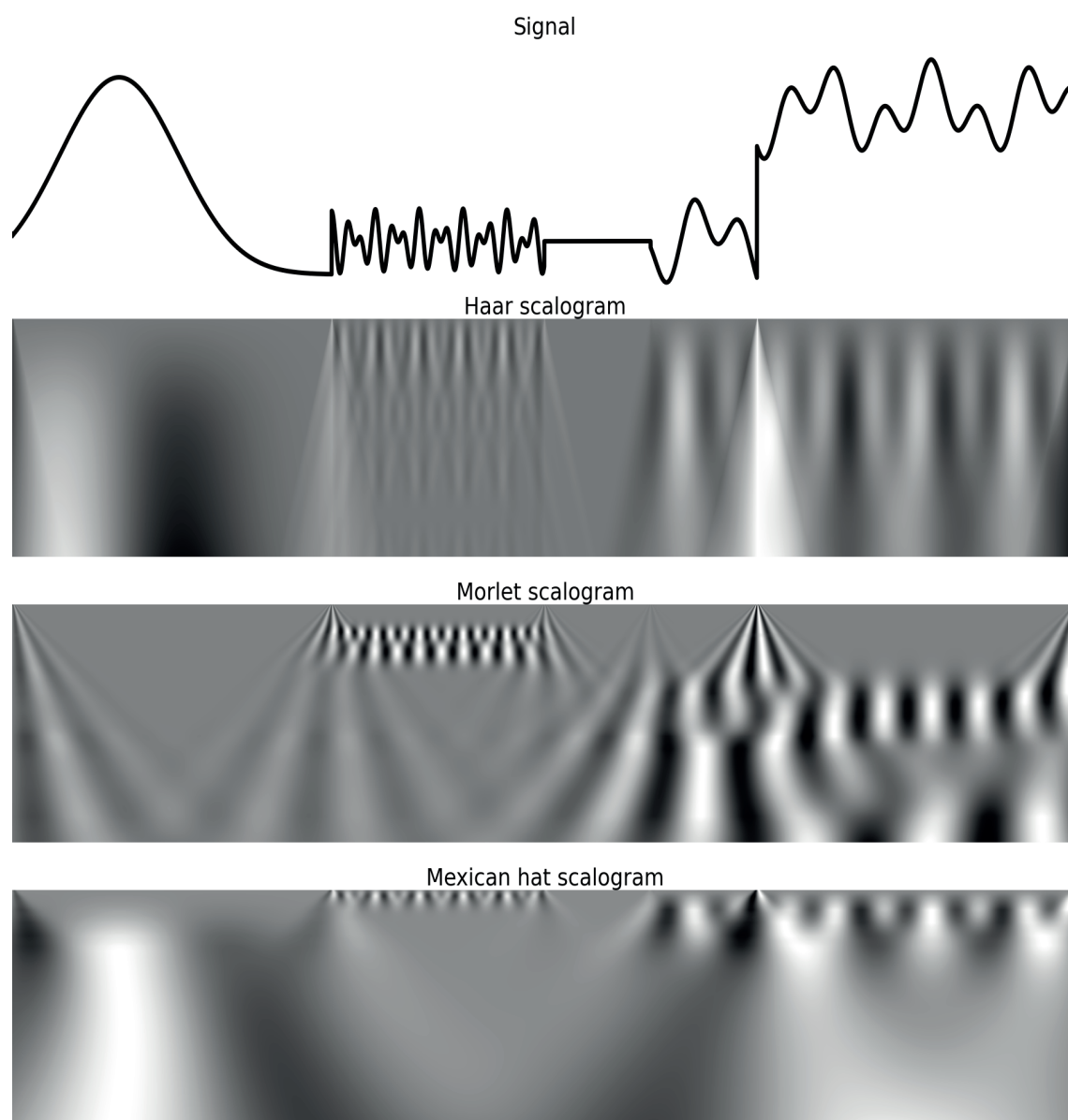
Figure 2.7 – Scalograms of a piecewise smooth signal (*top panel*) with, from top to bottom, Haar, Morlet and Mexican hat wavelets.

The magnitudes of the wavelet coefficients are normalised to the maximum over a given scale for better graphical representation.

From the scalograms, we see that Morlet's transform allows to discriminate between various modes of a locally periodic signal, but creates oscillation, around the discontinuities that are of the same nature as oscillations one would observe when reconstructing a step function with sinusoids. This is emphasised when looking at the specific scales we chose to represent in figure 2.8. Haar's wavelet, on the contrary, uses only one coefficient at a given scale to

represent discontinuities. The Mexican hat wavelet show a concentration of coefficient around the Gaussian feature on the left hand side of the signal's panel, illustrating its efficiency at representing blobs. We also see on the bottom panel of figure 2.8 that Mexican hat wavelets captures the large scale variations of the signal.
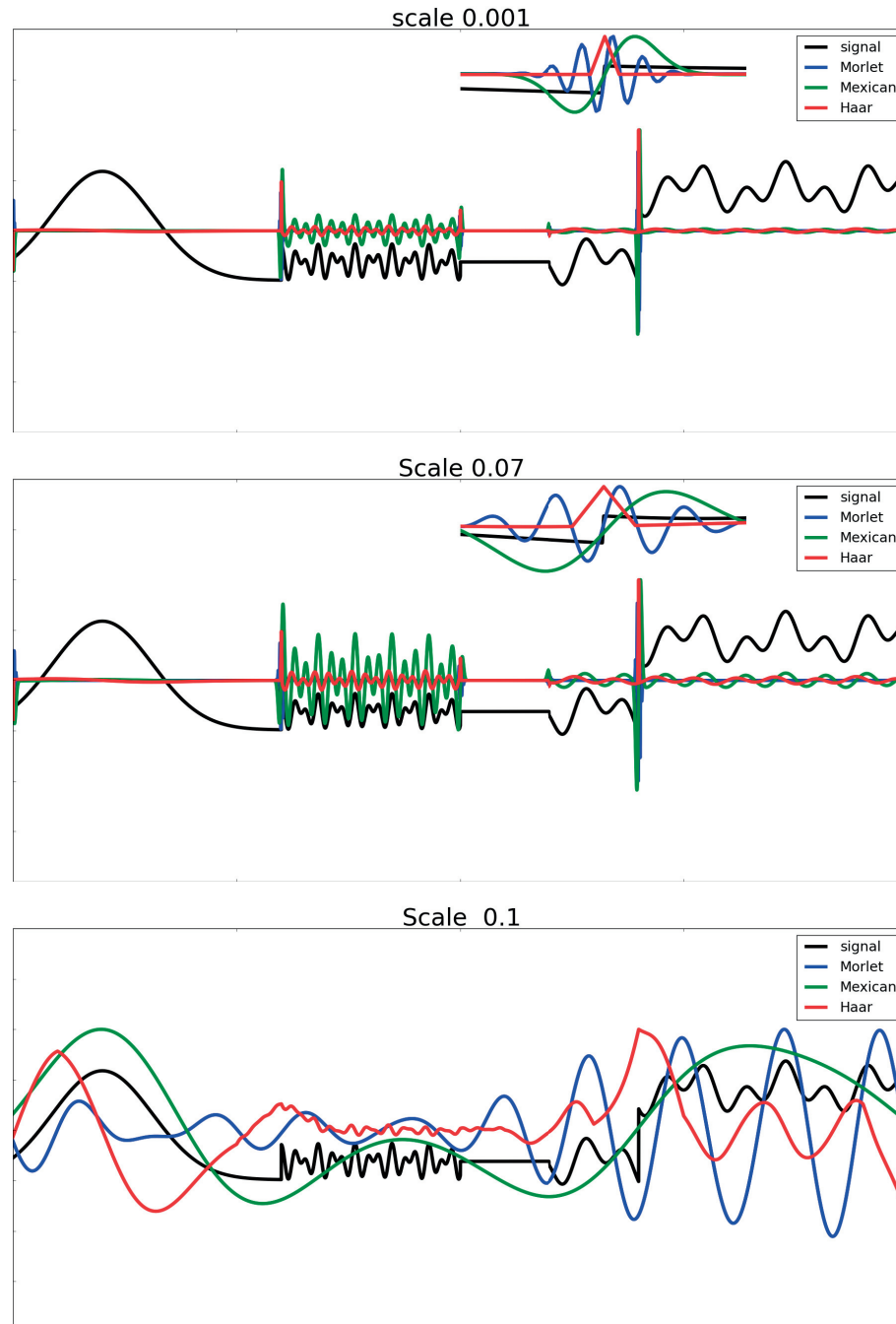
Figure 2.8 – Wavelet decomposition of a signal (*in black*) at various scales using Haar's (*in red*), Morlet's (*in blue*) and Mexican hat (*in green*) wavelets.

2.3.2 Discrete wavelet transform

Multiresolution analysis

In practical signal processing applications, signals are rarely continuous, but instead, are sampled on a grid. The definition of the wavelet transform as convolution product remains valid for discrete signals and, under careful considerations with regard to the sampling of the signal and the wavelet function, it is possible to apply a discretised version of equation 2.34. However, large datasets and higher [than one] dimensional signals require fast and invertible computation algorithms. To this end, S. Mallat developed the multiresolution analysis framework (Mallat, 1989), that allows to see wavelets as the difference between approximations of a function at two consecutive resolutions. The aim of the following paragraphs is to clarify and formalise this previous sentence.

Let f be a function of $L_2(\mathbb{R})$. Then, $f(2t)$ is an approximation of $f(t)$ at scale 2, $f(4t)$ is an approximation of $f(t)$ at scale 4 and we see that the values taken by $f(4t)$ are included among the values taken by $f(2t)$, which are included among the values taken by $f(t)$. Generalising this to all scales, we find that, $\forall j \in \mathbb{Z}$, approximation at scale 2^{j+1} of a function is included in its approximation at scale 2^j, thus defining a series of subsets, V_j, of $L_2(\mathbb{R})$ such that V_j contains the approximation of all functions of $L_2(\mathbb{R})$ at scale 2^j, i.e. $f(2^j) \in V_j$, and verifies: $\forall j \in \mathbb{Z}, \quad \{0\} \subset V_{j+1} \subset V_j \subset L_2(\mathbb{R})$.

The set of $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution approximation of $L_2(\mathbb{R})$. Each element $v_j \in V_j$ is a subsampled version of a function in $L_2(\mathbb{R})$ at resolution 2^j. Each function of $L_2(\mathbb{R})$ can be approximated as closely as desired by increasing the resolution infinitely, i.e.

$$\lim_{j \to -\infty} V_j = \underset{j\in\mathbb{Z}}{\cup} V_j = L_2(\mathbb{R}). \tag{2.43}$$

Because of the inclusion of V_{j+1} in V_j, it is possible, for all j, to find the orthogonal complement W_{j+1} to V_{j+1} that verifies $V_{j+1} \oplus W_{j+1} = V_j$. Therefore, for a given approximation of f at scale 2^{j+1}, it is possible to find $w_{j+1} \in W_{j+1}$, orthogonal to $f(2^{j+1})$ such that $f(2^{j+1}) + w_{j+1} = f(2^j)$. The complement w_{j+1} to $f(2^{j+1})$ contains the details of f at scale 2^j.

An important result relative to subspaces $V_{j\ j\in\mathbb{Z}}$ and $W_{j\ j\in\mathbb{Z}}$ from Mallat (1989), theorems 1 and 3, is that there exists two functions, $\Phi, \Psi \in L_2(\mathbb{R})^2$, such that:

$$\forall j \in \mathbb{Z}, \left\{ \Phi_{j,n} = \frac{1}{\sqrt{2^j}} \Phi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{n\in\mathbb{Z}} \tag{2.44}$$

is a basis of V_j and

$$\forall j \in \mathbb{Z}, \left\{ \Psi_{j,n} = \frac{1}{\sqrt{2^j}} \Psi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{n\in\mathbb{Z}} \tag{2.45}$$

is a basis of W_j. Following Mallat (1989)'s nomenclature, function Φ is called the scaling function of the multiresolution approximation and Ψ is an orthogonal wavelet function. Because of equation 2.43, we know that $\underset{j \in \mathbb{Z}}{\cup} V_j = L_2(\mathbb{R})$. As $\{\Phi_{j,n}\}_{n \in \mathbb{Z}}$ is an orthogonal basis of V_j, it follows that the family $\{\Phi_{j,n}\}_{n,j \in \mathbb{Z}^2}$ is a orthogonal basis of $L_2(\mathbb{R})$. By noticing that

$$V_{j-1} = V_j \oplus W_j = V_{j+1} \underset{i \in \mathbb{N}^*}{\oplus} W_{j+i}, \tag{2.46}$$

and by going to the limit $j \to \infty$, it follows that the set $\{W_j\}_{j \in \mathbb{Z}}$ is dense in $L_2(\mathbb{R})$. As a result, $\{\Psi_{j,n}\}_{n,j \in \mathbb{Z}^2}$ is also an orthogonal basis of $L_2(\mathbb{R})$.

Fast computation of orthogonal wavelets

Now that the multiresolution analysis framework has been introduced, we are going to see how it allows us to perform fast, invertible computations of wavelet transforms. From the previous realisation that $\{\Phi_{j,n}\}_{n \in \mathbb{Z}}$ and $\{\Psi_{j,n}\}_{n \in \mathbb{Z}}$ are respectively orthogonal basis of V_j and W_j, it follows that the approximation $f_j \in V_j$ of a function $f \in L_2(\mathbb{R})$ can be decomposed over these two basis as:

$$f_j = \sum_{n \in \mathbb{Z}} c_j[n]\Phi_{j,n} \tag{2.47}$$

and

$$f_j = \sum_{n \in \mathbb{Z}} d_j[n]\Psi_{j,n}, \tag{2.48}$$

where $c_j[n]$ and $d_j[n]$ are respectively the inner products of f with vectors $\Phi_{j,n}$ and $\Psi_{j,n}$. by definition, since $V_j \oplus W_j = V_{j-1}$, both V_j and W_j are in $V_j - 1$, and therefore every element of their bases are in V_{j-1}, resulting in:

$$\forall (j,n) \in \mathbb{Z}^2, (\Psi_{j,n}, \Phi_{j,n}) \in V_{j-1}. \tag{2.49}$$

As such, elements of basis of V_j and W_j can be decomposed on V_{j+1}, using equation 2.47, which raises equations:

$$\Phi_{j,n} = \sum_{m \in \mathbb{Z}} h_j[m]\Phi_{j-1,m} \quad and \tag{2.50}$$

$$\Psi_{j,n} = \sum_{m \in \mathbb{Z}} g_j[m]\Phi_{j-1,m}, \tag{2.51}$$

where:

$$h_j[n] = \langle \Phi_{j,n} | \Phi_{j-1,n} \rangle = h_1[j - 2n] \quad and \tag{2.52}$$

$$g_j[n] = \langle \Psi_{j,n} | \Phi_{j-1,n} \rangle = g_1[j - 2n]. \tag{2.53}$$

In these decompositions, h_1 and g_1 can be seen as discrete filters, that are fully determined by Φ and Ψ.

By applying the inner product by f to both sides of equations 2.50 and 2.51, respectively, we have:

$$\langle f|\Phi_{j,n}\rangle = \sum_{m\in\mathbb{Z}} h_1[m-2n]\langle f|\Phi_{j-1,m}\rangle, \tag{2.54}$$

$$\langle f|\Psi_{j,n}\rangle = \sum_{m\in\mathbb{Z}} g_1[m-2n]\langle f|\Phi_{j-1,m}\rangle. \tag{2.55}$$

Finally, by using notations from equations 2.47 and 2.48 in equations 2.54 and 2.55, we have:

$$c_j[n] = \sum_{m\in\mathbb{Z}} h_1[m-2n]c_{j-1}[m], \quad and \tag{2.56}$$

$$d_j[n] = \sum_{m\in\mathbb{Z}} g_1[m-2n]c_{j-1}[m]. \tag{2.57}$$

These two relations tell us that wavelet coefficients can be computed by applying a discrete filter to approximation coefficients c_{j-1}, which can, in turn be derived by applying another discrete filter to approximation coefficients c_{j-2}. We see by recursion, that wavelet coefficients can in fine be computed simply from any approximation coefficient c_i and from filters h_1 and g_1, which only depend on the scaling and wavelet function. The right hand side of equations 2.56 and 2.57 can be seen as a discrete convolution product by a version of h_1 down-sampled by a factor 2. Therefore, equations 2.56 and 2.57 can be rewritten as:

$$c_j[n] = [\bar{h}_1 * c_{j-1}]\downarrow_2, \quad and \tag{2.58}$$

$$d_j[n] = [\bar{h}_1 * c_{j-1}]\downarrow_2. \tag{2.59}$$

In signal processing application, signals are measured by taking regular samples of a continuous natural signal such as electromagnetic wave or acoustic wave. Such sampling constitutes in itself an approximation of a continuous function at a given scale, given by the sampling rate. As a result, we can choose $j = 0$ to be the sampling of the signal's measurement and therefore, the measured signal becomes c_0 in equation 2.56. Following equations 2.56 and 2.57, we can construct a diagram that shows the derivation of wavelet coefficients as shown in figure 2.9. Figure 2.9 also serves as an elementary block that can be chained up in order to compute wavelet coefficients to any scale j. This strategy describes the fast pyramidal algorithm (see Starck et al., 2015). Because physical, measured signal are limited in bandwidth, the number of wavelet scales that can possibly be computed is limited. This is easily illustrated by the fact that the scaling coefficients are down-sampled by a factor 2 at each application of the block from figure 2.9. Therefore, the number of scales becomes large enough so that the scale of the

scaling function 2^j becomes larger than the number of samples in c_0, which puts an upper limit on j. In order for the wavelet representation to give a complete description of c_0, we must therefore include the complementary to all detail coefficients, i.e. the scale approximation coefficient at scale $J = log_2(n)$, with n, the number of samples in c_0.
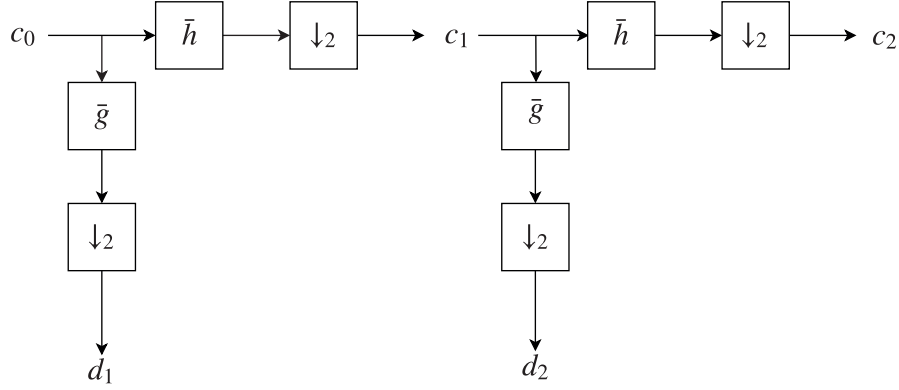
Figure 2.9 – Schematic view of a fast orthogonal wavelet transform.

The wavelet transform of a signal c_0 is therefore the set $\{d_0, d_1, \cdots, d_{log_2(n)}, c_J\}$ of all detail coefficients along with the approximation coefficients, which we call, the coarse scale. We have shown that fast computation of the discrete wavelet transform could be achieved by building a filter bank, from the scaling and wavelet functions, and by iteratively filtering and downsampling the approximation coefficients. This strategy presents the advantage to avoid computing inner products of a signal by every single translated and rescaled version of the mother wavelet.

Exact inverse transform

In the orthogonal case, approximation coefficients at scale 2^j can be reconstructed from detail and approximation coefficients at scale 2^{j+1} from:

$$c_j[n] \quad = \quad \sum_{m \in \mathbb{Z}} h[m+2n]c_{j+1}[m] + g[m+2n]d_{j+1}[m] \tag{2.60}$$

$$= \quad [c_{j+1}]_{\uparrow_2} * h[n] + [d_{j+1}]_{\uparrow_2} * g[n] \tag{2.61}$$

It was shown in Vetterli (1986) that an exact recovery could also be achieved in the case of bi-orthogonal decompositions that admit a filter bank $(h_1, g_1, \tilde{h}_1, \tilde{g}_1)$ that verifies the following

conditions in the Fourier domain:

$$\hat{h}_1^*\left(\nu + \frac{1}{2}\right)\hat{\tilde{h}}_1(\nu) + \hat{g}_1^*\left(\nu + \frac{1}{2}\right)\hat{\tilde{g}}_1(\nu) = 0 \tag{2.62}$$

and

$$\hat{h}_1^*(\nu)\hat{\tilde{h}}_1(\nu) + \hat{g}_1^*(\nu)\hat{\tilde{g}}_1(\nu) = 1, \tag{2.63}$$

where $\tilde{h}_1$ and $\tilde{g}_1$ are dual to h_1 and g_1. In this case, for bi-orthogonal wavelet bases, the exact recovery of the approximation of a function at scale j from wavelet coefficients w_{j+1} and approximation c_{j+1} is given by:

$$c_j[n] = [c_{j+1}]_{\uparrow_2} * \tilde{h}[n] + [d_{j+1}]_{\uparrow_2} * \tilde{g}[n]. \tag{2.64}$$

Cascading equation 2.64 across all scales allows to reconstruct c_0 from the set $\{d_0, d_1, \cdots, d_{log_2(n)}, c_J\}$.

A word on orthogonal wavelets

From the previous definition of orthogonal wavelets, it should appear that each detail level w_j belongs to the subset V_{j-1}. As such, each wavelet level d_j contains 2^j times less coefficients than c_0 and the large scale approximation level c_J contains $n/2^J$ coefficients. It follows that the total number of samples in the set $\{d_0, d_1, \cdots, d_J, c_J\}$ is equal to n, the number of samples in c_0. Such a decomposition is called a decimated wavelet decomposition. Decimation causes wavelet decompositions to only represent a signal 's contribution at scale 2^j in $n/2^j$ points, while one might be interested in extracting large scale information at each location. This lack of translation invariance also causes artefacts to appear when reconstructing a signal after modifying its wavelet coefficients.

To overcome this issue, the continuous wavelet transform in its discretised version provides a viable but time consuming alternative, due to the lack of a fast transform algorithm. In the following section, we will describe a class of redundant wavelet transform that admits fast computation algorithms along with inverse transforms.

On an additional note, it is possible to extend the multiresolution framework to higher dimensions by using the separability of wavelet functions across directions. The wavelet decomposition in higher dimension will therefore be performed by successively applying filter banks to each dimension. To illustrate this, we show the haar 2-D decomposition of comet 67-P/Churyumov-Gerasimenko taken by the ROSETTA mission[2] in figure 2.10. It is important to notice that the wavelet decomposition has as many coefficients as there are pixels (samples) in the original image c_0 of figure 2.10a. The top left corner of image 2.10b shows the

---

[2](image credit: Comet 67P on 19 September 2014 NavCam mosaic" by ESA,Rosetta,NAVCAM, CC BY-SA IGO 3.0. Licensed under CC BY-SA 3.0-igo via Wikimedia Commons)

(a) Chury comet

(b) Haar wavelet decomposition of Chury
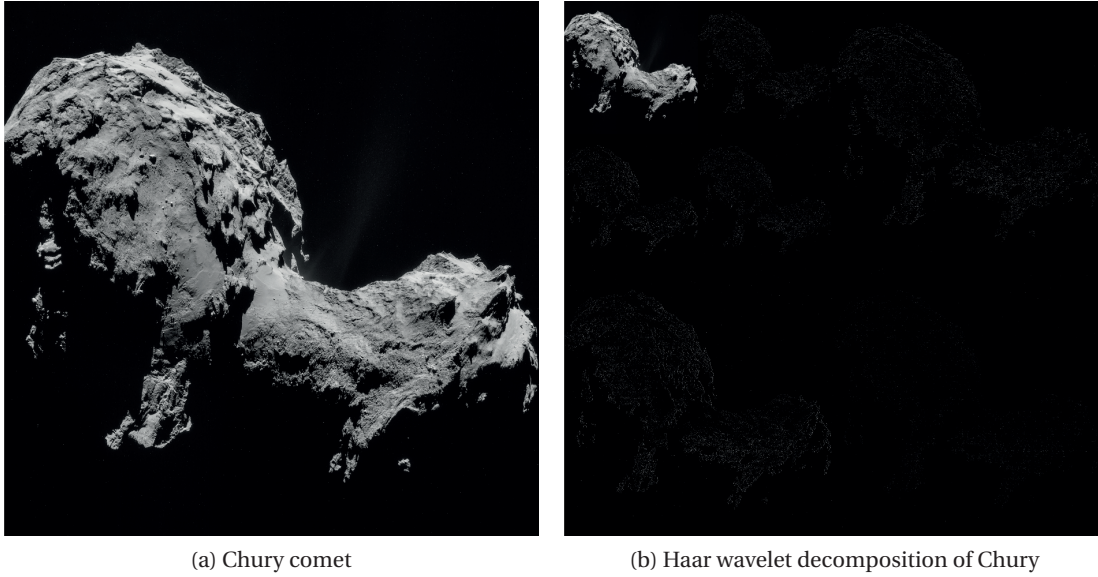
Figure 2.10 – Image of comet 67-P (left) and its decomposition with 2 dimensional Haar's wavelet (right).

last computed 2-D approximation level. The bottom right panel shows the first detail level, obtained by applying filter h_1 to c_0 in both directions. The bottom left and top right panels show respectively the detail-approximation and the approximation-detail coefficients as they are obtained by applying to c_0 respectively h_1 in the horizontal direction and g_1 in the vertical direction; and g_1 in the horizontal direction and h_1 in the vertical direction. Furthermore, on image 2.10b, we see that most of the textures that form image 2.10a are retained in the small number of coefficients constituted by the approximation scale and the details are contained in the detail coefficients.

Given that we already gave a detailed explanation of orthogonal wavelets in 1 dimension, we chose to leave it to the reader to read about their extension to higher dimensions (see Mallat, 1989; Starck et al., 2015, for more on the subject), and will address the case of 2-D transforms through the following study of isotropic redundant wavelet transform.

2.3.3 Undecimated wavelet transform

Undecimated wavelet transform (UWT) serves the purpose of representing a signal at each scale $\{2^j\}_{j \in [\![1,J]\!]}$, and at each location, by opposition to decimated wavelet transforms that represent scale 2^j in $n/2^j$ locations. This results in redundant wavelet representations that contain as many as $n \times J$ coefficients in the 1-D case.

Getting rid of the decimation when computing wavelet decompositions, imposes that we need to be able to apply filters at a given scale at every single location of the sampling. This is performed by applying a "stretched" version of the filters, such that the detail coefficients

at scale 2^j are obtained by convolving the approximation coefficients at scale 2^{j-1} by a filter h_1^j, generated from filter h_1, where $2^j - 1$ coefficients set to 0 are inserted between each coefficient of h_1. This process, dubbed the "a trous" (french for "with holes") algorithm is described in details in Holschneider et al. (1989); Shensa (1992). In (Shensa, 1992), the author showed that the coefficients from the discrete wavelet transform could be recovered from the UWT's coefficients at location $\{2^j k\}_{k \in [\![0, n/2^j]\!]}$ of the corresponding scale. The decomposition equations and inverse transform for the UWT case, thus simply become:

$$c_j[n] \quad = \bar{h}_1^j * c_{j-1}, \tag{2.65}$$

$$d_j[n] \quad = \bar{h}_1^j * c_{j-1}, \tag{2.66}$$

$$c_j[n] \quad = [c_{j+1}]_{\uparrow_2} * \tilde{h}^j[n] + [w_{j+1}]_{\uparrow_2} * \tilde{g}^j[n] \tag{2.67}$$

Another advantage of the UWT, is that because of the loss of the decimation, the de-aliasing condition from equation 2.62 does not need to be met. Instead, the only requirement put on the filter bank is the exact recovery condition from equation 2.63. The relaxation of the de-aliasing condition allows more flexibility when designing filter banks for UWT as we will see with the Starlet transform.

The Starlet transform

The Starlet transform is an undecimated isotropic wavelet transform that was developed to provide adequate decomposition for astronomical data in particular (Starck & Murtagh, 2006). A key concept to this representation is the isotropy of the decomposition. As the Universe does not have a preferred direction "a priori", it is important that decompositions of images of the Universe be carried out on analysis functions that do not have a preferred direction. For that reason, the filters must be symmetric ($h = \bar{h}$ and $g = \bar{g}$) and the wavelet and scaling functions need to be isotropic in dimensions higher than one.

The choice of the scaling function for Starlets was motivated by the need to have a function that would represent well the isotropy and the smoothness of astronomical objects and was designed to be a B-Spline of third order, defined as:

$$\Phi(x) = \frac{1}{12}(|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3). \tag{2.68}$$

Because the motivation for the development of Starlets stems from astronomical imaging, the scaling function has to be defined in two dimensions. In the following, we will write x and y, the arguments of two-dimensional functions respectively in the horizontal and vertical directions. The scaling function is designed to provide separable filters. Although this is not a

necessary condition, the separability allows to apply 1 dimensional filters to each direction rather than multi dimensional filters, which a time costly option for large dataset. The 2 dimensional scaling function for the Starlet transform is therefore defined as:

$$\Phi_{2D}(x, y) = \Phi(x)\Phi(y), \tag{2.69}$$

with this scaling function, the one dimensional discrete filter h_1 writes as:

$$h_1^0 = \frac{1}{16}[1, 4, 6, 4, 1]. \tag{2.70}$$

Therefore, in 2 dimensions the filter is:

$$h_{2D}^0 = h_1^{0T} h_1^0 = \frac{1}{256} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}. \tag{2.71}$$

For the definition of the wavelet to be complete, Starlets need a wavelet function with its associated filter g_1. In Starck et al. (2007) the authors showed that exact reconstructions could be achieved for any pair of even-symmetric filters (h, g) such that $g = \delta - h$, where:

$$\delta = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.72}$$

Such filter g corresponds to a wavelet function defined as the difference between two scaling functions at consecutive resolutions. Therefore, the wavelet function of the Starlet decomposition is the function Ψ_{2D} that verifies:

$$\forall (x, y) \in \mathbb{R}^2, \quad \frac{1}{4}\Psi_{2D}\left(\frac{x}{2}, \frac{y}{2}\right) = \Phi_{2D}(x, y) - \frac{1}{4}\Phi_{2D}\left(\frac{x}{2}, \frac{y}{2}\right). \tag{2.73}$$

The 1-D profile of both the scaling and starlet function are shown in figure 2.11.

With this simple and elegant definition, it comes that wavelet coefficients can be computed by performing the difference between two sets of approximation coefficients, such that:

$$w_j = c_{j-1} - c_j. \tag{2.74}$$

It follows that the reconstruction of a signal is simply the summation of all wavelet coefficients

(a) Scaling function, Φ (b) Starlet function, Ψ
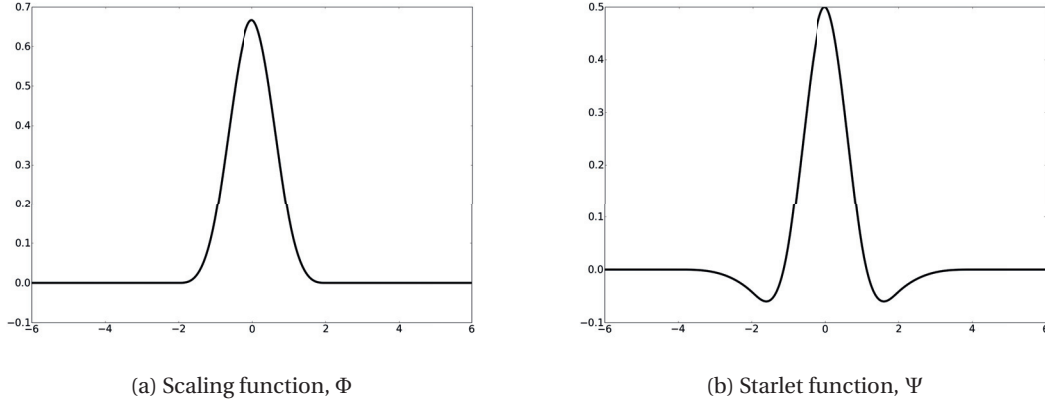
Figure 2.11 – Profile of the Starlet scaling and wavelet function.

with the last approximation scale:

$$c_0 = c_J + \sum_{j=1}^{J} w_j, \tag{2.75}$$

Leading to reconstruction filters $\tilde{h}_1 = \tilde{g}_1 = \delta$.

The result of the Starlet transform of an two dimensional signal (e.g. an image) c_0 with $n_p \times M_p$ samples (pixels) is a set of wavelet coefficient $\{w_j\}_{j \in [\![1,J]\!]}$ along with the last approximation scale, also called the coarse scale, c_J. Each element w_j and c_J is a set of $n_p \times M_p$ coefficients that stand for the variations of c_0 at each location at the corresponding scales. We show in figure 2.12 the Starlet decomposition of a spiral galaxy from the Hubble Frontier Fields survey. Observations were conducted in several bands. We present here the composite RGB image made from filters f435W, f606W and f814W of the ACS instrument on board HST. The Starlet decomposition is performed independently in each band and for each scale, we show the composite image made from the Starlet decomposition of each filter at the same scale. More on the subject of colour images in chapter 4.

Second generation Starlet transform

With the relaxation of the de-aliasing condition in redundant transforms, the design of a filter bank only has to satisfy the exact reconstruction equation (2.63). With this convenient rule in mind, one might be tempted to design filter banks that are more adapted to specific data representation. In astronomy, for instance, where most of the signals we measure consist in counting photons, we expect only positive features. For that reason, it is important to have reconstructions that favour positive reconstructions. In the case of Starlets, the wavelet function has negative features (which are required by the zero average condition on wavelet functions) and the reconstruction operator consists in summing wavelet coefficients with coarse scale

(a) Original image

(b) Scale 1

(c) Scale 2

(d) Scale 3

(e) Scale 4

(f) Scale 5

(g) Coarse scale
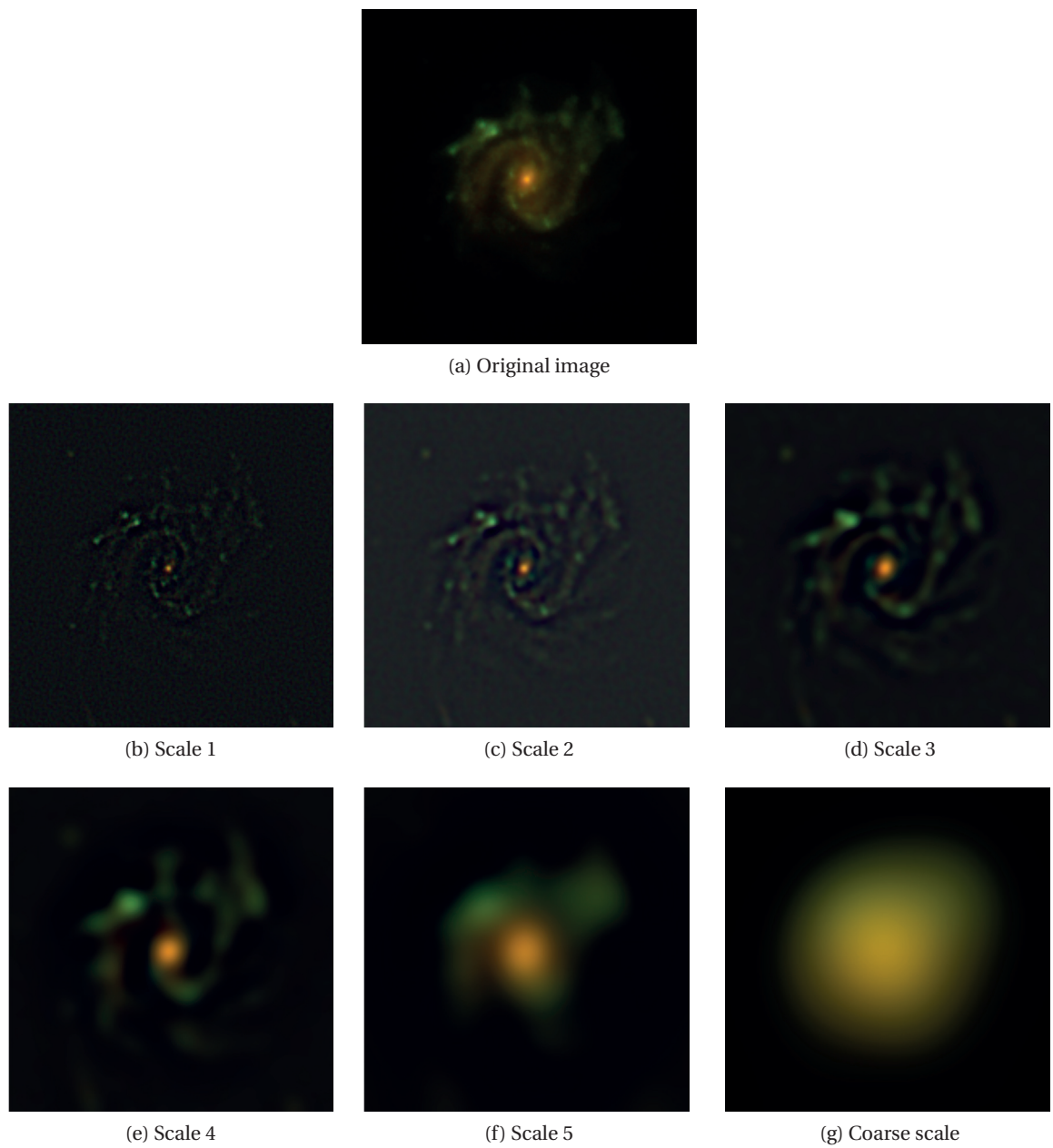
Figure 2.12 – Starlet decomposition of a spiral galaxy.

coefficients, which might result in the wavelet's negative features to show on the reconstructed image, causing negative rings around positive features. In order to avoid this effect, Starck et al. (2007) proposed a Starlet decomposition that admits a positive reconstruction filter under the form $\tilde{h}_{2D} = h_{2D}$, with $h_{2D} = h_{2D}^0$ from equation 2.71. With this reconstruction filter, the filter bank that verifies the exact reconstruction criteria, write as:

$$g_{2D} \quad = \delta - h_{2D} * h_{2D}, \tag{2.76}$$
$$\tilde{h}_{2D} \quad = h_{2D}, \tag{2.77}$$
$$\tilde{g}_{2D} \quad = \delta. \tag{2.78}$$

We notice that the shape of the starlet functions in 1-D (see figure 2.11) is very similar to the shape of emission lines recovered by spectrographs. With this simple realisation in mind, we illustrated how starlets could be used to analyse sprectra of galaxies, as recovered from the MUSE (Multi Unit Spectroscopic Explorer) integral-field spectrograph, by filtering noisy spectra, thus enabling easy identification of spectral features for spectral redshift estimation. The description of this application is detailed in appendix A.

2.4   Component separation

A common problem in signal processing consists in separating components that have been mixed to form a single observation. Such problem can be illustrated by the example of the "cocktail party". Let us picture a cocktail party with a large attendance. Several microphones are distributed across the room that all retransmit through the same speaker. When several microphone holders decide to deliver a speech at the same time through their respective microphones, how does one disentangle their voices and separate each of their discourses?

In astronomy, deep observations result in images that see overlapping objects of different nature:

- Stars of our own galaxy occult the light from background galaxies,

- Diffuse emissions from our galaxy shielding background diffuse emission,

- Close by galaxies hiding images of galaxies at higher redshift,

- Different stellar populations emitting at common wavelength,

- Compact objects embedded in the glare of larger object, at the same or at different redshifts,

to state only a few. In the rest of this dissertation, we will use the term blending to refer to images of overlapping objects with different nature.

Since precision measurements can usually only be achieved by isolating the objects of interest, we are very often constrained to either throw away data that show blending, or deal with the problem of component separation, which is, in general, not trivial. In this section, we will outline a few of the strategies that have been used across this thesis work to solve a wide range deblending problems.

2.4.1 Definition

Component separation problems are also called de-mixing problems, and, more specifically in astronomy, deblending problems. They consist in assuming that a matrix Y containing n_s observations (columns) of size n_o (number lines) each, can be written as a sum of n_c matrices $\{X_1, X_2, \cdots, X^{n_c}\}$ of the same size $n_o \times n_s$ such that:

$$Y = \sum_{k \leq n_c} X^k. \tag{2.79}$$

In deblending problems in astronomy, it is very common that the terms sources and components be confused. At this point, we would like to make it clear that each matrix X^k will be called a component in the rest of this dissertation. The term source will refer to compact astrophysical objects that form a single surface brightness. As such, a component may contain several sources (e.g., several galaxies with the same colour, see chapter 4) and a source might be decomposed into several components (e.g., a galaxy with several stellar populations/colours across its surface brightness, see figure 2.12 and chapter 4). We extend the term component to the linear transformations, S_k, of components such that $X^k = A S_k$, where A is a (known) matrix of coefficients.

The goal of solving a deblending problem is to reconstruct components $\{X^i\}_{i \leq n_c}$ such that $Y - \sum_{i \leq n_c} X^k = 0$. In practice, and in all the component separation problems we will consider, the observations are contaminated by additive noise Z. As a result, the goal becomes to reconstruct the components X^k such that:

$$\{\tilde{X}^k\}_{i \leq n_c} = \underset{X^1, X^2, \cdots, X^{n_c}}{argmin} \frac{1}{2} \|Y - \sum_{k \leq n_c} X^k\|_2^2. \tag{2.80}$$

Unfortunately, due to the large number of degrees of freedom in the reconstruction of the components, equation 2.80 admits an infinity of solutions. Most of the work in component separation, therefore consists in understanding the properties of the components that allow to discriminate between them and to provide a separation criteria.

2.4.2 Principal and independent component analysis

Principal component analysis

A first (in the historical sense) solution to equation 2.80, comes in the form of a decomposition into independent orthogonal components, which is provided by the principal component analysis (PCA F.R.S., 1901; Jolliffe, 2011). Let us assume that our observables Y are made of a series of n_y vectors ,$\{y_i\}_{i \in [[1,n_y]]}$ such that each vector y_i can be seen as a linear combination of orthogonal, independent vectors $\{v_j\}_{j \in [[1,n_v]]}$, such that:

$$\forall i \in [[1, n_y]], y_i = \sum_j a_{i,j} v_j, \tag{2.81}$$

where coefficient $a_{i,j}$ is a real number that gives the contribution of component v_j to vector y_i. In matrix form, equation 2.81 becomes:

$$Y = AV, \tag{2.82}$$

where the i-th column of matrix Y contains vector y_i, the j-th column of matrix V contains vector v_j and the element of the i-th line and j-th column of matrix A is the coefficient $a_{i,j}$[3] . This factorisation of the component separation problem is called a blind source separation (BSS) problem. Most of the problems of source separation encountered in astrophysics can be expressed as a BSS, in particular, when dealing with multi-channel data (Y is a collection of images).

Such decomposition can be done by taking the singular value decomposition (SVD) of Y, which consist in factorising Y such that:

$$Y = U\Omega W^T, \tag{2.83}$$

where U and W^T are orthonormal matrices and Ω is a diagonal matrix with positive coefficients sorted in decreasing order along the diagonal. The ordering of the coefficients comes from the fact that the SVD of a matrix is linked to the eigenvalue decomposition of the covariance of Y:

$$T^Y Y = W\Sigma W^T, \quad with \tag{2.84}$$
$$\Sigma = \Omega^T \Omega, \tag{2.85}$$

meaning that the diagonal elements of Ω correspond to the square root of the eigenvalues of $Y^T Y$ and W is the matrix formed by its eigenvectors.

---

[3]We notice that, with this formulation, the element at position $[i, j]$ of component X^k from equation 2.79, now writes $X^k_{[i,j]} = \begin{cases} a_{i,j} V_{j,m} & \text{if} \quad m = k \\ 0 & \text{otherwise} \end{cases}$.

With this formulation, we can write the factorisation $Y = AV$ in terms of:

$$A = U\Omega, \tag{2.86}$$
$$V = W^T, \tag{2.87}$$

Which provides a natural decomposition in to orthogonal components for Y.

In practice, because the diagonal values of Ω are sorted in decreasing order, it is very common to truncate the decomposition at a certain number of eigenvalues without a significant loss of information. This will be expanded on further in chapter 3 where we use PCA to separate images foreground smooth galaxies from compact textured background images.

Independent component analysis

It is easy to notice that from this decomposition, many other decompositions (A', V') can be found that verify $Y = A'V'$. Explicitly, for any invertible matrix I, the decomposition $(A' = AP, V' = P^{-1}V)$ is a decomposition of Y. While not all transformation matrices P might provide a valuable decomposition in terms of component separation (e.g, scaling and permutations), they provide a tool to investigate more general decompositions than the ones restricted to orthonormal decompositions such as PCA.

This non-orthogonal generalisation of PCA, called independent component analysis (ICA) relies on the separability of components (PV) based on their statistical independence (Jutten & Herault, 1991; Comon, 1994). Meaning that ICA aims at finding invertible matrices P that maximises the independence of the vector of PV based on a given criterion.

A number of algorithms have been proposed in the literature that provide ICA for various independence criteria: Hyvarinen (1999); Cardoso (1999); Bell & Sejnowski (1995) .

2.4.3 Morphological component analysis

With their pioneering work, Zibulevsky & Pearlmutter (2001) paved the way for a new separation criteria beyond independence: sparsity. They show that multiple mixtures (lines in Y in equation 2.82) of the same components (lines in V in equation 2.82) can allow the reconstruction of the components based on their sparsity in a dictionary Φ. This can be viewed as solving a linear inverse problem with an assumption of sparsity, but where the linear operator (here, the mixing matrix A) is not known and needs to be estimated.

Morphological diversity

Separability based on sparsity was then extended to give birth to the concept of morphological diversity (Starck et al., 2004, 2005a). Morphological diversity recognises that several compo-

nents of a same mixture might not be sparse in the same dictionary. This powerful notion allowed the authors to formulate the morphological component analysis (MCA) framework. MCA allows to recover mixed components from only one mixture, provided that each component can be sparsely decomposed in one dictionary, but not in the dictionaries of other components. This implies that each components has to present morphological characteristics that allow to discriminate between them provided that we can express a basis that represent these morphologies.

Formally, in MCA, the mixture is a one-dimensional vector Y that can be decomposed as

$$
\begin{aligned}
Y &= \sum_{i \leq n_c} x_i & (2.88) \\
&= \sum_{i \leq n_c} \Phi_i \alpha_i, & (2.89)
\end{aligned}
$$

where $\{x_i\}_{i \leq n_c}$ are the components we aim to reconstruct, $\{\Phi_i\}_{i \leq n_c}$ are the dictionaries in which the components $\{x_i\}_{i \leq n_c}$ are sparsely represented and $\{\alpha_i\}_{i \leq n_c}$ are the coefficients of the decompositions of vectors $\{x_i\}_{i \leq n_c}$ in their respective dictionaries. Because of the sparsity assumption, the solutions for equation 2.89 are obtained by solving the constrained problem:

$$
\{\tilde{\alpha}_i\}_{i \leq n_c} = \underset{\{\alpha_i\}_{i \leq n_c}}{argmin} \frac{1}{2} ||Y - \sum_{i \leq n_c} \Phi_i \alpha_i||_2^2 + \sum_{i \leq n_c} \lambda_i ||\alpha_i||_1. \tag{2.90}
$$

Problem 2.90 is very similar to the synthesis formulation of linear inverse problems 2.31, but has the particularity of requiring optimisation over several variables that each require different thresholds λ_i. A baseline algorithm for solving problem 2.90 consists in alternating optimisation steps over each component α_i as shown in algorithm 1, where n_{iter} is the number of alternating steps (iterations). In this algorithm, the threshold that accounts for the sparsity of the solution is updated at every iteration. The threshold λ^0 is set to start high such that the most salient features of each component are reconstructed in their respective dictionaries. The threshold is then decreased to account for the full reconstruction of each coefficient.

Several decrease strategies can be devised for the threshold. The starting threshold is often set to be high enough such that only the highest coefficient in the sparse domain, across all components. It is possible to design a decrease such that the threshold value decreases following a decreasing analytical function (usually, linear or exponential decrease), but data driven strategies have also been devised that allow to accelerate the convergence of the algorithm, in particular, in cases where the contrast between components is high. Such data driven strategies consist in considering contamination between components as a noise and to keep the threshold right above this noise level (Donoho et al., 2012), or to decrease the threshold right above the value of the component that has the second highest coefficient across all components[4] (Bobin et al., 2007a). In my experience, I found useful to combine these two

---

[4]This strategy ensures that, at every iteration, only the component that contributes the most to the residuals

---

Algorithm 1 MCA algorithm

1: **procedure** $\text{MCA}(Y, n_{iter}, \Phi_i\}_{i \leq n_c})$
2:     **Initialisation of** $x_i^0, \lambda^0 \forall i$
3:     **for** $k = 1$ **to** n_{iter} **do**
4:         $R \leftarrow Y - \sum_{i \leq n_c} \Phi_i \alpha_i^k$
5:         **for** $i \leftarrow 1$ **to** n_c **do**
6:             $x_i^k = x_i^{k-1} + R$
7:             $\alpha_i^k \leftarrow ST_{\lambda^{k-1}}(\Phi_i^T x_i)$
8:             $x_i^k \leftarrow \Phi_i \alpha_i^k$
9:         **end for**
10:        **Update** $\lambda^k, \quad \forall i$
11:     **end for**
12:     **return** $\{x_i^k\}_{i \leq n_c}$
13: **end procedure**

---

approaches by setting a analytic linear decrease but allowing the coefficient to decrease faster in cases of high contrasts. This strategy has the advantage of avoiding unnecessary iterations at high contrast and spending more iterations on the careful reconstruction of features at low contrast. In some cases, it is even possible to keep the threshold at the minimum value in cases of high sparsity of the component in their respective dictionaries.

A very schematic, but nonetheless intuitive way of interpreting this algorithm can be drawn by considering the very simple problem of separating spheres and tiles mixed in a bag. Let us consider a large number of identical tiles and spheres mixed in a bag as represented in figure 2.13. With the radius of the spheres is contained between the smallest and largest dimensions of the tiles, how can I separate the spheres from the tiles? Knowing the morphological properties of the objects, a simple solution consists in throwing the content of the bag on a sieve of circular apertures with the same radius as the spheres. When enough spheres have been through the sieve and the number of tiles is much larger than the number of spheres left, we take the rest to a sieve of rectangular apertures with the same dimensions as the tiles. By alternating between the circular and rectangular sieve, we should end up with a perfect separation between spheres and tiles.

We illustrate this by performing a decomposition between low and high frequency features, in the Starlet domain, in images of galaxies for the purpose of measuring surface brightness fluctuations (SBF Jensen et al., 2015; Cantiello et al., 2018), in the frame of the Next Generation Virgo cluster Survey (NGVS; PI Laura Ferrarese). Results can be found in appendix B.

We developed a modified version of the MCA interlaced with a linear inverse problem to model lens systems, which we present in chapter 5.

---

are reconstructed. Compared to analytical strategies, in case of high contrast between components, this has the advantage of skipping many iterations that would otherwise have been spent reconstructing the same component.
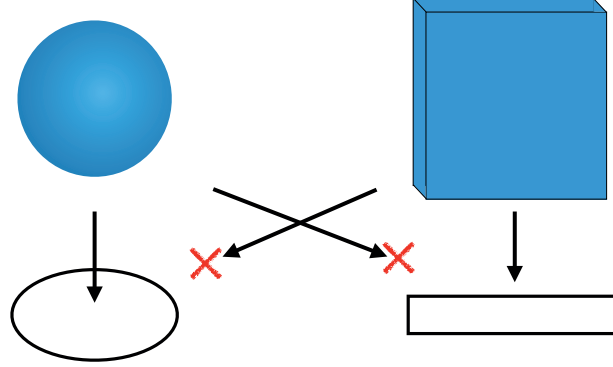
Figure 2.13 – Schematic interpretation of morphological component analysis.

Generalised morphological component analysis and beyond

With the works of Bobin et al. (2006); Bobin et al. (2007b); Bobin et al. (2008), the MCA framework was extended to multi-frame data to produce the generalised morphological component analysis (GMCA). GMCA addresses the same problem as equation 2.82, where the aim is to recover a set of components (lines in X) that are mixed in different unknown linear combinations represented by the elements of the mixing matrix A in a set of multi-frame observations (lines of Y). As in MCA, each component in X is sparse in a given dictionary. With this formulation, A are X solution of the constrained optimisation problem:

$$\tilde{A}, \tilde{X} = \underset{A,S}{arg\,min} \frac{1}{2} ||Y - AX||_2^2 + \lambda \sum_{i<n_c} ||\Phi_i X_i||_1, \tag{2.91}$$

Where X_i is the i-th line of matrix X and Φ_i is the dictionary where X_i is sparse. With GMCA, the component separation benefits from the separation criteria provided by the multiframe observations (Zibulevsky & Pearlmutter, 2001) combined with the separation from morphological diversity. In essence, MCA can be seen as special cases of GMCA, which corresponds to the case where Y is a one dimensional vector (single frame observation).

Solving a GMCA problem also includes solving the problem in A. As a result, GMCA algorithms alternate between minimisations over X at fixed A and minimisations over A at fixed X, as shown in algorithm 2

In the last decade, several similar techniques have emerged that are more adapted to solve other specific source separation problems:

Algorithm 2 MCA algorithm

1: **procedure** GMCA($Y, n_{iter}, \Phi_i\}_{i \le n_c}$)
2:     **Initialisation of** $A, X, \lambda^0 \forall i$
3:     **for** $k = 1$ **to** n_{iter} **do**
4:         $X \leftarrow \underset{X}{argmin} \frac{1}{2} ||Y - AX||_2^2 + \lambda \sum_{i < n_c} ||\Phi_i X_i||_1$
5:         $A \leftarrow \underset{X}{argmin} \frac{1}{2} ||Y - AX||_2^2$
6:     **end for**
7:     **Update** λ
8:     **return** $\{x_i^k\}_{i \le n_c}$
9: **end procedure**

- Local-GMCA (L-GMCA Bobin et al., 2013; Bobin et al., 2011) uses the principles of GMCA on adaptive patches of the sky and locally estimates the mixing matrix, it was successfully applied to CMB the reconstruction problem in Bobin et al. (2014);

- non-negative GMCA (nGMCA Rapin et al., 2013, 2014) accounts for the positivity of the observables in direct space (Paatero & Tapper, 1994; Donoho & Stodden, 2004) along with the sparsity of the sources in transformed space to perform blind source separation;

- Adaptive MCA (AMCA Bobin et al., 2015a) aims at solving source separation problems for partially correlated components;

- Deconvolution in GMCA (DecGMCA Jiang et al., 2017) for separation problems mixed with ill-posed deconvolution problems.

- robust GMCA (rGMCA Chenot et al., 2015), a separation algorithm robust to outliers.

Taking inspiration from these algorithms, techniques were recently developed in Moolekamp & Melchior (2017); Melchior et al. (2018) that rely on other linear constraints, thus extending component separation to an even more general framework. Using alternative direction method of multipliers (ADMM) , the authors showed that one can combine any number of linear constraints of any nature (sparsity, positivity, symmetry, monotonicity $\cdots$) to provide adequate separation criteria.

2.5 Outline

In this chapter, we have introduced the tools that form the basis for understanding most of the work done in this thesis with regard to linear optimisation. The key concepts that are extensively referred to in this work are those of sparse optimisation in the wavelet domain and component separation. I provided the theoretical framework that allows to build efficient algorithms of minimisation in this context by introducing proximal calculus and by giving an overview of the concept of wavelet decomposition. Morphological component analysis is the

starting ground of most of the applications we will present here. Even if the algorithms developed to solve specific problems related to strong gravitational lensing, show mild deviation from the baseline algorithms, they all rely on the basic ideas formulated in algorithms 1 and 2. In chapter 3, we will see how principal component analysis allowed us to facilitate automated detection of gravitational arcs and rings and laid the grown for future development of more advanced machine learning-based methods. Chapter 4 uses a simplified version of GMCA to separate blended images of astronomical objects with different colours in the optical domain. In chapter 5, we show how MCA can be used in single band images to separate lensed from non-lensed objects by solving a linear inverse problem of lens inversion. With MCA being a particular case of GMCA, we will see that it is possible to mix both approaches of colour and lensing component separation.

3 Automated Strong Lens Finding

Contents

While strong gravitational lenses at galaxy scale have the potential to bring a multitude of information about distant objects and about cosmology alike, it remains a rare event. With most galaxy-scale lens systems being only a few arseconds in size on the plane of the sky and the sky counting 64800 square degrees with billions of light sources, finding them constitutes a challenging problem. In this chapter, I review the different approaches to finding galaxy-scale strong lenses and show the result of the application of a method I designed prior to the beginning (see section 3.3) of this thesis to the CFHTL Survey in section 3.3.

The method we proposed and that we present in this chapter relies on the use of principal component analysis to learn the shape of lens galaxies, remove them and search lensed

features in the residual image. While PCA can be considered as the level zero of machine learning, we only used it as a preprocessing tool to facilitate our lens search, but not as a selection criteria for lens system candidates, due to its limited capacities. With the revolution of deep learning algorithms (LeCun et al., 2015) and convolutional neural networks (CNN Lecun et al., 1998; Fukushima, 1980), machine learning became robust enough to produce a robust selection criteria, and the idea was applied in a multitude of lens finding algorithms as illustrated in Metcalf et al. (2018) and references therein.

3.1 The Problem of Finding Lenses

Strong gravitational lenses come in a wide variety of configurations and observational characteristics that each hold different information. As we have seen in chapter 1, lensed sources with QSOs allow us to probe the cosmological parameter space, and each extended lensed source allows us to probe the distribution of dark matter from substructure scale to galaxy-scale.

3.1.1 Lens diversity

Lensed quasars can assume various configurations: the multiplicity of observable quasar images usually ranges from 2 to 4. The configuration of a quadruply imaged quasars comes in cusp, cross or fold as illustrated in figure 3.1 by lens systems RXJ1131, HE0435 and WFI2033 respectively. In some systems, the host galaxy contributes significantly to the system's surface brightness (lenses B1608 and RXJ1131). In the case of extended background sources, images may form as rings or arcs. The morphology of the lens galaxy is also a factor to take into account when searching for lens systems. With lenses being slightly more likely to form around massive early type galaxies, a few examples of gravitational lenses have been found in spiral galaxies (Treu et al., 2011; Huchra et al., 1985), where the lensed images tend to form inside the bulge of the lens.

Of course, each of these configurations come with a variety of image separations and luminosities. As a result of this variety of morphologies, most lens search algorithms are specialised into finding a specific type of lenses. One major difficulty faced when searching for strong gravitational lenses, arises from the difficulty to identify lensed features in the glare of lens galaxies. For that reason, many of the techniques we are going to review in the next section rely on the careful subtraction of the lens light.

3.1.2 Expectations

The coming years should see the coming on line of several large optical surveys such as DES, LSST or Euclid, that will map large portions of our sky. In Collett (2015), the author predicts the number of lens system that should be observed based on simulations of realistic lenses and compares his prediction of about existing survey, CFHTLS with the results of Gavazzi et al.
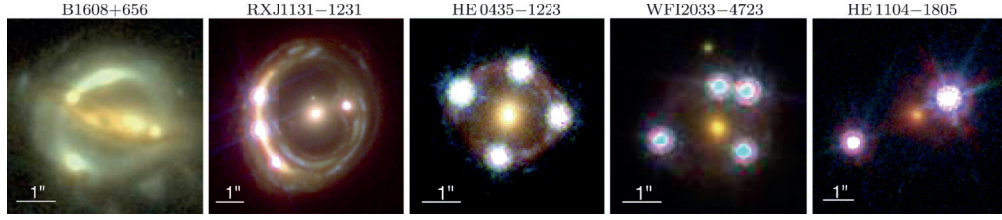
Figure 3.1 – Five lens systems from the H0LICOW configuration (Suyu et al., 2017)

(2014) and Sonnenfeld et al. (2013). We summarise in table 3.1 the number of expected lenses in each survey ($N_{predicted}$) along with the sky coverage of the survey (Ω). It is important to notice that Euclid is expected to find more than twice as many lens systems as LSST, which has a larger sky coverage. This is due to Euclid being a spaceborne probe, which resolution should allow to detect lenses systems with smaller separation and Einstein radii.

Survey	$\Omega\,(deg^2)$	$N_{predicted}$
CFHTLS	150	250
DES	5000	2300
LSST	20000	120000
Euclid	15000	280000

Table 3.1 – Prediction for lens discoveries in upcoming optical surveys from Collett (2015).

With so many gravitational lenses to be found among several orders of magnitude more light sources, we need to devise strategies to search for gravitational lenses in a reasonable time.

3.2 State of the Art in Lens Finding

Efforts towards the development of large scale search for strong lenses started in the early 2000's and are now booming with the rise of machine learning and the urge of upcoming surveys.

3.2.1 Visual inspection

The first natural method for finding strong gravitational lenses naturally comes from visual inspection of preselected sources. In Browne et al. (2003), the authors searched for gravitational lenses through the ~ 16000 radio sources of the VLA cosmic lens all-sky survey. After pre-selection based on spectral features, ~ 8000 sources were inspected among which 149 lens candidates were selected. In the optical, Faure et al. (2008) searched for lenses among the ~ 280000 source of the COSMOS survey. Similarly to what was done in the radio domain, they performed a pre-selection of ~ 10000 candidates based on their luminosity, photometric

redshift and SED. In a following paper, Jackson (2008) realised the daunting effort of manually inspecting each of the 285423 sources of the COSMOS survey. The author reports an average 8000 to 10000 objects inspected per hour, which amounts to 28 hours of inspection for the full survey to be covered. Given that the COSMOS survey is 1.64 square degrees and assuming a Euclid-like survey of 15000 square degrees with similar properties as the COSMOS survey in terms of depth and resolution, it would take 256097 hours, or 29 years of continuous inspection to go through the whole survey. To overcome this major issue, a multiprocessing approach of visual inspection was proposed by way of the Space Warps project (Marshall et al., 2016; More et al., 2016) where the authors relied on citizen science to process the 430000 tiles of the CFHTLS survey.

3.2.2 Automation of imaging searches

Due to the increasing size and number of large surveys, the past decade has seen the multiplication of efforts to produce automated or semi automated methods to find strong gravitational lenses. While it has proven so far, challenging to produce fully automated methods, most techniques focus on reducing the number of potential candidates that require visual inspection to a few thousand images.

Feature extraction

Because strong gravitational lensing produces characteristic elongated and distorted features, several algorithms have focused on the detection and extraction of such features. The `arcfinder` algorithm (Alard, 2006; Seidel & Bartelmann, 2007; More et al., 2012) identifies arc-like, elongated features in images of lens candidates. In Gavazzi et al. (2014); Joseph et al. (2014); Paraficz et al. (2016), the authors focus on the identification of rings around elliptical galaxies that are subtracted prior to the ring identification.

lens modelling search

Another approach consists in assessing whether a lens candidate system can be explained by a lens model. This is the approach chosen by Marshall et al. (2009), where each early type galaxy is considered as a potential lens system. Then the surface brightness of the candidate is fitted with a gravitational lens model. The adequacy of the fit gives an indication of how likely a system is to be a lens based on the quality os the fit of actual known lens systems. The `CHITAH` algorithm (Chan et al., 2015) searches multiply imaged QSOs as multiple point sources around elliptical galaxies that are previously subtracted based on colour difference imaging. The identification of multiple point sources as a lens system is then performed by assessing whether the position of the lensed quasars could be explained using a lens model. In Sonnenfeld et al. (2018), the authors combine the arc and ring identification feature to the lens modelling search techniques to give a better identification criteria of lensed arcs and rings.

3.2.3 spectroscopic search

Other than imaging, several methods rely instead on the identification of features in spectroscopic data. Due to the small angular separation between a lens light profile and its potential lensed source, it is possible to identify spectral lines that are not coherent with the spectra of single objects, but instead, can be explained by the superimposition of two objects at different redshifts (Bolton et al., 2006; Willis et al., 2006; Brownstein et al., 2012).

A particularly interesting application of such technique was implemented to specifically target systems where QSOs act as lenses. In Courbin et al. (2010, 2012); Meyer et al. (2017) the authors identify spectral emission lines in the spectra of pre-selected QSOs that could only be explained by a higher redshift source.

3.2.4 The dawn of machine learning

With the rapid development of algorithms and techniques for machine learning, this field has revolutionised the search for strong gravitational lenses. In the recent strong lens finding challenge (Metcalf et al., 2018), out of the 24 submissions 19 can be coined as machine learning-based techniques. While most of them (15 submissions) rely on convolutional neural networks (Jacobs et al., 2017; Petrillo et al., 2017; Lanusse et al., 2018; Schaefer et al., 2018) to account for the information of all pixels in the image set. Other techniques focus on learning features that are specific to gravitational lenses using Gaussian mixture models (Ostrovski et al., 2017), support vector machine (SVM, Hartley et al., 2017) or neural networks on pre-selected arc or ring-like features (Bom et al., 2017).

3.2.5 Results of the lens finding challenge

With the large number of CNN-based submission and the growing interest for this approach, it is no surprise that the methods described in Lanusse et al. (2018) and Schaefer et al. (2018) be amongst the top ranking techniques in terms of compromise between completeness and purity of the selection. The SVM technique of Hartley et al. (2017) and the CNN of Lanusse et al. (2018) ranked the highest in terms of true positive ratio, meaning that these methods provide the highest completeness in their selection. It is interesting to notice that human inspection was also used in this challenge, where two human classifiers managed the inspection of the the 100000 images in under 48 hours. Interestingly, most machine learning methods performed better than human inspection due to these techniques providing better levels of confidence on images where humans would be hesitant. This is due to the adequacy between the training set and the test set. However, the "human classifier" was the only one to correctly identify an exotic lens system with two lensed sources, which morphology was not included in the training set.

3.3 A PCA-based Deblender for Lens Finding

3.3.1 Lens light subtraction

An important issue faced when searching for galaxy-scale strong lens systems is the sometimes small angular separation between the lens and the lensed source, which results in the blending of the light profiles. In many approaches, the authors chose to remove the foreground galaxies in a survey before searching for lensed features in the residuals. This assumes a pre-selection of potential lens galaxies, which is usually done based on criteria of size, luminosity and colour.

Several approaches have been used to subtract the light profile of a lens candidate. The most common one comes from colour difference imaging. Given that source galaxies at higher redshift tend to be star forming galaxies, while foreground lenses are often early galaxies, source galaxies are expected to be bluer than red lens galaxies. Using this principle, Gavazzi et al. (2014) performs lens search on the subtraction of bands $g - i$ in the CFHTL Survey while Chan et al. (2015) performs the same operation with the subtraction of bands $g - z$.

As we will see in chapter 4, stating that a source is bluer than an other one does not necessarily defines an exclusive relation with regard to light contribution to imaging filters. This means that a blue source might contribute to the surface brightness of the system in a red filter. As a result, the band subtraction approach can reduce the signal of an object in the detection image, which can reduce the detection power especially at low signal to noise ratio.

In Brault & Gavazzi (2015) the authors propose to select early type galaxies that usually present smooth, elliptical surface brightness profiles. The smoothness and ellipticity assumption allows the authors to remove the foreground lens light profile by fitting an elliptical light profile using the `galfit` software (Peng et al., 2002, 2010). While the procedure of fitting elliptical profiles to each galaxy in a survey can be time consuming, it has the advantage of allowing lens light subtraction in single-band imaging, where band subtraction techniques require multi-band observations.

3.3.2 Learning light profiles from PCA

In an attempt to provide an efficient method for galaxy-scale strong lensing detection, I developed a lens finding algorithm that relies on the subtraction of the lens light profile in single band imaging surveys.

The method performs principal component of a preselected cube of galaxy images with consistent sizes and ellipticities. The PCA is used to build a basis of morphological features, for which the first few components will represent the most common features in the data. As a result, the reconstruction of an image projected on the first few vectors of this basis will build a simple profile made of the most common preselected features in the survey: smooth elliptical galaxies. The residuals between the reconstruction and the original image shows

objects hidden in the glare or in the vicinity of elliptical galaxies. The details of the method are given in the preprint of a paper published in *Astronomy & Astrophysics* (Joseph et al., 2014).

In this paper, we also describe two methods for strong lens features identification in the PCA subtracted image. In the first technique, that I developed, performs a polar coordinate transform of the PCA-subtracted images. In this the polar system, a ring-like feature would be seen as a line, which is detected by identifying peaks in the projection of the polar transformed image along the angular axis.

The second method developed by co-authors Philippa Hartley and Neal Jackson identifies objects in the PCA-subtracted images using the `Sextractor` (Bertin & Arnouts, 1996) software. The characteristics of the extracted features are then scanned to look for radial alignments and extreme elongations. In this paper, we present of a comparison of both methods on simulated data provided by the `Bologna Lens Factory` project[1].

---

[1] https://bolognalensfactory.wordpress.com

A PCA-based automated finder for galaxy-scale strong lenses

R. Joseph[1], F. Courbin[1], R.B. Metcalf[2], C. Giocoli[2,3,4], P. Hartley[5], N. Jackson[5], F. Bellagamba[2], J.-P. Kneib[1], L. Koopmans[6], G. Lemson[7] M. Meneghetti[3,4,8], G. Meylan[1], M. Petkova[2,9,10], and S. Pires[11]

[1] Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland
[2] Dipartimento di Fisica e Astronomia - Universita di Bologna, via Berti Pichat 6/2, I-40127 Bologna, Italy
[3] INAF - Osservatorio Astronomico di Bologna, via Ranzani 1, 40127, Bologna, Italy
[4] INFN - Sezione di Bologna, viale Berti Pichat 6/2, 40127, Bologna, Italy
[5] Jodrell Bank Centre for Astrophysics, School of Physics & Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[6] Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands
[7] Department of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 München, Germany
[8] Jet Propulsion Laboratory, 4800 Oak Grove Dr., La Canada-Flintridge, CA 91011, USA
[9] Max-Planck-Institut für Astrophysik, D-85748 Garching, Germany
[10] Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany
[11] Laboratoire AIM, CEA/DSM-CNRS-Universite Paris Diderot, IRFU/SEDI-SAP, Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France

ABSTRACT

We present an algorithm using Principal Component Analysis (PCA) to subtract galaxies from imaging data, and also two algorithms to find strong, galaxy-scale gravitational lenses in the resulting residual image. The combined method is optimized to find full or partial Einstein rings. Starting from a pre-selection of potential massive galaxies, we first perform a PCA to build a set of basis vectors. The galaxy images are reconstructed using the PCA basis and subtracted from the data. We then filter the residual image with two different methods. The first uses a curvelet (curved wavelets) filter of the residual images to enhance any curved/ring feature. The resulting image is transformed in polar coordinates, centered on the lens galaxy center. In these coordinates, a ring is turned into a line, allowing us to detect very faint rings by taking advantage of the integrated signal-to-noise in the ring (a line in polar coordinates). The second way of analysing the PCA-subtracted images identifies structures in the residual images and assesses whether they are lensed images according to their orientation, multiplicity and elongation. We apply the two methods to a sample of simulated Einstein rings, as they would be observed with the ESA Euclid satellite in the VIS band. The polar coordinates transform allows us to reach a completeness of 90% and a purity of 86%, as soon as the signal-to-noise integrated in the ring is higher than 30, and almost independent of the size of the Einstein ring. Finally, we show with real data that our PCA-based galaxy subtraction scheme performs better than traditional subtraction based on model fitting to the data. Our algorithm can be developed and improved further using machine learning and dictionary learning methods, which would extend the capabilities of the method to more complex and diverse galaxy shapes.

Key words. Methods: data analysis – Gravitational lensing: strong – Galaxies: surveys

1. Introduction

With the many ongoing or planned sky surveys taking place in the optical and near-IR, gravitational lensing has become a major scientific tool to study the properties of massive structures at all spatial scales. On the largest scales, in the weak regime, gravitational lensing constitutes a crucial cosmological probe (e.g. Heymans et al. 2013; Frieman et al. 2008). On smaller scales, weak galaxy-galaxy lensing allows us to study the extended halo of individual galaxies or of groups of galaxies (e.g. Simon et al. 2012) and to constrain cosmology (e.g. Mandelbaum et al. 2013; Parker et al. 2007).

In the strong regime, when multiple images of a lensed source are seen, gravitational lensing offers an accurate way to weigh galaxy clusters (Bartelmann et al. 2013; Hoekstra et al. 2013; Meneghetti et al. 2013; Kneib & Natarajan 2011, for reviews), galaxy groups (e.g. Foëx et al. 2013; Limousin et al. 2009) and individual galaxies (e.g. Brownstein et al. 2012; Treu et al. 2011; Bolton et al. 2006). However, all strongly lensed systems known today, combined together, represent only hundreds

of objects. Wide field surveys have the potential to produce samples three orders of magnitude larger, allowing us to study statistically dark matter and its evolution in galaxies as a function, e.g. of morphological type, mass, stellar and gas contents (see Gavazzi et al. 2012; Ruff et al. 2011; Sonnenfeld et al. 2013b,a). For example, Pawase et al. (2012) predicts that a survey like Euclid will find at least 60000 galaxy-scale strong lenses. To find and to use them efficiently, it is vital to devise automated finders that can produce samples of lenses with high completeness and purity and with a well defined selection function. Note that the lenses of Pawase et al. (2012) are source selected. There is no volume-limited sample of lens-selected systems, so the number 60000 systems is given here only to give an order of magnitude of the number of objects that future wide-field surveys will have to deal with.

Several automated robots exist to find strong lenses. Among the best ones are `Arcfinder` (Seidel & Bartelmann 2007), which was primarily developed to find large arcs behind clusters and groups, and the algorithm by Alard (2006) used by

1
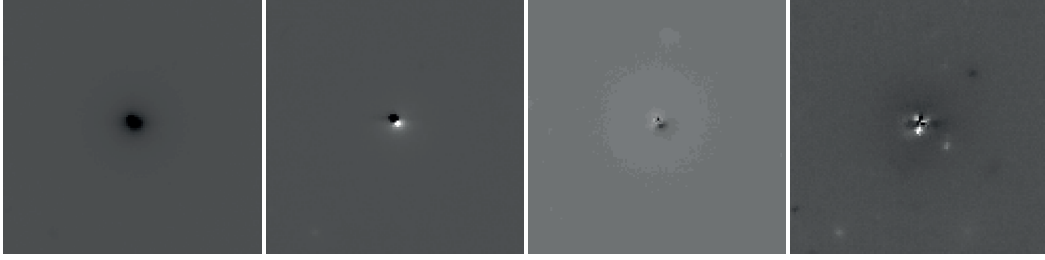
Fig. 1. Examples of PCA components obtained using 1000 simulated galaxies from the Bologna Lens Factory (see Sect. 4).

Cabanac et al. (2007) and More et al. (2012), to look for arcs produced by individual galaxies and groups in the CFHT Strong Lensing Legacy Survey. Other automated robots consider any galaxy as a potential lens and predict where lensed images of a background source should be before trying to identify them on the real data (Marshall et al. 2009). In order to detect lenses with small Einstein radii or with faint rings, most of these algorithms rely on foreground lens subtraction (e.g. Gavazzi et al. 2012). So far, this subtraction has been performed through model fitting. An example of a ring detector is given in Sygnet et al. (2010) which selects objects with possible lensing configuration according to their lensing convergence, estimated from the Tully-Fisher relation. This algorithm relies on photometric information but requires a visual check of a large number of candidates.

In the present paper, we propose a "lens finder" which uses single-band images to find full and partial Einstein rings based on purely morphological criteria. The algorithm uses as input a pre-selection of potential lens galaxies, hence producing so-called "lens-selected" samples. The present work sets the basis of an algorithm using machine learning techniques. Although focused on finding Einstein rings, it can be adapted to other types of lenses, such as those consisting of multiple, relatively point-like, components.

This paper is organised as follows. In Sections 2 and 3 we outline our algorithm and introduce the principles behind each step of the process. In section 4 we show the performance of our algorithm using a set of simulations designed to reproduce Euclid images in the optical. We discuss the completeness and purity of our algorithm as a function of signal-to-noise (SNR) and caustic radius of the lensing systems. Section 5 shows results of our galaxy subtraction algorithm compared to those of `galfit` software (Peng et al. 2011) on images from the CFHT optical imaging of SDSS stripe 82 and Section 6 summarizes our main results.

2. A new automated lens finder

2.1. Principle of the algorithm

By construction, lens-selected samples display bright foreground lenses and faint background sources, otherwise the pre-selection of the lenses based on morphological type, luminosity and color would not be possible. As a consequence, faint Einstein rings are hidden in the glare of the foreground lenses, which must be properly removed before any search for lensed rings can be undertaken. An efficient "lens finder" therefore involves two main steps: 1- removal of the lens galaxy, 2- identification of rings in the lens-subtracted image.

A traditional way of subtracting galaxies is to fit a two dimensional elliptical profile to the data, e.g. as done with the

`galfit` software (Peng et al. 2011). While this is sufficient to characterize the main morphological properties of galaxies, it turns out to be insufficient to detect faint arcs seen superposed on bright galaxies with a significant level of resolved structures.

One way to circumvent the problem is to build an empirical light model from the sample of galaxies itself, i.e. to use machine learning techniques such as Principal Component Analysis (PCA; Jolliffe 1986). The sparsity and the diversity in terms of shape of the *lensed objects* (rings, arcs, multiple images) prevents them from being well enough represented in the basis, hence allowing for an accurate separation of lenses and sources. This has already been used to find lensed sources from PCA decomposition of quasar spectra (e.g. Courbin et al. 2012; Boroson & Lauer 2010). We adopt now a similar strategy to analyse images.

Once the foreground lenses have been properly removed, we analyse the residual rings using methods described in Section 3. The main steps of the algorithm can be summarized as follows:

1. Pre-selection of the galaxies with a predefined range of shape parameters (size, ellipticities, magnitudes, colors, etc.)
2. Building the PCA basis either from the selected sample of galaxies or from an adapted training set.
3. Reconstruction of the central galaxies and subtraction from the original images.
4. Detection of lensed objects, either using island finding (groups of adjacent pixels) or a polar transform or the residual image.

2.2. Selection of galaxies

The first step of this method is to build stamp images of galaxies in which to look for lensed objects. This step strongly depends on the specific sample considered and may take advantage of algorithms such as `SExtractor` (Bertin & Arnouts 1996).

For the PCA decomposition to work well, a compromise has to be found between the number of objects used to build the PCA basis, the size of the objects in pixels, and the range in shape parameters. The more complex the galaxies are, the more galaxies should be included in the training set, i.e. the sparsity of the problem has to be evaluated carefully.

For relatively simple galaxy shapes, like elliptical galaxies, the pre-selection may focus on galaxies with similar sizes and ellipticities, which ensures better morphological similarities. This usually results in a satisfactory subtraction of the lens galaxy with only few PCA components. However, the window in which the sizes and ellipticities are chosen has to be wide enough to allow a full representation of any shapes of galaxies in this range.

2
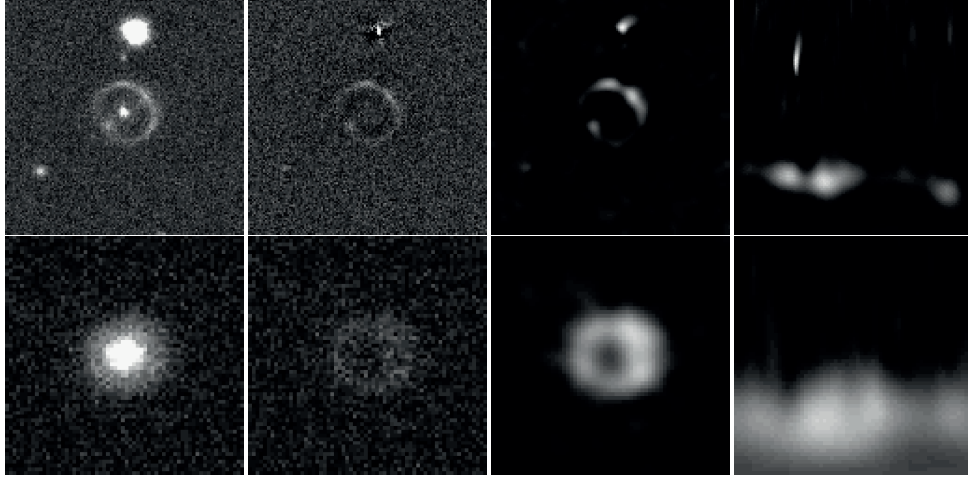
Fig. 2. Illustration of the ring finding process for two simulated Einstein rings from the Bologna Lens Factory (Sect. 4). For each row, from left to right are shown 1- an example of simulated Einstein ring (64×64 pixels), along with its lens galaxy, 2- the lensed ring after PCA subtraction of the foreground galaxy, 3- the result of curvelet denoising, 4- the polar transform of the ring revealing a well visible horizontal line which position along the y-axis gives a measurement of the radius of the Einstein ring.
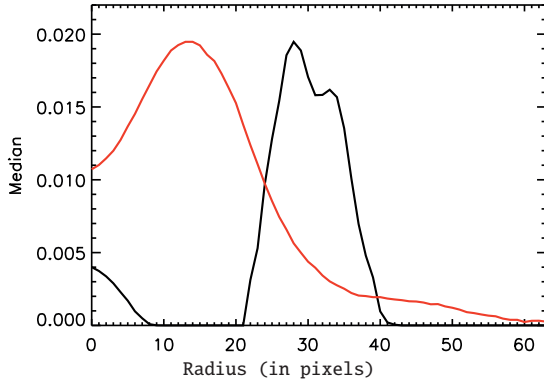
Fig. 3. Median pixel values along the pixel rows of the curvelet-filtered images shown in the third column of Fig. 2. The black line corresponds to the top row of Fig. 2 and the red line corresponds to the bottom row. A simple thresholding scheme allows us to detect the spike and to measure directly the size of the Einstein ring (see text).

The choice of this selection window is discussed later when applying the method to specific data.

Computational time is an important parameter to consider as well. Building the PCA basis involves finding the eigenvectors and the eigenvalues of a $n^2 \times N_{gal}$ matrix, where n is the number of pixels per stamp and where N_{gal} is the number of stamps in the training set.

2.3. Building the PCA basis

Before computing the PCA basis, we rotate all the galaxies in the training set so that their major axes are all aligned and we cen-ter the galaxies in each stamp image. The rotation is performed using a polynomial transformation and a bilinear interpolation. This restricts further the parameter space to be explored and is fully justified given that position angle of galaxies on the sky distribute in a random way: the position angle cannot be a principal component. Note that we do not apply any other re-scaling, e.g. of parameters such as ellipticity, which do not distribute in a random way.

Any companions to the galaxies used to build the PCA basis are a possible source of artefacts. Companion galaxies are frequent enough to have an important weight in the final basis. This can result in removing part of the lensed object at the end of the process or, conversely, to create fake lensed objects.

In order to avoid this effect, we select only galaxies with no bright companions or with companions far away from the center of light. This method results of course in reducing the size of the PCA basis. To include more "companion-free" galaxies, one often has to widen the original selection function, at least in surveys of limited volume, and this may results in a PCA basis less representative of the considered sample. The selection also involves reducing the efficiency of the removal of galaxies with companions. In order to search for strong lensing around that peculiar kind of morphologies, one can devise a masking strategy, but this has not been considered in the present study.

The PCA analysis is computed by building a matrix $\mathbf{X_b}$ in which each of the n columns is an image from the basis set, reshaped as a vector of size n^2. A singular value decomposition is performed on the covariance matrix of the elements of the basis, $\mathbf{X_b}$, which boils down to find V, and W verifying

$$\mathbf{X_b^T X_b} = \mathbf{VWV^T}, \tag{1}$$

where $\mathbf{W}$ is a diagonal matrix. The singular value decomposition of $\mathbf{X_b}$ is written

$$\mathbf{X_b} = \mathbf{U\Omega V^T}, \tag{2}$$

with $\mathbf{\Omega^2} = \mathbf{W}$, and $\mathbf{U}$ the matrix of the eigenvectors for the decomposition of $\mathbf{X_b}$. Therefore, the eigenvectors $\mathbf{Ei}$ can be re-

3

covered from the singular value decomposition of the covariance matrix

$$\mathbf{Ei} = \mathbf{X_b V^t W^{-1/2}}. \tag{3}$$

The decomposition of an $n \times n$ image of galaxy reshaped as a column vector, $\mathbf{X_{set}}$ (not necessarily in the basis) can now be decomposed as

$$\alpha_{\mathbf{set}} = \mathbf{Ei^T X_{set}}, \tag{4}$$

where $\alpha_{\mathbf{set}}$ is a $\mathbf{N_{gal}}$-sized vector of PCA coefficients that represents the image $\mathbf{X_{set}}$.

A partial reconstruction of the image is done by using only the k-first coefficients of the PCA, i.e. the k most significant coefficients. The estimated reshaped image is

$$\tilde{\mathbf{X}}_{\mathbf{set}} = \mathbf{Ei}_{[0..n^2,0..k]} \alpha_{\mathbf{set}[0..k]}. \tag{5}$$

As the basis does not represent anything but the variations in shapes of the central parts of the galaxies, they will be the only reconstructed objects. The remaining companions are much less represented in the PCA basis. Rare structures such as Einstein rings or multiply imaged objects are very little represented in the PCA basis. Using a limited number of PCA coefficients during the reconstruction will therefore create images of lens galaxies without any significant lensed structure potentially present in the original data. The reconstructed PCA images can therefore be subtracted from the original data in order to unveil the lensing structures, when present. Fig. 1 displays examples of the first PCA coefficients for the simulated Einstein rings described in Section 4.

In order to evaluate the quality of reconstruction in an objective way, we compute the reduced χ^2 (per pixel) of the reconstruction in some circular area S containing N_S pixels:

$$q = \frac{1}{N_S} \sum_{i=1}^{N} \left[\frac{d_i - m_i}{\sigma_i^2} \right]^2 \tag{6}$$

where d_i are the pixels in the original image along with their photometric error σ_i, and where m_i are pixel values as predicted by the PCA model/reconstruction. The radius of the circular area S can be chosen to match the mean size of the galaxies in the sample.

A critical step in the PCA reconstruction is the choice of the number of PCA coefficients. If all of the coefficients are used, the reconstruction will include elements of the basis that represent the noise, hence resulting in an overfitting of the data and to an apparent smoothing of the residual image obtained after subtraction of the galaxy. This can be damaging when trying to detect faints rings and arcs. Conversely, if the number of coefficients is insufficient the central galaxy will be only partially removed leaving significant and undesired structures in the residual image.

In Section 4, we describe a way to choose the number of PCA coefficients in an objective way, using the reduced χ^2 and we illustrate the effect of this choice using a set of simulated Einstein rings, as they would be seen with the ESA Euclid satellite (Laureijs et al. 2011).

3. Finding the lensed images, arcs and rings

Once a galaxy is removed from the image, the second step is to search for any residual lensed signal. In this paper, we focus on partial or full Einstein rings. We investigate two different approaches. The first one uses a curvelet filter (Starck et al. 2002),
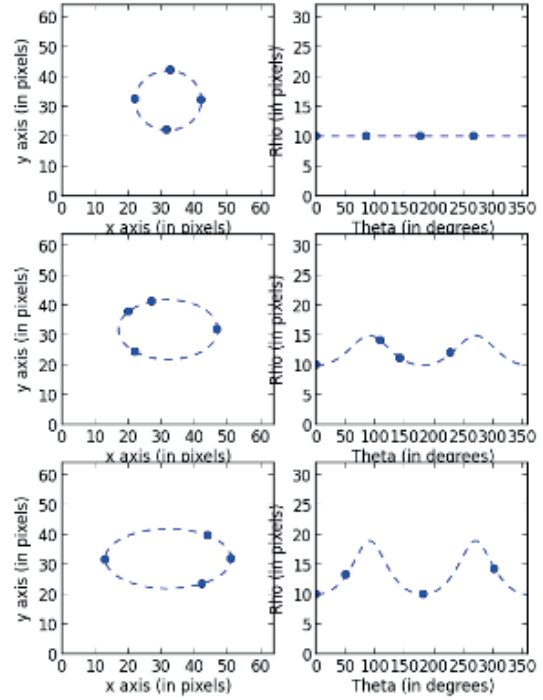
Fig. 4. *Left panels*: schematic view of rings (dashed line) and multiple images (blue dots along the ring tracks). *Right panels*: their corresponding transform in polar coordinates.

optimized to enhance any arc-like structure, on images reshaped in a polar grid. The second method uses SExtractor (Bertin & Arnouts 1996) to identify remaining sources in the residuals and to assess whether they are lensed images according to their orientation and elongation.

3.1. Polar transform

A simple way to detect full or partial rings can be devised by turning the Cartesian coordinate system of the data into the polar one. The polar coordinates (ρ, θ) are chosen so that the origin is the center of the galaxy that has been removed using the PCA decomposition. The polar-transformed image is built by creating a new grid of pixels and by asking, for each pair of (ρ, θ) coordinates, the value of the pixels in the original (x, y) Cartesian grid. This involves an interpolation process giving the pixel intensities $I_{\text{pol}}(\rho, \theta)$ as a function of the pixel intensities in the original image $I(x, y)$, with the standard relations $x = \rho \cos(\theta)$ and $y = \rho \sin(\theta)$.

By construction, the polar transform centered on the lens galaxy barycenter, turns a circle into a line, as illustrated in Fig. 2. The problem of ring detection is then reduced to a problem of line detection. The polar image's columns are collapsed into a vector containing the median value of each column. If the original image contains a ring, this vector will present a spike, whose position directly gives the radius of the ring, as illustrated in Fig. 3. In practice, we define a threshold that determines if the maximum of the vector stands for a ring or not. Figs. 2 & 3 show the different steps of the ring detection.

4

R. Joseph et al.: PCA-based lens finder

As the rings are not always perfectly circular but elliptical, their shape in polar coordinates can deviate significantly from a straight line, as is the case in Fig. 2. In most cases, looking for straight lines in polar coordinates is sufficient to detect rings, at least for moderate ellipticities. However, it is possible to refine the detection criterion by fitting an ellipse in polar coordinates,

$$\rho(\theta) = \frac{ab}{\sqrt{(b\cos\theta)^2 + (a\sin\theta)^2}}, \tag{7}$$

where a and b are the semi-major and semi-minor axes of the ellipse and where the origin of the system is centered on the lensing galaxy. In order to find point-source components superposed to the rings (or simply lensed point sources), one can add simple Gaussian profiles to the fit or the actual instrumental/atmospheric PSF. Alternatively, one can implement the detection scheme of Meneghetti et al. (2008) to find brightness fluctuations along the arcs. Different typical lensing configurations are shown to illustrate this in Fig. 4.

3.2. Island finding: the use of SExtractor parameters

An alternative method for assessing the presence of lensed structure in fields is to characterise all sources in the field, and use the measured parameters of these sources in order to identify patterns among them. This process begins with the use of SExtractor to identify sources in the field above a signal-to-noise threshold. The flux, ellipticity, tangentiality (closeness of the position angle to $90°$ to a vector from the field centre to the object), and distance from the field centre are measured. In addition, flux islands (which may contain one or more SExtractor components) are identified and the third moments of the flux distribution are measured. Third moments are sensitive to bent or arc-like structures, which are hard to detect from single components alone. For the current purpose, we define a combination of third moments ζ as:

$$\zeta = \frac{1}{2}\log_{10}\left[(\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2\right], \tag{8}$$

where

$$\mu_{mn} = \sum_{n,m} d(x,y)x^m y^n \tag{9}$$

where $d(x,y)$ is the data value in terms of offsets x and y from the brightest pixel in the island. This statistics, as a combination of third moments is sensitive to bending and is also invariant under scaling and rotation.

A Point score is then assigned to each component according to the elongation of the component and its tangential orientation with respect to the field centre. In addition, components with similar radii are weighted upwards in the point score allocation, and components which are part of an island with significant third moment are also weighted up. Specifically, the point score is given by the following procedure, using free parameters p_i where necessary:

- Each component, unless it has a flux less than a threshold p_0, is assigned a point score of $10\epsilon^2 exp(-t^2/p_1^2)$, where $\epsilon \equiv a/b$ is its elongation and t is the difference between its tangentiality and the angle tangential to the radius vector to the point. In general, we use Gaussian penalty functions where we wish to select for a value close to one which would be expected for lensing, and power laws for quantities which we wish to maximise. The ϵ^2 dependence results from a limited amount

of experimentation by hand, although such dependencies can ideally be optimized on a larger sample.

- The point score of any component within a factor of p_2 in radius from its neighbour is multiplied by $(1.0 + N/p_3) * exp[-(r-1)^2/p_4^2]$, where N is the number of points assigned to the neighbour, and r is the ratio of their distances from the centre of the field. This selection favours multiple lensed images at the same radius, although the selection will have more effect if the individual images are themselves elongated and tangential.
- If a component is part of an island with third moment $\zeta > p_5$, its point score is multiplied by $[1 + (\zeta - p_5)]^2$.

The six parameters p_i are then optimized on a small training set of lenses before being applied to the dataset. A variable point-score threshold can be used for lens detection, completeness generally being achieved at the expense of purity of the resulting sample.

4. Application to Euclid-like simulated images

The "lens finder" described in Sect. 2 is designed to process large imaging data sets. Although the pre-selection of the galaxies to be searched for lensing may require color information, the new algorithm proposed in this paper can be applied to single-band data to perform a purely morphological search. In the following, we evaluate the performances of the method using simulated images of Einstein rings, as they would be seen with the ESA Euclid satellite (Laureijs et al. 2011).

The image simulations are provided through the Bologna Lens Factory (BLF) project[1]. This is a project dedicated to performing lensing simulations and providing realistic mock data for a large variety of lensing studies from large scale weak lensing, to galaxy cluster lensing and strongly lensed quasars. For the purposes of this work, images were created to specifically mimic the expected Euclid images in the visible instrument, as described in Laureijs et al. (2011). The pixel size is $0.1''$ and the PSF is Gaussian with a Full-Width-Half-Maximum (FWHM) of $0.18''$. The surface brightness is translated into photon counts taking into account the expected instrumental throughput in the VIS band. Background counts from zodiacal light are added, assuming a brightness equal to 22.8 $mag/arcsec^2$. Noise is then calculated taking care of Poisson statistics, flat-field error and read-out (Meneghetti et al. 2008). The lensing and image construction is done with the GLAMER lensing code (Metcalf & Petkova 2013; Petkova et al. 2013). The pre-lensed galaxy surface brightness models and mass distribution are provided by the Millennium Run Observatory (MRObs; Overzier et al. 2013). Each galaxy is represented by a bulge and a disk component whose properties are predicted by a semi-analytic galaxy evolution model. The mass distribution consists of halos identified in the Millennium Nbody simulation.

The lensing simulations were done as follows. The halos in the catalog are represented by NFW halos (Navarro et al. 1997) with Singular Isothermal Ellipsoids (SIEs) in their centers to represent the baryonic galaxy. This model has been shown to fit observed Einstein rings well (Gavazzi et al. 2007). The NFW profile is fit to the mass and peak circular velocity of the halo found in the Millennium simulation. The mass and velocity dispersion of the SIE component is set by the stellar mass to halo mass relation of Moster et al. (2010) and the Faber-Jackson relation (Faber & Jackson 1976). The lensed image of every source

---

[1] www.bolognalensfactory.wordpress.com

5
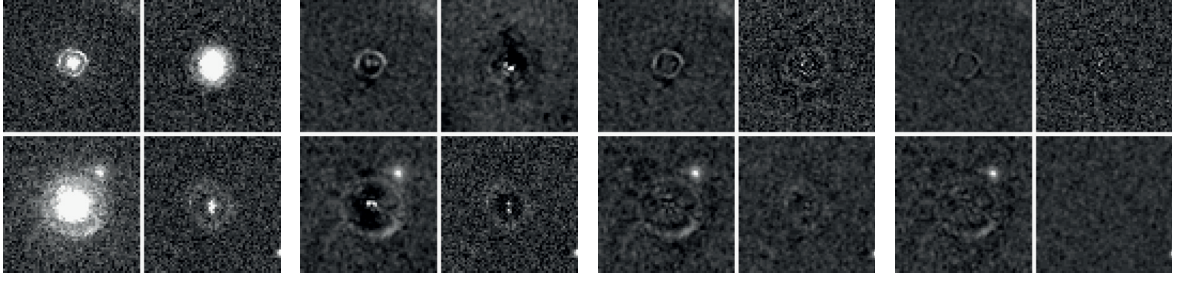
Fig. 5. Result of the galaxy removal on four of our simulated Einstein rings. The left hand side panel displays the four original images. From left to right, the other panels display galaxy removals when 10, 50 and 200 PCA coefficients are used. The reduced χ^2 are respectively $q = 1.74$, $q = 1.00$ (i.e. optimal number of coefficients), and $q = 0.9$.
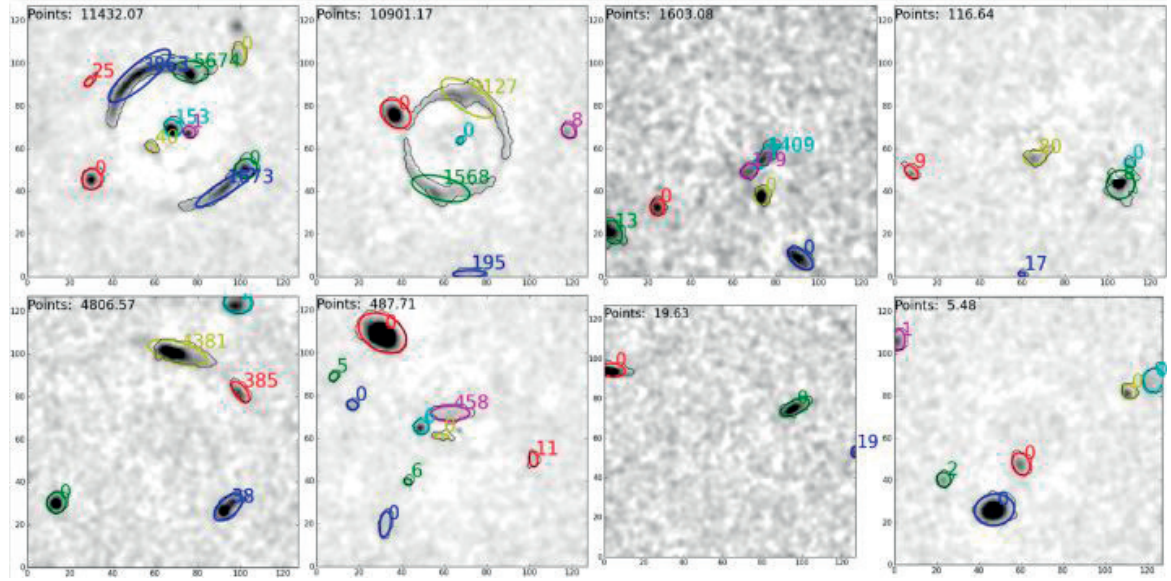
Fig. 6. Results of the island-finding algorithm. Each panel shows the residual image after the PCA galaxy subtraction, with the point score of each component given separately, and the total point score at the top (see text). The top row shows systems which have lenses, and is ordered so that the highest point-score is on the left and the lowest on the right. Objects with high ellipticity and high curvature, tangential to the radius vector from the centre of the image, are highly preferred; lens systems without such objects are hard to recognise by eye and also tend to attract a lower point score. The bottom row shows a sample of non-lenses, again ordered by point score. High point-score objects are generally those in which chance coincidences produce configurations which mimic the presence of lensing.

within a 0.1 deg^2 light cone down to 28th magnitude in I band is constructed and put into a master image. This image contains only a few strongly lensed objects because the source density is small enough that it is rare to have a visible object within a caustic. To boost the number of strong galaxy-galaxy lenses, all the critical curves and their associated caustics in the field are found for a source redshift of $z_s = 2.5$ and a source galaxy is moved to be near the caustic. The sources are taken randomly from galaxies within the light cone at a similar redshift. Then the lensed image of this source is constructed and added to a 200×200 pixel cutout stamp from the master image. Images with and without the added source are provided and an image with only the added, lensed source are provided. All images are

provided with and without the noise and PSF effects. A catalog of all the critical curves and caustics is also provided with their locations and properties such as average radius and area.

Since we are not concerned with predicting the statistical properties of the lenses in this paper, many of the precise details of these simulations are not important (for example the distribution of source and lens redshifts, morphologies, luminosities, etc.). The performance of the PCA lens finder will be stated in terms of the signal-to-noise ratio of the Einstein ring so the simulations are only required to represent the variety of expected lenses and not their precise distribution.

The set of Euclid simulation images consists of 3000 galaxies with a full or partial background Einstein ring and of a train-

ing set of 1250 galaxies with no lensing. Adding more galaxies to the training set does not change significantly the PCA basis. Among the 1250 non-lensing galaxies of the training set, 1000 are used to build the PCA basis in order to search for lensing in the 3250 images, 3000 of which containing Einstein rings. Note that with real data, the training set can be the whole data set itself, as galaxies with lensing features are rare.

Building the PCA basis for the 1000 Euclid galaxies, which are 128 pixels on-a-side, takes about 40 minutes on a single processor. Using this PCA basis, doing the galaxy reconstruction and subtraction takes less than a minute more for the whole data set, i.e. 3250 images. In terms of cpu, the PCA method is therefore well tractable and applicable to large data sets.

4.1. Quality of the central galaxy reconstruction

The quality of the PCA reconstruction depends on 3 main factors: 1- the range in galaxy sizes, 2- the presence of companions near the galaxies used to build the PCA basis, 3- the number of PCA coefficients to be used.

In order to minimize the parameter space to explore, all galaxies are first centred on the central pixel of the FITS stamp and rotated so that their long-axis aligns with the image rows. If necessary, the resulting images are zero-padded and trimmed to a common size. In the present case we use 128×128 pixels.

In order to minimize the contamination of the PCA basis by companions to the galaxies in our sample, we only select the stamps that have no companion brighter than 50% of the maximum brightness of the main galaxy in a range of less than 10 pixels to the patch's center, i.e. 1″ given the Euclid pixel size of 0.1″.

To estimate the number of PCA components, we carry out different reconstructions with an increasing number of PCA coefficients. We stop adding coefficients when reaching an acceptable quality, i.e. when there is no residual above the noise level. A good reduced χ^2 is when q, (Eq. 6) remains between 1 and 1.5, i.e. when the mean χ^2 per pixel is on average close to 1σ. Indeed, if the pixels in the residuals are highly correlated due to a reconstruction that includes coefficients representative of the noise, the reduced χ^2 becomes smaller than 1. Conversely, when the residuals contain important patterns due to an insufficient reconstruction, q is significantly larger than 1. This is illustrated in Figs. 7 & 8 for the specific case of our Euclid simulation, where a good reconstruction is achieved for a number of PCA coefficients of about 50, i.e. the minimum number of coefficients required to reach $q \sim 1$.

4.2. The effect of galaxy sizes

Even for relatively smooth light distributions, like early type galaxies, a careful balance must be found between the number of galaxies in the training set and the range in galaxy sizes. We investigate in the following the influence of the distribution of the galaxies in sizes for the specific case of our Euclid simulations.

To do so, we bin the sample in galaxy sizes, keeping 100 galaxies per bin and we build the PCA basis for each bin of size, i.e. like in Fig. 7. Note that rescaling the galaxies in R_{eff} is also an alternative, but we try as much as we can to avoid alter the data before building the PCA basis. Rescaling in R_{eff} may be considered for small samples of galaxies that cannot be binned in galaxy size. The images are then reconstructed using different number of coefficients. The quality of reconstruction, estimated
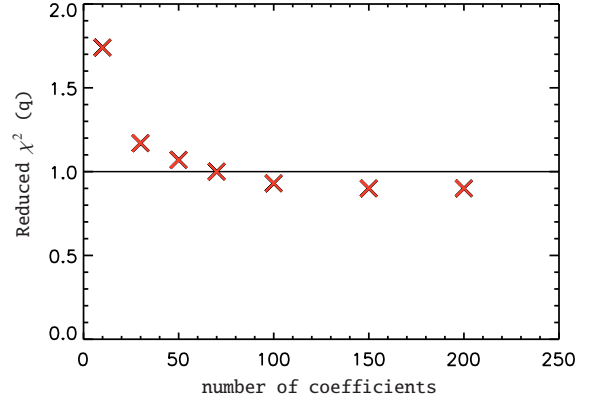
Fig. 7. Reduced χ^2, as a function of the number of coefficients used in the reconstruction. Only 50-70 coefficients are needed to reach a reduced χ^2 of $q \sim 1$ in the case of our Euclid simulations.

using the median q factor over all images of the sub-sample, is then evaluated. Fig. 7 suggests that 50-70 coefficients is an optimal number to reach a reduced χ^2 close to 1.

Fig. 8 shows how q rises when galaxies are getting bigger than a semi-major axis bigger than 3 pixels. As big galaxies are less represented in the PCA basis, because of their scarcity, their reconstruction is less accurate, hence leading to a larger χ^2.

It is therefore very important to carefully select the range of size that we want to investigate when building the PCA basis and to ensure that a sufficient number of galaxies are available to represent the full variety of structures in the sample/bin. Indeed, for bigger galaxies, where Einstein rings are more likely to be found, the number of objects contributing to the basis is reduced, simply because big galaxies are rare.

4.3. Completeness and purity

In order to evaluate the efficiency of the algorithm, we perform tests of detection on simulated images for which the signal-to-noise ratio and the caustic radius of the lensing galaxies are known. For this study we use a set of 3000 simulated full rings from the BLF. With these realistic Euclid-like ring images and the associated noise images we can compute the SNR for each Einstein ring:

$$SNR = \frac{S}{\sigma \sqrt{N_i}}, \qquad (10)$$

where N_i is the number of non-zero pixels in the noise free ring image, σ is the rms noise per pixel and S is the total flux in the ring. The analysis of the simulated images is done by building a PCA basis using 1000 galaxies from a set of non lensing galaxies. The detection algorithms, described in Section 3 are then applied to the 3000 images with lensing and to the 250 images without lensing . The island finding algorithm has been trained on a set of 167 images of lensed rings provided by the BLF, together with another set of 200 images which did not contain lenses. The parameters were optimized here, and then re-optimized on the dataset itself. The output of the process is compared with the known answer from the simulations to evaluate the completeness and the purity of the derived lens catalogues.
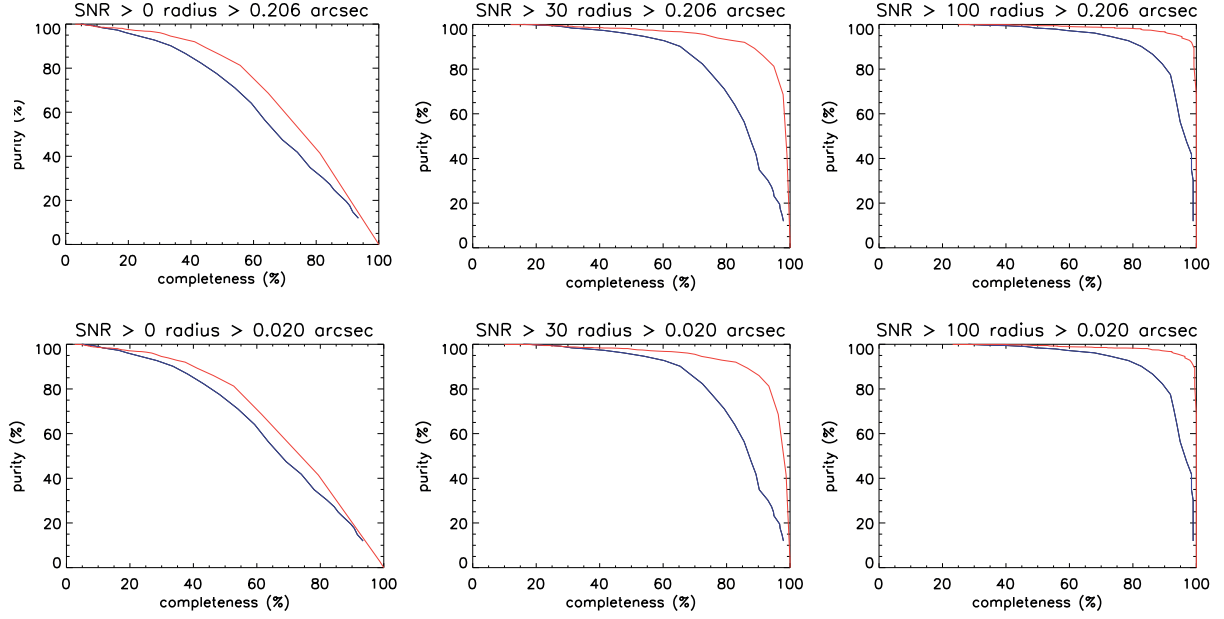
Fig. 9. Completeness as a function of purity for different thresholds of Einstein radii (expressed in terms of critical curve here) and signal-to-noise ratio with the two methods described in Sect. 3: polar transform (in red) and island finding (in blue). The minimal radius in the sample is $r = 0.02''$, which means that the top left panel shows the results over the whole sample.
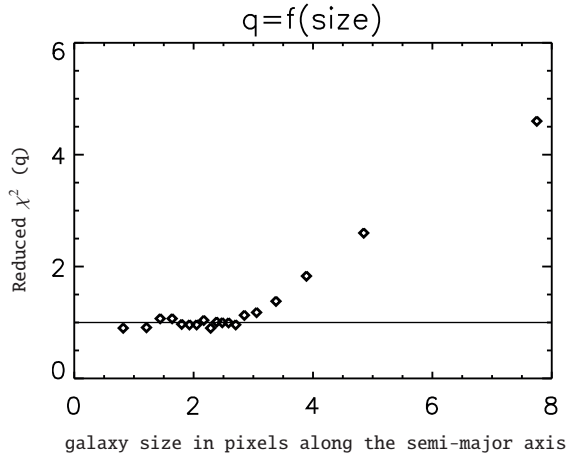
Fig. 8. Quality of the reconstruction of the simulated Euclid lenses as a function of the average size of the galaxies in pixels, as measured with SExtractor. The pixel size of the images matches that of Euclid, i.e. 0.1″. As big galaxies are rare, they are less well represented in the PCA basis and they are therefore less well modeled.

As the fraction of non-lens images in the sample is small compared to reality, we rather define the purity as the fraction of non-lens images that have not been detected instead of the

fraction of true positive among all the detected lensed images:

$$\text{Purity} = 1 - \frac{N_{\text{false positive}}}{N_{\text{false positive}} + N_{\text{true negative}}}. \quad (11)$$

The completeness is expressed as the fraction of actual lens images that have been detected over the whole sample of lenses:

$$\text{Compl.} = \frac{N_{\text{true positive}}}{N_{\text{true positive}} + N_{\text{false positive}}}. \quad (12)$$

Fig. 9 shows the purity as a function of completeness for both methods. Different thresholds in signal-to-noise ratio and critical curve for the lensing have been considered. Although both methods are comparable at low completeness, at high completeness levels the SExtractor algorithm generally leads to lower purity, corresponding to more false positives. This problem appears worse at high signal-to-noise levels, because the number of false positive detections in the non-lens sample remains constant while the number of true positives declines. This is likely to be due to the attempt to preserve at least some sensitivity to only marginally extended components, corresponding for example to quadruply imaged sources of modest extent. The algorithm is therefore more vulnerable to chance alignments between external components; work is under way to alleviate this problem, and particularly to use colour information to distinguish between genuine and chance alignments. In the context of the present work, we stick to single-band detections. The results tend to show that we can detect rings almost independently on the radius. For instance, with the polar transform method and a signal-to-noise ratio higher than 30, one can reach a completeness of 90% for a purity of 86%.
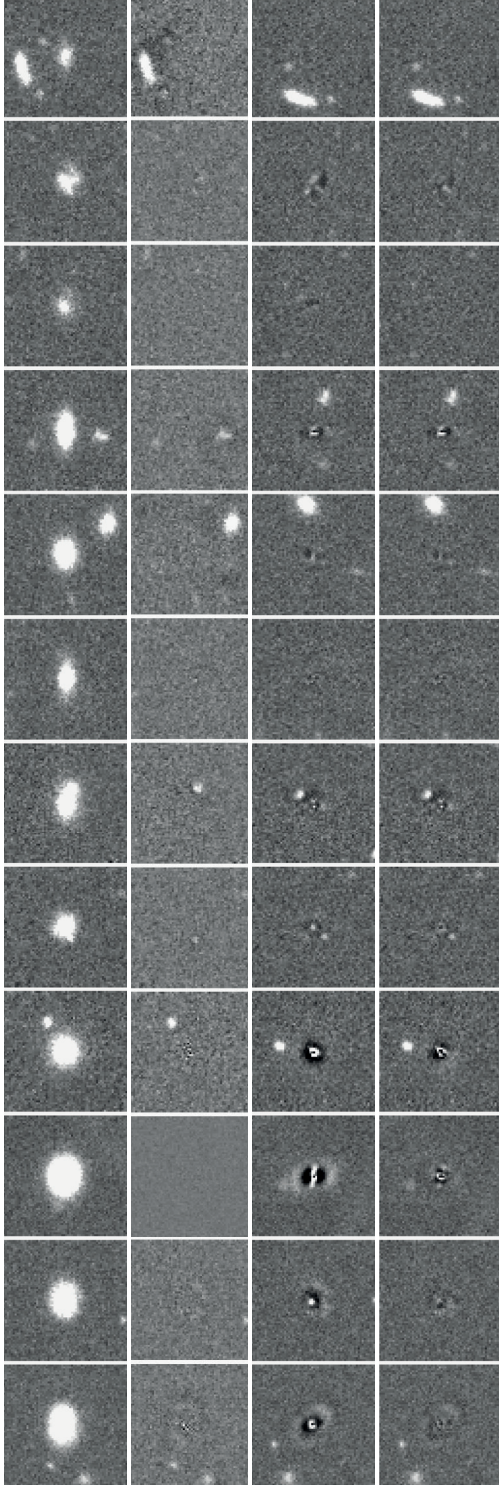
Fig. 10. Comparison of different galaxy-removal schemes applied to deep CFHT images. The first columns shows the original image. The second shows the residual image after subtraction of a PCA reconstruction of the galaxy. The third and fourth columns show the subtraction of a single and double elliptical sersic profile respectively, using `GALFIT3`. Note that the PCA-subtracted images are rotated by construction of the PCA basis but the `GALFIT`-subtracted images are not, in order to avoid interpolation when not mandatory.

5. Application to real data

In the above, we test our lens finder on simulated images that mimic Euclid images in the VIS band. An obvious question is whether the algorithm performs in a satisfactory way on real data. While carrying out a ring search on a large data set is outside the scope of this paper, we can nevertheless test how our PCA decomposition of galaxies compares with other more traditional ways of removing lensing galaxies.

In order to do that, we use the deep and sharp optical images taken with MEGACAM at the CFHT to map SDSS stripe 82. Following the same procedure as with the Euclid simulations, we set the optimal number of PCA coefficients by checking that we can actually reach reduced χ^2 $1 < q < 1.2$ depending on the seeing and on the physical size of the galaxies we want to subtract.

In Fig. 10, we compare our galaxy subtraction with that done in other lens searches using single or double Sersic profiles (e.g. Vegetti et al. 2012; Lagattuta et al. 2010). Not surprisingly, the subtraction with Sersic profiles performs rather well with low SNR galaxies or with small galaxies, but leaves significant residuals for large galaxy sizes. As these residuals often take the shape of a ring, they may lead to large numbers of false positives in a ring search.

The experiment we carry out here with real data uses only 1 single field of the CFHT data of stripe 82, i.e. 1 square degree out of the 180 available. This means that the PCA decomposition uses only a limited number of large galaxies. As a consequence, using the whole 180 fields has the potential to improve further the galaxy subtraction, while profile fitting will always be limited to the information in one single galaxy and does not benefit from the global information on the shape of galaxies from a whole data set. In other words, increasing the survey size, not only increases the number of potential lenses, but also increases the density of galaxies per bin of size, hence improving the quality of the PCA basis.

6. Conclusion

The two lens finder algorithms developed here all rely on a good subtraction of lensing galaxies with machine learning methods; different ideas for ring detection then allow objects with different properties to be detected on the residual images:

- The polar transform method enhances the signal in the residual image by applying curvelet denoising and uses a polar transform of the images to turn the problem of a circle detection to a line detection. It is designed to detect full or partial rings with or without ellipticity.
- The "Island finding algorithm" uses `SExtractor` to detect structures in the PCA-subtracted images and to determine whether they correspond to lensed sources according to their elongation, orientation and bending. This algorithm is expected to be more efficient in finding partial arcs and multiple images.

The method is successfully applied to Euclid-like simulations. With the polar transform method, a completeness of 90% is reached for data where the signal-to-noise in the Einstein ring is at least 30. The same simulations show that the purity of the derived ring sample reaches 86% of the non lensed galaxies detected as false positives.

The galaxy subtraction algorithm occurs to be efficient when applied to real data as well: our tests with CFHT images of SDSS

Stripe 82 surpasses in quality the subtraction obtained with direct model fitting.

In future work, ways to increase the purity of the algorithms will be investigated by using adapted dictionary learning (e.g. Beckouche et al. 2013) for galaxy subtraction. The strength of those machine learning methods should allow us to build bases adapted to more complicated problems, such as the subtraction of galaxies in clusters to detect rings produced by multiple galaxies. Better morphological selection based on PCA "clustering" or beamlet analysis (e.g. Donoho & Huo 2002) can be used to discriminate ring-like shapes, to classify rings and arcs and to carry out galaxy classification in general, as has been done in the past with quasar spectra (Boroson & Lauer 2010) and, more recently, with galaxy multi-band photometry (Wild et al. 2014).

References

Alard, C. 2006, astro-ph/0606757
Bartelmann, M., Limousin, M., Meneghetti, M., & Schmidt, R. 2013, Space Sci. Rev., 177, 3
Beckouche, S., Starck, J. L., & Fadili, J. 2013, A&A, 556, A132
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Boroson, T. A. & Lauer, T. R. 2010, AJ, 140, 390
Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41
Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, A&A, 461, 813
Courbin, F., Faure, C., Djorgovski, S. G., et al. 2012, A&A, 540, A36
Donoho, D. L. & Huo, X. 2002, in Lect. Notes Comput. Sci. Eng., Vol. 20, Multiscale and Multiresolution Methods Theory and Applications, ed. et al. & T. J. Barth (Springer), 149–196
Faber, S. M. & Jackson, R. E. 1976, ApJ, 204, 668
Foëx, G., Motta, V., Limousin, M., et al. 2013, arXiv1308.4674
Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385
Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., & Ruff, A. 2012, ApJ, 761, 170
Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, ApJ, 667, 176
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
Hoekstra, H., Bartelmann, M., Dahle, H., et al. 2013, Space Sci. Rev., 177, 75
Jolliffe, I. T. 1986, Principal Component Analysis (Berlin; New York: Springer-Verlag)
Kneib, J.-P. & Natarajan, P. 2011, A&A Rev., 19, 47
Lagattuta, D. J., Auger, M. W., & Fassnacht, C. D. 2010, ApJ, 716, L185
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv1110.3193
Limousin, M., Cabanac, R., Gavazzi, R., et al. 2009, A&A, 502, 445
Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544
Marshall, P. J., Hogg, D. W., Moustakas, L. A., et al. 2009, ApJ, 694, 924
Meneghetti, M., Bartelmann, M., Dahle, H., & Limousin, M. 2013, Space Sci. Rev., 177, 31
Meneghetti, M., Melchior, P., Grazian, A., et al. 2008, A&A, 482, 403
Metcalf, R. & Petkova, M. 2013, submitted, arXiv:1312.1128
More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Overzier, R., Lemson, G., Angulo, R. E., et al. 2013, MNRAS, 428, 778
Parker, L. C., Hoekstra, H., Hudson, M. J., van Waerbeke, L., & Mellier, Y. 2007, ApJ, 669, 21
Pawase, R. S., Faure, C., Courbin, F., Kokotanekova, R., & Meylan, G. 2012, arXiv1206.3412
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2011, GALFIT: Detailed Structural Decomposition of Galaxy Images, ascl1104.010
Petkova, M., Metcalf, R., & Giocoli, C. 2013, submitted, arXiv:1312.1536
Ruff, A. J., Gavazzi, R., Marshall, P. J., et al. 2011, ApJ, 727, 96

Seidel, G. & Bartelmann, M. 2007, A&A, 472, 341
Simon, P., Schneider, P., & Kübler, D. 2012, A&A, 548, A102
Sonnenfeld, A., Gavazzi, R., Suyu, S. H., Treu, T., & Marshall, P. J. 2013a, ApJ, 777, 97
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013b, ApJ, 777, 98
Starck, J.-L., Candès, E., & Donoho, D. 2002, ITIP 11, 131–141
Sygnet, J. F., Tu, H., Fort, B., & Gavazzi, R. 2010, A&A, 517, A25
Treu, T., Dutton, A. A., Auger, M. W., et al. 2011, MNRAS, 417, 1601
Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341
Wild, V., Almaini, O., Cirasuolo, M., et al. 2014, ArXiv1401.7878

3.4   Application to the CFHTL Survey

In a subsequent paper, of which a pre-print is reproduced in the following pages, we improved the PCA deblending technique by applying a masking strategy to bright companions. In our previous paper (Joseph et al., 2014), we described how galaxies with companions had to be avoided when building the PCA basis set. In this new iteration of the method, bright companions are masked when selecting galaxies to build the basis set, which allows for more galaxies to be selected in the building of the basis without losing in terms quality of the reconstruction. Bright companions are also masked in test set images before applying the PCA reconstruction. This strategy was proven very efficient at removing central artefacts in the PCA-subtracted images.

The new version of the lens finding algorithm, developed by co-authors Andrea Tramacere and Danuta Paraficz is a hybrid of the previously developed techniques. First a `DBSCAN` procedure (Easter et al., 1996; Tramacere & Vecchio, 2013) is ran on the PCA-subtracted images that identifies clusters of connected bright pixels. This allows to make a selection of ring apertures that contain lensed source candidates and, instead of searching for a line in the polar domain, fit a circle to the source image candidates.

This improved algorithm was then applied to the CFHTLS data. The algorithm produced a sample of 1098 lenses that we visually inspected to converge to a sample 109 new lens candidates along with 86 known lens systems discovered in previous lens searches.

The PCA Lens-Finder: application to CFHTLS

D. Paraficz[1], F. Courbin[1], A. Tramacere[2], R. Joseph[1], R.B. Metcalf[3], J.-P. Kneib[1], P. Dubath[2], D. Droz[1], F. Filleul[1], D. Ringeisen[1], C. Schäfer[1]

[1] Laboratoire d'Astrophysique Ecole Polytechnique Fédérale de Lausanne (EPFL) Observatoire de Sauverny CH-1290 Versoix
[2] Department of Astronomy, University of Geneva, Ch. d'Écogia 16, CH-1290 Versoix, Switzerland
[3] Dipartimento di Fisica e Astronomia - Universita di Bologna, via Berti Pichat 6/2, 40127, Bologna, Italy

May 2, 2018

ABSTRACT

We present the results of a new search for galaxy-scale strong lensing systems in CFHTLS *Wide*. Our lens-finding technique involves a preselection of potential lens galaxies, applying simple cuts in size and magnitude. We then perform a Principal Component Analysis of the galaxy images, ensuring a clean removal of the light profile. Lensed features are searched for in the residual images using the clustering topometric algorithm DBSCAN. We find 1098 lens candidates that we inspect visually, leading to a cleaned sample of 109 new lens candidates. Using realistic image simulations we estimate the completeness of our sample and show that it is independent of source surface brightness, Einstein ring size (image separation) or lens redshift. We compare the properties of our sample to previous lens searches in CFHTLS. Including the present search, the total number of lenses found in CFHTLS amounts to 678, which corresponds to ~4 lenses per square degree down to $i(AB) = 24.8$. This is equivalent to ~ 60.000 lenses in total in a survey as wide as Euclid, but at the CFHTLS resolution and depth.

Key words. gravitational lensing, galaxies, surveys

1. Introduction

Strong gravitational lensing occurs when light rays emitted by a distant source are deflected by the potential well of a foreground mass (Einstein 1916). If the latter is compact enough i.e. above the lensing critical surface mass density, multiple images of the background source are formed. Because strong lensing has only simple dependence on its geometry and fundamental physics (general relativity), thus its applications in cosmology and in the study of galaxy formation and evolution are straightforward and becoming more and more numerous.

Strong gravitational systems have been used in recent years to address key scientific questions. In particular, strong lensing consists of a powerful tool to map the total mass (dark and luminous) in galaxies, independently of their dynamical state or star formation history (e.g. Treu et al. 2002; Rusin et al. 2003; Treu & Koopmans 2004; Rusin & Kochanek 2005; Sonnenfeld et al. 2012; Bolton et al. 2012). Also, thanks to strong lensing, small and dark satellite galaxies have even been found and weighted (e.g. Metcalf & Madau 2001; Dalal & Kochanek 2002; Treu & Koopmans 2004; Koopmans et al. 2006; Jiang & Kochanek 2007; Gavazzi et al. 2007; Treu et al. 2010; Auger et al. 2010; Bolton et al. 2012; Sonnenfeld et al. 2012; Vegetti et al. 2012; Oguri et al. 2014). Applications in cosmology using the time delay method (e.g. Refsdal 1964) between the multiply-lensed images of distant quasars are also becoming of increasing interest thanks to intensive photometric monitoring programs like COSMOGRAIL (e.g. Vuissoz et al. 2008; Courbin et al. 2011; Rathna Kumar et al. 2013; Tewes et al. 2013). In combination with state-of-the-art modelling tools, these time delays can be used to constrain the cosmological parameters both with precision and accuracy (e.g. Suyu 2012; Suyu et al. 2010, 2013, 2014). Even without time delay measurements, large samples of galaxy-scale strong lenses can help to constrain cosmology, as

Cao et al. (2015) did, using 118 systems to constrain the dark energy equation of state parameter, w.

The above applications of strong lensing are possible because: 1. significantly large samples of strong lensing systems are now available, 2. some of the discovered systems have specific, rare properties making them particularly effective in delivering astrophysical or cosmological constraints. The ongoing (DES, KIDS) and planned wide field surveys of the next decade (Euclid, LSST, WFIRST) will continue to revolutionise the field, by making available hundreds of thousands of new strong lenses (e.g. Pawase et al. 2014; Collett 2015), i.e. ~3 orders of magnitude larger than the current sample sizes.

Early systematic searches for strong lenses took advantage of the so-called lensing magnification bias, i.e. the fact that a lensed source is seen brighter because it is lensed. These source-selected lensing system samples were built by looking for multiple images among samples of optically bright quasars (e.g. Surdej et al. 1987; Magain et al. 1988). This was followed up in a more systematic way with a search in the Hamburg-ESO bright quasar catalogue (Wisotzki et al. 1993, 1996, 1999, 2002, 2004; Blackburne et al. 2008), in the SDSS with the Sloan Quasar Lens Survey (SQLS; Inada et al. 2003, 2012; Oguri et al. 2006, 2008; Inada et al. 2007) as well as in other wide-field optical observations (e.g. Winn et al. 2000, 2001, 2002b,a). Similarly, strong lens searches were also carried out in the radio in the FIRST survey (Gregg et al. 2000) and in the CLASS survey (Myers et al. 2003; Browne et al. 2003). More recently, the same strategy was adopted at millimeter wavelengths with the South Pole Telescope (SPT Hezaveh et al. 2013), and at sub-millimeter wavelengths with the Herschel satellite like H-ATLAS (Negrello et al. 2010; González-Nuevo et al. 2012; Bussmann et al. 2013) and HerMES (Conley et al. 2011; Gavazzi et al. 2011; Wardlow et al. 2013).

1

Source-selected samples of lensing systems are mostly composed of lensed quasars. Searches for non-quasar lensed galaxies are generally carried out by preselecting a sample of potential lensing galaxies and by looking for lensed images or spectra in their background (Ratnatunga et al. 1999; Fassnacht et al. 2004; Moustakas et al. 2007; Cabanac et al. 2007; Belokurov et al. 2007; Faure et al. 2008; Marshall et al. 2009; Pawase et al. 2014; More et al. 2016). These lens-selected samples are best constructed using spectra where sets of emission lines at two (or more) distinct redshifts are looked for. The method was pioneered by (Willis et al. 2005, 2006) with their Optimal Line-of-Sight Survey, soon followed by the SloanLens ACS Survey (SLACS, e.g., Bolton et al. 2006; Treu et al. 2006; Koopmans et al. 2006; Gavazzi et al. 2007, 2008; Bolton et al. 2008; Auger et al. 2009) and by the BOSS Emission-Line Lens Survey (BELLS; Brownstein et al. 2012). The SLACS sample on its own has about 100 confirmed gravitational lenses in the redshift range $0.1 < z < 0.4$ with HST imaging (e.g. Bolton et al. 2006; Auger et al. 2009). The main advantage of the spectroscopic approach is that the redshifts of the lens and of the source are readily obtained, along with the stellar velocity dispersion in the lens (e.g. Treu & Koopmans 2004; Koopmans et al. 2006; Auger et al. 2010). Moreover, if the source has strong emission lines, then the light from the lens and the source can easily be separated.

In the imaging, on the other hand, the source is often hidden in the lens glare, thus it cannot be detected so easily. For this reason carrying out an imaging lens search requires careful image processing to efficiently remove the lens light and unveil any faint background lensed galaxy. Such techniques start to be implemented, and will become increasingly important with the development of large sky surveys like DES, KIDS, Euclid, the LSST and WFIRST.

Based on two-band imaging, Gavazzi et al. (2014) have devised a method to detect faint blue arcs behind foreground redder galaxies. They extend their technique to multi-band lens modelling (Brault & Gavazzi 2015) and they apply them to the CFHTLS (Canada-France-Hawaii Telescope Legacy Survey) data (Cuillandre et al. 2012). A second method was introduced by Joseph et al. (2014) that can work both in single-band and multi-band. It is based on a principal component analysis (PCA; Jolliffe 1986) of the full imaging dataset to subtract the image of galaxies, even in the presence of complex structures. The residual image can then be used to search for lensed sources. In this paper, we use the method of Joseph et al. (2014) to extend the sample of known galaxy-scale strong lenses in CFHTLS.

The paper is organized as follows. In Sect. 2, we provide a brief description of the observational material and of the sample selection technique from the object catalogues for the CFHTLS. In Sect. 3, we describe the lens-finding algorithm based on PCA and its improvements. In Sect. 4, we present the list of our new lens candidates and compare it to previous results from other lens searches in the same area of sky, i.e., the CFHTLS fields. In Sect. 5, we discuss the completeness of the sample based on lens simulation. Finally, in Sect. 6, we provide a summary of the main conclusions from this work and provide a short outlook for future progress.

Throughout this work, we assume $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70$ kms^{-1}Mpc^{-1}. All magnitudes are in the AB system (Oke & Gunn 1983).

2. Observational material and sample selection

Our main goal in this work is to complement and extend current samples of galaxy-scale strong lens candidates, starting with the
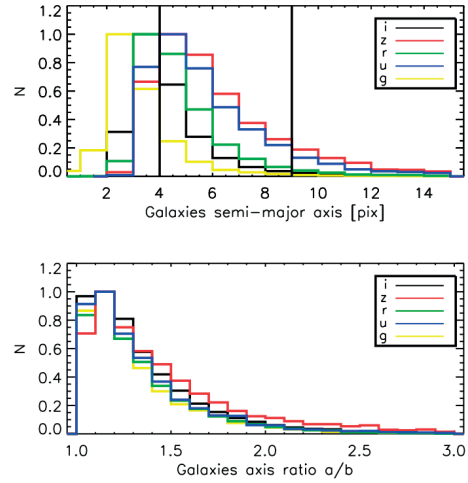
Fig. 1. Distribution of the semi-major axes and of their ratio, for our preselections of galaxies in the full CFHTLS sample (Cuillandre et al. 2012). The vertical lines mark our size cut-off of galaxies. The pixel size is $0''187$.

public imaging data from the CFHTLS. To do so, we use the new technique proposed by Joseph et al. (2014).

2.1. CFHTLS data

The Canada-France-Hawaii-Telescope Legacy Survey is a large program consisting of 500 observing nights between January 2003 and early 2009, divided into two parts. The *Deep* survey has 4 ultra-deep pointings and the *Wide* survey has 171 intermediate-depth pointings. Because strong lensing systems are rare, we need to use the widest possible survey area, i.e. the *Wide* part of CFHTLS.

The *Wide* CFHTLS (Cuillandre et al. 2012) covers 155 deg^2 divided into four independent fields observed in five bands. The limiting point-source AB magnitude (80% completeness) for each band is $u = 25.2$, $g = 25.5$, $r = 25.0$, $i = 24.8$, $z = 23.9$. For each band the mean seeing is FWHM= $0''85$, $0''79$, $0''71$, $0''64$, and $0''68$, respectively. The pixel size of the CFHTLS data is $0''187$.

In this paper, we use the T0007[1] final release of the CFHTLS with improved flat-fielding and photometric calibration techniques developed by the Supernova Legacy Survey (SNLS) team in collaboration with the CFHT. This release benefits from two types of photometric catalogues: 1. source catalogues for individual images (i.e. the `.ldac` files in the T0007 package) that we use to build our samples of lens galaxies in each filter separately, 2. merged source catalogues produced from the g, r, and i images that we use to infer the colour information of our lenses. The full description of the CFHTLS-T0007 release can be found in Cuillandre et al. (2012).

---

[1] See details at http://www.cfht.hawaii.edu/Science/CFHLS/
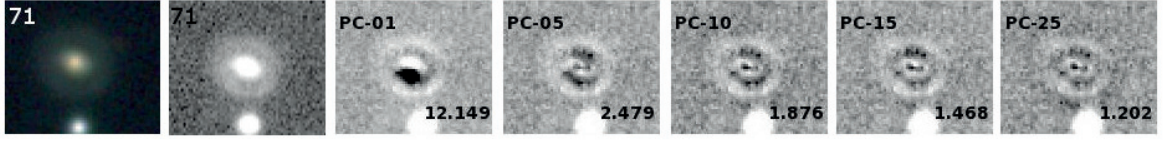
Fig. 2. Illustration of the PCA reconstruction of a galaxy image. The first two panels show the original image (colour and single-band). Each of the other panels shows the residual image between the data and the reconstructed galaxy using respectively 1, 5, 10, 15 and 25 principal components (PCs). The value of the corresponding reduced χ^2 is given in the lower right corner.

2.2. Sample selection

We preselect galaxies among the full CFHTLS source catalogues for individual images produced by TERAPIX using SEXTRACTOR (Bertin & Arnouts 1996). PCA requires uniform in size, morphology and brightness sample of elliptical galaxies, and since size and morphology of a galaxy changes between different bands, thus we create five catalogues, one for each of the five CFHTLS filters, independently. A given galaxy can therefore appear in several of the catalogues. Before using the PCA subtraction technique of Joseph et al. (2014), we apply the following selection criteria to each of the five catalogues:

- We use only the objects classified as non-stellar by SEXTRACTOR (CLASS_STAR > 0.98) and we apply a detection threshold of 10σ for each band, that is $u = 23.9$, $g = 24.3$, $r = 23.5$, $i = 23.5$, $z = 22.4$. Fainter objects would make difficult targets for future spectroscopic follow-up.
- We apply a cut on the semi-axis ratio, $a/b < 3$, which includes most of the early-type galaxies (Park et al. 2007), but rejects most of the spurious objects like those "created" by diffraction patterns of bright stars.
- We apply an (angular) size cut-off. Small galaxies, with a semi-major axis $a < 4$ pixels are excluded. Any arc in their vicinity can be detected without subtracting the light of the foreground galaxy. We also remove galaxies with $a > 9$ pixels. These galaxies are rare and therefore poorly modeled with the PCA technique (Joseph et al. 2014). Since we want to ensure uniformity of galaxy shapes in a group, galaxy sizes are computed separately in each band. Our final selection therefore spans sizes in the range $4 < a < 9$ pixels. Figure 1 displays the distribution in semi-major axis for the full sample in all the CFHTLS bands.

This leaves us with a pre-selection of early-type and late-type galaxies. However, spiral arms can be mistakenly taken for lensed arcs, resulting in false positives. To avoid this, we further restrict the sample to only elliptical galaxies. This can be achieved by either using a galaxy classifier based in morphological features in images (e.g. the ASTErIsM software by A. Tramancere, et al., 2015 submitted) or by applying a colour selection. We adopt the latter strategy, selecting galaxies with $(g - i) > 1.0$ within a 3″ aperture, following Gavazzi et al. (2010). Obviously, some of the potential lenses are missed by this selection, but this is the price to pay to remove spiral galaxies efficiently.

For each selected object, we create an image stamp centred on the galaxy. Since rotation is not a principal component, we also apply a rotation to each stamp to align the major axes of all galaxies. In doing so, we use a polynomial transformation and a bilinear interpolation. We note that we do not apply any other re-scaling. Instead, to ensure final uniformity of the PCA basis, we take advantage of the very large sample and we split it in five
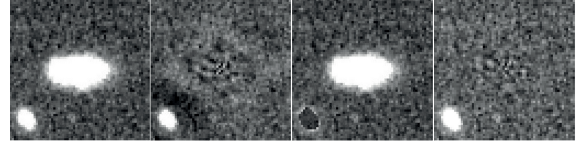
Fig. 3. Illustration of our masking strategy. *From left to right*: a) image of a galaxy from our sample, b) subtraction without using a mask during the reconstruction process, leaving ring-like artefacts, c) masked image used for the reconstruction process, d) resulting residual image, without any artefact.

bins of galaxies sizes. The five groups are defined by the galaxies semi-major axis as follows: (1): $a = [4 - 5]$, (2): $a = [5 - 6]$, (3): $a = [6 - 7]$, (4): $a = [7 - 8]$, and (5): $a = [8 - 9]$ pixels.

3. Lens-finding algorithm and improvements

Our algorithm by nature finds bright lensing galaxies. In such samples the lensed source is often hidden in the glare of the foreground galaxy, which must be properly removed before any search for lensed structures can be carried out. Our PCA-based lens-finder therefore includes two steps: 1. subtraction of the central galaxy from the original images/stamps, using PCA image reconstruction and 2. detection of lensed extended objects (arcs, rings) in the residual images.

3.1. Removal of the lensing galaxy: the PCA approach

Traditional ways of subtracting galaxies in imaging data are either to fit a elliptical profile to the data with, for example, the `galfit` software (Peng et al. 2010; Cabanac et al. 2008) or to subtract aperture-scaled images in two different bands (Gavazzi et al. 2014). As galaxies are not perfect elliptical profiles, these approaches often leads to significant flux residuals that prevents the detection of faint background lensed objects and they produce large amounts of false positives.

Our solution to this problem is presented in Joseph et al. (2014) [2], where we build an empirical galaxy light model from the sample of galaxies itself using a principal component analysis. PCA decomposition of a dataset allows one to recognise any similarity among its elements: the elements in the dataset are converted into another set of variables called principal components (PC), which are orthogonal and ordered so that the first PC has the largest possible variance, the second PC has the second largest variance, and so on. The details of our PCA technique are provided in Joseph et al. (2014).

---

[2] PCA script available at `https://github.com/herjy/PiCARD`

Paraficz: The PCA Lens-Finder: application to CFHTLS
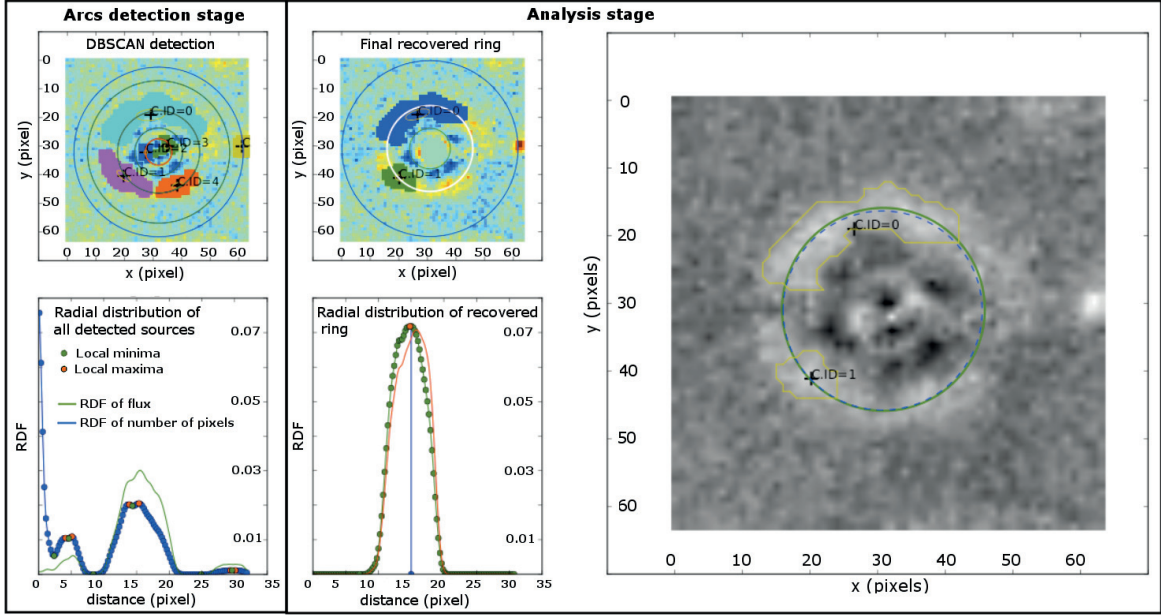
Fig. 4. Analysis of the PCA-subtracted images. **Arc detection stage.** *Top panel*: DBSCAN detection of the sources. *Bottom panel*: radial distribution of the detected sources. The green solid line represents the pixel flux radial distribution, while the blue line represents the radial distribution of the number of pixels in the sources. **Analysis stage.** *Top panel*: All sources smaller than a critical angular size (e.g. the PSF size) are removed, the remaining sources are merged together in the final ring (indicated with white line). *Bottom panel* The red solid line represents the pixel flux radial distribution, while the green line represents the radial distribution of the number of pixels in the sources. *Right panel*: Residual image, where the green solid line shows the best circle fit to the final ring. The yellow lines show the contour of the components of the final ring.

A critical step in the PCA reconstruction is the choice of the number of PC coefficients. If all coefficients are used, the reconstructed image is identical to the original image. This clearly leads to overfitting of the data and noise and will simply remove all structures of interest, like faint lensed rings and arcs. To circumvent this problem, the galaxy image needs to reconstructed using only a limited number of coefficients. Obviously, there might be an optimal number of coefficients to be used to avoid over-fitting or under-fitting of the data. This optimal number of PC depends on the diversity in shape among the galaxies in the sample, i.e. the range in galaxy sizes, the presence of companions near the galaxies used to build the PCA basis, and it also depends on the number of objects used to build the PCA basis. Fig. 2 gives an illustration of a galaxy image reconstructed from a different number of PCs. To evaluate the overall quality of reconstruction of a galaxy image in an objective way, we compute the reduced χ^2 in the image after subtraction of the galaxy (Eq. 6 in Joseph et al. 2014). An ideal reconstruction gives a reduced χ^2 close to one. Of course lensed features in the image do produce signal in the residual image, a cut-off in χ^2 has to be defined that ensures an adequate tradeoff between clean removal of the foreground galaxy and non-removal of any potential lensed feature. Finding this cut-off is a subjective part of our procedure and final result for false to true positives ratio will strongly depend on the adopted value. In the present case we choose $\chi^2 \sim 1.4$ for the CFHTLS data (see Fig. 2) and for the simulations used to evaluate completeness of the procedure (see Section 5).

The PCA technique described in Joseph et al. (2014) works well for isolated galaxies. In practice, however, galaxies often have companions, either physical or the result of line-of-sight effects. Companions affect the results both when building the PCA basis from the galaxy sample and when reconstructing the image of a given galaxy. The first problem (as described already in Joseph et al. 2014) is easily avoided by computing the PCA basis on a subset of galaxies with no bright companions. The large size of our galaxy sample allows us to do that in practice. However, companions are present in the images of the galaxy we want to reconstruct and subtract. In the case of reconstruction of an image of a galaxy with bright companions, residuals might feature artifacts that mimic a ring (Fig. 3). To avoid this problem we simply mask bright companion before the reconstruction process. Images reconstructed in this manner are then subtracted from original, non-masked images creating residuals that are now free of the false rings. Figure 3 illustrates the improvement over a non-masked image.

To apply the masking to all the companion sources in the image stamps, we use the DBSCAN algorithm implemented in the ASTErIsM software (Tramacere et al 2015 submitted). We identify the object at the centre of the stamp, as our source of interest that will be not masked. We iterate among all the remaining sources, and we mask all the sources with an integrated flux larger than the half of the central galaxy's integrated flux and all the sources with a distance from the central galaxy smaller than the half of the central galaxy's radius. The masked pixel fluxes are replaced with flux values randomly sampled from the background pixels flux distribution.

4

3.2. Looking for lensed features in the residual image

Once the galaxies have been properly removed from all preselected image stamps, we can now search for lensing features in these residual images. To avoid too many false positives we choose to search for only arc-like features. This was done using the cluster/island detection algorithm described in detail and tested in Tramacere et al (2015, in prep). Below, we provide a short overview of the procedure. This method is based on the application of the `DBSCAN` clustering topometric algorithm [3] (Easter et al. 1996; Tramacere & Vecchio 2013), which extracts *sources* in image stamps by defining density-based *clusters* of connected pixels. The method consists in the following two steps (see Fig. 4).

1. **Arcs and rings detection stage**
   - An initial list of sources is extracted using the `DBSCAN` algorithm (Top-left panel of Fig. 4);
   - The shape of each source is determined and each source is flagged to be: arc-like, ring-like, ellipse-like, circular/point-like;
   - All point-like sources are removed, leaving us with a list of candidate lensed sources;
   - For each stamp in the candidate list, we compute the radial distribution of the sources, and we determine the minima and the maxima in this radial distribution. This is presented on the bottom-left panel of Fig. 4, where the green filled circles represent the local minima, and the red filled circles represent the local maxima. The first minima in the innermost ring sets the radius for the internal disk, indicated as a red circle in the top-left panel of Fig. 4. These allow us to partition the stamps in circular areas.
   - All the sources within the internal disk (red circle in the left-top panel of Fig. 4) are masked, i.e. all the corresponding pixels are set to the background flux level (central-top panel of Fig. 4);
   - All the sources in the candidate list are assigned to a single circular area for each sources, enabling us to easily measure the angular size and the orientation of the sources, α_r, with regard to the radial direction;
   - All sources smaller than a critical angular size (e.g. the PSF size) are removed; only rings with at least one source meeting criteria are kept;
   - The sources are then merged together in the final ring to preform the analysis (see central panel in Fig. 4)

2. **Analysis stage**
   - Once we have a *final* ring, we fit a circle to the distribution of pixels in the recovered ring and we measure a centroid position, barycentre and radius, R (green circle in the left panel of Fig. 4).
   - We also fit a circle to each of the ring components that have an arc-like shape (blue dashed line in the right panel of Fig. 4), and we check that the circle is contained within the final ring best-fit circle.
   - We assign a quality factor to the ring, which is determined by the total angular coverage of the ring θ_{tot}, and the displacement d, between the ring circle best fit centroid, and the ring barycentre

$$q_f = \frac{\theta_{tot}}{2\pi} \frac{1}{\exp(\frac{d-R}{R/f}) + 1} \qquad (1)$$

where the larger the value of f, the more conservative the quality factor is. In this work we adopt minimum $q_f = 0.1$ to flag an object that are a possible lens.

We apply this automated procedure to the five bands of CFHTLS imaging data after PCA subtraction of the foreground object, leading to 1 098 lens candidates passing all above criteria.

4. CFHTLS results

In the following, we describe our main results using the PCA-finder. This includes a visualisation step by five of the authors, allowing us to define three subsamples of lens candidates depending on how likely the candidates are to be real lenses. The characteristics of our new sample are compared with previous lens searches in the CFHTLS.

4.1. Visual inspection

The automatically selected 1 098 candidates were visually inspected to identify obvious spurious objects and to refine our lensing classification. We rank each object in one of the following categories:

- **A**: an almost definite lens with a striking image configuration that is typical for lensing;
- **B**: probable lens, but the candidate would need follow-up with spectroscopy or more imaging;
- **C**: possible lens, but with low probability of being confirmed, either because of low signal-to-noise (S/N) or because the potential lensed images are single or consist of short arcs that could still be compatible with edge-on galaxies or chain galaxies;
- **0**: not a lens, spurious detection or spiral galaxy mimicking an arc or an Einstein ring. Objects in this category are false-positives.

The visual classification is a time-consuming process. However, the workload remains reasonable in the case of the CFHTLS, which requires a few hours of human time to inspect the 1 098 candidates. The classification is performed both on the true-colour images using the g, r, and i bands and on the residual images. This is done using the FITS images, enabling us to easily and quickly explore the full dynamical range of the data. This classification is made independently by five of the authors: D.P., J.-P.K., R.J., F.C., P.D. Out of all the systems, we select those objects that are classified as definite lenses by at least one individual initial judgment. All authors then needed to agree on a final classification. Figures A.1-A.5 in the Appendix and Tables 1-2 present all the lens candidates that we rank with the grade *A* and grade *B*.

Our visual inspection shows that the most frequent contaminants are face-on spiral galaxies, ring galaxies and polar ring galaxies. Face-on spirals mistakenly taken by the PCA-finder as lens candidates are easily identified by eye because their spiral arms point towards the bulge, while lensed arcs are tangentially aligned with respect to their lens galaxy.

More problematic are ring-like galaxies in general and polar ring galaxies in particular. These rare composite galaxies consist of a gas-poor, early-type galaxy (typically S0 galaxies) surrounded by a blue gaseous ring with ongoing star formation (see Fig. 5). The most widely accepted explanation for the formation of polar ring galaxies is that accreted gas settles onto orbits

---

[3] `DBSCAN` algorithm available at a (Easter et al. 1996; Tramacere & Vecchio 2013)

Paraficz: The PCA Lens-Finder: application to CFHTLS

ID	RA	DEC	z	$R_{\rm eff}$	g	r	i	R_E	Quality
				pix	mag	mag	mag	pix	
1	30.2905	-6.3474	$0.548^{+0.587}_{-0.507}$	4.53	21.28 ± 0.02	20.21 ± 0.01	19.61 ± 0.01	18.3	B
2	30.3615	-10.7597	$0.563^{+0.607}_{-0.521}$	4.17	23.37 ± 0.07	21.96 ± 0.05	21.17 ± 0.02	15.9	A
3	30.4522	-7.5357	$0.393^{+0.437}_{-0.335}$	4.51	22.38 ± 0.04	21.47 ± 0.03	21.02 ± 0.02	33.7	B
4	30.7655	-4.4937	$0.387^{+0.423}_{-0.353}$	3.90	21.56 ± 0.02	20.31 ± 0.01	19.73 ± 0.01	12.1	B
5	30.9987	-8.3652	$0.495^{+0.529}_{-0.459}$	4.48	21.48 ± 0.02	20.30 ± 0.01	19.76 ± 0.01	9.3	B
6	31.0361	-9.6104	$0.409^{+0.463}_{-0.363}$	3.83	22.00 ± 0.02	20.70 ± 0.01	20.03 ± 0.01	16.9	A
7	31.2852	-3.9099	$0.252^{+0.294}_{-0.227}$	5.22	20.56 ± 0.01	19.76 ± 0.01	19.31 ± 0.01	7.06	A
8	31.4770	-6.4598	$0.442^{+0.473}_{-0.411}$	5.32	20.98 ± 0.02	19.69 ± 0.01	19.17 ± 0.01	7.8	B
9	31.9736	-8.8432	$0.289^{+0.359}_{-0.243}$	5.30	20.55 ± 0.01	19.83 ± 0.01	19.47 ± 0.01	11.6	B
10	32.2221	-6.9186	$0.902^{+0.939}_{-0.866}$	6.40	22.28 ± 0.04	21.82 ± 0.05	20.91 ± 0.03	13.5	A
11	32.3970	-8.3013	$0.611^{+0.641}_{-0.580}$	4.23	21.85 ± 0.03	20.63 ± 0.01	19.81 ± 0.01	8.6	A
12	32.4703	-6.5295	$0.389^{+0.414}_{-0.351}$	3.90	20.00 ± 0.01	19.12 ± 0.00	18.65 ± 0.00	9.03	A
13	32.5096	-3.7956	$0.556^{+0.595}_{-0.515}$	7.29	21.60 ± 0.03	20.48 ± 0.02	19.86 ± 0.01	9.4	A
14	32.6591	-7.4773	$0.478^{+0.504}_{-0.453}$	3.05	22.93 ± 0.05	21.44 ± 0.02	20.64 ± 0.01	30.0	A
15	32.8441	-4.3681	$0.731^{+0.759}_{-0.703}$	4.36	23.13 ± 0.08	21.81 ± 0.03	20.49 ± 0.02	17.4	A
16	32.9734	-5.9950	$0.139^{+0.190}_{-0.059}$	5.98	19.29 ± 0.01	18.69 ± 0.01	18.31 ± 0.00	6.6	B
17	33.0833	-7.9352	$0.451^{+0.477}_{-0.428}$	7.58	21.06 ± 0.01	20.26 ± 0.01	19.64 ± 0.01	6.7	A
18	33.1342	-6.6479	$0.450^{+0.481}_{-0.421}$	6.13	20.94 ± 0.01	20.01 ± 0.01	19.63 ± 0.01	9.2	A
19	33.1958	-5.8338	$0.809^{+0.853}_{-0.768}$	5.96	21.76 ± 0.02	21.16 ± 0.03	20.36 ± 0.02	6.3	B
20	33.6128	-9.0673	$0.698^{+0.724}_{-0.670}$	4.08	22.44 ± 0.04	21.25 ± 0.02	20.17 ± 0.01	8.3	A
21	33.6250	-9.1754	$0.398^{+0.433}_{-0.366}$	5.64	21.16 ± 0.01	20.01 ± 0.01	19.59 ± 0.01	10.1	A
22	33.8107	-4.7156	$0.346^{+0.378}_{-0.313}$	5.55	20.40 ± 0.01	19.32 ± 0.01	18.83 ± 0.01	10.3	A
23	33.9600	-4.4247	$0.388^{+0.417}_{-0.359}$	5.23	20.32 ± 0.01	19.31 ± 0.01	18.88 ± 0.01	9.4	B
24	34.9904	-6.5704	$0.486^{+0.510}_{-0.464}$	3.54	22.72 ± 0.04	21.26 ± 0.02	20.43 ± 0.01	13.8	A
25	35.0485	-6.8143	$0.489^{+0.525}_{-0.454}$	5.72	20.85 ± 0.01	19.90 ± 0.01	19.45 ± 0.01	6.5	A
26	35.0763	-5.6397	$0.709^{+0.749}_{-0.676}$	5.34	22.29 ± 0.03	21.66 ± 0.04	21.09 ± 0.03	6.2	B
27	35.1759	-8.1834	$0.361^{+0.384}_{-0.337}$	4.90	20.51 ± 0.01	19.53 ± 0.01	19.18 ± 0.01	7.2	A
28	35.3647	-9.9535	$0.772^{+0.815}_{-0.732}$	4.32	23.06 ± 0.06	22.19 ± 0.05	21.13 ± 0.03	21.9	A
29	35.5374	-5.6453	$0.457^{+0.496}_{-0.420}$	6.31	21.06 ± 0.01	20.21 ± 0.01	19.93 ± 0.01	7.5	A
30	37.0163	-5.8651	$0.400^{+0.448}_{-0.356}$	2.66	22.78 ± 0.03	21.68 ± 0.03	21.27 ± 0.02	7.6	B
31	37.1982	-3.9803	$0.680^{+0.711}_{-0.651}$	9.24	21.66 ± 0.04	20.70 ± 0.03	19.93 ± 0.02	9.8	B
32	37.5014	-7.8604	$0.591^{+0.630}_{-0.549}$	3.85	22.96 ± 0.05	21.79 ± 0.03	21.01 ± 0.02	12.7	B
33	37.5045	-5.7003	$0.559^{+0.594}_{-0.525}$	4.96	21.31 ± 0.02	20.16 ± 0.01	19.54 ± 0.01	7.6	A
34	38.0929	-3.7355	$0.798^{+0.848}_{-0.762}$	9.60	22.03 ± 0.05	21.38 ± 0.05	20.41 ± 0.03	10.5	B
35	38.2284	-5.3160	$0.710^{+0.760}_{-0.665}$	4.57	23.10 ± 0.05	22.36 ± 0.06	21.73 ± 0.05	10.7	A
36	132.1016	-5.1126	$0.412^{+0.439}_{-0.388}$	7.33	21.02 ± 0.01	19.99 ± 0.01	19.50 ± 0.01	7.1	A
37	132.1377	-4.8329	$0.774^{+0.804}_{-0.745}$	2.18	22.99 ± 0.04	21.86 ± 0.03	20.65 ± 0.01	13.3	A
38	132.5373	-4.2216	$0.222^{+0.257}_{-0.186}$	4.96	19.66 ± 0.00	18.70 ± 0.00	18.25 ± 0.00	9.8	A
39	133.2203	-3.9328	$0.430^{+0.458}_{-0.401}$	8.14	20.49 ± 0.01	19.49 ± 0.01	19.12 ± 0.01	10.6	A
40	133.5310	-4.1230	$0.670^{+0.698}_{-0.641}$	8.29	22.18 ± 0.04	21.22 ± 0.04	20.44 ± 0.02	14.5	A
41	133.7787	-3.8646	$0.715^{+0.745}_{-0.687}$	8.10	21.60 ± 0.02	20.67 ± 0.03	19.85 ± 0.01	10.4	A
42	134.4040	-2.8884	$0.687^{+0.726}_{-0.660}$	5.44	22.05 ± 0.03	21.37 ± 0.02	20.76 ± 0.02	7.2	B
43	135.0560	-3.0676	$0.608^{+0.648}_{-0.566}$	5.33	22.53 ± 0.05	21.33 ± 0.04	20.54 ± 0.02	9.1	A
44	135.2780	-1.8642	$0.353^{+0.385}_{-0.322}$	6.13	20.47 ± 0.01	19.51 ± 0.01	19.11 ± 0.01	11.2	A
45	135.4850	-2.5134	$0.689^{+0.716}_{-0.663}$	5.91	22.16 ± 0.04	21.22 ± 0.02	20.49 ± 0.02	6.7	A
46	208.9030	57.0818	$0.392^{+0.420}_{-0.366}$	5.15	21.42 ± 0.02	20.26 ± 0.01	19.74 ± 0.01	6.4	A
47	208.9420	57.1261	$0.406^{+0.434}_{-0.383}$	6.76	20.32 ± 0.01	19.33 ± 0.01	18.92 ± 0.01	11.0	A
48	209.2140	54.2889	$0.574^{+0.616}_{-0.515}$	5.09	21.45 ± 0.02	20.64 ± 0.01	20.18 ± 0.01	5.4	A
49	209.3525	55.6741	$0.398^{+0.425}_{-0.374}$	7.07	20.29 ± 0.01	19.31 ± 0.01	18.93 ± 0.01	14.2	B
50	209.3780	53.4301	$0.370^{+0.399}_{-0.329}$	4.96	20.44 ± 0.01	19.53 ± 0.01	19.08 ± 0.01	8.2	A
51	209.6380	55.8449	$0.376^{+0.399}_{-0.352}$	4.92	20.02 ± 0.00	18.99 ± 0.00	18.55 ± 0.00	9.8	A
52	209.7398	57.0189	$0.303^{+0.343}_{-0.256}$	6.29	20.66 ± 0.01	19.44 ± 0.01	18.93 ± 0.01	14.1	B
53	209.7620	53.3673	$0.331^{+0.374}_{-0.277}$	4.86	21.10 ± 0.02	19.77 ± 0.01	19.21 ± 0.01	12.2	A
54	209.8280	57.4606	$0.396^{+0.432}_{-0.363}$	5.92	21.60 ± 0.02	20.38 ± 0.01	19.87 ± 0.01	11.7	A
55	209.8940	54.8880	$0.456^{+0.487}_{-0.424}$	4.40	20.61 ± 0.01	19.81 ± 0.01	19.45 ± 0.01	7.1	B
56	209.8970	56.7132	$0.307^{+0.339}_{-0.274}$	5.00	20.28 ± 0.01	19.31 ± 0.00	18.85 ± 0.01	7.6	B
57	209.9210	56.1383	$0.370^{+0.400}_{-0.340}$	8.58	20.52 ± 0.01	19.41 ± 0.01	18.92 ± 0.01	11.2	A
58	210.0069	56.9977	$0.377^{+0.398}_{-0.352}$	7.27	20.52 ± 0.01	19.58 ± 0.01	19.19 ± 0.01	9.7	B

Table 1. List of grade-*A* and grade-*B* new lens candidates in CFHTLS. The photometric redshifts *z*, the effective radius $R_{\rm eff}$ and the magnitudes are the ones provided by Coupon et al. (2009).

that are more frequently contained either within the equatorial or polar planes. Since the polar rings are blue and nearly perpen- dicular to the semi-major axis of their central hosts, they closely

6

Paraficz: The PCA Lens-Finder: application to CFHTLS

ID	RA	DEC	z	$R_{\rm eff}$	g	r	i	R_E	Quality
				pix	mag	mag	mag	pix	
59	210.3022	56.2394	$0.412^{+0.446}_{-0.382}$	4.81	20.93 ± 0.01	19.58 ± 0.01	18.99 ± 0.01	14.7	B
60	210.3220	57.3084	$0.382^{+0.412}_{-0.348}$	7.61	20.08 ± 0.01	19.47 ± 0.01	19.17 ± 0.01	9.1	A
61	210.3420	57.0673	$0.810^{+0.905}_{-0.757}$	2.73	22.92 ± 0.06	22.52 ± 0.05	21.98 ± 0.05	7.9	B
62	210.5270	53.4316	$0.564^{+0.601}_{-0.529}$	7.80	21.29 ± 0.03	20.06 ± 0.01	19.36 ± 0.01	8.9	A
63	210.5496	57.5600	$0.759^{+0.792}_{-0.727}$	9.35	21.89 ± 0.05	20.97 ± 0.03	20.11 ± 0.02	14.8	A
64	210.5840	51.7352	$0.214^{+0.250}_{-0.175}$	4.40	19.79 ± 0.00	18.92 ± 0.00	18.47 ± 0.00	9.4	A
65	211.4080	57.6165	$0.306^{+0.340}_{-0.271}$	6.51	20.33 ± 0.01	19.43 ± 0.01	19.03 ± 0.01	8.9	A
66	211.8142	57.1322	$0.322^{+0.353}_{-0.289}$	6.09	20.02 ± 0.01	18.94 ± 0.00	18.50 ± 0.00	11.6	A
67	211.8690	52.6938	$0.485^{+0.516}_{-0.448}$	6.44	21.18 ± 0.02	20.03 ± 0.01	19.56 ± 0.01	7.1	A
68	211.9780	56.2218	$0.387^{+0.412}_{-0.364}$	7.21	21.08 ± 0.01	19.99 ± 0.01	19.56 ± 0.01	8.9	A
69	212.0175	56.2446	$0.370^{+0.400}_{-0.339}$	5.01	21.05 ± 0.01	19.88 ± 0.01	19.36 ± 0.01	10.1	A
70	212.1570	52.3579	$0.855^{+0.899}_{-0.819}$	6.39	21.78 ± 0.03	21.05 ± 0.03	20.16 ± 0.02	7.7	A
71	212.2455	51.8158	$0.343^{+0.373}_{-0.315}$	7.97	21.02 ± 0.02	20.55 ± 0.02	20.50 ± 0.02	11.2	A
72	212.3657	53.5918	$0.421^{+0.453}_{-0.389}$	5.84	20.84 ± 0.01	19.66 ± 0.01	19.16 ± 0.01	13.1	B
73	212.6040	54.0908	$0.420^{+0.457}_{-0.383}$	5.63	20.84 ± 0.01	19.68 ± 0.01	19.25 ± 0.01	12.4	A
74	212.7290	54.9406	$0.469^{+0.510}_{-0.429}$	4.91	21.75 ± 0.02	20.53 ± 0.01	19.89 ± 0.01	9.7	B
75	212.8450	51.6687	$0.499^{+0.536}_{-0.455}$	4.70	21.66 ± 0.02	20.41 ± 0.01	19.80 ± 0.01	6.5	A
76	213.0320	52.9143	$0.507^{+0.560}_{-0.463}$	6.89	21.07 ± 0.02	20.28 ± 0.01	19.86 ± 0.02	8.9	A
77	213.1650	53.9570	$0.436^{+0.469}_{-0.403}$	7.82	20.64 ± 0.01	19.53 ± 0.01	19.05 ± 0.01	8.2	A
78	213.4510	51.7295	$0.350^{+0.384}_{-0.311}$	7.57	19.92 ± 0.01	19.01 ± 0.01	18.58 ± 0.01	13.6	A
79	213.5430	52.8470	$0.393^{+0.424}_{-0.366}$	6.62	20.86 ± 0.01	19.83 ± 0.01	19.40 ± 0.01	7.3	A
80	213.6000	57.6236	$0.529^{+0.559}_{-0.496}$	5.81	21.84 ± 0.03	20.69 ± 0.02	20.19 ± 0.01	6.4	A
81	213.9140	54.8451	$0.232^{+0.266}_{-0.202}$	6.64	18.79 ± 0.00	18.05 ± 0.00	17.65 ± 0.00	11.9	A
82	214.4110	56.3307	$0.590^{+0.619}_{-0.553}$	5.54	21.76 ± 0.02	20.76 ± 0.02	20.04 ± 0.01	8.3	B
83	214.5100	57.3730	$0.570^{+0.609}_{-0.535}$	6.41	21.15 ± 0.02	20.19 ± 0.01	19.74 ± 0.01	7.7	A
84	214.5255	54.2536	$0.738^{+0.775}_{-0.707}$	4.95	22.23 ± 0.03	21.22 ± 0.03	20.21 ± 0.01	18.3	A
85	214.9620	51.8585	$0.682^{+0.709}_{-0.658}$	3.83	22.01 ± 0.02	21.16 ± 0.02	20.45 ± 0.02	8.9	B
86	215.3410	56.2251	$0.546^{+0.583}_{-0.508}$	5.86	21.08 ± 0.01	20.23 ± 0.01	19.78 ± 0.01	7.4	A
87	215.6690	57.0355	$0.433^{+0.467}_{-0.401}$	4.90	20.54 ± 0.01	19.41 ± 0.01	18.91 ± 0.01	9.2	A
88	216.3770	56.4335	$0.508^{+0.538}_{-0.475}$	7.41	20.86 ± 0.01	19.62 ± 0.01	19.01 ± 0.01	11.1	A
89	216.5700	55.1213	$0.629^{+0.655}_{-0.602}$	5.90	21.73 ± 0.02	20.66 ± 0.02	19.80 ± 0.01	9.2	A
90	216.7250	56.1682	$0.240^{+0.268}_{-0.210}$	6.96	19.97 ± 0.01	19.27 ± 0.01	18.90 ± 0.01	13.7	B
91	217.0550	54.8198	$0.855^{+0.894}_{-0.814}$	4.50	23.40 ± 0.08	22.12 ± 0.04	20.97 ± 0.02	18.4	B
92	217.1570	55.4547	$0.651^{+0.679}_{-0.622}$	6.09	22.16 ± 0.03	21.09 ± 0.02	20.21 ± 0.02	8.4	B
93	217.4450	54.6213	$0.639^{+0.667}_{-0.609}$	6.02	21.55 ± 0.02	20.84 ± 0.02	20.28 ± 0.02	6.5	B
94	217.9957	55.7248	$0.465^{+0.507}_{-0.427}$	8.14	19.95 ± 0.01	19.04 ± 0.00	18.69 ± 0.01	9.9	A
95	218.4500	57.6522	$0.307^{+0.339}_{-0.274}$	6.68	19.74 ± 0.01	18.70 ± 0.00	18.26 ± 0.00	11.8	A
96	218.9394	55.9681	$0.734^{+0.758}_{-0.711}$	7.15	21.03 ± 0.01	20.25 ± 0.01	19.78 ± 0.01	11.4	B
97	218.9660	57.6901	$0.616^{+0.656}_{-0.583}$	6.94	21.96 ± 0.05	20.99 ± 0.03	20.26 ± 0.02	12.4	B
98	330.2529	2.2095	$0.250^{+0.285}_{-0.218}$	5.42	19.84 ± 0.01	18.94 ± 0.01	18.48 ± 0.01	11.3	A
99	330.6014	3.9024	$0.316^{+0.347}_{-0.284}$	5.65	20.17 ± 0.01	19.03 ± 0.01	18.56 ± 0.00	11.8	A
100	330.6080	2.1078	$1.008^{+1.062}_{-0.951}$	6.13	21.79 ± 0.03	21.39 ± 0.04	20.57 ± 0.03	7.2	B
101	331.3547	0.9742	$0.621^{+0.650}_{-0.588}$	5.34	22.00 ± 0.03	20.95 ± 0.02	20.26 ± 0.01	8.3	B
102	331.6466	2.2712	$0.334^{+0.375}_{-0.297}$	3.88	21.51 ± 0.02	20.06 ± 0.01	19.49 ± 0.01	21.1	B
103	331.8250	3.5431	$0.411^{+0.450}_{-0.380}$	6.03	21.45 ± 0.02	20.40 ± 0.02	20.00 ± 0.01	7.9	B
104	331.8586	1.4529	$0.373^{+0.396}_{-0.347}$	7.66	20.39 ± 0.01	19.48 ± 0.01	19.14 ± 0.01	7.7	A
105	332.0030	2.6561	$0.466^{+0.502}_{-0.426}$	4.09	21.76 ± 0.02	20.64 ± 0.02	20.15 ± 0.01	10.7	B
106	332.3005	3.7471	$0.270^{+0.298}_{-0.239}$	4.36	21.95 ± 0.02	21.19 ± 0.02	20.77 ± 0.02	14.8	A
107	332.3815	-0.2096	$0.472^{+0.509}_{-0.429}$	8.11	20.38 ± 0.01	19.40 ± 0.01	19.09 ± 0.01	8.4	B
108	334.0200	1.8810	$0.764^{+0.794}_{-0.735}$	7.28	21.75 ± 0.03	20.98 ± 0.03	20.14 ± 0.02	11.2	A
109	335.5896	-0.2775	$0.291^{+0.324}_{-0.257}$	5.69	20.14 ± 0.01	19.12 ± 0.01	18.62 ± 0.00	13.3	B

Table 2. List of grade-*A* and grade-*B* new lens candidates in CFHTLS. The photometric redshifts z, the effective radius $R_{\rm eff}$ and the magnitudes are the ones provided by Coupon et al. (2009).

resemble Einstein rings that are produced by gravitational lensing.

In our visual procedure, we attempt to classify an object as a ring galaxy if i) the ring structure has an ellipticity $\epsilon > 0.2$, or ii) the ring shows a surface brightness close to constant. Our criterion on the ellipticity comes from the fact that only very extreme, rare, and rather unphysical lens galaxies or systems with extreme external shear can create a strongly elongated Einstein ring. Very elliptical Einstein rings are therefore not expected.

In fact, none are known in the current literature. On the contrary, gaseous rings around polar ring galaxies can be strongly elliptical, simply due to orientation effects. The second condition, i.e. constant surface brightness, comes from the fact that Einstein rings are never fully symmetric and that lensed sources have structures, i.e. a bulge, spiral arms, etc. Ring galaxies have more uniform light distributions across the gaseous ring.

Using the above criteria, the PCA-finder provides a list of 1 098 lens candidates that split, after visual classification, into:
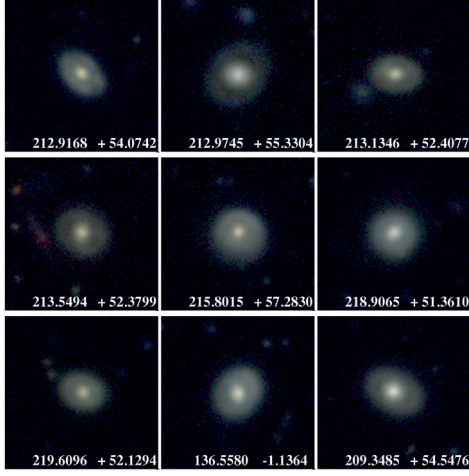
7

Paraficz: The PCA Lens-Finder: application to CFHTLS

Fig. 5. Examples of objects producing false positives in our lens search and that we classify as ring-like galaxies or polar ring galaxies (see text). Among 1 098 lens candidates, we identify 274 of these ring-like galaxies.

– 70 grade-*A* candidates (Tables 1 & 2),
– 39 grade-*B* candidates (Tables 1 & 2),
– 183 grade-*C* candidates (Table A.2 of the Appendix),
– 274 ring-galaxies or polar-ring galaxies (Table A.1 of the Appendix).

All our newly discovered grade-*A* and grade-*B* lens candidates are shown in Figs. A.1- A.5 of the Appendix [4]. Despite the visual classification, all above candidates would need spectroscopic and/or high-resolution photometric follow-up, which is beyond the scope of this paper. In the following we compare the properties of our sample of 109 (new) grade-*A* and grade-*B* lenses with other lens samples found in the CFHTLS data.

4.2. Comparison with previous searches

The CFHTLS data comprise all desirable survey properties for a lens search. They have been extensively explored in the past with a range of automated lens-finders, leading to very different lens samples. This clearly illustrates that no single technique can detect all the types of lenses at once and that current lens-finders are complementary. Some favour specific types of lensing object, such as spiral or elliptical galaxies, and others may select only massive lenses, e.g. by pre-selecting lenses as galaxy groups or clusters. Other favour a given source geometry, e.g. point sources (AGNs, quasars) or extended arcs and rings. Here, we give a brief summary of previously published CFHTLS lens samples and we attempt to understand why the PCA-finder method finds some but not all lenses from the published samples.

A lens sample that is significantly different from the one in the present paper is provided by Elyiv et al. (2013) and Sygnet et al. (2010). On the one hand Elyiv et al. (2013) searched for gravitational lens candidates among the optical counterparts of X-ray-selected QSOs/AGNs. The authors visually inspect a sample of 5 500 optical counterparts of X-ray point-like sources identified in the XMM-LSS imaging of the CFHTLS W1 field.

[4] FITS images of our lens candidates are also available at https://github.com/herjy/PiCARD

ArcFinder by More et al. (2016)			
ID	RA	DEC	*i*	z_{phot}
1	30.6619	-6.5823	19.54	0.37
2	30.8351	-7.5808	19.45	0.59
3	33.8459	-7.6065	20.89	1.05
4	35.2351	-7.7199	20.51	0.71
5	35.8142	-6.4851	19.21	0.55
6	36.5298	-4.4573	17.97	0.17
7	132.0986	-4.1209	18.85	0.51
8	134.4546	-1.2169	18.26	0.29
9	210.4371	53.0360	19.61	0.56
10	214.8007	53.4365	19.11	0.69
11	214.8219	51.2913	18.72	0.47
12	217.9695	57.4769	20.19	0.83
13	217.1451	52.2185	19.94	0.52
14	217.5027	55.7799	19.12	0.55
15	330.8709	2.0886	19.37	0.38
16	331.2788	1.7844	19.15	0.46
17	333.2789	-0.5103	18.81	0.69
18	333.5784	1.1761	18.84	0.74

ArcFinder by Maturi et al. (2014)			
1	33.5688	-5.0548	21.00	0.37
2	34.9856	-6.0341	20.50	0.42
3	36.4030	-4.2549	22.10	0.56
4	37.2865	-5.3320	22.40	0.37
5	209.2597	52.5104	23.00	0.38
6	209.6937	52.3495	23.40	0.35
7	210.0883	52.2626	21.20	0.76
8	335.5734	0.2007	21.70	0.51

Table 3. Strong lenses found using the two different ArcFinders (see text), and that we also find in the present work with the PCA-finder.

They find three good gravitational lens candidates. Sygnet et al. (2010), on other hand, look for lensing events produced only by massive edge-on disk galaxies. In their search, they preselect only highly elongated objects with $0.7 > \epsilon > 0.9$. Their final sample, which also involves a visual inspection, has 16 lens candidates. The PCA-finder neither looks for point-like multiple images nor for elongated lenses, thus we do not expect our search to recover any of those published lenses.

To the best of our knowledge, there are four lens searches similar to ours in CFHTLS (Gavazzi et al. 2014; More et al. 2012; Maturi et al. 2014; More et al. 2016). More et al. (2012) built a sample of lenses using ArcFinder with a setting such that only systems with arc radii larger than 2″ are kept in the sample. Their lens sample with large Einstein radii therefore predominantly selects group and cluster-scale lenses. ArcFinder measures the second order moments of the flux distribution in pixels within small regions of the sky to estimate the direction and extent of local elongation of features. Then, a set of thresholds on feature properties such as the area, length, width, curvature and surface brightness were used to select arc-like candidates. The search was carried out in the *g* band which is the most efficient wavelength to find typical lensed features. The ArcFinder final sample consists of 55 promising lenses out of a total of 127 lens candidates, which are selected from both CFHTLS *Wide* and *Deep* fields. The PCA-finder recovers 16 out of these 127 candidates. This low fraction of recovered systems is somewhat expected since the majority of the systems found by ArcFinder consist of multiple lensing galaxies, which are not recoverable by our method. PCA-finder detects arcs and rings that are centred on single lensing galaxies, any of the lensing features around

Paraficz: The PCA Lens-Finder: application to CFHTLS

RingFinder by Gavazzi et al. (2014)			
ID	RA	DEC	i	z_{phot}
1	31.0368	-6.2019	19.92	0.440
2	31.3527	-9.5065	19.46	0.697
3	32.7569	-8.9320	20.67	0.562
4	33.2527	-8.7196	19.26	0.471
5	33.9505	-3.7979	19.7	0.577
6	34.6119	-7.2910	20.02	0.474
7	35.2352	-7.7199	20.48	0.688
8	35.6735	-5.6477	19.50	0.502
9	36.4030	-4.2549	19.60	0.631
10	36.5152	-9.7643	18.30	0.229
11	36.6384	-3.8179	20.08	0.652
12	36.7455	-8.0105	19.06	0.450
13	37.1431	-8.7207	19.08	0.493
14	37.9618	-4.2917	19.69	0.838
15	38.6843	-6.8091	20.32	0.728
16	133.3229	-2.0543	20.51	0.706
17	133.7865	-3.1020	20.64	0.613
18	134.3794	-1.0678	18.72	0.660
19	136.5196	-3.9364	19.51	0.776
20	210.0897	51.5229	19.82	0.523
21	210.0947	54.9680	20.01	0.703
22	210.1774	56.0118	19.26	0.568
23	210.6061	56.6629	20.32	0.662
24	210.9173	56.7688	19.54	0.689
25	211.0588	51.7374	19.69	0.645
26	211.1062	52.0850	18.82	0.522
27	211.3248	54.5971	20.48	0.726
28	211.8857	54.5689	19.03	0.411
29	212.2298	52.7479	19.98	0.492
30	213.9302	52.4597	19.17	0.445
31	214.8219	51.2913	18.72	0.468
32	215.0140	52.5271	20.99	0.510
33	215.1154	54.1452	18.54	0.421
34	215.1325	52.9728	18.87	0.461
35	215.1830	54.8169	19.97	0.727
36	215.8413	57.3786	19.11	0.611
37	216.0988	52.5648	18.31	0.277
38	216.1354	55.0055	19.39	0.451
39	217.8799	57.1606	18.71	0.454
40	218.5875	54.6375	20.29	0.728
41	219.1557	54.9436	19.05	0.360
42	219.7768	54.6502	19.33	0.378
43	330.8709	2.0886	19.37	0.380
44	331.2789	1.7845	19.15	0.460
45	332.1596	3.0189	18.51	0.302
46	333.3725	0.4932	18.91	0.483
47	333.4006	0.1964	20.34	0.623
48	333.4959	0.9046	18.27	0.370
49	335.4535	1.2618	18.35	0.346
50	335.5735	0.2008	19.13	0.421

Table 4. Strong lenses discovered with the RingFinder and also found by the PCA-finder.

Space Warps by More et al. (2016)			
ID	RA	DEC	i	z_{phot}
1	31.6759	-9.8669	20.8	0.2
2	32.1339	-4.5542	21.0	1.0
3	33.1051	-8.8697	19.5	0.8
4	135.5794	-5.6566	0.0	0.0
5	211.5958	52.1617	20.3	0.7
6	216.5869	56.2323	19.5	0.5
7	216.7205	56.0016	0.0	0.0
8	217.3907	56.4277	19.0	0.5
9	217.7351	57.4084	19.3	0.7
10	219.2150	53.1183	19.2	0.6

Table 5. Galaxy-scale lensed systems found in the context of the Space Warps project that are also detected by our PCA-finder.

creased the CFHTLS sample of gravitational lenses. They apply their method to the CFHTLS Archive Research Survey (CARS; Erben et al. 2009) data, which covers 37 square degrees, to verify its efficiency and to detect new gravitational arcs. Table 3 lists the lenses found both by the PCA-finder and by Maturi et al. (2014).

Gavazzi et al. (2014) use their RingFinder tool to search for galaxies lensed by massive foreground early-type galaxies. The principle of RingFinder is similar to ours: they select all early-type galaxies from CFHTLS and then subtract them from the images to find lensing features. There are two main differences between our work and Gavazzi et al. (2014): the way the lenses are subtracted from the images and the way the residual images are analysed. To remove the central galaxy, Gavazzi et al. (2014) subtracts the PSF-matched i-band images from the g-band images. On the residual image, they looked for excess flux in the g-band to search for compact lensing signal, i.e. multiply-imaged point sources, rings and arcs. In total 2 524 objects passed their automatic selection procedure. These are visually inspected, leading to a total of 330 lens candidates, out of which 42 were ranked as good quality lenses and 288 were in their medium-quality category. In addition to the main sample of Gavazzi et al. (2014), another 71 candidates were reported to have been detected by earlier versions of the RingFinder. From the main sample of RingFinder, during their follow-up campaign, they confirmed 33 lenses. Out of the 330 medium and high quality candidates found with RingFinder, 50 are also found by our PCA-finder (Table 4).

Finally, the most recent CFHTLS lens search is known as Space Warps by More et al. (2016) and is fully based on a visual detection and classification of lensing systems by humans, namely "citizen" that volunteer to inspect the CFHTLS colour images. They report the discovery of 29 promising new lens candidates out of a total 59 candidates, based on about 11 million classifications performed by motivated citizen scientists. The goal of the blind lens search was to identify lens candidates missed by automated searches. This type of massive visual search enables us to catch a larger diversity in lens and source properties than automated searches (see also Pawase et al. 2014, for an example of a visual search in the HST database). Our PCA-finder recovers ten out of the 29 best candidates found in Space Warps. These are listed in Table 5.

Our PCA-finder cannot be expected to recover all the RingFinder and Space Warps lenses, owing to the different levels of incompleteness of the different searches but also because of the pre-selection of galaxies in the PCA-finder search. We target early-type galaxies as potential lenses and we apply a cut in size

groups or cluster of galaxies are lost. Table 3 lists the lenses found both by the PCA-finder and with ArcFinder.

Maturi et al. (2014) devised an automated lens-finder based on the colour statistics of arcs using a model for the spectral energy distribution (SED) of high redshift galaxies, the lensing optical depth, and the data depth. They therefore find lensed sources not only based on their morphology, but also from their colour, selecting the colours corresponding to sources providing the largest possible lensing cross-section. Using this procedure, which combines the ArcFinder created by Seidel & Bartelmann (2007), with a fine-tuned colour selection, they significantly in-

9

for the lens galaxy (4-9 pixels). At least part of the lens candidates from the previous searches do not meet this size cut. We also note that the PCA-finder is not optimized to find multiply-imaged point sources, which are very well spotted visually in Space Warps and with RingFinder.

Finding gravitational lenses is a complex task, thus no single lens finding method is perfect, each method has advantages over the other. It may be the case that a single method may never be the best method for optimising completeness and purity. As a consequence, it is not surprising that in spite of the many previous extensive lens searches in CFHTLS, we still manage to find new, interesting candidates. The PCA-finder, despite being very close conceptually to the RingFinder, has two major advantages. First, it is applied efficiently on single-band data. We therefore apply it independently to all the bands. In this way the search is not restricted to a limited range of source colours. Second, the resulting lens subtraction leaves very few artefacts, hence allowing us to spot fainter lensed features closer to the lens centre.

We note that the 109 good lens candidates listed in Tables 1 & 2 are completely new. We also list 183 new objects in Table A.2 of the Appendix that we rank as possible lenses, but that certainly require follow-up with either deeper imaging or spectroscopy or both.

4.3. Sample properties

We now compare various properties of our lens candidates with previous samples from CFHTLS. We emphasise that these comparisons use lens candidates that are not yet confirmed and that our results are therefore only indicative.

We use the CFHTLS photometric catalogues from Coupon et al. (2009) and generated with the Le Phare software (Ilbert et al. 2006). The accuracy of the photometric redshifts of galaxies for the *Wide* survey with magnitudes $i < 21.5$ is $\sigma(\Delta z/(1 + z)) = 0.032$. Figure 6 shows the distributions in apparent magnitude, Einstein radii and redshift for our PCA-finder sample of 70 grade-*A* plus 39 grade-*B* new gravitational lens candidates. These are shown together with the same distributions for Space Warps (More et al. 2016), for the ArcFinder (More et al. 2012), and for the RingFinder (Gavazzi et al. 2014). We find that the median of the lens redshift distribution for the PCA sample is $z_{\rm PCA} = 0.48 \pm 0.17$ and it is lower than redshift for all the other known lenses, which is $z_{\rm arcs} = 0.52 \pm 0.20$ (including giant arcs which systematically have larger redshifts). The median of *i*-band magnitude of our sample is $m_i = 19.63$, which turns into a median absolute magnitude of our sample $M_g = -21.90 \pm 0.745$. These magnitudes are K-corrected following Coupon et al. (2009).

The Einstein radii displayed in Fig. 6 are estimated from the position of the multiply-lensed images. R_E is taken to be half the averaged values of the angular separation between images. The distribution of image separations can be used to probe the average density profile of the lens population (Oguri et al. 2006; More et al. 2012). We find that the average Einstein radius for our new lenses is $R_E = 1.9 \pm 0.8''$ which is, as expected, smaller than for the ArcFinder candidates, which have $R_E = 4.0''$. This is also lower than for Space Wraps which have $R_E = 1.9''$. For comparison the SLACS lenses have $R_E = 2.2''$ and the RingFinder lenses is $R_E = 1.4''$ (see Sonnenfeld et al. 2013).
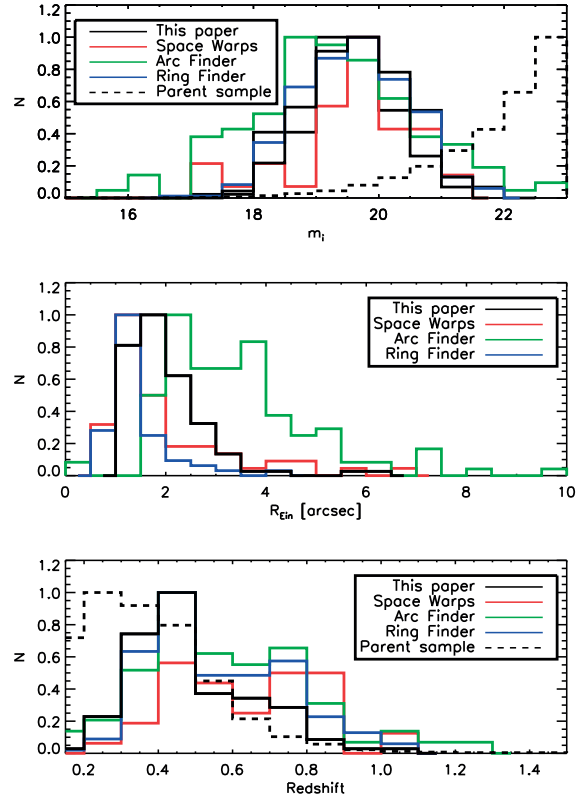

Fig. 6. *i*-band magnitude, Einstein radius, and redshift distributions of our lens candidates with *A* and *B* grades (black line). These are compared with the same distributions for other lens samples found in CFHTLS *Wide*: the one from the Space Warps program (red line; More et al. 2016), from the ArcFinder (green line; More et al. 2012) and from the RingFinder (blue line; Gavazzi et al. 2014). When applicable, we also show the distribution for the parent sample, i.e. our preselection of potential lens galaxies.

5. Simulations and completeness

An evaluation of the completeness of our sample can be done in two ways, using realistic image simulations or using a sample of already known lenses. The latter approach has been attempted in the previous section, but has a clear limitation: the reference sample of known lenses has its own completeness and purity. Moreover, different lens-finders are not necessarily optimised to find the same lens types and the parent samples (i.e. after the pre-selection) are not the same. As was shown in the previous section, one lens-finder can be very efficient at finding objects with low lens/source luminosity contrast, another one may be specialised in finding large arc-like structures and others may find better lensed point sources. For all these reasons, we choose to use simulated images for our completeness estimation.

In this section we evaluate the performances of our method using simulated images of Einstein rings, as they would be seen with the CFHT. We have made an attempt to generate lenses that are as realistic as possible within the requirements of the

10

Paraficz: The PCA Lens-Finder: application to CFHTLS
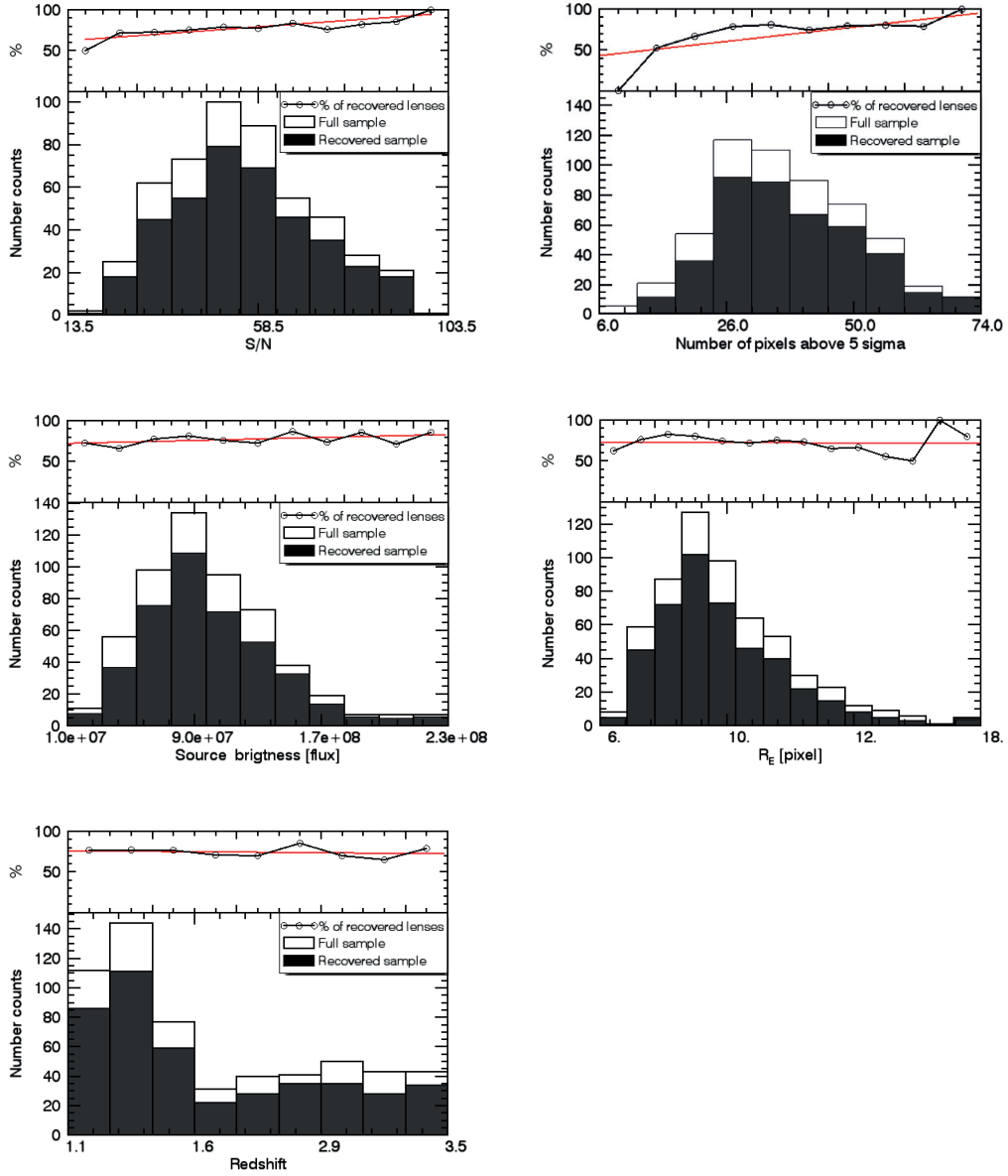
Fig. 7. Properties of the lensed systems for the simulated sample. Each plot in the lower panels shows the distributions of selected parameters of the full sample of simulated systems and for the simulated objects we identify as lenses (true positives). The top panels give the ratio of the two, i.e. the completeness per bin of the selected parameter. The red line is a linear regression to guide the eye.

PCA lens-finder. In the following, we also describe some of the properties of our simulated sample.

5.1. The lens and source simulation

The image simulations are provided by the Bologna Lens Factory (BLF). The BLF setup was chosen to match the proper-

ties of gravitational lenses expected in the CFHTLS wide fields by adding fake lensed objects to real images.

The lensing simulations were done as follows. A dark matter halo catalogue within a light cone extending out to $z = 4$ is taken from the Millennium Run Observatory (Overzier et al. 2013). This contains all the halos found within the Millennium cosmological simulation that were more massive than 10^{10} $M_\odot$, which

11

Paraficz: The PCA Lens-Finder: application to CFHTLS

should include the hosts of all the observable strong lenses. The light cone covers 1.6 square degrees of sky. Each halo is represented within the lensing code by a Navarro, Frenk & White (NFW) halo plus a singular isothermal ellipsoid (SIE) in its centre to represent the baryonic component. The mass of the baryonic component is determined using the halo mass vs. stellar mass relation of Moster et al. (2010) and the velocity dispersion is set by the Tully-Fisher relation (Bell & de Jong 2001). Once the light cone is assembled, all the caustics in this light cone with Einstein radii $R_E > 0.''05$ are located, for a series of six source planes running from $z = 1$ to $z = 3.5$, using the GLAMER lensing code (Metcalf & Petkova 2014; Petkova et al. 2014). GLAMER shoots rays through the light cone and identifies regions of the source plane that will be strongly lensed. The code then adaptively shoots rays in these areas at higher and higher resolution to resolve the critical curves and caustic curves of each prospective lens.

The simulated image of a lensed source is then added to a real image of a galaxy that is randomly taken from the preselected CFTHLS data. This step was unavoidable, bearing in mind that the PCA lens extraction is solely based on the self-similarity of the foreground galaxies. Since the simulations need to be as close to real CFHTLS data as possible, we thus draw galaxies from the preselected targets in size, magnitude, colour, etc (see Sect. 2.2). Galaxies selected for the simulations in such manner, contain all the relevant limits that we face in real data. This gives us several advantages: 1. the simulation naturally includes the noise properties of the original data, 2. the level of complexity in galaxy shapes is well representative of the real data, far beyond the reach of analytical galaxies models, 3. the simulation includes blending effects with companion galaxies. Even though this approach gives us a reliable way to estimate the completeness of our sample, it does not allow us to evaluate its purity. This requires a priori knowledge of which galaxies are and which are not acting as lenses. Since the galaxies are drawn randomly from the real data, they can potentially contain lensed features that would affect the results. While a visual inspection would probably solve this problem, it could however bias the completeness.

We note that the distribution of lens properties should be generally reproduced in the simulations, but that it is not necessary to reproduce the statistics of the lenses to high precision for our purposes. In the next section, we characterise the lens-detection efficiency in terms of various parameters such as S/N, Einstein radius, etc. It is necessary that the simulations fully cover the range of these parameters, but not that they reproduce the predicted distribution of parameters precisely. These simulations meet this requirement.

5.2. Completeness of the new lens sample

Our simulations include 600 systems, which match well the properties of the galaxies we preselected in CFHTLS (Sect. 2.2), as well as the noise properties of the images. We run the PCA lens-finder in the exact same way as we do on the real data, excluding the last step of a visual inspection. Our results are summarised in Fig. 7, where we compare the distribution of some of the most important observational parameters for the full population of the simulated lenses and for the population of simulated systems, actually identified as such by the PCA lens-finder. The ratio between the two histograms in Fig. 7 gives the completeness.

In our analysis, we estimate the completeness as a function of the total S/N in the (lensed) source, the number of pixels of the

source above the noise level (5σ), the source surface brightness, R_E (taken as half the averaged values of the angular separation between images) and the source redshift. We note that the calculation of the source S/N includes a noise contribution from the lensing galaxy, which can significantly impact the detection when the source and the lens overlap. Using our lens simulations we find that:

– Not surprisingly the completeness of the sample improves with increasing S/N, reaching at least 80% as soon as $S/N > 50$. Even for the lowest source S/N in the sample, the completeness is still above 50% and this number increases to 70% when $S/N > 20$. Of course, within a given S/N bin the sample spans a large range of Einstein ring size and source/lens luminosity contrast, but overall the completeness achieved by the PCA lens-finder is very high.
– The completeness depends strongly on the number of pixels above a given luminosity threshold of the lensed source. This affects the angular size of the lensed image and therefore also our ability to determine its shape (ring, full, or partial arc). Of course the more spatial resolution elements in a lensed source, the better it can be classified.
– The completeness has slightly weaker dependence on source surface brightness than on S/N of the lensed source. This shows that we fail to detect some lenses with arcs and rings that fall into the glare of the lensing galaxy. This also shows that central lensing galaxy removal with PCA method is not perfect and has an impact on the lens search.
– Importantly, there is little or no dependence of the completeness with respect to R_E or to the source redshift. This suggests that the PCA lens-finder is capable of providing unbiased samples of systems spanning a broad range of masses and redshifts, which is desirable for galaxy formation and evolution studies based on strong gravitational lensing.

6. Summary and conclusions

We have implemented a novel method, PCA-finder, for the automated detection of galaxy-scale strong gravitational lensing to a heavily explored survey, namely, the 155 square degree of imaging data of the CFHTLS *Wide*. With the PCA lens-finder we discover 109 (70 grade *A* and 39 grade *B*) brand new gravitational lens candidates. The discovery of such a large number of new lens candidates missed by other searches proves PCA-finder to be a powerful tool in discovering lenses.

The search was carried out in four steps. In Step 1, we create a uniform data cube consisting of small image stamps centred on preselected early-type galaxies. In Step 2, we subtract central galaxies from the image stamps. In Step 3, we analyse residual images created in Step 2 to look for lensing features with the DBSCAN method. In Step 4, a sample of 1 098 candidates are selected from this automated procedure. Finally, Step 5 is a visual inspection of the lens candidates by five authors of this paper (D.P., J-P.K., R.J., F.C., P.D.). Following this last step, all candidates are allocated *A*, *B*, or *C* grades.

In this paper, we present the new PCA-finder lenses and compare it with the previously known samples from the CFHTLS, namely, SPACE WRAPS, RINGFINDER and ARCFINDER. Our main results can be summarised as follows:

– PCA-finder works well as a discovery engine for gravitational lenses.
– We present a sample of 70 grade-*A* and 39 grade-*B* new gravitational lens candidates, and additional 183 grade-*C* worth

12

noticing, but with no strong evidence for lensing. We rediscover 86 lens candidates from various samples published in the literature.

– The PCA-finder selects lens systems whose statistical properties are well comparable with the RingFinder and ArcFinder samples, including the range of lens redshifts, magnitudes, and image separation.

– We also find 274 potential ring galaxies or polar ring galaxies.

– We use a sample of simulated lenses tailored to the CFHTLS *Wide* data to verify the completeness of our automated method.

The discovery of many new lens candidates through the first PCA-finder lens search illustrates the strength of the method, since we find lens candidates that other algorithms missed. Upcoming and planned wide field imaging surveys such as the DES, HSC, KIDS, Euclid and the LSST will produce a great amount of data. Reliable automated algorithms together with citizen blind search will be necessary to find lenses in these very large surveys. As shown in this paper, one approach for finding lenses from the entire survey data may not be sufficiently complete and pure. Thus, combining robotic methods for preselection with the citizen science approach for visual screening might be a good strategy for finding lenses in these large imaging surveys. For samples that are sufficiently cleaned by the automated part of the pipeline, the human time spent on the final classification remains acceptable.

References

Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, ApJ, 705, 1099
Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511
Bell, E. F. & de Jong, R. S. 2001, ApJ, 550, 212
Belokurov, V., Evans, N. W., Moiseev, A., et al. 2007, ApJ, 671, L9
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Blackburne, J. A., Wisotzki, L., & Schechter, P. L. 2008, AJ, 135, 374
Bolton, A. S., Brownstein, J. R., Kochanek, C. S., et al. 2012, ApJ, 757, 82
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Brault, F. & Gavazzi, R. 2015, A&A, 577, A85
Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, MNRAS, 341, 13
Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41
Bussmann, R. S., Pérez-Fournon, I., Amber, S., et al. 2013, ApJ, 779, 25
Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, A&A, 461, 813
Cabanac, R. A., Valls-Gabaud, D., & Lidman, C. 2008, MNRAS, 386, 2065
Cao, S., Biesiada, M., Gavazzi, R., Piórkowska, A., & Zhu, Z.-H. 2015, ApJ, 806, 185
Collett, T. E. 2015, ApJ, 811, 20
Conley, A., Cooray, A., Vieira, J. D., et al. 2011, ApJ, 732, L35
Coupon, J., Ilbert, O., Kilbinger, M., et al. 2009, A&A, 500, 981
Courbin, F., Chantry, V., Revaz, Y., et al. 2011, A&A, 536, A53
Cuillandre, J.-C. J., Withington, K., Hudelot, P., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8448, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0
Dalal, N. & Kochanek, C. S. 2002, ApJ, 572, 25
Easter, M., Kriegel, H., Sander, J., & Xu, X. 1996, In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining
Einstein, A. 1916, Annalen der Physik, 356, 639
Elyiv, A., Melnyk, O., Finet, F., et al. 2013, MNRAS, 434, 3305
Erben, T., Hildebrandt, H., Lerchster, M., et al. 2009, A&A, 493, 1197
Fassnacht, C. D., Moustakas, L. A., Casertano, S., et al. 2004, ApJ, 600, L155
Faure, C., Kneib, J.-P., Covone, G., et al. 2008, ApJS, 176, 19
Gavazzi, G., Fumagalli, M., Cucciati, O., & Boselli, A. 2010, A&A, 517, A73
Gavazzi, R., Cooray, A., Conley, A., et al. 2011, ApJ, 738, 125
Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144
Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 677, 1046
Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, ApJ, 667, 176
González-Nuevo, J., Lapi, A., Fleuren, S., et al. 2012, ApJ, 749, 65
Gregg, M. D., Becker, R. H., Brotherton, M. S., et al. 2000, ApJ, 544, 142
Hezaveh, Y. D., Marrone, D. P., Fassnacht, C. D., et al. 2013, ApJ, 767, 132
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Inada, N., Becker, R. H., Burles, S., et al. 2003, AJ, 126, 666
Inada, N., Oguri, M., Becker, R. H., et al. 2007, AJ, 133, 206
Inada, N., Oguri, M., Shin, M.-S., et al. 2012, AJ, 143, 119
Jiang, G. & Kochanek, C. S. 2007, ApJ, 671, 1568
Jolliffe, I. T. 1986, Principal component analysis
Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, A&A, 566, A63
Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599
Magain, P., Surdej, J., Swings, J.-P., Borgeest, U., & Kayser, R. 1988, Nature, 334, 325
Marshall, P. J., Hogg, D. W., Moustakas, L. A., et al. 2009, ApJ, 694, 924
Maturi, M., Mizera, S., & Seidel, G. 2014, A&A, 567, A111
Metcalf, R. B. & Madau, P. 2001, ApJ, 563, 9
Metcalf, R. B. & Petkova, M. 2014, MNRAS, 445, 1942
More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38
More, A., Verma, A., Marshall, P. J., et al. 2016, MNRAS, 455, 1191
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Moustakas, L. A., Marshall, P., Newman, J. A., et al. 2007, ApJ, 660, L31
Myers, S. T., Jackson, N. J., Browne, I. W. A., et al. 2003, MNRAS, 341, 1
Negrello, M., Hopwood, R., De Zotti, G., et al. 2010, Science, 330, 800
Oguri, M., Inada, N., Clocchiatti, A., et al. 2008, AJ, 135, 520
Oguri, M., Inada, N., Pindor, B., et al. 2006, AJ, 132, 999
Oguri, M., Rusu, C. E., & Falco, E. E. 2014, MNRAS, 439, 2494
Oke, J. B. & Gunn, J. E. 1983, ApJ, 266, 713
Overzier, R., Lemson, G., Angulo, R. E., et al. 2013, MNRAS, 428, 778
Pawase, R. S., Courbin, F., Faure, C., Kokotanekova, R., & Meylan, G. 2014, MNRAS, 439, 3392
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Petkova, M., Metcalf, R. B., & Giocoli, C. 2014, MNRAS, 445, 1954
Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, A&A, 557, A44
Ratnatunga, K. U., Griffiths, R. E., & Ostrander, E. J. 1999, AJ, 117, 2010
Refsdal, S. 1964, MNRAS, 128, 307
Rusin, D. & Kochanek, C. S. 2005, ApJ, 623, 666
Rusin, D., Kochanek, C. S., Falco, E. E., et al. 2003, ApJ, 587, 143
Seidel, G. & Bartelmann, M. 2007, A&A, 472, 341
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2012, ApJ, 752, 163
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98
Surdej, J., Magain, P., Swings, J.-P., et al. 1987, Nature, 329, 695
Suyu, S. H. 2012, MNRAS, 426, 868
Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70
Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201
Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJ, 788, L35
Sygnet, J. F., Tu, H., Fort, B., & Gavazzi, R. 2010, A&A, 517, A25
Tewes, M., Courbin, F., Meylan, G., et al. 2013, A&A, 556, A22
Tramacere, A. & Vecchio, C. 2013, Astronomy and Astrophysics, 549, A138
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
Treu, T., Koopmans, L. V., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 640, 662
Treu, T. & Koopmans, L. V. E. 2004, ApJ, 611, 739
Treu, T., Stiavelli, M., Casertano, S., Møller, P., & Bertin, G. 2002, ApJ, 564, L13
Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341
Vuissoz, C., Courbin, F., Sluse, D., et al. 2008, A&A, 488, 481
Wardlow, J. L., Cooray, A., De Bernardis, F., et al. 2013, ApJ, 762, 59
Willis, J. P., Hewett, P. C., & Warren, S. J. 2005, MNRAS, 363, 1369
Willis, J. P., Hewett, P. C., Warren, S. J., Dye, S., & Maddox, N. 2006, MNRAS, 369, 1521
Winn, J. N., Hewitt, J. N., Patnaik, A. R., et al. 2001, AJ, 121, 1223
Winn, J. N., Hewitt, J. N., Schechter, P. L., et al. 2000, AJ, 120, 2868
Winn, J. N., Lovell, J. E. J., Chen, H.-W., et al. 2002a, ApJ, 564, 143
Winn, J. N., Morgan, N. D., Hewitt, J. N., et al. 2002b, AJ, 123, 10
Wisotzki, L., Christlieb, N., Liu, M. C., et al. 1999, A&A, 348, L41
Wisotzki, L., Koehler, T., Kayser, R., & Reimers, D. 1993, A&A, 278, L15
Wisotzki, L., Koehler, T., Lopez, S., & Reimers, D. 1996, A&A, 315, L405
Wisotzki, L., Schechter, P. L., Bradt, H. V., Heinmüller, J., & Reimers, D. 2002, A&A, 395, 17
Wisotzki, L., Schechter, P. L., Chen, H.-W., et al. 2004, A&A, 419, L31

Paraficz: The PCA Lens-Finder: application to CFHTLS

Appendix A: Observational data

In the following we provide colour stamps for our 109 lens candidates classified as grade-A or grade-B (Figs. A.1-A.5). We also provide a list of the objects we classify as ring-like galaxies (Table A.1) or as a grade-C lens candidates (Table A.2).

14

ID	RA	DEC	ID	RA	DEC	ID	RA	DEC	ID	RA	DEC
1	30.7113	-4.1200	70	208.8451	56.5793	139	212.0681	51.7086	207	216.0082	55.7079
2	32.0140	-7.3349	71	208.8577	55.2961	140	212.0702	54.6277	208	216.1492	53.3707
3	32.7019	-11.091	72	208.8897	56.7042	141	212.0781	53.1058	209	216.1771	56.1187
4	32.9684	-10.640	73	208.9984	57.6036	142	212.1310	56.5988	210	216.2110	53.6238
5	33.3126	-4.5772	74	209.1030	54.2619	143	212.1942	52.2143	211	216.3065	57.5956
6	34.0313	-6.8846	75	209.1158	53.3728	144	212.2530	53.6661	212	216.3152	54.6290
7	34.6449	-6.2790	76	209.1330	57.7878	145	212.3300	52.7731	213	216.4620	56.7888
8	35.3084	-9.6229	77	209.1440	53.1748	146	212.3553	55.9526	214	216.6035	54.0957
9	36.2642	-6.6648	78	209.1640	57.7423	147	212.4420	52.1788	215	216.6826	52.8987
10	37.7604	-5.9290	79	209.1907	56.3655	148	212.4687	51.9757	216	216.9283	57.2185
11	132.1819	-2.5873	80	209.2600	57.224	149	212.4777	56.5225	217	217.1591	55.7337
12	132.2712	-2.7720	81	209.2658	53.3428	150	212.6073	56.4597	218	217.2200	55.6851
13	132.5795	-3.1395	82	209.3485	54.5476	151	212.6626	57.4131	219	217.2537	55.4910
14	132.5827	-2.1625	83	209.3499	54.3683	152	212.6708	52.8456	220	217.3800	55.4627
15	132.5857	-4.6938	84	209.4350	56.9008	153	212.8555	55.0499	221	217.4268	53.9160
16	132.6970	-3.9655	85	209.4560	56.9913	154	212.9167	54.0741	222	217.4859	51.7736
17	132.7400	-1.0218	86	209.5048	57.6376	155	212.9404	57.5752	223	217.5249	52.7104
18	132.7620	-4.4793	87	209.6050	56.5502	156	212.9745	55.3304	224	217.5592	56.9327
19	132.9841	-1.9516	88	209.6506	56.1349	157	213.0125	51.3368	225	217.6319	57.7677
20	133.0210	-4.6829	89	209.7750	53.7688	158	213.1040	53.7781	226	217.7050	55.9099
21	133.1227	-1.7640	90	209.8462	56.0677	159	213.1345	52.4077	227	217.8106	57.2403
22	133.1827	-1.0132	91	209.8569	53.1140	160	213.1900	53.6537	228	217.9196	54.5902
23	133.2438	-3.7043	92	210.0242	53.2755	161	213.2580	53.1980	229	218.0436	54.5195
24	133.5939	-4.9943	93	210.0400	54.7716	162	213.2939	57.2640	230	218.0770	54.9706
25	133.6931	-2.7340	94	210.0919	52.5197	163	213.3095	52.4831	231	218.3619	55.3975
26	133.7100	-5.0446	95	210.1264	54.9506	164	213.5020	56.2184	232	218.5645	52.6184
27	133.7260	-3.9403	96	210.2546	52.4485	165	213.5440	57.3132	233	218.6160	57.6086
28	133.7880	-1.4389	97	210.2719	57.5983	166	213.5494	52.3798	234	218.6570	57.5602
29	133.9045	-4.3638	98	210.3997	57.7682	167	213.7320	52.3263	235	218.8210	55.4744
30	133.9124	-2.3331	99	210.4121	53.2592	168	213.7877	52.2259	236	218.8813	54.4891
31	133.9648	-4.7568	100	210.4300	51.2151	169	213.8380	57.1487	237	218.9065	51.3609
32	134.1370	-3.0186	101	210.4339	55.3332	170	213.9270	56.0135	238	219.3108	54.4433
33	134.1507	-1.3852	102	210.4797	51.2445	171	213.9639	51.3241	239	219.4381	56.3716
34	134.2170	-3.9283	103	210.4847	57.6922	172	213.9691	51.9893	240	219.5293	53.7316
35	134.2991	-4.3455	104	210.4876	53.2545	173	213.9950	51.5452	241	219.6096	52.1294
36	134.4163	-2.7038	105	210.4968	55.0219	174	214.0239	57.0921	242	219.6133	52.4955
37	134.4683	-3.6448	106	210.5020	53.1556	175	214.1700	54.0163	243	220.2990	57.7254
38	134.5020	-3.2156	107	210.5154	52.3060	176	214.2040	55.4237	244	330.1120	1.5701
39	134.5101	-3.6909	108	210.6305	51.6363	177	214.4082	57.2170	245	330.2540	3.6684
40	134.5972	-2.1057	109	210.7200	52.8003	178	214.4454	51.5445	246	330.2720	1.8044
41	134.7143	-3.3857	110	210.7309	52.3358	179	214.5368	53.1975	247	330.4520	3.6380
42	134.8220	-4.1856	111	210.7704	53.7681	180	214.6275	56.8027	248	330.5939	1.2562
43	134.9660	-4.5596	112	210.8229	51.8781	181	214.6400	55.6239	249	330.8630	4.4709
44	135.0381	-3.3306	113	210.8282	56.4585	182	214.6506	54.2784	250	330.8980	2.2835
45	135.1403	-3.8596	114	210.9047	53.0148	183	214.6966	56.8436	251	331.3460	1.3502
46	135.3235	-5.5642	115	210.9170	53.8914	184	214.7002	53.3850	252	331.6890	1.0121
47	135.3369	-1.1031	116	211.0441	52.8848	185	214.7420	56.4807	253	331.8380	2.8518
48	135.3720	-1.5786	117	211.0709	57.2175	186	214.8210	56.2483	254	331.9330	2.3048
49	135.5498	-1.4060	118	211.1030	52.1859	187	215.0682	56.3292	255	332.0440	3.5210
50	135.5651	-2.0581	119	211.1110	57.0090	188	215.0730	56.531	256	332.1489	1.9636
51	135.5864	-2.2146	120	211.1478	52.4499	189	215.0828	54.9762	257	332.2340	3.9315
52	135.8341	-1.1346	121	211.2522	52.0962	190	215.2030	56.333	258	332.3800	0.5319
53	135.9001	-2.1360	122	211.2820	56.6721	191	215.2367	52.8727	259	332.6879	1.7517
54	135.9940	-3.2371	123	211.2950	53.9345	192	215.2490	51.4651	260	333.1200	0.8459
55	136.0742	-4.1396	124	211.4487	57.3926	193	215.3003	57.7346	261	333.1270	2.4188
56	136.1734	-2.0613	125	211.5459	56.7882	194	215.3255	57.5031	262	333.1709	0.3183
57	136.2500	-4.7148	126	211.5520	56.0455	195	215.3361	55.3473	263	333.2590	-0.8181
58	136.2506	-2.2494	127	211.6047	51.6347	196	215.3542	57.4317	264	333.4410	0.4691
59	136.3568	-2.6909	128	211.6335	52.0406	197	215.4756	56.7486	265	333.9400	1.5399
60	136.5063	-5.1453	129	211.7230	54.7437	198	215.5070	56.9580	266	334.1099	1.0777
61	136.5580	-1.1363	130	211.7706	54.6133	199	215.6347	54.5268	267	334.3880	1.2319
62	136.5629	-4.7081	131	211.7880	53.3319	200	215.6390	51.7208	268	334.4719	1.3240
63	136.5651	-2.1079	132	211.8378	52.5960	201	215.7157	55.2452	269	334.4760	2.7216
64	136.6360	-4.9990	133	211.8520	54.4707	202	215.8014	57.2830	270	334.6690	-0.4942
65	136.6770	-1.4028	134	211.8555	53.8529	203	215.9419	51.6791	271	335.2009	1.1618
66	136.7488	-1.6531	135	211.8660	54.7334	204	215.9590	51.6543	272	335.2869	0.1860
67	136.7710	-1.2799	136	211.8971	54.1434	205	215.9689	57.2974	273	335.3330	1.0047
68	208.6711	57.7116	137	211.9844	54.6894	206	216.0003	54.6786	274	335.5480	0.5849
69	208.7878	56.4358	138	212.0647	52.3376						

Table A.1. List of ring-like galaxies.

ID	RA	DEC	ID	RA	DEC	ID	RA	DEC	ID	RA	DEC
1	30.7690	-4.3707	47	37.9131	-8.4071	93	210.3239	57.0069	139	213.6360	53.4336
2	31.1471	-6.8370	48	132.6063	-1.6624	94	210.3239	57.0069	140	213.8290	51.5396
3	31.1845	-9.2302	49	132.8670	-1.7824	95	210.3830	52.9747	141	213.8425	54.5827
4	31.1953	-7.4963	50	133.0932	-5.5540	96	210.4185	51.9295	142	213.8930	55.9188
5	31.3286	-9.4541	51	133.1093	-2.1114	97	210.4440	56.0728	143	214.0553	56.3335
6	31.4334	-5.5922	52	133.6085	-3.3218	98	210.4970	55.0210	144	214.1416	54.2238
7	31.4347	-8.9391	53	133.8673	-4.4843	99	210.5960	56.7669	145	214.1440	52.1982
8	31.7222	-6.9676	54	134.1219	-2.8850	100	210.8169	56.3686	146	214.2251	53.2605
9	32.0885	-10.1001	55	134.3641	-3.8366	101	210.8650	54.0454	147	214.4718	56.4726
10	32.1265	-8.6989	56	134.4226	-5.5544	102	210.9850	53.6275	148	214.6783	52.0068
11	32.1489	-10.6963	57	134.9147	-1.7250	103	210.9875	52.7897	149	214.8440	52.0608
12	32.3864	-8.6895	58	135.0480	-4.2772	104	211.0765	56.2987	150	214.8870	55.7473
13	32.5458	-6.9613	59	135.3490	-2.7373	105	211.0960	52.7181	151	214.9260	56.2809
14	32.6089	-4.2655	60	135.3985	-4.6624	106	211.1021	56.0841	152	215.2367	52.8727
15	32.6312	-5.3873	61	135.4108	-4.9603	107	211.2520	52.0962	153	215.5535	52.0773
16	32.7019	-11.0916	62	135.5910	-2.0328	108	211.2966	53.9393	154	215.5858	52.3223
17	32.7654	-10.1586	63	135.7976	-3.1933	109	211.2975	52.6984	155	215.5870	52.5406
18	32.7713	-4.3339	64	135.8190	-1.4759	110	211.4610	56.5435	156	215.9215	53.1101
19	32.7958	-9.1606	65	135.8390	-4.7139	111	211.5791	51.5461	157	216.0830	56.5382
20	32.8151	-4.6442	66	135.8770	-3.3166	112	211.6760	56.8844	158	216.1280	53.7346
21	32.8263	-5.9572	67	135.8916	-4.7350	113	211.7245	54.9516	159	217.1294	53.2367
22	33.1106	-9.1819	68	136.1790	-2.0222	114	211.7933	57.7058	160	217.2020	57.1215
23	33.1371	-8.2071	69	136.3051	-1.7993	115	211.8378	52.5960	161	217.3254	51.2955
24	33.4489	-5.0069	70	136.4287	-4.3918	116	211.8980	54.3523	162	217.3837	57.0770
25	33.7189	-10.2549	71	136.6511	-4.1206	117	211.9132	56.5606	163	217.4874	53.4669
26	33.8125	-7.6329	72	136.6759	-3.4018	118	212.0111	54.5693	164	217.5479	53.9042
27	33.8846	-7.3768	73	136.7700	-3.6983	119	212.3363	53.7088	165	218.5887	53.5343
28	34.0291	-10.4792	74	208.5923	56.9917	120	212.5192	52.8386	166	218.9180	51.5978
29	34.0313	-6.8846	75	209.0410	55.1548	121	212.5344	53.5059	167	218.9306	51.6789
30	34.0420	-4.9278	76	209.0520	55.3643	122	212.5578	57.1932	168	218.9306	51.6789
31	34.2354	-7.3950	77	209.0883	54.6384	123	212.6180	54.6689	169	219.2150	53.1183
32	34.4127	-5.6183	78	209.1308	56.8001	124	212.7260	56.5193	170	219.2270	54.7119
33	34.7033	-6.3146	79	209.3442	56.0257	125	212.7260	54.2777	171	219.2620	53.0397
34	34.7611	-5.6861	80	209.4020	56.8134	126	212.7260	56.5193	172	219.4122	56.8565
35	34.7705	-8.1057	81	209.6070	52.7409	127	212.7370	53.3133	173	219.5995	53.3351
36	35.1727	-10.8737	82	209.6217	52.6670	128	212.8499	56.3970	174	219.6419	54.4867
37	35.4835	-10.6293	83	209.6358	56.3116	129	212.8877	56.1517	175	219.6730	56.9740
38	35.4901	-6.8625	84	209.8516	54.3894	130	212.9065	53.8578	176	331.1035	1.3934
39	35.5342	-8.2175	85	209.9950	56.2114	131	212.9810	52.5545	177	331.8447	4.2985
40	36.0965	-3.9120	86	210.0480	53.4897	132	213.1729	54.4224	178	331.9094	1.6015
41	37.1595	-8.0260	87	210.1192	56.0877	133	213.1870	55.2213	179	332.0385	1.5294
42	37.3517	-11.1634	88	210.1260	54.9506	134	213.3750	53.4311	180	332.0615	2.6110
43	37.3725	-9.9510	89	210.1920	55.5750	135	213.4316	53.0792	181	332.1574	3.3451
44	37.7770	-10.0155	90	210.2130	52.9077	136	213.5034	55.1556	182	332.7543	0.0248
45	37.8688	-9.2495	91	210.2437	56.8107	137	213.5379	52.4758	183	333.8364	0.9369
46	37.9131	-8.4071	92	210.3010	57.0831	138	213.5680	54.4716			

Table A.2. List of our grade-*C* candidates in CFHTLS.

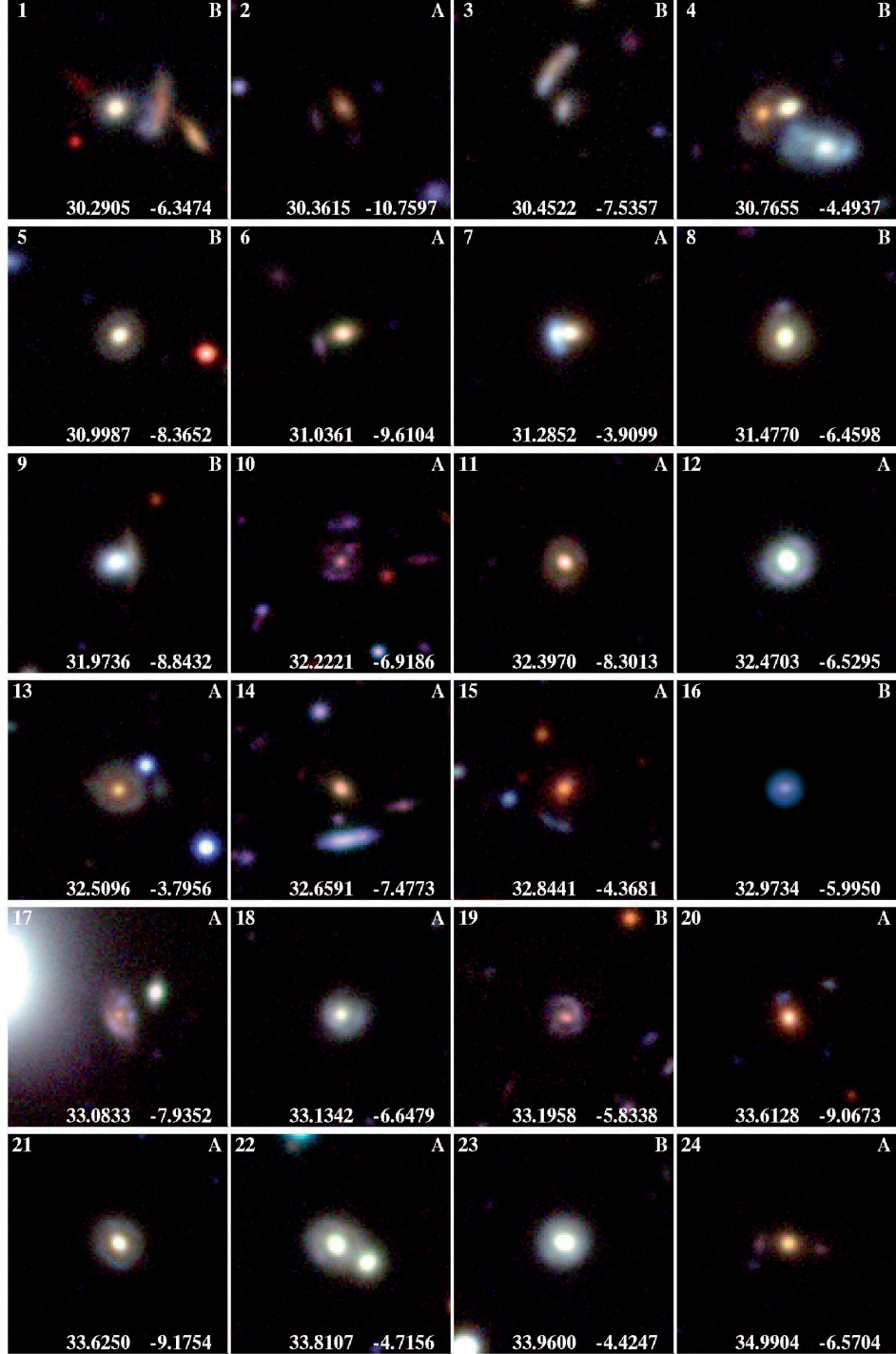**Fig. A.1.** Our new lenses found with the PCA-finder, with grades *A* and *B* (see text). The stamps are 18.7″ on-a-side.
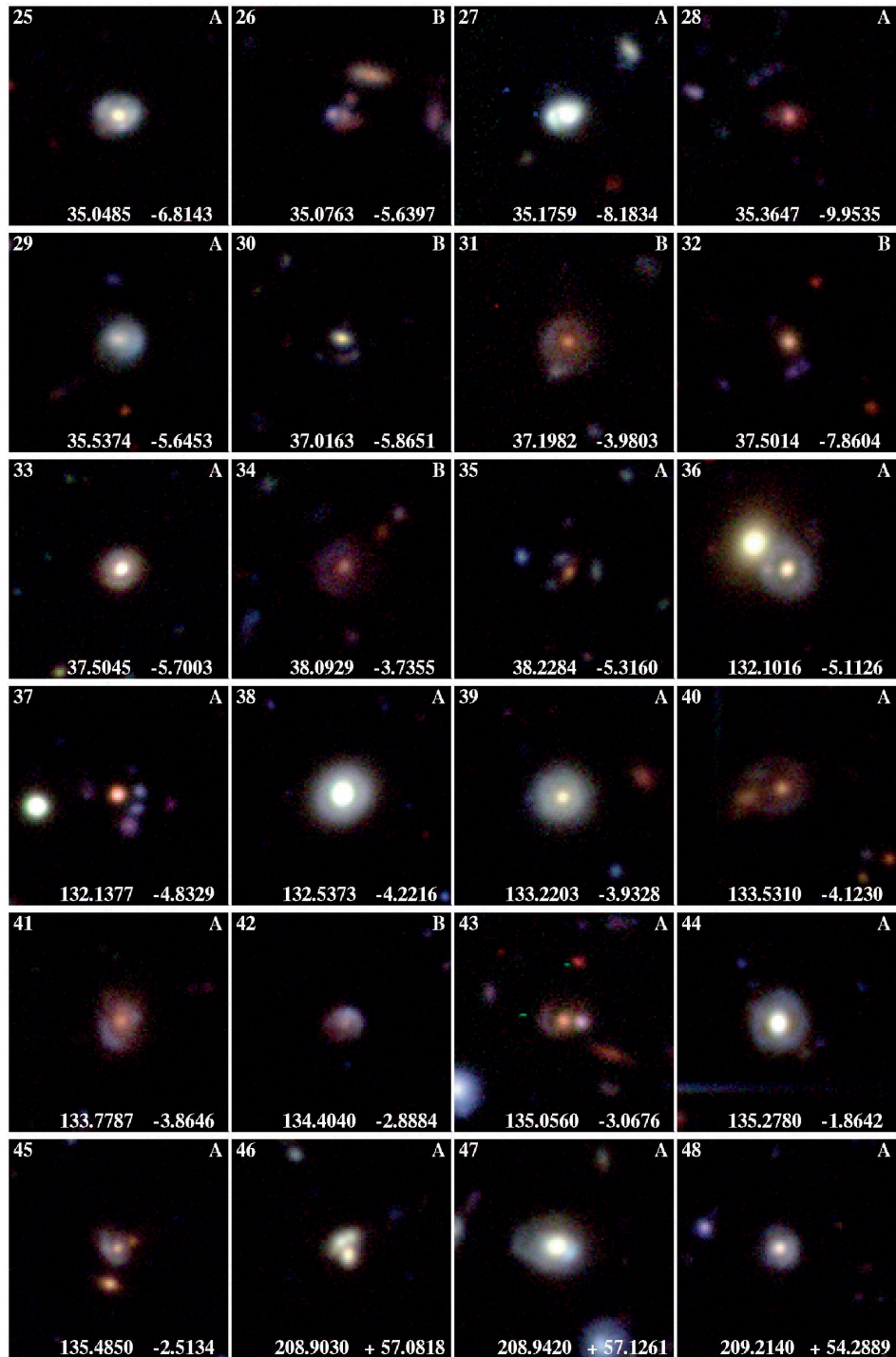
| 25	A	26	B	27	A	28	A
35.0485	-6.8143	35.0763	-5.6397	35.1759	-8.1834	35.3647	-9.9535
29	A	30	B	31	B	32	B
35.5374	-5.6453	37.0163	-5.8651	37.1982	-3.9803	37.5014	-7.8604
33	A	34	B	35	A	36	A
37.5045	-5.7003	38.0929	-3.7355	38.2284	-5.3160	132.1016	-5.1126
37	A	38	A	39	A	40	A
132.1377	-4.8329	132.5373	-4.2216	133.2203	-3.9328	133.5310	-4.1230
41	A	42	B	43	A	44	A
133.7787	-3.8646	134.4040	-2.8884	135.0560	-3.0676	135.2780	-1.8642
45	A	46	A	47	A	48	A
135.4850	-2.5134	208.9030	+ 57.0818	208.9420	+ 57.1261	209.2140	+ 54.2889

Fig. A.2. Our new lenses found with the PCA-finder, with grades *A* and *B* (see text). The stamps are 18.7″on-a-side.

106
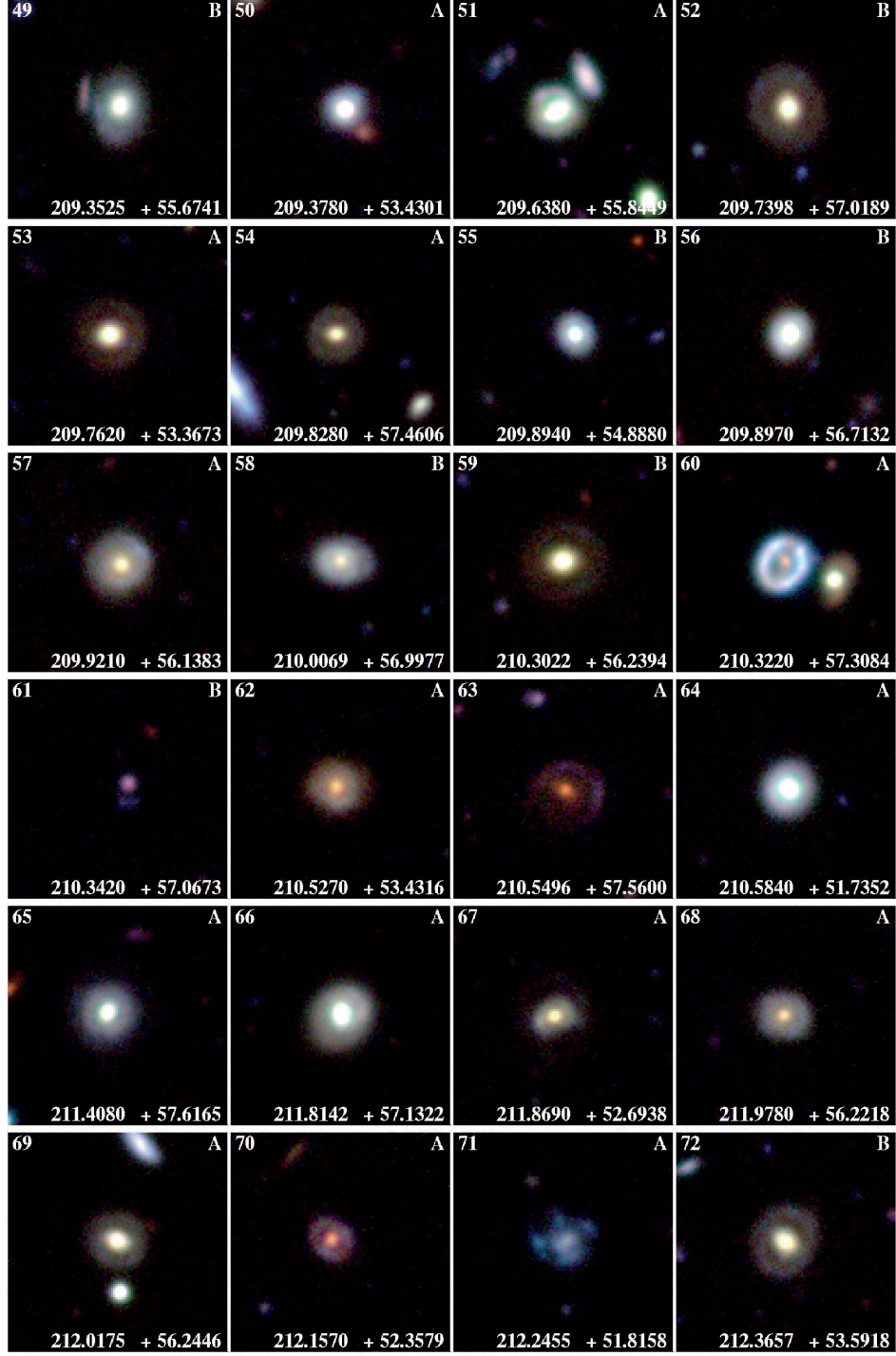
Fig. A.3. Our new lenses found with the PCA-finder, with grades *A* and *B* (see text). The stamps are 18.7″on-a-side.

Paraficz: The PCA Lens-Finder: application to CFHTLS
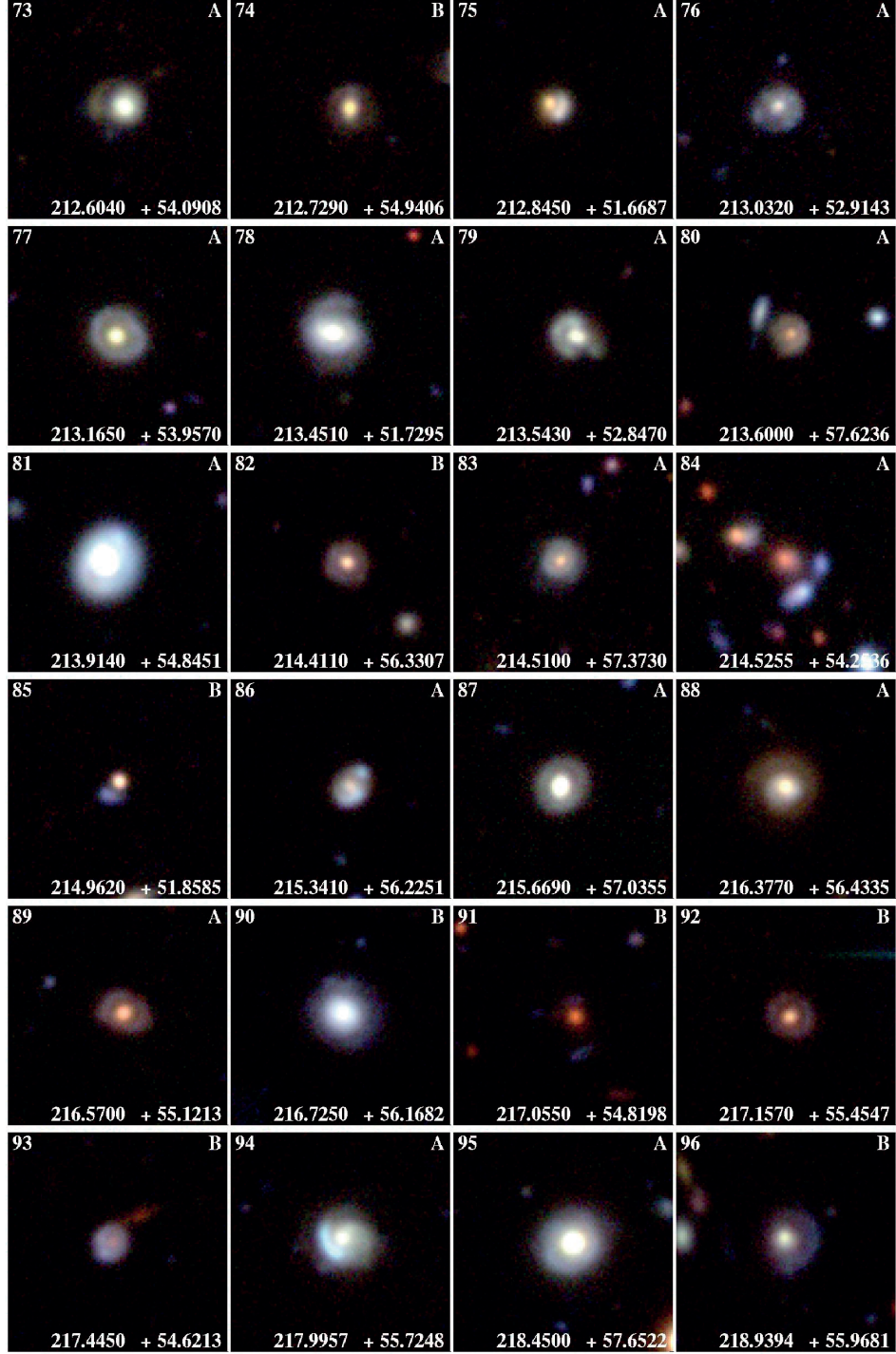
Fig. A.4. Our new lenses found with the PCA-finder, with grades *A* and *B* (see text). The stamps are 18.7″ on-a-side.
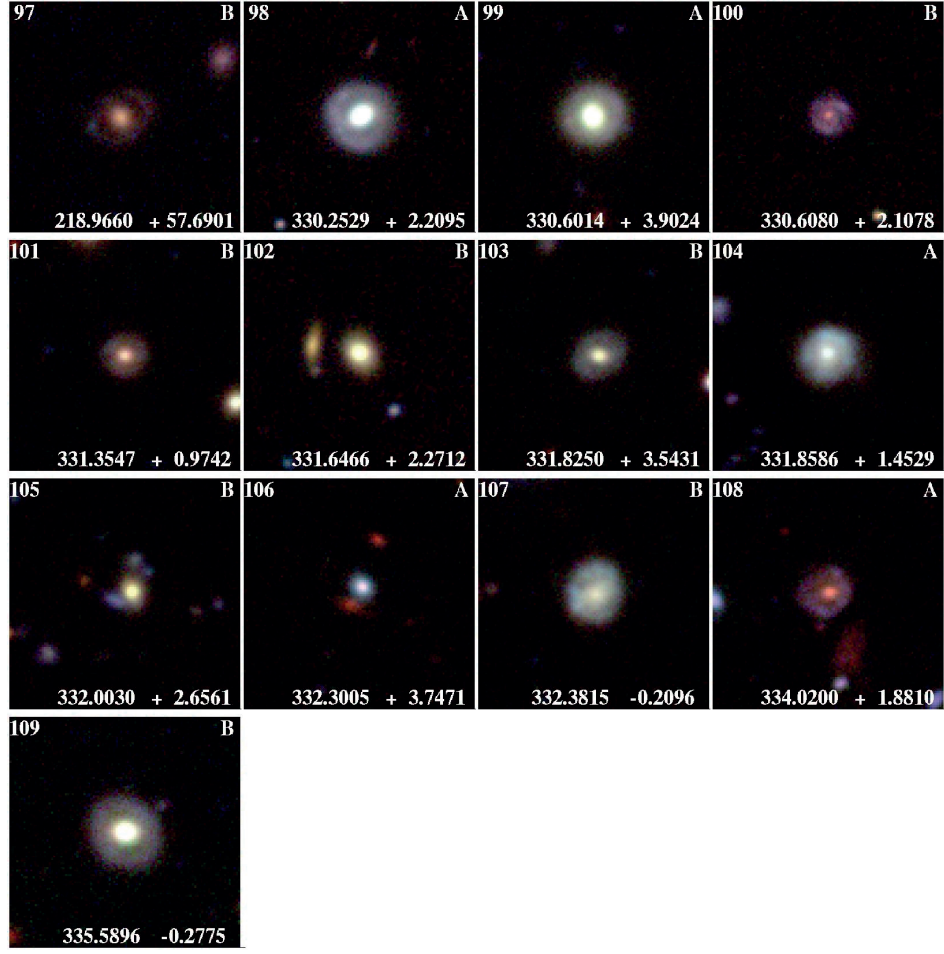
Fig. A.5. Our new lenses found with the PCA-finder, with grades *A* and *B* (see text). The stamps are 18.7″ on-a-side.

3.5 Outline

In this chapter, I detailed a technique for semi-automated lens search that relies on the subtraction of light profile of potential lens candidates to search for lensed structures in the glare of elliptical galaxies. We successfully applied our method to the CFHTL Survey, where we found 109 new lens systems that were not found in previous searches. The reason for the finding of new systems is imputed to the properties of the lens finding algorithm. This suggests that a thorough search for lenses should include several lens finding algorithms, which properties in terms of characteristics of expected lenses are well understood.

With the rapid growth of machine learning-based techniques, the feature extraction strategy that we developed to discriminate between lenses and non-lenses became quickly outdated. Indeed, most ML-based methods are fully automated and provide reliable lens candidate sets, while our feature extraction techniques required the visual inspection of a sample of lenses ten times bigger as our final selection. Such small purity of the selection makes the method intractable in the frame of large surveys where hundreds of thousands of lens systems are expected to be found. This realisation illustrate how important it is that we develop lens finding techniques that have high purity in order to avoid the visual inspection of large samples of candidates. Due to the large number of expected lenses in future surveys we could imagine enforcing purity while sacrificing to the completeness.

Nonetheless, the PCA subtraction method provides a reliable technique to model lens light distribution, which is necessary to the study of the lensed sources. While PCA constitutes a naive implementation of ML techniques, it paves the way for the potential development of the modelling of galaxies based on more sophisticated and robust ML strategies as shown by the works if Ravanbakhsh et al. (2016).

This first technique for strong gravitational lens deblending relies on the modelling of lens galaxies by use of the similarities between images of galaxies in a large survey. In essence, this approach makes little use of the properties of gravitational lenses, for the reason that it was developed to find them. As a result, this technique is able to remove the light profile of any galaxy on the plane of sky. In the following chapter, we will specify our deblending techniques to the address the specific problems posed by gravitational lens imaging by making use of the differences between lensed and lens galaxy light profiles.

4 Colour Deblending for Strong Gravitational Lensing

Contents

As we have seen, strong gravitational lenses form by the alignment of galaxies at different redshift. Because observing distant objects is also observing these objects as they were a long time ago, the farther an object, the younger it is. As a result, the background of a gravitational lens is always a galaxy at a younger stage of its evolution resulting in blue stellar population. Galaxies at a later stage of their evolution will naturally be composed of older stars that emit with a redder spectra. In addition to this, gravitational lenses are often around early type galaxies (e.g. the SLACS sample Bolton et al., 2006), which are galaxies that stopped forming stars early on in their evolution. These galaxies are often found in massive galaxy clusters and are, on average heavier than spiral galaxies, hence the higher likelihood for gravitational lens systems formation. The result of this is that in most cases gravitational lenses are composed

of a red elliptical foreground lens and a blue background spiral or irregular lensed source, as illustrated by the spectacular system in figure 4.1 [1].

Figure 4.1 – The "cosmic horseshoe" lens system.

In this image in particular, the separation between lens and source and source galaxy light profiles is clear due to the angular separation between the lens and the images of the source, but also due to the difference in colour between the objects. This latter observation was the starting point of a reflection on how to make our algorithms see in colours. Indeed, if our eye is able to make the difference between these objects due to their difference in colour, it should be possible to separate these light components based on their relative contributions to different colour filter imaging. This idea is not dissimilar to the idea presented in Gavazzi et al. (2014) and Chan et al. (2015), where tha authors subtracted together bands where the lens light profile dominates from bands where lens and source light profiles are mixed. While this is a perfectly acceptable strategy in the case of the search for lens galaxies, the approximation that one of the light sources emits specifically in one band while the other contributes to both might lead to the subtraction of part of the signal from the lensed source.

One of the motivations for the development of such method comes from the limitations met by other techniques. A very widespread approach consists in fitting elliptical profiles of surface brightness to the lens galaxy light profile. In most recent lens modelling applications, this fitting procedure is adequately performed while modelling the other components of the lens, namely, the mass distribution and the source light profile. In cases such as gravitational lensing by galaxy clusters however, it is impossible to implement such strategy due to the size and the complexity of the systems to model. It is therefore common to proceed to the subtraction of the light from the foreground galaxies prior to any analysis. This strategy has the

---

[1]Image credit: ESA/Hubble & NASA

drawback to neglect the light contribution from background source, which might result in an overestimation of the luminosity in the outer regions of the galaxy cluster members. Another limitation of this method comes from the number of components used in fitting the lens light profile. This requires an a exact knowledge of the number of light sources and components per source, and may lead to a large number of parameters to represent the total luminosity profile of the lens galaxies. In the case of the PCA approach, a key requirement to the application of the method is that the morphology of the galaxy to subtract be well represented in the survey where the image of interest originates from. In the case of very large surveys, this is usually not a problem due to the shear number and variety of the sources in the sky. However, when dealing with high resolution data with a limited field of view, the repetition of specific features of interest is not granted. In particular, in the case of galaxy clusters, the occurrence of large isolated cluster members in a survey is usually, thus limiting the number of galaxies the PCA can use to learn galactic features and shapes.

Another reason to take interest in the separation of galaxies based on their colours is that this criteria can be expressed in terms of a linear problem as we will demonstrate in the next sections. Due to the linearity of the problem, it is possible to find numerical solutions that do not depend on a predefined model for the light sources.

In this chapter, we will see how to express the difference in colours as a linear problem and use the results presented in section 2 to derive solutions. The method, called `MuSCADeT`, and the code we designed for this task are described in a published work, reproduced a s a pre-print in section 4.2. The method was applied to several data sets. One of these applications was used to contribute to the constraint of the mass model of cluster Abell 3827 which was published in the monthly notices of the royal astronomical society and is reproduced here as a pre-print in section 4.3.1. In section 4.3.2 we show the result of the application of `MuSCADeT` to the whole Hubble Frontier Fields (HFF, Lotz et al., 2017). The results of the colour separated components are then used to measure the luminosity of lensed background sources and give estimates of magnification ratios across the cluster.

4.1 Multi-Band Observations

In this section, I would like to give a justification for the linear model we build to represent colour data. While the method paper presented in the next section shows how we model mathematically multi-band data to extract colour information, I use the current section to give a better justification to the linear model by looking at how multi-band data are acquired in the dataset that we treated using `MuSCADeT`: the Hubble frontier fields.

4.1.1 The Hubble Frontier Fields

Initially designed for multi-purpose multi-band applications, we developed and tested `MuSCADeT` on the the Hubble frontier fields survey (Lotz et al., 2017) acquired with the Hubble space

Cluster	z	M_{vir} $(M_\odot)$	Lensing probability
Abell 2744	0.308	1.8×10^{15}	0.69 ± 0.07
MACS J0717	0.545	$2 - 3 \times 10^{15}$	0.84 ± 0.05
MACS j1149	0.543	2.5×10^{15}	0.60 ± 0.10
MACS J0416	0.396	1.2×10^{15}	0.63 ± 0.12
Abell 370	0.375	$\sim 1 \times 10^{15}$	0.90 ± 0.08
Abell S1063	0.348	1.4×10^{15}	0.69 ± 0.08

Table 4.1 – Mass, redshift and lensing probability of the Hubble frontier fields clusters.

telescope. The goal of this survey was to provide deep observations, down to AB magnitudes < 29.1, of 6 galaxy clusters, massive enough to produce high magnification images of distant background galaxies. The characteristics in terms of redshifts, and Virial mass estimated from lens modelling are given in table 4.1 (Lotz et al., 2017). In this table,the lensing probability is the the likelihood that a source at redshift $z = 9.6$ be magnified to AB magnitude < 27 in the F160W filter, which corresponds to a condition of observability of the source due to the targeted highest magnitude in filter F160W being 28.7. The lensing probability can be seen as a measure of strength of the lensing by the corresponding cluster.

The six galaxy clusters were observed in three optical bands: F435w, F606w and F814w with the HST/ACS camera and in four near infrared (NIR) bands: F105w, F125w, F140w and F160w of HST/WFC3/IR instrument. While one of the instruments (ACS or WFC3/IR) was pointing the cluster centre, the other instrument was left to point a parallel field of view in the outskirts of the cluster. In the rest of this chapter we will differentiate these fields of view by naming *cluster fields* the field of view centred on the cluster core, and *parallel field* the one centred on the outskirts region. In our analysis of the frontier fields, we focused on the optical images only, acquired with the ACS instrument, in order to avoid the difference in resolution between the two cameras.

4.1.2   Acquisition of filter images

The ACS instrument is equiped with three colour filters in the optical. The colour filters are band-pass filters which response is given in figure 4.2[2] where the cyan green and magenta curves give the dimensionless throughput of the F435w, F606w and F814 colour filters respectively as a function of wavelength. The ACS camera, when pointing towards a target, therefore captures the light received at the corresponding wavelength of the filters in proportion to the respective throughput at a given wavelength. Figure 4.2 also contains the spectra of three galaxies redshifted to the lens and source redshifts of the horseshoe lens. In blue are shown the templates of two spiral galaxies seen at redshift $z = 2.381$. The red curve shows a template of an elliptical galaxy seen at redshift $z = 0.44$. The spectral templates are credited to Gwyn

---

[2]Data for the figure were collected from http://www.stsci.edu/hst/acs/analysis/throughputs
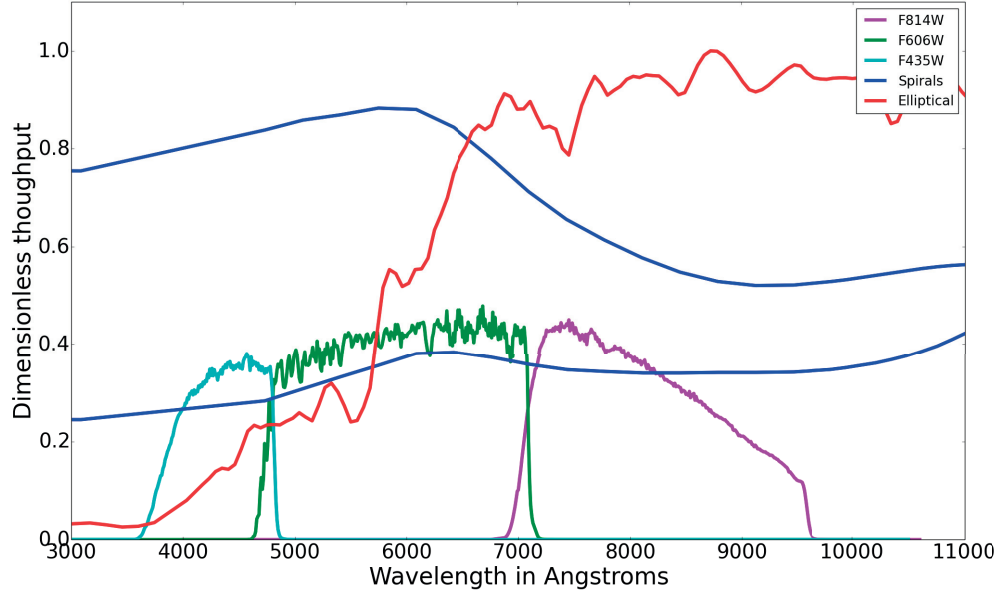
Figure 4.2 – HST/ACS filters and three galactic spectral templates at the redshifts of the horsehoe light sources.

(2012); Coleman et al. (1980)[3]. They are normalised such that the maximum of the curve is set to one for the sake of the representation.

This figure serves the purpose of showing that in a typical case of lensing by a red early type galaxy of a blue background source, each galaxy contributes to each of the three ACS optical bands. Observing a source through these three filters results in the acquisition of a cube of three images where each image corresponds to the intensity of an object red, green and blue colours. A colour image can then be formed by the superimposition of the three images seen in the corresponding colour as illustrated in figure 4.3. In figure 4.3 the top panels show the surface brightness of the horseshoe lens system in three different HST/ACS filters (F814w, F606w and F475w) seen in their respective corresponding colours. The bottom panel shows the combination of three colour images with the same colour-scale. With this, we see that the contribution of a galaxy's light profile to a given band is the weighting of its surface brightness by the integral of its spectra, weighted by the response of the corresponding filter. For a given light source, the weights of its surface brightness contribution to each band form the spectral energy distribution (SED) of the source.

---

[3]and recovered at http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/community/CFHTLS-SG/docs/extra/spec.html

4.2 Multi-Band Spectral Component Analysis Deblending Tool: `MuSCADeT`

In this section we reproduce the pre-print of the publication Joseph et al. (2016) that gives the details of the algorithm we developed to separate sources with different colours. The modelling of colours in multi-band images is explained in section 2. The algorithm was tested on simulated data as well as on a few postage stamps taken from the HFF. These tests revealed our method well suited for the separation of background lensed galaxies from foreground the light intra-cluster light (ICL) of early type, cluster member galaxies. In addition to this, the colour separation was applied to a cluster member spiral galaxy that presents a red bulge with extended blue spiral arms. We show that the spiral arms can be separated from the red component and reconstructed even at locations were the components are completely overlapping.

Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT): Deblending colourful objects

R. Joseph[1], F. Courbin[1], and J.-L. Starck[2]

[1] Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

[2] Laboratoire AIM, CEA/DSM-CNRS-Universite Paris Diderot, Irfu, Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France

ABSTRACT

We introduce a new algorithm for colour separation and deblending of multi-band astronomical images called `MuSCADeT` which is based on Morpho-spectral Component Analysis of multi-band images. The `MuSCADeT` algorithm takes advantage of the sparsity of astronomical objects in morphological dictionaries such as wavelets and their differences in spectral energy distribution (SED) across multi-band observations. This allows us to devise a model independent and automated approach to separate objects with different colours. We show with simulations that we are able to separate highly blended objects and that our algorithm is robust against SED variations of objects across the field of view. To confront our algorithm with real data, we use HST images of the strong lensing galaxy cluster MACS J1149+2223 and we show that `MuSCADeT` performs better than traditional profile-fitting techniques in deblending the foreground lensing galaxies from background lensed galaxies. Although the main driver for our work is the deblending of strong gravitational lenses, our method is fit to be used for any purpose related to deblending of objects in astronomical images. An example of such an application is the separation of the red and blue stellar populations of a spiral galaxy in the galaxy cluster Abell 2744. We provide a python package along with all simulations and routines used in this paper to contribute to reproducible research efforts. Codes can be found at `http://lastro.epfl.ch/page-126973.html`.

Key words. Methods: data analysis – Gravitational lensing: strong – Galaxies: surveys – stellar populations

1. Introduction

Astronomical objects are often seen merged or blended on the plane of the sky. This blending can be apparent, because objects at different distances are seen in projection on the plane of the sky or, real, because different objects at the same distance are physically overlapping.

Whatever the reason for the blending, reliable deblending techniques are mandatory for astrophysical projects to meet their scientific objectives. Among the many possible examples, blends of galaxies of different colours can impact performances of photometric redshift algorithms (e.g. Pérez-González et al. 2010; Bellagamba et al. 2012; Parker et al. 2012; Hsu et al. 2014) and conclusions of stellar populations studies (e.g. Yan et al. 2014). Obviously, blending also affects the determination of morphological properties of astronomical objects, for example the shape measurement of faint galaxies in weak lensing cosmological surveys (e.g. Chang et al. 2013; Arneson 2013). In strong gravitational lensing, deblending of the foreground lensing object from the background lensed sources is essential, for example as shown at galaxy scale by Gavazzi et al. (2007) and at cluster-scale by Massey et al. (2015). This is true for at least two reasons. First, one needs to map the visible mass in the lensing object precisely, either to use it as a prior to guide the lens modelling or to infer the mass-to-light ratio in the lens. Second, the image of lensed source must be isolated in the best possible way. Any faint extended arc-like structure, clump, or star-forming region must be seen precisely with minimum light contamination from the lensing object. Our ability to constrain the mass model is completely driven by the amount of details seen in the lensed source, which represent as many observational constraints.

Many of the current techniques to deblend astronomical objects are limited to analytical modelling of their light distribution either in single band (e.g. `PSFex`, Bertin 2011) or multi bands, sometimes including a simultaneous fit of many overlapping objects (`Megamorph`; Vika et al. 2013, 2015). Alternatively, some methods make use of high resolution images to flag blended objects and then measure them at different wavelengths using images of lower spatial resolution (e.g. Laidler et al. 2007). Popular softwares like `Sextractor` (Bertin & Arnouts 1996) use image segmentation to separate blends, which is a technique that was further improved by Zheng et al. (2015). Other techniques include machine learning, recently used in the area of strong gravitational lensing to subtract the light of bright galaxies and to unveil possible lensed background objects without invoking any analytical representation of the galaxies to subtract. This technique is based on a principal component decomposition of the galaxy images using large samples of single-band imaging data (Joseph et al. 2014). A step forwards is to use multi-band images to separate the objects in the lens plane and source plane, also using the colour information. Recent methods have started to make use of multi-band information and combine source position from high resolution images with profile fitting to deblend lower resolution bands (Merlin et al. 2015). Another example is given in Hocking et al. (2015) where neural networks are used to identify objects with different colours in multi-band images.

With the current burst of wide-field sky surveys (DES, KIDS, HSC, Euclid, LSST and WFIRST), data in many optical and

1

near-IR bands will become available. The present paper describes a technique taking advantage of these multi-band data to address the deblending problem using both colour information and a spatial prior, but not involving any analytical modelling of the light distribution of the lensed and lensing objects. Our work is based on a multi-channel extension of the morphological component analysis (MCA) presented in Starck et al. (2004). We illustrate the performances of the algorithm with numerical experiments and with real HST data of strong gravitational lenses.

This paper is organised as follows: In section 2, we introduce the mathematical model we use to understand and separate objects with different colours. In section 3, we describe the mathematical technique used to solve the problem of colour separation in our approach, that is to say morphological component analysis. In section 4, we detail our implementation of the MuSCADeT algorithm. Section 5 shows the performance of our algorithm on simulations that test realistic problems encountered in deblending. We apply our method to real astronomical images from the Hubble Space Telescope in section 6 with the galaxy clusters MACS J1149+2223 and Abell 2744. We compare our results with current model fitting methods. Section A provides useful information to reproduce our results from the code we made freely available.

2. Deblending and strong lensing

2.1. The source separation problem

We assume the observed data $\{y_i\}_{i=1,...,N_b}$ in the band i can be represented as

$$y_i[k] = \sum_{j=1}^{N_o} s_{i,j} o_j[k] + z_i[k], \qquad (1)$$

where o_j are the different observed sources, $s_{i,j}$ is the contribution of the j-th source in the observation y_i, N_o the number of sources, N_b the number of bands, z_i is an additive Gaussian noise, k is the pixel index ($k = 1...N_p$), and N_p is the number of pixels.

The parameter $s_{*,j}$ corresponds to the spectral energy distribution (SED) of the source o_j. The deblending problem consists in finding the different objects o_j, which is somewhat complicated since their SED are not known and even the number of objects is not known.

However, several galaxies may have similar colour properties and, therefore, share the same SED, so we can simplify Eq. 1 by considering the data containing only N_s groups of sources, such as, for instance early- and late-type galaxies, and we can restrict the deblending problem to only extract these two groups. We note x_j ($j = 1..N_s$) the image which contains the sum of all objects belonging to the group j, i.e. $x_j[k] = \sum_{l=1}^{N_o^{(j)}} o_l[k]$, where $N_o^{(j)}$ is the number of sources in the group j. We can write

$$y_i[k] = \sum_{j=1}^{N_s} a_{i,j} x_j[k] + z_i[k]. \qquad (2)$$

Even if this equation looks very similar to Eq. 1, it is in fact simpler since N_s is smaller than N_o. As a given component x_j contains several astrophysical sources, it also gives us more statistics to derive its SED. This linear mixture model can be recast in the following matrix form

$$Y = AX + Z, \qquad (3)$$
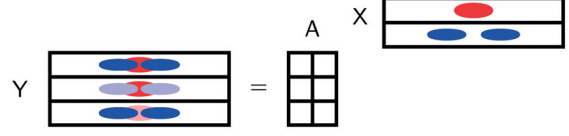
Fig. 1. Illustration of the blind source separation in the case of two sources. To make the figure simple, the images in each band are represented as lines in the Y matrix. Sources are lines in the X matrix. On the sketch we figure a red object in the first source and two blue objects in the second source. Matrix A contains the mixing coefficients that allow various combinations of elements of X to produce Y.

where Y is a $N_b \times N_p$ matrix, A is the SED mixing matrix, and X is the $N_s \times N_p$ matrix, which contains the components x_j.

To sum up, we consider that each band is a weighted mix of n_s colour components. In the statistical literature, each component is called a source (and it should not be mixed with an astrophysical source). This general problem is called a blind source separation problem (BSS), i.e. estimating both A and X knowing only Y. The weight for a given source is the value of the associated SED at the corresponding wavelength. Figure 1 illustrates the BSS in the case of two sources relative to two populations of galaxies (red and blue galaxies in the figure).

2.2. Determination of the mixing matrix A

The A matrix is central to modelling of multi-band data as it describes the contribution of the different sources to the images taken at different wavelengths. In practice, the elements of A are the SEDs of the objects in the sources X. They can be assumed, for example as template spectra for objects of a given type, or they can be measured because spectroscopic data are available for at least some of the objects in the field of view. In most cases, however, the matrix A needs to be estimated solely from the multi-band imaging data, Y. In order to do this, we use a method based on a principal component analysis (PCA; Jolliffe 1986) of the data.

We consider the multi-band data as an ensemble of vectors $\{y_{i=1..N_b}[k]\}$, where the pixel values in band i at the spatial location k are stored in $y_i[k]$, as previously. In other words, $\{y_{i=1..N_b}[k]\}$ is the measured SED at location k.

Following our definition of a source, two pixels belonging to a given source x have proportional SEDs. To build the mixing matrix A, one solution is to preselect obvious objects belonging to a same source and to average their SEDs, thus approximating the mixing coefficients corresponding to their source. A more subtle way to do this is to perform a principal component analysis of all SEDs belonging to bright objects and to look for proportional vectors. The details of this procedure can be summarised as follows:

– We select the brightest objects in all bands and perform the PCA of the SEDs at pixel locations with high signal to noise. In practice, this is done by applying a wavelet filtering of all bands. Also, to save computation time, we rebin images to 64×64 pixels in size.
– We perform a clustering analysis of the first two PCA components: the linearity and orthogonality of the PCA decomposition implies that proportional vectors see their respective PCA coefficients distributed along the same hyperplane

2

in the PCA space. In other words, vectors with proportional SEDs have their first two components, PC1 and PC2, distributed along lines in the PC1-PC2 space, as illustrated in Fig. 2,

- We identify the SEDs that have proportional PC1 and PC2 coefficients and average them. Coefficients that are judged too faint or too ambiguous (i.e. they could be a mix of both sources) are rejected,
- We store the resulting mean SEDs as a column in the A matrix.

This algorithm shows good results in identifying objects with different dominant colours (Fig. 2), but its capabilities in terms of deblending are rather limited when distinct sources spatially overlap. The above PCA analysis is only a spectral analysis. The MCA method proposed in this paper combines the strengths of the morphological analysis and spectral analysis to design a reliable deblending algorithm.

3. Morphological component analysis

The morphological component analysis (MCA) method (Starck et al. 2004; Starck et al. 2010) allows us to separate several components in a given image based on their morphological diversity. Indeed, it was shown that it is possible to disentangle two (or more) signals mixed into one observable, this is only based on the fact that each of those signals can be sparsely represented in their respective data representation domains, called dictionaries, but not in the other's. For instance, one could separate a periodic signal from a Gaussian profile in an image based on Fourier transform (associated with the periodic signal) and the wavelet transform (for the Gaussian profile). The projection of the mixing over the Fourier dictionary shows enhanced contribution from the periodic component whereas wavelet space shows higher coefficients for the Gaussian component.

3.1. Separation using a sparsity prior

MCA is based on the concept of sparse signal representation. A signal is sparse in a dictionary Φ when it can be well represented, such that $x = \Phi\alpha = \sum_i \phi_i \alpha_i$ and only few coefficients α are different from zero. Some dictionaries, such as Fourier or wavelets, have implicit fast transformation and reconstruction operators that allows us to derive the coefficients α from x (and also to derive x from α) efficiently without having the elements of matrix Φ in memory. In inverse problems, a sparse solution is imposed by adding an ℓ_0-norm penalisation term to the data fidelity attachment term. MCA is an iterative algorithm, which separates a single image Y into J components x_j, by solving

$$\min_{x_1,\dots,x_J} \sum_{j=1}^{J} \| \Phi^* x_j \|_0 \quad s.t. \quad \| Y - \sum_{j=1}^{J} x_j \|^2 \leq \sigma, \qquad (4)$$

where $\Phi^* x = \alpha$ and σ is the noise standard deviation of the noise. Full details can be found in Starck et al. (2010). This method has been used to extract filamentary clouds in Herschel data (André et al. 2010) or, more recently, to improve SNIa detection in the SuperNova Legacy Survey data set (Möller et al. 2015).

3.2. Multi-band dictionaries

As we have multi-band data, we need to use morpho-spectral diversity. The dictionary Φ_i related to a given component x_i is

therefore a tensorial product of a spectral dictionary $\mathcal{S}_j$ with the spatial dictionary Ψ_j, i.e. $\Phi_j = \mathcal{S}_j \Psi_j$

3.2.1. Spatial dictionary

In the case of strong lensing, the diversity between the components is mainly related to a different spectral morphology. Therefore we can reasonably use the same spatial dictionary, and in this strong lensing application we use the starlet dictionary (Starck et al. 2007). Starlet transform is an isotropic, undecimated wavelet transform that is computed using consecutive convolutions by a B−spline profile as a scaling function (Starck & Murtagh 2007). The resulting starlet representation is an overcomplete set of coefficients that represent variations in an image at different scales, and is particularly suited to represent astronomical images (Starck & Murtagh 2006).

3.2.2. Morpho-spectral dictionary

We note Ψ the starlet dictionary, we have $x_j = \Psi\alpha_j$ where α_j are the starlet coefficients of the jth source. A good choice for the morpho-spectral dictionary is to take $\Phi_j = a_j\Psi$, where a_j is the jth column of the matrix A. The data attachment term for multichannel data can be written as

$$L = \| Y - \sum_{j=1}^{J} \Phi_j \alpha_j \|^2 = \| Y - \sum_{j=1}^{J} a_j \Psi \alpha_j \|^2$$

$$= \| Y - \sum_{j=1}^{J} a_j x_j \|^2 = \| Y - AX \|^2 . \qquad (5)$$

The multichannel MCA hence consists in changing the data attachment term only, and we need to solve

$$\min_{X} \sum_{j=1}^{J} \| \Psi^* x_j \|_0 \quad s.t. \quad \| Y - AX \|^2 \leq \sigma. \qquad (6)$$

3.2.3. Lagrangian form and positivity

The sparse recovery problem can be formulated under an augmented Lagrangian form

$$\min_{X} \| Y - AX \|^2 + \sum_{j=1}^{J} \lambda_j \| \Psi^* x_j \|_0 \qquad (7)$$

and we can add a positivity constraint to the solution, so we need to solve

$$\min_{X} \| Y - AX \|^2 + \sum_{j=1}^{J} \lambda_j \| \Psi^* x_j \|_0 \quad s.t. \ \forall j, \ x_j \geq 0, \quad (8)$$

where λ_j accounts for the sparsity of each component x_j in its own morpho-spectral dictionary. The next section describes how this equation can be solved.

4. MuSCADeT algorithm

The strong lens sparse deblending iterative algorithm (MuSCADeT) is an extension of the MCA algorithm, consisting in applying at each iteration three main steps:

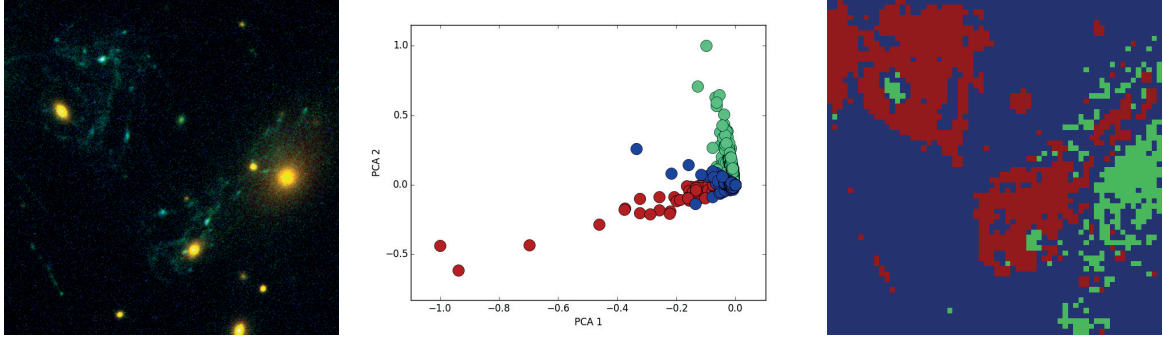1. Perform a gradient step: $U = X^{(n)} + \mu A^t(Y - AX^{(n)})$.

3

Fig. 2. Illustration of the PCA colour selection. *Left:* HST image of the "Refsdal lens" in the galaxy cluster MACS J1149.6+2223. *Middle:* distribution of the first two PCA coefficients. The red and green dots correspond to the coefficients attributed to the first and second sources by `MuSCADeT`, respectively. Blue dots are rejected coefficients. *Right:* corresponding spatial distribution of colours as detected via PCA.

2. Solve for each j: $min_{x_j^{(n+1)}} \| u_j - x_j^{(n+1)} \| + \lambda_j \| \Psi^* x_j^{(n+1)} \|_0$,
   and set to zero negative entries in $x_j^{(n+1)}$.

3. Decrease λ_j.

In this algorithm μ is the gradient step, derived from the 2-norm of matrix A (Higham 1992) such that $\mu = 2/\|A\|_2$.

This algorithm is also directly related to the proximal forward-backward algorithm (Combettes & Wajs 2005). A very nice aspect of this algorithm is that the minimisation involved in the second step do not require any iteration, and is obtained by $x_j^{(n+1)} = \Delta_{\Psi,\lambda_j}(u_j)$, where Δ_{Ψ,λ_j} is the operator which performs the starlet transform, hard thresholds the starlet coefficients, and reconstructs an image from the thresholded coefficients. Here, λ_j is the threshold that allows us to select only coefficients that are significant enough to represent the signal. Thresholds are updated at each iteration as described in the following paragraphs. Full details can be found in Starck et al. (2010). Pseudo-algorithm 1 shows the principle of this iterative scheme. MuSCADeT is an iterative process that alternates between a gradient step (line 6) and a filtering of the components in transformed space through iterative hard thresholding (line 9).

Algorithm 1 MuSCADeT algorithm

1: **procedure** MuSCADeT(Y, K, A, J, N_{iter})
2:     $\tilde{X} \leftarrow 0$
3:     **for** $0 < i \leq N_{iter}$ **do**
4:         $R = \mu(A^T(Y - A\tilde{X}))$
5:         **Update** λ
6:         $\tilde{X} \leftarrow \tilde{X} + R$
7:         $\lambda \leftarrow mi(\frac{\lambda-K}{N_{iter}-i-6}, MOM(X))$
8:         **for** $0 < j \leq J$ **do**
9:             $\tilde{x}_j \leftarrow \Delta_{\Psi\lambda_j}(\tilde{x}_j)$
10:         **end for**
11:     **end for**
12:     **return** $\tilde{X}$
13: **end procedure**

4.1. Thresholding strategy

Thresholding aims at selecting the coefficients in transformed space that allow us to reconstruct the desired signal only. In this case, this means that for a given component, we want to select coefficients above noise level that accounts for this component and not for the others. It is therefore crucial to devise an adequate method to adapt thresholds at each iteration of algorithm 1.

Since each iteration moves our solution for components closer to a good separation, the thresholds have to be decreased to capture fainter and fainter structures. A classical way is to operate a linear decrease for instance, where values for λ_j are linearly sampled between an initial threshold chosen high above noise levels and a sensitivity value K. In general, K is chosen between three and five. The sensitivity value three allows for good completeness of detected coefficients and five ensures a selection that is free from noise-related coefficients. Noise levels are computed using median absolute deviation (Donoho & Johnstone 1994). Although linear or exponential laws are well suited to such problems (Starck et al. 2004), we choose here to rely on a more adaptive strategy based on minimum of maximums (MOM, see Bobin et al. 2007):

At each iteration, we simply estimate the maximum coefficient of each component in its own morpho-spectral dictionary and choose the smallest maxima plus a margin as a threshold. If the result is smaller than the threshold given by a linear decrease, the threshold is updated with the value estimated from MOM as illustrated in line 7 of the `MuSCADeT` algorithm (1). Full details on this thresholding scheme can be found in Starck et al. (2010).

5. Tests on simulations

In this section, we present several tests conducted on simulated multi-band images that emulate realistic problems. We show that `MuSCADeT` is able to separate highly blended sources and is robust to the approximation made that several astronomical sources (e.g. galaxies in a cluster) with similar but, not identical, SEDs can be considered a single source.

5.1. Simulations

In all following simulations we generate objects with Gaussian profiles and realistic SEDs. For the first two simulations, all SEDs were extracted from real HST observations of the galaxy cluster MACS J1149+2223 cluster (Kelly et al. 2015) using the PCA method described in Sect. 2.2. Each simulation comprises seven bands. We also add a white Gaussian noise with standard
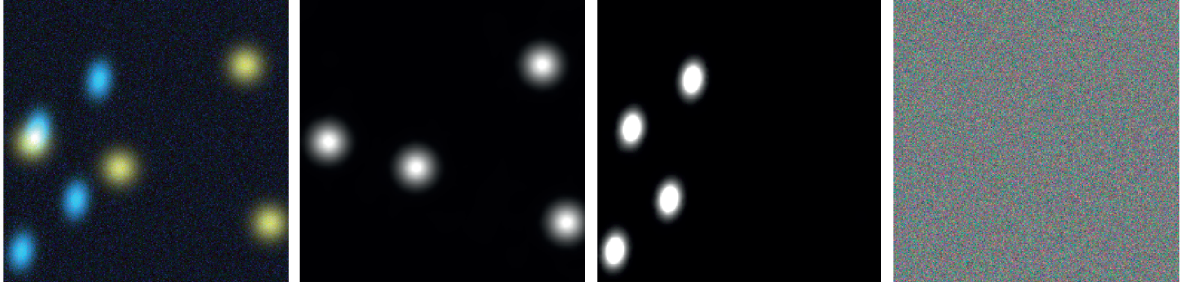
4

Fig. 3. Source separation with `MuSCADeT` in the case of a simple colour separation with SEDs estimated from PCA. *From left to right*: original simulated image of colour sources, first and second components (elements of X in eq. 3) as extracted by `MuSCADeT`, and the residual image after subtracting both components from the original image.
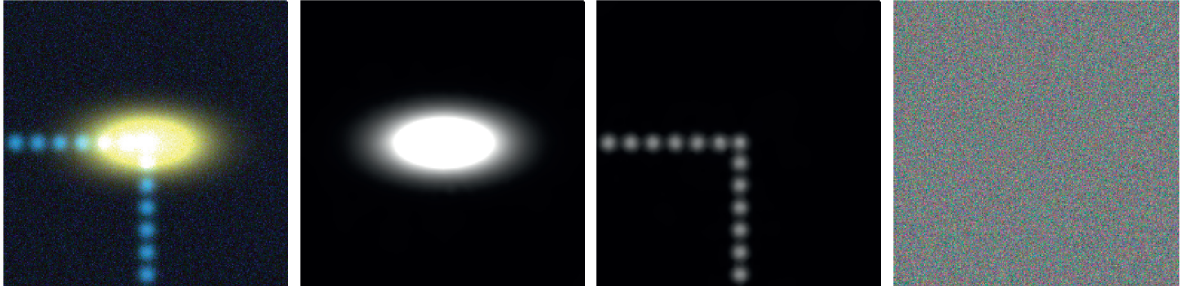

Fig. 4. Separation of blended sources with SEDs estimated from PCA (as in Fig. 3).
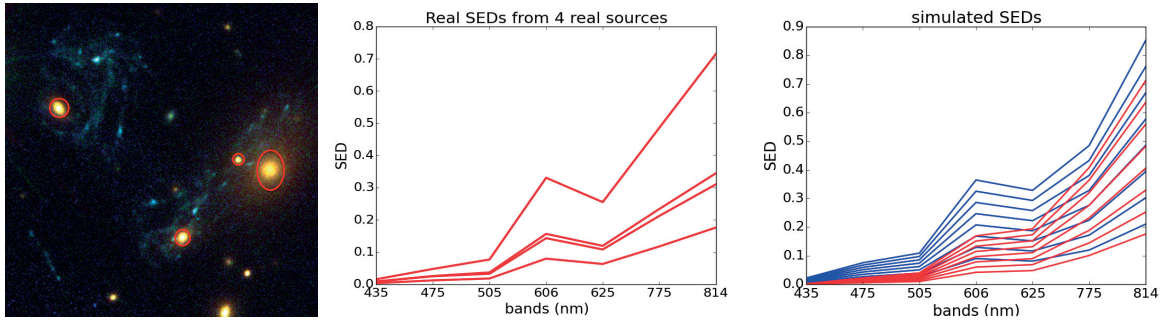

Fig. 5. *Left:* part of an HST image of the galaxy cluster MACS J1149+2223, where the four objects used to extract SEDs are indicated with red contours. *Middle:* extracted SEDs. Each curve corresponds to the SEDs of the galaxies circled in red in the first panel. *Right:* simulated SEDs. The seven red SEDs are used to produce the upper row of galaxies in Fig. 6. The blue SEDs correspond to the lower row of galaxies (see text).

deviation $\sigma = 0.01$. An example of the impact of higher noise levels on our image separation is given in appendix B.

We first apply `MuSCADeT` to data with four red objects with exactly the same SED and four blue objects with exactly the same SED. All objects have elliptical Gaussian profiles. Three of the simulated bands are used to produce the colour image in Fig. 3. Results of the separation after 100 iterations of `MuSCADeT` are shown in the other panels of fig 3. We observe no contamination between sources since no structure contaminates either component. Since the residuals present no structure, we also conclude that each component has been fully reconstructed.

The second simulation tests the deblending capacities of our algorithm in the ideal case where there is no SED variation be-

tween objects of a same component. We generate an extended elliptical Gaussian profile at the centre of the image affected with a red SED. Thirteen blue profiles are evenly spread between the centre and edges of the image such that profiles at the centre are completely blended. Results of the separation are shown in Fig. 4 along with a colour image of our simulated objects. No contamination between sources is visible either and residuals show no structure. We see in particular no difference between blue sources at the centre of the image (the most blended ones) and blue sources on the edges (less blended), meaning that the deblending of each profile is successful whether a profile is highly blended or not. We used 200 iterations of `MuSCADeT` to produce this result.

5
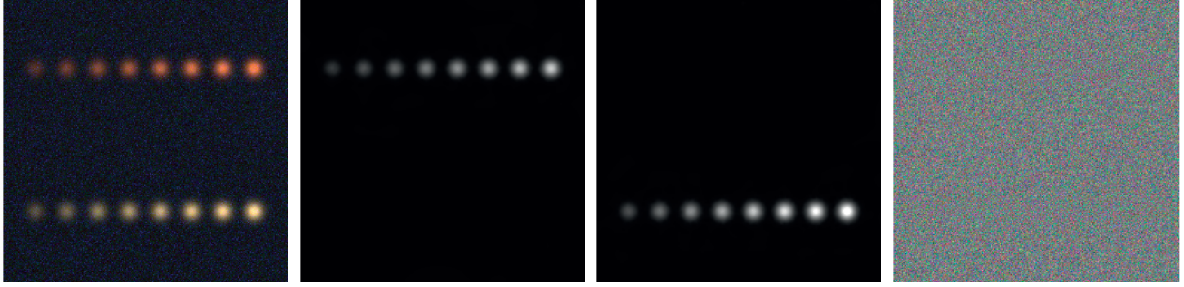
Fig. 6. Results of a separation of sources with poorly known SEDs (as in Fig. 3).

The last simulation tests the robustness of our algorithm to SED variation across objects from a same component. To account for realistic SED variation, we extracted the SEDs from four red galaxies in the MACS J1149+2223 cluster (see 5), which appear to have the same colour by integrating the flux in each galaxy profile. The resulting SEDs can be seen in the middle panel of fig 5. We recorded the slopes of the SEDs and generated a set of eight slopes linearly spread between the maximum and minimum slope estimated. These slopes are then applied to SEDs extracted from cluster MACS J1149+2223 via PCA (see Fig. 5). This way, we have two sets of SEDs that account for red and blue sources and that mimic a range of variations as observed in real images. Sixteen Gaussian profiles (eight red, eight blue) are then generated and each of them is associated with one of the previously generated SEDs. The left panel of Fig. 6 shows three bands of the simulated images as RGB images. Figure 6 shows a colour image of our simulated images and the result of a separation by MuSCADeT. We see again that no structure appears in the residuals and no contamination is found between components. However, the great similarity between SEDs from different components forced us to increase the number of iterations of MuSCADeT to 5000 in order to obtain such results, thus increasing computation time. In each, we used the full algorithm described in this paper including an automated estimation of SEDs through PCA.

6. Application to real data

6.1. Lens-source separation on MACS J1149+2223

We now apply MuSCADeT to real multi-band data and we compare the performances to traditional model fitting.

We use the deep HST data set of the galaxy cluster MACS J1149+2223 ("Refsdal SN"; Kelly et al. 2015) to carry out this experiment, we show bright cluster members producing strongly lensed images of a distant spiral galaxies with clumpy structures. Our goal is to separate the data into two sources containing the foreground lens galaxies and background lensed object(s). As the cluster contains many member galaxies and as the background galaxy has complex structure, the deblending task is challenging, making this data set a good test for our method.

MACS J1149+2223 has been observed with the ACS in seven bands: F435W, F475W, F555W, F606W, F625W, F775W and F814W (proposal ID: 12068, principal investigator: M. Postman), providing a good spectral coverage for MuSCADeT to work. The data are publicly available from the STScI website[1]

and drizzled so that the combined frames in each band have the same orientation and pixel scale.

We estimate both SEDs in the mixing matrix using our PCA technique. The result after 2000 iterations of MuSCADeT is shown in Fig. 7, where the two separated sources are shown in the middle panels. In the two lower panels of the figure, we also show the result of the subtraction of source 1 and source 2 from the original data. The colour scale in these images is the same as in the original data. The overall residual image, i.e. with both sources subtracted from the data, is shown with $\pm 5\sigma$ cut levels.

The source separation works very well, with the exception of a few objects with "green" SEDs. One such object is visible in the centre of the image, resulting in a signal in both red and blue sources. This is an intrinsic limitation to our algorithm that separates objects using a limited number of sources, each one with its own SED. Although the SEDs do not need to be known perfectly (as shown in our tests with simulated images), objects with SEDs falling "in between" the SEDs allocated to each source, may lead to inaccurate separation. A possible mitigation strategy is to add extra sources to the decomposition with for example a blue, a green, and a red SED. Deciding whether to do this or not depends on the exact scientific application.

In our example, we do not take the PSF convolution into account. A variation of the PSF with wavelength can introduce artefacts in the separation, especially for objects whose angular size is comparable with the size of the PSF. The central parts of galaxies show such structures and, indeed, in the present data small structures are seen in the residual image at the location of the foreground galaxies. Introducing the PSF convolution can be done in principle, and it might be needed at least for some applications, but at a cost of increased computation time and complexity in the minimisation process.

6.1.1. Comparison with profile fitting

A popular way to carry out source separation in galaxy clusters is to fit two-dimensional profiles to the cluster members and to subtract from the data. We apply such a procedure to the HST data of the MACS J1149+2223 cluster and compare with MuSCADeT. In this exercise, we identified nine red objects with elliptical profiles that we fit using the galfit (Peng et al. 2002) software. Each red object is fitted with either one or two Sersic profiles depending on its morphology. Fitting is performed in each of the three bands used to build the colour image (F475W, F606W, and F814W) using the same initial conditions. Since no PSF was used in applying MuSCADeT to this cluster, we used a Dirac function as an input PSF for galfit. The result is presented in the
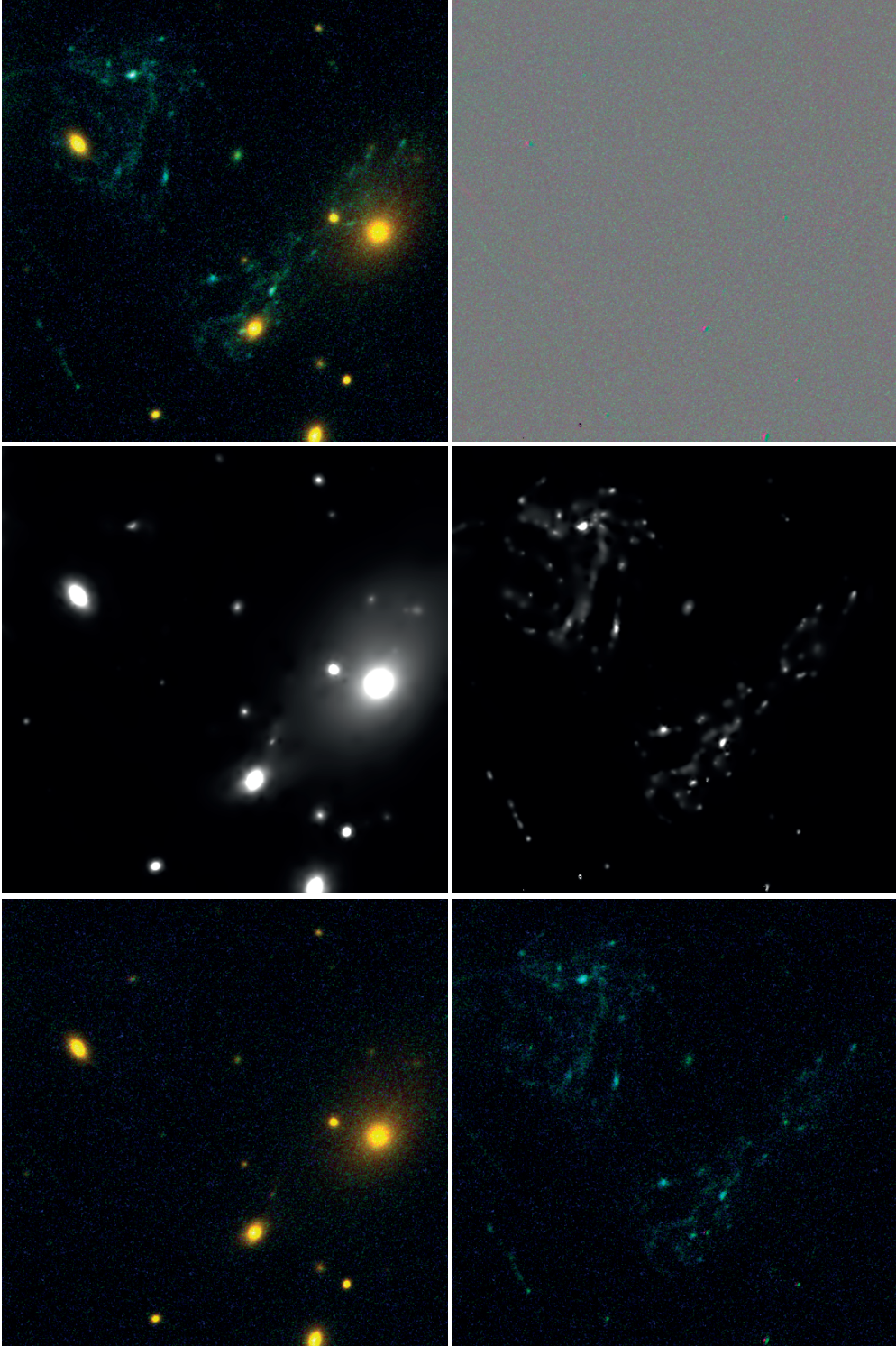
---

[1] https://archive.stsci.edu/

6

Fig. 7. Application of MuSCADeT to the MACSJ1149+2223 cluster. *Top left:* colour image generated using the F475W, F606W, and F814W bands. Our goal is to separate the foreground cluster galaxies (red) from the background lensed galaxy (blue). *Middle:* the two sources extracted with 2000 iterations of MuSCADeT. *Bottom left:* original image minus the blue source found by MuSCADeT. *Bottom right:* same as the bottom left, with the red source subtracted. *Top right:* residual image obtained after subtraction of the two estimated sources from the original data.

7
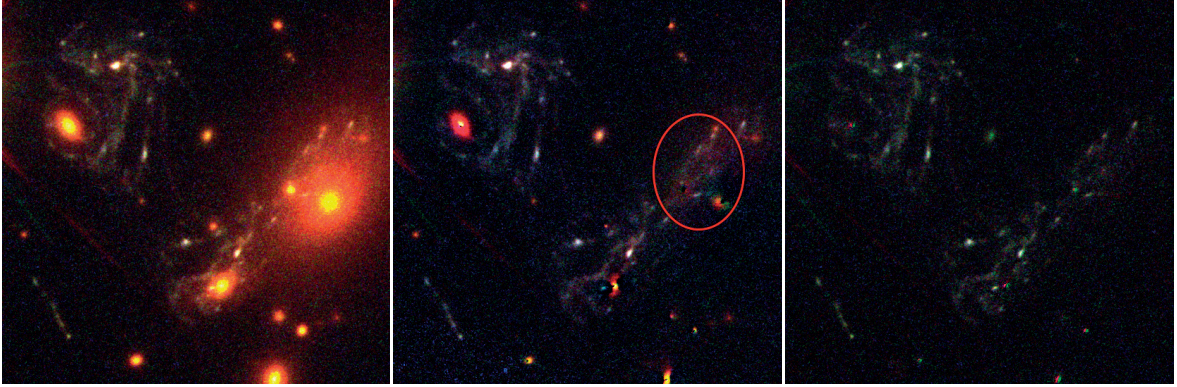
Fig. 8. Comparison between the `MuSCADeT` and that `galfit` separations. The original colour image of MACS J1149+2223 is shown on the left, followed by the `galfit` subtraction of the galaxy members and the subtraction using `MuSCADeT`. The colour cuts are exactly the same in all three images. The red ellipse in the middle panel indicates an area where the `galfit` fit leads to under-subtraction of a galaxy halo.

middle panel of Fig. 8 along with the `MuSCADeT` result. We note that:

- Profile fitting leaves significant artefacts in the central parts of the fitted galaxies even when a double Sersic model is used.
- Profile fitting does not model the extended halos of galaxies well, as shown in the circled region in middle panel of Fig. 8. This is both because extended halos are not simple analytical profiles and because the many background lensed structures influence the fit. In principle, these structures can be masked but 1- designing the mask can be difficult and is time consuming, 2- in some parts of the image there is no clean area with no contamination by the background object. The mask would take the majority of the data.
- The top left lensed image appears to be more blue after `galfit`'s run than with `MuSCADeT`. This could mean that `MuSCADeT` performs badly at extracting extended halo. However, it would imply that too much signal from the red source has been attributed to the blue source. Therefore, when subtracting the blue source from the original images one would see holes in the extended profiles of the red galaxies that are not observed here (see lower left panel of Fig. 7). This implies that `galfit` is overfitting the extend halo to compensate for blue structures, as pointed out in the previous note, thus removing part of the flux from blue sources.
- The human time involved in profile fitting can be a limiting factor for large data sets. The user has to decide where to put a galaxy and to find a strategy to estimate the initial guesses for the many parameters involved in the fit. `MuSCADeT` is fully automated procedure with only one parameter to be chosen, i.e. the sensitivity value K involved in the threshoding scheme.

6.2. Bulge-disk separation on spiral galaxy in Abell 2744

Another possible application of our algorithm is the separation of coloured disk and bulge components in a spiral galaxy. To illustrate this, we use a spiral galaxy in the galaxy cluster Abell 2744, which was imaged with the HST as part of the Hubble Frontier Fields programme.

Abell 2744 was observed with the ACS in three bands: F435W, F606W, and F814W, resulting in the colour image in Fig. 9(Proposal ID: 11689, principal investigator: R. Dupke). The data are publicly available from the STScI website[2] and drizzled so that the combined frames in each band have the same orientation and pixel scale.

Spiral galaxies represent an obvious test bench for our algorithm as they are composed of a red bulge of older stars and of a blue disk dominated by young stars. The ability to separate both colour components in an unbiased way allows us to trace stellar populations in galaxies and/or highlight morphological relations between bulge and disk shapes. Numerous examples can be found in the literature of studies where bulge and disk components are separated morphologically, using profile-fitting techniques featuring Sersic, exponential, or DeVaucouleurs functions in either single or multiple bands (e.g. Vika et al. 2014; Maltby et al. 2012; Pastrav et al. 2013; Kendall et al. 2011).

With `MuSCADeT`, we are able to capture the morphology of the young and old stellar populations of the spiral galaxy in Abell 2744 under the single assumption that they do not display the same colour. Our decomposition in Fig. 9 shows an elliptical red bulge with extended and smooth features along the spiral arms, which elliptical profile-fitting methods would fail to model. The separation of the blended blue and red components in Fig. 9 is overall excellent except for slight cross-contamination of the two colour channels due to features smaller than the PSF size. Adding the PSF in our separation technique is under development. However, even without this refinement, the decomposition presented in our example would be impossible to achieve with profile-fitting techniques given the morphological complexity of the galaxy.

7. Conclusion

We have developed a new model-independent tool to deblend the images of astronomical objects based on their colour contrast. This is made possible via modern techniques for image analysis based on sparsity of objects in an appropriate dictionary.

More specifically, we created a morpho-spectral dictionary combining Starlets and SEDs to separate images of galaxies with

---

[2] https://archive.stsci.edu/pub/hlsp/frontier/

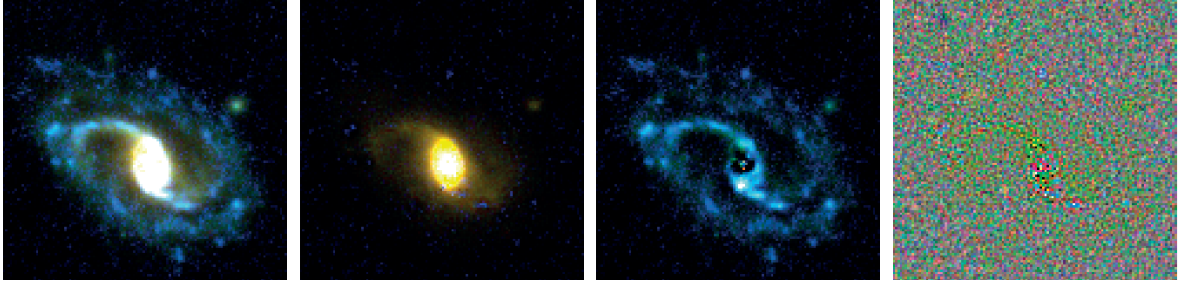8

R. Joseph et al.: MCA-based lens-source separator

Fig. 9. Separation of the red and blue stellar populations in a spiral galaxy. From left to right we show the original HST colour image of a spiral galaxy in the galaxy cluster Abell 2744, the same colour image after subtraction of the blue component estimated with `MuSCADeT`, the colour image after subtraction of the red component estimated with *MuSCADeT*, and the residual image after subtraction of both components. The three first images have the same colour cuts. The residuals are shown with cut levels set to five times the noise level in each colour channel.

different colours. We show from simulated data that our algorithm is robust against very strong blending as well as against SED variations among objects belonging to the same colour component (source) of our model. SED variations across the field result in an increase of the computing time by one order of magnitude with respect to the ideal case where all objects in a given source have the same SED. This does not hamper a reliable source separation, however.

The method is successfully applied to the deep HST images of the galaxy cluster MACS J1149+2223, in which we separate the foreground red galaxies from background blue clumpy lensed galaxies, and to Abell 2744, where we separate the red and blue stellar populations of a spiral galaxy. This is done in an automated way and with better efficiency that with standard profile fitting. All codes used to produce the results presented here are made freely available for the sake of reproducible research and to make our tool usable for the community.

Future developments of our method include accounting for the PSF in each band and including explicit SED variations across the field of view. These SED variations multiply the complexity of the problem by $N_p \times N_s \times N_b$, but the effect of the increased complexity can likely be minimised by using sparse priors and physical constrains on the SED profiles. Also, the increased computation time resulting from the extra complexity should be partially compensated by a reduced number of iterations.

The deblending method described here was devised specifically to address the problem of object deblending in the case of strong lensing systems. Deblending is essential in this case to see small and faint details in the lensed structures, which are free of contamination by the bright foreground. The range of applications of the method is nevertheless much broader. Among the numerous possible applications are the identification of starforming regions in galaxies, model-independent bulge-disk decompositions of galaxies, or even the improvement of photometric redshift in large sky surveys where blending can be a serious issue given the depth of the data.

References

André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Arneson, R. A. 2013, PhD thesis, University of California, Irvine
Bellagamba, F., Meneghetti, M., Moscardini, L., & Bolzonella, M. 2012, MNRAS, 422, 553
Bertin, E. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bobin, J., Starck, J.-L., Fadili, J., Moudden, Y., & Donoho, D. 2007, Image Processing, IEEE Transactions on, 16, 2675
Chang, C., Jarvis, M., Jain, B., et al. 2013, MNRAS, 434, 2121
Combettes, P. L. & Wajs, V. R. 2005, Multiscale Modeling & Simulation, 4, 1168
Donoho, D. L. & Johnstone, J. M. 1994, Biometrika, 81, 425
Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, ApJ, 667, 176
Higham, N. J. 1992, Numer. Math, 62, 511
Hocking, A., Geach, J. E., Davey, N., & Sun, Y. 2015
Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, ApJ, 796, 60
Jolliffe, I. T. 1986, Principal Component Analysis (Berlin; New York: Springer-Verlag)
Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, A&A, 566, A63
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123
Kendall, S., Kennicutt, R. C., & Clarke, C. 2011, MNRAS, 414, 538
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Maltby, D. T., Hoyos, C., Gray, M. E., Aragón-Salamanca, A., & Wolf, C. 2012, MNRAS, 420, 2475
Massey, R., Williams, L., Smit, R., et al. 2015, MNRAS, 449, 3393
Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, A&A, 582, A15
Möller, A., Ruhlmann-Kleider, V., Lanusse, F., et al. 2015, J. Cosmology Astropart. Phys., 4, 41
Parker, A., Bard, D., & Bragg, A. 2012, in APS Ohio Sections Spring Meeting Abstracts, C1002
Pastrav, B. A., Popescu, C. C., Tuffs, R. J., & Sansom, A. E. 2013, A&A, 553, A80
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Pérez-González, P. G., Egami, E., Rex, M., et al. 2010, A&A, 518, L15
Starck, J., Elad, M., & Donoho, D. 2004, Advances in Imaging and Electron Physics Series
Starck, J.-L., Fadili, J., & Murtagh, F. 2007, Image Processing, IEEE Transactions on, 16, 297
Starck, J.-L. & Murtagh, F. 2006, Astronomical Image and Data Analysis
Starck, J.-L. & Murtagh, F. 2007, Astronomical image and data analysis (Springer Science & Business Media)
Starck, J.-L., Murtagh, F., & Fadili, M. 2010, Sparse Image and Signal Processing (Cambridge University Press)
Vika, M., Bamford, S. P., Häußler, B., & Rojas, A. L. 2014, MNRAS, 444, 3603
Vika, M., Bamford, S. P., Häußler, B., et al. 2013, MNRAS, 435, 623
Vika, M., Vulcani, B., Bamford, S. P., Häußler, B., & Rojas, A. L. 2015, A&A, 577, A97
Yan, H., Stefanon, M., Ma, Z., et al. 2014, ApJS, 213, 2

9

Zheng, C., Pulido, J., Thorman, P., & Hamann, B. 2015, MNRAS, 451, 4445

Appendix A: Reproducible research

In the spirit of carrying out reproducible research, we make pubic all codes and resulting products describes in this paper. Table A.1 lists all products that will be made available along with this paper. The `MuSCADeT` itself is made available as well as input files and routines needed for all benchmark tests and for the application to real data. The routines provide simple examples of how to execute the `MuSCADeT` algorithm. We encourage potential users to modify them as they wish for their own scientific applications.

Appendix B: Deblending noisy data with known SED

We show here that the `MuSCADeT` algorithm is capable of separating blended sources even in the case of high noise levels. We generate simulations of blended objects as in Fig. 4, but with a noise level ten times larger. The high noise levels along with the strong blending make it hard for the PCA estimator to estimate a good SED for the separation.

Although the PCA technique fails in such conditions, the main feature of our algorithm, which is the morphological component analysis-based inversion, still manages to estimate good sources. The price to pay in this case is that the SEDs must be known. Figure B.1 shows the result of a separation performed by `MuSCADeT` on very noisy data using known SEDs, showing that our algorithm is still able to separate sources. Our PCA SED estimator might replaced in the near future to cope with noisy data. For the present example, we decided to show the residuals after separation of both colour components (middle panels of Fig. B.1) to show that we estimate the sources down to the noise level.

10

R. Joseph et al.: MCA-based lens-source separator

Product name	Type	Description
Software products:		
MuSCADeT	python package	includes MuSCADeT implementation and visualisation tools
Routines:		
Example_simple.py	code (python)	routines to reproduce Fig. 3.
Example_big.py	code (python)	routines to reproduce Fig. 4.
Example_real.py	code (python)	routines to reproduce Fig. 6.
Example_refsdal.py	code (python)	routines to reproduce Fig. 7.
Example_2744.py	code (python)	routines to reproduce Fig. 9.
Example_SNR.py	code (python)	routines to reproduce Fig. B.1.
Example_nottoosimple.py	code (python)	An other example of a MuSCADeT run on simulations.
Simulations:		
Cube.fits	fits data cube	cube with all simulated images for each benchmark
Simu_A.fits	fits table	table with the simulated spectra used in our simulations

Table A.1. List of products made available in this paper in the spirit of reproducible research. All above material is available here: http://lastro.epfl.ch/page-126973.html.
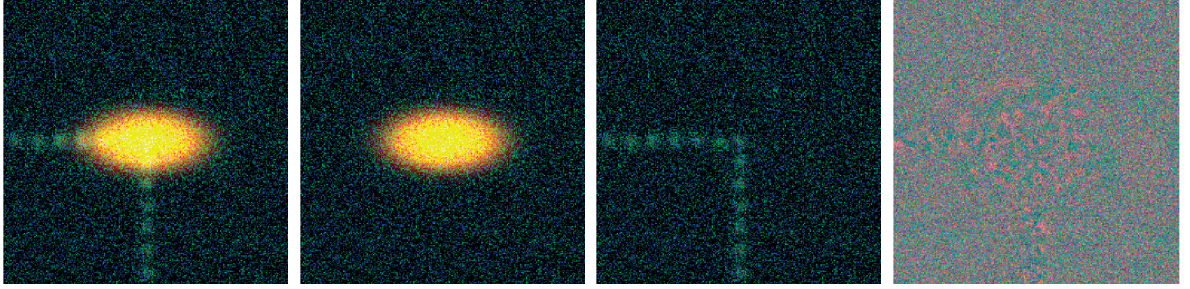
Fig. B.1. Separation of blended sources at low S/N. From left to right are shown the original simulated images, the original image after subtraction of the blue component as estimated from MuSCADeT, the original image after subtraction of the red component and the residual image after subtraction of both components.

11
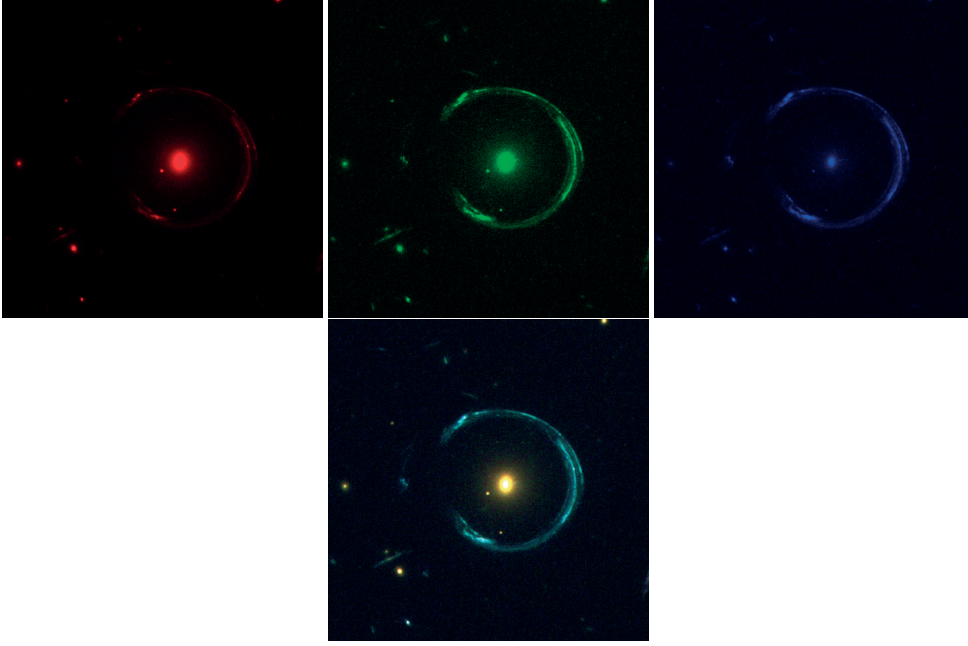
Figure 4.3 – The generation of a colour image from the superimposition of three colour filter images.

We attract the reader's attention on the fact that this publication and the baseline `MuSCADeT` algorithm neglects the convolution of the images by the PSF of the telescope. Because the size of the telescope's diffraction figure depends on the frequency we should correct for the PSF variation across bands. In our applications to HST the small size of the PSF in the optical domain lead us to make the approximation that the PSF was constant in order to decrease the computation time and to avoid the steps of estimating the PSF for each HST band.

4.3   Application to Cluster Lensing

In cluster lensing large fields of views are occupied by bright red cluster-member galaxies that act as lenses for background, predominantly blue sources. Some of these background sources end up to extremely high redshift galaxies that are magnified by the cluster and that would otherwise not be observable (Oesch et al., 2016). In order to understand how these background sources are magnified, it is paramount that we be able to build accurate models of the mass distribution in these clusters. The mass distribution of clusters is also a excellent probe for the nature of dark matter. As we have seen in chapter 1, merging clusters allow us to put constraints on the dark matter cross section (Harvey et al., 2015). Some extremely massive clusters can help us challenge or confirm our current models for dark matter by showing possible discrepancies between the number of structures found in simulations and observational data.

In order to establish these lens models, it is necessary to identified multiply imaged lensed

sources that will inform us about the deflection angles met by these images. While multiply imaged sources are confirmed by spectral identification of common features, the first method for the selection of background sources is the visual inspection and pre-selection of lensed objects. Efforts are currently being made to create automated methods for multiply-imaged source identification (Mahler et al., 2015), that rely on the matching of the SEDs of image pairs. Nevertheless, these two approaches rely on the careful subtraction of the foreground ICL. In visual inspection, the subtraction of the foreground red objects helps identifying potentially lensed sources hidden in the glare of the bright cluster member. For the SED matching technique, the ICL constitutes a major source of pollution for the SEDs of background objects.

So far, several techniques exist to remove the foreground component. A common approach consists in fitting smooth elliptical profiles to the galaxy cluster members surface brightnesses as recently done in Merlin et al. (2016). In Mahler et al. (2018), the authors remove the median value in running windows of 1.3" on-a-side. This is the method used in the automated SED matching technique. Livermore et al. (2017) identifies cluster members and subtract their profile at different frequencies in the starlet domain. This approach actually supports ours in the sense that the authors of Livermore et al. (2017) use the fact that background sources and ICL are at different frequencies, at different positions. In `MuSCADeT` this property of the signals complements the criteria provided by colour to separate objects.

In the following sections, we show several applications of `MuSCAeT` to various HST datasets.

4.3.1 Revealing hidden lensed sources

In this first application, we used `MuSCADeT` in a study of the dark matter distribution of cluster Abell 3827. In a previous publication (Taylor et al., 2017), for which the pre-print is reproduced in appendix C, we used `MuSCADeT` to remove the ICL of cluster Abell 3827 and identify potential lensed sources at the background. This cluster is composed of four massive, bright elliptical galaxies and a background extended blue source appears imaged four times around the cluster. The subtraction revealed two low SNR objects behind the center of the cluster and identified as part of demagnified counter-images of the lensed source. The detection of these objects was supporting two observations made with the multi-unit spectroscopic explorer (MUSE) instrument at the VLT. In a previous publication (Massey et al., 2015, see figure 3), the subtraction of the ICL by way of an elliptical light profile fitting had revealed no demagnified central image of the source, leading to the central image detected as a low SNR source by a first MUSE observation, to be discarded. With MUSE and the ICL subtraction on HST data agreeing to the presence of the central image, this detection was included in the lens modelling of the cluster, thus improving the constraints.

In the previous publications Massey et al. (2015) and Taylor et al. (2017) had concluded to an offset between the centres of the dark matter halos and the center of the light profile of their corresponding galaxies as well as a low significance skewness of the dark matter profiles, which would have been indicative of self-interacting dark matter. In the latter publication

presented here, where ALMA (Atacama large millimeter/submilimeter array) data helped identifying multiple images, we found no evidence of such offset and skewness.

In these two publications (Taylor et al., 2017; Massey et al., 2018), my contribution was restricted to the application of `MuSCADeT` to the HST images. In Taylor et al. (2017), `MuSCADeT` was applied on the postage stamp containing the cluster, and the SEDs were taken from the stamp directly. Due to the brightness of the cluster member galaxies and to two Milky-way stars being in the field of view, the performance of `MuSCADeT` was not optimal and part of the light from the ICL remained as a contaminant in the resulting image (Taylor et al., 2017, figure 1). In the second publication (Massey et al., 2018), the SEDs were optimised to represent the red component by running the PCA estimator of the SED on regions presenting isolated red and blue components. While the first publication is somewhat more relevant to my work, due to this being the first published application of `MuSCADeT`, I incorporate the second publication to the core of the thesis due to this being a non-detection paper. While the non detection of a feature (in this case, offset and skewness), has a lesser and less spectacular impact, in the sense that it did not challenge our current cosmological models, it is nonetheless worth reporting. This constitutes a way of encouraging and supporting the reporting of non-detections, that still contribute to our state of knowledge and report the production of relevant experiments.

Dark matter dynamics in Abell 3827:
new data consistent with standard Cold Dark Matter

Richard Massey[1,2]⋆, David Harvey[3], Jori Liesenborgs[4], Johan Richard[5],
Stuart Stach[2], Mark Swinbank[2], Peter Taylor[6], Liliya Williams[7],
Douglas Clowe[8], Frédéric Courbin[3], Alastair Edge[2], Holger Israel[1,9],
Mathilde Jauzac[1,2,10], Rémy Joseph[3], Eric Jullo[11], Thomas D. Kitching[4],
Adrienne Leonard[12], Julian Merten[13,14], Daisuke Nagai[15], James Nightingale[2],
Andrew Robertson[1], Luis Javier Romualdez[16,17], Prasenjit Saha[18], Renske Smit[19],
Sut-Ieng Tam[2] & Eric Tittley[20]

[1] Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK
[2] Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK
[3] École Polytechnique Fédérale de Lausanne, 51 Chemin des Maillettes, Observatoire de Sauverny, Versoix, CH-1290, Switzerland
[4] Expertisecentrum voor Digitale Media, Universiteit Hasselt, Wetenschapspark 2, B-3590, Diepenbeek, Belgium
[5] Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France
[6] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
[7] School of Physics & Astronomy, University of Minnesota, 116 Church Street SE, Minneapolis, MN 55455, USA
[8] Department of Physics and Astronomy, Ohio University, 251B Clippinger Labs, Athens, OH 45701, USA
[9] Fakultät für Physik, LMU München, Scheinerstr. 1, 81679 München, Germany
[10] Astrophysics and Cosmology Research Unit, University of KwaZulu-Natal, Durban 4041, South Africa
[11] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille), UMR 7326, 13388, Marseille, France
[12] FlightGlobal, RBI UK, World Business Center 2, Newall Road, Heathrow Airport, TW6 2SF UK
[13] Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
[14] INAF, Osservatorio Astronomico di Bologna, via Pietro Gobetti 93/3, 40129 Bologna, Italy
[15] Department of Physics, Yale University, New Haven, CT 06520, USA
[16] Institute for Aerospace Studies, University of Toronto, Canada
[17] Centre for Advanced Instrumentation, Durham University, South Road, Durham DH1 3LE, UK
[18] Physik-Institut, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
[19] Kavli Institute of Cosmology, Cambridge University, Madingley Road, Cambridge CB3 0HA, UK
[20] Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

ABSTRACT

We present integral field spectroscopy of galaxy cluster Abell 3827, using ALMA and VLT/MUSE. It reveals an unusual configuration of strong gravitational lensing in the cluster core, with at least seven lensed images of a single background spiral galaxy. Lens modelling based on HST imaging had suggested that the dark matter associated with one of the cluster's central galaxies may be offset. The new spectroscopic data enable better subtraction of foreground light, and better identification of multiple background images. The inferred distribution of dark matter is consistent with being centered on the galaxies, as expected by ΛCDM. Each galaxy's dark matter also appears to be symmetric. Whilst we do not find an offset between mass and light (suggestive of self-interacting dark matter) as previously reported, the numerical simulations that have been performed to calibrate Abell 3827 indicate that offsets and asymmetry are still worth looking for in collisions with particular geometries. Meanwhile, ALMA proves exceptionally useful for strong lens image identifications.

Key words: dark matter — astroparticle physics — galaxies: clusters: individual: Abell 3827 — gravitational lensing: strong

⋆ e-mail: r.j.massey@durham.ac.uk

2   R. Massey et al.

1   INTRODUCTION

Determining the properties of dark matter has become a priority of astrophysics and particle physics. In the standard ΛCDM cosmological model, dark matter has significant interactions with standard model particles through only the gravitational force (e.g. Massey, Kitching & Richard 2010; Kneib & Natarajan 2011). It therefore neither emits nor absorbs light, and appears invisible. Nonetheless, over the course of cosmic history, dark matter's gravitational attraction assembled the Universe's large-scale structure, and governed the evolution of galaxies. Dark matter has pulled together both ordinary and dark material into a series of collisions – then eventual mergers – between ever-larger structures (Vogelsberger et al. 2014; Schaye et al. 2015).

Several particle physics theories of dark matter predict additional forces *between* dark matter particles, hidden entirely within the dark sector (Peter et al. 2012; Cyr-Racine & Sigurdson 2013). The most direct way to measure these hypothesised forces is to observe the trajectory of dark matter during collisions with other dark matter. In effect, astrophysical mergers can be treated as enormous particle colliders (Clowe et al. 2004, 2006; Bradač et al. 2008; Merten et al. 2011; Clowe et al. 2012; Dawson et al. 2012; Gastaldello et al. 2014; Chon & Böhringer 2015; Ng et al. 2015; Harvey et al. 2015; Jee et al. 2016; Golovich et al. 2016, 2017; Kim, Peter & Wittman 2017; Monteiro-Oliveira et al. 2017). In simulated mergers assuming ΛCDM, the (non-interacting) dark matter remains tightly bound near stars (Schaller et al. 2015). If dark-sector forces exist, simulations of mergers predict dark matter to temporarily lag behind stars, which serve as collisionless test particles (Randall et al. 2008; Massey, Kitching & Nagai 2011; Dawson et al. 2013; Harvey et al. 2014; Kahlhoefer et al. 2014; Robertson et al. 2017a,b). In some simulations, the distribution of dark matter is also stretched into asymmetric tails (Kahlhoefer et al. 2014).

Two properties of galaxy cluster Abell 3827 (RA=22h 01′ 49″.1, Dec=−59° 57′ 15″, z=0.099, De Plaa et al. 2007; Carrasco et al. 2010; Williams & Saha 2011), make it uniquely interesting for studies of dark matter dynamics. First, the cluster core contains four similarly-bright galaxies. They must be undergoing a simultaneous, high speed merger, because this amount of substructure is unique: most clusters have reached a steady state with only a single Brightest Central Galaxy. Second, directly behind the cluster core lies a spiral galaxy ($z = 1.24145 \pm 0.00002$; Massey et al. 2015) that is rich in morphological structure. The background spiral galaxy has been gravitationally lensed by the cluster, and its multiple images wrap around all four of the central galaxies. These images can be used to infer the spatial distribution of (dark plus stellar) mass in the cluster and its galaxies.

One of Abell 3827's central galaxies lies very close to a set of gravitationally lensed images, so the distribution of its mass is particularly well constrained. Analysis of the gravitational lensing in optical imaging suggested that this galaxy's dark matter is offset by $1.62^{+0.47}_{-0.49}$ kpc from its stars (Massey et al. 2015), and possibly asymmetric (Taylor et al. 2017). This could have been caused by a dark sector force with interaction cross-section $\sigma/m \gtrsim 1\mathrm{cm}^2/\mathrm{g}$, where m is the (unknown) mass of the dark matter particle (Kahlhoefer et al. 2015). The most difficult part of this analysis was the iden-

tification of features in the faint, background spiral, right next to a very bright foreground galaxy (see Appendix B in Massey et al. 2015).

In this paper, we present new Integral Field Unit (*i.e.* 2D) spectroscopy of Abell 3827 at near-IR and millimetre wavelengths: where the foreground cluster is faint, but the background spiral galaxy remains bright. We describe the new data in section 2. We describe our analysis techniques in section 3, and reconstruct the spatial distribution of dark matter in section 4. We discuss the consequences of our results in section 5. Throughout this paper, we adopt a cosmological model with $\Omega_{\mathrm{M}} = 0.3$, $\Omega_{\Lambda} = 0.7$ and $H_0 = 70\,\mathrm{km/s/Mpc}$, in which $1''$ corresponds to $1.828\,\mathrm{kpc}$ at the redshift of the cluster. Adjusting this cosmological model perturbs the inferred physical distances, and the absolute normalisation of inferred masses.

2   DATA

2.1   Pre-existing imaging

Broad-band imaging of Abell 3827 has been obtained by the Gemini telescope at optical wavelengths (Carrasco et al. 2010) and by the *Hubble Space Telescope* (*HST*; programme GO-12817) in the F336W (UV), F606W and F814W (optical) and F160W (IR) bands (Massey et al. 2015).

This revealed four similarly-bright elliptical galaxies (N1–N4) within 10 kpc radius, and a background lensed spiral galaxy (with a red bulge and blue spiral arms), whose multiple images are threaded throughout the cluster core. In this paper, we exclusively use the HST imaging. As described in Taylor et al. (2017), we reveal the background lensed galaxy by fitting and subtracting foreground emission from the five brightest cluster galaxies and two Milky Way stars using the MuSCADeT method (Joseph, Courbin & Starck 2016) (figure 1).

2.2   ALMA integral field spectroscopy

In October 2016, we obtained a 5.2 hour observation of Abell 3827 with the Atacama Large Millimetre Array (ALMA; programme 2016.1.01201.S). The band 3 data sample frequencies 89.9-93.8 GHz and 101.8-105.6 GHz with spectral resolution 15.6 MHz (47.8 km/s). Observations were conducted with 44 12 m antennae in the C40-6 configuration. Flux and bandpass calibration were obtained from J2056−4714, and the phase calibrator was J2208−6325.

Data were reduced using CASA software v4.7.2 (McMullin et al. 2007). Spectral data cubes were created using the CLEAN algorithm, with channel averaging and natural weighting to maximise sensitivity. This yielded a synthesised beam of $\sim0.48'' \times 0.39''$, and a 1σ noise level of $0.08\,\mathrm{mJy/beam}$ for each 31.3 MHz channel. In addition, to minimise potential extended flux being resolved out, we created a second spectral cube with a (u, v) taper applied that yielded a synthesised beam of $\sim 0.87'' \times 0.82''$ and 1σ noise level of $0.15\,\mathrm{mJy/beam}$.

The background $z = 1.24$ galaxy is visible in emission from the 230.5 GHz CO(2-1) transition, redshifted to 102.8 GHz (figure 2). However, the emission is fainter than
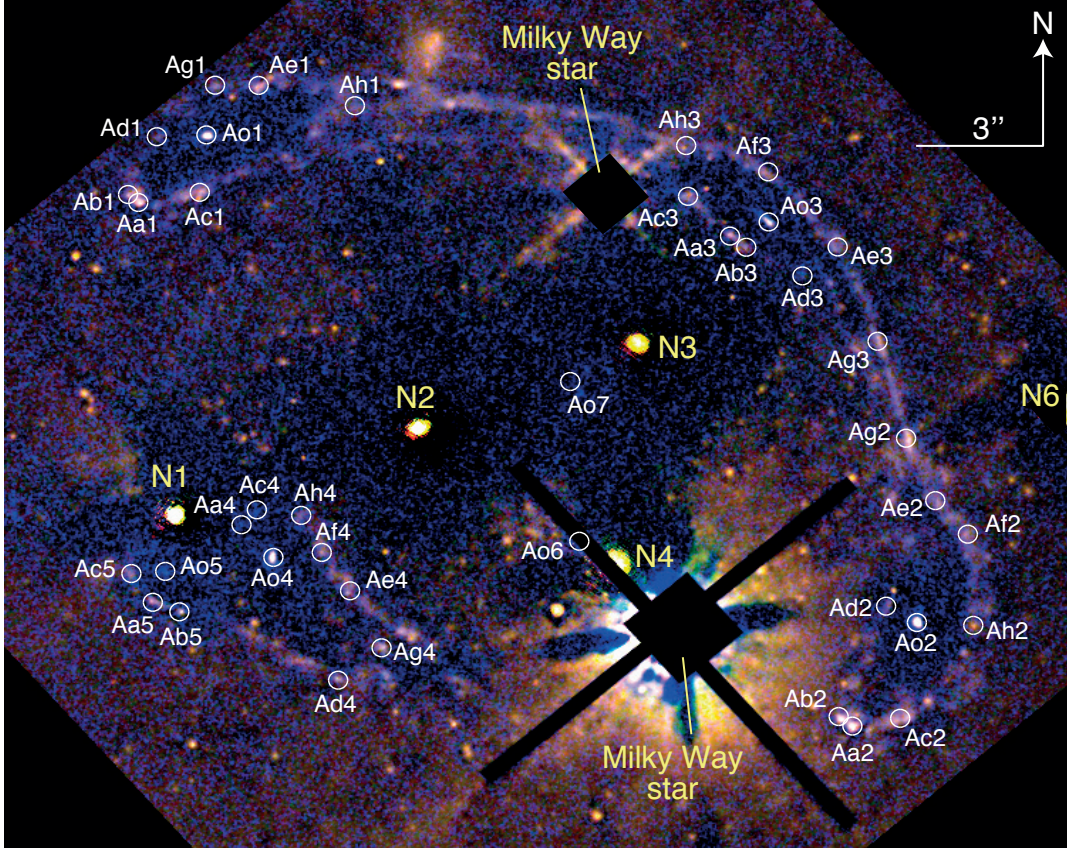
Figure 1. *Hubble Space Telescope* image of the core of Abell 3827 in F814W (red), F606W (green) and F336W (blue) bands. Light from two foreground stars and five foreground galaxies (labelled in yellow) has been subtracted to reveal the background lens system. The colour scale is linear. Multiply-imaged components of the background spiral galaxy, identified either in this image or in ALMA/MUSE data are labelled in white. In our cosmological model, $3'' = 5.5$ kpc at the redshift of the cluster.

expected from an extrapolation of near-IR emission (a somewhat indirect chain using [OII] emission to estimate star formation rate and hence far-infrared luminosity, then using the Solomon & Vanden Bout 2005 relation to predict CO luminosity). Our exposure time was therefore only just sufficient to detect spatial structure in the line emission; no continuum emission is detected beneath the foregrounds.

2.3 VLT/MUSE integral field spectroscopy

In June 2016, we obtained a 4 hour integration of Abell 3827 using the Multi-Unit Spectroscopic Explorer (MUSE) Integral Field Unit (IFU) spectrograph (Bacon et al. 2010) on the European Southern Observatory (ESO) Very Large Telescope (VLT). We combined these data (programme 295.A-5018) with a pre-existing 1 hour exposure from programme 294.A-5014. All the observations were obtained in dark time, with V-band seeing better than $0.7''$ and good atmospheric transparency. The data sample wavelengths 475.0-935.1 nm with 0.125 nm binning and spectral resolution $R{=}4000$ at the red end.

© 2017 RAS, MNRAS **000**, 1–10

Data were reduced using v1.0 of the ESOREX pipeline, which extracts the spectra, applies wavelength and flat-field calibration, then forms the data cube. Each hour on sky included $3\times$ 20 minute exposures, dithered by $\sim 10''$. We aligned the individual exposures by registering the images of stars, then removed cosmic rays and pixel defects, and stacked the exposures using the EXP_COMBINE routine. Flux calibration was achieved using ESO standard stars which were reduced in an identical manner.

The background galaxy is visible in emission from the [OII] $\lambda 3726.8, 3729.2$ line doublet, redshifted to 835.5 nm. In each spatial pixel, we model the spectrum of foreground continuum emission as a low-order polynomial either side of 835.5 nm. We subtract this foreground emission, then integrate the remaining line flux as an [OII] narrow-band image (figure 3). We also use a two-Gaussian model to fit the [OII] doublet line ratio (3728.9/3726.2), line-of-sight velocity, and line width. Both components of the line are assumed to have the same width, and the measurement of spectral line width is corrected for instrumental broadening.

133

4   *R. Massey et al.*
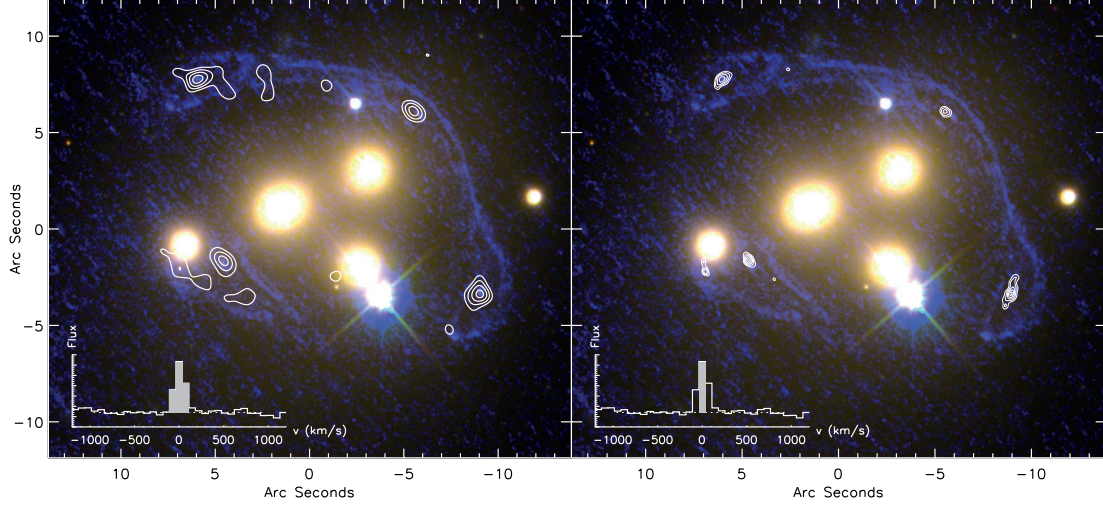
Figure 2. ALMA detection of CO(2-1) emission in the lensed spiral, as contours overlaid on the HST image from figure 1, before foreground subtraction. *Left:* CO(2-1) emission collapsed over $\pm 100\,\mathrm{km/s}$ from the systemic redshift and (u, v) tapered to a $0.8''$ beam, to show the full emission. The 1σ noise level is $0.15\,\mathrm{mJy/beam}$, and contours show $3, 4, 5, 6\sigma$. *Right:* CO(2-1) emission from a single, central ALMA channel, at natural $0.47''$ resolution, to identify multiple images of the source's bulge. The 1σ noise level is $0.08\,\mathrm{mJy/beam}$, and contours show $4, 5, 6, 7\sigma$. The inset spectra have a linear scale and include a dotted line at zero flux.
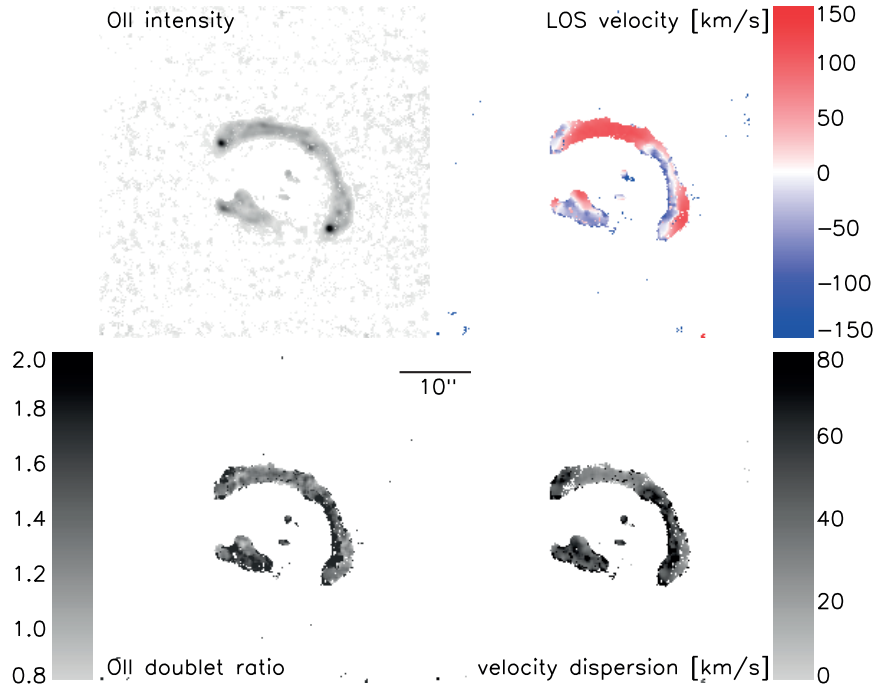
Figure 3. MUSE data compressed into a 2D narrow-band image of [O II] doublet line emission from the lensed spiral galaxy, after subtraction of the foreground continuum emission *(top left)*. To cross-identify regions of the galaxy in a way that is independent of the spatially varying lensing flux magnification, the other panels show parameters of a model fitted to the line doublet's spectral energy distribution in each spatial pixel where line emission is detected with S/N>3. The parameters are the line's local line-of-sight velocity *(top right)*, the flux ratio between the doublet's two components *(bottom left)* and the spectral line width *(bottom right)*.

3 ANALYSIS

3.1 Strong lens image identifications

The multiply-imaged background source is a spiral galaxy consisting of a red bulge 'Ao' inside a ring of star formation knots 'Aa'–'Ah'. Its rotational support is apparent from the ~ 200 km/s velocity gradient apparent across the galaxy in the MUSE data (and present at low S/N in ALMA data, but not shown in figure 2).

Features in the observed image have been variously identified as multiple versions of the background galaxy's bulge or star forming regions (Williams & Saha 2011; Massey et al. 2015; Taylor et al. 2017). Many of these features are deeply embedded within the light from foreground galaxies. After foreground subtraction using GALFIT (Peng et al. 2010), and based on its apparent colour and morphology, Massey et al. (2015) identified a point-like source immediately south of galaxy N1 as the fifth (sorted by arrival time in the best-fit mass model) multiple image of knot Aa.

Our new data indicate that this identification was incorrect. Our ALMA data show that the feature south of N1 is at the same systemic velocity and similar CO(2-1) flux as the source's central bulge. The feature's line-of-sight velocity is also inconsistent with that of star formation knot Aa. Our MUSE data also support a new interpretation that the feature is an additional image of the bulge, which we now call Ao5. This identification of the source's central bulge implies that images of knots Aa and Ac must be further southeast. The ALMA data is too low S/N to detect them, and the MUSE data have only barely sufficient angular resolution, but candidate features can be seen in HST imaging after our improved foreground subtraction using MuSCADeT (Joseph, Courbin & Starck 2016). These features were hidden behind the foreground emission from N1, and are fainter than the foreground cluster's many globular clusters. Indeed, the chain of three or four sources between Ao4 and Ao5 appears to be an unfortunate alignment of foreground globular clusters, confusingly unrelated to the background source.

Building upon this new hypothesis, and incorporating additional features resolved by ALMA and ordered by MUSE, a new set of multiple-image identifications Ao and Aa–h become clear (Table 1). This configuration of multiple image identifications was not amongst those considered in Appendix B of Massey et al. (2015). We shall now demonstrate that this new configuration yields a model of the lens's mass distribution with higher Bayesian evidence and better consistency with observed lensed image positions.

3.2 Mass model

To ensure that we can draw robust conclusions, we use two independent algorithms to infer the mass distribution in the lens. Both have been tested in a blind analysis of strong lensing data for which the true mass distribution is known (Meneghetti et al. 2017). First, we use LENSTOOL v6.8.1 (Jullo et al. 2007). Its parametric mass models may not capture all the complexity of a real mass distribution, but it allows quantities of scientific interest (such as the position of dark matter) to be parameterised explicitly and to be fitted directly from data. Second, we use GRALE (Liesenborgs, De Rijcke & Dejonghe 2006). This 'freeform' method possesses

Table 1. Locations of multiply-imaged components of the background spiral galaxy. Images Aon are the bulge, and images A[a–h]n are knots of star formation in the spiral arms.

Name	RA	Dec
Ao1	330.47479	−59.94358
Ao2	330.46649	−59.94665
Ao3	330.46828	−59.94411
Ao4	330.47407	−59.94623
Ao5	330.47529	−59.94634
Ao6	330.47044	−59.94614
Ao7	330.47054	−59.94514
Aa1	330.47559	−59.94400
Aa2	330.46725	−59.94732
Aa3	330.46871	−59.94421
Aa4	330.47443	−59.94605
Aa5	330.47546	−59.94652
Ab1	330.47571	−59.94395
Ab2	330.46741	−59.94726
Ab3	330.46852	−59.94428
Ab5	330.47515	−59.94658
Ac1	330.47487	−59.94394
Ac2	330.46669	−59.94726
Ac3	330.46920	−59.94396
Ac4	330.47424	−59.94596
Ac5	330.47571	−59.94634
Ad1	330.47537	−59.94359
Ad2	330.46685	−59.94656
Ad3	330.46784	−59.94446
Ad4	330.47327	−59.94701
Ae1	330.47420	−59.94327
Ae2	330.46627	−59.94589
Ae3	330.46745	−59.94428
Ae4	330.47315	−59.94644
Af2	330.46589	−59.94610
Af3	330.46826	−59.94381
Af4	330.47348	−59.94620
Ag1	330.47471	−59.94327
Ag2	330.46661	−59.94550
Ag3	330.46694	−59.94488
Ag4	330.47276	−59.94681
Ah1	330.47305	−59.94340
Ah2	330.46583	−59.94667
Ah3	330.46922	−59.94364
Ah4	330.47372	−59.94599

more flexibility to represent a real mass distribution and, by inferring unphysical distributions, to highlight errors in e.g. source image identification. However, interpretation is later required to extract quantities of scientific interest.

3.2.1 LENSTOOL

Our LENSTOOL mass model consists of dark matter in one cluster-scale Pseudo-Isothermal Elliptical Mass Distribution (PIEMD; Limousin et al. 2005; Elíasdóttir et al. 2007), plus the four bright galaxies' stellar and dark matter with respectively Hernquist and Pseudo-Isothermal Skewed Potential (PISP; Taylor et al. 2017) distributions. A PISP distribution reduces to a PIEMD if its skewness $s = 0$. We also fit a PISP component to the dark matter associated with faint member galaxy N6, but assume it has negligible stellar mass and skewness to reduce parameter space. Including

6   *R. Massey et al.*

mass associated with galaxies farther from the cluster core yields indistinguishable results but slows the analysis dramatically, so we omit them. Finally, we allow an external shear (e.g. Hogg & Blandford 1994).

Parameters of the dark matter components are adjusted to reduce the rms of distances between the source galaxy's predicted and observed positions in the image plane, $\langle \text{rms} \rangle_i$. The parameters' posterior probability distribution function (PDF) is explored by a Markov Chain Monte Carlo (MCMC) iteration, with a constant proposal distribution after a burn-in phase (LENSTOOL's RUNMODE=3) and priors identical to those in Taylor et al. (2017). For example, the location and amount of each galaxy's dark matter is given a flat prior $2''$ either side of its stars. Taylor et al. (2017) reported failed convergence of skewness parameters; this has been solved by a much longer Markov Chain that samples the PDF 100,000 times, and by ensuring that the skewness angle ϕ_s wraps far from any peak in the PDF. We assume statistical uncertainty of $0.5''$ on the location of Ao6 and Ao7, which are detected only in ground-based data, and $0.15''$ on the location of every other image.

Parameters of the stellar mass components are derived from GALFIT fits to flux in the F606W band, with the flux converted into mass via Bruzual & Charlot (2003) models, assuming a Chabrier (2003) initial mass function, solar metallicity, and a single burst of star formation at redshift $z_\text{f} = 3$. These parameters are fixed during the optimisation.

3.2.2   GRALE

Our GRALE mass model incorporates a grid of approximately 1300 Plummer spheres (Plummer 1911) in a $50'' \times 50''$ region centered on (RA: 330.47043, Dec: -59.945804). An iterative procedure adaptively refines the grid in dense regions, and uses a genetic algorithm to adjust the mass in each Plummer sphere. The genetic algorithm optimises the product of (a) the fractional degree of overlap between multiple images of the same source in the source plane and (b) a fitness measure penalising the presence of false counter-images in regions where they are not observed.

We run twenty mass reconstructions with different random seeds. n total, this produces 26786 optimised Plummer spheres. We average the inferred mass distributions; their rms provides an estimate of statistical error.

4   RESULTS

Inferred mass maps are presented in figure 4. Results from LENSTOOL and GRALE are now more consistent with each other. They also provide a better fit to the data than they were when assuming the source identifications of Massey et al. (2015) (whose LENSTOOL model had $\langle \text{rms} \rangle_i = 0.26''$). The new parameters of LENSTOOL's best-fit model are presented in Table 2. This model achieves $\langle \text{rms} \rangle_i = 0.13''$, or $\chi^2 = 31.7$ with 29 degrees of freedom, likelihood $\log(\mathcal{L}) = 59.9$, and Bayesian evidence $\log(\mathcal{B}) = -11.5$.

Central images Ao6 and Ao7 have split GRALE's previous reconstruction of a bimodal cluster (consisting of N1 plus everything else) into four distinct mass concentrations around each galaxy. There is no reason for the genetic algorithm to prefer either, yet the new model is more physical.
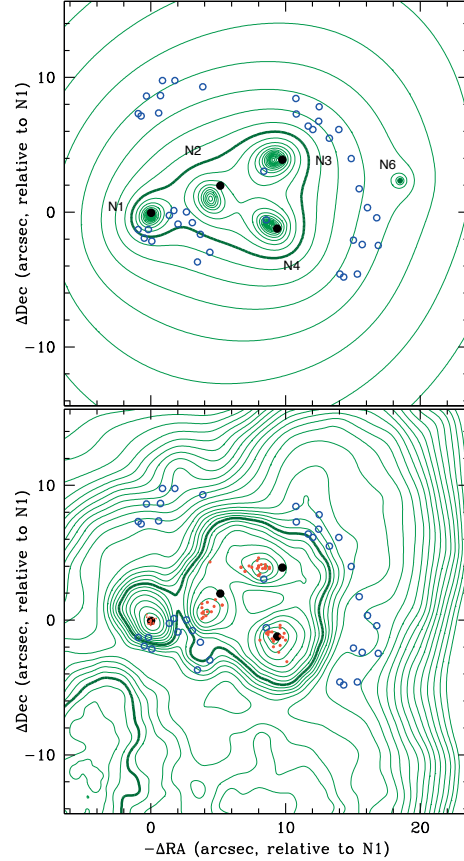
Figure 4. *Top:* Map of total mass in the cluster core, reconstructed using LENSTOOL, and averaging over the posterior PDF. Green contours show the projected mass density, spaced logarithmically by a factor 1.15; the thick contour shows convergence $\kappa = 1$ for $z_{c\ell} = 0.099$ and $z_\text{A} = 1.24$ ($\Sigma_\text{crit} = 1.03\,\text{g/cm}^2$). Blue circles show the lensed images. Black dots show cluster ellipticals N1–N4. *Bottom:* Total mass, as in the top panel but reconstructed via GRALE. Red dots show local maxima in individual realisations of the mass map.

Adding the central images also creates a prediction (from both LENSTOOL and GRALE) for a diffuse trail of source emission southwest of Ag4, including counter-images of Ab, Ad, and Ag. These are possibly demagnified and observed, but the whole area is unclear in HST imaging because of the bright foreground star and confusion with globular clusters. Both models predict demagnified images of the star formation knots tightly packed around Ao6, and loosely packed around Ao7, as are visible in MUSE data but with insufficient confidence to be used as input constraints.

Our LENSTOOL analysis suggests a $\sim 2\sigma$ statistical significance for the offset of N1. However, the absolute value of the offset is far smaller than in Massey et al. (2015), and its significance disappears entirely when combining with our GRALE analysis and allowing for systematic, model-induced biases of up to $\sim 0.21''$ for this configuration of lenses

Table 2. Parameters of LENSTOOL's best-fit mass model. Quantities in square brackets were fixed during optimisation. Errors on other quantities show 68% statistical confidence limits, marginalising over uncertainty in all other parameters. Stellar mass components are modelled as Hernquist profiles, with their mass, scale radius and ellipticity calculated from F606W broad-band emission. Dark matter components are modelled as PIEMDs with a 1D velocity dispersion, core and cut radii, and ellipticity; or PISPs with an additional skewness. Positions are given in arcseconds relative to (R.A.: 330.47518, Dec.: -59.945985), except galaxies' dark matter components, which are relative to the position of their stars. Angles are anticlockwise from West. The external shear is $(1.47^{+0.97}_{-0.01})\%$, at angle $(92^{+24}_{-94})°$.

		$x\,['']$ / $\Delta x\,['']$	$y\,['']$ / $\Delta y\,['']$	Mass $[M_\odot]$ / σ_v [km/s]	$r_{sc}\,['']$ / $r_{core}\,['']$	$r_{cut}\,['']$	ϵ	$\phi_\epsilon\,[°]$	s	$\phi_s\,[°]$
N1	stars	[0.00]	[0.00]	$[1.00 \times 10^{11}]$	[0.53]		[0.12]	[61]		
	dark matter	$0.09^{+0.10}_{-0.09}$	$-0.28^{+0.13}_{-0.12}$	166^{+9}_{-11}	[0.10]	[40]	$0.12^{+0.25}_{+0.00}$	56^{+52}_{-18}	$0.14^{+0.08}_{-0.28}$	102^{+12}_{-106}
N2	stars	[5.13]	[2.00]	$[2.47 \times 10^{11}]$	[0.79]		[0.17]	[39]		
	dark matter	$-0.81^{+0.19}_{-0.20}$	$-0.59^{+0.33}_{-0.29}$	170^{+13}_{-18}	[0.10]	[40]	$0.38^{+0.01}_{-0.25}$	129^{+16}_{-22}	$0.10^{+0.09}_{-0.15}$	41^{+94}_{-23}
N3	stars	[9.75]	[3.93]	$[2.76 \times 10^{11}]$	[0.33]		[0.05]	[31]		
	dark matter	$-0.57^{+0.14}_{-0.14}$	$0.08^{+0.24}_{-0.16}$	214^{+6}_{-14}	[0.10]	[40]	$0.14^{+0.07}_{-0.08}$	14^{+18}_{-8}	$-0.09^{+0.09}_{-0.07}$	41^{+67}_{-27}
N4	stars	[9.32]	[-1.12]	$[2.06 \times 10^{11}]$	[1.37]		[0.39]	[127]		
	dark matter	$-0.54^{+0.34}_{-0.11}$	$0.40^{+0.09}_{-0.20}$	206^{+7}_{-15}	[0.10]	[40]	$0.32^{+0.33}_{+0.00}$	144^{+12}_{-65}	$0.12^{+0.11}_{-0.12}$	104^{+53}_{-58}
N6	stars	[18.60]	[2.43]	[0]						
	dark matter	[0.00]	[0.00]	61^{+13}_{-27}	[0.10]	[40]	[0.00]	[0]	[0]	[0]
Cluster	dm	$8.61^{+0.89}_{-0.90}$	$-0.28^{+1.04}_{-0.79}$	842^{+77}_{-89}	30^{+5}_{-7}	[1000]	$0.50^{+0.07}_{-0.15}$	62^{+2}_{-2}	[0]	[0]

(Massey et al. 2015). The mass peak reconstructed by GRALE outside the cluster core imposes an external shear near N1 consistent with that fitted by LENSTOOL.

Statistical errors on the position of dark matter associated with N2, N3 and N4 are tightened by our new detection of central images Ao6 and Ao7. They would be dramatically improved if more of the source galaxy's structure could be seen in the central images (e.g. with deeper ALMA data). However, the position of N2's dark matter shows a large scatter in our current GRALE analysis, and can change in a LENSTOOL analysis if the prior is adjusted on the position of the cluster-scale halo. In the MCMC chain of our LENSTOOL analysis, the positions of N3 and N4 are degenerate with each other. Furthermore, we have an *a priori* expectation that only N1 is sufficiently close to space-resolution lensed images to be constrained with kiloparsec accuracy (even when the lens identifications are unambiguous Harvey, Kneib & Jauzac 2016, and they may still not all be correct here).

The inferred location of the dark matter associated with each galaxy N1–N4 appears consistent with the location of its stars. Deeper ALMA or *HST* observations would clarify the status of N2, N3 and N4. However, given various parameter degeneracies in our current analysis, and the potential for systematic errors at a level comparable to their offsets, we cannot here conclude that any offset is physically significant.

The total mass of the dark matter components of galaxies N1–N4 is formally $1.47^{+0.16}_{-0.19}$, $1.54^{+0.24}_{-0.31}$, $2.44^{+0.14}_{-0.31}$ and $2.26^{+0.16}_{-0.32} \times 10^{12}\,M_\odot$, and that of the cluster-scale halo is $2.79^{+0.53}_{-0.56} \times 10^{14}\,M_\odot$ (see equation 10 of Limousin et al. 2005). However, these calculations depend approximately linearly on our unconstrained choice of r_{cut}.

De-lensed images of the background galaxy, assuming the best-fit LENSTOOL model, are presented in figure 5; results from GRALE are similar. It is a ring galaxy reminiscent

of the $z = 1.67$ lensed source in Zwicky cluster Cl0024+1654 (Colley, Tyson & Turner 1996; Jones et al. 2010). Its central component is by far the brightest in CO(2-1) emission. A large reservoir of dusty, molecular gas in a galaxy's bulge would be unusual at $z = 0$, but not at $z = 1.24$, when bulges are still forming stars. Assuming LENSTOOL's best-fit mass model, the luminosity-weighted amplification of its [OII] emission is $\mu = 144$, summing over all the images. Taking into account this amplification, its apparent [OII] luminosity implies an total star-formation rate of $\sim 1\,M_\odot/\text{yr}$, using the Kennicut (1998) calibration and a Chabrier (2003) initial mass function. Canonical dust extinction of about $A_V \sim 1$ magnitude could raise this by a factor 2–3. Other than its role in gravitational lensing due to its location behind a cluster, it is not an intrinsically unusual galaxy.

5 CONCLUSIONS

Previous studies of galaxy cluster Abell 3827 (Williams & Saha 2011; Massey et al. 2015) imaging suggested that the dark matter associated with at least one of its galaxies is offset from its stars. This is predicted by simulations of self-interacting dark matter in which the exchange particle is light (Harvey et al. 2014; Kahlhoefer et al. 2014; Robertson et al. 2017b). Prompted by this potentially exciting result, further simulations (Kahlhoefer et al. 2015) suggested that the offset could be observable in (rare) systems where a massive galaxy intersects a cluster's Einstein radius, and its 3D motion happens to be near the plane of the sky. A strongly lensing merger between two field galaxies has shown a similar offset (Shu et al. 2016).

In this paper, ALMA has proved an exceptional tool to identify background lensed images, with high spatial resolution at wavelengths where foreground galaxy clusters are virtually transparent. Whilst there is no guarantee that we have perfected the source identifications in Abell 3827, it is

8   *R. Massey et al.*

now possible to construct lens models with residuals that are consistent with noise, and robust between very different modelling approaches. The consistency between parametric and non-parametric lens models lends confidence to the conclusions. Indeed, both ALMA data and deviations from physically expected mass distributions in a free-form mass reconstruction could be a powerful discriminator between future source identifications.

Our new analysis shows that there is no statistically significant offset between galaxies and their dark matter in Abell 3827, projected onto the plane of the sky. Galaxy N1 is best constrained. Assuming statistical errors only, its offset in our LENSTOOL model is $0.29^{+0.12}_{-0.13}$ arcseconds or $0.54^{+0.22}_{-0.23}$ kpc. Following Kahlhoefer et al. (2015)'s reasoning that any offset requires dark matter self-interactions to balance a gravitational restoring force that can be calculated, this measurement implies an interaction cross-section $(\sigma/m)\cos(i) = 0.68^{+0.28}_{-0.29}\,\mathrm{cm}^2/\mathrm{g}$, where i is the inclination of the galaxy's 3D motion with respect to the plane of the sky. That this angle is unknown makes it difficult to infer an upper limit on σ/m from this system without further information.

Nonetheless, the unusual configuration of Abell 3827, with four bright central galaxies and a background spiral galaxy with complex morphology is multiply-imaged between them, makes it still interesting for studies of dark matter dynamics. Regardless of possible particle interactions, as a galaxy enters a cluster, its dark matter halo is gradually stripped via tidal gravitational forces. Simulations disagree about the timescale and the orbits on which dark matter stripping occurs in the inner tens of kiloparsecs (Diemand, Kuhlen & Madau 2007; Peñarrubia, McConnachie & Navarro 2008; Wetzel, Cohn & White 2009; Bahé et al. 2012), but this dissipation is a key ingredient in semi-analytic models of galaxy formation (e.g. Dariush et al. 2010). Observations of dark matter mass loss in galaxies entering a galaxy cluster from the field (Mandelbaum et al. 2006; Limousin et al. 2007, 2012; Parker et al. 2007; Natarajan et al. 2009; Gillis et al. 2013; Niemiec et al. 2017) have never been followed inside ∼1 Mpc, and measurements of strong lensing clusters with multiple central galaxies, like those in table 2, could constrain this for the first time.

Figure 5. *Top:* De-lensed images of the $z = 1.24$ source galaxy, after foreground subtraction as in figure 1 and assuming the best-fit lens model produced by LENSTOOL in table 2. Each panel is $1.5'' \times 1.5''$, and centered on (RA=22h 01′ 53″0, Dec=−59° 56′ 44″). Results from GRALE are similar. *Bottom:* Re-lensed version of the above realisation of source A3, the centre of the triple. The predicted brightness of the central images Ao6 and Ao7 changes sightly if other versions of the source are re-lensed.

REFERENCES

Bacon R. et al. 2010, SPIE 7735, 773508

Bruzual G. & Charlot S., 2003, MNRAS 344, 1000

Bahé Y., McCarthy I., Crain R. & Theuns T. 2012 MNRAS 424, 1179

Bradač M. et al. 2008 ApJ 687, 959

Carrasco E. et al. 2010 ApJ 715, 160

Chabrier G. 2003 PASP 115, 763

Chon G. & Böhringer H. 2015, A&A 574, 132

Clowe D., Gonzalez A. & Markevitch M. 2004 ApJ 604, 596

Clowe D., Bradač M., Gonzalez A., Markevitch M., Randall S., Jones C. & Zaritsky D. 2006 ApJL 648, 109

Clowe D. Markevitch M., Bradač M., Gonzalez A., Massey R. & Zaritsky D. 2012 ApJ 758, 128

Colley W., Tyson J. & Turner E. 1996 ApJL 461, 83

Cyr-Racine F.-Y. & Sigurdson K. 2013 PhRvD 87, 3515

Dariush A., Raychaudhury S., Ponman T., Khosroshahi H., Benson A., Bower R. & Pearce F. 2010 MNRAS 405, 1873

Dawson W. et al. 2012 ApJL 747, 42

Dawson W. 2013, ApJ, 772, 131

De Plaa J., Werner N., Bleeker J., Vink J., Kaastra J. & Méndez M. 2007 A&A 465, 345

Diemand J., Kuhlen M. & Madau P. 2007 ApJ 667, 859

Elíasdóttir Á. et al. 2007, arXiv:0710.5636

Gastaldello F. et al. 2014 MNRAS Lett. 442, 76

Gao L., De Lucia G., White S. & Jenkins A. 2004 MNRAS 352, L1

Gillis B. et al. 2013 MNRAS 431, 1439

Golovich N., Dawson W., Wittman D., Ogrean G., van Weeren R. & Bonafede A. 2016 ApJ 831, 110

Golovich N., van Weeren R., Dawson W., Jee W. & Wittman D. 2017, ApJ in press (arXiv:1703.04803)

Harvey D. et al. 2014 MNRAS 441, 404

Harvey D. Massey R., Kitching T., Taylor A. & Tittley E. 2015 Science 347, 1462

Harvey D., Kneib J.-P. & Jauzac M. 2016 MNRAS 456

Hogg D. & Blandford R. 1994 MNRAS 268, 889

Jauzac M. et al. 2014 MNRAS 443, 1549

Joseph R., Courbin F. & Starck J.-L. 2016 A&A 589, 2

Jee M., Dawson W., Stroe A., Wittman D., van Weeren R., Brüggen M., Bradac M. & Röttgering H. 206 ApJ 817, 179

Jones T., Swinbank A., Ellis R., Richard J. & Stark D. 2010 MNRAS 404, 1247

Jullo E., Kneib J.-P., Limousin M., Elasdóttir Á., Marshall P. & Verdugo T. 2007 NJPh 9, 447

Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. & Sarkar S. 2014 MNRAS 437, 2865

Kahlhoefer F., Schmidt-Hoberg K., Kummer J. & Sarkar S. 2015 MNRAS Lett. 452,54

Kennicutt, R. C. 1998 ARA&A 36, 189

Kim S., Peter A. & Wittman D. 2017 MNRAS 469, 1414

Kneib J.-P. & Natarajan P. 2011 A&A Rev. 19, 47

Liesenborgs J., De Rijcke S. & Dejonghe H. 2006 MNRAS 367, 1209

Limousin M., Kneib J.-P. & Natarajan P., 2005 MNRAS 356, 309

Limousin M., Kneib J.-P., Bardeau S., Natarajan P., Czoske O., Smail I., Ebeling H. & Smith G. 2007 A&A 461, 881

Limousin M. et al. 2012 A&A 544, 71

Mahdavi A., Hoekstra H., Babul A., Balam D. & Capak P. 2007 ApJ 668, 806

Mandelbaum R. Seljak U., Kauffmann G., Hirata C. & Brinkmann J. 2006 MNRAS 368, 715

Massey R., Kitching T. & Richard J. 2010 Rep. Prog. Phys. 73, 086901

Massey R., Kitching T. & Nagai D. 2011 MNRAS 413, 1709

Massey R. et al. 2015 MNRAS 449, 3393

McMullin J. P. , Waters B., Schiebel D., Young W. & Golap K. 2007 Astronomical data analysis software and systems XVI 376, 127

Meneghetti M. et al. 2017 MNRAS submitted, arXiv:1606.04548

Merten J. et al. 2011 MNRAS 417, 333

Monteiro-Oliveira R., Cypriano E., Machado R., Lima Neto G., Ribeiro A., Sodré L. & Dupke R. 2017 MNRAS 466, 2614

Nagai D. & Kravtsov A. 2005 ApJ 618, 557

Natarajan P. , Kneib J.-P., Smail I., Treu T., Ellis R., Moran S., Limousin M. & Czoske O. 2009 ApJ 693, 970

Ng K. et al. 2015 MNRAS 453, 1531

Niemiec A. et al. 2017 arXiv:1703.03348

Parker L., Hoekstra H., Hudson M., van Waerbeke L. & Mellier Y. 2007 ApJ 669, 21

Peñarrubia J., McConnachie A. & Navarro J. 2008 ApJ 672, 904

Peng C., Ho L., Impey C. & Rix H.-W. 2010 AJ 139, 2097

Peter A. et al. 2012 arXiv:1208.3026

Plummer H. 1911 MNRAS 71, 460

Randall S., Markevitch M., Clowe D., Gonzalez A. & Bradač M. 2008 ApJ 679, 1173

Robertson A., Massey R. & Eke V. 2017a, MNRAS 465, 569

Robertson A., Massey R. & Eke V. 2017b, MNRAS 467, 4719

Schaller M., Robertson A., Massey R., Bower R. & Eke V. 2015, MNRAS Lett. 453, 58

Schaye J. et al. 2015, MNRAS 446, 521

Shu Y., Bolton A., Moustakas L., Stern D., Dey A., Brownstein J., Burles S. & Spinrad H. 2016, ApJ 820, 43

Solomon P. & Vanden Bout P. 2005 ARA&A 43, 677

Taylor P., Massey R., Jauzac M., Courbin F., Harvey D., Joseph R. & Robertson A. 2017 MNRAS 468, 5004

Vogelsberger M. et al. 2014 Nature 509, 177

Wetzel A., Cohn J. & White M. 2009 MNRAS 395, 1376

Williams L. & Saha P. 2011 MNRAS 415, 448

10   *R. Massey et al.*

This paper has been typeset from a TEX/ LATEX file prepared
by the author.

Filter	MACS J0416		MACS J0717		MACS J1149		Abell 2744		Abell 370		Abell S1063	
	Red	Blue	Red	Blue	Red	Blue	Red	Blue	Red	Blue	Red	Blue
F814w	0.599	0.342	0.666	0.399	0.655	0.393	0.594	0.415	0.571	0.331	0.598	0.348
F606w	0.373	0.473	0.311	0.463	0.324	0.495	0.373	0.496	0.394	0.482	0.371	0.463
F435w	0.027	0.183	0.022	0.137	0.020	0.110	0.032	0.087	0.034	0.186	0.029	0.187

Table 4.2 – SEDs used in the colour subtraction of the Hubble frontier fields.

4.3.2 A Colourful View of the Hubble Frontier Fields

In this section we present our colour separation performed with `MuSCADeT` on all the HFF cluster fields and the parallel fields for the MACS clusters. Due to the large size of the fields of view (~ 7000 pixels on-a-side per band), the images were separated in overlapping patches 1024 pixels on-a-side, and `MuSCADeT` was ran on each patch independently. The SED is estimated once for each cluster using a pre-selected patch with identifiably red and blue separate components. The same SEDs were used for each patch of a same FoV (field of view) and for each field of a target (cluster and parallel FoV). in the following pages, we show, in figure 4.4 - 4.30 the original images, Y, of the FoV in the three HST bands of interest here: F814w (red), F606w (green), F435w (blue) , along with the subtractions, R_b and R_r, of the blue and red components respectively. Using the notations from section 4.2, the subtraction form the red component corresponds to the cube described by

$$R_r = Y - A[:,0]S[0,:] \qquad (4.1)$$
$$R_b = Y - A[:,1]S[1,:], \qquad (4.2)$$

where $A[i,:]$ is the i-th column of the mixing matrix A, i.e. the SED of the corresponding component, and $S[i,:]$ is the i-th line in S, i.e. the surface brightness of the i-th component. The original image and the colour component residuals R_b and R_r are display with the same colour scales. The normalised SEDs of each FoV are given in table 4.2. We show the residuals $Y - AS$ in appendix D. The legwork of applying `MuSCADeT` to the HFF has to be credited to , at the time, grad. student Martin Millon.

Due to the limit size imposed by the format of the document, the images might apper in a lower resolution. High resolution images of the deblended Hubble frontier fields may be found at colour-deblending-of-the-hubble-frontier-fields.

MACS J0416 cluster
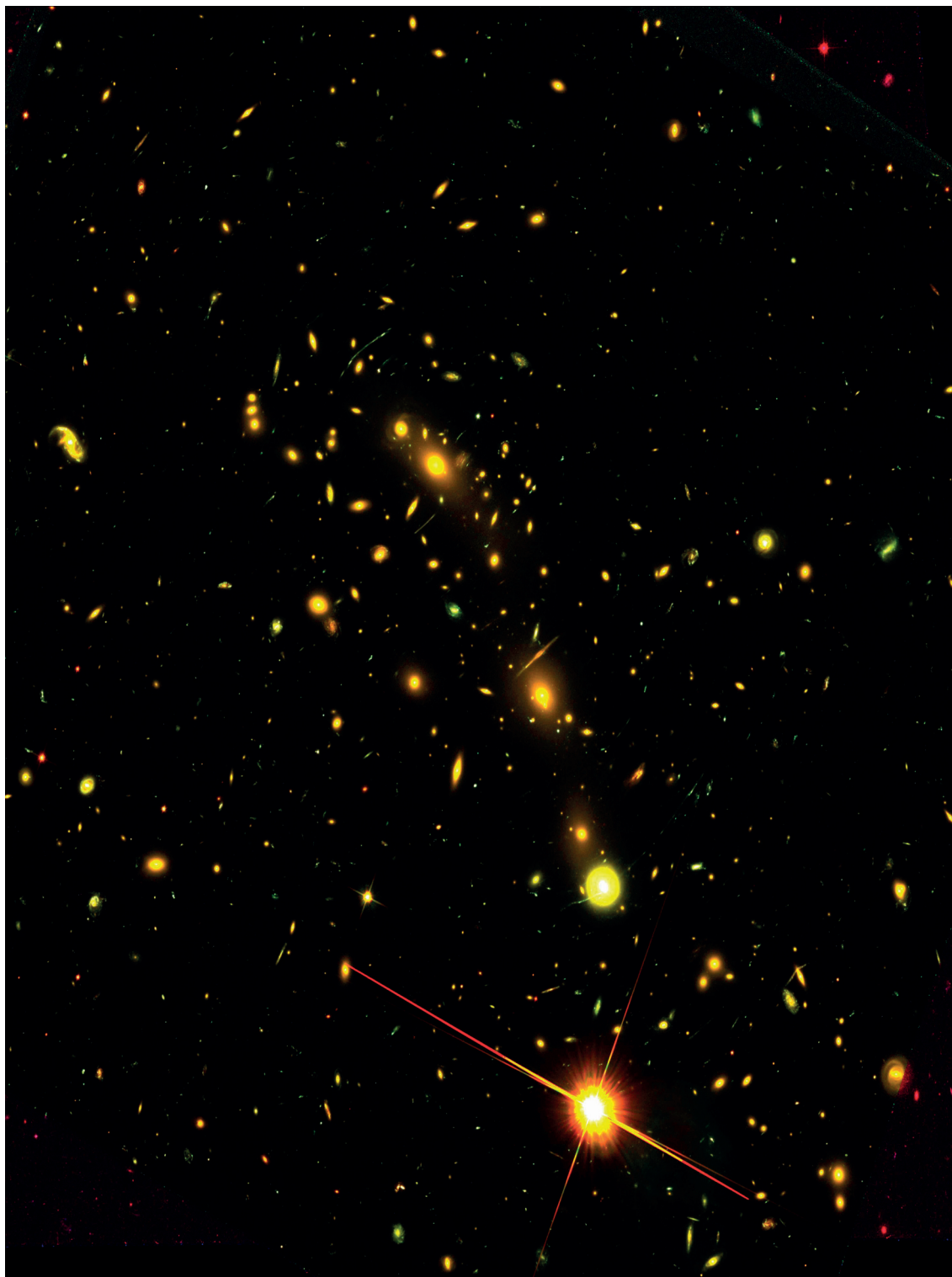
Figure 4.4 – Composite RGB image of cluster field MACS J0416.
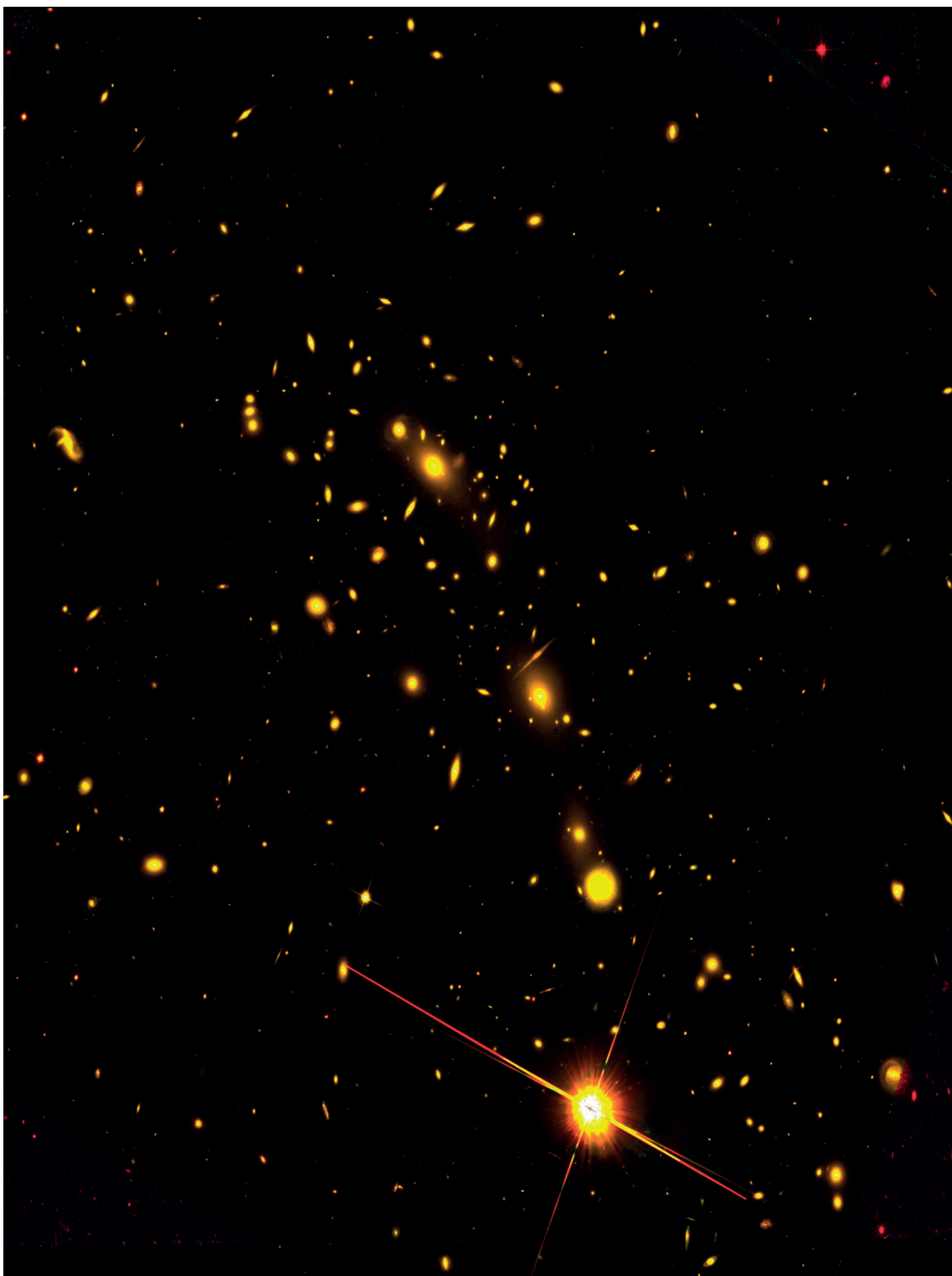
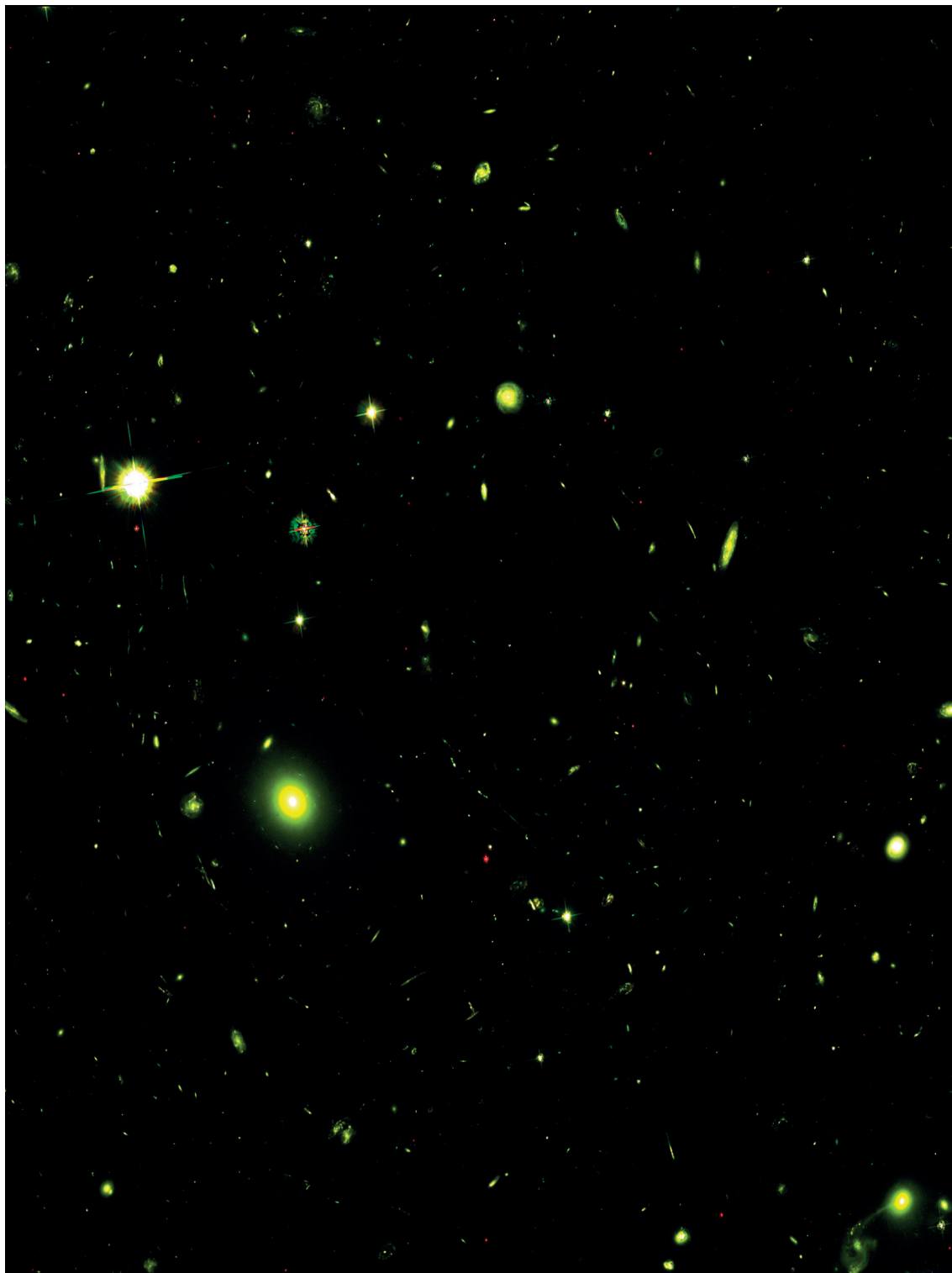Figure 4.5 – Composite RGB image of the subtraction of the blue component from cluster field MACS J0416.
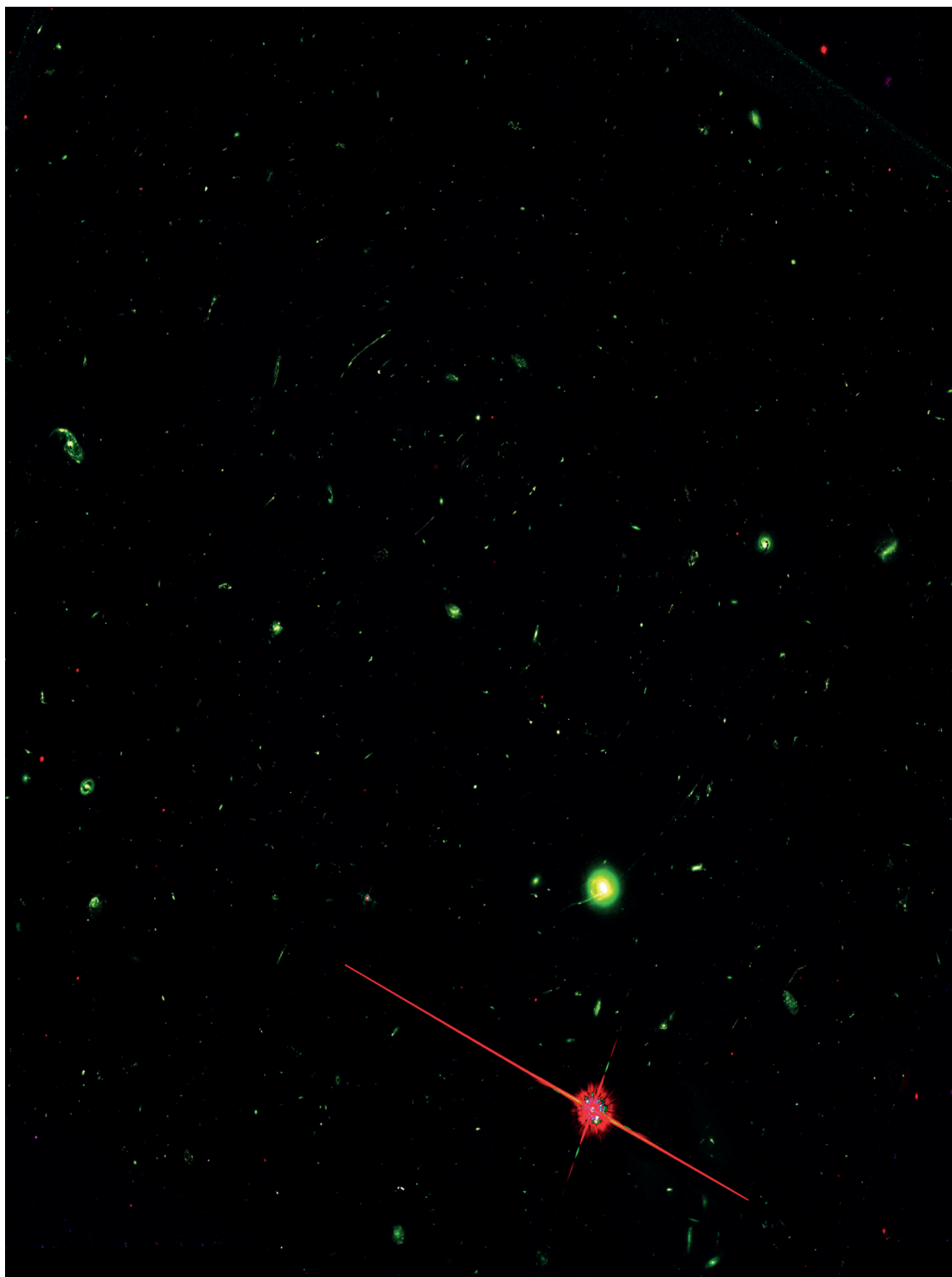
Figure 4.6 – Composite RGB image of the subtraction of the red component from cluster field MACS J0416.
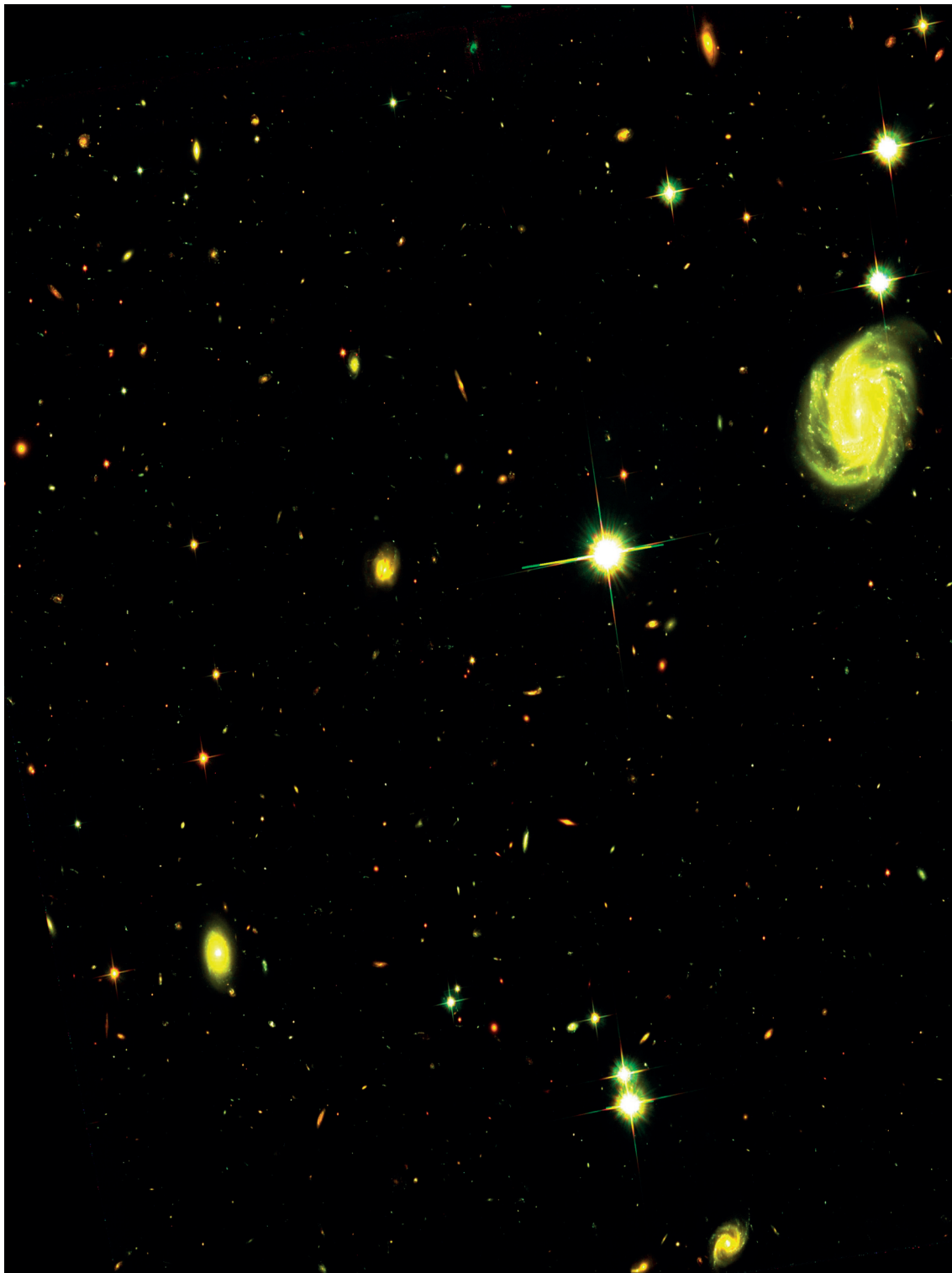
MACS J0416 parallel
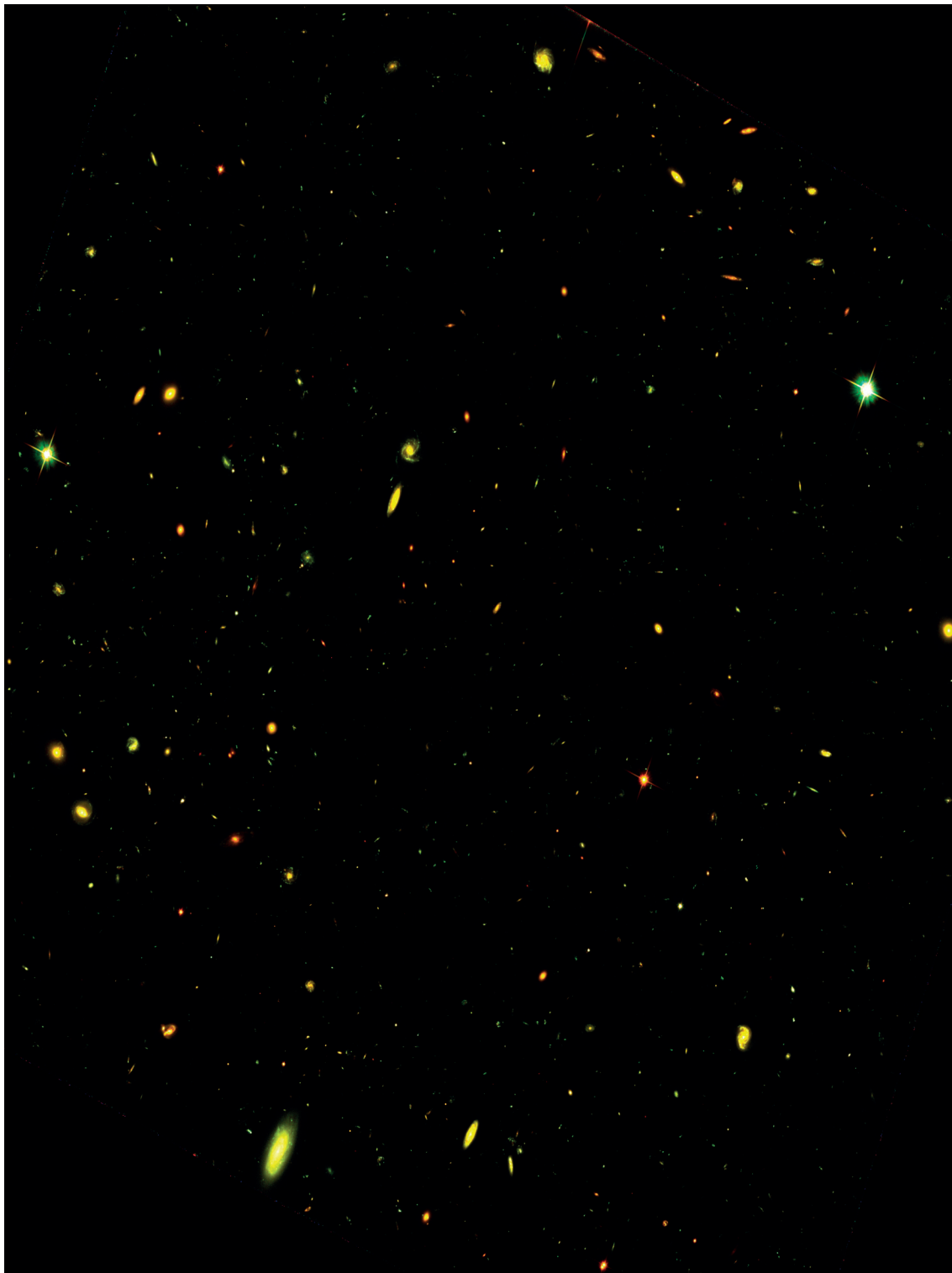
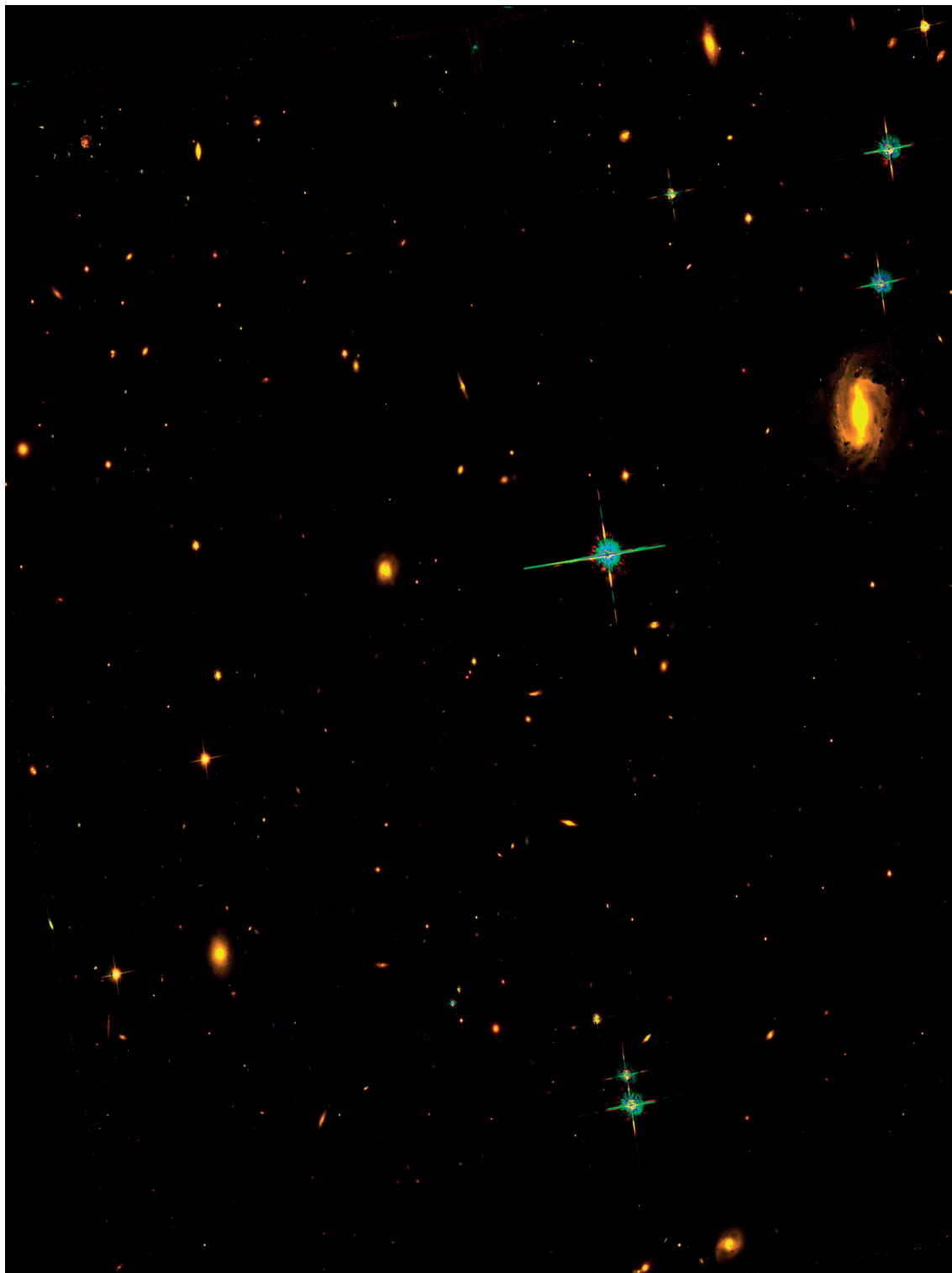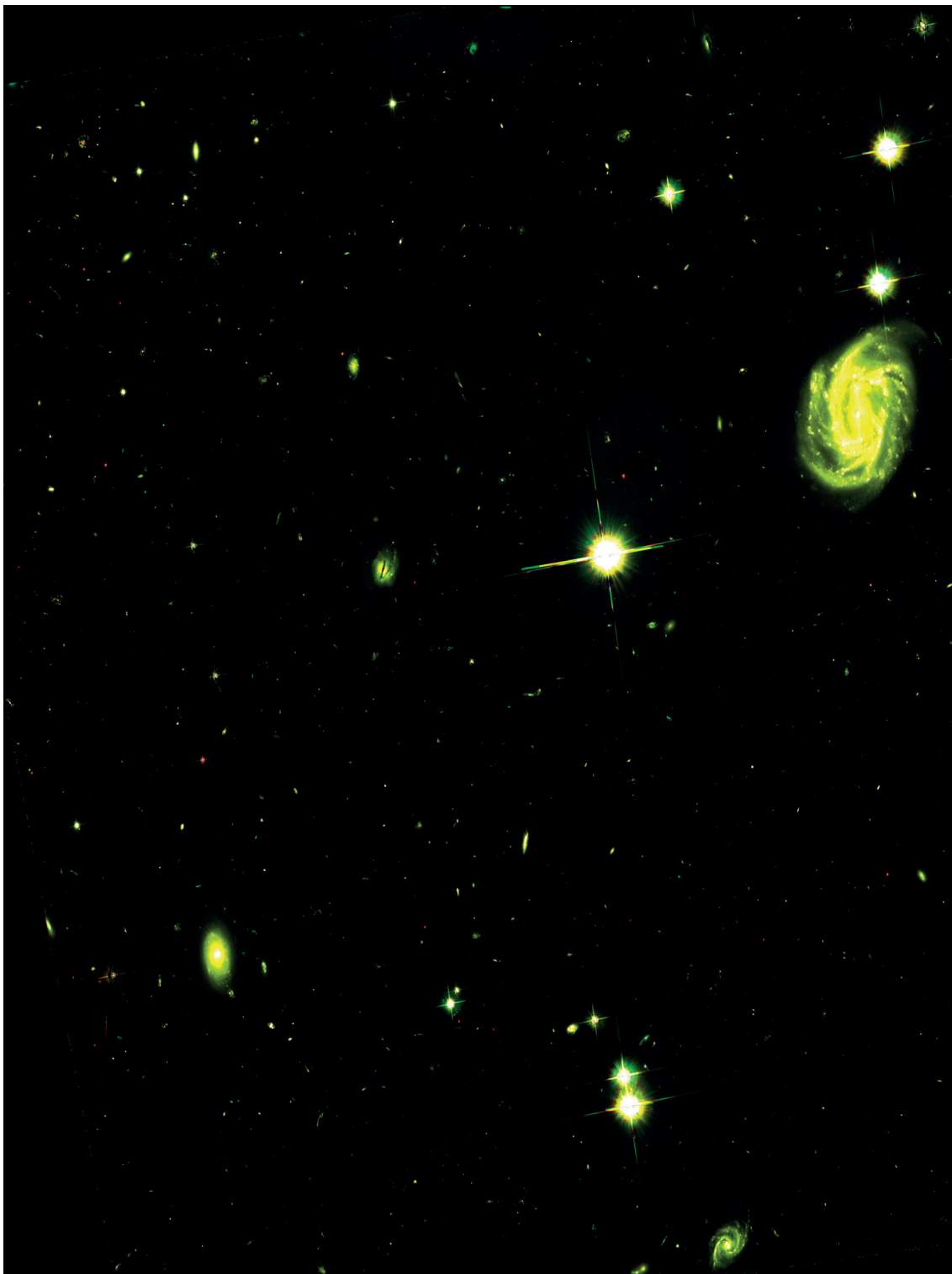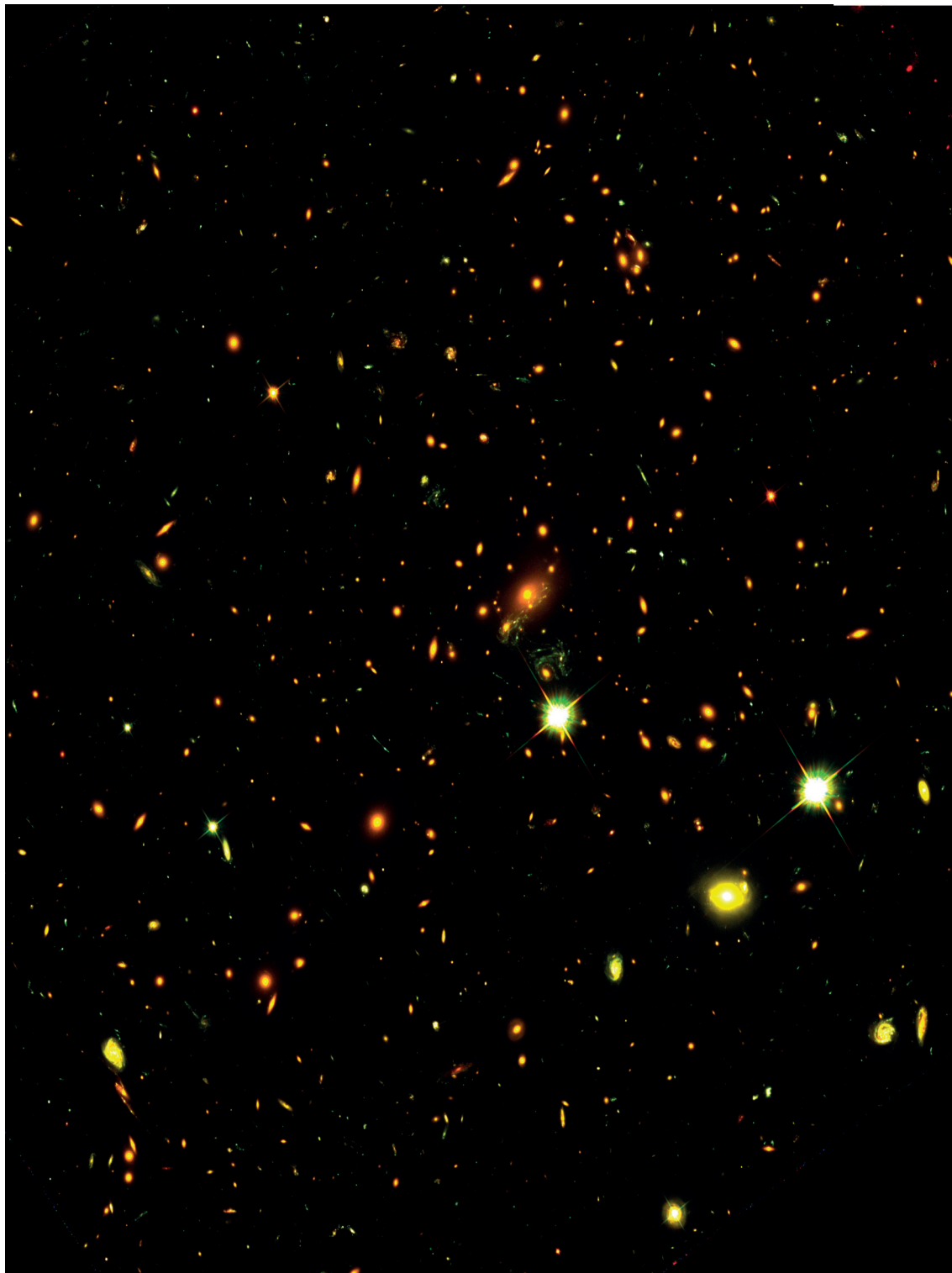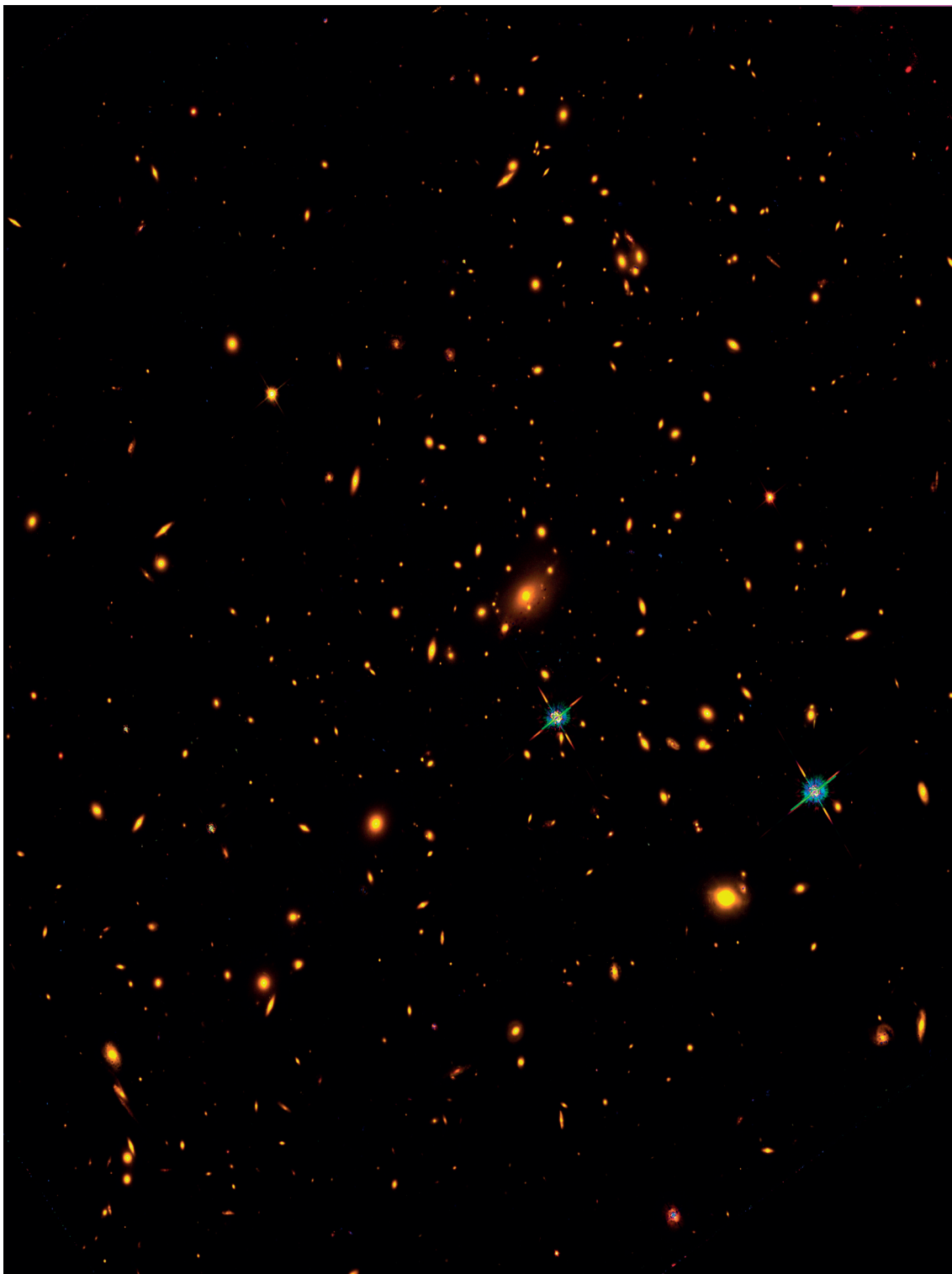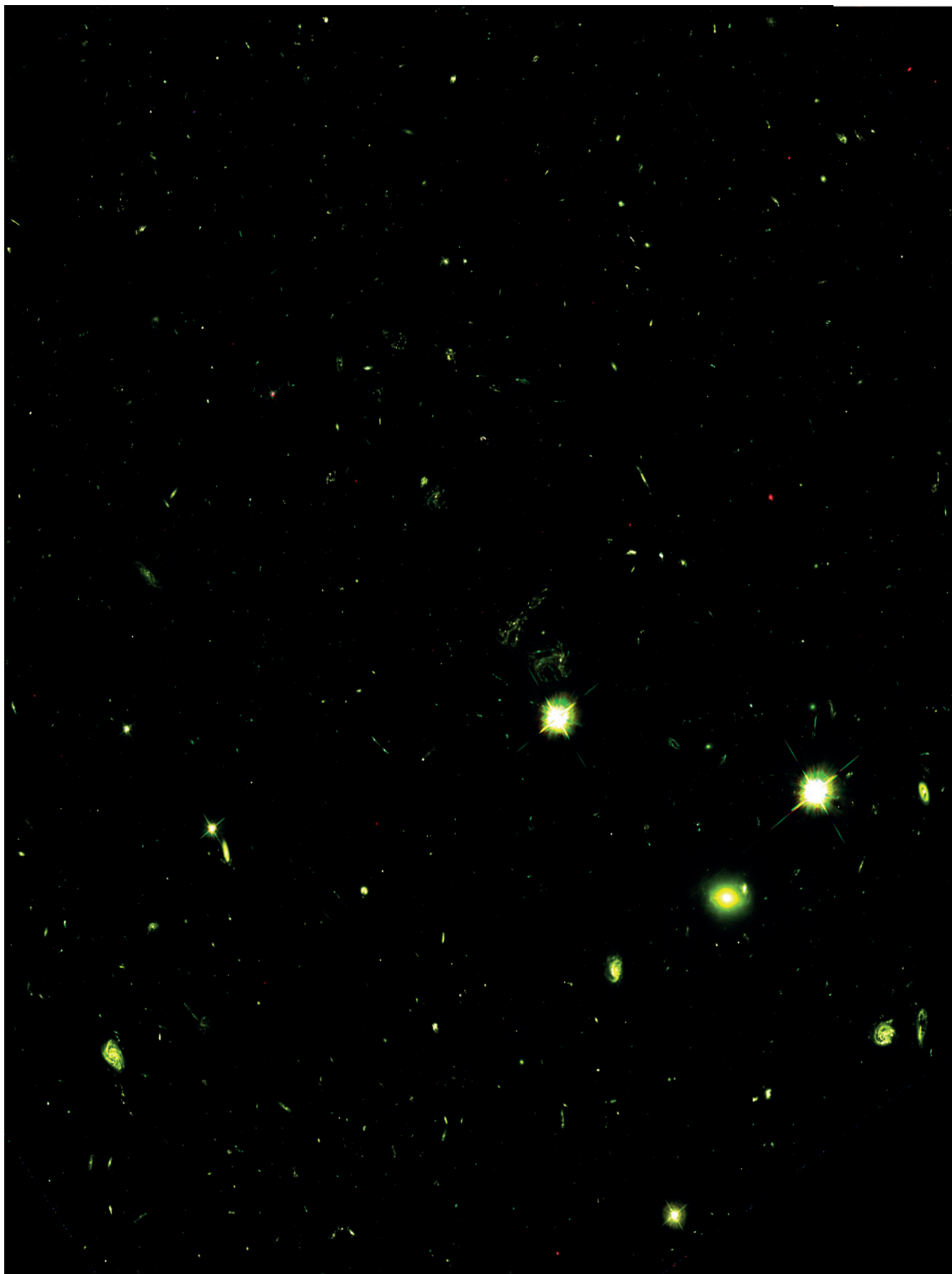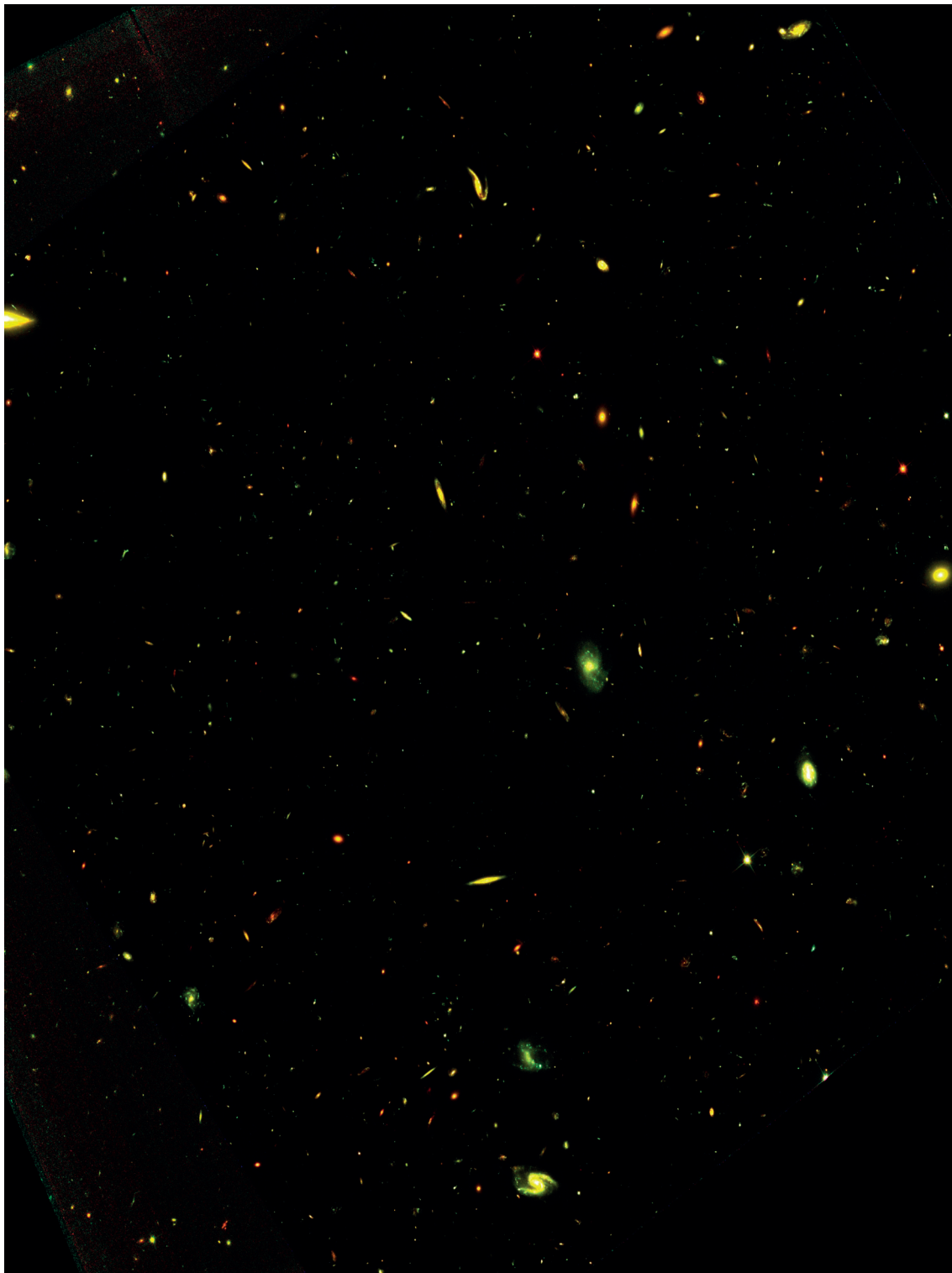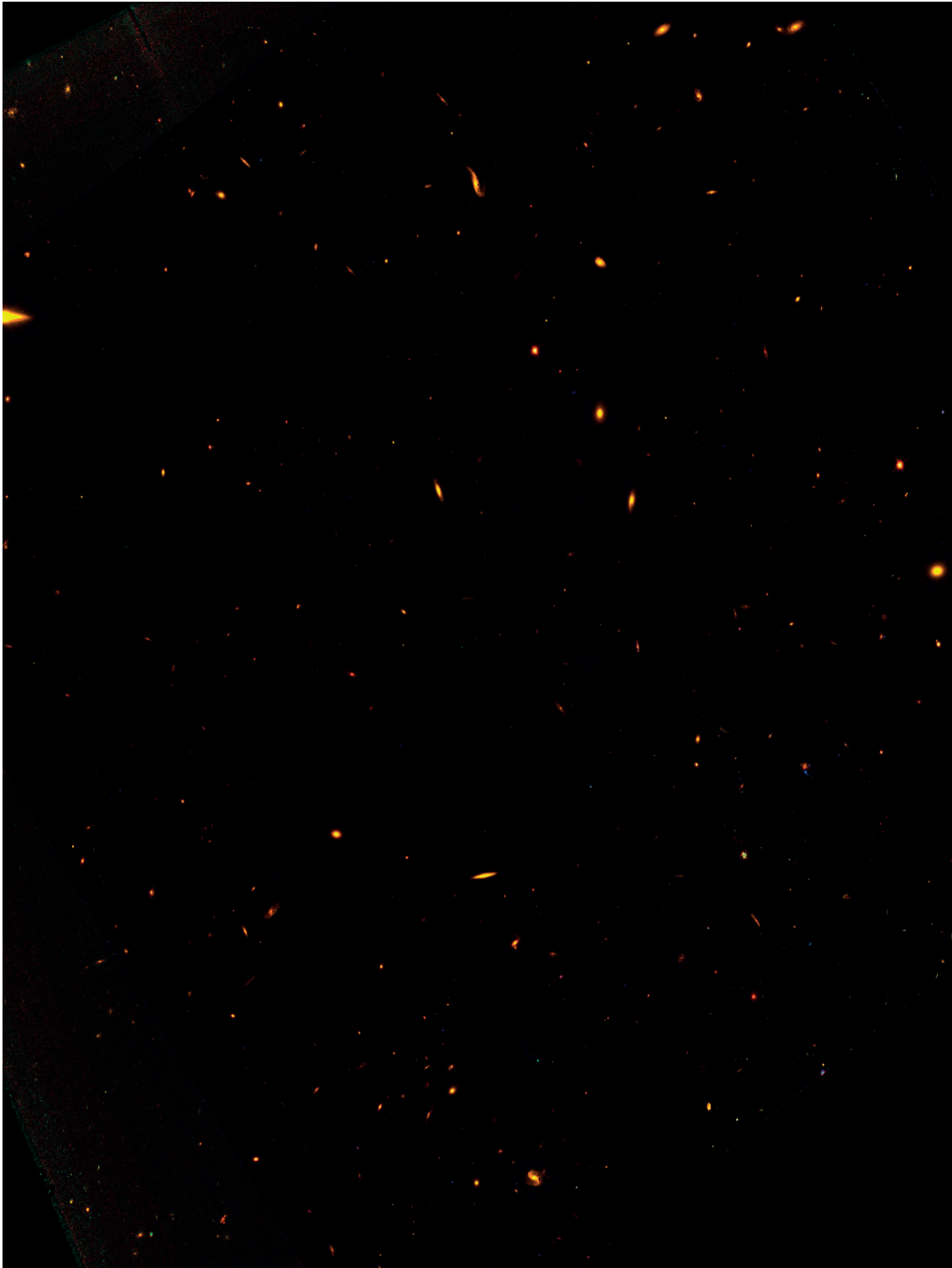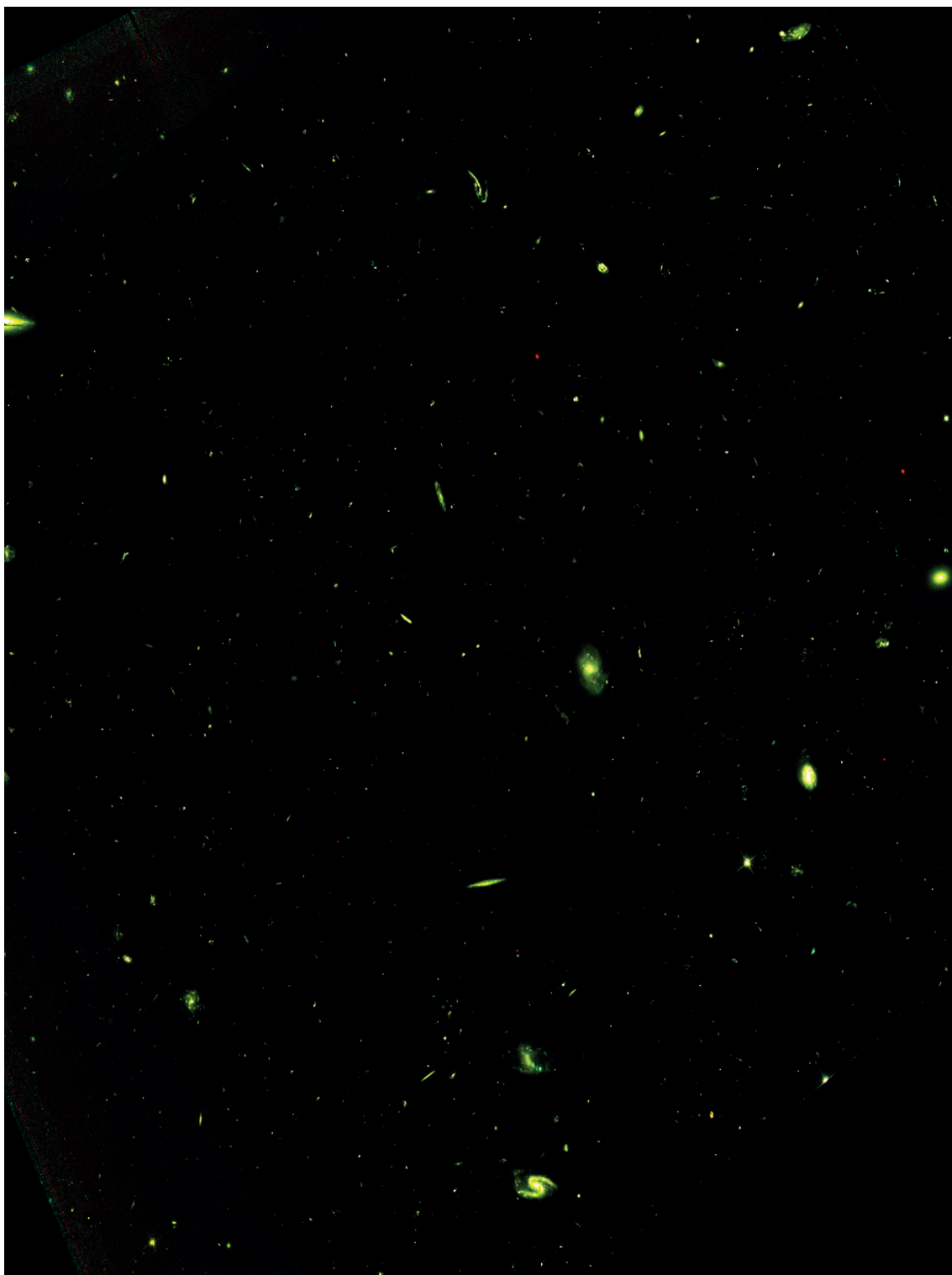
Figure 4.7 – Composite RGB image of parallel field MACS J0416.
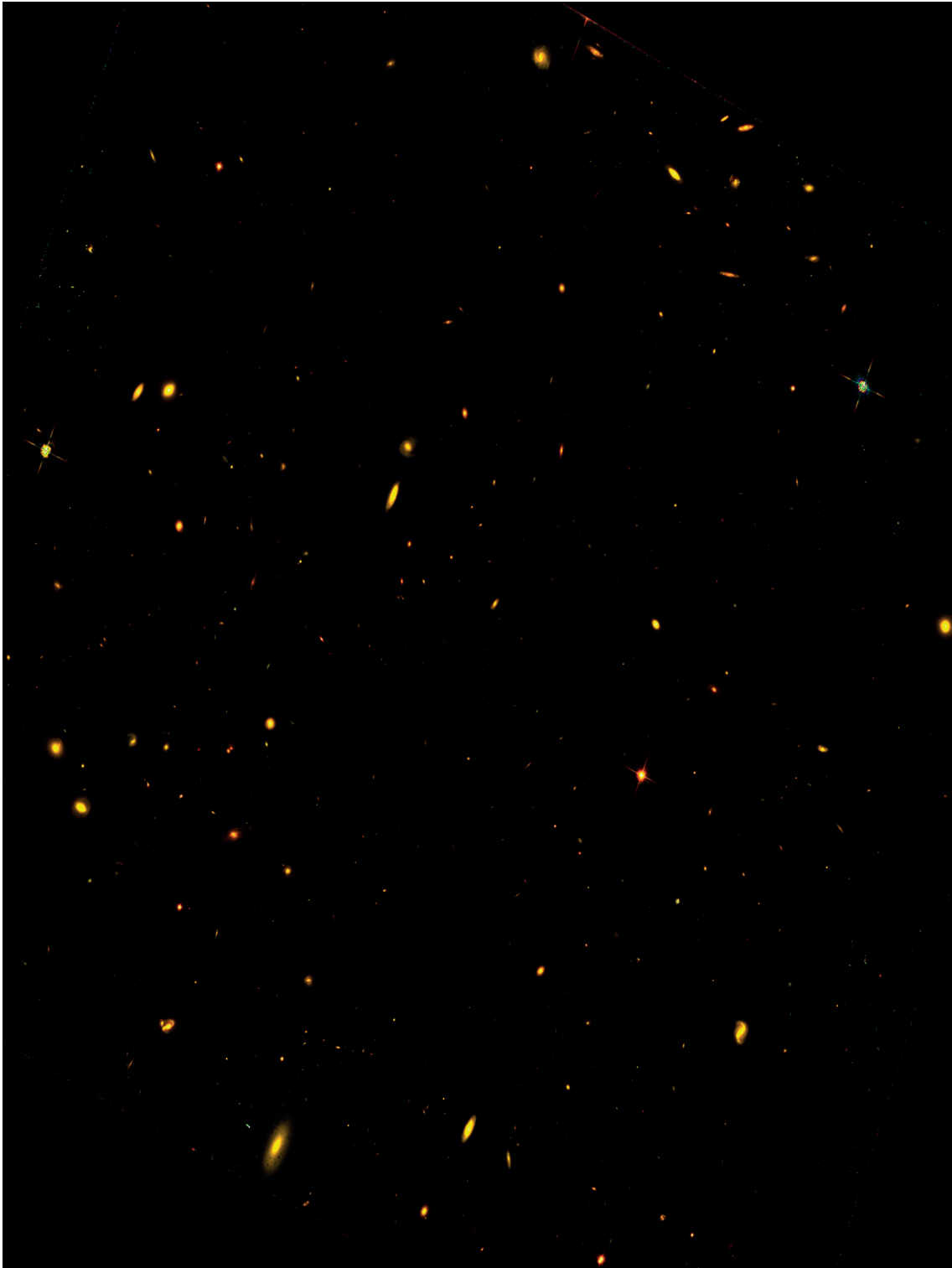
Figure 4.8 – Composite RGB image of the subtraction of the blue component from parallel field MACS J0416.
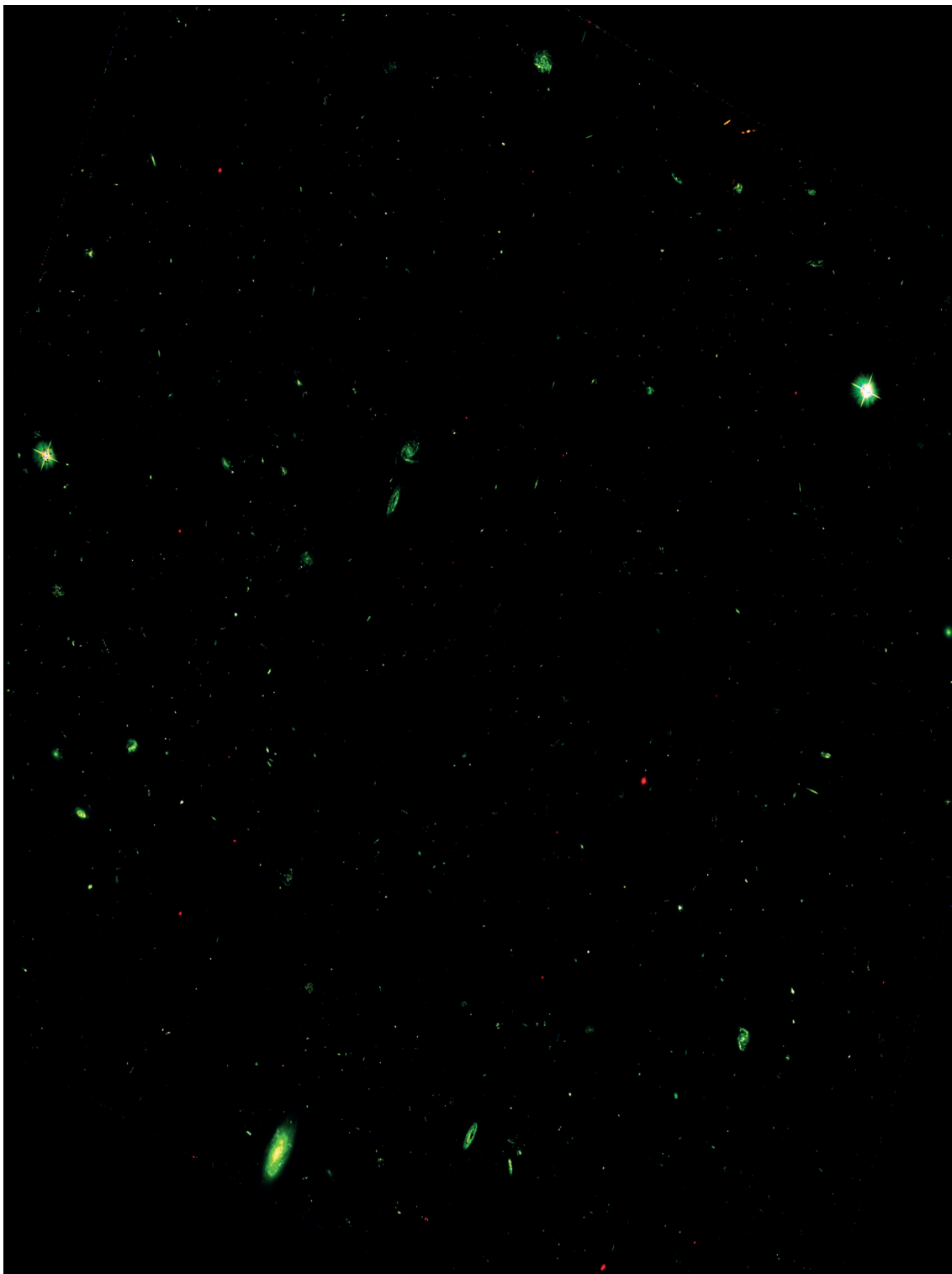
Figure 4.9 – Composite RGB image of the subtraction of the red component from parallel field MACS J0416.

MACS J0717 cluster
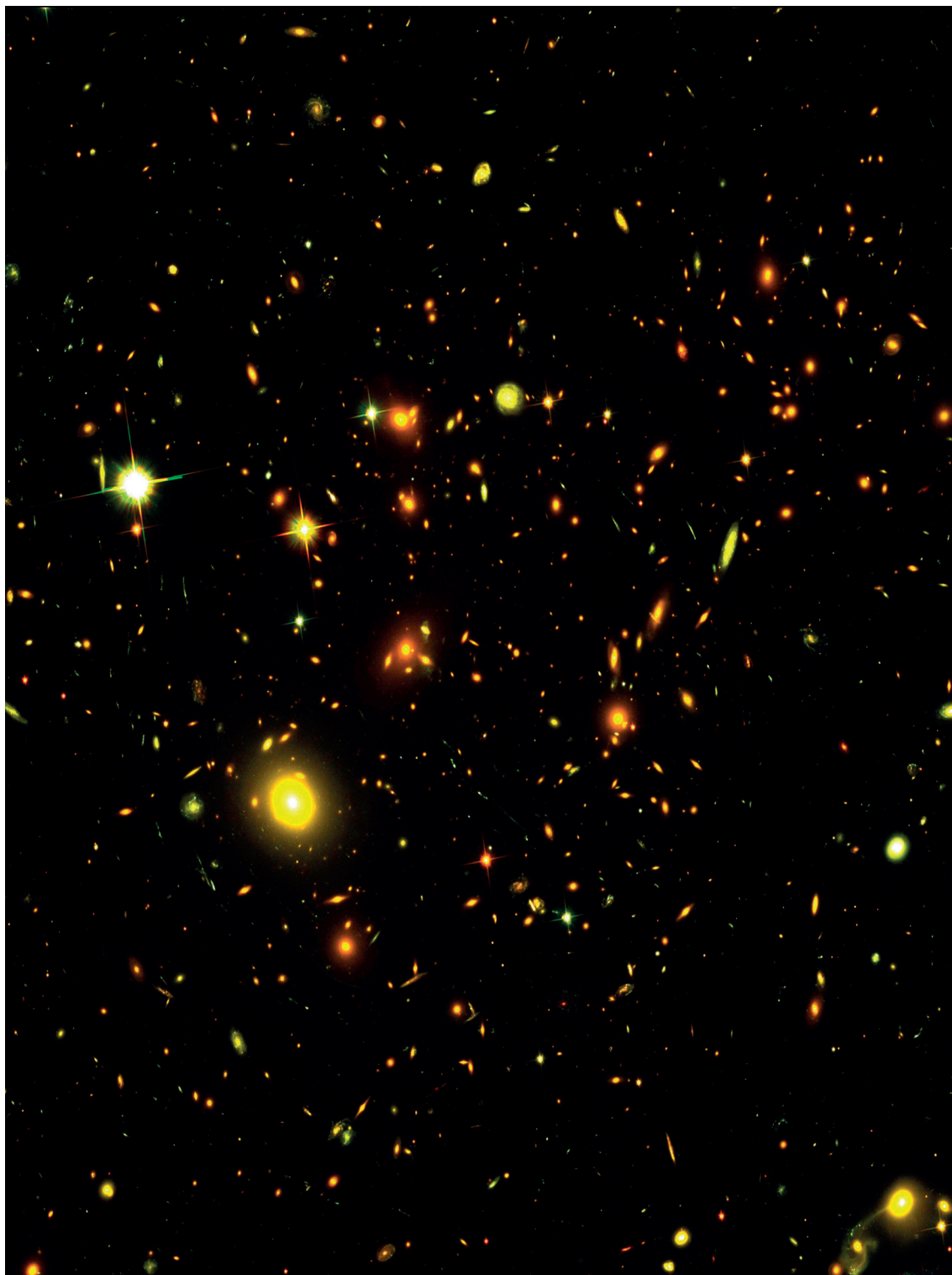
Figure 4.10 – Composite RGB image of cluster field MACS J0717.
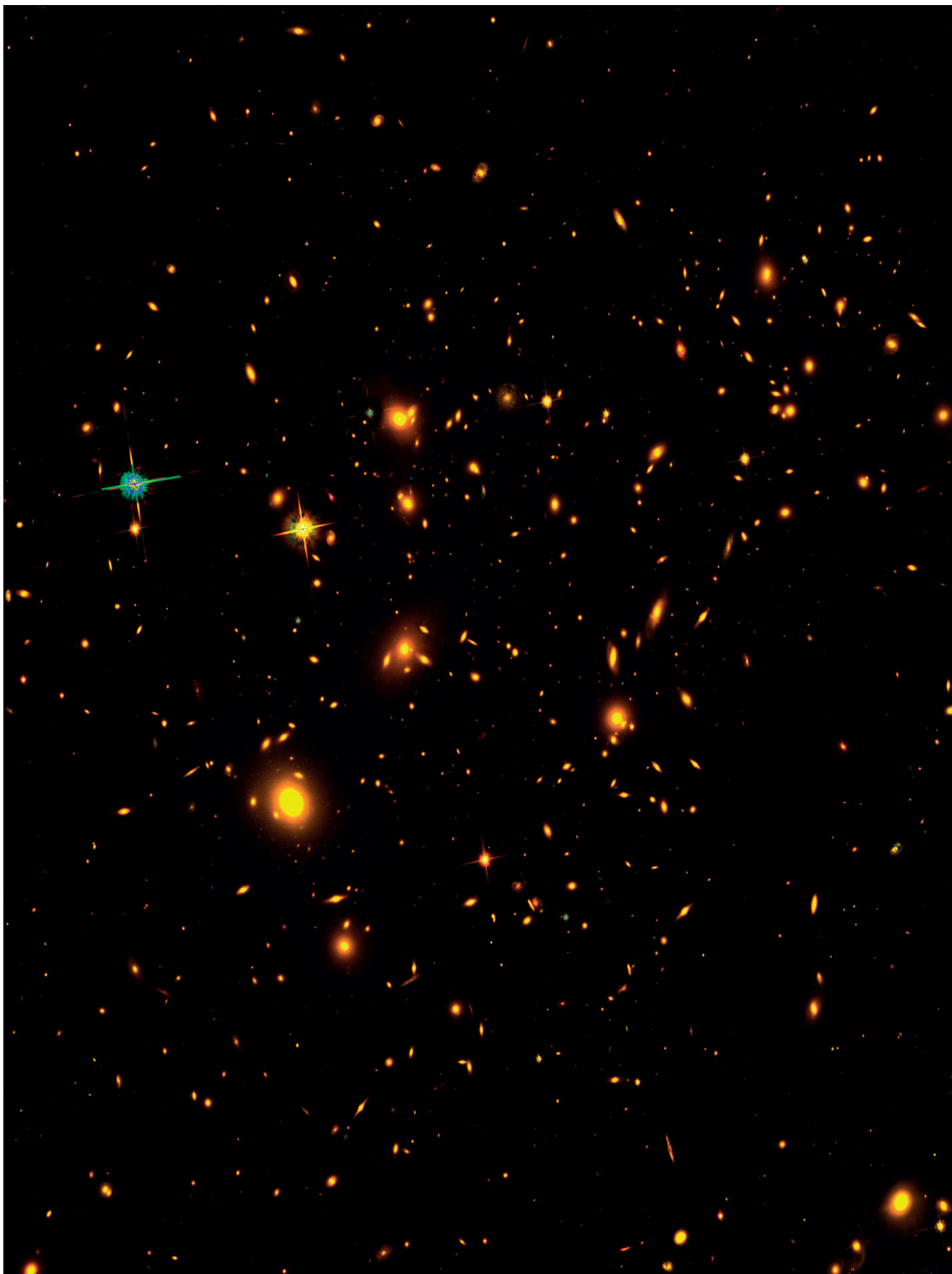
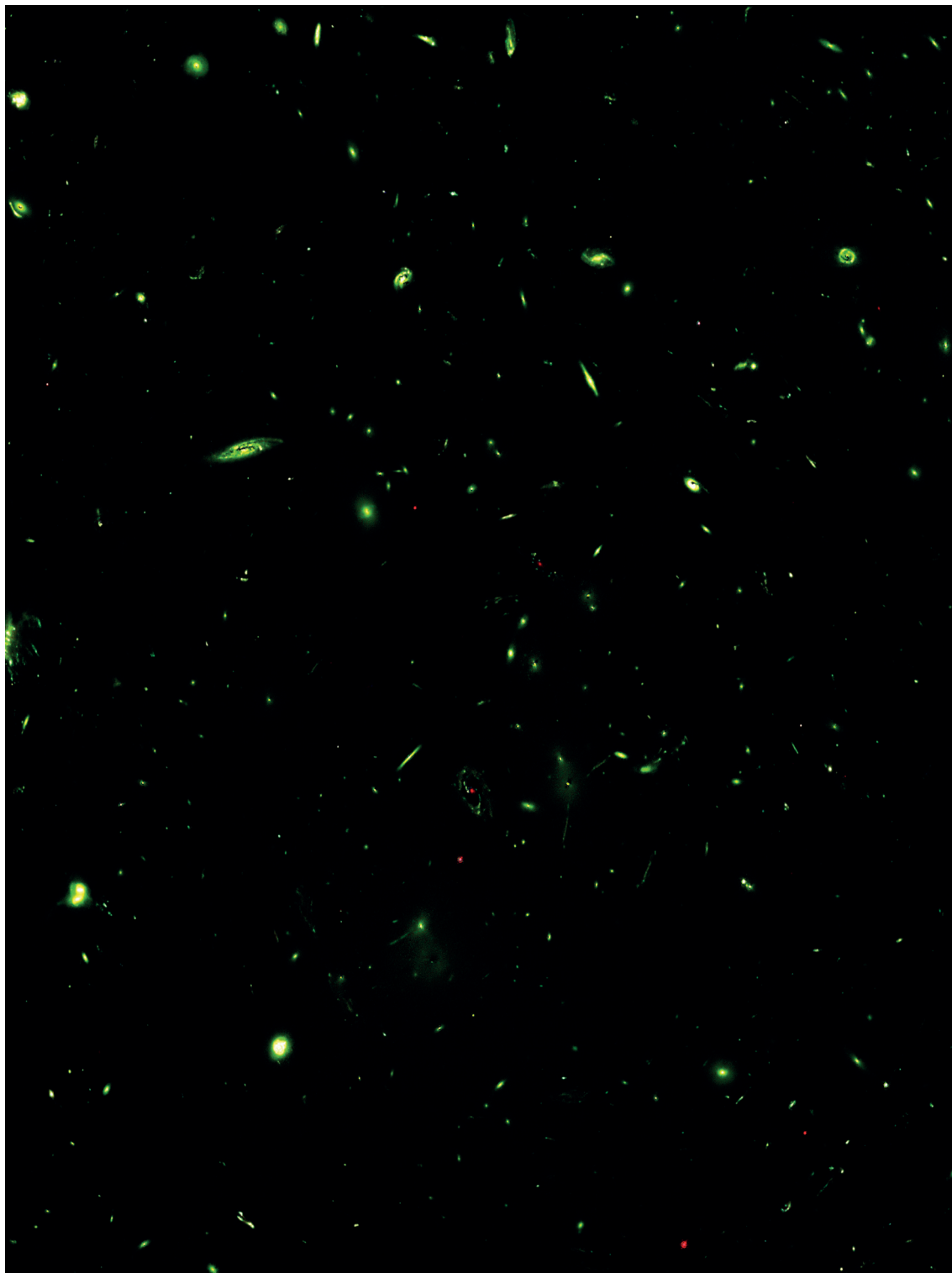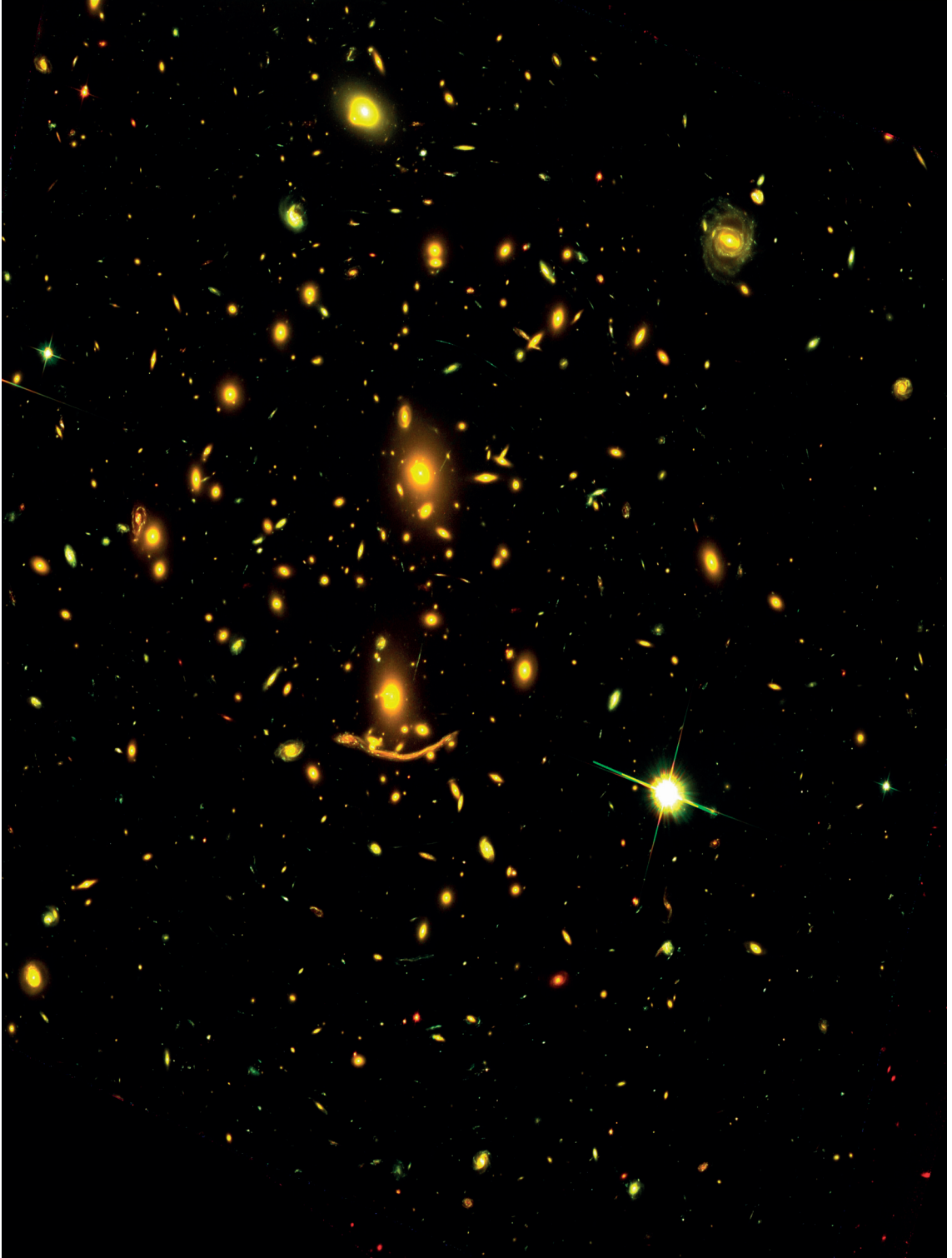Figure 4.11 – Composite RGB image of the subtraction of the blue component from cluster field MACS J0717.

Figure 4.12 – Composite RGB image of the subtraction of the red component from cluster field MACS J0717.

MACS J0717 parallel

Figure 4.13 – Composite RGB image of parallel field MACS J0717.

Figure 4.14 – Composite RGB image of the subtraction of the blue component from parallel field MACS J0717.

Figure 4.15 – Composite RGB image of the subtraction of the red component from parallel field MACS J0717.

MACS J1149 cluster

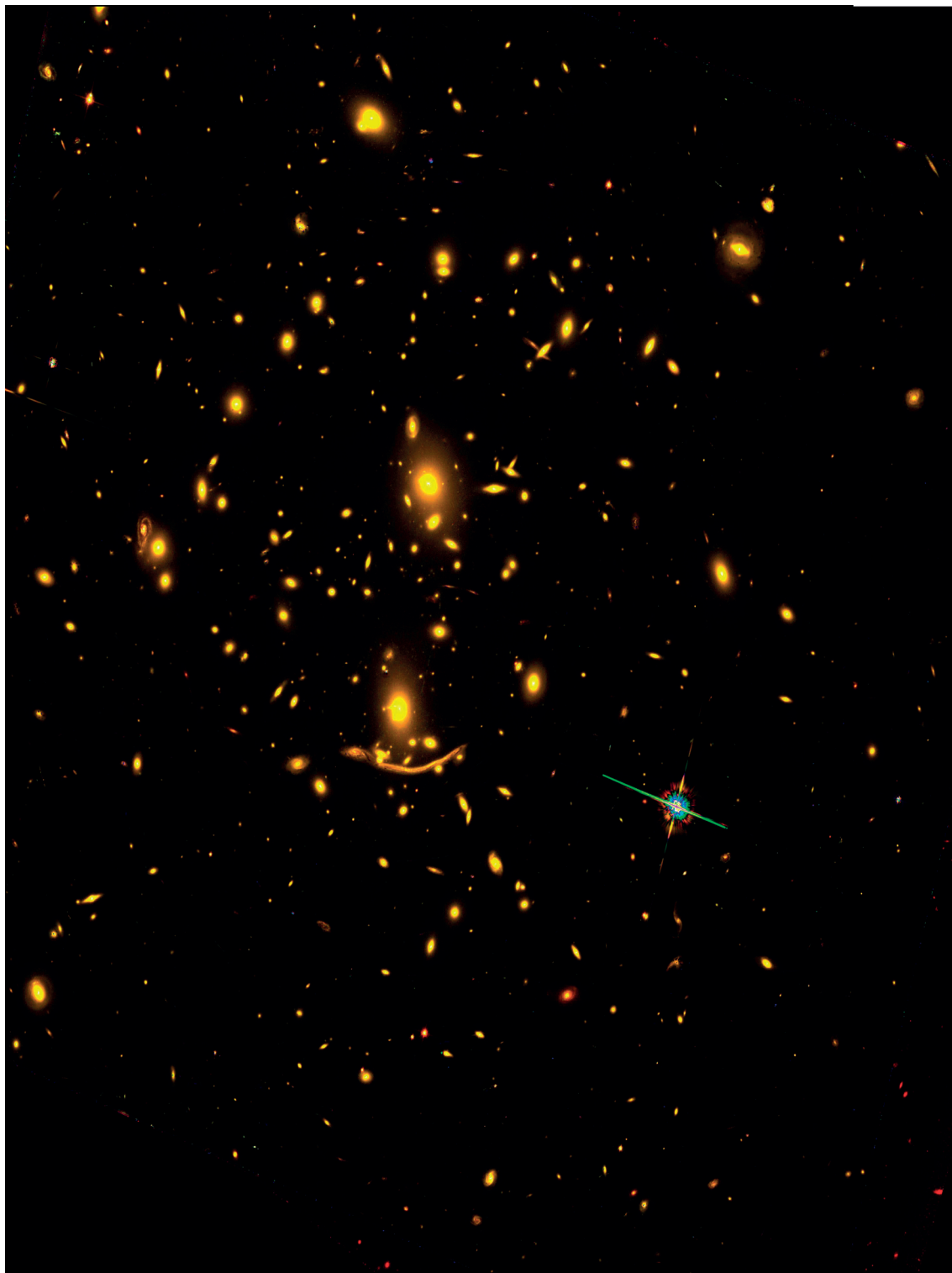Figure 4.16 – Composite RGB image of cluster field MACS J1149.

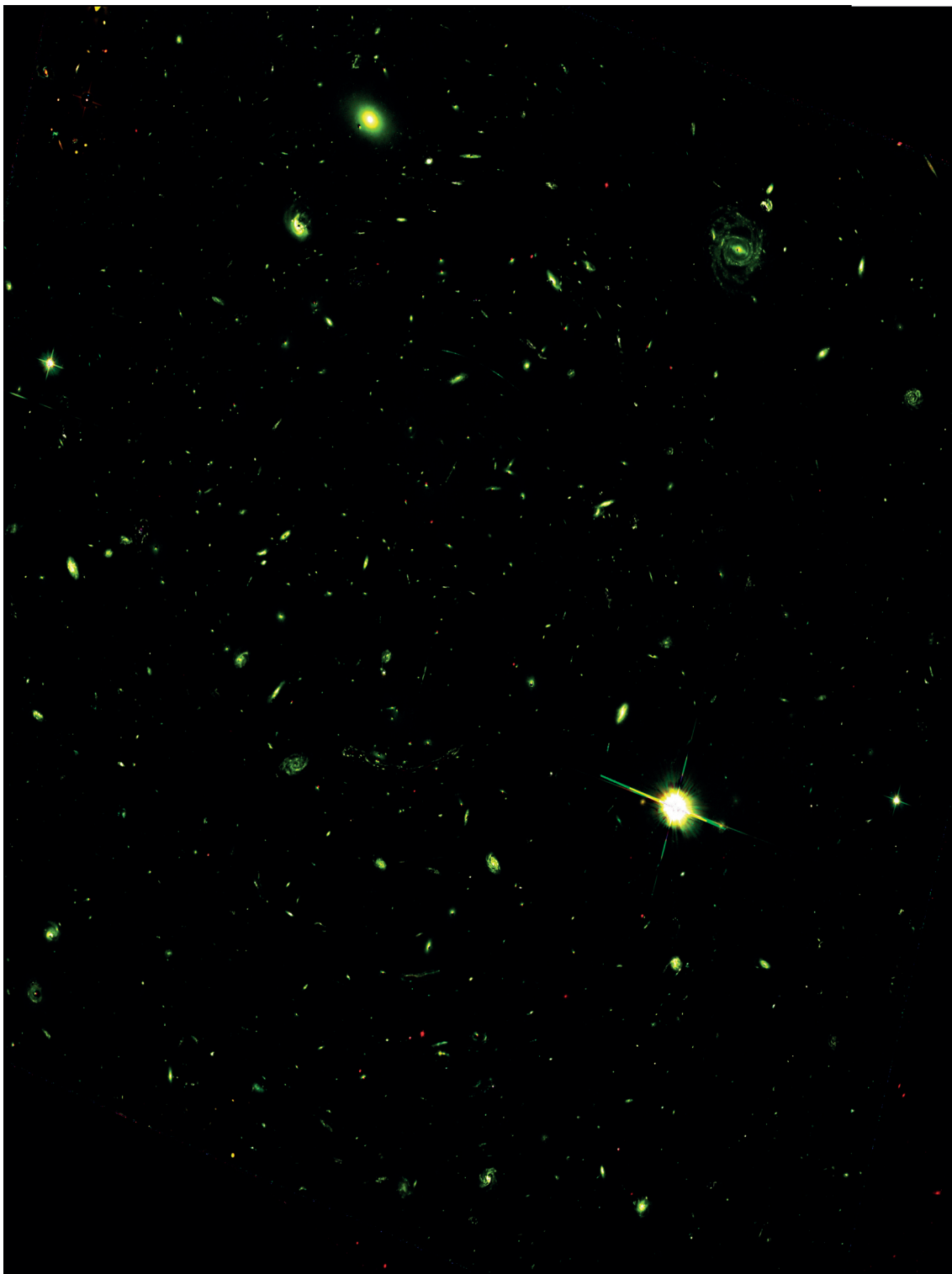Figure 4.17 – Composite RGB image of the subtraction of the blue component from cluster field MACS J1149.

Figure 4.18 – Composite RGB image of the subtraction of the red component from cluster field MACS J1149.

MACS J1149 parallel

Figure 4.19 – Composite RGB image of parallel field MACS J1149.

Figure 4.20 – Composite RGB image of the subtraction of the blue component from parallel field MACS J1149.

Figure 4.21 – Composite RGB image of the subtraction of the red component from parallel field MACS J1149.
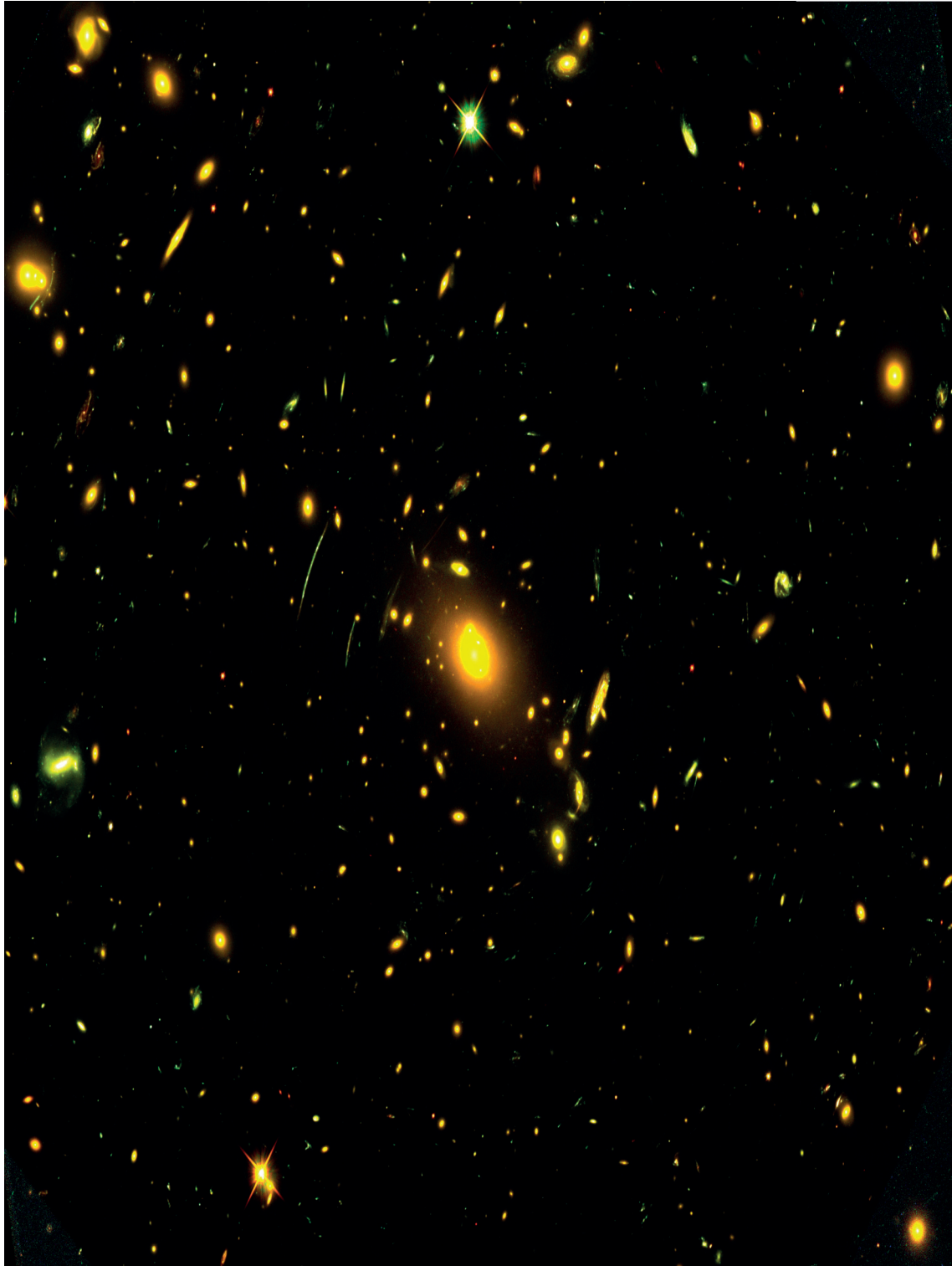
Abell 2744 cluster
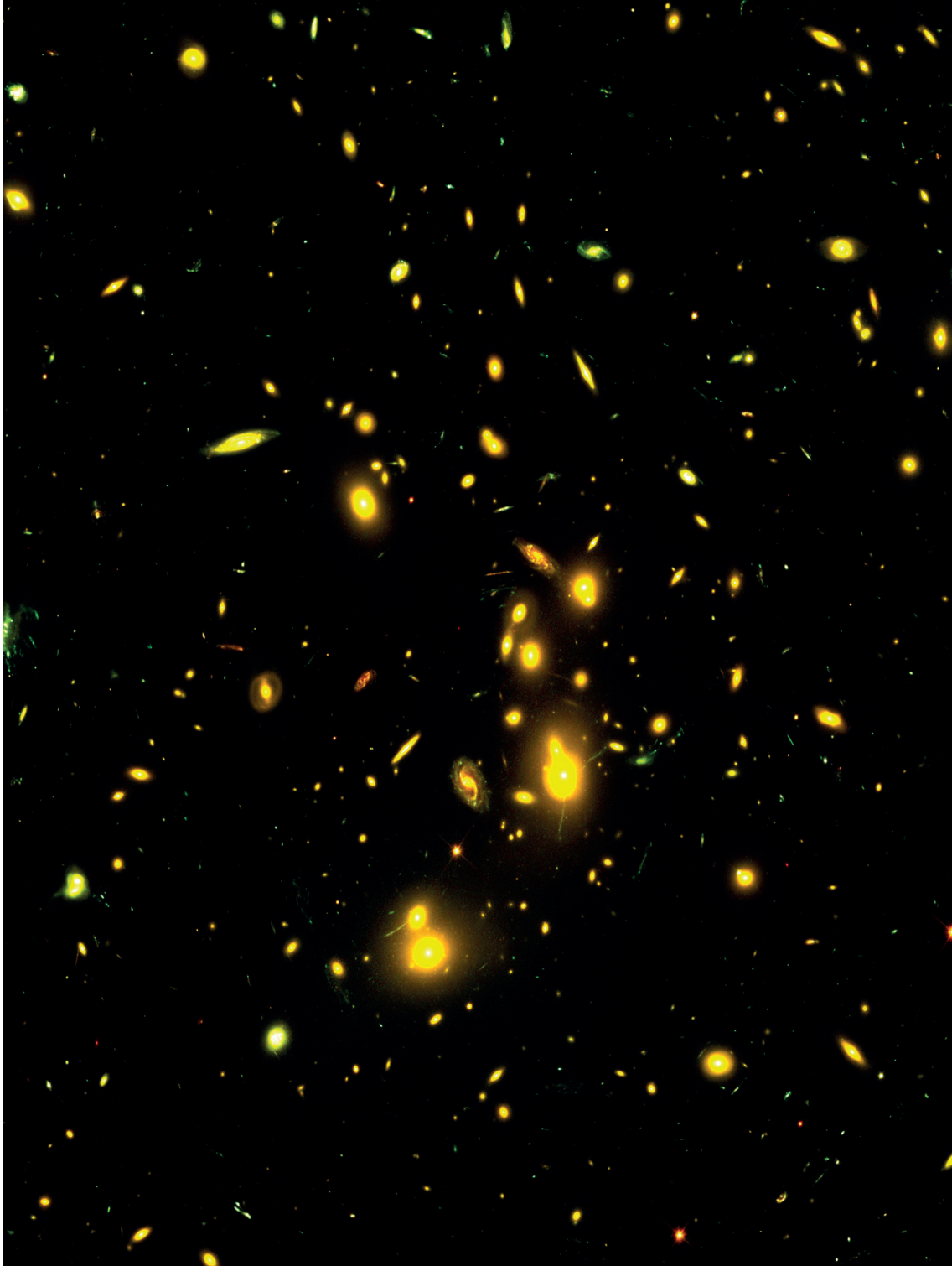
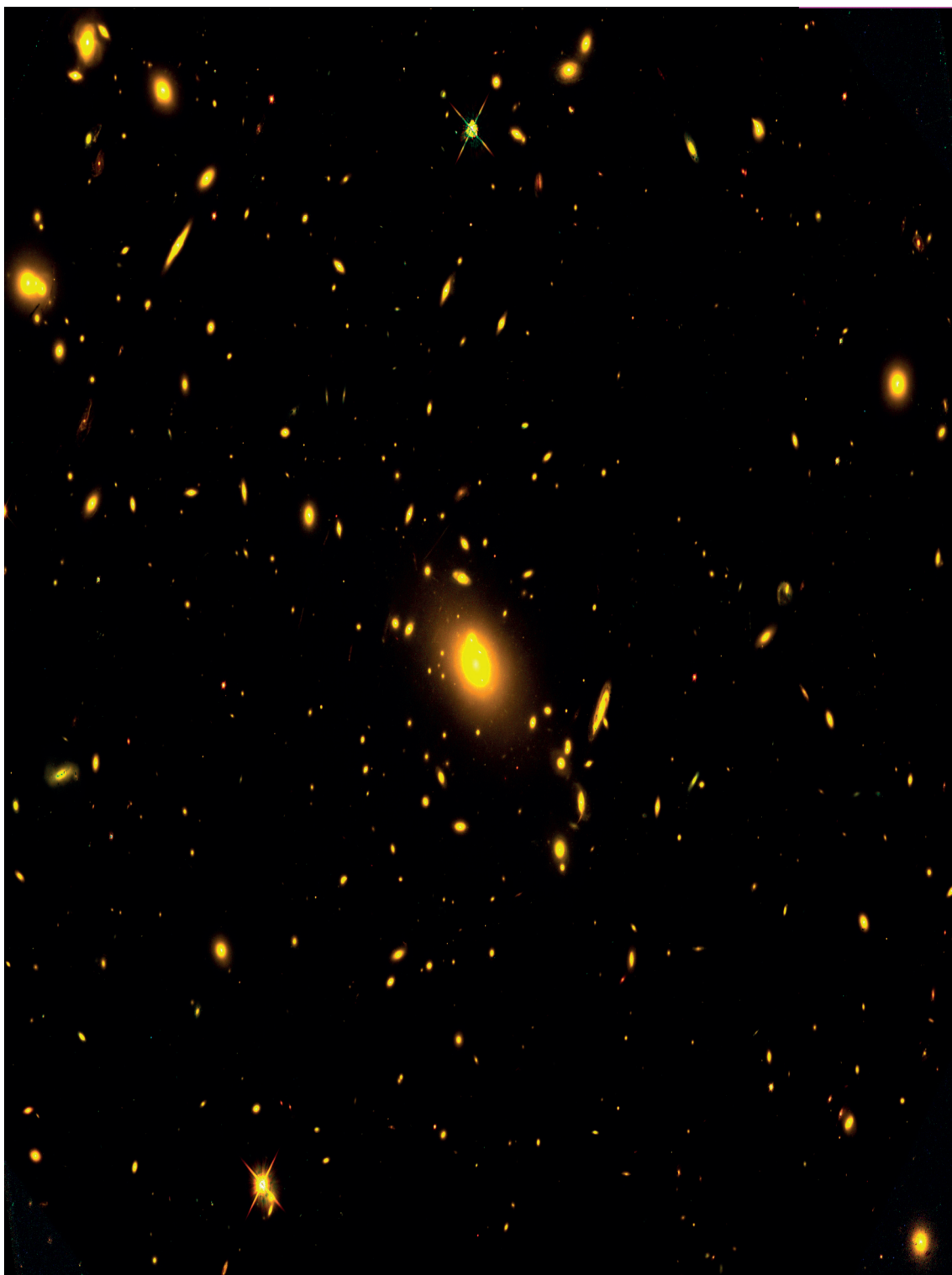Figure 4.22 – Composite RGB image of cluster field Abell 2744.
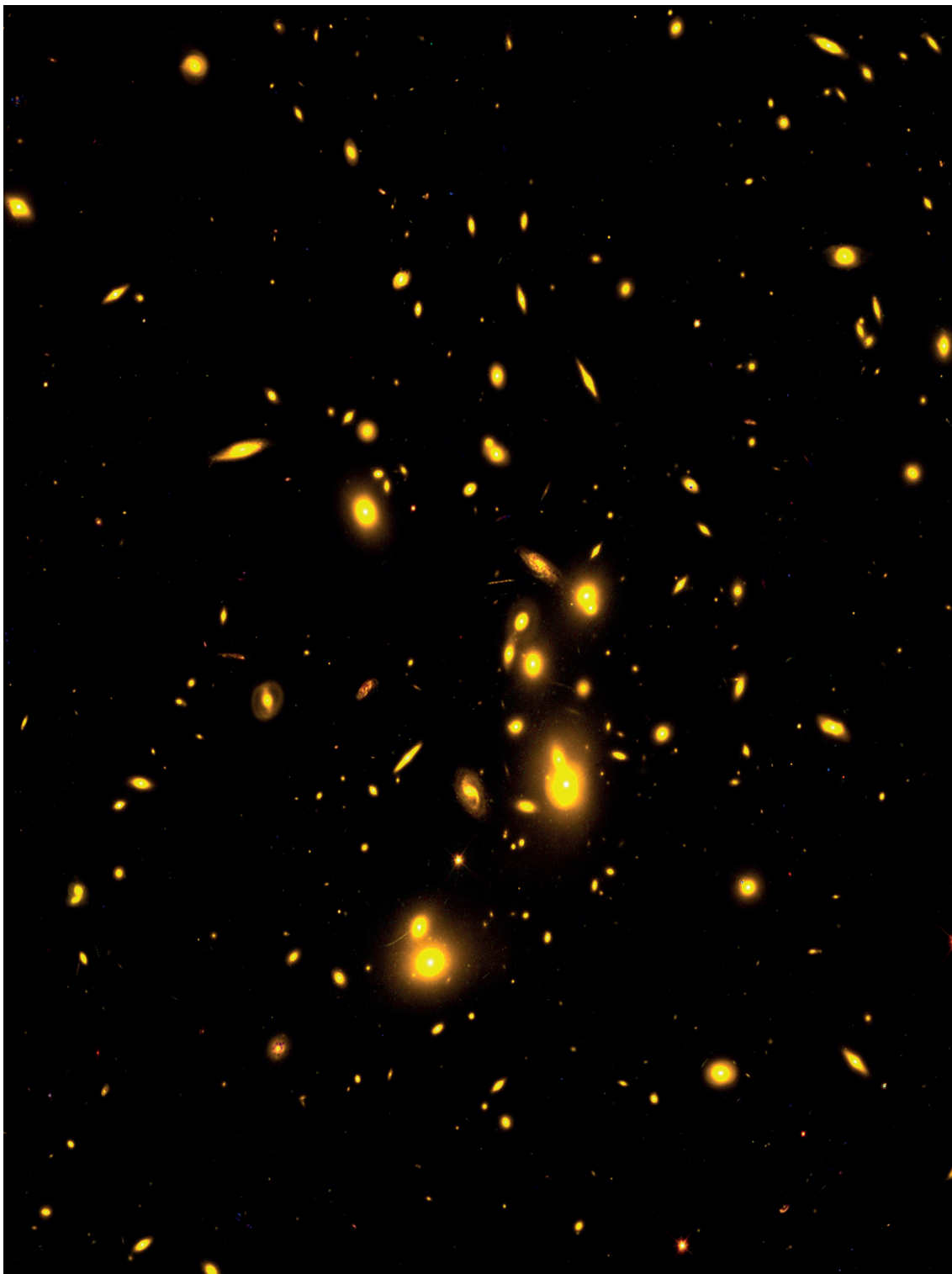
Figure 4.23 – Composite RGB image of the subtraction of the blue component from cluster field Abell 2744.

Figure 4.24 – Composite RGB image of the subtraction of the red component from cluster field Abell 2744.

Abell 370 cluster

Figure 4.25 – Composite RGB image of cluster field Abell 370.

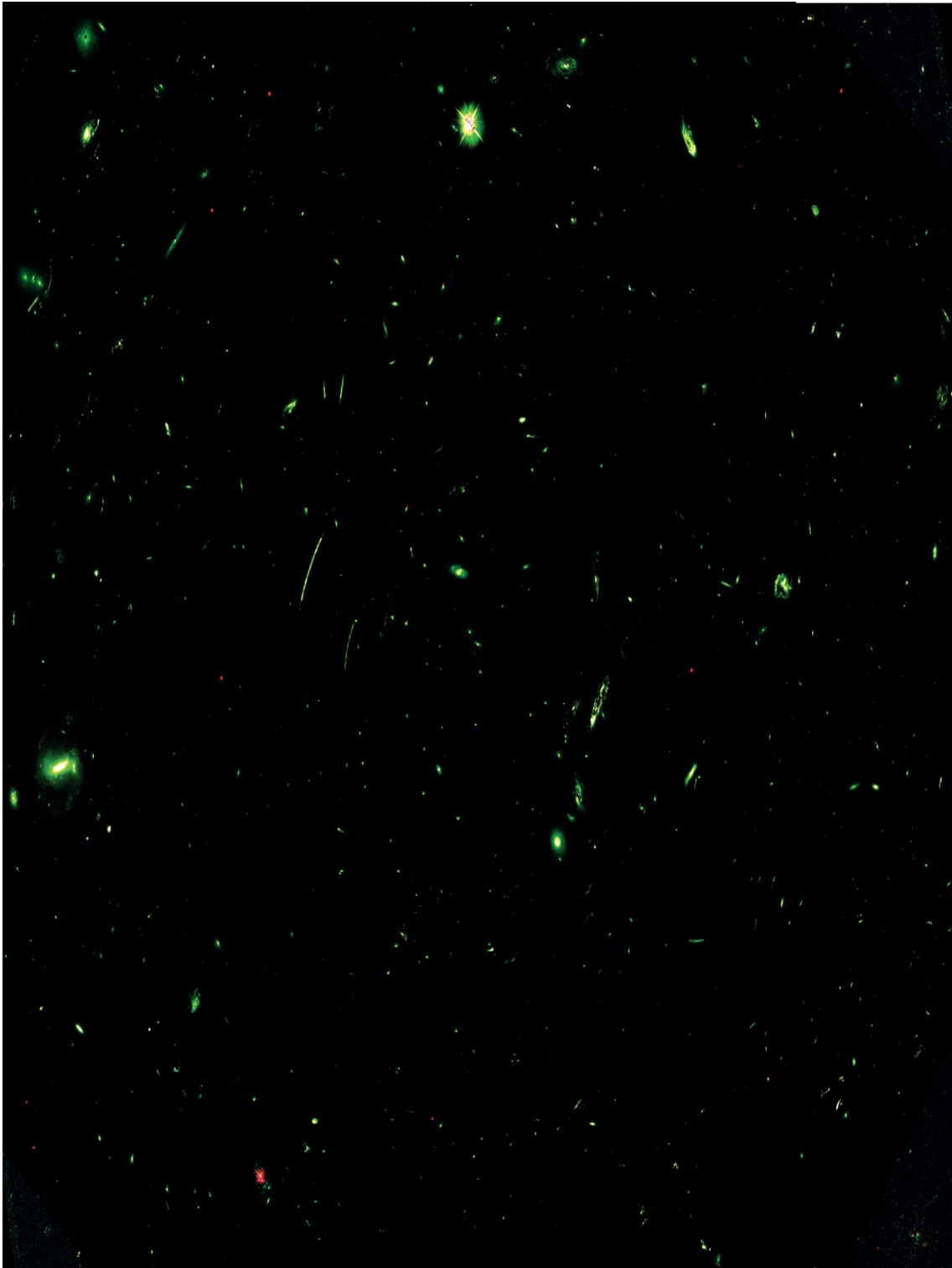Figure 4.26 – Composite RGB image of the subtraction of the blue component from cluster field Abell 370.

Figure 4.27 – Composite RGB image of the subtraction of the red component from cluster field Abell 370.

Abell S1063 cluster

Figure 4.28 – Composite RGB image of cluster field Abell S1063.

Figure 4.29 – Composite RGB image of the subtraction of the blue component from cluster field Abell S1063.

Figure 4.30 – Composite RGB image of the subtraction of the red component from cluster field Abell S1063.

4.4 Magnification in the HFF

With the ICL removed from the HFF, we see the background blue source in R_b appear with little to no contamination from the cluster light. In this section, I investigate how these subtractions can be used to measure the flux in images of multiply imaged objects. An accurate measure of the flux of lensed objects could potentially give access to the magnification ratio between images, which could, in turn, be used to constrain cluster mass models.

4.4.1 Magnification ratios

As we have seen in chapter 1, magnification is a direct consequence of lensing and depends on the deflection potential. At each location on the plane of the sky, except from critical lines, it is possible to compute the magnification:

$$\mu(\boldsymbol{\theta}) = \frac{1}{(1 - \kappa(\boldsymbol{\theta}))^2 - \gamma_1(\boldsymbol{\theta})^2 - \gamma_2(\boldsymbol{\theta})^2} \tag{4.3}$$

where κ, γ_1 and γ_2 can be expressed as derivatives of the gravitational potential. In a first order approximation, we can consider that a lensed source imaged at position $\boldsymbol{\theta}$ sees its flux multiplied by a factor $\mu(\boldsymbol{\theta})$. As a consequence, magnification is not an observable quantity as the observed flux form a lensed source F_L is the product:

$$F_L(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}) F_0, \tag{4.4}$$

where F_0 is the intrinsic flux of the background source. An alternative to measuring the magnification is to measure the magnification ratio between two images of the same source at positions $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. In this case, the magnification ratio r_μ is simply the ratio of the fluxes of two multiple images of a same source:

$$r_\mu = \frac{F_L(\boldsymbol{\theta}_1)}{F_L(\boldsymbol{\theta}_2)} = \frac{\mu(\boldsymbol{\theta}_1) F_0}{\mu(\boldsymbol{\theta}_2) F_0} = \frac{\mu(\boldsymbol{\theta}_1)}{\mu(\boldsymbol{\theta}_2)} \tag{4.5}$$

4.4.2 The *Ares* simulated cluster

To perform this study, I used the artificial images of the *Ares* galaxy cluster. *Ares* is a simulated galaxy cluster created for the purpose of the lens modelling challenge (Meneghetti et al., 2017). In this paper the authors simulated two clusters, *Ares* and *Hera*. Given that *Ares* has a higher convergence and was generated to form 242 images of 85 sources, while *Hera* only contains 65 images of 19 sources, I chose to focus on the study of the *Ares* cluster. The goal of this study is to see how well one can recover the magnification ratios between multiple images of a cluster and to see if measuring fluxes from MuSCADeT's R_b image could bring better estimates of the flux ratios. The simulations are emulations of the HFF survey. As a result, we could conduct the same separation on *Ares* as what we did with all the HFF images. We used the simulated images for bands F814w, F606w and F435w to separate between red and blue components of

Ares. The results are shown in appendix E.

Using a simulated cluster allows to compare the computed magnification ratios to the actual ratios, and to compare the result of the subtraction of the foreground ICL to the lensed images without ICL as they are generated in the simulation. To compute the magnification ratios, we extract postage stamps of 40 pixels on-a-side of the lensed sources based on the catalogs provided by Meneghetti et al. (2017). For each multiply-imaged source, three postage stamps are extracted and we compute the flux for each postage stamp as the sum of the pixel intensities over all the pixels in the stamp. The three stamps of interest are as follows:

- P_0 is the postage stamp of an image taken in the raw image in 3 bands (Y),

- P_{blue} is the postage stamp of the raw image in three bands after subtraction of the red components (R_b),

- P_{sl} is the postage stamp provided by the simulation team of the lensed images of the multiply imaged objects seen in three bands (these include noise). These postage stamps are the groundtruth for the surface brightness of the lensed images.

Each postage stamp is taken in the three HST bands in order to maximise the signal to noise of the flux measurement. We recently noticed that the F606w band for stamps P_{sl} was in fact exactly the same as the F435w band that we were provided with, which should, strictly speaking, not be the same. As a result, in the rest of this analysis, we discard completely band F606w and will perform the analysis again with the three bands once we receive the corrected F606w image.

4.4.3 Comparing magnification ratios

Sanity check on the flux errors

Before computing the magnification ratios, I verify that the result of `MuSCADeT`'s subtraction allows a better estimate of the flux of the lensed sources. For each stamp, I compute the relative flux errors:

$$err_{P_R} = \frac{|(f(P_R) - f(P_{sl}))|}{f P_{sl}} \tag{4.6}$$

$$err_{P_{blue}} = \frac{|(f(P_{blue}) - f(P_{sl}))|}{f P_{sl}}, \tag{4.7}$$

where $f(P)$ is the flux in a postage stamp P. The mean error for stamps P_R is $m_R = 0.439$ with a standard deviation $\sigma_R = 28.9$, while for the `MuSCADeT` subtracted stamps, the mean error is $m_{blue} = -0.583$ and the standard deviation of the error is $\sigma_{blue} = 8.58$. While both mean errors are similar, the standard deviation of the error is much larger that in the case of the

original image than in the image where the ICL has been subtracted, which tends to show that `MuSCADeT` successfully reduced the contamination by the flux by foreground galaxies. I bring to the reader's attention that here, the fluxes were computed from the noisy stamps provided by the simulation, even for P_{sl}. This can explain the somewhat large dispersion of the error on the fluxes even for the `MuSCADeT` subtracted images. In the next section, we compare our measured magnification ratios to the actual magnification ratio provided by the simulation.

Measures of flux ratio

Due to the simulated cluster *Ares* having been generated for a lens modelling challenge, all the lensing properties were known by the simulation team. As a result, thanks to the authors of Meneghetti et al. (2017), I was able to use the position and the matching (knowledge of the source from which each image formed) of multiply imaged sources as well as the magnification map of the cluster.

I computed the actual magnification of each image of the lensed sources by recording the value of the magnification at the location given by the multiple images catalogue in the cluster's magnification map. For each background source, the first lensed image in the catalogue was used as a reference image from which the magnification ratios were computed. For instance, for source i with 4 multiple images, $\{I_{i.1}, I_{i.2}, I_{i.3}, I_{i.4}\}$ with respective magnifications $\{\mu_{i.1}, \mu_{i.2}, \mu_{i.3}, \mu_{i.4}\}$, the magnification ratios are computed as:

$$r_{\mu_2} = \frac{\mu_{i.2}}{\mu_{i.1}} \tag{4.8}$$

$$r_{\mu_3} = \frac{\mu_{i.3}}{\mu_{i.1}} \tag{4.9}$$

$$r_{\mu_4} = \frac{\mu_{i.4}}{\mu_{i.1}} \tag{4.10}$$

Empirical magnification ratios were also computed from the flux ratios of multiply imaged sources by taking the same reference image for each source. Three empirical magnification ratios are computed here:

- the "raw magnification ratio" r_{μ_R} computed directly from the non-processed stamps P_R,

- the "blue magnification ratio" r_{μ_B} computed from the stamps where the red component was removed by way of `MuSCADeT` P_{blue}

- the "simulated magnification ratio" r_{μ_S} computed from the simulated images of the lensed sources only.

The ground-truth magnification ratios computed from the magnification map are noted r_{μ_0}. In figure 4.31, we show the results of this study in terms of magnification error. The

absolute difference between true and empirical magnification ratios is shown against the true magnification of the non-reference image used in the ratio. Each point shows the averaged error and dispersion over a bin of 2 in magnification. It appears that at large magnifications, $\mu > 30$, the magnification ratio error increases. This is due to the difference between the ways empirical and true magnification ratios are computed. The true magnification ratio is computed by taking the value of the magnification at the center of the magnified image, while empirical magnification ratios are computed by taking the flux ratios of two images, which assumes that a magnification is constant over the whole surface of the image surface brightness. Around critical lines, the magnification profile can take a high slope, making the assumption of constant magnification unrealistic. In these cases, the averaged magnification can be smaller than the actual magnification, resulting in errors in the magnification ratio. We also notice that their are no error-bars for several points at large magnification. This is due to the rarity of sources at large magnifications, where, in most cases, only one source is found per bin of two magnifications. The lower panel of figure 4.31 focuses on the low magnification region of the plot.

Another observation from the error bars of figure 4.31 is that, as expected, the magnification ratio errors seem to be much smaller for the deblended images than for the original ones. However, we also notice that, in some cases, the error bars are smaller for the deblended images than for the simulated images of lensed sources only. This should not be the case due to the simulated images of lensed sources being the closest observations from the truth. Also, errors should be larger in `MuSCADeT` due to the possibility of background sources to be red [4] and therefore be removed by `MuSCADeT`, and to the possibility for blue components from cluster member galaxies to pollute the deblended images of background galaxies. This will have to be investigated further before any publication of the results. Among the few possibilities that I have considered so far are the discarding of band F606w, small number statistics, error on the measurement of the flux due to very low SNR and elongated sources, errors on the computation of fluxes from deblended images due to stretched lensed images being poorly represented by `MuSCADeT`. At this point I do not see how any of these effects could explain the systematically smaller errorbars the deblended images, but the careful inspection of the multiply imaged systems with these sources of error in mind might help finding an explanation.

A summary of the distribution of magnification errors is given in figure 4.32. Each histogram shows the number of multiple images systems per bin of magnification ratio error. The distribution was truncated at small magnification ratio errors ($|r_{\mu_0} - r_\mu| < 10$) for the sake of representation. The distributions in these histograms are as expected: the distributions of errors in the deblended and the simulated lensed images are similarly peaked around 0. This observations hints at the possibility that the previous discrepancy be explained by small number statistics. The distribution of errors on magnification from the raw image is flatter than the other distributions and shifted towards positive ratio errors. The flattening

---

[4]While background lensed galaxies are expected to be mostly blue, very high redshift galaxies and dust-obscured galaxies can be a source of red background galaxies
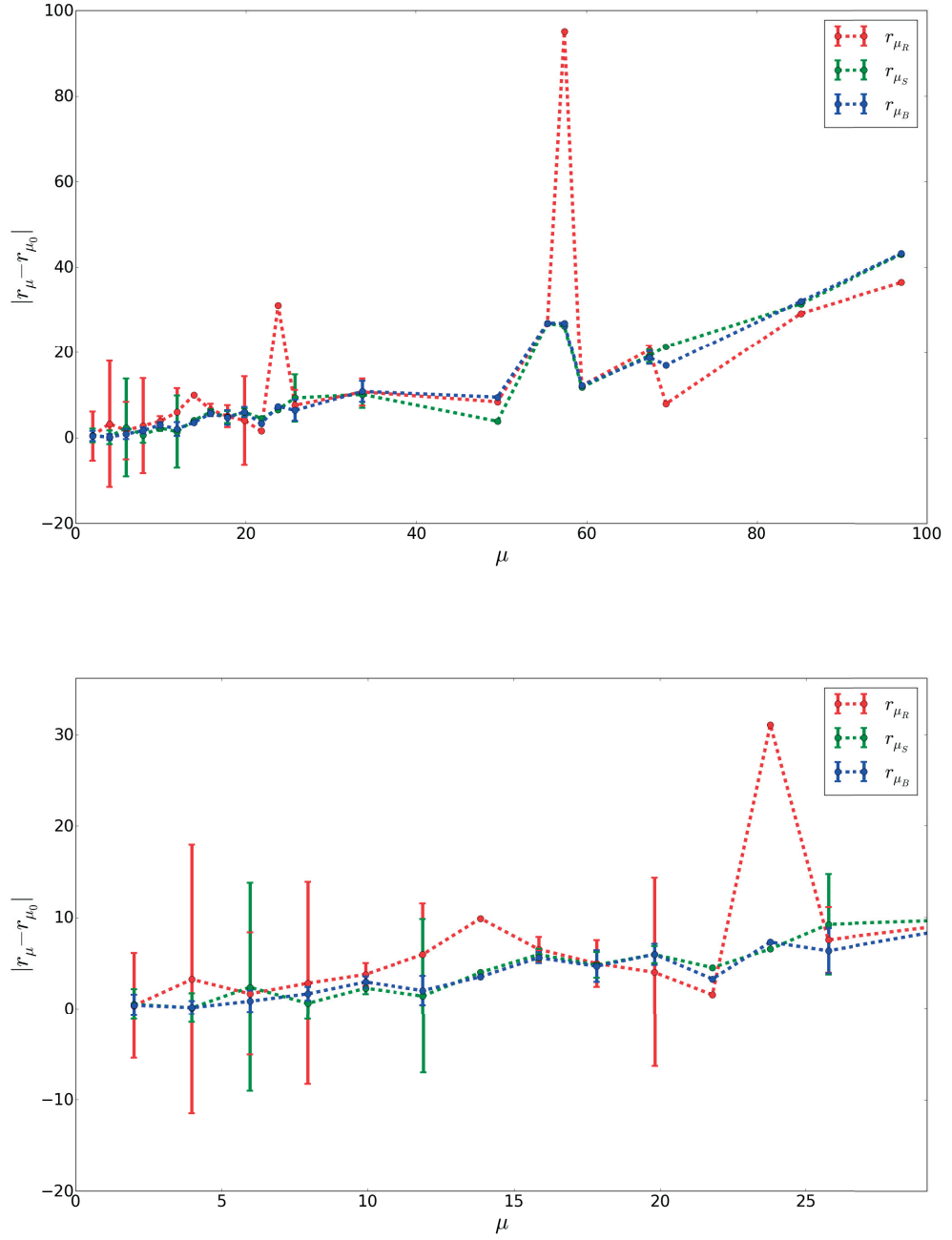
Figure 4.31 – Binned error on magnification ratios as a function of magnification.

illustrates the lesser accuracy of the measurement due to the ICL. The shift towards positive errors can be explained by the fact that the light from foreground galaxies takes precedence in the contribution to the measured flux, thus smoothing the magnification ratios towards smaller values.
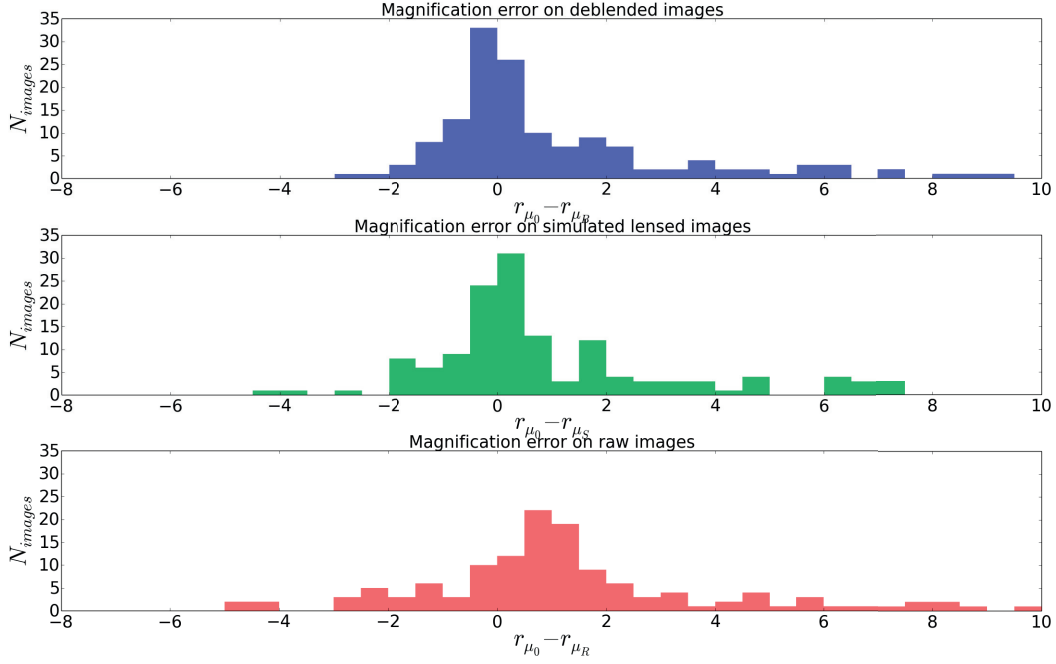
Figure 4.32 – Histograms of the magnification error distributions for the three empirical magnification ratio measurements.

Application to real data

With the help of my grad. student, we applied our method to galaxy cluster MACS J0416, for which the deblending is shown in figure4.6 and 4.5, and compared our measured magnifications to the predictions made be different lens modelling teams. The details of the various models can be found in Priewe et al. (2017) and references therein. We used the positions of multiple images identified in Jauzac et al. (2014) to compute our empirical magnification ratios from `MuSCADeT`'s deblended images. The authors of Jauzac et al. (2014) report 69 lensed sources forming a total of 194 lensed images, out of which 26 have a spectroscopic redshift. When available, the spectroscopic redshifts were used, but in other cases, we relied on the redshifts predicted by Jauzac et al. (2014), for consistency, to compute the magnifications predicted by other models. Magnification maps by modellers were computed using the online tool https://archive.stsci.edu/prepds/frontier/lensmodels/[5].

The empirical magnification ratios from `MuSCADeT` were compared with those obtained by the various modellers by computing the correlation coefficients between them. The correlation coefficient between vector r_{μ_B} containing the magnification ratios from the deblended images
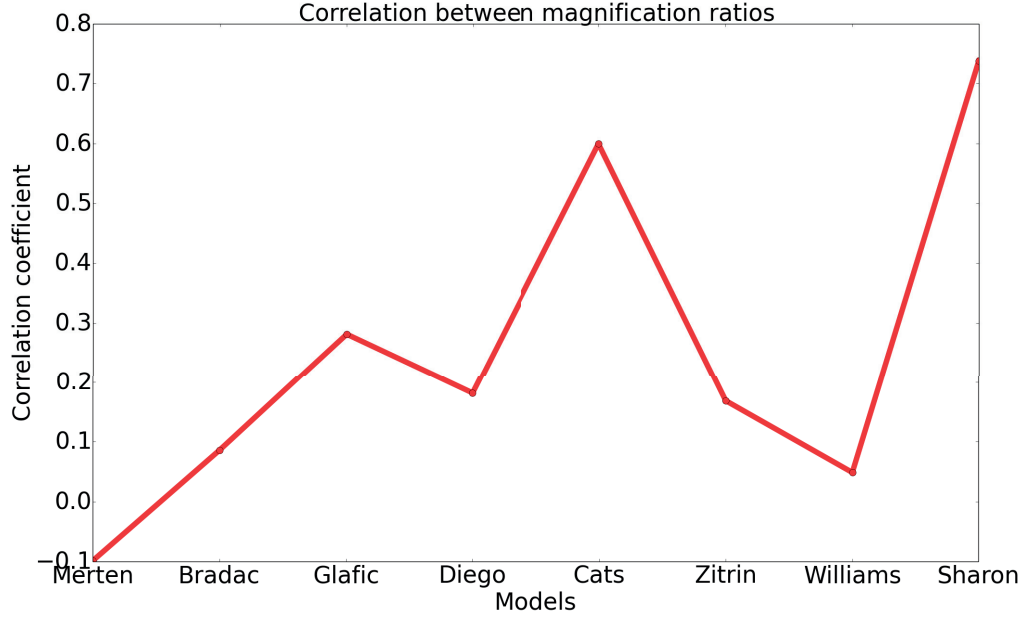
Figure 4.33 – Correlation coefficient between various predictions for magnification ratios in MACS J0416 and empirical values derived from `MuSCADeT`.

and vector r_{μ_M} containing the magnification ratios derived for a given model M is given by:

$$C_{BM} = \frac{cov(r_{\mu_B}, r_{\mu_M})}{\sqrt{cov(r_{\mu_M}, r_{\mu_M}) cov(r_{\mu_B}, r_{\mu_B})}}, \tag{4.11}$$

where cov is the covariance operator.

In figure 4.33 we show the correlation coefficient between our empirical measurement and the prediction of the different teams considered here. We find that the team that performed well with this correlation coefficient criteria are also the teams that performed well on the magnification recovery in the lens modelling challenge (See Meneghetti et al., 2017, figures 20 and 25 for comparison).

Conclusion on magnification ratios

This method for measuring magnification ratios based on the subtraction of the ICL from `MuSCADeT` seems to raise more accurate measurements of the flux of background lensed sources. Several aspects of the method need to be improved and better understood before claiming a robust method for magnification ratio determination. Computation of the flux from postage stamp summation should be replaced by more robust methods for photometric extraction such as `sextractor` (Bertin & Arnouts, 1996). Tests need to be conducted that include all the bands used in the extraction, photometry of the deblended images should be

compared with noiseless photometry of the original lensed sources and finally we need to quantify how magnification ratios could be used to constrain lens mass models.

Nevertheless, our preliminary results on simulations and real data show are encouraging us to investigate whether magnification ratio measurements could be used to, at least, assess the performance of a mass and at best, contribute imposing constraints on mass models.

4.5   Outline

In this chapter, I presented a method for the separation of colour components in astronomical multi-band images, based on their difference in SEDs and on the sparse representation of each component in the starlet domain. Our technique is very similar to the AMCA method that has the advantage of optimising for the mixing matrix. Due to the correlation between the SEDs of astronomical sources in the visible domain, we found that the pre-selection of the mixing matrix by way of a principal component analysis of the SEDs in each pixel of a multi-band image, allowed the algorithm to converge to a solution for the colour separation. First designed for the purpose of galaxy-scale strong lenses deblending, the method has proven very efficient at removing intra-cluster light in HST imaging of massive galaxy clusters.

After the successful application of the method to the Hubble frontier fields, I showed that this method could contribute to strong gravitational lens modelling in at least two ways:

- First, by removing the ICL, `MuSCADeT` reveals hidden objects behind galaxy clusters that would otherwise not be visible. This was shown in published work (Taylor et al., 2017; Massey et al., 2018), and I leave it to the reader to find new background lensed sources in HFF from the images provided in this manuscript. Colour images as well as the source fits files used to create these images will be made publicly available online in the near future.

- Second, by allowing better photometry for background lensed sources after deblending. I showed how the improved photometric measurements could be used to assess the accuracy of magnification maps prediction with the potential to help imposing constraints on cluster mass models.

Another result of the application of `MuSCADeT` to the Hubble frontier is the possibility for the deblending of different stellar components inside single galaxies. While this is out of the scope of this thesis work, I show some preliminary results on the subject in appendix F.

With the growth in number and size of upcoming large surveys, the need for automated methods for deblending is dire, and we see more examples of colour-based deblending technique appear. Recently, Jiang et al. (2017) proposed a method for joint deconvolution and separation of hyper-spectral data based on sparse decomposition of the components, and Moolekamp & Melchior (2017) and Melchior et al. (2018) proposed a generalisation of the GMCA approach

that includes convex constraints other than positivity and sparsity.

5 Joined Lensed Source Reconstruction and Deblending

Contrary to popular belief, the modelling of gravitational lenses to determine the mass distribution of a lens is not a "black art".

— Christopher Kochanek

Contents

With my short-term goal being to develop an automated method for strong gravitational lens inversion based on linear optimisation, I decided to break down the whole problem of lens modelling. In this chapter I will tackle what I consider to be the first step in modelling galaxy-galaxy strong gravitational lenses: the reconstruction of the lensed source light profile in source plane. Across this chapter, I will mostly refer to notions of strong gravitational lensing that were introduced in chapter 1. I recommend a careful reading of section 1.3 in particular for a better understanding of the base concepts behind the building of the linear problem I present in section 5.1.

Some of the modelling techniques currently in use consider a full analytical lens mass and light distribution (Kneib et al., 2011b; Bellagamba et al., 2017; Oguri, 2010). Others use a semi-analytic approach, where the source is pixelated and regularized but where the lens has

an analytical representation (Dye & Warren, 2005; Warren & Dye, 2003; Suyu et al., 2006; Vegetti & Koopmans, 2009), or where the lens is represented on a pixelated grid with regularisation or assumptions on its symmetry (e.g. Coles et al., 2014; Nightingale et al., 2018). Further work in this direction involves adaptive pixel grids to represent the source (Nightingale et al., 2018; Nightingale & Dye, 2015; Vegetti & Koopmans, 2009) or an analytical decomposition of the source on a predefined dictionary as was done in Birrer et al. (2015), where the authors used shapelets (Refregier, 2003).

In this chapter, I address the problems of source reconstruction and deblending as a single linear inverse problem. By using Starlets , I am able to use sparse regularisation over the lens and source light profiles in their respective planes. Sparsity with starlets has the advantage of performing model-independent reconstructions of smooth profiles and allows for deterministic expression of the regularisation parameter. Because the lensed source can be represented using only a limited number of starlet coefficients, the pixel grid can be almost as thin as desired and the reconstructed source is denoised and deconvolved from the instrumental Point Spread Function (PSF). In chapter 4, we have seen how deblending could be achieved from discriminating between colours in multi-band images. In many cases, multiple bands might not be available. This new approach has the advantage to allow for the deblending of strong gravitational lenses in single-band images.

This chapter is an adaptation of a paper submitted for publication to *Astronomy & Astrophysics*, which is intended as a proof of concept to show how convex optimisation under a sparsity prior on the source light profile can be used as an adequate minimisation technique for lens modelling. The scope of this chapter remains limited to the modelling of light distribution alone and to the potential of using morphological component analysis to provide a new framework for lens modelling.

5.1 Source reconstruction given a known lens mass

In this section, I recall the basics of gravitational lensing formalism I will use to express the back-projection of lensed images to source plane and write the linear equations that form the basis of the minimisation problem I solve in the next section.

I note $\boldsymbol{\theta}$ the angular position on the sky of an object seen through a gravitational lens (image plane coordinates), with intrinsic angular position $\boldsymbol{\beta}$ (source plane coordinates). I recall here that the mapping from source to image plane is described by the lens equation:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{5.1}$$

The problem of inverting Eq. 5.1 from photometric observations only (meaning only $\boldsymbol{\theta}$ is known) is a non-linear and highly under-constrained problem with two unknowns: The

source's position, $\boldsymbol{\beta}$, and the convergence map of the lens. In the case of extended sources, the goal of lens inversion is to recover the light profile of a lensed galaxy as seen in the source plane, which implies being able to calculate the flux at each position $\boldsymbol{\beta}$ knowing the flux at position $\boldsymbol{\theta}$.

In practise, current techniques for lens inversion rely on an iterative process that consists in successively reconstructing the source profile brightness and the κ map. In an effort of developing an automated, model independent method for lens inversion, I choose to decompose the problem. In this chapter, I address the problem of source light profile reconstruction given a known mass, in which case, the problem is linear. In Warren & Dye (2003), the authors express the mapping between source and image surface brightness using an operator F_κ, such that the observed surface brightness of a lensed galaxy can be written:

$$Y = F_\kappa S + Z, \tag{5.2}$$

where Y is the observed surface brightness, flattened as a vector with length N_{pix}. Vector S is the unknown source surface brightness vector in the source plane with length N_{ps}. F_κ is a $N_{ps} \times N_{pix}$ matrix where, following Warren & Dye (2003)'s formalism, element $f_{i,j}$ is the jth pixel in image plane of the mapping of a source that has only its ith pixel set to one. In other words, F_κ indicates which pixels from the source plane have to be combined to predict the value of a pixel in the image plane. The elements of F_κ are entirely determined by the mass density distribution κ. Vector Z is an additive noise map. I consider Z as a white Gaussian noise with standard deviation σ, but the method can easily be extended to Poisson statistics, or more generally to noise with known root mean square.

5.1.1 Pixel-to-pixel mapping

As illustrated in Fig 1 of Wayth & Webster (2006), a square pixel in the image plane is a diamond shaped pixel in source plane, with a total area depending on the magnification at the pixel location. Although this phenomenon should strictly be accounted for, I choose to make the approximation that each photon hitting a pixel whether in source or image plane hits at the centre of the pixel. This way, part of the information provided by the continuous distribution of photons over the whole surface of a pixel is lost, but this approximation has the advantage of avoiding correlating the noise when back projecting from the image plane to the source plane. Furthermore, using a small pixel size limits this imperfect modelling and allows to compensate for the variation in light profile. This inverse problem being ill-posed, it admits no unique and stable solution, hence calling for a regularisation to solve it. Increasing the pixel size, and therefore losing resolution can be seen as a naive regularisation. I will show how advanced regularisation techniques can be used efficiently to address this problem.

5.1.2 Projection and back-projection between source and lens planes

To compute the elements of F_κ for each of the N_{ps} pixels in the source plane, I associate a pixel in the image plane by shooting a photon from the centre, $\boldsymbol{\beta}$ of a source plane pixel and recording the position(s) $\boldsymbol{\theta}$, given by Eq. 5.1, of the pixel(s) in the image plane where the deflected photon will hit. This boils down to recording the positions $\boldsymbol{\theta}$ where $\boldsymbol{\beta} + \boldsymbol{\alpha}(\boldsymbol{\theta}) - \boldsymbol{\theta} = 0$. The element of F_κ at position(s) (β, θ) is(-are in the case of a multiply imaged pixel) set to one to indicate the mapping. Since F_κ is a sparse matrix with only very few non-zero coefficients, I choose to only store the positions (β, θ) that map into one another in order to save memory and hence, computation time in the following steps.

The projection of a source profile into a lensed profile in the image plane is then performed by allocating to each image pixel the sum of the intensities of the corresponding source pixels according to F_κ. Conversely, back-projection is performed by allocating to each source pixel the average value of all its lensed counterparts according to F_κ. This ensure conservation of surface brightness between the source and lens planes.

5.2 Linear development of strong gravitational lens imaging

In real imaging data of strong gravitational lenses, the problem of finding the delensed light profile of a lensed galaxy is harder than solving Eq. 5.2, which is already non-trivial. First, one has to include the impulse response of the instrument that acquired the image. This effect corresponds to a convolution of the images described by $F_\kappa S$, by the point spread function (PSF). Let the linear operator H account for the convolution by a known PSF, Eq. 5.2 becomes:

$$Y = HF_\kappa S + Z, \tag{5.3}$$

which is the problem one has to solve when dealing only with the lensed light profile of a source, assuming that the light profile from the foreground lens galaxy has been perfectly removed prior to the analysis.

In practise, images of strongly lensed galaxies are contaminated by light from a foreground lens galaxy, G. Taking this into account, Eq. 5.3 then writes:

$$Y = H(F_\kappa S + G) + Z. \tag{5.4}$$

When Z is a white gaussian noise, solving Eq. 5.4 reduces to finding S and G such that:

$$||Y - H(F_\kappa S + G)||_2^2 < \epsilon \tag{5.5}$$

Where ϵ accounts for the precision of the reconstruction and depends on the noise level. Given that an image has N_{pix} pixels and that I aim at finding an N_{pix}-sized galaxy profile and an N_{ps}-sized source profile, further constraints are required on the light profiles. Classically, the light distribution of the lens is approximated by an analytic profile such as a Sersic or deVaucouleur profile.

Reconstructing the source light profile being an ill-posed problem , where unknowns largely outnumber the number of observables, several strategies have been investigated in the literature, for instance: adaptive pixel grids (Dye & Warren, 2005), negentropy minimisation (Warren & Dye, 2003; Wayth & Webster, 2006), Bayesian inference over the regularisation parameters of the source (Suyu et al., 2006), perturbative theory (Alard, 2009) or model profile fitting (Bellagamba et al., 2017). Although these methods have their own advantages and disadvantages, only few of them are able to reconstruct complex sources without degrading the resolution of the output. In Birrer et al. (2015), the authors used a family of functions to reconstruct the source light profile with promising results. Here I propose to push this idea further by exploiting a redundant family of functions that is very well suited to represent galaxies, and that possesses properties of redundancy allowing for the use of sparse regularisation.

5.2.1 Source reconstruction in absence of light from the lens

I propose a new approach to solve Eq. 5.3. Given that galaxies are compact and smooth objects, their decomposition over the starlet dictionary (Starck et al., 2007) will be sparse, meaning that only a small number of non-zero starlet coefficients will contain all the information in a galaxy image. This property allows to constrain the number of coefficients used in starlet space to reconstruct galaxy profiles, therefore offering a powerful regularisation to this problem.

Sparse regularisation

Assuming a signal is sparse in a dictionary Φ, the solution to an inverse problem like in equation. 5.3 is the solution that uses the least number of coefficients in the Φ dictionary while minimising the square error between the observables and the reconstruction. In a more formal way, sparsity is enforced by minimising the ℓ_1-norm of the decomposition over Φ of a signal known to be sparse in this dictionary. In addition, because the mapping of an image from lens plane to source plane does not conserve shapes, the edges of the image in lens plane does not match the borders of the image in the source plane, leaving parts of the source image unconstrained as they map into pixels outside the field of view of the lens plane. Let me call $\mathbb{S}$

the set of pixels in source plane that have an image in lens plane. I impose that the coefficients of the solution outside set $\mathbb{S}$ be set to zero. This allows me to write the problem of finding S as an optimisation problem of the form:

$$\tilde{\alpha}_S = \underset{\alpha_S}{argmin}||Y - HF_\kappa \Phi \alpha_S||_2^2 + \lambda||W \odot \alpha_S||_1 \qquad (5.6)$$

where Φ is the starlet dictionary, α_S are the starlet coefficients of S such that $\alpha_S = \Phi^T S$. The operator $\odot$ is the term by term multiplication operator, and W is a vector of weights that serves the purpose of setting to zero all coefficients outside $\mathbb{S}$, while keeping the ℓ_1-norm constraint from biasing the results (more on that in the following paragraphs). In practice, minimising the ℓ_1-norm of a vector is done by soft-thresholding the vector. This consists in decreasing by a positive factor λ the absolute value of all its coefficients and by setting to zero the coefficients smaller than λ, as shown in the following equation:

$$ST_\lambda(x) = \begin{cases} sign(x) \times (|x| - \lambda) & \text{if } |x| > \lambda \\ 0 & \text{otherwise} \end{cases} \qquad (5.7)$$

The regularisation parameter λ controls the trade-off between fitting the observed data and enforcing sparse solutions. From the definition of equation 5.7, it appears that solutions derived with soft-thresholding will present a bias due to the subtraction by λ. In order to mitigate this effect, I use the reweighting scheme from Candes et al. (2008). In order to prevent the most significant coefficients from being truncated, I multiply the regularisation parameter λ by:

$$W = \frac{2}{1 + exp(-10(\lambda - \alpha_0))}, \qquad (5.8)$$

where α_0 is the solution of equation 5.6 with $W = 1$. With this definition for W, the coefficients that are much larger than λ are less affected by soft thresholding than others. Values of W for coefficients outside $\mathbb{S}$ are set to infinity, hence naturally ensuring that the corresponding wavelet coefficients are set to zero.

This approach can be used to recover the source light profile in systems involving a faint foreground lens galaxy, a large Einstein radius, or when a reliable deblending of the lens and source light profiles is available prior to the source reconstruction scheme presented here.

5.2.2 Source reconstruction and deblending of the foreground lens light profile

In a more general case, one has to deal with the separation between the lens and source light profiles. Although several techniques allow for their separation prior to the analysis of the lens system (e.g. Joseph et al., 2014, 2016; Brault & Gavazzi, 2015), each of these methods have limitations, in the sense that they require specific inputs (field of view, or multiband images) or do not take into account the lensed source profile when fitting the lens, resulting in potential biases. Another approach consists in fitting an analytic lens light profile while reconstructing the lens mass density profile and the source (Birrer et al., 2015; Tessore et al., 2016). Here, I propose a solution to reconstruct and separate the lens and source light profiles using the fully linear framework provided by morphological component analysis (henceforth, MCA, Starck et al., 2005b).

Very importantly, a galaxy is sparser in starlets in its own plane (source or image), meaning that, given a mapping F_κ, with $\kappa(\theta) > 1$ at several positions θ, between source and lens plane, a galaxy in the source plane will be sparser than its projection to the lens plane. Likewise, a lens galaxy in the lens plane will be sparser than its projection to the source plane. I justify and illustrate this statement in Sect. 5.4 using simulated lenses.

Morphological component analysis allows for separation of two mixed components in a signal, based on the fact that each component can be sparsely represented in its own dictionary but not in others. Here, a dictionary is a collection of atoms (vectors) that, together, form a generative set of $\mathbb{R}^{N_{ps}}$. In the context of lens source separation, the explicit dictionaries are the starlet transform of a back-projection in source plane on one hand and the starlet transform for the lens, in lens plane, on the other hand.

I can therefore iteratively project the mixed signals in their own respective dictionaries, impose a sparsity constrain on each projection and therefore reconstruct the corresponding components separately. As seen in Sect. 5.2.1, sparsity is imposed by minimising the ℓ_1-norm of both decompositions.

5.2.3 Optimisation problem

In mathematical terms, the aforementioned MCA problem boils down to finding the model $\{\hat{S}, \hat{G}\}$ that provides the best approximation of the data set Y according to equation 5.4, while minimising the ℓ_1-norms of the starlet coefficients α_S, and α_G, with $\alpha_G = \Phi^T G$. This writes:

$$\hat{\alpha_S}, \hat{\alpha_G} = \underset{\alpha_S, \alpha_G}{arg\,min} \quad ||Y - H(F\Phi\alpha_S - \Phi\alpha_G)||_2^2 + \lambda_S ||W_S \odot \alpha_S||_1 \tag{5.9}$$
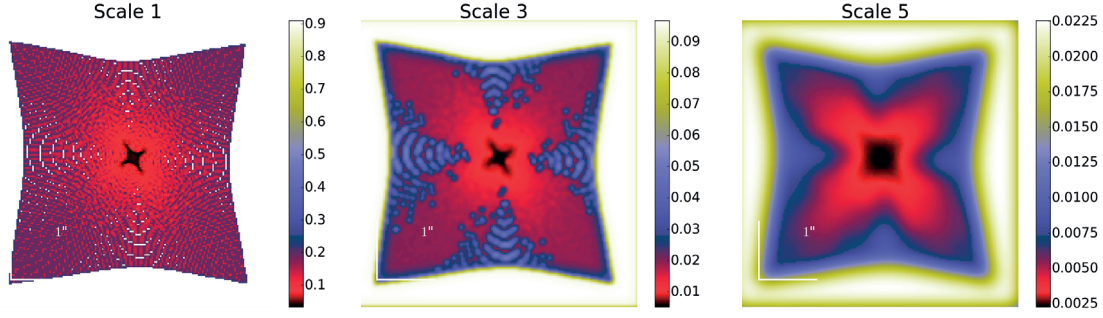$$+ \lambda_G ||W_G \odot \alpha_G||_1,$$

Figure 5.1 – Noise levels in source plane (λ_S) for three starlet scales (scale 1, 3 and 5) out the five computed for 100×100 pixels images with noise standard deviation $\sigma = 1$.

Here, λ_S and λ_G account for the sparsity of α_S and α_G respectively. Similarly to equation 5.6, W_S and W_G are weights that play the same role as in equation 5.8. I describe the calculation of these values in section 5.3.1.

Since, my main interest is to fully reconstruct the source S, it is not necessary to compute the fully deconvolved vector G. Instead, I only estimate the convolved vector $G_H = HG$ so that I extract the convolved foreground lens galaxy G_H and decrease the computational time by avoiding several convolution steps of over G when solving equation 5.9, which becomes:

$$\hat{\alpha}_S, \hat{\alpha}_{G_H} = \underset{\alpha_S, \alpha_{G_H}}{argmin} \quad ||Y - HF\Phi\alpha_S - \Phi\alpha_{G_H}||_2^2 + \lambda_S ||\alpha_S||_1 \qquad (5.10)$$

$$+ \lambda_{GH} ||\alpha_{G_H}||_1.$$

5.3   Method: the `SLIT` algorithms

In this section, I describe the two algorithms, `SLIT` and `SLIT_MCA` that I implemented to solve equation 5.6 (no lens light) and 5.10 (full light reconstruction problem) respectively.

5.3.1   Source delensing: `SLIT` algorithm

Starting with the simpler case of solving equation 5.3, I made use of the fast iterative soft thresholding (FISTA) algorithm (Beck & Teboulle, 2009). This iterative algorithm is similar to a forward backward (Gabay, 1983) algorithm with an inertial step that accelerates the convergence. I show the pseudo-algorithm for one iteration in algorithm 3. It consists in a step of gradient descent (steps 2 and 3) followed by a soft-thresholding of the starlet coefficients of the source (step 4), which acts as a sparse regularisation. Step 6 aims at pushing forward the

solution in the direction of smaller error, which accelerates the convergence. The process is repeated until convergence, as shown in algorithm 4

Algorithm 3 FISTA iteration

1: **procedure** FISTA$(Y, \alpha_i, H, F, \xi_i, t_i, \lambda, W)$
2: $\quad R \leftarrow \mu F^T H^T (Y - HF\Phi\alpha_i)$
3: $\quad \gamma \leftarrow \xi_i + \Phi^T R$
4: $\quad \alpha_{i+1} \leftarrow ST_{\lambda \odot W}(\gamma)$
5: $\quad t_{i+1} \leftarrow \frac{1+\sqrt{1+4t_{S_{i-1}}^2}}{2}$
6: $\quad \xi_{i+1} \leftarrow \alpha_i + \frac{t_i - 1}{t_{i+1}}(\alpha_{i+1} - \alpha_i)$
7: $\quad$ **return** $\xi_{i+1}, \alpha_{i+1}, t_{i+1}$
8: **end procedure**

In this algorithm, Y is the original image of a lensed galaxy, ξ_i and γ are local variables used to carry local estimates from one iteration to another, α_i is the starlet decomposition of the estimated source at iteration i and t_i gives the size of the inertial step. This sequence has been chosen to ensure that the cost function convergence is bounded by the Euclidian distance between the starting point for S and a minimiser of the cost function. This is explained in more details in Chambolle & Dossal (2015) (thm. 1) and Beck & Teboulle (2009) (thm 4.1). The gradient step μ is chosen to be $\mu = (||HF\Phi||_s^2)^{-1}$, with $||.||_s$ being the spectral norm of a matrix, defined by:

$$||M||_s = \max_{x \neq 0} \frac{||Mx||_2}{||x||_2}. \tag{5.11}$$

Function ST_λ is the soft-thresholding operator described by equation 5.7.

Algorithm 4 SLIT

1: **procedure** SLIT$(Y, H, F, \lambda, Niter, W)$
2: $\quad \xi_0, \alpha_0, t_0 \leftarrow 0, 0, 1$
3: $\quad$ **for** $0 < i \leq N_{iter}$ **do**
4: $\quad\quad \xi_{i+1}, \alpha_{i+1}, t_{i+1} \leftarrow FISTA(Y, \alpha_i, H, F, \xi_i, t_i, \lambda, W)$
5: $\quad$ **end for**
6: $\quad S = \Phi\alpha_{N_{iter}}$
7: $\quad W = \frac{2}{1+exp(-10(\lambda - \alpha_0))}$,
8: $\quad$ **return** S, W
9: **end procedure**

Parameter λ has to be chosen with care as it accounts for the sparsity of the solution. In practice, λ is a threshold that is applied to each starlet coefficient of the solution in order to reduce its ℓ_1-norm in starlet space. In the present case, given the presence of noise in the input data Y, it is important to chose a threshold above the noise levels. This is done by propagating the noise levels in image Y to the starlet coefficients α_i. The starlet transform being an undecimated multi-scale transform, coefficients α_i can be ordered as a set of images, each image representing the variations in the data at different scales. Therefore, it is necessary

to estimate how noise levels translate from the data to each scale of the starlet transform. In the current implementation, the noise from image Y also has to be propagated through the H^T and F^T operators as shown in step 2 of algorithm 3. Because the convolution H^T correlates the noise in the data and the back-projection to source plane induces varying multiplicity of the delensed pixels across the field of view, it is necessary to estimate a different threshold λ at each pixel location in each scale of the starlet transform of the source. In practice, for measurements affected by noise with known covariance Σ, noise standard deviation in the starlet domain of the source plane are given by the square root of the diagonal elements of

$$\Sigma_S = \Phi^T F_\kappa^T H^T \Sigma H F_\kappa \Phi \tag{5.12}$$

In the case of the "a trou" (french for "with whole", Holschneider et al., 1989; Shensa, 1992) algorithm, which relies on filter bank convolution to perform the starlet transform, the elements of Φ are never explicitly calculated. Instead, the noise standard deviation at scale s and pixel p in the starlet domain of the source plane is given by:

$$\Xi_{s,p}^2 = \Delta_s^2 * \Gamma_{i,j,(i=j)}^2, \tag{5.13}$$

Where Δ_s is the starlet transform of a dirac function at scale s, $*$ is the convolution operator, and $\Gamma_{i,j,\{i=j\}}$ is the vector containing the diagonal elements of $F_\kappa^T H^T \Sigma H F_\kappa$.

The result is the noise level in source space at each location and each scale of the starlet transform. By construction, the last scale contains the coarse details in the image and is left untouched in the thresholding process. Figure 5.1 shows the noise levels in source plane calculated from a simulation where the surface mass density is a singular isothermal ellipsoid (SIE). The PSF is a simulated Tiny Tim PSF (Krist et al., 2011), for the F814W filter of the ACS/WFC instrument on the HST and noise standard deviation is set to 1. The original image is a 100×100 pixels image which is decomposed in 6 starlet scales, i.e. the maximum number of scales that can possibly be computed, given the size of the image.

5.3.2   Deblending and source de-lensing: `SLIT_MCA` algorithm

In real data, the source and lens light profiles are blended, i.e. the light of the lens impacts the quality of the source reconstruction. Handling the deblending and de-lensing simultaneously can be done using MCA.

In classical source separation problems where two components are to be separated, solutions are obtained through MCA, by performing a gradient step and by alternatively regularising over each component in its own dictionary. In the present case, solving equation 5.4 requires solving an inverse problem each time one aims at reconstructing the quantity $HF_\kappa S$. This inverse problem corresponds to solving equation 5.3 for which I already presented a solution in algorithm 4.

Our MCA algorithm is therefore an iterative process that consists in alternatively subtracting a

previous estimation of G_H and S to the data:

$$D_S = Y - G_H \tag{5.14}$$

and

$$D_G = Y - HF_\kappa S, \tag{5.15}$$

as detailed in algorithm 5. At each iteration, the previous subtractions D_S and D_G are used to estimate S and G_H respectively. Estimating S requires running the full SLIT algorithm on D_S until convergence. Estimating G_H at a given iteration is simply done by running one single iteration of the FISTA algorithm on D_G with inputs F_κ and H being identity matrices. I found empirically that using the projections of S and G_H on the subsets of vectors with positive values (meaning that all negative coefficients are set to zero), faster convergence towards more realistic solutions could be achieved. Although this is not a formal positivity constraint since I do not apply positivity on the solutions themselves, I found that in practice this leads to galaxy profiles with less negative structures which is not a physical feature found in galaxy light profiles.

Estimating λ_G is as crucial as estimating λ_S but is much simpler given that there is no inverse problem to solve in this case. The threshold λ_G only depends on the noise level in the image. Given that I impose sparsity in starlet space, noise levels still have to be evaluated at each scale of the starlet transform. To do so, I simply compute how a unitary signal in direct space translates into starlet space and multiply it by the noise standard deviation. In other words, I take the starlet transform of a Dirac function and compute the 2-norm of each scale of the starlet transform. This tells how energy is distributed into starlet space. For a decomposition over 6 starlet scales, the values obtained for the first five scales in order of increasing scale are: $\lambda_G = [0.891, 0.200, 0.086, 0.041, 0.020]$. As in the previous section, the last scale is left untouched. The obtained values are then multiplied by a scalar that accounts for the desired detection level in units of noise. The scalar is often chosen to be between 3 and 5σ as seen previously. A detection at 3σ will produce very complete but noisy reconstruction of the signal, when 5σ will lead to a more conservative reconstruction of the most high signal-to-noise ratio features only. The obtained thresholds are applied uniformly across each scale.

5.4 Numerical experiments with simulations

In the following, I illustrate the performances of the algorithms with numerical experiments that mimic different observational situations.

Algorithm 5 SLIT MCA algorithm

1: **procedure** $SLIT\_MCA(Y, H, F_\kappa, N_{iter}, N_{subiter}, \lambda_S, \lambda_G)$
2:     $\tilde{S} \leftarrow 0$
3:     $\tilde{G}_H \leftarrow 0$
4:     $[\xi_{S0}, \xi_{G0}] = [0, 0]$
5:     $[\alpha_{S0}, \alpha_{G0}] = [0, 0]$
6:     $[t_{S0}, t_{G0}] = [1, 1]$
7:     **for** $0 < i \leq N_{iter}$ **do**
8:         $D_S \leftarrow Y - G_H$
9:         $S \leftarrow SLIT(D_S, H, F_\kappa, \lambda_S, N_{subiter})$
10:         $D_G \leftarrow Y - HF_\kappa S$
11:         $\xi_{Gi}, \alpha_{Gi}, t_{Gi} \leftarrow FISTA(D_G, \alpha_{Gi-1}, I_d, I_d, \xi_{Gi-1}, t_{Gi-1}, \lambda_G, 1)$
12:         $G_H \leftarrow \Phi^T \alpha_{Gi}$
13:     **end for**
14:     **return** $\tilde{S}, G$
15: **end procedure**

5.4.1 Creating realistic simulated lenses

In order to make the simulations as realistic as possible, I use galaxy light profiles extracted from deep HST/ACS images taken in the F814W filter. The images are part of the Hubble Frontier Fields program and the specific data[1] were taken from the galaxy cluster Abell 2744 (Lotz et al., 2017). I selected various patches, each containing a galaxy that is used to represent a lens or a source. The HFF images were cleaned using starlet filtering with a 5-sigma threshold. Source galaxies have been chosen to display significant structures over a broad range of scales, that I aim at recovering with the lens inversion methods. The lens light profile is smoother, as expected for a typical massive early-type galaxy, e.g. like in the SLACS samples (Bolton et al., 2008).

To generate the simulations I then lens the sources following the recipe in Sect. 5.1.2, using various lens mass profiles. Then, the lens light is added and convolved with a PSF created with the Tiny Tim software (Krist et al., 2011) for the ACS/WFC and the F814W filter.

The images shown in this section were created from images taken with the ACS/WFC instrument on HST. Flux units are showed in e^- and pixels in image plane are 0.05 arc-seconds on-a-side.

5.4.2 Plane-wise sparsity of galaxy light profiles

The `MCA-SLIT` algorithm consists in projecting the mixture image of the lens galaxy and the lensed galaxy, back and forth between the source and lens plane, and thresholding the starlet coefficients of each projection. My hypothesis is that the starlet thresholding favours, in each

---

[1]the frames were recovered from the HFF site at http://www.stsci.edu/hst/campaigns/frontier-fields/FF-Data

plane, the corresponding galaxy: source galaxy in source plane and lens galaxy in lens plane. This assumes that a lens galaxy in lens plane is well reconstructed with only a few starlets coefficients, while a source galaxy projected to lens plane is not. Conversely, it implies that a source galaxy in source plane is sparser in starlet domain than a lens galaxy projected to source plane.

To verify this hypothesis, I selected 167 images of galaxies from cluster MACS J0717 from the HFF survey, in the F814W filter of the ACS/WFC camera of HST. I used the HFF-DeepSpace catalogue by Shipley et al. (2018) to select galaxies with a semi-major axis of at least 5 pixels and a flag at 0 in the F814W filter to insure that the galaxies are isolated in their stamp. I performed the starlet decomposition of these images along with their projections to source and lens plane using three different mass profiles (SIE, SIS and elliptical power law) with realistic draws of the lens parameters. For each of these starlet decompositions the $p\%$ smallest coefficients are set to zero and the image in pixel space in reconstructed. I then compute the error on the reconstruction as a function of p. The resulting curve is the non-linear approximation error (NLA; see Starck et al., 2015), shown in Fig. 5.2. The NLA can be used as a metric for the sparsity of a galaxy profile, in the sense that a sparse galaxy will see its NLA decreasing rapidly with p.

In Fig. 5.2, we see in particular that the NLA of galaxies (in red) decreases faster than the one of the lensed source (in green). This means that when keeping only a small percentage of the highest coefficients in starlets (say 10%) of the decomposition of an image Y, will reconstruct well the lens galaxy but not the lensed source. When comparing the NLA of a galaxy (red curve in Fig. 5.2) to that of its projection to source plane (cyan curve in Fig. 5.2), we see that the NLAs of both profiles are very similar, making it difficult to disentangle between them. In practice, the reconstructed source images can be contaminated with features belonging to projections of the lens galaxy if not converged properly. On the bright side, since lens galaxy light profiles are being reconstructed in lens plane very efficiently, the signal from lens galaxy in source plane will decrease very rapidly with iterations.

5.4.3 Testing `SLIT` and `SLIT_MCA` with simulations

I apply my algorithms to a set of simulated images and show comparisons of reconstructions with a state-of-the-art method: `lenstronomy` (Birrer & Amara, 2018). In the present chapter, my goal is to show the potential of MCA-based algorithms as a simultaneous source reconstruction and source-lens deblending technique. All the tests therefore assume that the mass density profile of the lens is known, as well as the PSF. Unless stated otherwise, the following images were simulated with white Gaussian noise with standard deviation σ. I define the
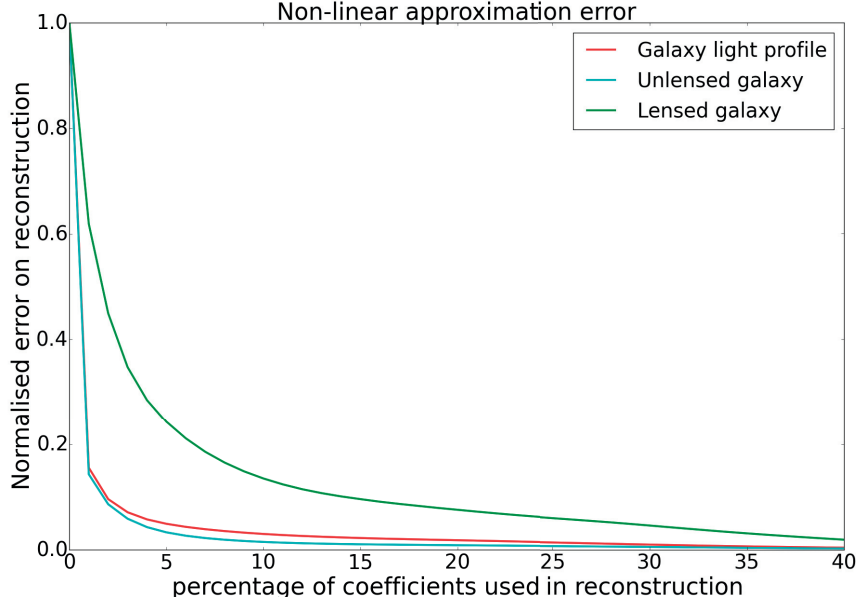
Figure 5.2 – Normalised non-linear approximation (NLA; 5.4.2) of galaxies projected in source and lens plane. The red curve stands for the average NLA of galaxy images that can be seen as source or lens galaxies. The cyan curve stands for the average NLA of the same galaxies once projected from lens to source plane. The green curve stands for the NLA of the same galaxies projected to lens plane.

signal to noise ratio (SNR) of an image I with N_{pix} pixels as:

$$SNR = \frac{1}{N_{pix}\sigma^2} \sum_{N_{pix}} I^2 \qquad (5.16)$$

Case 1: simulation with no lens light

I first reconstruct an image of a lensed galaxy with no foreground light. The simulation contains Gaussian white noise with $SNR = 50$. I used 50 iterations of Algorithm 4. The results are presented in Fig 5.3 illustrating the quality of the reconstruction, and leading to negligible residuals.

In this simulation, there are 4 times more pixels in the source plane image than in the lens plane, i.e. in image plane, the observable is an image of 100 by 100 pixels, while the source is reconstructed on a 200 by 200 pixels.

192
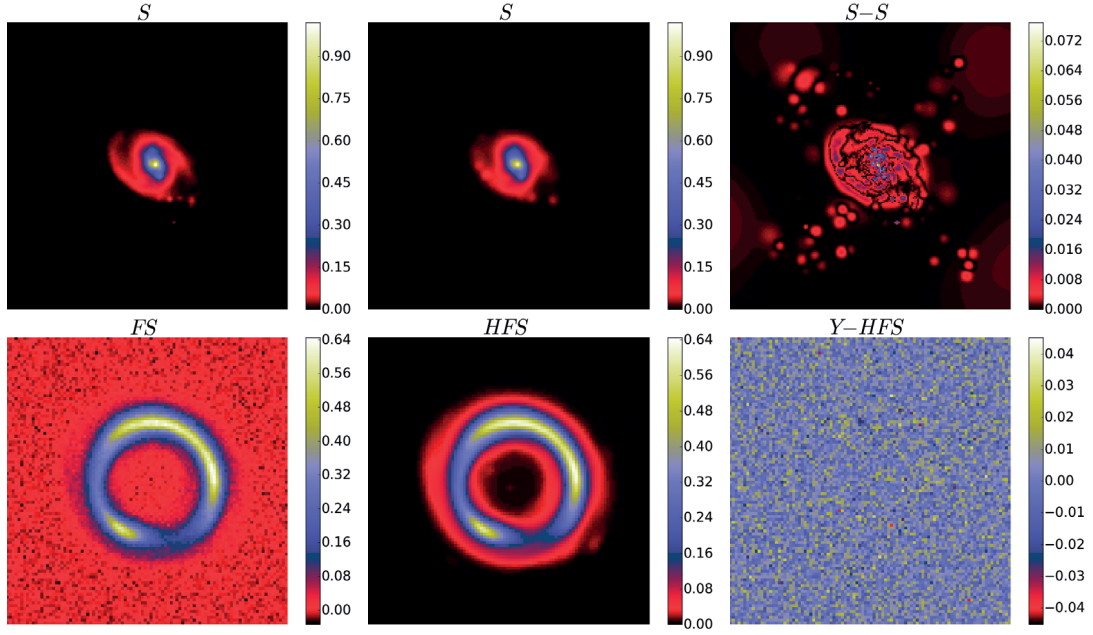
Figure 5.3 – Application of the SLIT algorithm to a simulated lensing system in the simple case where there is no light from the lensing galaxy. *Left:* the simulated source is shown on the top while its lensed and noisy version is shown on the bottom. Both include a PSF convolution. *Middle:* source recovered with the SLIT algorithm and the lensed version of it. Note that both are still convolved with the PSF. *Right:* the difference with the true source (top) and the residuals in the lens plane (bottom). The original and reconstructed images are displayed with the same colour cuts. The residuals in the bottom right panel are shown with $\pm 5\sigma$ cuts.

Case 2: simulation with both lens and source light

I then applied Algorithm 5 to the simulated images of a full lens system. In this case both the convolved lens light profile and the source light profile (Fig 5.4) are recovered. Sparsity is enforced in each solution by using enough iterations of the algorithm to perform an efficient separation. A difficulty here is to chose the numbers of iterations and sub-iterations such that both components converge to a sparse solution. In these experiments, $N_{iter} = N_{subiter}$ ensures similar quality in the reconstructions of both components.

The results show no structure in the residuals and visually good separation between the lens and the source as well as a good reconstruction of the source without significant leakage between the two. However, the residuals in the first three lines of the Figure show that the source flux was slightly overestimated at larger scales, while the lens galaxy was slightly underestimated. The amplitude of the phenomenon reaches no more than 10σ of the noise level given the amplitudes displayed in the last column.
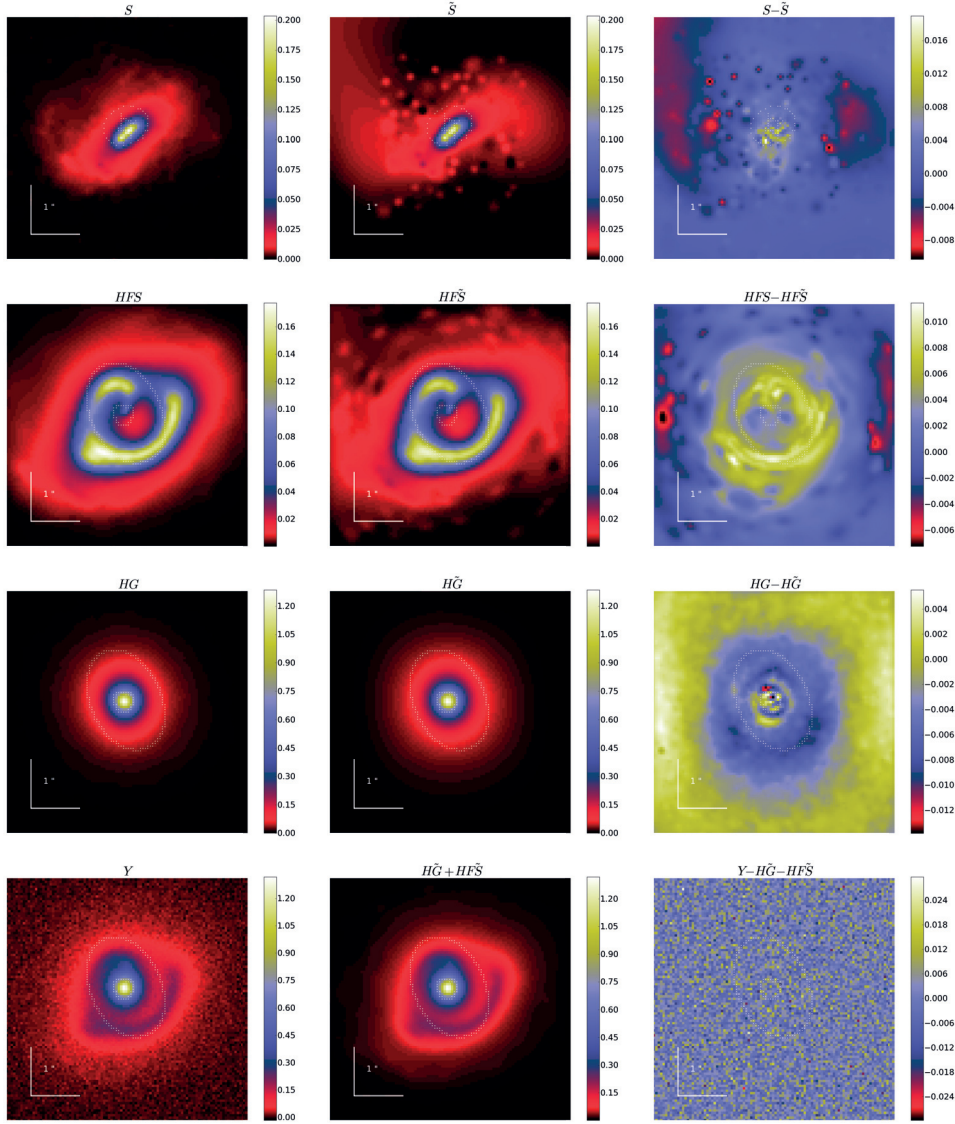
Figure 5.4 – Illustration of the `SLIT_MCA` algorithm with simulated data. *Left:* simulated ground truths. From top to bottom are shown the original un-lensed source, its lensed version also convolved with the PSF, the lensing galaxy (convolved with the PSF) and the full simulated system where noise had been added. *Middle:* the output of the `SLIT_MCA` algorithm. *Right:* the differences between the two previous panels. The original and reconstructed images are displayed with the same colour cuts. The residuals in bottom right panel are shown with cuts set to $\pm 5\sigma$. White dots show the positions of pixels crossed by critical lines in lens plane and by caustics in source plane.

5.4.4 Comparison with `lenstronomy`

I tested my source reconstruction technique on three other simulations with various source morphologies and lens mass profiles, including one generated with the `lenstronomy` package.

Image number	Source Origin	Source processing	Factor	PSF	Noise	Lens model
1	NASA, ESA jpeg	Shapelets	5	Gaussian	P+G	SPEP
2	HFF fields	$\star G_-$	4	Gaussian	SNR = 10	SIE
3	HFF fields	$\star G_-$	2	Gaussian	SNR = 100	SIS

Table 5.1 – Description of the simulated images. Symbol $\star G_-$ stands for convolution by a Gaussian kernel and subtraction of the median (see text). Column factor stands for the resolution factor between source and lens plane. For instance, in image 1, the source has 5 times more pixels on the side than the image. Gaussian PSFs were used in all three images with a FWHM of 2 pixels, P+G stands for gaussian Poisson mixture. Gaussian noise with $\sigma = 2$ was used. The Poisson noise value at pixel i is drawn from a Gaussian distribution with $\sigma = \sqrt{f_i}$, f_i being the flux in pixel i. Details of the lens profiles are given in appendix G.1

I compared the reconstructions of three lensed sources with the ones computed by S. Birrer, the author of `lenstronomy`. In order to avoid favouring one method over the other, we tried as much as possible to chose representations for true sources that do not correspond to either code's decomposition for reconstructing sources. Sources generated with `SLIT` were extracted from HFF images and convolved with a Gaussian kernel with a full width at half maximum of five pixels. This produces a smooth version of the noisy HFF images to which I then subtract the median value of the image in order to set the sky background to zero. All remaining negative values in the image are set to zero. The image generated with `lenstronomy` uses a source from a jpeg image of NGC 1300 from NASA, ESA. The image resolution is degraded by a factor 25 and decomposed over the shapelet dictionary (Refregier, 2003) using enough coefficients (11476) to accurately recover the morphology of the image. Despite `lenstronomy` relying on shapelets to solve the source inversion problem, the number of coefficients that it is possible to recover in the reconstruction is much smaller than the number of coefficients used in generating the true source. Therefore, the basis set of the reconstruction is different from the one used in generating the true source. The three systems tested here were made from sources with different morphologies and different lens profiles table 5.1. In this exercise I test the methods on simulated images that were generated with different procedures. This comparison therefore allows to show how robust these techniques are to the underlying mapping between source and lens plane.

In order to compare the results of both methods I show the resulting reconstructions of the runs in Figs. 5.5 and 5.6 and use three metrics:

- Quality of the residuals ,$R(\tilde{S})$, given by the relative standard deviation of the residuals for a model of the source , $\tilde{S}$:

$$R = std\left(\frac{Y - HF\tilde{(S)}}{\sigma}\right). \tag{5.17}$$

In cases of Gaussian and Poisson mixture noise, $\sigma = \sqrt{\sigma_G^2 + f}$, where σ_G is the standard deviation of the Gaussian component and f is a 2-D map of the flux in the true noiseless

195

model for image Y. Given that definition, the closest R is from one, the better the reconstruction.

- Quality of the source reconstruction relative to the true source , S_{true}, estimated with the Source Distortion Ratio (SDR, Vincent et al., 2006). The SDR is the logarithm of the inverse of the error on the source light profile, weighted by true flux of the source. As a result, the higher the SDR, the better the reconstruction of the source. The SDR is computed as:

$$SDR(\tilde{S}) = 10log_{10} \frac{||S_{true}||}{||S_{true} - \tilde{S}||}. \tag{5.18}$$

- Computation time

These two metrics of quality and source residuals, were chosen to provide a measure of the quality of both reconstruction in both source and lens plane. While the SDR of the sources is the most informative metric with regards to the quality of the reconstruction of the source, it is also necessary to ensure that both methods are able to reconstruct similarly well the observables, hence the role of metric R. The evaluation of these metrics for both methods are given in Fig 5.7.

The residuals in Fig 5.5, as well as the results for $R(\tilde{S})$ in Fig 5.7 show that both codes achieve similar quality of reconstruction. While `lenstronomy` leaves a little bit more signal in the residuals, in particular in cases of smooth sources generated with SLIT, SLIT, on the contrary tends to create false detections at noise level at locations where the actual signal is zero, resulting in a slight over-fitting. In the case of `lenstronomy`, over-fitting of outer regions with no signal is prevented by the fact the method relies on shapelets, which are localised around the centre of the images provided that the number of coefficients used in the reconstruction is kept small. The down-side of that strategy is that shapelets hardly represent companion galaxies in the source such as the one on the right side of image 3 in Fig. 5.6. To circumvent this problem, one needs to use a second set of shapelets positioned at a different location to represent a second displaced light component. This has deliberately not been done in this comparison. From looking at the reconstructions of image 1 and the corresponding SDRs, it appears that `lenstronomy` performs slightly better at reconstructing truth images generated with high resolution and detailed features. Despite both reconstructions show the same levels of details, false detections in SLIT at locations where the truth signal is 0 contribute diminishing artificially its SDR. Also, the large number of pixels in the source compared to the number of pixels in the observations (25 times more pixels in the source than in the image) makes the problem highly under-constrained. In `lenstronomy`, this is overcome by computing a small number of shapelet coefficients and displaying them on an arbitrarily thin grid. With SLIT, I optimise for each coefficient in the starlet dictionary, which means, more pixels in source plane equals more unknowns.

Regarding computational time, while `lenstronomy` runs in less than a second, a typical SLIT
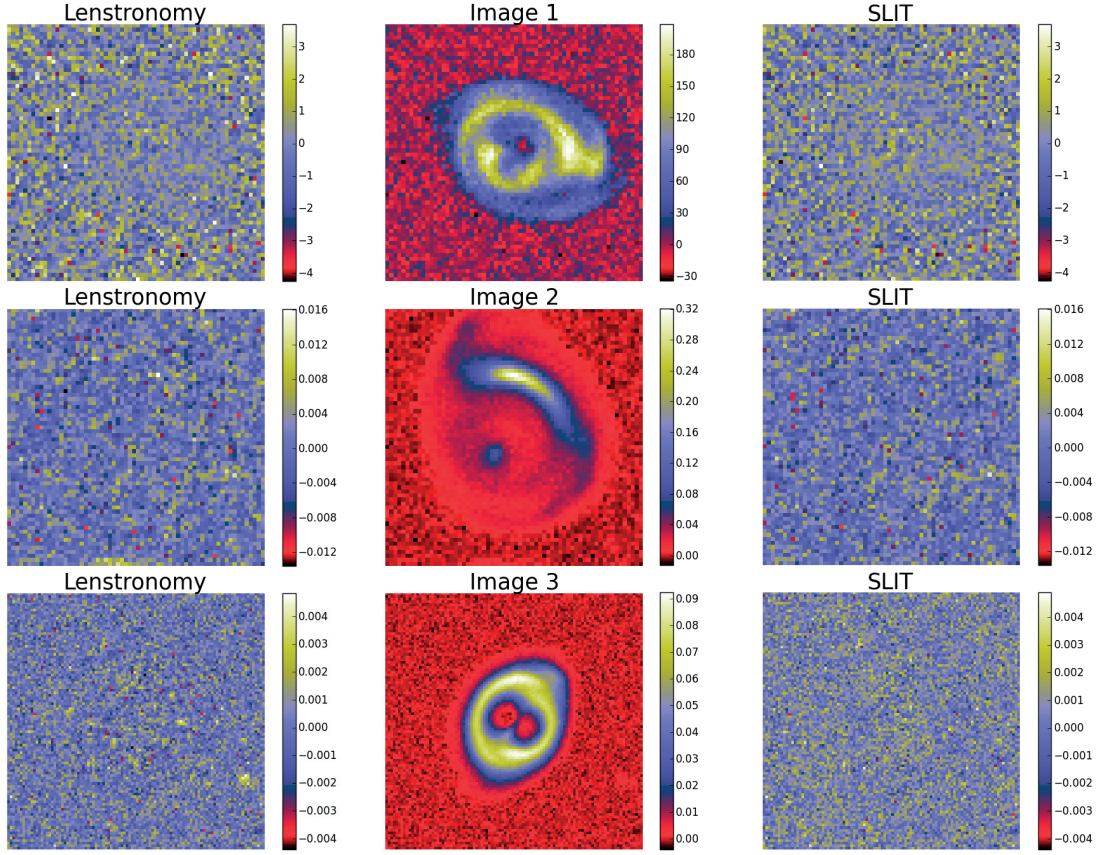
Figure 5.5 – Reconstructions with `Lenstronomy` and `SLIT` in image plane. The middle panels represent respectively from top to bottom, simulated images 1, 2 and 3. The left panels show the corresponding residuals after reconstruction with `lenstronomy`, while the right panels show the residuals obtained with `SLIT`.

run for examples such as the ones provided above will last between 250 and 500 seconds. For a full run of `SLIT_MCA`, this number is multiplied by a factor 5 to 10. While this is a current weakness of the algorithm, I am confident that optimised packages for starlet transforms and linear optimisation as well as parallel computing will allow to lower these numbers by at least a factor 10.

5.4.5 Lens parameter optimisation

In order to assess whether this technique holds the potential to be used in a full lens modelling context, I test its sensitivity to the density slope γ. In Koopmans et al. (2006, 2009), the authors showed real lens galaxies have, on average, a total mass density profile with a density slope $\gamma \sim 2$, i.e. isothermal. In Fig. 1 of Koopmans et al. (2009), the authors show that the posterior probability distributions of a sample of 58 strong lenses are maximised for a density slope between 1.5 and 2.5, which sets the limits for the values of γ investigated here.
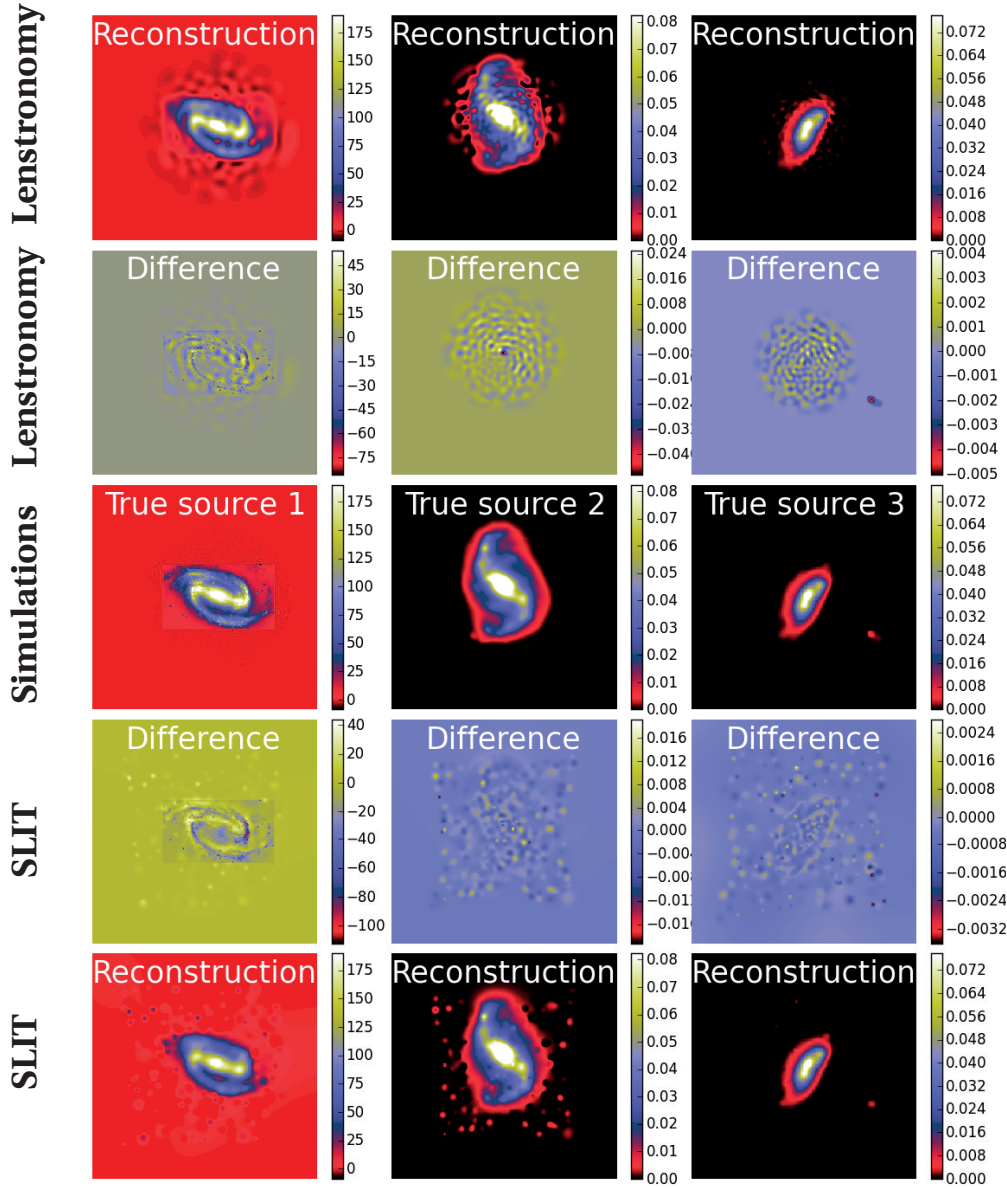
Figure 5.6 – Reconstructions with `Lenstronomy` and `SLIT` in source plane. Panels from the middle row show the true sources used to generate respectively simulated images 1,2 and 3. The first row show the source reconstruction from `lenstronomy`. The second row show the difference between the true sources and the sources reconstructed by `lenstronomy`. The last row shows the source reconstruction from `SLIT`. The penultimate row shows the difference between true sources and sources reconstructed with `SLIT`. Panels between reconstructed and true images, show the difference between the two for the corresponding technique.

In this analysis, I generate a lens system (shown in Fig 5.9) with a power law mass density profile with $\gamma = 2$. The light profiles for the lens and source galaxies were drawn from HFF images. The PSF is a gaussian profile with a FWHM of 2 pixels. I create 100 realisations of additive Gaussian at SNR 100 for this system and try to reconstruct, with `SLIT MCA`, each of
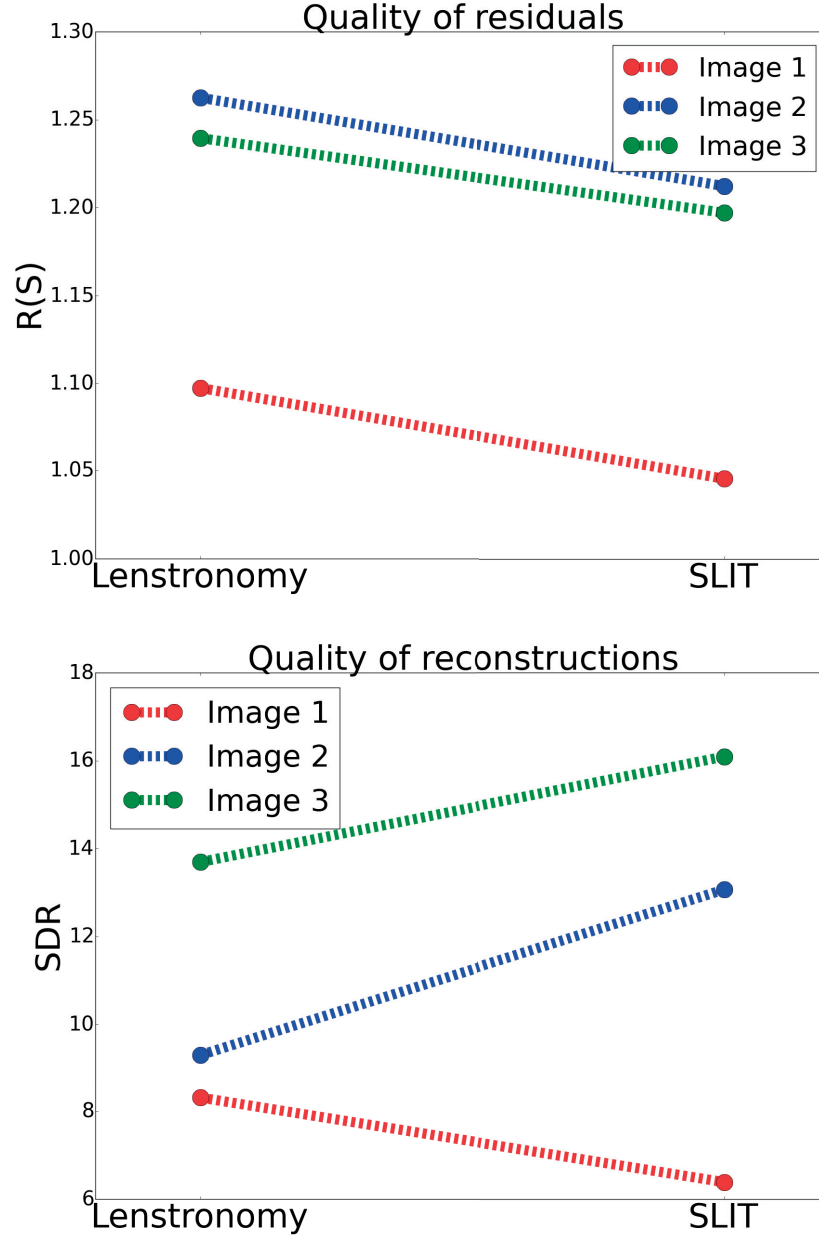
Figure 5.7 – Quantitative comparison of `Lenstronomy` and SLIT reconstructions. *Left panel*: quality of the residuals according to equation 5.17. *right panel*: SDR of the reconstructions according to equation 5.18.

these realisations with several density slopes $\tilde{\gamma}$ ranging from 1.5 to 2.5. The results are shown in Fig 5.8. The true and the reconstructed profiles are shown in appendix G.2.

Fig 5.8 shows that the actual morphologies of S and G_H are recovered very accurately ($SDR \sim$ 20) for $\gamma = 2$. In a real case study, the truth for S and G_H light distributions are not known,
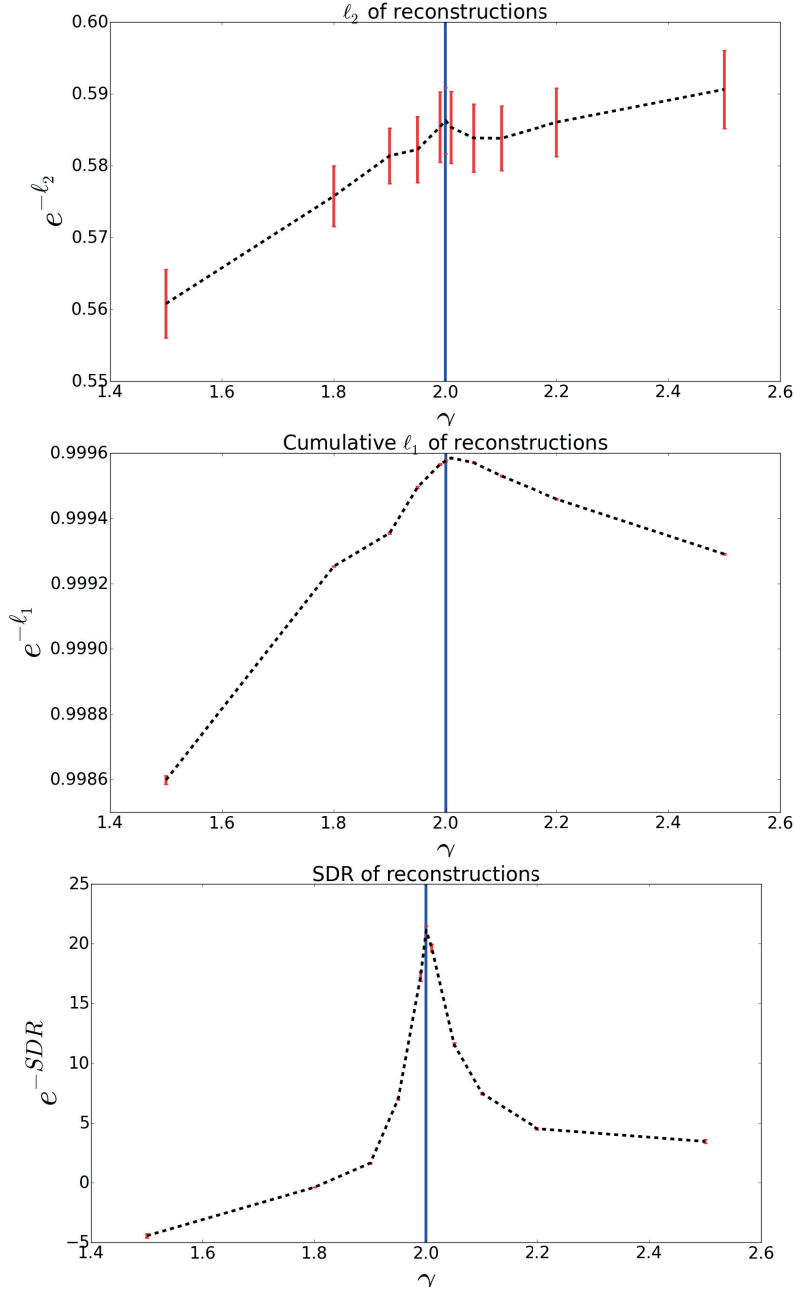
Figure 5.8 – Metrics of the reconstructions of system in Fig 5.9 as a function of mass density slope. The Top left panel shows the cumulative SDR of the source and galaxy light profile reconstruction. The top right panel shows the average of the residuals as $exp(-||Y_H F_{\kappa(\bar\gamma)} \Phi \alpha_S + \Phi \alpha_{G_H}||_2^2)$ over 100 noise realisations. The bottom panels displays the cumulative ℓ_1-norm of α_S and α_{G_H} as $exp(-(\lambda_S ||\alpha_S||_1 + \lambda_G ||\alpha_{G_H}||_1))$. The error bars show the standard deviation of these metrics over 100 noise realisations. The blue line shows the truth value γ.

therefore, it is impossible to use the SDR to discriminate between lens mass model parameters. Instead, I have to rely on quantities derived from the observations or on properties of the
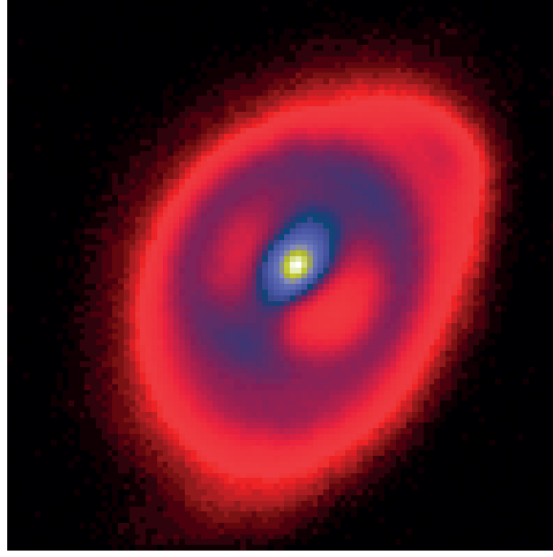
Figure 5.9 – Light profile of a simulated lens system (lens and lensed source light profiles) generated with a power law mass profile with $\gamma = 2$.

reconstructed profiles. The top right panel of Fig. 5.8 shows the likelihood defined as:

$$exp(-\mathscr{L}(\gamma) = ||Y_H F_{\kappa(\tilde{\gamma})}\Phi\alpha_S + \Phi\alpha_{G_H}||_2^2, \tag{5.19}$$

of a lens model with slope $\tilde{\gamma}$ to be the right model. Because of the strategy I chose for `SLIT_MCA`, which consists in optimising alternatively over the source and lens light profiles, the algorithm is very likely to estimate light profile models that will minimise the residuals in the image, even in cases where the mass model does not correspond to the truth. Hence the relative flatness, compared to the error bars, of the likelihood profile from Fig 5.8. In particular, we observe that, for $\tilde{\gamma} > 2.2$, a likelihood as high or higher than the likelihood at $\gamma = 2$ is achieved, despite the lens model being wrong. This is caused by the extreme steepness of the mass density profile, which causes `SLIT_MCA` to model the source as its lensed version back in source plane. With these results, one would think `SLIT_MCA` unfit to be used in a full lens modelling framework. The strength of the algorithm lies in its potential to find the sparsest solution to a problem of lens light modelling. Since wrong mass models introduce artefacts in the reconstructions of light models, their ℓ_1-norm is significantly higher than in the case of a reconstruction with a true model where light profiles are smooth. The bottom panel of Fig 5.8 shows the cumulative ℓ_1-norms of α_S and α_{G_H} as an argument of the likelihood. In this case, the metric is maximised around the true value for $\tilde{\gamma}$. Despite the sparsest solution being found for $\tilde{\gamma} = 2.01$, while the truth value for γ sits at 2.00, this result is still in the error bars estimated in Koopmans et al. (2009) over a sample of 58 lenses.

5.5   Outline

In this chapter, I built a fully linear framework to separate the lens and source light profiles in strong lensing systems, while reconstructing the source light profile as it was prior to the lensing effect at fixed lens mass model.

The first algorithm, `SLIT` (Sparse Lens Inversion Technique), decomposes the source plane on a basis of starlets. This allows us to represent the source in a non-analytical way, hence providing a sufficiently large number of degrees of freedom to capture any small structure in the data. `SLIT` applies to lensed systems where the lens light has been removed and for a fixed mass profile.

The second algorithm, `SLIT_MCA` (SLIT Morphological Component Analysis), also applies in the case of a fixed mass model but is able to deblend the lens light from the source as it reconstructs it in the source plane and deconvolves it from the instrumental PSF. The separation of lens and source light profiles in `SLIT_MCA` relies on the principle of morphological component analysis, but uses the distortion introduced by lensing itself as a way of discriminating between lens and source features. As is the case for `SLIT`, `SLIT_MCA` does not use any analytical form for either the source or the lens representation. Both algorithms account for the instrumental PSF.

I tested the algorithms with simulated images, showing that very accurate source reconstructions and separations can be obtained, within the observational noise. I identify several advantages of this approach:

- the lens and source light profiles are pixelated numerical profiles, allowing a large number of degrees of freedom in their reconstruction;

- the code implementing the algorithms is fully automated due to a careful automated computation of the regularisation parameter. It does not require any prior or assumption about the light profiles, either for the source or the lens;

- the performances of the algorithms are robust against pixel size in the sense that arbitrarily small pixel sizes can be used without leading to noise amplification or artefacts. The pixel size of both the source and lens can be an order of magnitude smaller than the PSF without negative impact on the results. On the contrary, adopting very small pixel sizes allows for detailed reconstruction of the source;

- the `python` code is public and easy to use.

- the linear framework of the algorithms presented here opens the possibilities for the developement of a full lens modelling technique based on non-negative matrix factorisation (NMF, Lee & Seung, 1999; Paatero & Tapper, 1994). This technique could allow to solve equation 5.4 in F_κ, G, S simultaneously. The knowledge of F_κ would in turn allow us to recover free form solutions for the density mass profile κ. We showed in Fig 5.8

how our strategy favoured the reconstruction of the sparsest solution, hinting at the possibility to be able to find a model for F_κ that will favour the sparsest set of lens and source profiles. Because this problem introduces another degree of complexity due to the number of unknowns and the increasing degeneracies of the solutions, we delay this study to a later publication.

The main limitations of the algorithm, at this time, is its very high computational time which scale to dozens minutes if not an hour when estimating very large source light profiles with `SLIT_MCA`. Although this makes the method unefficient as a possible minimiser in a Monte-Carlo Markov Chain sampling strategy, I am confident that upcoming optimised packages for linear optimisation will increase the speed of the algorithm. Also, the motivation behind the development of this technique is to be able to estimate free form lens mass models from methods such as non negative matrix factorisation, which requires less evaluations of `SLIT_MCA`.

6 Conclusion

The study of strong gravitational lens systems requires robust algorithmics at several levels, from the detection of gravitational lenses to the extraction of lensing parameters, through the preprocessing and deblending of lens images. In this thesis I have tried to express these problems as linear inverse problems when possible and to solve them as such. On the side of strong lens finding, principal component analysis has proven to be a strong asset allowing better detection by subtracting the light of potential foreground lens galaxies. Since the creation of this technique, much efforts have been done that rely on machine learning techniques that are not limited to learning the shape of lens galaxies, but that also learn the morphologies of background lensed sources.

In the last two chapters, I have described two methods for deblending strong gravitational lenses. While the first method relies on colour differences between lenses and sources in multi-band images, the second one uses the fact that a lensed source is the distorted version of an otherwise sparse galaxy light profile in source plane to perform deblending in single-band images. I hope the reader will have noticed that these two techniques provide independent discrimination criteria between lensed and unlensed objects can be combined in adequate data sets. Indeed, the morphological argument about the sparsity of galaxy profiles in their own referential still holds in multi-band images. It is therefore possible to express the light profile of a strong gravitational lens system as a combination of equations 3 in section 4.2 and equation 5.2 to form:

$$Y = H\left(A_1 G + A_2 F_\kappa S\right) + Z. \tag{6.1}$$

In equation 6.1, A_1 is the first column the mixing matrix from `MuSCADeT`'s formulation and A_2 is the second column. Vectors A_1 and A_2 therefore contain respectively the SEDs of the red foreground lens galaxy and the blue background lensed galaxy. Matrix H accounts for the convolution of each band by their corresponding PSF.

Developing an automated method for strong gravitational lens modelling requires being able to solve equation 6.1 in A_1, A_2, G, S and F_κ. While this appears as a complex problem to solve, we have established in chapters 4 and 5 that both lensing (matrix F_κ) and colour (A_1 and A_2) allowed to separate between lens and source. Combining these two independent separation criteria should lead to an improved joint separation and reconstruction of lens and source galaxies, at least at fixed mass distribution.

With regard to the estimation of lens mass models, while it is not possible to solve equation 6.1 as a linear problem in κ, it is possible to solve it for F_κ. Because F_κ is the matrix that gives the mapping between lens and source pixels, it is possible to assume the sparsity of each column of this matrix, which gives us constraints to factorise the matrix product $F_\kappa S$ even with F_κ and S unknown. Due to the global problem being widely underconstrained, I consider using a starting point for the matrix F_κ, derived from an initial guess of an analytical SIS profile, for instance. Deriving F_κ does not give us the surface mass density distribution, but by recording the pixel matching established in this matrix, one can derive the deflection angle $\boldsymbol{\alpha}$ at positions $\boldsymbol{\theta}$ where lensed images of a background source appeared. From this point, computing the surface mass distribution κ is a deconvolution problem given by equation 1.10, which can be solved as a linear inverse problem. Such method could allow the computation of numerical, free-form source and mass profiles. One of the advantages of free-form mass profiles is their sensitivity to massive substructures that would be extracted without having to account for them explicitly in the formulation of the problem.

While there are still active efforts to be made to improve the robustness and automation of strong lensing analysis algorithms, the recent discoveries of a lensed supernovae and a magnified individual star have opened the door to new science with gravitational lensing. Lensed supernova allow to multiply the occurrence of the rare event that constitutes a supernova explosion and allows repeated observations of the explosion with the possibility to record the beginning of the explosion, provided that our algorithms for predicting the reappearance of an object are precise enough. Magnification of individual stars, allows to probe the first population of stars, but also has the potential to probe the distribution of matter at small scales through possible microlensing event. Coming data-sets provided by large surveys and long term monitoring of lensed quasars also have the potential to revolutionise our field by providing better statistics and constraints on our models for dark matter and cosmology in general.

A Filtering of MUSE spectra

The following paper is a simple application of wavelet filtering, developed for the needs of filtering MUSE spectra. The goal is to differentiate between features in MUSE spectra that belong to noise and features that represent emission lines.

IFS-RedEx, a redshift extraction software for integral-field spectrographs: Application to MUSE data

Markus Rexroth[1], Jean-Paul Kneib[1,2], Rémy Joseph[1], Johan Richard[3], Romaric Her[1,4]

[1] *Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland*
[2] *Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France*
[3] *Univ Lyon, Univ Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France*
[4] *ISAE-SUPAERO, Université de Toulouse, 10 Avenue Edouard Belin, 31400 Toulouse, France*

ABSTRACT

We present IFS-RedEx, a spectrum and redshift extraction pipeline for integral-field spectrographs. A key feature of the tool is a wavelet-based spectrum cleaner. It identifies reliable spectral features, reconstructs their shapes, and suppresses the spectrum noise. This gives the technique an advantage over conventional methods like Gaussian filtering, which only smears out the signal. As a result, the wavelet-based cleaning allows the quick identification of true spectral features. We test the cleaning technique with degraded MUSE spectra and find that it can detect spectrum peaks down to $S/N \approx 8$ while reporting no fake detections. We apply IFS-RedEx to MUSE data of the strong lensing cluster MACSJ1931.8-2635 and extract 54 spectroscopic redshifts. We identify 29 cluster members and 22 background galaxies with $z \geq 0.4$. IFS-RedEx is open source and publicly available.

Key words: Techniques: Imaging spectroscopy – Techniques: Image processing – Galaxies: clusters: individual: MACSJ1931.8-2635 – Galaxies: high-redshift

1 INTRODUCTION

Astrophysical research has benefited greatly from publicly available open source software and programs like SExtractor (Bertin & Arnouts 1996) and Astropy (Astropy Collaboration et al. 2013) have become standard tools for many astronomers. Their public availability allows researchers to focus on the science and to reduce the programming overhead, while the open source nature facilitates the code's further development and adaptation. In this spirit, we developed the Integral-Field Spectrograph Redshift Extractor (IFS-RedEx), an open source software for the efficient extraction of spectra and redshifts from integral-field spectrographs[1]. The software can also be used as a complement to other tools such as the Multi Unit Spectroscopic Explorer (MUSE) Python Data Analysis Framework (mpdaf)[2].

Our redshift extraction tool includes a key feature, a wavelet-based spectrum cleaning tool which removes spurious peaks and reconstructs a cleaned spectrum. Wavelet

transformations are well suited for astrophysical image and data processing (see e.g. Starck & Murtagh 2006 for an overview) and have been successfully applied to a variety of astronomical research projects. To name only a few recent examples, wavelets have been used for source deblending (Joseph et al. 2016), gravitational lens modeling (Lanusse et al. 2016) and the removal of contaminants to facilitate the detection of high redshift objects (Livermore et al. 2016).

The paper is designed as follows: Sections 2 and 3 present the spectrum and redshift extraction routines of IFS-RedEx. In section 4, we describe and test the wavelet-based spectrum cleaning tool. In section 5, we illustrate the use of our software by applying it to MUSE data of the strong lensing cluster MACSJ1931.8-2635 (henceforth called MACSJ1931). We summarize our results in section 6.

2 SPECTRUM EXTRACTION & CATALOG CLEANING

It is advantageous to combine Integral-Field Unit (IFU) data cubes with high resolution imaging, as this allows us to detect small, faint sources which might remain undetected if we used the image obtained by collapsing the data cube

---

[1] The software can be downloaded at http://lastro.epfl.ch/software
[2] Available at https://git-cral.univ-lyon1.fr/MUSE/mpdaf

along the wavelength axis (henceforth called white-light image) for source detection. For example, Bacon et al. (2015) used this combination in their analysis of MUSE observations of the *Hubble* Deep Field South. Therefore we exploit this case in the following, but in principle the software can be used without high resolution data. IFS-RedEx uses the center positions of stars provided by the user to align the IFU and high resolution images. It utilizes a SExtractor (Bertin & Arnouts 1996) catalog of the high-resolution data to extract the spectra and the associated standard deviation noise estimate for each source from the data cube. It extracts the signal in an area with a radius of 3 to 5 data cube pixels, depending on the SExtractor full width at half maximum (FWHM) estimate. Sources with FWHM < 2 high resolution pixels are discarded as these are typically spurious detections, e.g. due to cosmic rays.

IFS-RedEx shows the user each source and extraction radius overplotted on the high resolution image and the IFU data cube. The user can now quickly examine each detection and decide to either keep it in the database or to remove it, for example because it is too close to the data cube boundary and suffers from edge effects.

The tool also supports line emission and continuum emission catalogs. These are for example created by the MUSELET[3] software, which uses narrow-band images to perform a blind search for the respective signal. IFS-RedEx displays the detected sources and their extraction radius of 3 pixels on the IFU data cube. The user labels sources which cannot be used, e.g. because the signal is only a spurious detection in one pixel or it is too close to the image boundary. The spectra and noise of the good sources are automatically extracted.

Finally, the cleaned SExtractor, line emission, and continuum emission catalogs are merged into a master catalog. In this step, the sources are displayed on the high-resolution image so that the user can decide if the MUSELET and SExtractor detections are part of the same source. This visual inspection is more reliable than an automatic association and the number of sources is typically small enough for a manual inspection in reasonable time.

3   REDSHIFT EXTRACTION

Each 1D spectrum is displayed in an interactive plot and a second window shows the corresponding high resolution image, see figure 1. The position of sky lines with a flux $\geq 50 \times 10^{-20} \ \mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}}$ are labeled in green. The sky line fluxes are taken from Cosby et al. (2006). IFS-RedEx also lists the emission line identifications from MUSELET if available.

The user can now adjust the position of the emission and absorption line template by changing the source

---

[3] MUSELET is part of the mpdaf package. A tutorial and the documentation are available at http://mpdaf.readthedocs.io/en/latest/muselet.html

redshift. Once the template matches the source spectrum, the right redshift is found. IFS-RedEx has several features to facilitate the correct identification of spectral features. The user can zoom in and out, overplot the noise on the spectrum, smooth the signal with a Gaussian filter and perform a wavelet-based spectrum cleaning, see figure 1. When IFS-RedEx plots the noise, it shows the standard deviation around an offset. The offset is calculated by smoothing the spectrum signal with a Gaussian with $\sigma = 100$ pixels. Thus the noise is centered on the smoothed signal and it follows signal drifts. The wavelet cleaning is described in detail in the next section. As can be seen in figure 1, it reconstructs the shape of the reliable spectrum features and suppresses the noise. The Gaussian filter only smears out the signal. Thus the wavelet-based reconstruction makes it easier to distinguish true from false peaks.

Finally, the user can fit a Gaussian to the most prominent spectral line. IFS-RedEx combines the error of the fitted center position with the wavelength calibration error from the IFU data reduction pipeline into the final statistical redshift error. The software creates a final catalog with all source redshifts and errors. In addition, it produces a document with all spectral feature identifications and high resolution images for later use, e.g. for verification by a colleague.

4   WAVELET-BASED SPECTRUM RECONSTRUCTION

4.1   Wavelet transform algorithms

The wavelet-based cleaning algorithm reconstructs only spectral features above a given significance threshold. For this purpose, we use the "à trous" wavelet transform with a B_3-spline scaling function of the coordinate $x \in \mathbb{R}$,

$$\phi(x) = \frac{1}{12}(|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3), \quad (1)$$

which is well suited for isotropic signals such as emission lines (Starck et al. 2007; Starck & Murtagh 2006; Holschneider et al. 1989). In contrast to a Fourier transform, wavelets possess both frequency and location information. We note that the measured spectrum signal is discrete and not continuous and we denote the unprocessed, noisy spectrum data c_0, where the subscript indicates the scale s, and its value at pixel position l with $c_{0,l}$. We assume that $c_{0,l}$ is the scalar product of the continuous spectrum function $f(x)$ and $\phi(x)$ at pixel l. Now we can filter this data, where each filtering step increases s by one and leads to c_{s+1}, which no longer includes the highest frequency information from c_s. The filtered data for each scale is calculated by using a convolution. The coefficients of the convolution mask h derive from the scaling function,

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_l h(l)\phi(x-l), \quad (2)$$

and they are (1/16, 1/4, 3/8, 1/4, 1/16) (Starck & Murtagh 2006). By noting that $h(k)$ is symmetric (Starck et al. 2007), we have

$$c_{s,l} = \sum_k h(k)c_{s-1,l+2^{s-1}k} \quad (3)$$

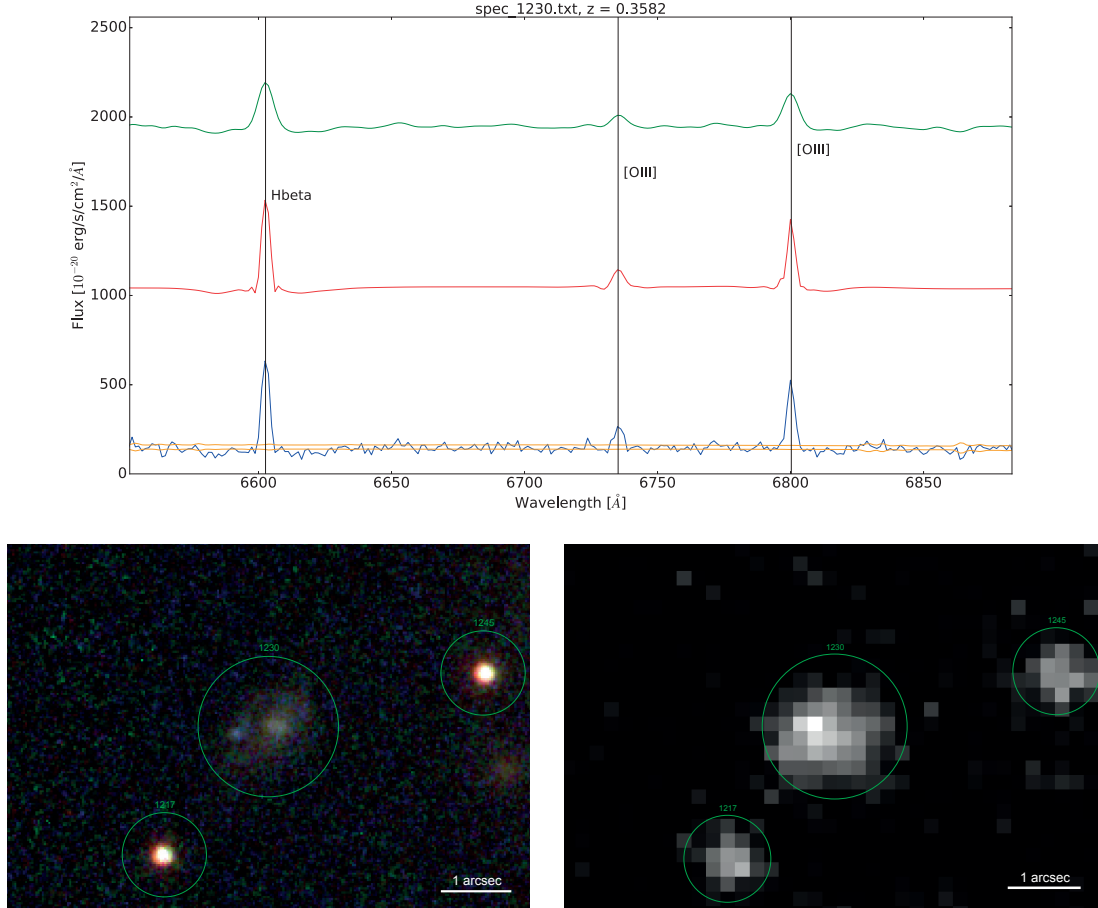# Appendix A. Filtering of MUSE spectra

Figure 1. Top: Interactive spectrum plot. The user can apply Gaussian filtering (green), wavelet cleaning (red) and plot the noise (yellow) to distinguish real from spurious features of the data (blue). The offsets of the plots can be adjusted. The noise shows the upper and lower standard deviation around the smoothed signal (see text). Bottom left: IFS-RedEx displays the high resolution color image for each source to facilitate the redshift extraction. The respective source is always at the image center and labeled with the number of the source's spectrum file, here 1230. Bottom right: MUSE data cube slice at 6799.97 Angstrom corresponding to the high resolution image on the left. In the spectrum fitting step the data cube is typically not needed and thus it is not displayed by default, but it can be quickly loaded via the DS9 interface if required.

and we define the double-convolved data on the same scale by

$$cd_{s,l} = \sum_k h(k)c_{s,l+2^{s-1}k}. \tag{4}$$

The wavelet coefficients are now given by

$$w_{s,l} = c_{s-1,l} - cd_{s,l}, \tag{5}$$

and they include the information between these two scales (Starck et al. 2016). A low scale s implies high frequencies and vice versa. The final wavelet transform is the set $\{w_1, \ldots, w_L, c_L\}$, where L is the highest scale level we use, and it includes the full spectrum information. We impose an upper limit for L depending on the spectrum wavelength range and resolution: $L \leq \log_2((P-1)/(H-1))$, where P is the number of pixels of the spectrum signal and H the length of $\boldsymbol{h}$, which is in our case $H = 5$. Otherwise s could become so large that the filtering equation 3 would require data outside of the wavelength range. We compute the wavelet transform according to algorithm 1 and we transform back into real space by using algorithm 2 (Starck et al. 2016).

The cleaning in wavelet space is performed following Starck & Murtagh (2006): We transform a discretized Dirac δ-distribution to obtain the wavelet set $\{w_1^\delta, \ldots, w_L^\delta\}$. Subsequently, we convolve each squared $\boldsymbol{w}_s^\delta$ with the squared standard deviation spectrum noise extracted from

---

Algorithm 1 Transform the spectrum into wavelet space

Require: Spectrum c_0 (= set of discrete spectrum pixels $\{c_{0,l}\}$), highest scale level L, convolution mask h
Output: Wavelet transform of spectrum $\{w_1, \ldots, w_L, c_L\}$
1: **Procedure** WAVELET_TRANSFORM(c_0, L):
2: $s \leftarrow 0$
3: **while** $s < L$ **do**
4:     $s \leftarrow s + 1$
5:     $c_{s,l} \leftarrow \sum_k h(k) c_{s-1,l+2^{s-1}k} \ \forall l$
6:     $cd_{s,l} \leftarrow \sum_k h(k) c_{s,l+2^{s-1}k} \ \forall l$
7:     $w_{s,l} \leftarrow c_{s-1,l} - cd_{s,l} \ \forall l$
8: **end while**
9: **return** $\{w_1, \ldots, w_L, c_L\}$

---

the IFU data cube and take the square root of the result. This gives us the noise coefficients in wavelet space.

In the next step, we build the multiresolution support $\mathbf{M}$, which is a $(L+1) \times P$ matrix. We compare the absolute value of the signal and noise wavelet coefficients at each pixel, $w_{s,l}$ and $w_{s,l}^N$. We take a threshold T set by the user, for example 5 for a 5σ cleaning in wavelet space, and set the corresponding matrix entry in $\mathbf{M}$ to 1 if $|w_{s,l}| \geq T|w_{s,l}^N|$, and 0 otherwise. Note that for $s = 1$, we use a higher threshold of $T + 1$, as this wavelet scale corresponds to high frequencies, where we expect the noise to dominate. The matrix coefficients for the smoothed signal c_L are automatically set to 1.

Now we perform the cleaning: We set all $w_{s,l}$ associated with a vanishing $\mathbf{M}$ value to zero and transform back into real space to obtain a first clean spectrum. However, there is still some signal to be harnessed in the residuals. Therefore we subtract the clean spectrum from the full spectrum to obtain the residual spectrum, and we compare its standard deviation, σ_{res}, with the standard deviation of the full spectrum (in the first iteration) or of the residual used in the previous iteration (all subsequent iterations), which we indicate in both cases with σ_{prev}. If $|(\sigma_{\text{prev}} - \sigma_{\text{res}})/\sigma_{\text{res}}| > \epsilon$, we transform the residual spectrum into wavelet space, set wavelets with vanishing $\mathbf{M}$ values to zero, transform back into real space, and add the resulting signal to obtain our new clean signal. Note that the same multiresolution support as before is used. Subsequently, we calculate again the residual and continue until the ϵ criterion is no longer fulfilled and all the signal has been extracted. The value of ϵ is set by the user and must satisfy the condition $0 < \epsilon < 1$. Algorithm 3 summarizes this cleaning procedure.

4.2 Testing the wavelet-based reconstruction

To test our software, we use the spectrum of the brightest cluster galaxy (BCG) from our MUSE data set described in the next section. MUSE provides both the spectrum signal and a noise estimate over the full wavelength range. The original spectrum can be considered clean due to its

---

Algorithm 2 Transformation from wavelet to real space

Require: Wavelet transform of spectrum $\{w_1, \ldots, w_L, c_L\}$, highest scale level L, number of spectrum pixels P, convolution mask h
Output: Spectrum in real space c_0 (= set of discrete spectrum pixels $\{c_{0,l}\}$)
1: **Procedure:** WAVELET_BACKTRANSFORM($\{w_1, \ldots, c_L\}$):
2: $S \leftarrow c_L$
3: **for all** $s \in \{1, \ldots, L\}$ **do**
4:     **for all** $l \in \{1, \ldots, P\}$ **do**
5:         $C_l \leftarrow \sum_k h(k) \, S_{l+2^{L-s}k}$
6:     **end for**
7:     $S \leftarrow C + w_{L+1-s}$
8: **end for**
9: $c_0 \leftarrow S$
10: **return** c_0

Algorithm 3 Signal cleaning in wavelet space

Require: Spectrum c_0 (= set of discrete spectrum pixels $\{c_{0,l}\}$), σ_{spec} (= vector with standard deviation noise estimate for each spectrum pixel), highest scale level L, number of spectrum pixels P, cleaning threshold T, cleaning parameter ϵ ($0 < \epsilon < 1$)
Output: Cleaned spectrum S_{clean}
1: **Procedure** CLEAN_SIGNAL(c_0, σ_{spec}, L, T, ϵ):
2: $\{w_1^\delta, \ldots, w_L^\delta, c_L^\delta\} \leftarrow$ WAVELET_TRANSFORM(δ-dist., L)
3: **for all** $w_s^\delta \in \{w_1^\delta, \ldots, w_L^\delta\}$ **do**
4:     $w_s^N \leftarrow \sqrt{w_s^{\delta 2} * \sigma_{\text{spec}}^2}$
5: **end for**
6: $\{w_1, \ldots, w_L, c_L\} \leftarrow$ WAVELET_TRANSFORM(c_0, L)
7: $\mathbf{M} \leftarrow \mathbf{0}_{L+1,P}$ // Multiresolution support matrix
8: **for all** $s \in \{1, \ldots, L+1\}, l \in \{1, \ldots, P\}$ **do**
9:     **if** $s == 1$ **and** $|w_{s,l}| \geq (T+1)|w_{s,l}^N|$ **then**
10:        $M_{sl} \leftarrow 1$
11:    **else if** $1 < s \leq L$ **and** $|w_{s,l}| \geq T|w_{s,l}^N|$ **then**
12:        $M_{sl} \leftarrow 1$
13:    **else if** $s == L + 1$ **then**
14:        $M_{sl} \leftarrow 1$
15:    **end if**
16: **end for**
17: $S_{\text{clean}} \leftarrow \mathbf{0}_P$, $\sigma_{\text{prev}} \leftarrow 0$, $res \leftarrow c_0$
18: $\sigma_{\text{res}} \leftarrow \text{std}(res)$
19: **while** $|(\sigma_{\text{prev}} - \sigma_{\text{res}})/\sigma_{\text{res}}| > \epsilon$ **do**
20:    $\{w_1^{\text{res}}, \ldots, c_L^{\text{res}}\} \leftarrow$ WAVELET_TRANSFORM(res, L)
21:    **for all** $s \in \{1, \ldots, L\}, l \in \{1, \ldots, P\}$ **do**
22:        **if** $M_{sl} == 0$ **then**
23:            $w_{s,l}^{\text{res}} \leftarrow 0$
24:        **end if**
25:    **end for**
26:    $res_{\text{clean}} \leftarrow$ WAVELET_BACKTRANSFORM($\{w_1^{\text{res}}, \ldots, c_L^{\text{res}}\}$)
27:    $S_{\text{clean}} \leftarrow S_{\text{clean}} + res_{\text{clean}}$
28:    $res \leftarrow c_0 - S_{\text{clean}}$
29:    $\sigma_{\text{prev}} \leftarrow \sigma_{\text{res}}$
30:    $\sigma_{\text{res}} \leftarrow \text{std}(res)$
31: **end while**
32: **return** S_{clean}

Appendix A. Filtering of MUSE spectra

very high signal-to-noise. We rescale it to simulate fainter sources at low signal-to-noise. We calculate the rescaling factor R by looking at the highest spectrum signal peak and dividing the associated MUSE noise estimate by this signal. This results in $R \approx 0.0015$. We investigate three cases, namely a good, an intermediate, and a low signal-to-noise case, where we rescale the full signal spectrum by $10R$, $5R$, and $2R$ respectively. Subsequently we add Gaussian noise simulating the real noise estimate of the MUSE data cube. For each spectral wavelength pixel l we obtain the realized noise by drawing from a Gaussian probability distribution with a standard deviation equal to the MUSE standard deviation noise estimate at this pixel. We repeat this process 10 times to obtain spectra with different noise realizations.

We calculate the signal-to-noise of six emission lines by summing over their respective wavelength ranges,

$$\frac{\text{S}}{\text{N}} = \frac{\sum_l \text{signal}_l}{\sqrt{\sum_{l'} \text{std}_{l'}^2}}, \qquad (6)$$

where std_l is the MUSE standard deviation noise estimate at pixel l. We will refer to the lines according to their wavelength order, i.e. the first line is situated at the lowest wavelength and the last line at the highest. We apply our wavelet cleaning software to the spectra using the MUSE noise estimate and different wavelet parameters as input. We investigate 5σ and 3σ cleaning and ϵ parameters of 0.1, 0.01, and 0.001. The cleaning procedure is fast and takes about 1 second per spectrum on a laptop. Figure 2 shows reconstructed spectra for the three different signal-to-noise cases. Note that the last two emission lines in the true spectrum are actually comprised of merged individual lines. As can be seen in figure 2, the wavelet tool can detect if a line consists of two merged lines and reconstruct them correctly if their signal-to-noise is high enough. If it is too low, it will reconstruct them as a single line.

For all 90 spectra which we analyzed with a 5σ wavelet reconstruction, we find no fake detections of emission lines. For signal-to-noise larger than 20, all 6 test emission lines are detected. For S/N between 10 and 20, all emission lines but the third are found. The third peak is no longer recovered due to its proximity to the fourth peak, which has typically a twice larger S/N value. In general, the wavelet software might reconstruct two close-by peaks as a single peak unless they have each a sufficiently large signal. When the signal-to-noise of both peaks was similar, both the third and the fourth emission line were detected and reconstructed. For emission lines with low signal-to-noise values between 5 and 10, we can reconstruct the stronger lines with S/N $\gtrsim$ 8, while the weaker peaks remain typically undetected. However, as the bottom plot in figure 2 shows, even weaker peaks can occasionally be reconstructed.

Emission lines modeled with a wavelet reconstruction do sometimes not reach the full peak height of the signal, in particular for high ϵ values, and their tails can suffer from ringing effects which might be due to the wavelet shape, see for example the first emission line of the intermediate S/N case in figure 2. For low signal-to-noise emission lines
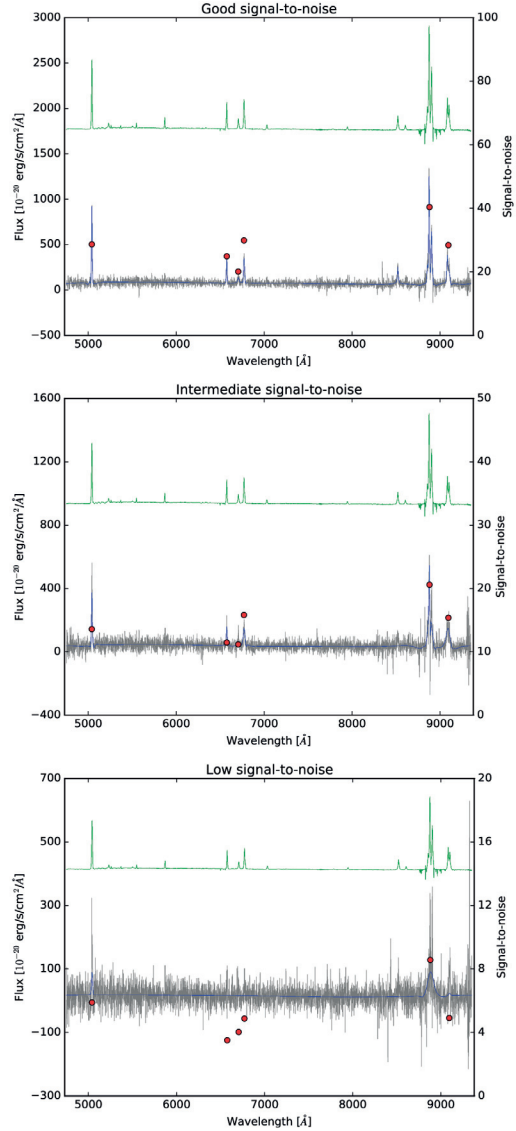
Figure 2. Reconstructed spectra for three different signal-to-noise scenarios using 5σ cleaning and $\epsilon = 0.01$. The true spectrum is offset and shown in green, the noisy spectrum is displayed in gray and the reconstructed signal in blue. Red dots indicate the signal-to-noise of the respective emission lines. The wavelet tool detects peaks with $S/N \gtrsim 8$. However, if a low S/N emission line is located very close to a high S/N line, it is possible that it will not be recovered (middle plot, 3rd emission line from the left). The wavelet reconstruction can occasionally even find lines with $S/N < 8$ (bottom plot, 1st and 6th emission line), but typically they will not be detected (bottom plot, 2nd - 4th emission line).
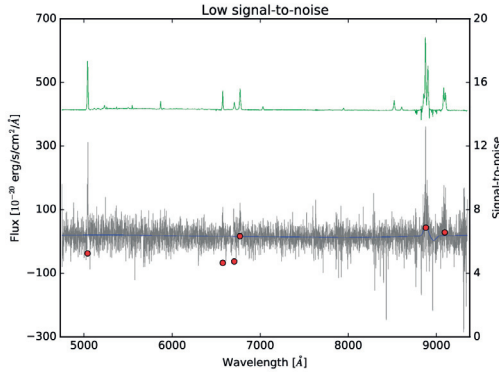
Figure 3. In low signal-to-noise (S/N ≤ 10) cases, ringing effects can occasionally lead to signal dips with similar amplitude as the signal peaks of the reconstructed emission line. Therefore care has to be taken not to mistake these effects for absorption lines. This might be ameliorated by re-running the wavelet reconstruction with a different setup. The colors have the same meaning as in figure 2 and the reconstruction was performed with a 5σ cutoff and $\epsilon = 0.01$.

(S/N ≤ 10), care has therefore to be taken not to mistake the signal dip due to ringing effects as an absorption signal, as the ringing effect might occasionally have a similar (negative) amplitude as the signal peak of the reconstructed emission line, see figure 3. When this effect occurs in practice, it might be improved by changing the wavelet setup, e.g. by lowering the ϵ value. A lower ϵ is designed to detect a larger fraction of the signal peak and should thus increase its height. However, care has to be taken as a lower ϵ might also lead to stronger ringing effects.

The 3σ wavelet reconstruction recovered more emission lines than the 5σ cleaning, but it also produced false detections. We therefore adopted a conservative approach and used the 5σ wavelet cleaning when applying the code to real data.

Finally, we compared the noise free emission line shapes with the reconstructed ones. We find that the shape reconstruction is generally good, but the reconstructed line shape and height recovered from the noisy data can differ from the original, clean ones, in particular in low signal-to-noise scenarios. Therefore we use the wavelet cleaning only to distinguish true from false spectrum peaks, and we perform all data operations such as fitting a Gaussian to obtain the centering error on the real, noisy data.

5   APPLICATION TO MUSE DATA: MACSJ1931

We apply IFS-RedEx to our data set of the strong lensing cluster MACSJ1931 obtained with MUSE (Bacon et al. 2010) on the Very Large Telescope (VLT). We combine our data with the publicly available *Hubble Space Telescope* (HST) imaging from the Cluster Lensing And Supernova survey with Hubble (CLASH, Postman et al. 2012). The

cluster is part of the MAssive Cluster Survey (MACS), which comprises more than one hundred highly X-ray luminous clusters (Ebeling et al. 2010, 2001).

The core of MACSJ1931 (z = 0.35) was observed with MUSE on June 12 and July 17 2015 (ESO program 095.A-0525(A), PI: Jean-Paul Kneib). The 1 x 1 arcmin^2 field of view was pointed at α = 19:31:49.66 and δ = -26:34:34.0 (J2000) and we observed for a total exposure time of 2.44 hours, divided into 6 exposures of 1462 seconds each. We rotated the second exposure of each exposure pair by 90 degrees to allow for cosmic ray rejection and improve the overall image quality. The data were taken using the WFM-NOAO-N mode of MUSE in good seeing conditions with FWHM ≈ 0.7 arcseconds.

We reduced the data using the MUSE pipeline version 1.2.1 (Weilbacher et al. 2014, 2012), which includes bias and flat-field corrections, sky subtraction, and wavelength and flux calibrations. The six individual exposures were finally combined into a single data cube and we subtracted the remaining sky residuals with ZAP (Soto et al. 2016). The wavelength range of the data cube stretches from 4750 to 9351 Å in steps of 1.25 Å. The spatial pixel size is 0.2 arcseconds.

We used the HST data for MACSJ1931 obtained as part of the CLASH program (Zitrin et al. 2015) in the bands F105W, F475W, F625W, and F814W with a spatial sampling of 0.03 arcsec/pixel. The HST data products are publicly available on the CLASH website[4].

We use only redshift identifications which we consider secure because we see e.g. several lines or a clear Lyα emission line shape. We extract 54 sources with redshifts ranging from 0.21 to 5.8. Among them, 29 are cluster members with $0.3419 \leq z \leq 0.3672$ and 22 are background sources with $0.4 \leq z \leq 5.8$. A table of all sources with spectroscopic redshifts is presented in the companion paper Rexroth et al. 2017 (in preparation), in which we use the data to improve the cluster lens model. Figure 4 shows a histogram of the source distribution in redshift space.

6   SUMMARY

We describe IFS-RedEx, a public spectrum and redshift extraction pipeline for integral-field spectrographs. The software supports SExtractor catalogs as well as MUSELET narrow-band detection catalogs as input. The pipeline has several features which allow a quick identification of reliable spectrum features, most notably a wavelet-based spectrum cleaning tool. The tool only reconstructs spectral features above a given significance threshold. We test it with degraded MUSE spectra and find that it can detect spectral features with S/N ≳ 8. We find no fake detections in our test. Finally, we apply IFS-RedEx to a MUSE data cube of the strong lensing cluster MACSJ1931 and extract 54 spectroscopic redshifts.
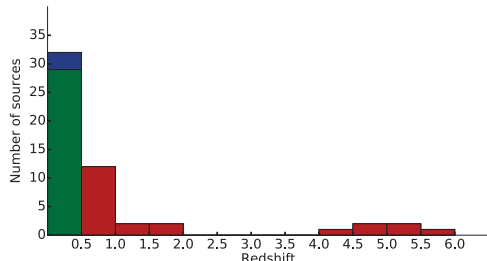
---

[4] https://archive.stsci.edu/prepds/clash/

Appendix A. Filtering of MUSE spectra

Figure 4. Source distribution in redshift space. Background sources are colored in red, cluster members in green and the remaining objects in blue.
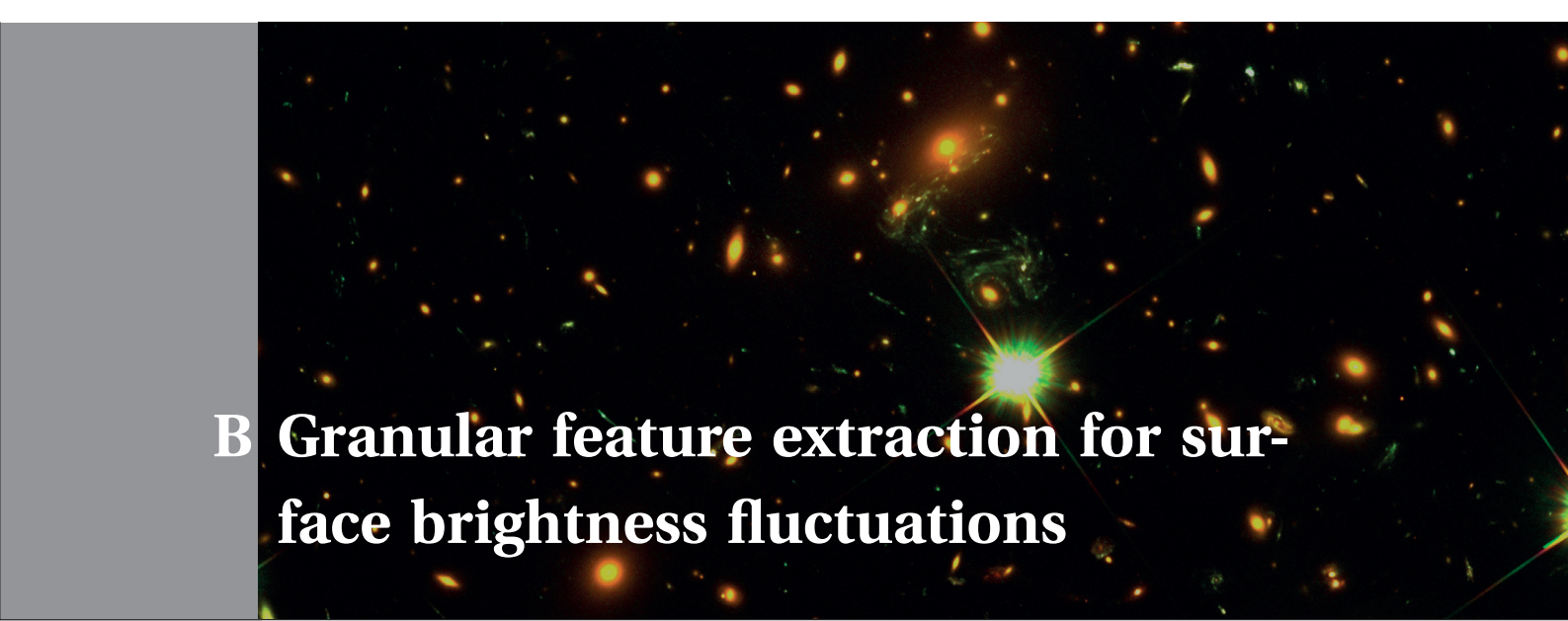
ACKNOWLEDGEMENTS

REFERENCES

Astropy Collaboration et al., 2013, A&A, 558, A33
Bacon R., et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III. p. 773508, doi:10.1117/12.856027
Bacon R., et al., 2015, A&A, 575, A75
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Cosby P. C., Sharpee B. D., Slanger T. G., Huestis D. L., Hanuschik R. W., 2006, Journal of Geophysical Research (Space Physics), 111, A12307
Ebeling H., Edge A. C., Henry J. P., 2001, ApJ, 553, 668
Ebeling H., Edge A. C., Mantz A., Barrett E., Henry J. P., Ma C. J., van Speybroeck L., 2010, MNRAS, 407, 83
Holschneider M., Kronland-Martinet R., Morlet J., Tchamitchian P., 1989, in Combes J.-M., Grossmann A., Tchamitchian P., eds, Wavelets. Time-Frequency Methods and Phase Space. p. 286
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python [Online; accessed 16.12.2016], http://www.scipy.org/
Joseph R., Courbin F., Starck J.-L., 2016, A&A, 589, A2

Lanusse F., Starck J.-L., Leonard A., Pires S., 2016, preprint, (arXiv:1603.01599)
Livermore R. C., Finkelstein S. L., Lotz J. M., 2016, preprint, (arXiv:1604.06799)
Postman M., et al., 2012, ApJS, 199, 25
Soto K. T., Lilly S. J., Bacon R., Richard J., Conseil S., 2016, MNRAS, 458, 3210
Starck J.-L., Murtagh F., 2006, Astronomical Image and Data Analysis, doi:10.1007/978-3-540-33025-7.
Starck J.-L., Fadili J., Murtagh F., 2007, IEEE Transactions on Image Processing, 16, 297
Starck J.-L., Murtagh F., Fadili J., 2016, Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis. Cambridge University Press
Tody D., 1986, in Crawford D. L., ed., Proc. SPIEVol. 627, Instrumentation in astronomy VI. p. 733
Weilbacher P. M., Streicher O., Urrutia T., Jarno A., Pécontal-Rousset A., Bacon R., Böhm P., 2012, in Software and Cyberinfrastructure for Astronomy II. p. 84510B, doi:10.1117/12.925114
Weilbacher P. M., Streicher O., Urrutia T., Pécontal-Rousset A., Jarno A., Bacon R., 2014, in Manset N., Forshay P., eds, Astronomical Society of the Pacific Conference Series Vol. 485, Astronomical Data Analysis Software and Systems XXIII. p. 451 (arXiv:1507.00034)
Zitrin A., et al., 2015, ApJ, 801, 44
van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science and Engineering, 13, 22

This paper has been typeset from a TeX/LaTeX file prepared by the author.

B Granular feature extraction for surface brightness fluctuations

Surface brightness fluctuation (SBF) is a technique for measuring distances to galaxies based on the mottling observed on high signal to noise images of galaxies light profile (Tonry & Schneider, 1988; Tonry & Schechter, 1990). The mottling is due to the variation of the number of stars from pixel-to-pixel. This results in a signal at the scale of the PSF in the Fourier domain. This signal can be measured by taking the power spectrum of the image after removing the possible contamination from the low frequency signal coming from the smooth profile of the galaxy. To this end, I developed a code called `Granulet` that can remove the smooth component of galaxies. To do this I developed a simple MCA algorithm based on the discrimination between low and high frequencies in the starlet domain.

B.1 Filtering low and high frequencies

The algorithm is an adaptation from algorithm 1, where we aim at extracting two components that are both sparsely represented in the starlet domain, but at different scales. To use this criteria as a discrimination between low and high frequency signal, I designed two dictionaries, Φ_1 and Φ_2 defined as: $\Phi_1 = D_L \Phi_S$ and $\Phi_2 = D_H \Phi_S$, where Φ_S is the starlet dictionary and D_L and D_H are respectively low and high-pass filters. A possible choice for the high-pass filter is to set to one all coefficients at a starlet scale below scale s and to 0 all coefficients in scales above s. In practice, we found that using a sigmoid profile for the discrimination between low and high frequencies prevented the contamination of the low frequencies by borderline low frequencies belonging to small compact objects such as small background galaxies. The low-pass filter is simply, at each starlet scale: $D_L(s) = 1 - D_H(s)$. At this early stage of the method, the scale limit s is still an arbitrary parameter chosen by the user to optimise the separation. In figure B.1, I show an example of the low and high-pass filter gains as a function of starlet scales.

In the following figures, B.2 and B.3, I show the result of the application of `Granulet` on two galaxies (*top left panels*) where we see the clear separation between high frequencies (*lower left panels*), high frequencies (*lower right panels*) and the residual noise (*top right panels*).
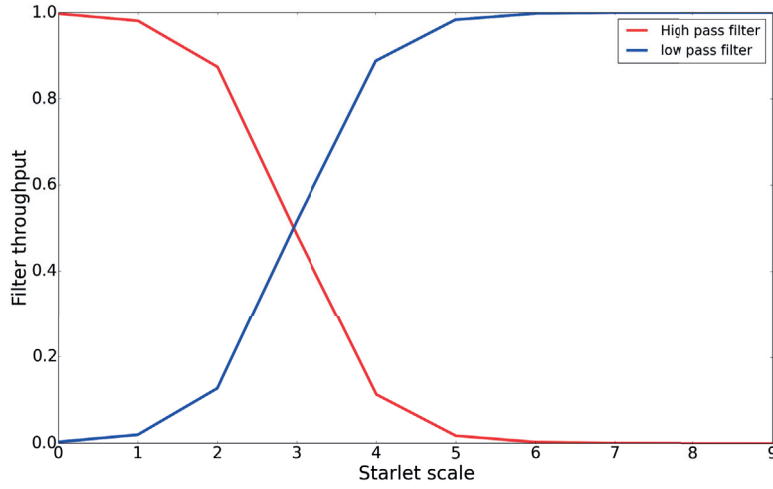
Figure B.1 – SBF low and high-pass filters.

B.2   Removing dust features

In another application, we applied our method to a galaxy that shows dust features in absorption in the image of galaxy vcc1615. Absorption by dust can be seen as an negative component, due to the absorption, added to the light profile of an otherwise smooth galaxy. In figure B.4, I show the result of applying `Granulet` directly to the image of vcc1615. The result shows negative features in the residuals, but we also notice absorption features in the smooth component.

In order to circumvent this issue, I add a third component to the modelling of the data for which I impose a negativity criteria, meaning that all the pixels above zero in this component are set to zero. One of the disadvantages of this approach is that the negative component might try to compensate for the positive components, especially around extremely bright structures, leading to overestimated positive components and unnecessary negative components around positive structures. While this has to be accounted for before being able to produce any consistent result with such approach, I applied this naive modification to the algorithm, which led to the results in figure B.5. In this figure, the bottom panel shows the dust in absorption, meaning that the intensity in a pixel is a measure of how much light is absorbed due to dust. While we see ring-like structures form around bright objects in the field of view, we also see that the morphology of the dust structures are well reproduced, the low frequency component is smoother and the residuals present less structures that can be imputed to dust.

(a) Original image

(b) Residuals

(c) High frequencies
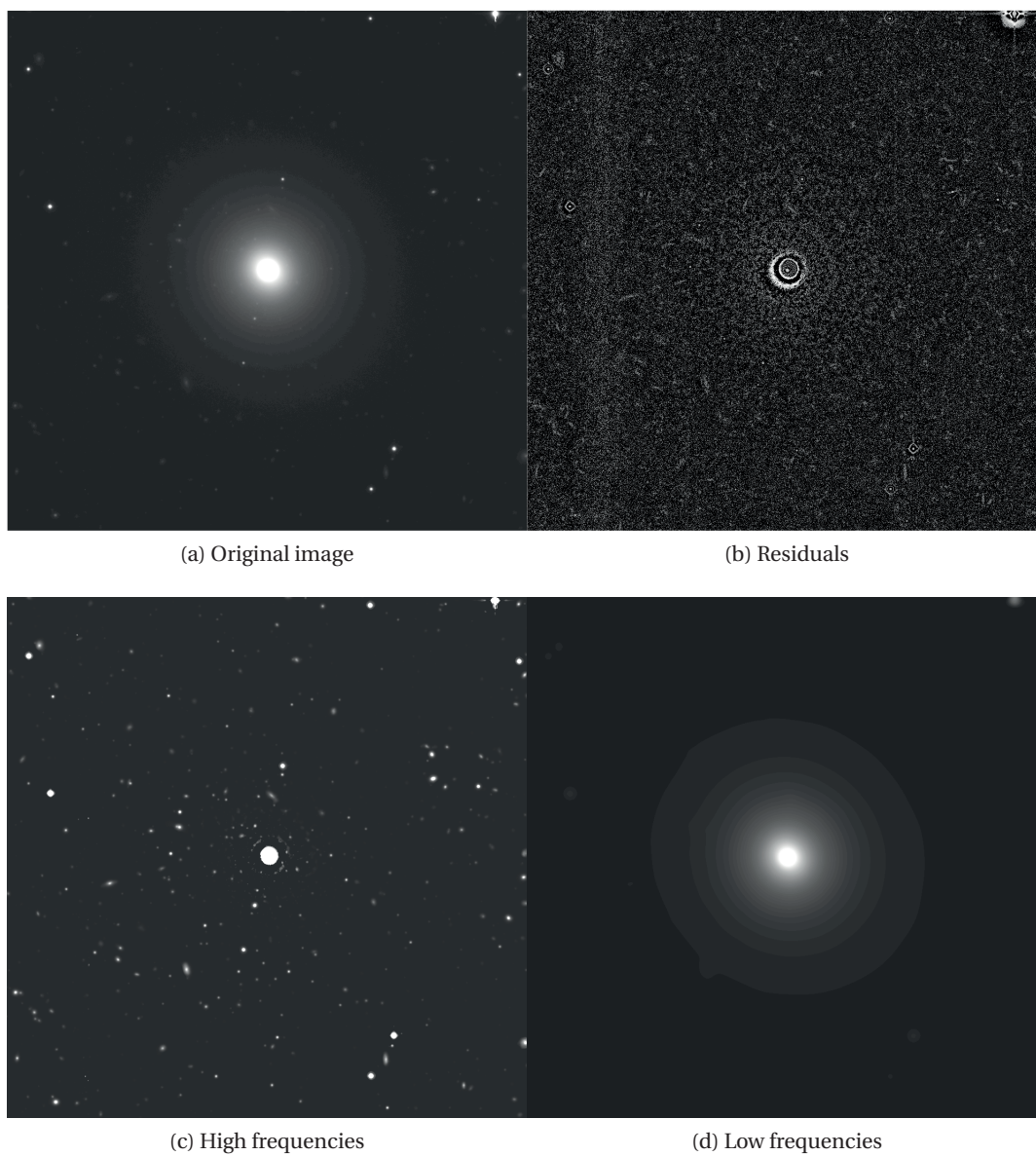
(d) Low frequencies

Figure B.2 – Low-high frequency separation in galaxy vcc1025

(a) Original image

(b) Residuals

(c) High frequencies
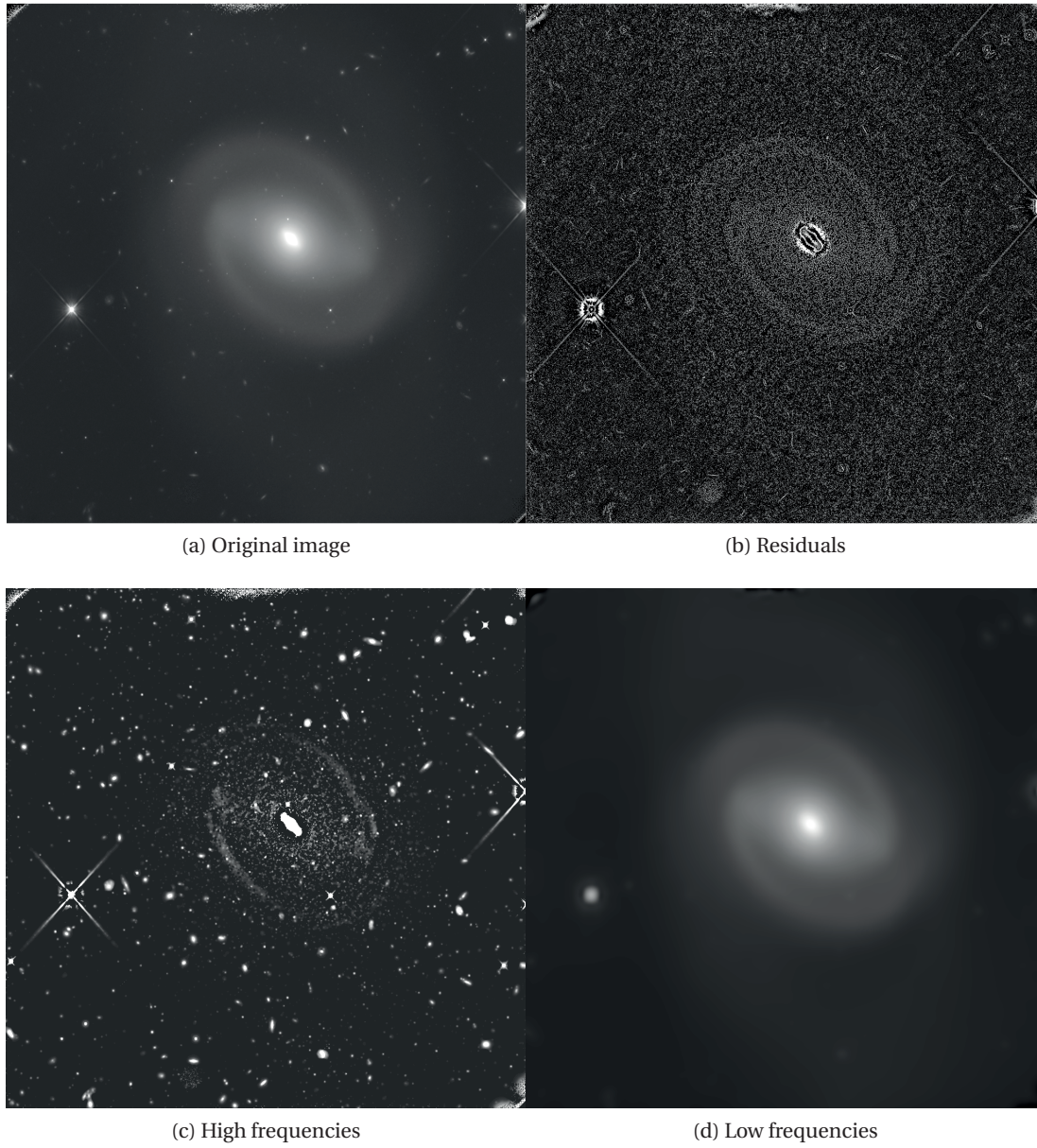
(d) Low frequencies

Figure B.3 – Low-high frequency separation in galaxy n0495

(a) Original image

(b) Residuals

(c) High frequencies
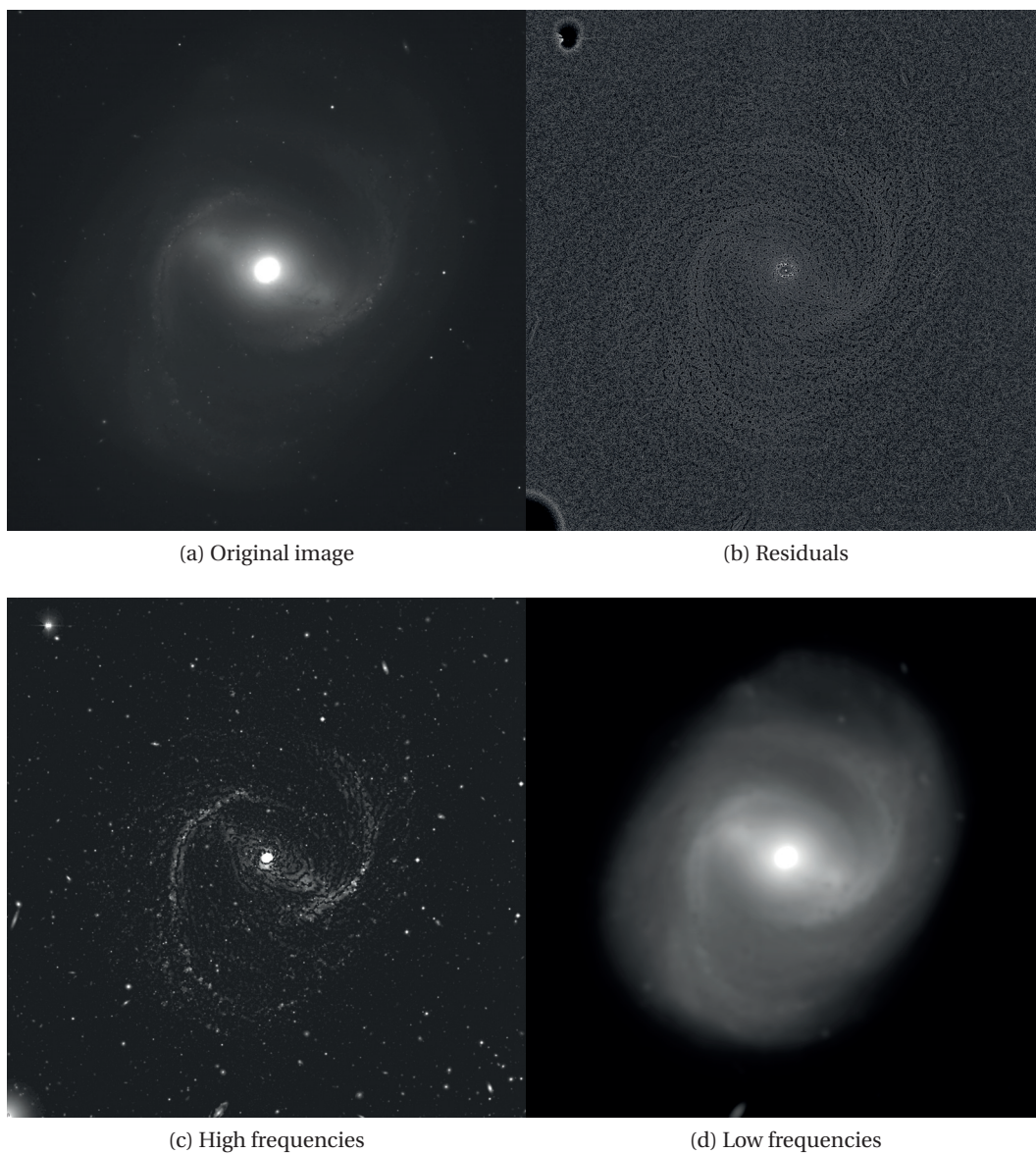
(d) Low frequencies

Figure B.4 – Low-high frequency separation in galaxy vcc1615
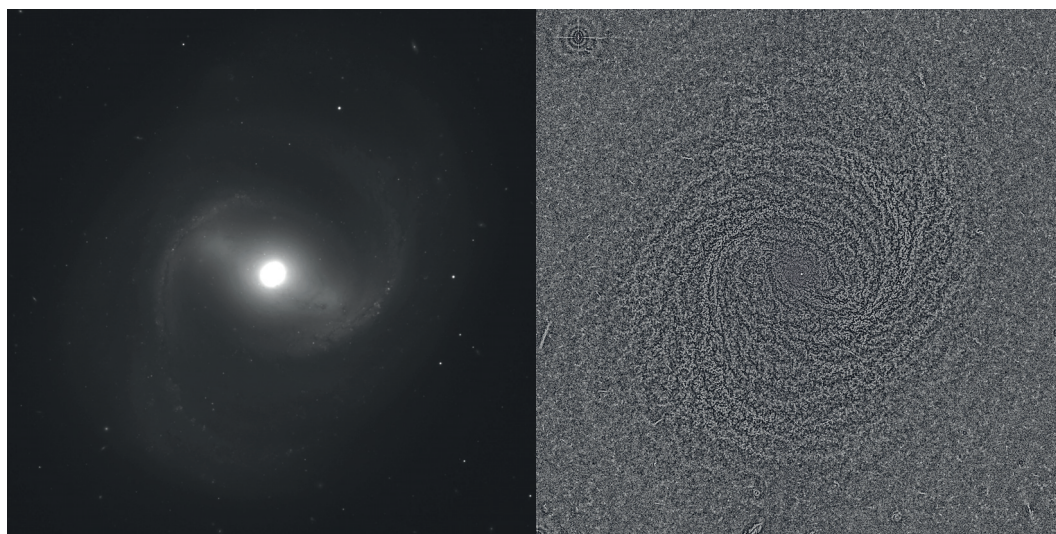
(a) Original image

(b) Residuals
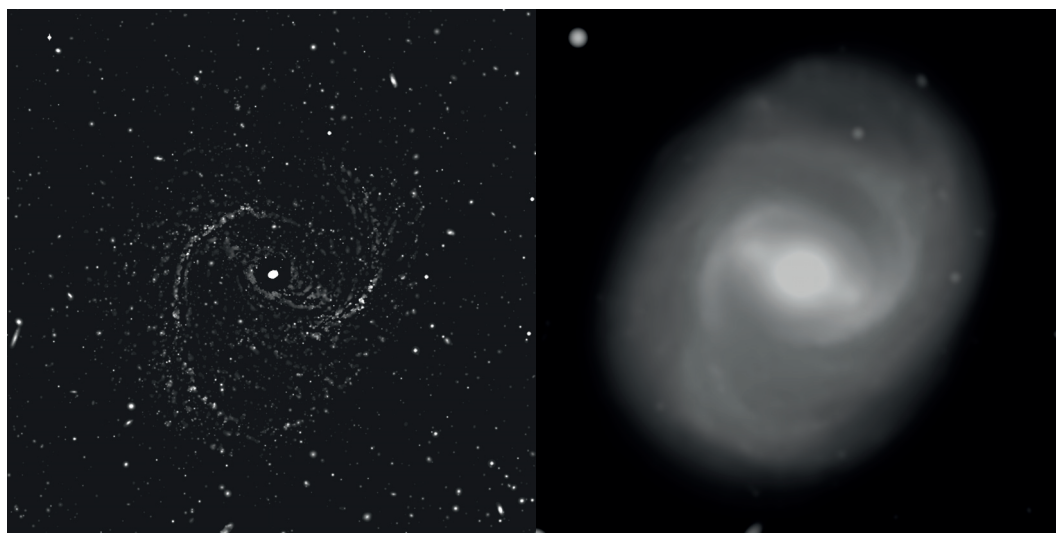
(c) High frequencies

(d) Low frequencies
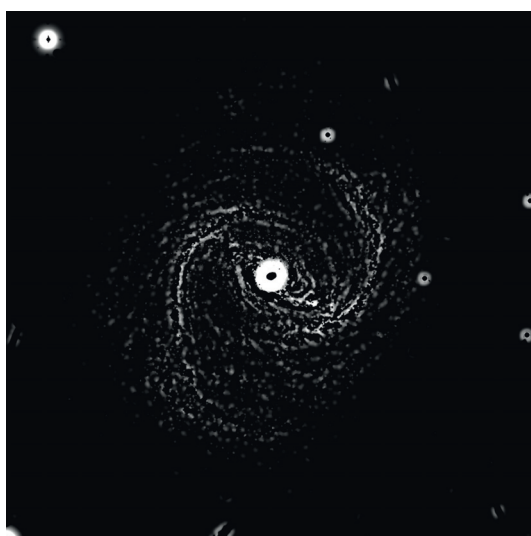
(e) Dust absorption

Figure B.5 – Low-high frequency and negative component separation in galaxy vcc1615

C Colour deblending of Abell 3827

A test for skewed distributions of dark matter, and a possible detection in galaxy cluster Abell 3827

Peter Taylor[1,2]*, Richard Massey[1,3], Mathilde Jauzac[1,3,4], Frédéric Courbin[5],
David Harvey[5], Rémy Joseph[5] & Andrew Robertson[3]

[1] *Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK*
[2] *Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK*
[3] *Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[4] *Astrophysics and Cosmology Research Unit, School of Mathematical Sciences, University of KwaZulu-Natal, Durban 4041, South Africa*
[5] *Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne, Observatoire de Sauverny, CH-1290 Versoix, Switzerland*

ABSTRACT

Simulations of self-interacting dark matter (SIDM) predict that dark matter should lag behind galaxies during a collision. If the interaction is mediated by a high-mass force carrier, the distribution of dark matter can also develop asymmetric dark matter tails. To search for this asymmetry, we compute the gravitational lensing properties of a mass distribution with a free *skewness* parameter. We apply this to the dark matter around the four central galaxies in cluster Abell 3827. In the galaxy whose dark matter peak has previously been found to be offset, we tentatively measure a skewness $s = 0.23^{+0.05}_{-0.22}$ in the same direction as the peak offset. Our method may be useful in future gravitational lensing analyses of colliding galaxy clusters and merging galaxies.

Key words: dark matter — astroparticle physics — galaxies: clusters: individual: Abell 3827 — gravitational lensing: strong

1 INTRODUCTION

Most of the mass in the Universe is dark matter (e.g. Planck Collaboration 2016). Dark matter appears invisible, because it does not interact (or interacts very weakly) with Standard Model particles including photons.

As the nature of dark matter remains unknown, there is no reason to a priori assume a particular theory of its origin. The wide range of proposed dark matter models predict different spatial distributions, particularly on small scales. Dark matter particles that interact with each other (SIDM) were proposed in (Spergel & Steinhardt 2000) to explain small scale discrepancies between observations and simulations of collisionless dark matter. In the SIDM paradigm, energy transfer between particles makes the centre of galaxies (Vogelsberger et al. 2012) and galaxy clusters (Rocha et al. 2013) more circular and less dense, potentially resolving the core/cusp problem. Small substructures can also be erased – leading to the observed underabundance of galaxies in the Local Group, relative to simulations. During mergers between galaxies or galaxy clusters, dark matter interactions transfer momentum between the colliding dark matter haloes (Randall et al. 2008; Kahlhoefer et al. 2014;

Robertson et al. 2017; Kim et al. 2016; Robertson et al. 2016). These scatterings can temporarily separate dark matter from its associated galaxies. Such dark matter lags behind the galaxies, toward the position of diffuse gas that is slowed by ram pressure (Clowe et al. 2004; Lage & Farrar 2014; Harvey et al. 2015). Scattering processes during collisions can be seperated into two types: frequent low momentum transfer and infrequent high momentum transfer. These will have different qualitative behaviours.

Frequent low momentum transfer scattering will cause an effective drag force, which if greater than the gravitational restoring force, will seperate the entire DM halo from the galaxy during collisions. Crucially there will be no tail of scattered DM particles escaping the potential well (Kahlhoefer et al. 2014). Numerous studies have placed constraints on the cross-section of DM in this regime. Measuring dark matter galaxy offsets on a sample of 72 merging clusters, (Harvey et al. 2015) found $\tilde{\sigma}/m_{DM} < 0.5\,\mathrm{cm^2g^{-1}}$. Constraints from the Bullet Cluster place $\tilde{\sigma}/m_{DM} \gtrsim 1.2\,\mathrm{cm^2g^{-1}}$ (Kahlhoefer et al. 2014), while constraints from an offset galaxy in Abell 3827 yields $\tilde{\sigma}/m_{DM} \gtrsim 2.0\,\mathrm{cm^2g^{-1}}$ (Kahlhoefer et al. 2015).

In contrast infrequent high momentum transfer scattering (mediated by a high-mass force carrier, for example) will cause a small fraction of scattered particle to leave the potential well on the trailing side. Shortly after collision, this

* e-mail: `peterllewelyntaylor@gmail.com`

will appear as a tail of scattered DM particles (see Figure 5 in Kahlhoefer et al. 2014). Although the peak of the DM distribution will remain coincident with the galaxy, the tail of DM particles will lead to an apparant shift in the centre (Kahlhoefer et al. 2014).

Gravitational lensing offers the most direct way to map the spatial distribution of dark matter, and hence to infer its particle properties. Gravitational lensing refers to the deflection of light rays passing near any mass, including dark matter. Thanks to this deflection, (unrelated) objects behind dark matter appear characteristically distorted, or even visible along more than one (curved) line-of-sight. Even though dark matter is invisible, it is possible to invert this process and infer where it must be, by undistorting the observed images, or ray-tracing multiple images back onto each other.

Galaxy cluster Abell 3827 ($22^h\,01'\,49''.1\,-59°\,57'\,15''$, redshift z=0.099) is particularly well suited for this kind of study. It gravitationally lenses a z=1.24 galaxy with spiral arms and several knots of star formation that can be treated as independent background sources (Massey et al. 2015, hereafer M15). While most clusters contain only one brightest central galaxy, Abell 3827 contains four equally-bright galaxies within its central 10 kpc (Carrasco et al. 2010; Williams & Saha 2011). This highly unusual configuration means that some of the galaxies appear close to gravitationally lensed images. Thus, under parametric model assumptions, the distribution of the dark matter can be measured. Because of the cluster's relative proximity (in terms of gravitational lensing), it is possible to resolve small spatial offsets between the distribution of dark matter and stars in the foreground galaxies.

In this paper we present a new parametric lensing approach to search for the predicted asymmetry in the distribution of dark matter during mergers. A previous search looked for residuals after subtracting the symmetric component (Harvey et al. 2017), but that may be less sensitive because a tail of scattered particles shifts the best-fit position of the symmetric component backwards, thus removing some of the residual. We instead construct a single halo model with a free *skewness* parameter that qualitatively captures the asymmetry found in high momentum transfer scattering simulations. We implement and distribute this model in the publicly available LENSTOOL software[1] (Jullo et al. 2007). We test it on both mock data, where the skewness of the lens is known *a priori*, and on Abell 3827. Section 2 describes existing observations of Abell 3827. Section 3 introduces our new parametric lens model. Section 4 contains an analysis of Abell 3827. To be consistent with M15 we assume throughout this paper a flat ΛCDM cosmology with Ω_M=0.3, Ω_Λ=0.7 and H_0=70 km s^{-1} Mpc^{-1}. At the redshift of Abell 3827, $1''$ corresponds to 1.828 kpc.

2   DATA

Broad-band imaging of Abell 3827 has been obtained from the Gemini telescope at optical wavelengths (Carrasco et al. 2010) and from the *Hubble Space Telescope* (*HST*) in the UV, optical and near-infrared (M15). This revealed four

---

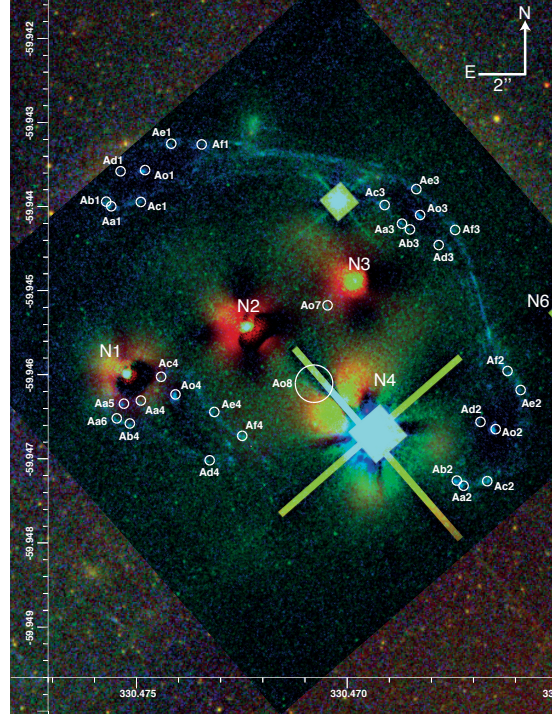[1] http://projets.lam.fr/projects/lenstool/wiki

Figure 1. *Hubble Space Telescope* image of galaxy cluster Abell 3827 in F160W (red), F814W (green) and F330W (blue) bands, after using MuSCADeT (Joseph et al. 2016) to fit and subtract foreground emission. Residual emission from two Milky Way stars has been masked out, and remains visible at low level around the four bright central galaxies N1–N4. Circles show multiple image identifications, with the radius of the circle reflecting uncertainty on their positions (Ao8 has only been detected from the ground).

similarly-bright elliptical galaxies within 10 kpc of each other, plus a background spiral galaxy, whose multiply-lensed images are threaded throughout the cluster core (figure 1).

Integral Field Unit spectroscopy has been obtained from the VLT. An initial 1 hour exposure with the *Multi-Unit Spectroscopic Explorer (MUSE)* identified four main groups of lensed images, and suggested two low S/N peaks as candidates for a demagnified central image (M15). A subsequent additional 4 hour exposure (programme 295.A-5018; Massey et al. in prep.) confirms both candidates (Ao7 at RA: 330.47047, Dec: −59.945183, Ao8 at RA: 330.47079, Dec: −59.946112). Indeed, Ao7 is also visible in *HST* imaging, after using the MuSCADeT multiwavelength method (Joseph et al. 2016) to estimate and subtract bright foreground emission (Figure 1). We therefore use all the images identified by M15, plus the two new ones.

3 LENS MODELLING METHOD

We shall model the distribution of mass in the galaxy cluster as a sum of cluster-scale plus galaxy-scale halos (following e.g. Limousin et al. 2007a; Jauzac et al. 2014), each a perturbation around the Pseudo Isothermal Elliptical Mass Distribution (PIEMD; Kassiola & Kovner 1993).

3.1 Pseudo-Isothermal Skewed Potential

The 2D surface mass density Σ of a *circularly symmetric* pseudo-isothermal mass distribution, projected along a line of sight, is:

$$\Sigma(\mathbf{r}) \equiv \frac{\sigma_0^2 \, r_{\rm cut}}{2G \left(r_{\rm cut} - r_{\rm core} \right)} \left(\frac{1}{\sqrt{r_{\rm core}^2 + r^2}} - \frac{1}{\sqrt{r_{\rm cut}^2 + r^2}} \right), \quad (1)$$

where σ_0 is the 1D velocity dispersion, and where $r_{\rm core}$ ($r_{\rm cut}$) is an inner (outer) radius. To convert this into a PIEMD with ellipticity $\epsilon = \frac{a-b}{a+b} \geqslant 0$, where a and b are the semi-major and semi-minor axes respectively, Kassiola & Kovner (1993) apply their coordinate transformation (2.3.6):

$$x \to x_{\rm em} = \frac{x}{1+\epsilon} \,, \qquad y \to y_{\rm em} = \frac{y}{1-\epsilon} \,. \quad (2)$$

This maps a circle onto an ellipse centered at the origin, with its major axis along the x axis. In general, including a rotation to set the major axis at angle ϕ_ϵ, this can be expressed in polar coordinates as:

$$r^2 \to r_{\rm em}^2 = \frac{r^2}{\left(1-\epsilon^2\right)^2} \left[1 + \epsilon^2 - 2\epsilon \cos\left(2(\theta - \phi_\epsilon)\right) \right]. \quad (3)$$

The angle α by which a light ray is deflected as it passes near the lens, and the equivalent 2D gravitational potential ψ can be computed by integrating the density distribution:

$$\begin{aligned} \alpha(\mathbf{r}) &= \frac{4G}{c^2} \frac{D_l D_{ls}}{D_s} \int \Sigma(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^2} \, {\rm d}^2\mathbf{r}' \\ \psi(\mathbf{r}) &= \frac{4G}{c^2} \frac{D_l D_{ls}}{D_s} \int \Sigma(\mathbf{r}') \log|\mathbf{r} - \mathbf{r}'| \, {\rm d}^2\mathbf{r}' \,, \end{aligned} \quad (4)$$

where D_l, D_s and D_{ls} are the angular diameter distance from the observer to the lens, observer to the source, and lens to the source respectively. For general mass distributions, these integrals are difficult to solve – but closed forms have been found for the PIEMD, using techniques from complex analysis that exploit its elliptical symmetry (Bourassa & Kantowski 1975).

A related halo model is the Pseudo Isothermal Elliptical Potential (PIEP; Kassiola & Kovner 1993). In this, the coordinate transformation is applied to a circular potential ψ (rather than the density). It is then mathematically easier to obtain the deflection angle and density via differentiation:

$$\begin{aligned} \alpha(\mathbf{r}) &= \nabla\psi(\mathbf{r}) \\ \Sigma(\mathbf{r}) &= \frac{c^2}{8\pi G} \frac{D_s}{D_l D_{ls}} \nabla^2 \psi(\mathbf{r}) \,. \end{aligned} \quad (5)$$

In detail, the PIEP potential ψ is transformed so that:

$$\psi(x,y) \to \psi'(x,y) \equiv \psi\left(x',y'\right). \quad (6)$$

The first and second derivatives can then be computed with

applications of the chain rule. For example, the first x-derivative of the potential is:

$$\psi_x' = \left(\psi_{x'}\left(x',y'\right) x_x' + \psi_{y'}\left(x',y'\right) y_x' \right)\big|_{(x,y)}, \quad (7)$$

where the subscript denotes partial differentiation. The resulting mass distribution is not the same as a PIEMD, because of the way the coordinate transformation propagates through the chain rule (or back up the integrals in equation 4). For large ϵ, the mass distribution corresponding to a PIEP has undesirable features including concave (peanut-shaped) isodensity contours (Kassiola & Kovner 1993).

3.2 Pseudo-Isothermal Skewed Mass Distribution

To perturb the mass distribution in a way that resembles the behaviour of SIDM in numerical simulations (see figure 5 of Kahlhoefer et al. 2014), we apply a further coordinate transformation that maps a circle onto an ellipse with its focus (rather than centre) at the origin:

$$r^2 \to r'^2 = \frac{r^2 \left(1 - s^2\right)^{3/2}}{\left(1 + s\cos\left[\theta - \phi_s\right]\right)^2} \quad (8)$$

with s being the third eccentricity such that $s = \sqrt{1 - b^2/a^2}$, and the power $3/2$ being introduced to preserve area. Note the asymmetric $\cos(\theta)$ terms rather than the $\cos(2\theta)$ terms in the mapping described by equation (3).

We apply this transformation to the 2D gravitational potential corresponding to the PIEMD.[2] Analytic (albeit cumbersome) expressions for deflection angle and density can be readily calculated via differentiation (equation 5). We denote this the Pseudo Isothermal Skewed Potential (PISP); its isodensity contours are shown for various values of ϵ and s in Figure 2.

For any skew, the peak of the mass distribution always lies at the same position, so it will be possible to use the same metric as M15 to measure any offset between the most gravitationally bound stars and dark matter. The total mass changes slightly with increasing skew, but this can be recalculated after a fit.

Like the PIEP, the density distribution of the PISP exhibits undesired behaviour with large skews, because the coordinate transformation was applied to ψ, not Σ. Isophotes of the density distribution become concave, and the skew ellipticity can overwhelm the underlying ellipticity. We avoid these effects by restricting $|s| < 0.3$.[3] Since the PISP is invariant under transformations $s \to -s$ and $\phi_s \to \phi_s + \pi$, we fit parameters within the domain $s \in [-0.3, 0.3]$ and ϕ_s in some interval of length π. This ensures that the parameter space can be explored symmetrically about $s = 0$, allowing the case of zero skew to be recovered without bias. We set the edge of the domain of ϕ_s well away from any preferred

---

[2] We would ideally apply this transformation to the PIEMD mass distribution, but the relevant integrals (equation 4) do not contract to a simple form. A skewed mass distribution could also be derived from the potential corresponding to a PIEP. We choose to perturb the PIEMD so that we recover this widely-used mass distribution in the $s \to 0$ limit, and to minimise undesired convex curvature in density isophotes.

[3] See the Appendix for an alternative model that does not suffer from this effect, but has other disadvantages.
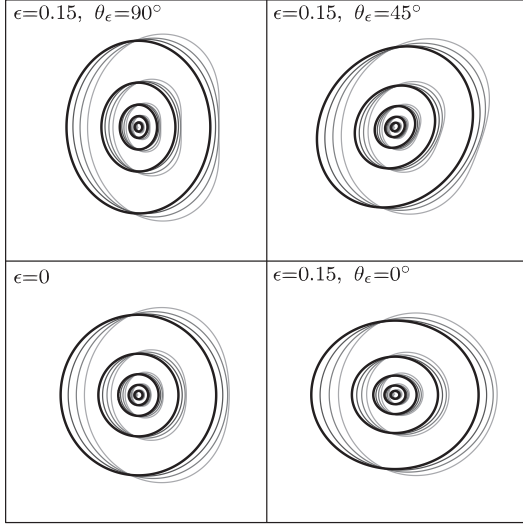
Figure 2. Isodensity contours of a PISP mass distributions. The core radius is about the same as the innermost density contour. Thick, black lines show an ordinary PIEMD; the bottom-left is circular, and the others have ellipticity ϵ=0.15, with the major axis at various angles. Thinner, grey lines show the same density profiles with skew s=0.1, 0.2, 0.3 to the right.

direction (in practice, having explored parameter space via a quick search), to make sure an MCMC sampler operates efficiently near regions of interest.

3.3   Testing an implementation in Lenstool

We have implemented the PISP as potential 813 in the publicly-available software Lenstool (Jullo et al. 2007). Given a parameterised mass distribution, and the location of background sources, Lenstool can compute the position of observed multiple images. Given the position of observed images, it can also use Markov-Chain Monte Carlo (MCMC) optimisation to fit parameters of the lensing mass distribution.

To test whether Lenstool can accurately recover a known input skew, we run two sets of tests. We first consider an isolated lens, with three background sources at different redshifts: the *example_with_images* configuration that is packaged with Lenstool. As a null test, we adopt the input mass distribution with skew $s_{\text{true}} = 0$. From the observed positions of multiple images, Lenstool successfully recovered a best fit (maximum likelihood) value $s = -0.0008^{+0.02}_{-0.02}$.

We then set skew $s_{\text{true}} = 0.2$ and $\phi_{s_{\text{true}}} = 1.6$. We set source positions by projecting one image of each source back to its source plane, then create a mock set of multiply lensed images by re-projecting this source forward through the lens. When fitting this mock data, Lenstool successfully recovers best-fit values $s = 0.2^{+0.001}_{-0.001}$ and $\phi_s = 1.6^{+0.04}_{-0.05}$.

Second, we test the recovery of input skews in a complex cluster lens with a mass distribution based on the

Abell 3827. Choosing one of the quadruply lensed background galaxy images, we repeat the procedure outlined above: projecting the light backwards and then forwards through a cluster lens with known mass distribution. The cluster is given the same parameters as our fiducial model for Abell 3827 (see §4.1), with the exception of the skew parameters. In this test, the dark matter associated with galaxy N1 is given skewness $s_{\text{true}} = 0.25$ and $\phi_{s_{\text{true}}} = 1.6$. As a null test, galaxies N2–N4 are set to have no skew, $s_{\text{true}} = 0$.

We run Lenstool with the same free parameters and priors as in §4.1). Within such a highly dimensional parameter space, we find that the best-fit values are sometimes noisy, for parameters that make only a small difference to the overall goodness of fit. However, the full posterior probability distribution function (PDF) is smooth and well-sampled. Hence, for the rest of this paper, we shall quote the modal peak and 68% width of the posterior PDF, which Lenstool also returns. This makes no difference for the simple model above, and successfully recovers $s = 0.24^{+0.04}_{-0.31}$ and $\phi_s = 1.6^{0.92}_{-0.99}$ for galaxy N1, and $s = 0.01^{+0.14}_{-0.13}, 0.07^{+0.10}_{-0.15}, 0.11^{+0.11}_{-0.16}$ for galaxies N2, N3, N4 (with very weakly constrained ϕ_s).

3.4   Prior bias for polar parameters

A skew is a two-component vector, and can be expressed in polar form as a magnitude $|s|$ and direction ϕ_s, or in Cartesian form as an amount in orthogonal directions $\{s_x, s_y\}$. We implemented the polar option, so that Lenstool's MCMC optimiser can explore a circularly symmetric region, with no preferred direction that could bias the inferred skew. Lenstool also defines ellipticities in this way, for the same reason.

Nonetheless, it may often be desirable to know the posterior probability distribution of skewness along e.g. a direction of motion, and perpendicular to that (i.e. the Cartesian form). The posterior probability distributions of skew and skew angle are returned by Lenstool (in RUNMODE=3) by the sampling density of the MCMC chain. This can be converted to the posterior of the skew in some direction ϕ by projecting and then weighting each sample by:

$$w = \frac{|s|}{\sqrt{0.3^2 - s^2 \cos^2(\phi_s - \phi)}}. \tag{9}$$

The numerator is the Jacobian to convert the area of parameter space from polar to Cartesian coordinates. The denominator corrects for prior bias, because the restriction $|s| \in [-0.3, 0.3]$ leads to a (semi-)circular prior on the projected skew.

4   STRONG LENS ANALYSIS OF ABELL 3827

We use the observed positions of lensed multiple images to fit a mass model of the cluster. Our choice of model parameters and their priors is based on those of M15, with some additional degrees of freedom. We assume $0.8''$ uncertainty on the position of lensed image Ao8, which has only been detected from the ground. We assume $0.2''$ uncertainty on the position of all other lensed images, which are identified by *HST*.

Appendix C. Colour deblending of Abell 3827

Table 1. Parameters of the fiducial mass model fitted by LENSTOOL. Quantities in square brackets are fixed. Errors on other quantities show 68% statistical confidence limits, marginalising over uncertainty in all other parameters. Stellar mass components are modelled as Hernquist profiles, with a mass (computed from flux in the F606W band), scale radius and ellipticity (fitted using GALFIT; galaxy N4 is contaminated by a nearby star). Dark matter components are modelled as PISPs, with a 1D velocity dispersion, core and cut radii, ellipticity and skewness. Positions are given in arcseconds relative to (R.A.: 4330.47515, Dec.: −59.945996), except galaxies' dark matter components, which are relative to the position of their stars. Angles are anticlockwise from East.

		x [''] Δx ['']	y [''] Δy ['']	Mass $[M_\odot]$ σ_v [km/s]	r_sc [''] r_core ['']	r_cut ['']	ϵ	ϕ_ϵ [°]	s	ϕ_s [°]
N1	stars	$[-0.06]$	$[0.04]$	$[1.00 \times 10^{11}]$	$[0.53]$		$[0.12]$	$[61]$	$0.21^{+0.06}_{-0.22}$	86^{+44}_{-44}
	dark matter	$-0.29^{+0.25}_{-0.14}$	$-0.71^{+0.30}_{-0.16}$	149^{+8}_{-12}	$[0.1]$	$[40]$	$0.02^{+0.33}_{-0.01}$	151^{+19}_{-116}		
N2	stars	$[5.07]$	$[2.05]$	$[2.46 \times 10^{11}]$	$[0.79]$		$[0.17]$	$[39]$	$0.03^{+0.11}_{-0.14}$	117^{+41}_{-80}
	dark matter	$-0.23^{+0.30}_{-0.16}$	$0.00^{+0.30}_{-0.30}$	182^{+29}_{-22}	$[0.1]$	$[40]$	$0.42^{+0.05}_{-0.22}$	23^{+32}_{-12}		
N3	stars	$[9.69]$	$[3.98]$	$[2.77 \times 10^{11}]$	$[0.33]$		$[0.05]$	$[31]$	$-0.02^{+0.08}_{-0.11}$	169^{+7}_{-109}
	dark matter	$-0.05^{+0.25}_{-0.25}$	$-0.06^{+0.18}_{-0.29}$	213^{+8}_{-10}	$[0.1]$	$[40]$	$0.49^{+0.01}_{-0.16}$	15^{+14}_{-8}		
N4	stars	$[9.26]$	$[-1.08]$	$[2.08 \times 10^{11}]$	$[1.37]$		$[0.39]$	$[127]$	$0.08^{+0.08}_{-0.09}$	147^{+21}_{-80}
	dark matter	$-1.35^{+0.39}_{-0.34}$	$0.51^{+0.35}_{-0.27}$	255^{+8}_{-10}	$[0.1]$	$[40]$	$0.02^{+0.25}_{-0.01}$	136^{+17}_{-28}		
N6	stars	$[18.54]$	$[2.47]$	$[0]$						
	dark matter	$[0]$	$[0]$	38^{+26}_{-25}	$[0.1]$	$[40]$	$[0]$	$[0]$	$[0]$	$[0]$
Cluster	dm	$5.53^{+1.46}_{-1.61}$	$2.33^{+1.97}_{-1.59}$	683^{+139}_{-75}	$30.12^{+9.23}_{-6.43}$	$[1000]$	$0.56^{+0.13}_{-0.10}$	63^{+2}_{-3}	$[0]$	$[0]$

4.1 Fiducial mass model

The cluster's large-scale mass distribution is modelled as a single PIEMD. Based on a comprehensive (but slow) initial exploration of parameter space, its position is given by a broad Gaussian prior with $\sigma = 2'' = 3.66$ kpc, centred on the position of galaxy N2. Flat priors are imposed on its ellipticity ($\epsilon < 0.75$), core size ($r_\mathrm{core} < 40''$) and velocity dispersion ($300 < \sigma_\mathrm{v} < 1000$ km/s). Its cut radius is fixed at $r_\mathrm{cut} = 1000''$, well outside the strong lensing region, i.e. away from any multiple image constraints.

Central galaxies N1–N4 are each modelled as a stellar component (which was not included in the fiducial model of M15), plus a dark matter one. Following Giocoli et al. (2012), the stellar components are modelled with Hernquist (1990) profiles:

$$\rho_\mathrm{star}(r) = \frac{\rho_s}{(r/r_s)\left(1 + r/r_s\right)^3}, \tag{10}$$

where the scale radius r_s is related to the half mass radius R_e, such that $R_e = r_s/0.551$, and the scale density $\rho_s = M_\mathrm{total}/\left(2\pi r_s^3\right)$. We fix the mass of the stellar component, and its half-mass radius, using the optical magnitudes and profiles measured by M15. These parameters are listed in Table 1.

The four central galaxies' dark matter components are now modelled as PISPs. We impose flat priors on their positions, in $4'' \times 4''$ boxes centred on their luminosity peaks, plus flat priors on their ellipticity ($\epsilon < 0.5$) and velocity dispersion ($v_\mathrm{disp} < 600$ km/s). We fix $r_\mathrm{cut} = 40'' = 73$ kpc (Limousin et al. 2007a).

Galaxy N6 is much fainter than the others, so we approximate its total mass distribution as a single PIEMD. This has a fixed position and ellipticity to match the light distribution, and only its velocity dispersion is optimised (with a flat prior $v_\mathrm{disp} < 500$ km/s).

We optimise the free parameters using LENSTOOL, with RUNMODE=3. This runmode is used to fully explore the
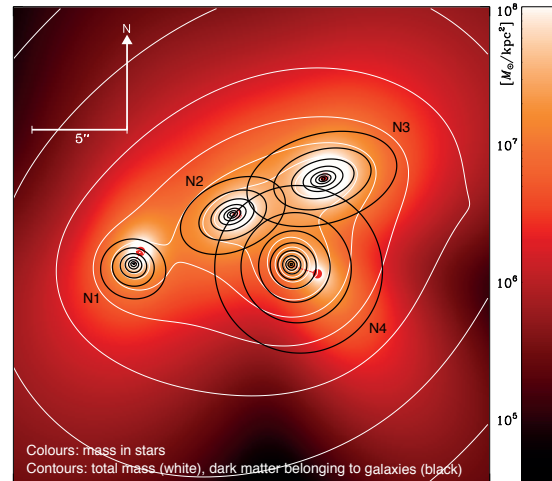

Figure 3. The best fitting mass distribution in the gravitational lens Abell 3827, integrated along our line of sight. For reference, the background colour scale shows the modelled stellar mass density. Red spots indicate the position of the luminosity peak in galaxies N1–N4. White isodensity contours show the total lensing mass of the cluster. The outermost contour corresponds to a projected density of $2 \times 10^9\,M_\odot/\mathrm{kpc}^2$, and values increase towards the centre by a factor of $2^{1/3}$=1.26. Black isodensity contours isolate each galaxy's dark matter component. The outermost contour corresponds to a projected density of $1.26 \times 10^9\,M_\odot/\mathrm{kpc}^2$ and values increase by a factor of $2^{2/3}$. The visible offset between stars and dark matter in galaxies N1 and N4 are both statistically significant; the asymmetry in the distribution of N1's dark matter is also significant.

posterior (Jullo et al. 2007). (Modal) maximum likelihood parameters are shown in Table 1, and the corresponding mass distribution is shown in Figure 3. The best fit model achieves a RMS offset between the observed and predicted positions of multiple images of $\langle\mathrm{rms}\rangle_i = 0.26''$. There are 54 constraints and 35 free parameters in our model. The modal $\chi^2/\mathrm{dof} = 67.1/19$ with a log likelihood of 8.18. The full posterior probability distribution for the dark matter associated with galaxies N1–N4 is shown in Figures 4 and 5.

4.2 Sensitivity to model choices

4.2.1 Stellar mass components

Galaxies definitely contain stars, and those stars have mass. Not accounting for this mass could bias the skew measurement. In an offset DM halo, the stellar mass will lead the DM peak, while any tail of DM particles will trail behind in the opposite direction. Not accounting for this stellar mass could weaken the skew measurement, or in the worst case scenario, if the stellar mass is greater than that of scattered particles, reverse the direction of the measured skew. We have explicitly modelled the stellar mass seperately to avoid any bias in the inferred skew. In practice, as in M15, we find that including the stellar mass component (or even multiplying/dividing its mass by a factor 2) does not change any other results, within their statistical errors.

4.2.2 Identification of new lensed images

Adding constraints from the two new lensed images Ao7 and Ao8 tightens constraints on nearby galaxies N3 and N4. These (demagnified) images are unresolved, and any of the features in the background spiral could be assigned to them. We have tried relabelling one or both of the demagnified images as either the bulge, Aon, or one of the two brightest knots of star formation, Aan or Abn. LENSTOOL's outputs are statistically indistinguishable. In all cases, the *entire* background spiral galaxy is predicted to be lensed onto both the northern and the southern demagnified images.

4.2.3 Mass in other cluster member galaxies

We also tested the impact of adding more cluster member galaxies to the mass model. These galaxies were identified using a colour-magnitude selection using the F814W and F606W2 HST/ACS band imaging. Source detection was done using SEXTRACTOR (Bertin & Arnouts 1996) in dual mode, with reference taken in the F814W-band. We then identified as cluster members all galaxies brighter than $mag_{\mathrm{F814W}} < 23$ and within 1σ of the red sequence best-fit:

$$(mag_{\mathrm{F814W}} - mag_{\mathrm{F606W}}) = 0.022 \times mag_{\mathrm{F606W}} - 1.129.$$

Our final cluster member catalogue contains 147 galaxies.

These galaxies are added to the mass model as small scale perturbers. We assume fixed cut radius and velocity dispersion, scaled by their luminosities in the F814W-band. This methodology has been successfully validated by Harvey et al. (2016), and adopted widely in previous work (e.g. Jauzac et al. 2014; Richard et al. 2014; Limousin et al. 2007b). To derive L^*, we use the K^* magnitudes obtained by Lin et al. (2006) as a funciton of cluster redshift. LENSTOOL

is then scaling the cut radius and velocity dispersion of each galaxies in our catalogue relative to a $K^* = 16.6$ galaxy with velocity dispersion $\sigma^* = 108.4 \pm 27.5$ km/s, and cut radius $r_{\mathrm{cut}}^* = 48.5 \pm 16.0$ kpc.

Including all cluster memeber galaxies in a re-optimised mass model significantly affects neither offset nor skewness measurements of dark matter associated with central galaxies N1–N4. By far the most affected measurement is the skewness of galaxy N1, which increases to $s = 0.28^{+0.01}_{-0.31}$. All other quantities remain consistent within random noise.

5 DISCUSSION

5.1 Galaxy N1

The previously-detected offset between galaxy N1's stars and dark matter persists at $> 3\sigma$ in our new analysis. Adding two free parameters for the asymmetry of its dark matter slightly increases uncertainty in its position. The modal offset is $(\Delta x, \Delta y) = \left(-0.22^{+0.25}_{-0.14}, -0.81^{+0.16}_{-0.17}\right)$ for an unskewed model, and $(\Delta x, \Delta y) = \left(-0.29^{+0.25}_{-0.14}, -0.71^{+0.30}_{-0.16}\right)$ if skewness is allowed. The consistency between these suggests that the measured position of the density peak is robust against the skew parameter probed here.

If the offset is entirely due to an effective drag force through frequent dark matter self-interactions, it implies a momentum-transfer interaction cross-section $\tilde{\sigma}/m_{\mathrm{DM}} \gtrsim 1\,\mathrm{cm}^2\mathrm{g}^{-1}$, assuming galaxy N1 is falling into the cluster for the first time (Kahlhoefer et al. 2015). In general, We agree with this interpretation[4], but note that the cross-section can be lower if the galaxy has completed multiple orbits; its current direction of motion is unknown.

We also find (at much lower statistical significance) that galaxy N1's dark matter is skewed in a direction consistent with the SIDM interpretation of its offset. This could be signs of a tail of scattered DM particles and would favour high momentum transfer scattering. However, the weak statistical significance of our result makes it impossible to rule out the low momentum transfer case. Figure 6 shows the posterior PDF of the skew onto the vector pointing from the DM peak to the stellar luminosity in the fiducial model, such that a positive skew corresponds to the direction predicted by SIDM. The peak of the posterior and 1σ errors are $s = 0.23^{+0.05}_{-0.22}$. If we individually project the skewness onto the offset direction individually in all MCMC samples, we find $s = 0.26^{+0.03}_{-0.22}$. Finally in our model that contained the additional 147 cluster galaxies we find $s = 0.28^{+0.01}_{-0.31}$. In all cases $\sim 70\%$ of the posterior PDF lies at $s > 0$. As the posterior peak is near the edge of the prior, which is chosen coservatively (see §3.2), a parametric halo model that does not break down for large skew parameters could result in a stronger detection.

---

[4] We have also repeated Kahlhoefer et al. (2015)'s calculation of $\tilde{\sigma}/m_{\mathrm{DM}}$ but integrating the effect of the restoring force on the *entire* distribution rather than just the peak. The difference is not significant.
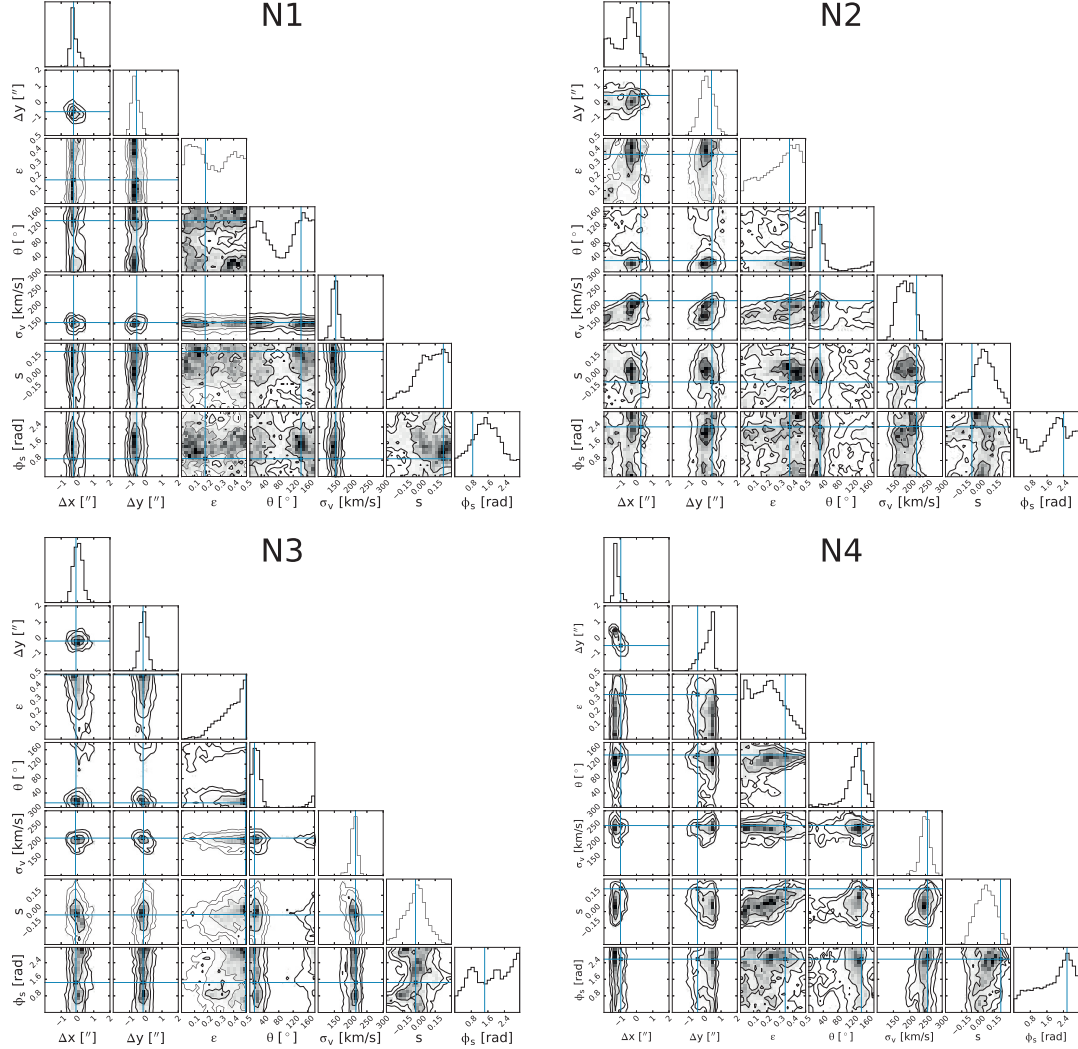
Figure 4. Posterior probability distribution of the distribution of dark matter associated with galaxies N1 (top left), N2 (top right), N3 (bottom left) and N4 (bottom right). Contours show the 68%, 95% and 99.7% contour levels. Blue lines indicate the best fit model, but note that this can be noisy, and we instead use the peak of the smoothed posterior distribution. Positions have been recentered such that $(x, y) = (0, 0)$ is the peak of the stellar luminosity. Offsets between stars and dark matter are measured at $>3\sigma$ for galaxies N1 and N4. A skew is detected at $>1\sigma$ for galaxy N1, in a direction consistent with the spatial offset.

5.2 Galaxies N2 and N3

The dark matter associated with galaxies N2 and N3 appears symmetric, and coincident with the stars. This result does not preclude offsets from existing along the line of sight. Furthermore, even if the galaxies are moving in the plane of the sky, they could be behind or in front of the most dense regions of the cluster core, and therefore passing through a less dense medium, experiencing less drag.

5.3 Galaxy N4

The measured position of N4's dark matter is intriguing. Accounting purely for statistical error bars, thanks to the confirmed positions of demagnified images, we find that galaxy N4 is offset from the galaxy's stars at the 3σ level. However, the offset position is mildly degenerate with the position of the cluster-scale dark matter (Figure 5), thus a flat prior on the cluster scale halo could lead to a different offset measurement. Furthermore, the measured ellipticity of the galaxy light is contaminated by light from an adjacent Milky Way star, and its position may also be.
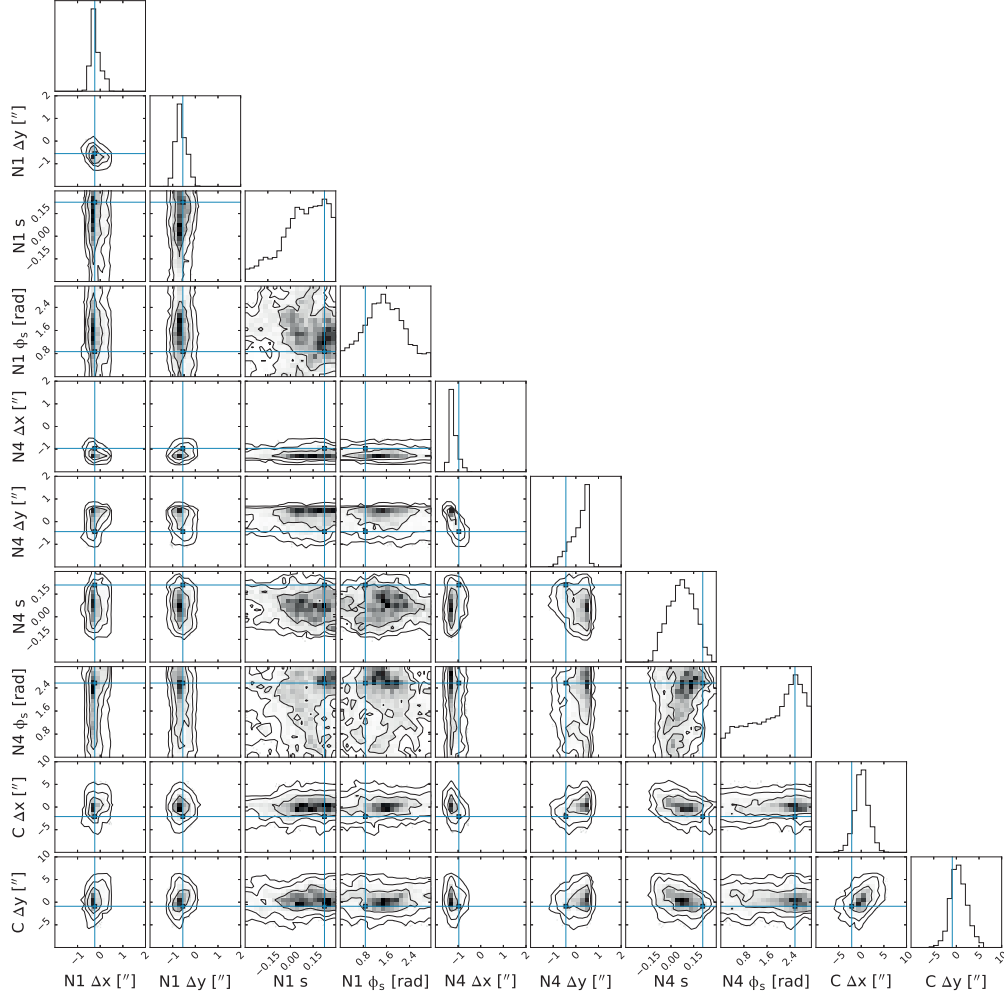
Figure 5. Posterior probability distribution showing (minimal) correlations between the position and asymmetry of the dark matter associated with key galaxies, and between those galaxies and the cluster-scale dark matter(denoted C). Contours show the 68%, 95% and 99.7% contour levels. Positions have been recentered as in Figure 4.

Despite the measured offset of dark matter from galaxy N4, it shows no sign of skewness. If the offset is spurious, as discussed above, then a tail is not expected. If the off-set is real, the lack of skew favours low momentum transfer scattering and (at very low S/N) is in mild tension with the skew detected in galaxy N1. Unknown systematics in the modelling of DM around either galaxy could be responsible. Nonetheless, there is also a possible physical explanation for this discrepancy. Galaxy N4 is in a higher density environment than galaxy N1, closer to the cluster core. It is possible that any tail of scatttered N4 particles has been tidally destroyed by the steeper gradient in gravitational potential.

6   CONCLUSION

We have developed a parametric lens models for asymmetrically skewed mass distributions. This can be used to search for scattered (self-interacting) dark matter in colliding systems. More generally, it will also be useful to investigate claims of dynamically-induced asymmetry (Prasad & Jog 2016; Chemin et al. 2016), or tidal tails (which are asymmetric if the size of a body is large compared to its distance from the centre of potential).

We have also presented a new model for the distribution of mass in galaxy cluster Abell 3827. Our choice of flat priors for the position of *all* galaxies' dark matter leads to a detected offset between a second galaxy's stars and its dark matter. New VLT/MUSE observations tighten the constraints on that offset. Neither measured offset changes

Appendix C. Colour deblending of Abell 3827
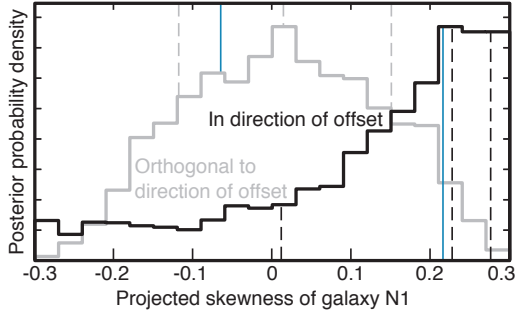
Figure 6. The posterior of the skew vectors in galaxy N1 projected onto the offset vector (black) and orthogonal to this (grey). The Jacobian has been accounted for and the priors have been adjusted as described in section 3.4. Dashed lines indicate the posterior peak and 1σ confidence intervals. The blue line indicates the (noisy) best fit value. There is a preference for a skew that is consistent with SIDM. No such preference is shown for a skew component in the orthogonal direction.

significantly if the models are allowed extra freedom to become skewed.

We find tantalising, but low significance evidence that the galaxy closest to multiply lensed images (and therefore the best constrained) has an asymmetric distribution of dark matter, skewed in the same direction as its offset from stars. We emphasise that our skew model, which captures the qualitative behaviour of scattered DM particles, is primarily motivated by mathematical convenience and that all skew measurements here are model dependent. More work will be needed to determine the significance of this result: whether it is physical, or an artefact of systematics in parametric lens modelling. Even in mock data where the true skew is known, skewness cannot be measured to high precision in a system as complex as Abell 3827. This is probably because the effect of skewness on lensed image positions is smaller than the effects of other free parameters.

A promising direction for future investigation may be provided by pairs of field galaxy in the SLACS survey, one of which has already been found to have an offset between dark and luminous matter (Shu et al. 2016). Whilst the SIDM model predicts a largest (most easily observable) offset in galaxies moving through a dense intracluster medium, it may be possible to more tightly constrain any asymmetry of dark matter in these simpler systems. If the directions of their dark matter tails correlate with the directions of their offsets, this evidence would support the hypothesis of SIDM with a high-mass mediator particle.

REFERENCES

Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393

Bourassa R., Kantowski R., 1975, The Astrophysical Journal, 195, 13

Carrasco E. R., et al., 2010, ApJL, 715, L160

Chemin L., Huré J.-M., Soubiran C., Zibetti S., Charlot S., Kawata D., 2016, A&A, 588, A48

Clowe D., Gonzalez A., Markevitch M., 2004, ApJ, 604, 596

Despali G., Giocoli C., Bonamigo M., Limousin M., Tormen G., 2016, arXiv preprint arXiv:1605.04319

Foreman-Mackey D., 2016, The Journal of Open Source Software, 24

Giocoli C., Meneghetti M., Bartelmann M., Moscardini L., Boldrin M., 2012, Monthly Notices of the Royal Astronomical Society, 421, 3343

Harvey D., Massey R., Kitching T., Taylor A., Tittley E., 2015, Science, 347, 1462

Harvey D., Kneib J.-P., Jauzac M., 2016, Monthly Notices of the Royal Astronomical Society, 458, 660

Harvey D., Robertson A., Massey R., Kneib J.-P., 2017, Monthly Notices, 464, 3991

Hernquist L., 1990, The Astrophysical Journal, 356, 359

Jauzac M., et al., 2014, Monthly Notices, 443, 1549

Joseph R., Courbin F., Starck J.-L., 2016, A&A, 589, A2

Jullo E., Kneib J.-P., Limousin M., Eliasdottir A., Marshall P., Verdugo T., 2007, New Journal of Physics, 9, 447

Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. T., Sarkar S., 2014, Monthly Notices, 437, 2865

Kahlhoefer F., Schmidt-Hoberg K., Kummer J., Sarkar S., 2015, Monthly Notices: Letters, 452, L54

Kassiola A., Kovner I., 1993, The Astrophysical Journal, 417, 450

Kim S. Y., Peter A. H. G., Wittman D., 2016, preprint, (arXiv:1608.08630)

Lage C., Farrar G., 2014, ApJ, 787, 144

Limousin M., Kneib J., Bardeau S., Natarajan P., Czoske O., Smail I., Ebeling H., Smith G., 2007a, Astronomy & Astrophysics, 461, 881

Limousin M., et al., 2007b, The Astrophysical Journal, 668, 643

Lin Y.-T., Mohr J. J., Gonzalez A. H., Stanford S. A., 2006, The Astrophysical Journal Letters, 650, L99

Massey R., et al., 2015, Monthly Notices, 449, 3393

Planck Collaboration 2016, A&A, 594, A13

Prasad C., Jog C. J., 2016, preprint, (arXiv:1610.01702)

Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., 2008, ApJ, 679, 1173

Richard J., et al., 2014, Monthly Notices of the Royal Astronomical Society, 444, 268

Robertson A., Massey R., Eke V., 2016, Monthly Notices of the Royal Astronomical Society, 465, 569

Robertson A., Massey R., Eke V., 2017, Monthly Notices, 465, 569

Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, Monthly Notices, 430, 81

Shu Y., Bolton A. S., Moustakas L. A., Stern D., Dey A., Brownstein J. R., Burles S., Spinrad H., 2016, The Astrophysical Journal, 820, 43

Spergel D. N., Steinhardt P. J., 2000, Physical Review Letters, 84, 3760

Vogelsberger M., Zavala J., Loeb A., 2012, Monthly Notices, 423, 3740

Williams L. L., Saha P., 2011, Monthly Notices of the Royal Astronomical Society, 415, 448

APPENDIX A: ALTERNATIVE METHOD TO GENERALISE LENS MASS DISTRIBUTIONS

Another way to introduce asymmetry is to apply a weighting function $w(\mathbf{r}; \{a_i\})$ to an elliptical lensing potential

$$\psi(\mathbf{r}) \rightarrow \psi'(\mathbf{r}) \equiv w(\mathbf{r}; \{a_i\})\, \psi(\mathbf{r}), \qquad (A1)$$

where $\{a_i\}$ are a set of parameters. The deflection and surface mass density are readily computed by differentiating.

We consider weighting functions of the form

$$w(\mathbf{r}; \{a_i\}) = 1 + s f(r, \theta) \qquad (A2)$$

where s is a skew parameter and $f(r, \theta)$ is written in polar coordinates. The second term acts as a perturbation away from elliptical symmetry of $O(s)$, with $s = 0$ corresponding to the elliptically symmetric case. We chose $f(r, \theta)$ to meet the following criteria:

• To ensure that the space about $s = 0$ is explored symmetrically in Lenstool, so that a non-zero skew is not artificially recovered, we require that $s f(r, \theta) = -s f(r, \theta + \pi)$.

• To avoid difficulties near the origin, we require $f(r, \theta)$ to be an increasing function of r. This is also physically motivated, as it is difficult to scatter particles from the centre of the potential well at $r = 0$.

• To ensure that the surface mass density remains positive (or becomes negative only for large r well outside any region of interest), we require $f(r, \theta)$ to be bounded.
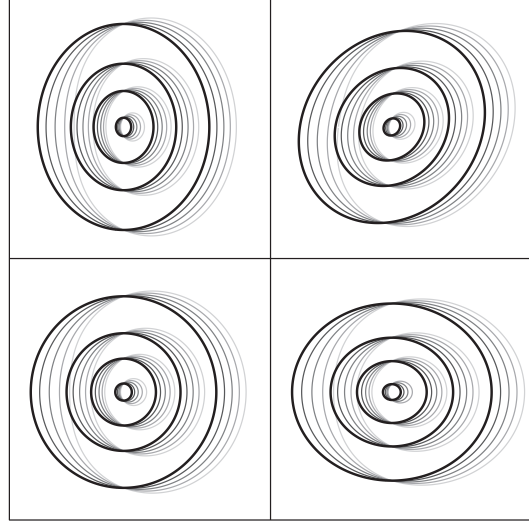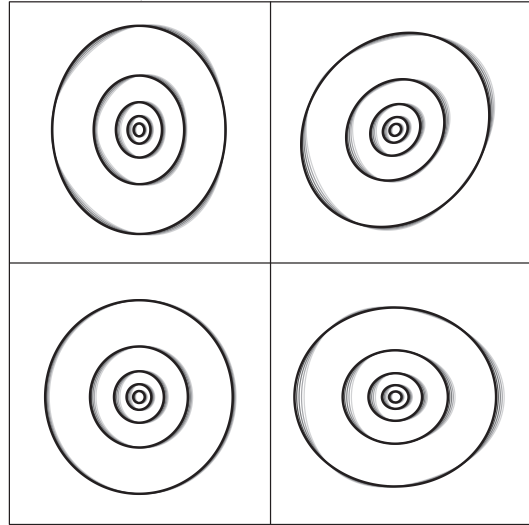

Figure A1. Isodensity contours of wPISP mass distributions, inside (top) and outside (bottom) the scale radius r_s/β. Thick, black lines show an ordinary PIEMD; the bottom-left is circular, and the others have ellipticity ϵ=0.15, at angles θ_ϵ=0°, 45° and 90° in the same order as in Figure 2. Thinner, grey lines show the same density profiles with skew s=0.1, 0.2, 0.3, 0.4 to the right.

A1 The Weighting Function Pseudo-Isothermal Skewed Potential

Meeting the above conditions we form the weighted Pseudo-Isothermal Skewed Potential (wPISP) by applying the weighting function

$$w(\mathbf{r}; s, r_s, \beta, \phi_s) = 1 + s \tan^{-1}\left(\beta \frac{r}{r_s}\right) \cos(\theta - \phi_s), \qquad (A3)$$

to the PIEMD potential, where r_s is a new scale radius, β sets the radial dependence of the skew, and ϕ_s is the skew angle.[5] This is now available as potential 812 in LENSTOOL.

The resulting surface mass densities are shown in Figure A1. The qualitative shape of the isodensity contours changes inside or outside the scale radius (owing to the sign change in second derivative of the inverse tangent). This feature could be used to isolate a behaviour that best matches numerical simulations, by fixing very large or very small r_s, or to capture more complex behaviour.

The total mass of a wPISP is identical to that of of a PIEMD. Since the weighting function is normalised by construction, the integrated mass density of a PIEMD and wPISP over a circular region are the same:

$$\frac{c^2}{8\pi G} \frac{D_s}{D_l D_{ls}} \int_{|\mathbf{r}|<R} \nabla^2 \psi \ dA = \frac{c^2}{8\pi G} \frac{D_s}{D_l D_{ls}} \int_{|\mathbf{r}|<R} \nabla^2 \left(w\psi\right) \ dA,$$

$$(A4)$$

where θ-dependence cancels. Taking the limit as $|\mathbf{r}| \to \infty$, the left hand side will converge to the total mass of a PIEMD with ellipticity ϵ, and the right to the mass of an equivalent wPISP. However, the position of the density peak varies slightly as a function of s. Care would need to be taken if using a wPISP to measure offsets of dark matter.

As with the PISP, this model breaks down for large values of s, since the weighting function has been applied to the potential and not the density. We have found that the value of s where this occurs is sensitive to β and the cut and core radii. For this reason, we recommend testing the boundaries of the parameter space for a breakdown of the desired skewed behaviour before substantial future work. Nonetheless, we tested the wPISP against the null hypothesis of the unskewed *example_with_images* system distributed with LENSTOOL (see §3.3). Fixing $\beta = 0.01$ and starting with a flat prior $s \in [0.3, 0.3]$, LENSTOOL successfully recovers skewness $s = 0.002^{+0.002}_{-0.002}$.

A2   Pseudo Isothermal Varying Ellipticity Mass Distribution

Despali et al. (2016) predict that, even with standard cold dark matter, the ellipticity of a cluster scale halo should change as a function of radius, becoming more elongated further from the centre. This prediction can be tested by using the weighting function formalism to add an extra parameter to halo models that mimics this behaviour. To achieve this, we combine a weighted sum of two different mass densities with different ellipticities into a Pseudo Isothermal Varying Ellipticity Mass Distribution (PIVEMD)

We suggest a mass density of the following form:

$$\Sigma\left(\mathbf{r}\right) = \Sigma_{\epsilon_1}\left(\mathbf{r}\right) w_1\left(\mathbf{r}\right) + \Sigma_{\epsilon_2}\left(\mathbf{r}\right) w_2\left(\mathbf{r}\right) \quad (A5)$$

where $\Sigma_{\epsilon_1}\left(\mathbf{r}\right)$ and $\Sigma_{\epsilon_2}\left(\mathbf{r}\right)$ are two elliptical profiles with ellipticity ϵ_1 and ϵ_2. All the other parameters for these two
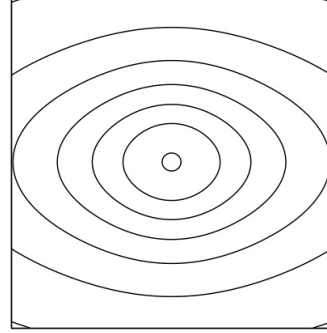

Figure A2. Isodensity contours for a radially varying weighted sum of two PIEMDs with different ellipticities.

densities should be shared. In this case, we find it most effective (and possible) to apply the weighting function directly to the mass density. To be computationally efficient within an MCMC loop, deflection angles must be computed once, using numerical integration, and stored in a look-up table.

The weighting functions w_i should meet the following criteria:

- To normalise the total mass, $w_1\left(\mathbf{r}\right) + w_2\left(\mathbf{r}\right) \equiv 1$, $\forall r$.
- So that one ellipticity dominates at small r and the other at large r, let $w_1\left(\mathbf{r}\right) \to 1$ as $r \to \infty$, $w_1\left(\mathbf{r}\right) \to 0$ as $r \to 0$, $w_2\left(\mathbf{r}\right) \to 0$ as $r \to \infty$ and $w_2\left(\mathbf{r}\right) \to 1$ as $r \to 0$.

Although this is quite a general set of conditions, we can take, for example

$$\begin{aligned} w_1\left(\mathbf{r}\right) &= 1 - e^{-\beta r} \\ w_2\left(\mathbf{r}\right) &= e^{-\beta r}, \end{aligned} \quad (A6)$$

where β controls the radial dependence. The resulting mass distribution for this weighted sum is illustrated in Figure A2.

---

[5] The inverse tangent form of the radial dependence is not physically motivated, and other functional forms may also work. While it is mathematically unnecessary to have two degenerate parameters r_s and β, to avoid computational divisions by zero, the distributed LENSTOOL implementation uses hardcoded $r_s = 0.1''$ and allows β to vary.

D HFF deblending residuals

This appendix shows the residuals of the deblending of the HFF clusters by `MuSCADeT`. A perfect separation by `MuSCADeT` would produce residuals with noise only. In this application, several features appear above the noise. First, milky way stars in the FoV create artefacts due to their extended features and due to the saturation of the CCD. Objects with extremely red and extremely blue SEDs are also poorly extracted by `MuSCADeT` due to their SEDs being to far from the ones `MuSCADeT` was given to extract.
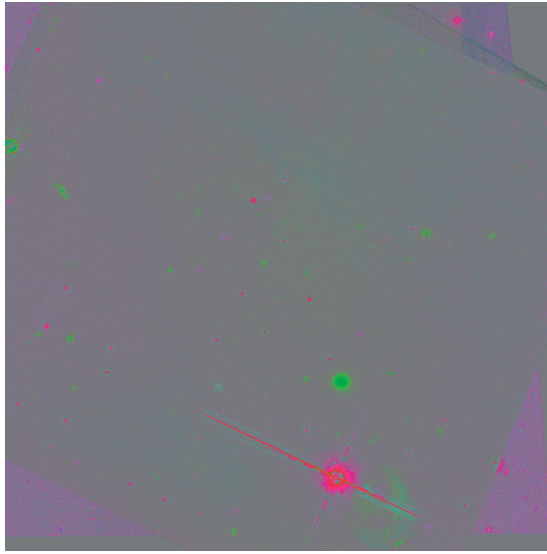
(a) MACS J0416 cluster
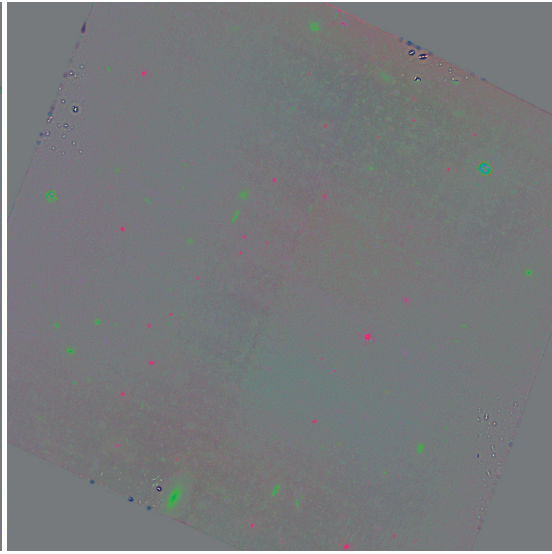
(b) MACS J0416 parallel
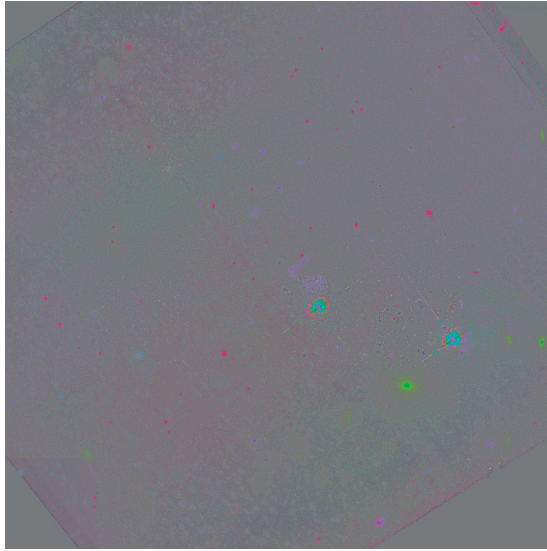
(c) MACS J0416 parallel
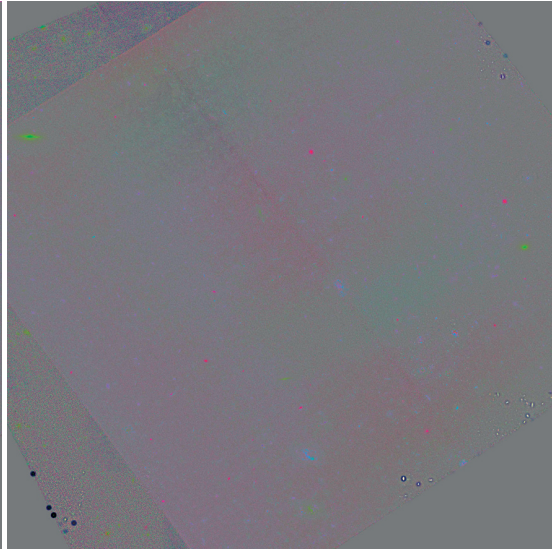
(d) MACS J0416 parallel
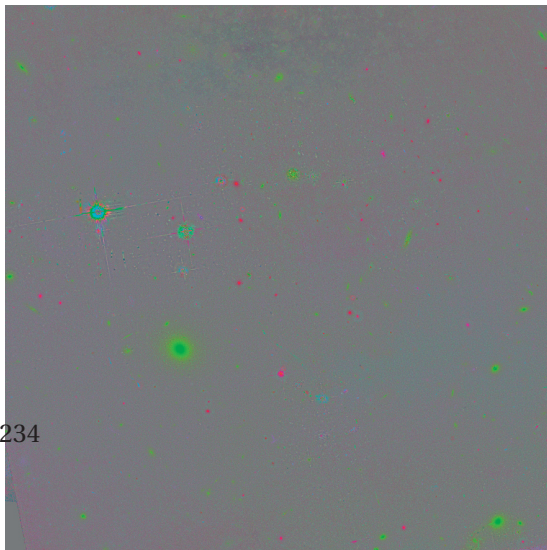
(e) MACS J0416 parallel
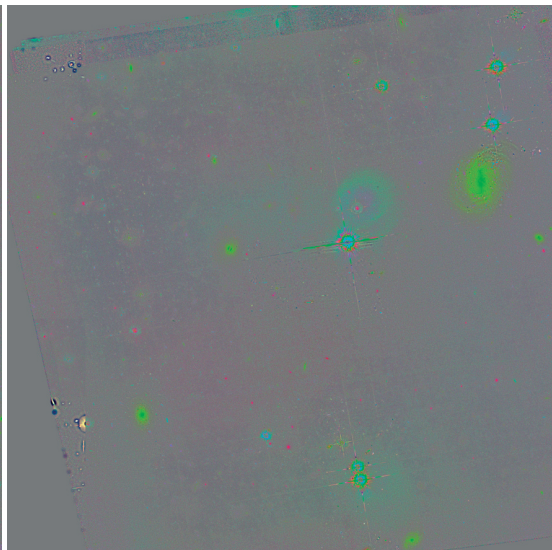
(f) MACS J0416 parallel

Figure D.1 – Residuals of the `MuSCADeT` decompositions of the HFF images.
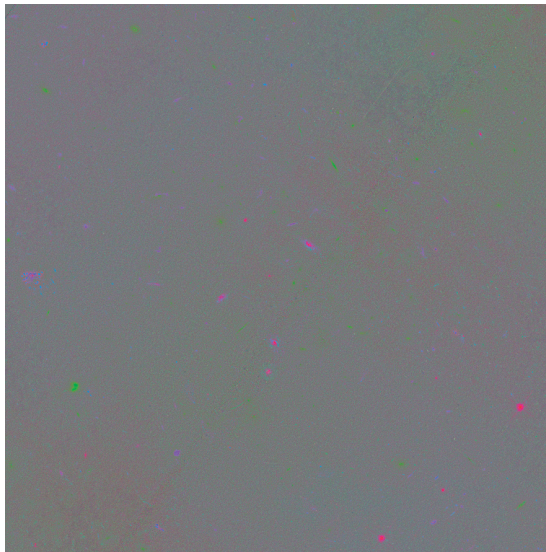
(a) Abell 2744
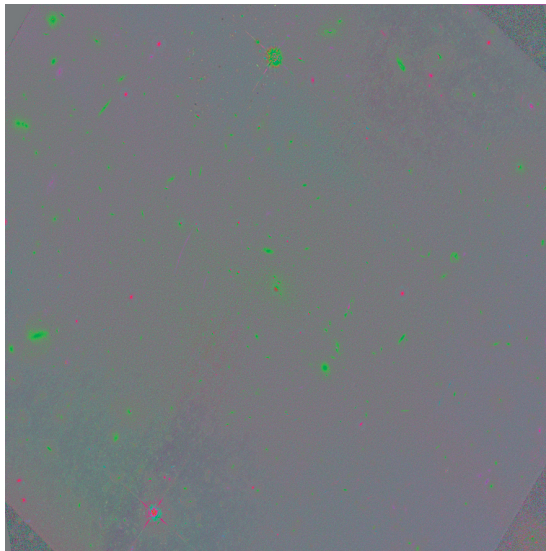
(b) Abell 1063 cluster
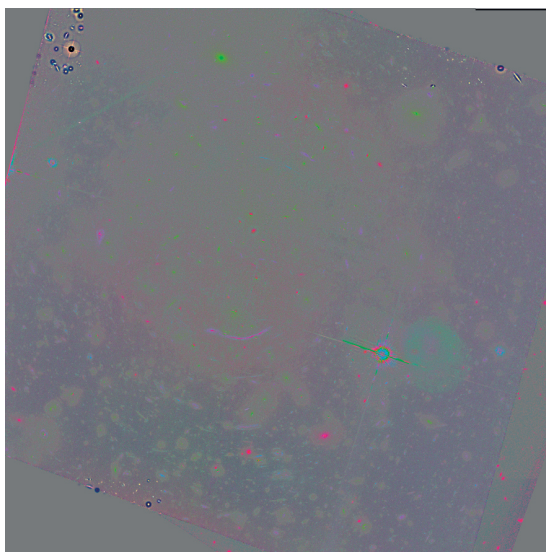
(c) Abell 370 cluster

Figure D.2 – Idem as figure D.1.

E Simulation deblending

In this appendix, I show the result of the separation of colour components in the HFF-like simulation of galaxy cluster *Ares* (Meneghetti et al., 2017) with `MuSCADeT`. The results of the separation are used to validate the accuracy of photometric measurements on background multiply imaged lensed sources. Accurate photometric measurements on these objects could allow to compute magnification ratios in a cluster, which could, in turn, be used to constrain mass models.
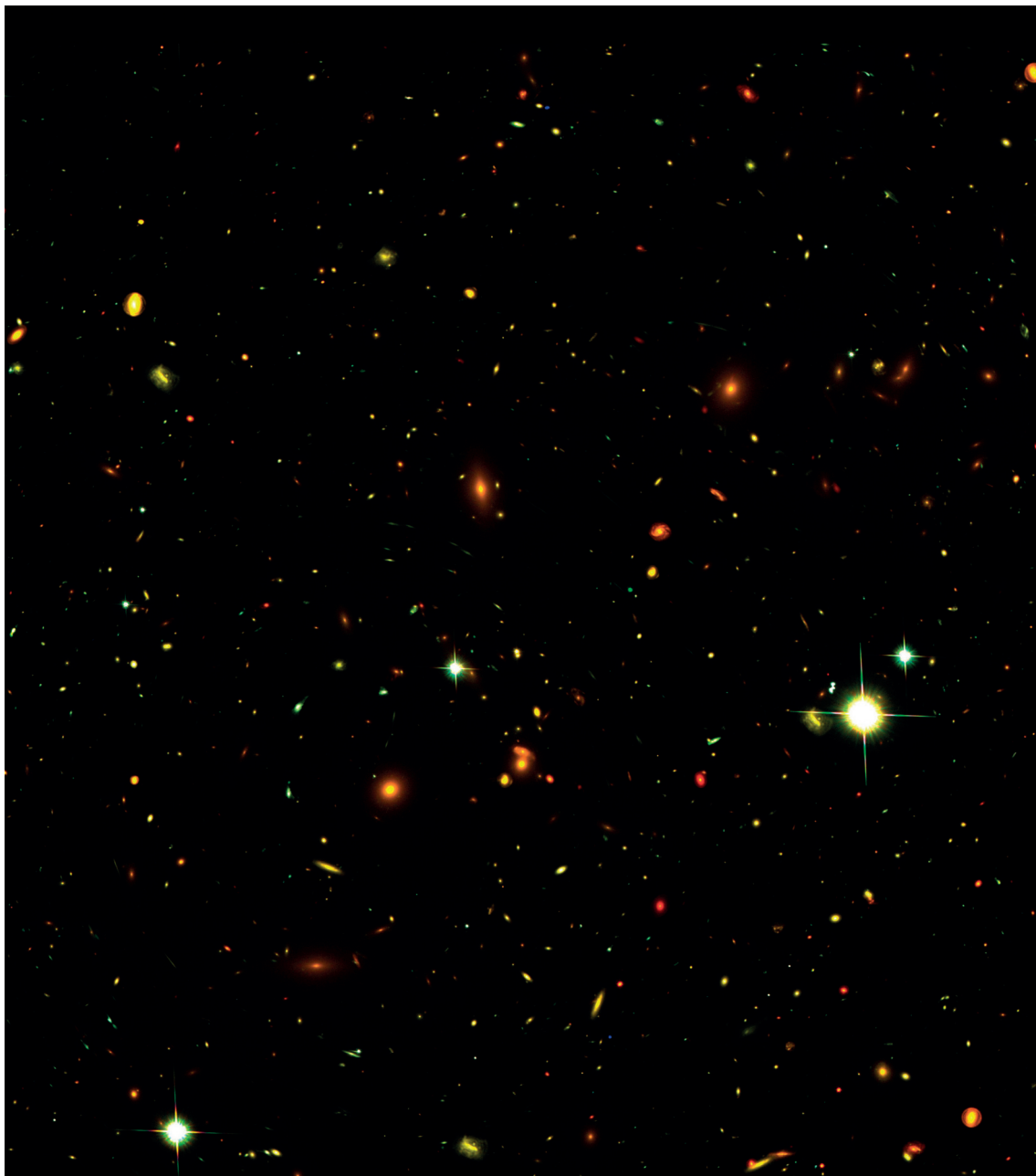
Figure E.1 – Composite RGB image of the Ares cluster
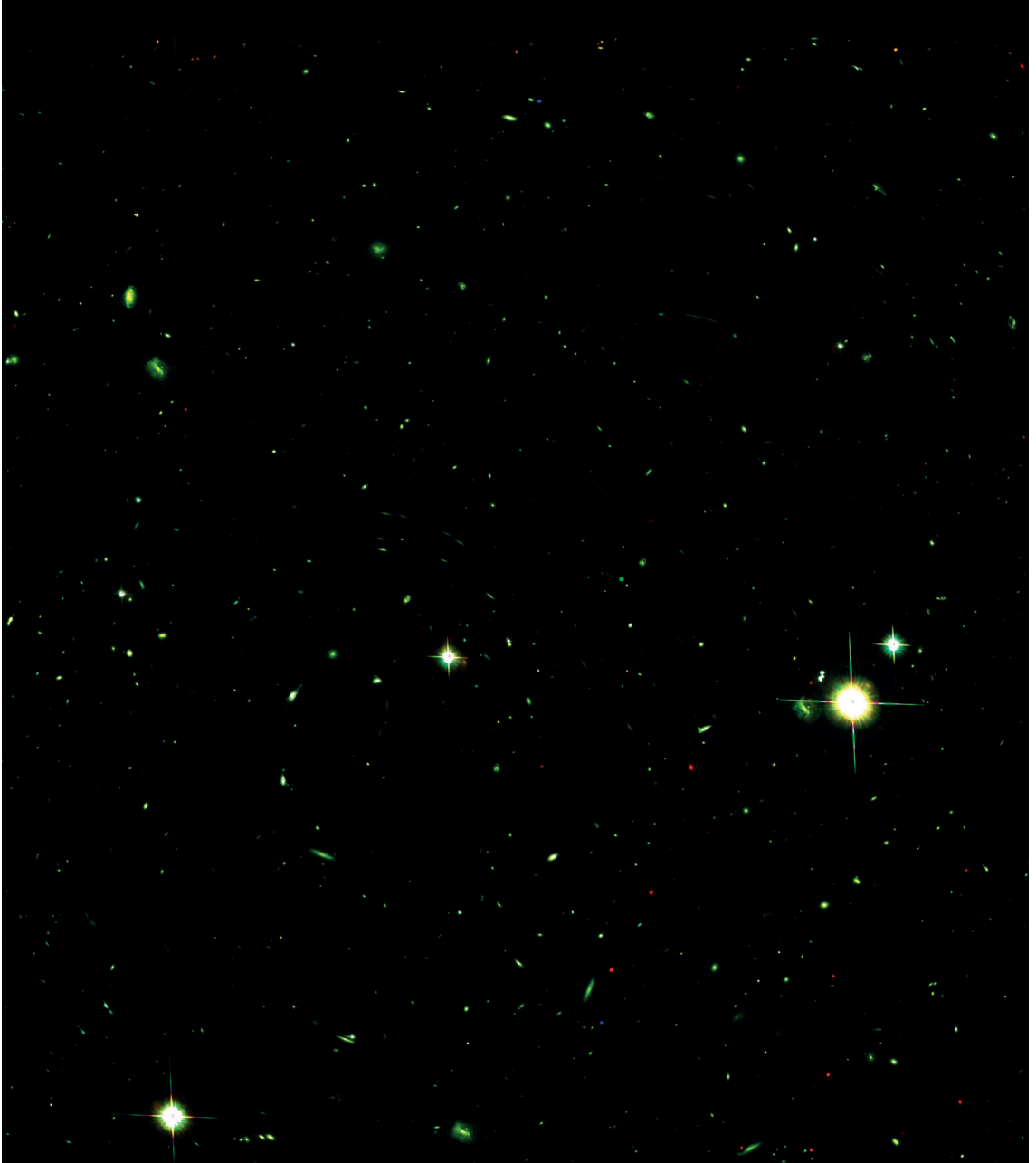
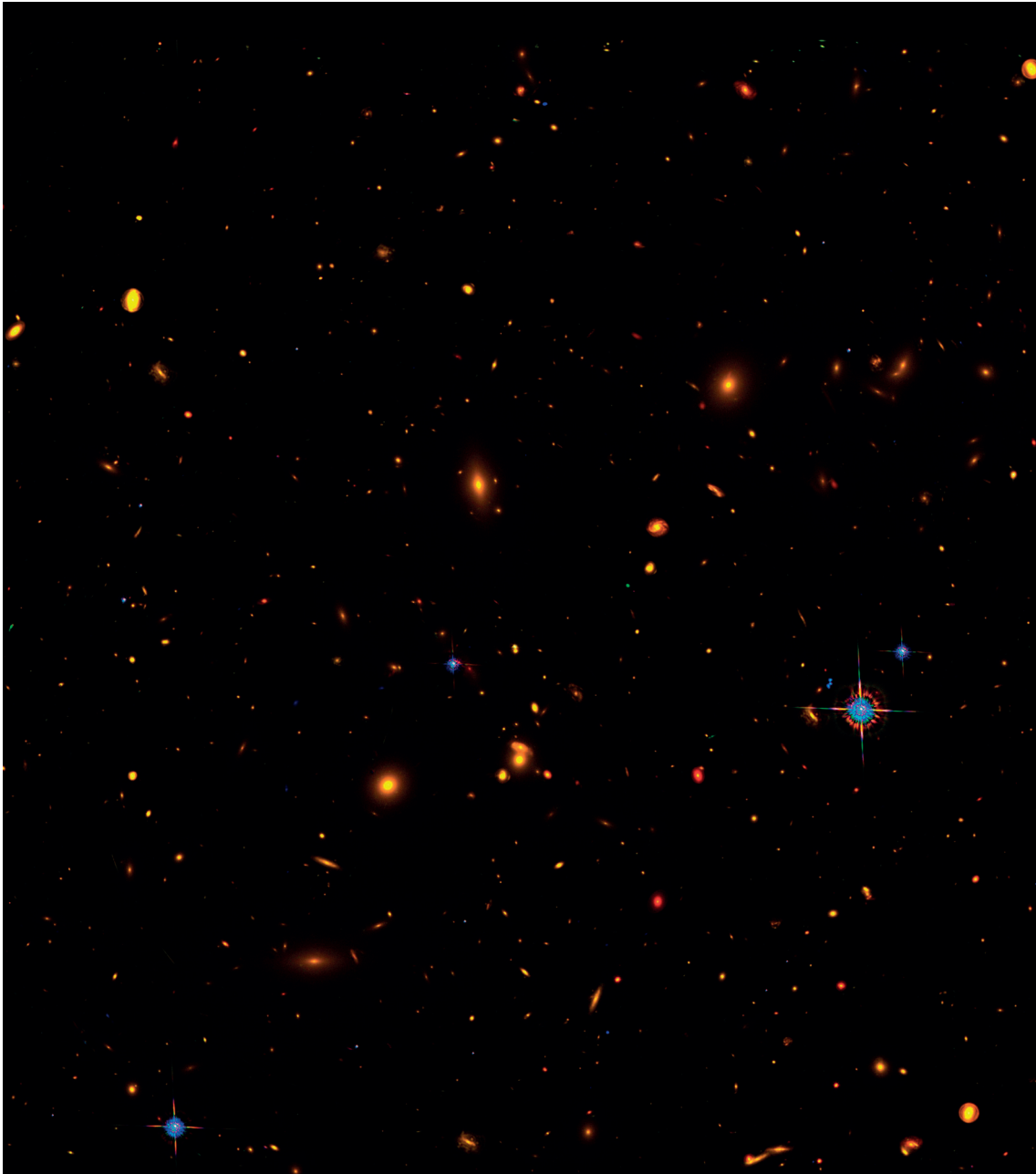Figure E.2 – Composite RGB image of the Ares cluster's blue sources, R_b.

Figure E.3 – Composite RGB image of the Ares cluster's red sources R_r.

F Galaxy populations from colour in the HFF

In this study of the `MuSCADeT` deblended HFF, I take interest in the contribution of light profiles of individual galaxies to the colour components as extracted by `MuSCADeT`. This is the preliminary stage of a study that intends to determine whether the components extracted using my method could bring new information about the understanding of galaxy population in massive galaxy clusters. The goal here is to understand what new information can be gained by using the colour separated components from `MuSCADeT` by making a comparison between the red and blue components of cluster and non-cluster-member galaxies. Given the early stage of the project, I show in this appendix the starting point of this investigation.

First, let us take a look at some of the star-forming galaxies in the parallel fields of HFF cluster 0416 after applying `MuSCADeT` to the whole image, as shown on figure F.1. In this figure, we see in each column, from top to bottom: the image of a galaxy in band F814w; the composite RGB image R_r, resulting in the subtraction of the blue component from Y; the original composite RGB image Y; the composite RGB image R_b, resulting from he subtraction of the red component from Y; and finally, the image in band F435w. We notice that the profile of the red component resulting from `MuSCADeT`'s extraction, is much smoother than the F814w or F435w bands. The features in band F435w are very similar to the ones found in the blue component of `MuSCADeT` with the difference that, visually, the SNR in F435w seems much higher than in the blue component. This is easily explained by the fact that the blue component image is the superimposition of three images representing the full surface brightness of blue stars, while F435w is only one image in a filter with a narrow band-pass (see figure 4.2). These images indicate that the result of `MuSCADeT`'s separation contains different information than the band images.

The properties of galaxies can be studied by looking at their distribution in a colour magnitude diagram. The colour of a galaxy can be expressed as the difference between its magnitude[1] in two different filters of interest and is usually represented as a function of the magnitude in one of these filters. It is common to represent colour-magnitude diagrams $B - R$ versus

---

[1]Magnitudes are defined as $Mag = -2.5 log_{10}(F) + 25$. Due to the logarithm, differences in magnitudes are equivalent to a ratio of fluxes.

R diagrams, where *B* is a galaxy's magnitude in a blue filter and *R* is its magnitude in a filter centred around a longer wavelength (red). In order to illustrate the difference of information between MuSCADeT and HST filters, with show HST the colour-magnitude diagram and the MuSCADeT colour-magnitude diagram. In the HST diagram, *B* is the magnitude in band F435w, Mag_{F435w} and R is the magnitude in band F814w, Mag_{F814w}. In the MuSCADeT diagram, *B* is the magnitude computed from the blue component, Mag_{Blue} and *R* is the magnitude in the red component Mag_{Red}. I show in figures F.2, F.3 and F.4 the HST diagrams (*top panel*) and the MuSCADeT diagrams (*bottom panel*) for the cluster fields of clusters MACS J0717, 0416 and 1149 respectively. The photometric redshifts of the galaxies were recovered from Shipley et al. (2018)'s catalogues and are shown in colours on the diagram. The redshift of the clusters are given in the captions. We limited our selection of galaxies to redshifts $z < 1$. In these diagrams, we clearly see the cluster galaxy members distributed as a line at $Mag_{F435w} - Mag_{F814w} > 2$ in the HST diagrams. In the MuSCADeT diagrams, we see the cluster members at $Mag_{Blue} - Mag_{Red} > 2$ and $Mag_{Red} < 31$, but a sub-population of the cluster members seems to extend to $Mag_{Blue} - Mag_{Red} > 5$, which is indicative of galaxies redder than the others.

These galaxies only appear redder thanks to MuSCADeT. If this trend is confirmed and correlates with other properties of the galaxies, such as their position in 3d, relative to the cluster, it could give more information about the evolution of galaxies inside the cluster, in particular with regard to their star formation history.
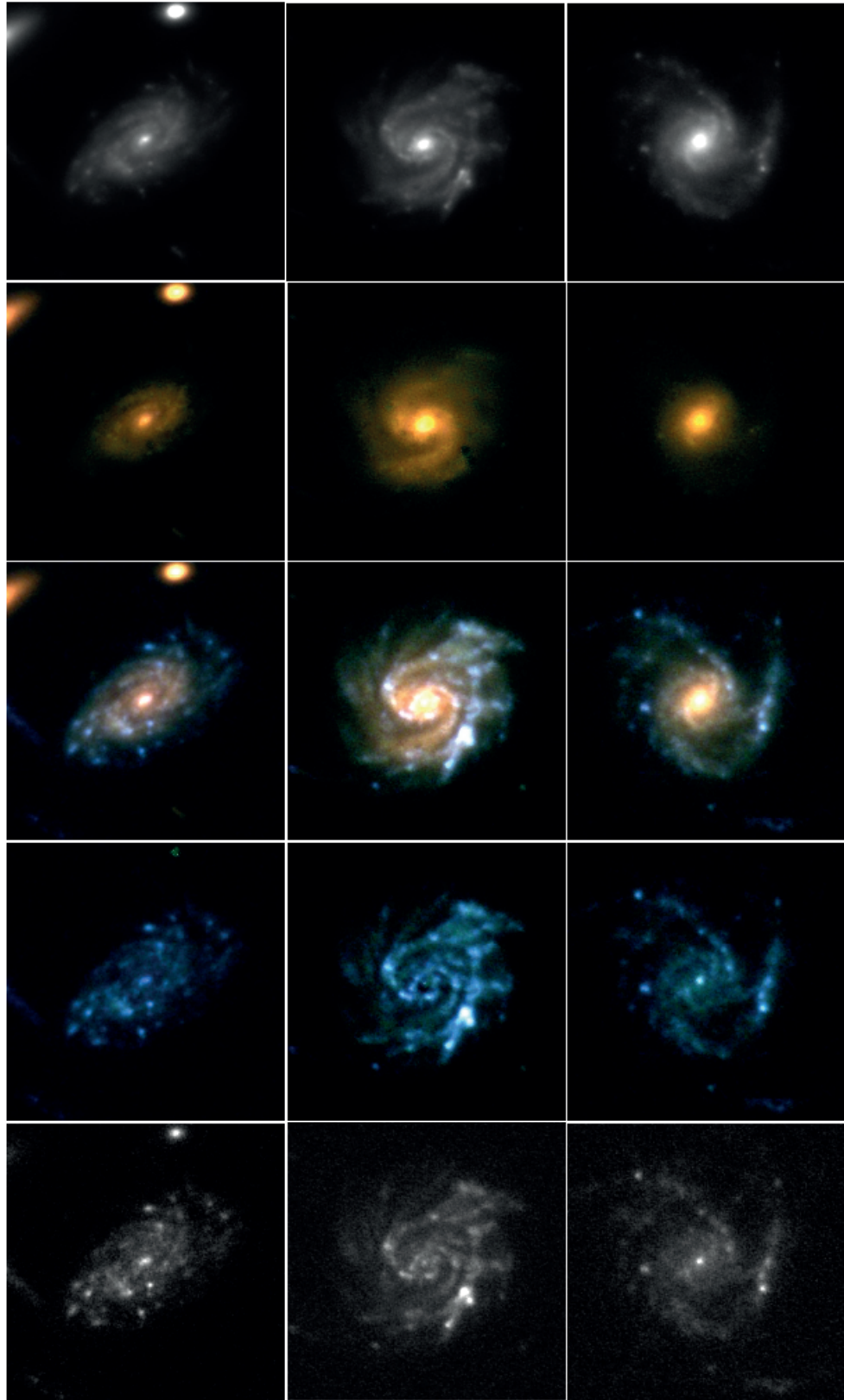
Figure F.1 – Examples of colour separation of colour components in individual galaxies of the HFF .
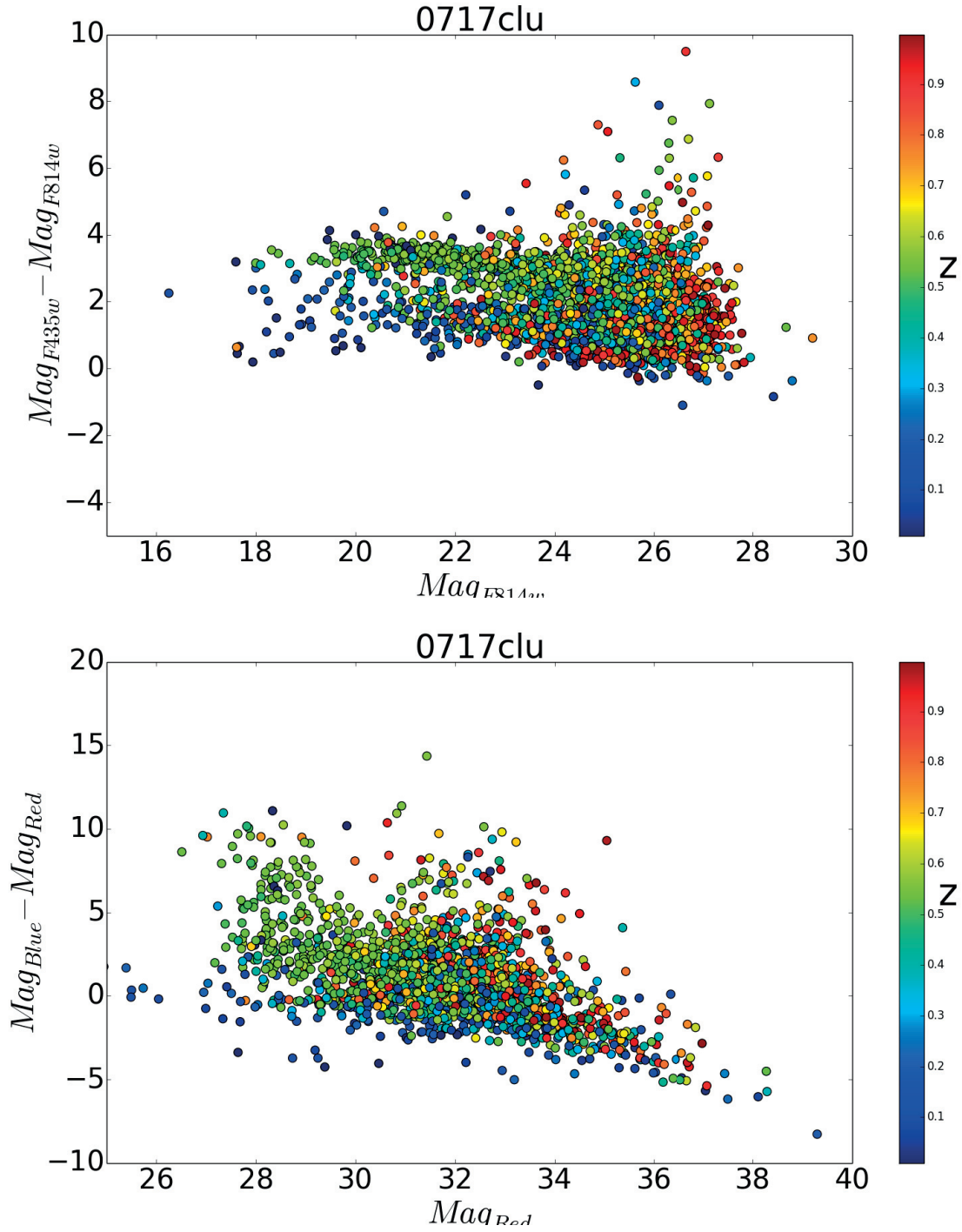
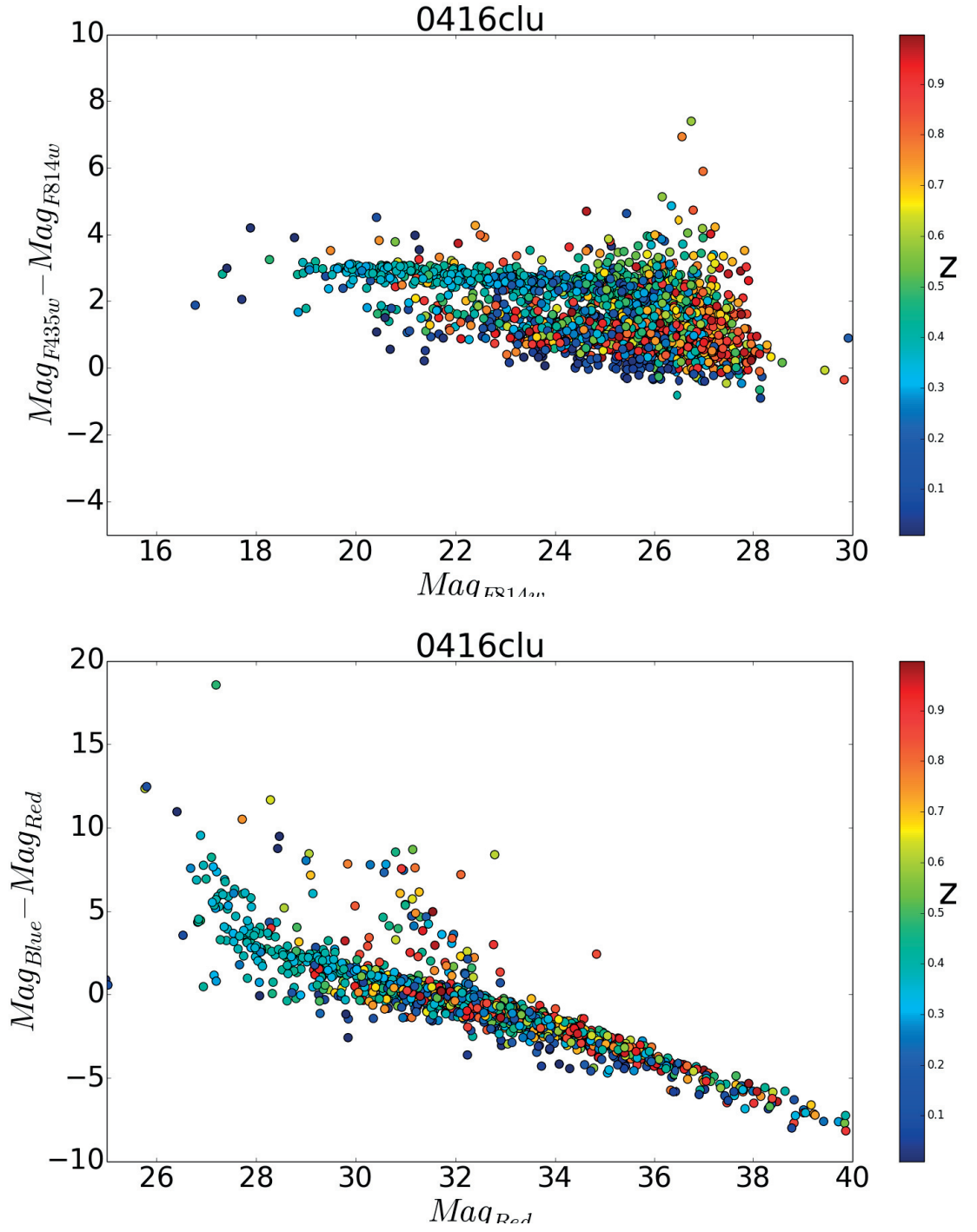Figure F.2 – Colour-magnitude diagrams of cluster 0717, $z_{clus} = 0.548$.

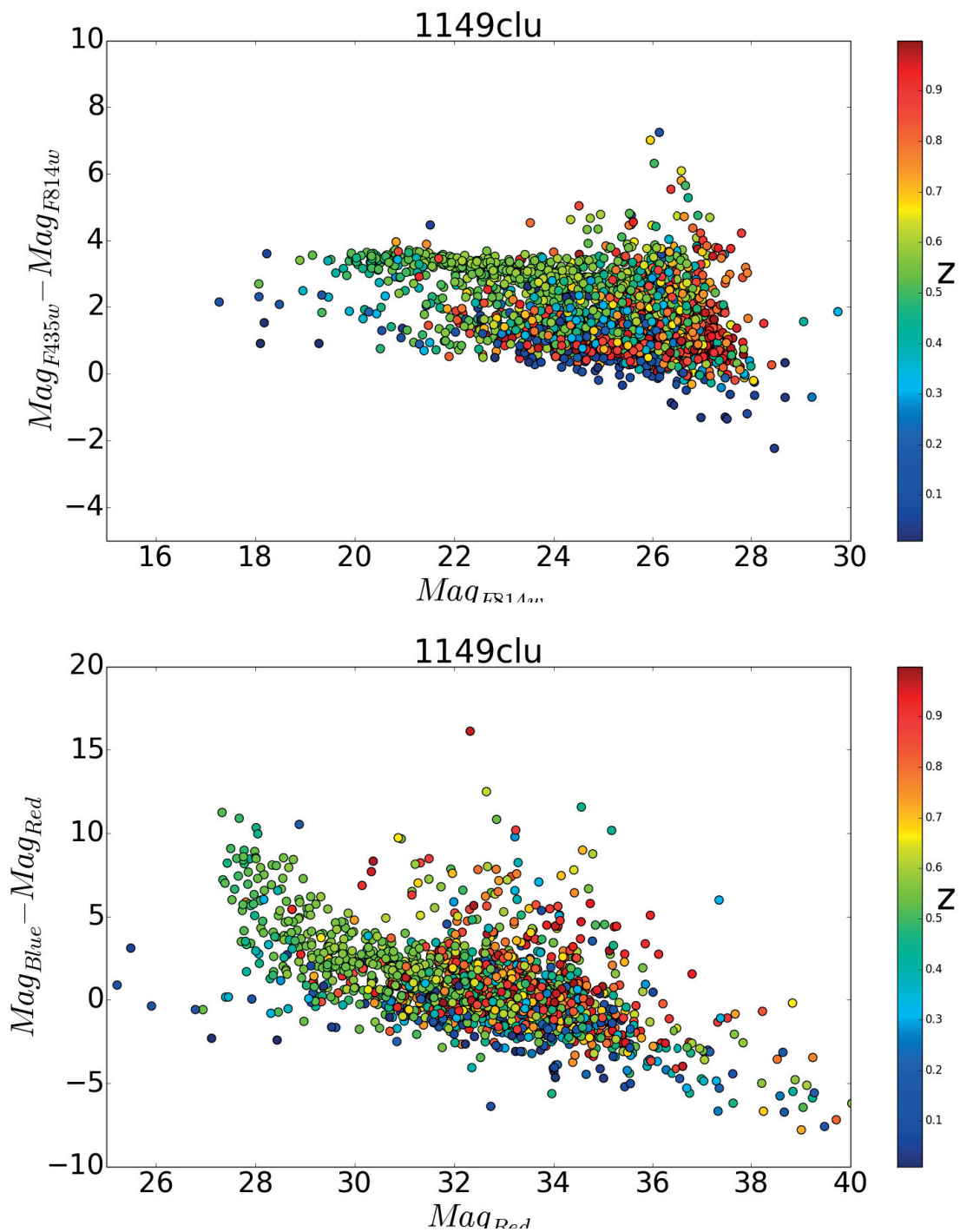Figure F.3 – Colour-magnitude diagrams of cluster 0416, $z_{clus} = 0.39$.

Figure F.4 – Colour-magnitude diagrams of cluster 1149, $z_{clus} = 0.544$.

G Supplementary material for SLIT

G.1 Lens mass density profiles

This section describes the profiles used to generate mass density distributions all throughout this publication. The following mathematical expressions give the computation of the mass density at location (x, y). In these formulas, κ_0 stands for the amplitude normalisation, γ is the slope of the profile, (x_0, y_0) is the center of the profile, q is the ellipticity, R_c is the core radius, and theta is the angle the semimajor axis of the profile forms with the horizontal axis of the image.

- Elliptical power law:

$$\kappa_{PL}(x, y) = \kappa_0 (2 - \frac{\gamma}{2}) \frac{q^{(\gamma - \frac{3}{2})}}{2(X^2 q^2 + Y^2)^{(\frac{\gamma-1}{2})}});\tag{G.1}$$

- Singular Isothermal Ellipsoid, SIE, which is another parametrisation of the Elliptical power law profile which we preferred in some examples:

$$\kappa_{SIE}(x, y) = \frac{\kappa_0^{\gamma-1}}{2(1 - \frac{1-q^2}{1+q^2})^{\frac{\gamma-1}{2}} (\frac{R_c^2}{1 - \frac{1-q^2}{1+q^2}} + X^2 + \frac{Y^2}{q^2})^{\frac{\gamma-1}{2}}}\tag{G.2}$$

- Sigular Isothermal Ellipsoid, SIS:

$$\kappa_{SIS}(x, y) = \frac{\kappa_0}{2\sqrt{(x - x_0)^2 + (y - y_0)^2}}\tag{G.3}$$

- softened power-law elliptical potential, SPEP (Barkana, 1998):

$$\kappa_{SPEP}(x, y) = \frac{\kappa_0^{\gamma-2}}{\gamma}(x^2 + \frac{y^2}{cos^2\theta} + R_c^2)^{\frac{\gamma-2}{2}} \tag{G.4}$$

$$\left((\gamma - 2)\frac{x^2 + \frac{y^2}{cos^4\theta}}{x^2 + \frac{y^2}{cos^2\theta} + R_c^2} + 1 + \frac{1}{cos^2\theta}\right) \tag{G.5}$$

With:

$$X = (x - x_0)cos(\theta) - (y - y_0)sin(\theta), \tag{G.6}$$
$$Y = (x - x_0)sin(\theta) - (y - y_0)cos(\theta). \tag{G.7}$$

G.2   Reconstructions for lens mass model optimisation

In this appendix, we show the average reconstructed lens and source light profiles as well as the corresponding truth profiles for reconstructions of a lens system with different slopes for a power law mass density profile. The profiles are shown in Fig G.1 & Fig G.2.
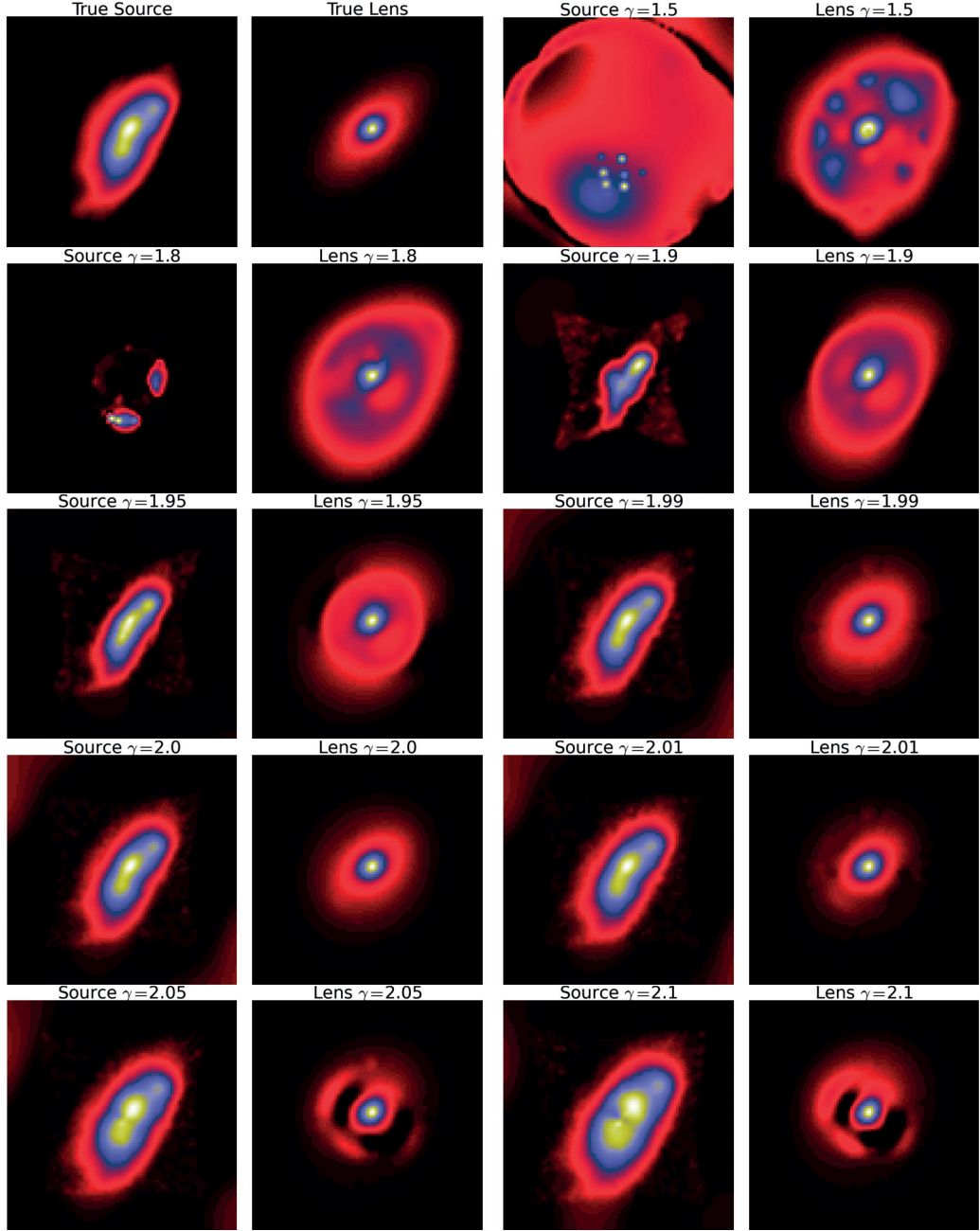
Figure G.1 – Reconstructions of lens and source light profiles for various values of mass density slope of a lens generated with a mass density slope of 2. The first two panels show the true source (left hand-side) and lens (right hand side) light profiles used to generate the simulated images. The other couples of panels from left to right and from top to bottom show the source and lens reconstructions for increasing values of $\tilde{\gamma}$.
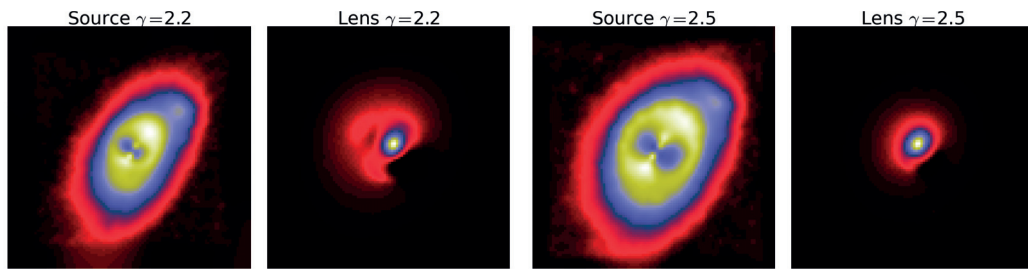
Figure G.2 – Continuation of Fig G.1.

Bibliography

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102

Adamo, A., Östlin, G., Bastian, N., et al. 2013, ApJ, 766, 105

Alard, C. 2006, ArXiv e-prints, astro

Alard, C. 2009, A&A, 506, 609

Arons, A. B. & Peppard, M. B. 1965, American Journal of Physics, 33, 367

Barkana, R. 1998, ApJ, 502, 531

Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291

Bayer, D., Chatterjee, S., Koopmans, L. V. E., et al. 2018, ArXiv e-prints, arXiv:1803.05952

Beck, A. & Teboulle, M. 2009, SIAM journal on imaging sciences, 2, 183

Bell, A. J. & Sejnowski, T. J. 1995, Neural computation, 7, 1129

Bellagamba, F., Tessore, N., & Metcalf, R. B. 2017, MNRAS, 464, 4823

Bernardi, M., Meert, A., Sheth, R. K., et al. 2013, MNRAS, 436, 697

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Birdi, J., Repetti, A., & Wiaux, Y. 2018, ArXiv e-prints [[arXiv]1801.02417]

Birrer, S. & Amara, A. 2018, ArXiv e-prints

Birrer, S., Amara, A., & Refregier, A. 2015, ApJ, 813, 102

Birrer, S., Amara, A., & Refregier, A. 2016, Journal of Cosmology and Astro-Particle Physics, 2016, 020

Bobin, J., Moudden, Y., Starck, J.-L., & Elad, M. 2006, Signal Processing Letters, IEEE, 13, 409

Bobin, J., Moudden, Y., Starck, J.-L., Fadili, J., & Aghanim, N. 2008, Statistical Methodology, 5, 307

Bibliography

Bobin, J., Rapin, J., Larue, A., & Starck, J.-L. 2015a, IEEE Transactions on Signal Processing, 63, 1199

Bobin, J., Starck, J., Fadili, J. M., Moudden, Y., & Donoho, D. L. 2007a, IEEE Transactions on Image Processing, 16, 2675

Bobin, J., Starck, J.-L., Fadili, J., & Moudden, Y. 2007b, Image Processing, IEEE Transactions on, 16, 2662

Bobin, J., Starck, J.-L., Sureau, F., & Basak, S. 2013, A&A, 550, A73

Bobin, J., Sureau, F., & Starck, J.-L. 2011, in Image Processing (ICIP), 2011 18th IEEE International Conference on, 1297–1300

Bobin, J., Sureau, F., & Starck, J.-L. 2015b, ArXiv e-prints [[arXiv]1508.07131]

Bobin, J., Sureau, F., Starck, J.-L., Rassat, A., & Paykari, P. 2014, A&A, 563, A105

Bolton, A. S., Brownstein, J. R., Kochanek, C. S., et al. 2012, ApJ, 757, 82

Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, The Astrophysical Journal, 638, 703

Bom, C. R., Makler, M., Albuquerque, M. P., & Brandt, C. H. 2017, A&A, 597, A135

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914

Bonvin, V., Tewes, M., Courbin, F., et al. 2016, A&A, 585, A88

Bradač, M., Schneider, P., Steinmetz, M., et al. 2002, A&A, 388, 373

Brault, F. & Gavazzi, R. 2015, A&A, 577, A85

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, Monthly Notices of the Royal Astronomical Society, 341, 13

Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, The Astrophysical Journal, 744, 41

Bullock, J. S. & Boylan-Kolchin, M. 2017, Annual Review of Astronomy and Astrophysics, 55, 343

Burke, W. L. 1981, ApJ, 244, L1

Candes, E. J., Romberg, J., & Tao, T. 2006, IEEE Transactions on Information Theory, 52, 489

Candes, E. J. & Tao, T. 2006, IEEE Transactions on Information Theory, 52, 5406

Candes, E. J., Wakin, M. B., & Boyd, S. P. 2008, Journal of Fourier analysis and applications, 14, 877

Cantiello, M., Jensen, J. B., Blakeslee, J. P., et al. 2018, ApJ, 854, L31

Cardoso, J.-F. 1999, Neural computation, 11, 157

Cava, A., Schaerer, D., Richard, J., et al. 2018, Nature Astronomy, 2, 76

Chambolle, A. & Dossal, C. 2015, Journal of Optimization Theory and Applications, 166, 968

Chambolle, A. & Pock, T. 2011, Journal of Mathematical Imaging and Vision, 40, 120

Chan, J. H. H., Suyu, S. H., Chiueh, T., et al. 2015, ApJ, 807, 138

Chan, J. Y. H., Leistedt, B., Kitching, T. D., & McEwen, J. D. 2017, IEEE Transactions on Signal Processing, 65, 5

Chatterji, S., Blackburn, L., Martin, G., & Katsavounidis, E. 2004, Classical and Quantum Gravity, 21, S1809

Chenot, C., Bobin, J., & Rapin, J. 2015, IEEE Signal Processing Letters, 22, 2172

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

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393

Coles, J. P., Read, J. I., & Saha, P. 2014, MNRAS, 445, 2181

Collett, T. E. 2015, ApJ, 811, 20

Combettes, P. & Wajs, V. 2005, Multiscale Modeling & Simulation, 4, 1168

Combettes, P. L. & Pesquet, J.-C. 2009, ArXiv e-prints [[arXiv]0912.3522]

Comon, P. 1994, Signal processing, 36, 287

Condat, L. 2013, Journal of Optimization Theory and Applications, 158, 460

Courbin, F., Eigenbrod, A., Vuissoz, C., Meylan, G., & Magain, P. 2005, in Gravitational Lensing Impact on Cosmology, Vol. 225, 297–303

Courbin, F., Faure, C., Djorgovski, S. G., et al. 2012, A&A, 540, A36

Courbin, F., Tewes, M., Djorgovski, S. G., et al. 2010, A&A, 516, L12

Dai, L., Venumadhav, T., Kaurov, A. A., & Miralda-Escudé, J. 2018, ArXiv e-prints, arXiv:1804.03149

Dalal, N. & Kochanek, C. S. 2002, ApJ, 572, 25

Diego, J. M., Broadhurst, T., Chen, C., et al. 2016, MNRAS, 456, 356

Donoho, D. & Stodden, V. 2004, in Advances in Neural Information Processing Systems 16, ed. S. Thrun, L. K. Saul, & B. Schölkopf (MIT Press), 1141–1148

Bibliography

Donoho, D. L. 2006, IEEE Transactions on Information Theory, 52, 1289

Donoho, D. L. & Huo, X. 2001, IEEE Transactions on Information Theory, 47, 2845

Donoho, D. L. & Johnstone, J. M. 1994, Biometrika, 81, 425

Donoho, D. L., Tsaig, Y., Drori, I., & Starck, J.-L. 2012, IEEE Transactions on Information Theory, 58, 1094

Dye, S. & Warren, S. J. 2005, ApJ, 623, 31

Dyer, C. C. & Roeder, R. C. 1980, ApJ, 238, L67

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

Easter, M., Kriegel, H., Sander, J., et al. 1996, in Proc. 1996 Int. Conf. Knowledge Discovery and Data Mining (KDD'96), Menlo Park: AAAI Press, 226–331

Eichner, T., Seitz, S., & Bauer, A. 2012, MNRAS, 427, 1918

Eigenbrod, A., Courbin, F., Vuissoz, C., et al. 2005, A&A, 436, 25

Einstein, A. 1905, Annalen der physik, 322, 132

Einstein, A. 1911, Annalen der Physik, 340, 898

Einstein, A. 1915a, Sitzungsber. preuss.Akad. Wiss, 47, 831

Einstein, A. 1915b, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, 778

Einstein, A. 1936, Science, 84, 506

Farrens, S., Starck, J.-L., & Ngolè Mboula, F. M. 2017, ArXiv e-prints

Faure, C., Kneib, J.-P., Covone, G., et al. 2008, The Astrophysical Journal Supplement Series, 176, 19

F.R.S., K. P. 1901, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2, 559

Fukushima, K. 1980, Biological Cybernetics, 36, 193

Gabay, D. 1983, Studies in Mathematics and Its Applications, 15, 299

Gabay, D. & Mercier, B. 1976, Computers & Mathematics with Applications, 2, 17

Gabor, D. 1946, Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of, 93, 429

Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175

Gilman, D., Birrer, S., Treu, T., & Keeton, C. R. 2017, ArXiv e-prints, arXiv:1712.04945

Goupillaud, P., Grossmann, A., & Morlet, J. 1984, Geoexploration, 23, 85 , seismic Signal Analysis and Discrimination III

Grillo, C., Eichner, T., Seitz, S., et al. 2010, ApJ, 710, 372

Grossmann, A. & Morlet, J. 1984, SIAM Journal on Mathematical Analysis, 15, 723

Gültekin, K., Cackett, E. M., Miller, J. M., et al. 2009, ApJ, 706, 404

Gwyn, S. D. J. 2012, AJ, 143, 38

Haar, A. 1910, Mathematische Annalen, 69, 331

Halkola, A., Hildebrandt, H., Schrabback, T., et al. 2008, A&A, 481, 65

Hartley, P., Flamary, R., Jackson, N., Tagore, A. S., & Metcalf, R. B. 2017, Monthly Notices of the Royal Astronomical Society, 471, 3378

Harvey, D., Kneib, J. P., & Jauzac, M. 2016, MNRAS, 458, 660

Harvey, D., Massey, R., Kitching, T., Taylor, A., & Tittley, E. 2015, Science, 347, 1462

Hewitt, J. N., Turner, E. L., Schneider, D. P., Burke, B. F., & Langston, G. I. 1988, Nature, 333, 537

Hezaveh, Y., Dalal, N., Holder, G., et al. 2016, Journal of Cosmology and Astroparticle Physics, 2016, 048

Hoag, A., Huang, K.-H., Treu, T., et al. 2016, ApJ, 831, 182

Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, P. 1989, in Wavelets. Time-Frequency Methods and Phase Space, ed. J.-M. Combes, A. Grossmann, & P. Tchamitchian, 286

Hopkins, P. F., Murray, N., & Thompson, T. A. 2009, MNRAS, 398, 303

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Huchra, J., Gorenstein, M., Kent, S., et al. 1985, AJ, 90, 691

Hyvarinen, A. 1999, IEEE Transactions on Neural Networks, 10, 626

Jackson, N. 2008, Monthly Notices of the Royal Astronomical Society, 389, 1311

Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, Monthly Notices of the Royal Astronomical Society, 471, 167

Bibliography

Jaki, S. L. 1978, Foundations of Physics, 8, 927

Jauzac, M., Clément, B., Limousin, M., et al. 2014, MNRAS, 443, 1549

Jauzac, M., Eckert, D., Schwinn, J., et al. 2016, MNRAS, 463, 3876

Jauzac, M., Richard, J., Jullo, E., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 1437

Jensen, J. B., Blakeslee, J. P., Gibson, Z., et al. 2015, ApJ, 808, 91

Jiang, M., Bobin, J., & Starck, J.-L. 2017, ArXiv e-prints, arXiv:1703.02650

Johnson, T. L., Sharon, K., Gladders, M. D., et al. 2017, ApJ, 843, 78

Jolliffe, I. 2011, in International encyclopedia of statistical science (Springer), 1094–1096

Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, A&A, 566, A63

Joseph, R., Courbin, F., & Starck, J. L. 2016, A&A, 589, A2

Jutten, C. & Herault, J. 1991, Signal processing, 24, 1

Kelly, P. L., Diego, J. M., Rodney, S., et al. 2018, Nature Astronomy, 2, 334

Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123

Kelly, P. L., Rodney, S. A., Treu, T., et al. 2016, ApJ, 819, L8

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, The Astrophysical Journal, 740, 102

Kneib, J.-P., Bonnet, H., Golse, G., et al. 2011a, LENSTOOL: A Gravitational Lensing Software for Modeling Mass Distribution of Galaxies and Clusters (strong and weak regime), Astrophysics Source Code Library

Kneib, J.-P., Bonnet, H., Golse, G., et al. 2011b, LENSTOOL: A Gravitational Lensing Software for Modeling Mass Distribution of Galaxies and Clusters (strong and weak regime), Astrophysics Source Code Library

Kochanek, C. S., Blandford, R. D., Lawrence, C. R., & Narayan, R. 1989, MNRAS, 238, 43

Koopmans, L. V. E. 2005, MNRAS, 363, 1136

Koopmans, L. V. E., Bolton, A., Treu, T., et al. 2009, ApJ, 703, L51

Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599

Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in Proc. SPIE, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J

Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895

Lanusse, F., Rassat, A., & Starck, J.-L. 2012, A&A, 540, A92

Lanusse, F., Starck, J.-L., Leonard, A., & Pires, S. 2016, A&A, 591, A2

LeCun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436 EP

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278

Lee, D. D. & Seung, H. S. 1999, Nature, 401, 788 EP

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Leonard, A., Lanusse, F., & Starck, J.-L. 2014, MNRAS, 440, 1281

Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113

Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97

Lynds, R. & Petrosian, V. 1986, in Bulletin of the American Astronomical Society, Vol. 18, 1014

Mahler, G., Richard, J., Clément, B., et al. 2018, MNRAS, 473, 663

Mahler, G., Richard, J., Clément, B., Lagattuta, D., & Patricio, V. 2015, Proceedings of the International Astronomical Union, 11, 779–780

Mallat, S. G. 1989, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674

Mao, S. & Schneider, P. 1998, MNRAS, 295, 587

Marshall, P. J., Hogg, D. W., Moustakas, L. A., et al. 2009, The Astrophysical Journal, 694, 924

Marshall, P. J., Treu, T., Melbourne, J., et al. 2007, ApJ, 671, 1196

Marshall, P. J., Verma, A., More, A., et al. 2016, Monthly Notices of the Royal Astronomical Society, 455, 1171

Massey, R., Harvey, D., Liesenborgs, J., et al. 2018, MNRAS, 477, 669

Massey, R., Williams, L., Smit, R., et al. 2015, MNRAS, 449, 3393

McEwen, J. D., Hobson, M. P., Mortlock, D. J., & Lasenby, A. N. 2007, IEEE Transactions on Signal Processing, 55, 520

Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, ArXiv e-prints, arXiv:1802.10157

Meneghetti, M. 2006, Lecture Notes (University of Heidelberg)

Meneghetti, M., Natarajan, P., Coe, D., et al. 2017, MNRAS, 472, 3177

Merlin, E., Amorín, R., Castellano, M., et al. 2016, A&A, 590, A30

Merten, J., Coe, D., Dupke, R., et al. 2011, MNRAS, 417, 333

Bibliography

Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2018, ArXiv e-prints, arXiv:1802.03609

Meyer, R. A., Delubac, T., Kneib, J.-P., & Courbin, F. 2017, ArXiv e-prints, arXiv:1711.01184

Moolekamp, F. & Melchior, P. 2017, ArXiv e-prints, arXiv:1708.09066

More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38

More, A., Verma, A., Marshall, P. J., et al. 2016, Monthly Notices of the Royal Astronomical Society, 455, 1191

Moreau, J. J. 1962, Comptes Rendus de l'Académie des Sciences (Paris), Série A, 255, 2897

Moreau, J.-J. 1965, Bull. Soc. Math. France, 93, 273

Murenzi, R. 1989, in Wavelets. Time-Frequency Methods and Phase Space, ed. J.-M. Combes, A. Grossmann, & P. Tchamitchian, 239

Narayan, R. & Bartelmann, M. 1996, ArXiv Astrophysics e-prints [astro-ph/9606001]

Newton, I. 1979, Opticks, or, a treatise of the reflections, refractions, inflections & colours of light (Courier Corporation)

Nierenberg, A. M., Treu, T., Wright, S. A., Fassnacht, C. D., & Auger, M. W. 2014, MNRAS, 442, 2434

Nightingale, J. W. & Dye, S. 2015, MNRAS, 452, 2940

Nightingale, J. W., Dye, S., & Massey, R. J. 2018, MNRAS

Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129

Oguri, M. 2010, PASJ, 62, 1017

Ohanian, H. C. 1983, ApJ, 271, 551

Ostrovski, F., McMahon, R. G., Connolly, A. J., et al. 2017, Monthly Notices of the Royal Astronomical Society, 465, 4325

Paatero, P. & Tapper, U. 1994, Environmetrics, 5, 111

Paraficz, D., Courbin, F., Tramacere, A., et al. 2016, A&A, 592, A75

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129

Pratley, L., McEwen, J. D., d'Avezac, M., et al. 2018, MNRAS, 473, 1038

Priewe, J., Williams, L. L. R., Liesenborgs, J., Coe, D., & Rodney, S. A. 2017, MNRAS, 465, 1030

Rapin, J., Bobin, J., Larue, A., & Starck, J.-L. 2013, Signal Processing Letters, IEEE, 61, 5620

Rapin, J., Bobin, J., Larue, A., & Starck, J.-L. 2014, Signal Processing Letters, IEEE

Ravanbakhsh, S., Lanusse, F., Mandelbaum, R., Schneider, J., & Poczos, B. 2016, ArXiv e-prints [[arXiv]1609.05796]

Refregier, A. 2003, MNRAS, 338, 35

Refsdal, S. 1964, MNRAS, 128, 307

Revaz, Y. & Jablonka, P. 2018, ArXiv e-prints, arXiv:1801.06222

Richard, J., Kneib, J.-P., Limousin, M., Edge, A., & Jullo, E. 2010, MNRAS, 402, L44

Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1980, ApJ, 238, 471

Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1978, ApJ, 225, L107

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. P. 2018, A&A, 611, A2

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schneider, P., Kochanek, C., & Wambsganss, J. 2006, Gravitational lensing: strong, weak and micro: Saas-Fee advanced course 33, Vol. 33 (Springer Science & Business Media)

Sebesta, K., Williams, L. L. R., Mohammed, I., Saha, P., & Liesenborgs, J. 2016, MNRAS, 461, 2126

Seidel, G. & Bartelmann, M. 2007, A&A, 472, 341

Shensa, M. J. 1992, IEEE Transactions on Signal Processing, 40, 2464

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Shipley, H. V., Lange-Vagle, D., Marchesini, D., et al. 2018, ApJS, 235, 14

Sluse, D., Surdej, J., Claeskens, J. F., et al. 2003, A&A, 406, L43

Soldner, J. 1804, Annalen der Physik, 370, 593

Somerville, R. S. & Primack, J. R. 1999, Monthly Notices of the Royal Astronomical Society, 310, 1087

Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29

Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98

Soucail, G., Fort, B., Mellier, Y., & Picat, J. P. 1987a, A&A, 172, L14

Soucail, G., Mellier, Y., Fort, B., Mathez, G., & Cailloux, M. 1987b, The Messenger, 50, 5

Bibliography

Springel, V. & Hernquist, L. 2003, Monthly Notices of the Royal Astronomical Society, 339, 289

Springel, V., Yoshida, N., & White, S. D. 2001, New Astronomy, 6, 79

Starck, J. ., Elad, M., & Donoho, D. L. 2005a, IEEE Transactions on Image Processing, 14, 1570

Starck, J., Fadili, J., & Murtagh, F. 2007, IEEE Transactions on Image Processing, 16, 297

Starck, J., Murtagh, F., & Fadili, J. 2015, Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis (Cambridge University Press)

Starck, J.-L., Donoho, D., & Elad, M. 2004, Redundant multiscale transforms and their application for morphological component separation, Tech. rep., CM-P00052061

Starck, J. L., Elad, M., & Donoho, D. L. 2005b, IEEE Transactions on Image Processing, 14, 1570

Starck, J. L., Moudden, Y., Abrial, P., & Nguyen, M. 2006, A&A, 446, 1191

Starck, J.-L. & Murtagh, F. 2006, Astronomical Image and Data Analysis (Springer-Verlag)

Starck, J.-L. & Murtagh, F. 2006, Astronomical Image and Data Analysis (Springer-Verlag)

Stockton, A. 1980, ApJ, 242, L141

Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590

Suyu, S. H. & Halkola, A. 2010, A&A, 524, A94

Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, The Astrophysical Journal, 750, 10

Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, MNRAS, 371, 983

Taylor, P., Massey, R., Jauzac, M., et al. 2017, MNRAS, 468, 5004

Tessore, N., Bellagamba, F., & Metcalf, R. B. 2016, MNRAS, 463, 3115

Tewes, M., Courbin, F., & Meylan, G. 2013, A&A, 553, A120

Tonry, J. & Schneider, D. P. 1988, AJ, 96, 807

Tonry, J. L. & Schechter, P. L. 1990, AJ, 100, 1794

Tramacere, A. & Vecchio, C. 2013, Astronomy & Astrophysics, 549, A138

Treu, T. 2010, Annual Review of Astronomy and Astrophysics, 48, 87

Treu, T., Dutton, A. A., Auger, M. W., et al. 2011, MNRAS, 417, 1601

Vegetti, S. & Koopmans, L. V. E. 2009, MNRAS, 392, 945

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969

Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341

Vegetti, S. & Vogelsberger, M. 2014, MNRAS, 442, 3598

Vetterli, M. 1986, Signal Processing (Elsevier), 10, 219

Vincent, E., Gribonval, R., & Fevotte, C. 2006, IEEE Transactions on Audio, Speech, and Language Processing, 14, 1462

Vū, B. C. 2013, Advances in Computational Mathematics, 38, 667

Wallis, C. G. R., Wiaux, Y., & McEwen, J. D. 2017, IEEE Transactions on Image Processing, 26, 5176

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

Warren, S. J. & Dye, S. 2003, ApJ, 590, 673

Wayth, R. B. & Webster, R. L. 2006, MNRAS, 372, 1187

Willis, J. P., Hewett, P. C., Warren, S. J., Dye, S., & Maddox, N. 2006, Monthly Notices of the Royal Astronomical Society, 369, 1521

Woiselle, A., Starck, J.-L., & Fadili, J. M. 2010, Applied and Computational Harmonic Analysis, 28, 171

Wright, A. H., Robotham, A. S. G., Driver, S. P., et al. 2017, MNRAS, 470, 283

Wuyts, E., Rigby, J. R., Gladders, M. D., & Sharon, K. 2014, ApJ, 781, 61

Zibulevsky, M. & Pearlmutter, B. A. 2001, Neural Computation, 13, 863

Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, ApJ, 762, L30

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Zwicky, F. 1937a, ApJ, 86, 217

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

Index

Rémy**Joseph**

288 avenue de Genève
01220
Divonne-les-Bains
France

remy.joseph@epfl.ch
remyjoseph.wordpress
github: herjy

Skills

Image processing, linear inverse problem solving, sparsity, **noise filtering**, Bayesian inference, **good oral communication**, machine learning, **programming**, OSs (macOSx, linux, Windows) , scientific writing, open science.

Languages

French: Mothertongue
English: fluent
German: Good oral and written comprehension
Chinese mandarin & French signed language: notions

Programming

Python, IDL, matlab, Latex, notions in C, java, R and html.

Interests

Bouldering, hiking, skiing, badminton and sports in general. Fencing: 20+ years of practice, first level coaching degree, competition, artistic fencing.

Experience

2017 **Reconstruction method for distorted images**   Design and implementation
Design and coding *of an algorithm for reconstruction of images of galaxies distorted by strong gravitational lensing effect.*

2016 **Colour separation of galaxies**   Design and implementation
Design and coding *of an algorithm for deblending of objects based on sparse decomposition of galaxies with different colours.*

2014 **Automated strong lens finding algorithm**   Design and implementation
Design and coding *of a new PCA-based technique for automated search of rare strong lenses in large data sets.*

Since 2014 **Public outreach speaker**   Public relations
Host for public visits of the observatory, including public outreach conferences on the observatory activities and public observations. Finalist at the EPFL MT180 tournament, 2017. (Watch the video on my website)

Since 2014 **Telescope observer and operator**   Data acquisition
More than 45 nights of photometric and spectrometric observations as observer and telescope operator on the 1.20m Euler telescope in La Silla, Chile.

Mentoring

Spring 2015 **Supervision of a master's project**
Colour separation of the Hubble Frontier Field images arxiv1603.00473. Rewarded by the Hausmann award for master's projects (5,000 CHF)

Spring 2014 **Supervision of two undergraduate students**
Search for strong gravitational lenses in the CFHTLS ground based survey, investigating **machine learning** *techniques (arxiv1605.04309,arxiv1403.1063)*

Since 2014 **Teaching assistant for undergraduate exercise session**
Astrophysics II: physics basics of astrophysics

Education

Since 2014 **Ph.D. candidate in Cosmology**   LASTRO | EPFLausanne
Linear optimisation for strong gravitational lens modelling (Defense scheduled for October 2018)

2013–2014 **M.Sc. Space Science and Engineering**   MSSL, University College London
Majoring in high energy astrophysics and plasma physics

2012–2013 **Year-long internship in cosmology**   CEA/Cosmostat & EPFL/LASTRO
On measurement of galaxy shapes for Euclid and design of an automated strong lens finder

2010–2012 **Engineering degree in Telecommunications**   Telecom Bretagne, France
Majoring in image processing

2007–2010 **Bachelor in fundamental science**   Lycée Claude Fauriel, France
Competitive preparation to Engineering school entrance exam, with major in physics and mathematics, minor in chemistry and philosophy.

Publications

- **R. Joseph**, F. Courbin, J.-L. Starck, S. Birrer
  Sparse Lens Inversion Technique (SLIT): lens and source separability from linear inversion of the source reconstruction problem
  Accepted A&A (09/2017), Arxiv 1809.09121

- **R. Joseph**, F. Courbin, R.B. Metcalf, C. Giocoli, P. Hartley, N. Jackson, F. Bellagamba, J.-P. Kneib, L. Koopmans, G. Lemson, M. Meneghetti, G. Meylan, M. Petkova, S. Pires
  A PCA-based automated finder for galaxy-scale strong lenses
  Astronomy & Astrophysics, V566, A63 (2014)

- **R. Joseph**, F. Courbin, , J.-L. Starck
  Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT): Deblending colourful objects
  Astronomy & Astrophysics, V589, A2 (2016)

- D. Paraficz, F. Courbin, A. Tramacere, **R. Joseph**, R.B. Metcalf, J.-P. Kneib, P. Dubath, D. Droz, F. Filleul, D. Ringeisen, C. Schäfer
  The PCA Lens-Finder: application to CFHTLS
  Astronomy & Astrophysics, V592, A75 (2016)

- P. Taylor, R. Massey, M. Jauzac, F. Courbin, D. Harvey, **R. Joseph**, A. Robertson
  A test for skewed distributions of dark matter and a possible detection in galaxy cluster Abell 3827
  Monthly Notices of the Royal Astronomical Society, V468, I4 (2017)

- R. Massey, D. Harvey, J. Liesenborgs, J. Richard, S. Stach, M. Swinbank, P. Taylor, L. Williams, D. Clowe, F. Courbin, A. Edge, H. Israel, M. Jauzac, **R. Joseph**, E. Jullo, T. D. Kitching, A. Leonard, J. Merten, D. Nagai, J. Nightingale, A. Robertson, L. Javier Romualdez, P. Saha, R. Smit, S. Ieng Tam, E. Tittley
  Dark matter dynamics in Abell 3827: new data consistent with standard Cold Dark Matter
  Submitted to Monthly Notices of the Royal Astronomical Society (2017)

- M. Rexroth, J.-P. Kneib, **R. Joseph**, J. Richard, R. Her
  IFS-RedEx, a redshift extraction software for integral-field spectrographs: Application to MUSE data
  Submitted to Monthly Notices of the Royal Astronomical Society

Talks

May 2016	**Deblending colorful objects: a morpho-spectral component analysis method for automated source separation**	Cosmo21: Chania, Greece
	Contributed talk	
Jully 2016	**Automated colour-based deblending of the Hubble Frontier Fields with MuSCADeT**	EWASS 2016: Athens, Greece
	Contributed talk	
June 2017	**Sparse methods for strong gravitational lensing: lens deblending and lensed source reconstruction**	Aosta , Italy
	Contributed talk	
Jully 2017	**Analysis of galaxies' colour components in the Hubble frontier fields clusters**	GECO-lam: Aix en Provence, France
	Contributed talk	