



SLITronomy: towards a fully wavelet-based strong lensing inversion technique

A. Galan, A. Peel, R. Joseph, F. Courbin, J.-L Starck

► To cite this version:

A. Galan, A. Peel, R. Joseph, F. Courbin, J.-L Starck. SLITronomy: towards a fully wavelet-based strong lensing inversion technique. *Astronomy and Astrophysics - A&A*, 2021, 647, pp.A176. 10.1051/0004-6361/202039363 . hal-03098818

HAL Id: hal-03098818

<https://hal.science/hal-03098818>

Submitted on 10 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SLITRONOMY: Towards a fully wavelet-based strong lensing inversion technique

A. Galan¹, A. Peel¹, R. Joseph², F. Courbin¹, and J.-L. Starck³

¹ Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
e-mail: aymeric.galan@epfl.ch

² Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

³ AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, 91191 Gif-sur-Yvette, France

Received 7 September 2020 / Accepted 3 December 2020

ABSTRACT

Strong gravitational lensing provides a wealth of astrophysical information on the baryonic and dark matter content of galaxies. It also serves as a valuable cosmological probe by allowing us to measure the Hubble constant independently of other methods. These applications all require the difficult task of inverting the lens equation and simultaneously reconstructing the mass profile of the lens along with the original light profile of the unlensed source. As there is no reason for either the lens or the source to be simple, we need methods that both invert the lens equation with a large number of degrees of freedom and also enforce a well-controlled regularisation that avoids the appearance of spurious structures. This can be beautifully accomplished by representing signals in wavelet space. Building on the Sparse Lens Inversion Technique (SLIT), we present an improved sparsity-based method that describes lensed sources using wavelets and optimises over the parameters given an analytical lens mass profile. We applied our technique on simulated HST and E-ELT data, as well as on real HST images of lenses from the Sloan Lens ACS sample, assuming a lens model. We show that wavelets allowed us to reconstruct lensed sources containing detailed substructures when using both present-day data and very high-resolution images expected from future thirty-metre-class telescopes. In the latter case, wavelets moreover provide a much more tractable solution in terms of quality and computation time compared to using a source model that combines smooth analytical profiles and shapelets. Requiring very little human interaction, our flexible pixel-based technique fits into the ongoing effort to devise automated modelling schemes. It can be incorporated in the standard workflow of sampling analytical lens model parameters while modelling the source on a pixelated grid. The method, which we call SLITRONOMY, is freely available as a new plug-in to the modelling software LENSTRONOMY.

Key words. gravitation – gravitational lensing: strong – methods: data analysis – techniques: image processing – galaxies: high-redshift – galaxies: structure

1. Introduction

Gravitational lensing offers a window into the nature of many fundamental properties of our Universe. The fortuitous alignment of a distant luminous object (i.e. source) and a massive foreground structure (i.e. lens) along the line of sight can produce striking visual distortions that allow us to constrain both astrophysical and cosmological parameters. By magnifying the distant source, strong lensing can reveal detailed structures that would otherwise be unresolvable. Cluster-scale lenses allow us to study the large-scale distribution of mass and the hierarchical evolution of dark matter halos (Natarajan et al. 2007; Bhattacharya et al. 2013; Han et al. 2015). Galaxy-scale lenses give insights into smaller scales, such as a substructure within the main deflector (e.g. Amara et al. 2006; Vegetti et al. 2010; Hezaveh et al. 2016; Birrer et al. 2017; Ritondale et al. 2019; Gilman et al. 2020), as well as population-level properties of galaxies when large samples can be uniformly selected (Bolton et al. 2006; Brownstein et al. 2012; Gavazzi et al. 2012). Strongly lensed quasars are particularly useful to constrain black hole evolution over a wide range of redshifts (e.g. Ding et al. 2021) and to derive high-precision constraints on the Hubble constant (Wong et al. 2020; Shajib et al. 2020a; Birrer et al. 2020, now referred as the TDCOSMO collaboration).

All of the above applications require lens modelling tools. If the underlying cosmology needed to compute distances to the lens and source is not exactly known, models are subject to degeneracies, to modelling assumptions (Xu et al. 2016; Sonnenfeld 2018; Blum et al. 2020), and/or to the inherent geometry and physics of the problem (Saha 2000; Schneider & Sluse 2013). As a result, for any given rescaling of the lens mass, there exists a corresponding rescaling of the lensed source that leads to identical lensing observables apart from the time delays. These degeneracies disappear if the underlying cosmology is known, but even in this case, lens modelling tools must, in practice, address several key points to work efficiently: (1) reliable reconstruction of the lens mass and light, (2) reliable reconstruction (de-lensing) of the source light, and (3) efficient deblending of the lens and source light.

In this work, we focus on the second point, namely source light reconstruction. Various techniques to model the light of the background source have been developed over the years. There are two main families of techniques, which we distinguish as ‘analytic’ and ‘pixel-based’ methods. While the former assumes a set of analytic functions through which the source is represented, the latter uses pixels directly to describe surface galaxy light distribution. We do not use the term ‘free-form techniques’

for pixel-based methods, as a pixelated surface brightness is already a constraining assumption compared to the continuous truth (e.g. Tagore & Jackson 2016).

There are many public software packages that implement analytic methods for strong lens analysis. For example, LENSTOOL (Kneib et al. 2011) and GLAFIC (Oguri 2010) provide specialised tools for cluster-scale lenses. For galaxy-galaxy lenses, dedicated software has been developed, such as GRAVLENS (Keeton 2001), PIXELEN (Williams & Saha 2004), AUTOLENS (Nightingale et al. 2018), and LENSTRONOMY (Birrer & Amara 2018). Typical analytic functions include variations of power-law or cored profiles for the lensing mass, while more elaborate ensembles of basis functions like shapelets (Bernstein & Jarvis 2002; Refregier 2003; Tagore & Keeton 2014; Birrer et al. 2015) can be used to capture the wider dynamic range of features often exhibited by high-redshift galaxies, especially when they are star-forming. While analytical methods are fast, they have two main disadvantages. First, complex light distributions like galaxies with resolved small-scale features or mergers are poorly modelled. Second, it requires lots of trial and error to infer the initial parameters describing the source shape, hence making it hard to automate these techniques.

Recent developments in deep learning have enabled neural networks, which can be seen as a collection of nested analytic functions of many (sometimes millions of) parameters, to become a viable tool for strong lens modelling. The work of Hezaveh et al. (2017) (and follow-up papers Perreault Levasseur et al. 2017; Morningstar et al. 2018, 2019), along with Pearson et al. (2019), have shown promising initial results on simulated data. Coupled with GPUs for training, neural networks can offer vast speed improvements compared to other techniques, although their applicability in practice hinges crucially on the quality of the training data. A key to modern neural network optimisation is automatic differentiation, which also facilitates the creation of fully differentiable pipelines capable of efficiently exploring the large parameter space of strong lens models (Chianese et al. 2020).

Within the family of pixel-based techniques, to which this work belongs, Kochanek & Narayan (1992) described one of the first methods, which is based on the CLEAN algorithm widely used for processing radio astronomy data (Högbom 1974). Later, Warren & Dye (2003) established the popular semi-linear inversion (SLI) formalism, where the (under-constrained) source surface brightness is optimised through regularised inversion. Its update to include adaptive gridding (Dye & Warren 2005) solved several issues inherent to the SLI method. Suyu et al. (2006) brought further improvements to the method using the Bayesian formalism to objectively find hyper-parameters that are required for the lens inversion (GLEE modelling code, Suyu & Halkola 2010). Various adaptive pixel grids have been tested to mitigate biases that can arise when inferring lens model parameters while still maintaining a tractable computation time. For instance, the algorithm by Vegetti & Koopmans (2009) uses a Delaunay tessellation, combined with an iterative scheme, to find deviations from a smooth gravitational potential, and Nightingale & Dye (2015) uses randomly initialised h -means clustering to define the adaptive grid. More recently, Rizzo et al. (2018) and Powell et al. (2021) improved the technique of Vegetti & Koopmans (2009) with the joint modelling of resolved kinematics and application to interferometry data.

The above-mentioned methods all primarily differ from each other in their constraints on the galaxy surface brightness distribution, independent of the imaging data. These constraints

restrict the large parameter space common to pixel-based methods, thus reducing degeneracies and improving convergence of the algorithm. However, they often lack a clear physical interpretation motivated by light distributions of real galaxies. Additionally, they introduce new hyper-parameters, whose optimal values can vary significantly depending on the data and are hence complex to handle.

Drawing from standard concepts in the field of image processing, Joseph et al. (2019, hereafter J19) – introduced improved assumptions on light distributions through a technique based on sparsity and wavelets called the Sparse Linear Inversion Technique (SLIT), which mitigates both issues mentioned above. The present paper introduces an updated and expanded open-source implementation of SLIT, which serves as a plug-in to the modelling software LENSTRONOMY. All codes described here are publicly available, and data along with python notebooks used to generate results and figures of this paper are available online¹.

As a validation of the algorithm, we reconstruct high-resolution source galaxies from of the Sloan Lens Advanced Camera for Surveys (SLACS) sample of strong galaxy-galaxy lenses (Bolton et al. 2006). This dataset has been extensively studied for constraining population-level properties of massive elliptical galaxies. However, little discussion has focused on the quality of source reconstructions, as it is commonly assumed that constraining deflector properties to the percent level can be achieved without high-fidelity source reconstructions. Based on the recent (re)modelling of a subset of the SLACS sample by Shajib et al. (2020b), we compare for the first time source reconstructions from analytical and pixel-based methods on this challenging dataset, consistently in the same modelling environment.

We further apply our modelling method to simulated data that is representative of the quality achievable by future thirty-meter-class telescopes. We show that our method is particularly well-suited to such extremely high resolution imaging data and that modelling the source light profile analytically quickly leads to an intractable problem in practice.

Our paper is organised as follows. We first review some basics of the strong gravitational lensing formalism and its linear approximation in Sect. 2. In Sect. 3 we give an overview of the method and theory behind the original SLIT method. We discuss key improvements in Sect. 4 and describe our new code implementation, which we call SLITRONOMY. In Sect. 5, we apply our technique to source reconstruction problems: first on simulated data, and then on a subset of SLACS lenses. We consistently compare the results in both cases with reconstructions using shapelets. To demonstrate the feasibility of the approach, we show sampled posteriors over lens model parameters. In Sect. 6, we proceed similarly on E-ELT simulated data and show how sparse reconstruction is suited to such high-resolution data. Finally we summarise and conclude in Sect. 7.

2. Theoretical aspects of strong gravitational lens inversion

2.1. Strong gravitational lensing

Gravitational lensing arises when a massive foreground object lies on or near the line-of-sight connecting an observer and a distant luminous source. The lensing effect is considered strong when the lens galaxy (or cluster of galaxies) is so massive that

¹ https://github.com/aymgal/SLITronomy-papers/tree/master/paper_I

it produces high order deformations, possibly leading to several images, of the background source². In this case, the source light rays are deflected from their original paths, and the lens equation (see below) displays several solutions.

In the following we write the redshift of the lensed source as z_s and the redshift of the lensing galaxy, also called the deflector, as z_d . We introduce the lensing potential $\psi(\theta)$, obtained by rescaling and projecting the three-dimensional gravitational potential on the lens plane, then evaluated at angular position θ on the coordinate grid. Light rays from the source are deflected by the deflection angle field α , which is directly related to the projected potential through its first derivatives: $\alpha = \nabla\psi$. The lens equation gives the angular coordinates of deflected light rays from a source at position β for every position θ :

$$\beta = \theta - \alpha(\theta). \quad (1)$$

The second derivatives of the potential give the dimensionless surface mass density, $\kappa = \nabla^2\psi/2$, of the deflector, also called the convergence.

Importantly, we note that Eq. (1) is, for most choices of deflection field α , non-linear with respect to the two coordinate variables of θ , even when the source position and gravitational potential are known. As a consequence, no root-finder algorithm is guaranteed to find a complete set of solutions if not initialised close to a suitable location in parameter space.

2.2. Linear approximation

Equation (1) allows us to determine the true (i.e. unlensed) position of a source given its lensed image and a model of the lensing mass. This inverse operation, however, requires solving an equation that may not have a general closed-form solution. Moreover, in the strong lensing regime, the equation is likely to have multiple solutions corresponding to duplicate and magnified images. For this reason, efficient modelling methods do not typically solve the lens equation for θ , but rather evaluate analytical profiles on a grid transformed by Eq. (1) or instead describe surface brightness directly on a grid in the source plane. The latter option is made possible with the linear framework developed in Warren & Dye (2003).

The lensing phenomenon smoothly defined by Eq. (1) can be formulated as a linear lensing operator F acting on an image s of source intensity values. In this formulation, s is a flattened vector of the underlying two-dimensional image. The action of F on s results in a lensed version of s . We note that while the mapping between the image and source planes is formulated as an algebraic linear operation, the intrinsic non-linear nature of lensing (i.e. image multiplicity) is retained. Additional observed features, such as light from the deflector, instrumental blurring, and other contaminants along the line of sight, can be straightforwardly included in a linear fashion. We can thus write a model y for strong lens imaging data as

$$y = H(Fs + g + \text{other components}) + \text{noise}, \quad (2)$$

where g is the surface brightness of the lens galaxy (as a flattened vector), and H accounts for seeing effects and/or blurring by the instrumental point spread function (PSF). Other components of the model can include, for instance, the multiple images of a background quasar or satellite galaxies near the main deflector. Several pixel-based methods to model strong lenses are

based on Eq. (2), such as the semi-linear inversion method (SLI, Warren & Dye 2003) and the works of Vegetti & Koopmans (2009) and Tagore & Keeton (2014). Software implementations can be found in GLEE (Suyu & Halkola 2010) and AUTOLENS (Nightingale et al. 2018). The present work relies on an independent implementation of the pixel-based formalism. Furthermore, our primary focus is on the source reconstruction problem at fixed lens model (i.e. fixed lensing operator F). The task of optimising F – ideally using a similarly flexible method – along with light components requires several iterations, as the lens model behaves non-linearly. We give in Sect. 5.3, however, an example of computing the lens model using the standard approach of sampling analytical parameters, solving the pixel-based source reconstruction at each iteration.

We note that an additional hyper-parameter³ is implicitly assumed in Eq. (2), due the pixelated nature of the linear approximation: the choice of source plane resolution, or alternatively the ratio of data to source resolution. For clarity, we introduce r_{pix} as

$$r_{\text{pix}} \equiv \frac{\text{source pixel size}}{\text{image pixel size}}. \quad (3)$$

As noted in Suyu et al. (2006), the source pixel size can be roughly related to the average lensing magnification. As it will be introduced later in the section, a large r_{pix} introduces more degrees of freedom in the model, which may not be constrained by the data alone. The choice of external constraints informed by the properties of the source light distribution allow for larger r_{pix} values while keeping the modelling accurate. The technique introduced in J19 and the further improvements we provide in this work have been designed to allow for exactly this.

2.3. An underconstrained problem

Given an observation \tilde{y} , our goal is to solve Eq. (2) by finding the most suitable s and g – ignoring any other luminous components – for a known H in the presence of noise. This amounts to minimising the difference between the data and our model y , commonly quantified by the function

$$\mathcal{L}(\tilde{y}, s, g) = \frac{1}{2} \|\tilde{y} - y\|^2, \quad (4)$$

where $\|\cdot\|$ is the Euclidean norm. Being strictly convex and differentiable, this norm is useful as a data fidelity term in optimisation algorithms.

Although mathematically simple, the linear formulation as described in Sect. 2.2 leads to a highly underconstrained problem. Data pixel coordinates mapped to the source plane according to Eq. (1) do not coincide anymore with a cartesian coordinate grid. The choice of how to discretise the source plane can introduce errors on the reconstruction. A common technique to reduce such errors, which we adopt, is to increase the resolution of the grid on which we reconstruct the source galaxy (i.e. increase r_{pix}). Combined with constraints that promote smoothness (see Sect. 2.4), this allows for more accurate conservation of surface brightness but also introduces more unconstrained degrees of freedom, as a subset of source pixels may not be mapped to any data pixel.

To solve the underconstrained problem, it is therefore necessary to somehow restrict the space of possible solutions. We

² Image multiplicity is sometimes viewed as a necessary condition for the lensing effect to be considered strong.

³ Depending on the method, other hyper-parameters can be introduced (e.g. controlling the mapping when using adaptive gridding strategies).

can do this by leveraging independent knowledge regarding the unknowns that is complementary to the imaging data itself. External knowledge of this sort serves to regularise the problem, and we denote the restrictive action of regularisation on a model component as $\mathcal{P}(\cdot)$. The augmented minimisation problem can then be written

$$\arg \min_{\mathbf{s}, \mathbf{g}} [\mathcal{L}(\tilde{\mathbf{y}}, \mathbf{s}, \mathbf{g}) + \mathcal{P}(\mathbf{s}) + \mathcal{P}(\mathbf{g})]. \quad (5)$$

Equation (5) can also be understood in a Bayesian context, with \mathcal{L} corresponding to the log-likelihood distribution and \mathcal{P} corresponding to log-prior probabilities based on, for example, physical arguments or a previous solution to the problem.

2.4. Regularisation strategies

Numerous strategies to regularise source galaxy reconstruction have been explored in the literature. A fundamental regularisation shared by all pixel-based techniques arises implicitly, as noted by [Tagore & Jackson \(2016\)](#), in choosing to express surface brightness on a pixelated grid rather than continuously. However, for complex sources, such as star-forming galaxies, still further regularisation is required.

The maximum entropy regularisation method (MEM) was used in [Wallington et al. \(1996\)](#) to recover astronomical sources from noisy images. The goal of MEM is to maximise the total Shannon-Jaynes entropy $\mathcal{P}(\mathbf{s}) \propto \sum_k s_k \ln s_k$ of the source pixels $\mathbf{s} = \{s_k\}$. This constraint accommodates incomplete data and ensures positivity of the solution, but it finds its minimum when all pixels s_k share the same value. This contradicts the goal of accurately reconstructing sources that show a large dynamical range, which is often the case in practice. In addition, the inversion problem becomes non-linear. To retain linearity, the similar ℓ_2 norm regularisation, that is to say $\mathcal{P}(\mathbf{s}) \propto \sum_k s_k^2$, is often used (e.g. [Warren & Dye 2003](#); [Suyu et al. 2006](#)), although non-negativity is no longer guaranteed.

Higher-order choices such as gradient and curvature regularisations are designed to penalise large first and second spatial derivatives, respectively, of the reconstructed image ([Suyu et al. 2006](#)). The smoothness of the solution increases with increasing regularisation order. Methods based on (irregular) adaptive grids have used algorithm-specific forms for $\mathcal{P}(\mathbf{s})$ (e.g. [Vegetti & Koopmans 2009](#); [Nightingale & Dye 2015](#)), and still other families of regularisation have been applied in a Bayesian framework ([Brewer & Lewis 2006](#)).

Machine learning techniques, such as neural networks, are also effective regularisation methods. More specifically, given a suitable dataset of galaxies that are close enough to real observations, supervised learning techniques are expected to extract complex structural information about the light distribution, which subsequently regularises the model (see e.g. [Morningstar et al. 2019](#); [Pearson et al. 2019](#); [Chianese et al. 2020](#)).

Each method has its advantages and disadvantages, and one may be more suitable than another depending on the quality of the data, the intrinsic morphology of the source and the science case. As simple convex functions, MEM and ℓ_2 norm regularisations have analytic derivatives that are useful in gradient descent algorithms but may overfit the data. Higher-order expressions reduce the risk of overfitting but may be time-consuming to compute and fail to properly account for local variations in the source. Moreover, regularisation from deep learning can be severely limited by the realism of training datasets. We describe our choice of regularisation in detail in Sect. 3.2.

3. The Sparse Linear Inversion Technique (SLIT)

This work aims to improve and expand the source inversion framework introduced in [J19](#), which pioneered the use of wavelets and sparse regularisation in strong lens modelling. Here we review and update the original problem and solution as implemented by, putting the regularisation method more into context.

3.1. Data model

The original Sparse Linear Inversion Technique (SLIT) algorithm introduced the following linear problem to solve ([J19](#)):

$$\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{g}_H + \mathbf{n}, \quad (6)$$

where \mathbf{y} , \mathbf{H} , \mathbf{F} ⁴, and \mathbf{s} retain their meanings from Eq. (2), ignoring other components. Here $\mathbf{g}_H \equiv \mathbf{H}\mathbf{g}$ is the galaxy lens light convolved with the instrumental PSF. Solving for \mathbf{g}_H instead of the deconvolved \mathbf{g} removes a potentially costly step in the modelling process and is anyway not needed for many scientific applications. We take \mathbf{n} to be white Gaussian noise, although it is possible to treat additive Poisson noise as well. For a fixed lens mass model, that is, known \mathbf{F} , the task is to jointly solve for the deconvolved and denoised source surface brightness in the source plane \mathbf{s} and the convolved lens surface brightness in the image plane \mathbf{g}_H . The algorithm can also optimise for \mathbf{s} alone when the lens light has been subtracted beforehand.

The original paper distinguished two versions of the algorithm: SLIT when solving only for \mathbf{s} and SLIT_MCA when solving for both \mathbf{s} and \mathbf{g}_H . The latter is based on the framework of morphological component analysis (MCA) developed for blind source separation ([Starck et al. 2005](#); [Bobin et al. 2008](#); [Joseph et al. 2016](#)). The technique seeks to disentangle superimposed signals based on their morphological properties when represented in a certain basis. This is relevant in strong lensing, as the deformed source galaxy light is often blended to some extent with the lens galaxy light. In SLIT_MCA, it is assumed that \mathbf{s} and \mathbf{g}_H can each be more sparsely represented in their respective planes by a particular wavelet transform, allowing the two components to be separated.

In this paper, we do not explicitly distinguish between the two algorithms. This is to emphasise that the new implementation is a single tool that adapts to the intended use. Where appropriate, we simply specify whether we solve for the source galaxy alone, for both the source and lens galaxies, or for the source galaxy and point sources (see Fig. 2 and Sect. 4.6 for details of point source modelling).

3.2. Sparsity and starlet regularisation

As summarised in Sect. 2.4, recent pixel-based reconstruction techniques make use of some variation of an adaptive grid for the source plane. In our work, we choose not to use an adaptive gridding scheme but instead a uniform source grid with resolution much higher than that of the data. This is for two main reasons: (1) our regularisation method is intrinsically multi-scale as we work in the wavelet space; (2) pixelated lensing operations and wavelet transforms are faster when applied on a uniform grid. We further allow the source plane grid to have the minimal size it can have given the data size, lens model, and (optionally) masking strategy.

⁴ We have dropped “ κ ” from the original notation \mathbf{F}_κ in [J19](#), since we do not build the lensing operator from the convergence map κ , but rather via deflection angles through ray-tracing.

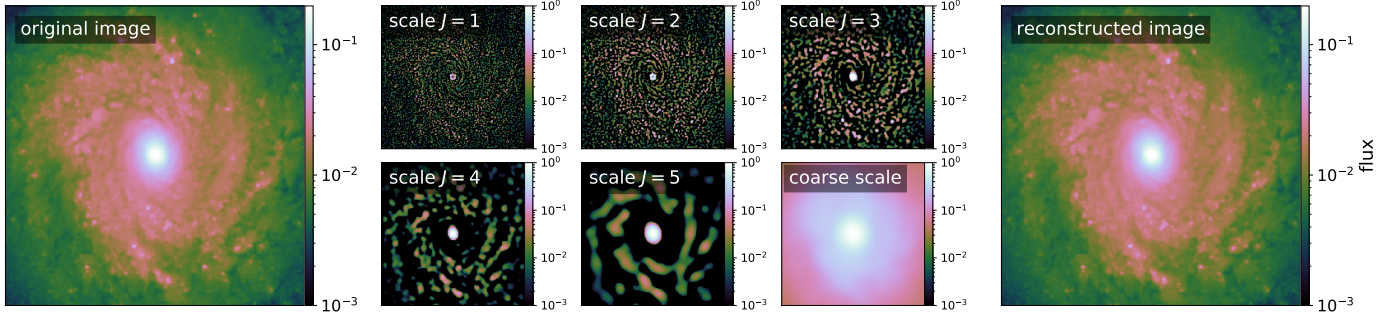


Fig. 1. Multiscale decomposition using the starlet transform. *Left panel:* original image, similar to source galaxies in Sects. 5 and 6. *Middle panels:* scales of the starlet decomposition using 6 spatial scales, and normalised ($\max = 1$) to enhance the contrast. In such a decomposition, a feature of a given scale is “brighter” in the corresponding image in the decomposition. Hence, keeping only highest pixels in each starlet scale filters out any insignificant signal over a large spectrum of spatial extents. All scale images have the same dimensions as the input image. *Right panel:* reconstructed image from all the decomposition scales, through co-addition of all decomposition scales. (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/visualize_starlet_transform.ipynb.)

Most regularisation strategies share the common difficulty of deciding how to set the strength of regularisation. The strength is controlled by a Lagrange parameter λ within \mathcal{P} that balances its importance relative to the data fidelity term \mathcal{L} (cf. Eqs. (5) and (7)). A further complication is that λ changes the effective number of degrees of freedom in a way that is difficult to quantify. This issue has motivated numerous uses of Bayesian arguments to objectively set λ (Suyu et al. 2006; Brewer & Lewis 2006; Dye et al. 2008; Vegetti & Koopmans 2009). Except for simple forms of regularisation, however, such as quadratic expressions, computing the Bayesian evidence often incurs significant computational overhead, as it requires integrals over a large parameter space volume.

We rely instead on sparsity in the wavelet domain to meaningfully set λ based on noise properties of the data. Sparse image reconstruction is a well-studied framework that has been used in a wide variety of astrophysical applications, such as the processing of weak-lensing (e.g. Lanusse et al. 2016; Peel et al. 2017) and radio interferometric data (e.g. Pratley et al. 2018), blind source separation of optical and radio sources (e.g. Joseph et al. 2016; Jiang et al. 2017), and recently in combination with deep learning (e.g. Sureau et al. 2020).

The success of wavelets in such applications stems from their multi-resolution property via the wavelet transform. Similarly to the way a signal is decomposed into its component frequencies by a Fourier transform, a wavelet transform decomposes a signal into both its frequency and spatial components. Figure 1 illustrates the multi-resolution decomposition of an image of a galaxy. A wavelet transform is defined by its basic functional wavelet form, which must obey certain mathematical properties. The particular choice of wavelet depends on the problem at hand. In this work we use the starlet transform (Starck et al. 2007), which is isotropic and undecimated, and has been shown to be suitable for treating astronomical images.

The starlet transform of an image returns J new images of equal size, each corresponding to a convolution of the original by a filtering kernel of a different scale. The j th kernel $j \in \{1, \dots, J-1\}$ amplifies features of the image on scales of 2^j pixels (see decomposition scales panels in Fig. 1) while preserving spatial locations. The final image in the decomposition corresponds to coarse-grained smoothing similar to a Gaussian filtering (see coarse scale panel in Fig. 1). One sets the number of scales in the transform by hand, the maximum value being limited by the number of pixels n_{pix} on the side of the image: $J_{\text{max}} = \lfloor \log_2 n_{\text{pix}} \rfloor$. We note that the number of decomposition

scales might also be optimised further for a specific system. We may explore this possibility in future work. Following standard notation (and that of J19), we write the linear starlet transform operator as Φ^\top and its inverse as Φ . The starlet transform of an image \mathbf{x} yields coefficients $\alpha_\chi = \Phi^\top \mathbf{x}$, and \mathbf{x} is recovered with $\mathbf{x} = \Phi \alpha_\chi$.

We regularise the reconstruction of both lens and source galaxies by applying a sparsity constraint on their starlet coefficients. This is motivated by the observation that astronomical images, including the light and mass distributions of galaxies, can be expressed with high fidelity using only a small number of coefficients in starlet space (e.g. Lanusse et al. 2016; Joseph et al. 2016; Peel et al. 2017). This property therefore defines a suitable regularisation for the underconstrained problem described by Eq. (5) in a model-independent way. The sparsity of a signal is enforced by minimising its ℓ_0 or ℓ_1 norm, in our case in the starlet domain. The former is generally preferred but difficult to use in practice, because it is not convex. SLIT uses the ℓ_1 norm, as it still promotes sparsity and being convex, guarantees convergence. We can now write the final expression of our priors as

$$\mathcal{P}(\mathbf{x}) = \lambda \left\| W_\chi \odot \Phi^\top \mathbf{x} \right\| + \iota_{\geq 0}(\mathbf{x}), \quad (7)$$

where \mathbf{x} stands for either \mathbf{s} or \mathbf{g}_H . We have introduced weights W_χ that adjust the starlet coefficients via an element-wise product \odot . This is necessary because the ℓ_1 norm (unlike the ℓ_0 norm) is known to bias the amplitude of the reconstructed signal, but this can be mitigated by an appropriate reweighting (Candes et al. 2007). We give more details in Appendix C on the implementation of the ℓ_1 -reweighting step. To avoid unphysical solutions, such as those containing negative pixels, we include as well the positivity constraint $\iota_{\geq 0}(\cdot)$, which goes to $+\infty$ if any coefficient of its argument is negative, set to zero otherwise, and is self-consistently folded in during optimisation.

Solving the convex minimisation problem of Eq. (5) subject to the constraints given by Eq. (7) is a difficult task because the ℓ_0 and ℓ_1 norm are not differentiable everywhere. A large

⁵ Formally one should write $\Phi_{(\mathbf{s}, \mathbf{g}_H)}^\top$ and $\Phi_{(\mathbf{s}, \mathbf{g}_H)}$, where the subscript in $\Phi_{(\mathbf{s}, \mathbf{g}_H)}^\top$ indicates if the operator is meant to be applied on the source or lens galaxy, with the corresponding number of scales $J_{(\mathbf{s}, \mathbf{g}_H)}$. To avoid cluttered expressions, we drop this subscript, and depending on which component the operator is applied on (\mathbf{s} or \mathbf{g}_H), the operator is the starlet operator whose number of decomposition is scales J_s or J_{g_H} , respectively. That is, $\Phi^\top \mathbf{s} \equiv \Phi_s^\top \mathbf{s}$ and $\Phi^\top \mathbf{g}_H \equiv \Phi_{g_H}^\top \mathbf{g}_H$.

family of methods to solve such minimisation problems rely on splitting the different constraints and applying them through more tractable mathematical operators. These are called proximal operators associated with each constraint and which can offer convergence guarantees depending on the algorithm and number of constraints. The proximal operator of the ℓ_1 norm is the soft-thresholding operation. The proximal operator of the $t_{\geq 0}(\cdot)$ corresponds to setting to zero all negative values. These operators are often applied after the step corresponding to a gradient descent update, which is the case in our algorithm.

As described above, the starlet transform gives a multi-resolution decomposition of the source being modelled. It encodes the spatial locations of features across many scales simultaneously. This is another advantage compared to classical regularisation strategies, like those introduced in Sect. 2.4. On one hand, classical strategies are “global” in that every image pixel contributes equally to the regularisation (through summation over the pixels), whereas starlet regularisation retains spatial locations. On the other hand, while MEM and the ℓ_2 norm are single-pixel measures, and gradient and curvature correlate flux over a few pixels, starlet regularisation incorporates correlations over a much wider range of scales, resulting in a higher fidelity reconstruction (see Fig. 1).

3.3. Regularisation strength

The value of the Lagrange parameter λ generically regulates the importance of regularisation compared to accuracy in reproducing the observation. A higher λ favours a smoother reconstruction at the cost of a worse fit to the data, whereas a lower λ gives more flexibility in matching the data at the pixel level but increases overfitting. In the context of sparse regularisation through the ℓ_0 or ℓ_1 norm, λ also has a clear interpretation: it effectively sets the statistical significance of the reconstruction in units of the noise, provided that the noise covariance is known and properly propagated to the starlet domain. We denote the propagated noise as σ_Φ (see J19, for its computation), hence $\lambda \equiv \lambda' \sigma_\Phi$ for a positive scalar λ' . For instance, $\lambda' = 3$ leads to a reconstruction at “ 3σ significance”. This understanding limits the range of interest for λ' values to typically ≥ 3 . In the algorithm, we start with a large λ' – automatically estimated from σ_Φ – and reduce it at each iteration until it reaches the target value. It is then held fixed for a small number of remaining iterations (typically 5).

We set $\lambda' = 3$ throughout this paper, as we have found it generally gives the smallest image plane residuals in our experiments, with no sign of overfitting. We also allow for a larger threshold applied specifically to the smallest starlet decomposition scale, to prevent spurious isolated pixels from entering the solution. As detailed in Sect. 3.2, the first starlet scale contains features with a spatial extent of $(2^j)_{j=0} = 1$ source pixel, which are likely to be noise. As a consequence, we set $\lambda' = 5$ for the smallest starlet scale only. We note that alternative solutions may exist, such as replacing the first starlet scale by a more suitable wavelet transform chosen such that isolated pixels are more penalised by the sparsity constraint (e.g. Lanusse et al. 2016, using a hybrid Battle-Lemarié+starlet transform).

A careful handling of the noise in the context of sparse modelling thus gives a simple way to choose the regularisation strength according to the desired reconstruction significance. A systematic comparison of the Bayesian evidence (Suyu et al. 2006; Dye et al. 2008; Vegetti & Koopmans 2009) is not necessary in our framework, which saves considerable computation time when solving the full SLIT problem (Eq. (8)). We refer to

J19 for further details concerning the computation of noise levels and their propagation through SLIT operations.

4. From SLIT to SLITRONOMY

The original SLIT algorithm presented in J19 is available as an open-source python code⁶, packaged with example scripts for illustrating its different features. There are two main difficulties with the code that we have sought to improve in this work. The first is that the computation time required for source reconstructions, and consequently joint source-lens reconstructions, is too large for tractable applications to real data. The second is that it is not practical to initialise the algorithm with results from the literature (e.g. lens model parameters), due to differences in coordinate systems and conventions. This is true for using the algorithm as a standalone tool, as well as in a workflow where model parameters first get optimised through analytical methods, then further refined using the SLIT algorithm.

We address these issues by introducing SLITRONOMY⁷, our revamped implementation of the SLIT algorithm that is significantly more efficient than its predecessor. It can be used as a plug-in to the open-source gravitational lens modelling software LENSTRONOMY⁸ (Birrer & Amara 2018). The framework of LENSTRONOMY allows us to access a number of practical and tested features that we exploit in our new development. The support of our code within LENSTRONOMY allows our sparse lens inversion method to be easily used by people already familiar with this software.

In the rest of this section, we describe a subset of key improvements and new features that we have brought to the original SLIT package. We note that not all features described in the following subsections are required for the results presented in Sects. 5 and 6.

4.1. Analysis and synthesis formulations

One can rewrite the general minimisation problem of Eq. (5) with the SLIT model from Eq. (6), and the regularisation strategy defined by Eq. (7). At this stage, there are two ways to solve the problem, depending on which variables one chooses to optimise. The ‘synthesis’ formulation solves the problem in the transformed domain, meaning that the actual variables are starlet coefficients of the source α_s and lens galaxies α_g . The output source and lens images, \mathbf{s} and \mathbf{g}_H , are then obtained by applying a final inverse starlet transform. In this case, one tries to find the sparsest solution as measured in starlet space. On the other hand, the ‘analysis’ formulation of the problem solves for \mathbf{s} and \mathbf{g}_H in direct space. In this case, one seeks a solution in the data space that simultaneously has a sparse representation in starlet space⁹.

In the original SLIT implementation, the model of Eq. (6) is optimised using the synthesis formulation. The main advantage of the synthesis formulation relies in the fact that the proximal operator of $\|\alpha_x\|$, where α_x are starlet coefficients of an image x , is exactly the soft-thresholding operator. This enables the use of simple and efficient algorithm such as the forward-backward algorithm (Combettes & Wajs 2005), without any approximation. In our implementation, we chose to solve the problem in

⁶ <https://github.com/herjy/SLIT>

⁷ <https://github.com/aymgal/SLITronomy>

⁸ <https://github.com/sibirrer/lenstronomy>

⁹ We note that the analysis and synthesis formulations are strictly equivalent only if $\Phi^T \Phi$ is the identity. As the starlet functions form an overcomplete basis, this condition is not met for the starlet transform.

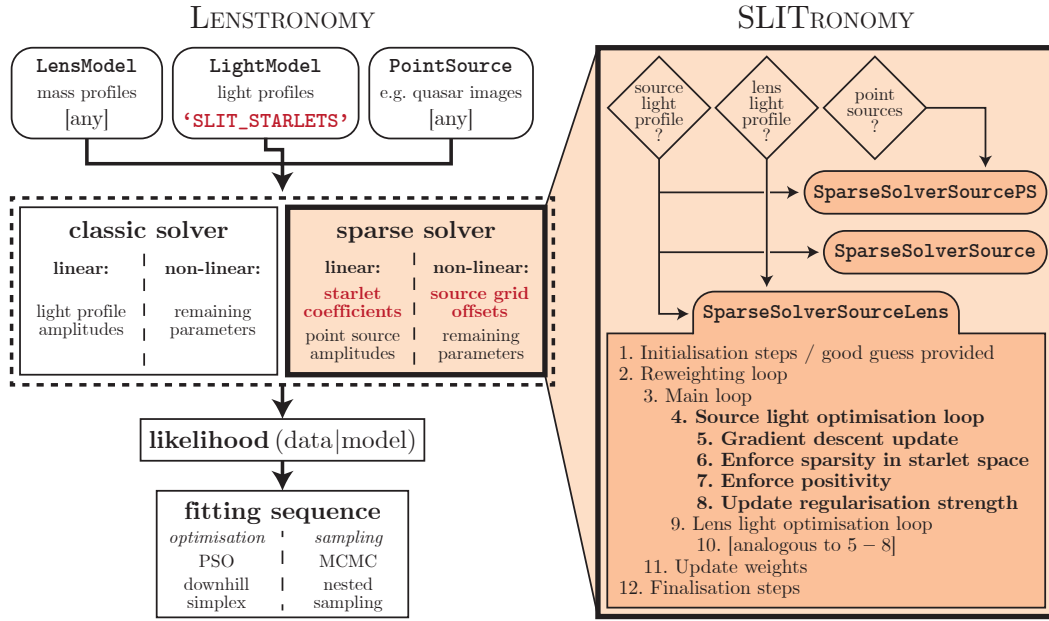


Fig. 2. Integration of SLITRONOMY into LENSTRONOMY. The global architecture of LENSTRONOMY is left unchanged. When the ‘SLIT_STARLETS’ profile is used (for modelling source or lens surface brightness), the image likelihood is computed through the sparse solver of SLITRONOMY, instead of the usual solver. Changes in LENSTRONOMY are indicated in red: a new light profile ‘SLIT_STARLETS’ has been introduced; linear amplitudes (originally defined for analytical profiles) are replaced by a sparse optimisation of individual coefficients in starlet space; source grid coordinate offsets are added as non-linear parameters, jointly optimised with, for example, lens model parameters. The integration of point source modelling is also illustrated. See Sect. 4 for details.

the analysis formulation. The problem of Eq. (5) is then written explicitly as:

$$\arg \min_{\mathbf{s}, \mathbf{g}_H} \frac{1}{2} \|\tilde{\mathbf{y}} - \mathbf{H}\mathbf{F}\mathbf{s} - \mathbf{g}_H\|^2 + \iota_{\geq 0}(\mathbf{s}) + \iota_{\geq 0}(\mathbf{g}_H) + \lambda \|\mathbf{W}_s \odot \Phi^T \mathbf{s}\| + \lambda \|\mathbf{W}_g \odot \Phi^T \mathbf{g}_H\|, \quad (8)$$

where \odot stands for the element-wise product, and we recall $\lambda \equiv \lambda' \sigma_\Phi$.

The difference with the synthesis formulation is that the variables are now \mathbf{s} and \mathbf{g}_H , that is, galaxy images in direct space. First, the number of effective free parameters (before regularisation is applied) is largely reduced when solving in direct space: it is simply the number of pixels N^2 as opposed to $J \times N^2$ for J decomposition scales. In this regard, optimisation in the analysis formulation is simpler and more stable. Second, the positivity constraint is simple to apply in direct space, because it does not require sub-iterations like in transformed space. Third, some specific features, such as those similar to a dirac function, are better represented in direct space, whereas they may require many more coefficients in transformed space. One drawback of the analysis formulation is that the sparsity constraint is not mathematically described by the exact soft-thresholding operation, compared to the synthesis formulation. In practice, however, the process of sequentially applying the starlet transform, the soft-thresholding operation, and then the inverse starlet transform does not prevent the algorithm from converging. There exist algorithms capable of solving Eq. (8) without this approximation (e.g. Cong & Vu 2011), but their implementation is left for a future version. As an additional improvement to the overall computation time, we use an optimised implementation of the starlet transform through the python package pySAP¹⁰ (Farrens et al. 2020).

4.2. Lensing operator

In SLITRONOMY, the implementation of the lensing operator (\mathbf{F} in Eq. (6)) has been improved for efficiency. When represented as a matrix, this operator is known to be extremely sparse, typically containing fewer than 1% of non-zero entries. In the original implementation, only non-zero entries were saved in memory to carry out lensing operations. In our new implementation, we use a faster and more memory efficient technique to build the operator based on its matrix representation and provided by the `scipy.sparse`¹¹ python package. In addition to storage optimisation, the matrix representation allows us to carry out the $\mathbf{F}\mathbf{s}$ operation and its inverse as plain matrix-vector products, which is considerably faster. We give quantitative measures of improved computation times in Table 1 corresponding to a configuration of 100×100 data pixels with twice the resolution in the source plane on a recent personal computer. This setup is typical, for example, of imaging data in the WCF3/F160W band of HST. To give an idea of the total computation time needed to solve problem (8), we also compare solving for \mathbf{s} and $\mathbf{s} + \mathbf{g}_H$ for the same configuration.

These speed improvements become crucial as soon as one wants to solve the more general problem of optimising the lens model in addition to the source reconstruction for fixed lens model (see Sect. 5.3). It is also highly beneficial for the more difficult problem of deblending and modelling both the source and lens galaxies. More detailed benchmarks are shown in Appendix F for various settings of the algorithm and data sizes.

4.3. Pixelated source surface brightness

The lens equation (Eq. (1)) gives the mapping between image plane coordinates and source plane coordinates, provided the

¹⁰ <https://github.com/CEA-COSMIC/pysap>

¹¹ <https://www.scipy.org>

Table 1. Speed up improvements of the new SLITRONOMY implementation, compared to the original SLIT code.

Operation	SLIT	SLITRONOMY	Speedup
Computation of the lensing operator, F	1.7 s	6 ms	300
Lensing of an image, Fx	200 ms	450 μ s	4500
Delensing of an image, $F^T y$	400 ms	120 μ s	3500
Solve for source light, s	~16 min	~1 s	1000
Solve for source and lens light, $s + g_H$	~2 h	~30 s	240

Notes. These numbers are valid for a typical configuration of 100×100 data pixels, with $r_{\text{pix}} = 2$ for source resolution, and $J_s = 6$ starlet scales. It can be for instance the modelling of a strong lens system imaged by HST in $F160W$ band.

deflection angles are known. When considering the surface brightness of galaxies as an ensemble of pixels, the lens equation is discretised, and one defines grids of pixels for the image plane and source planes. These grids need not necessarily be rectangular, but we only consider regular Cartesian grids in this work. When ray-tracing image plane pixels to the source plane through Eq. (1), the resulting coordinate is not necessarily centred on a source plane pixel. A simple way to deal with this while keeping a Cartesian coordinates for both planes is to interpolate over neighbouring source pixels to compute the values of the ray-traced image pixels. In our formalism, we consider this interpolation step as included in the lensing operator F .

In the original SLIT, a nearest-neighbour interpolation scheme was chosen for its simplicity of implementation, relying entirely on the regularisation (Eq. (8)) to fill unconstrained pixels. With LENSTRONOMY’s new implementation of the lensing operator as described above, more complex interpolation schemes now do not significantly increase the time required to build the matrix F . We have thus implemented a bilinear interpolation over source plane pixels similar to the one described in Treu & Koopmans (2004). In order to demonstrate visually the advantage of bilinear interpolation, we show in Fig. 3 the difference between the two types of interpolation for two different source plane resolutions. The bilinear interpolation leads to a much smoother light distribution that is closer to the true profile computed from analytical ray-tracing.

Another interesting observation is worth noting in Fig. 3. When the source plane resolution is significantly higher than the data resolution (twice, in this example), one can see “holes” in the source light after applying the de-lensing operator. This is typical of the linear approach to strong lensing on regular grids: not all source plane pixels get mapped to by an image plane pixel. This further illustrates the need for regularisation to solve the linear problem of Eq. (5), since it is intrinsically underconstrained. A regularisation based on wavelets is expected to be very efficient at filling these holes in the source plane, because wavelets provide a multi-resolution representation of the light distribution.

4.4. Lens model optimisation

The fact that SLITRONOMY is embedded in the LENSTRONOMY framework makes it convenient to use optimisation and sampling tools for lens model optimisation, in addition to the pixel-based reconstructions. For example, it is now simple to run a non-linear

optimisation or Markov chain Monte Carlo (MCMC) sampling over lens model parameters with a call to the SLITRONOMY solver for each proposed sample. As mentioned in Sect. 4.2, the runtime of a SLITRONOMY reconstruction with a fixed mass model for a typical HST cutout image is ~1 s when solving only for the source, and ~30 s when solving for the source and lens light together. While solving for the source at each iteration of an MCMC process is still tractable (see Sect. 5.3) solving for both components would be difficult given current computation speeds. Consequently, we see several possibilities to deal with this difficulty. First, by implementing further speed improvements of the $s + g_H$ solver. Deep learning techniques would likely accelerate the algorithm by replacing steps considered as bottlenecks (see e.g. Meinhardt et al. 2017; Sureau et al. 2020). Second, by using other non-linear solutions to explore the lens model parameter space. An interesting direction to explore is free-likelihood techniques with extreme data compression (see e.g. Alsing et al. 2019). And third, by pre-optimising the lens model using faster modelling techniques (analytical methods), so a smaller number of refinement steps are needed using starlet regularisation.

An extensive study of lens model parameter optimisation is dedicated to a future paper. We see the present work as one of the key steps needed to develop a full model for both the light and mass components.

4.5. Offset to source grid alignment

A difficulty shared by any pixel-based source reconstruction method is related to the definition of the source-plane coordinate grid. In particular, the alignment between image plane coordinates and source plane coordinates can be arbitrary and potentially lead to biases in lens model parameters depending on the choices made to address the problem. This is sometimes referred to as the “discretisation bias”, and several methods have been suggested to mitigate its effect, such as randomising the source plane initialisation (Nightingale & Dye 2015) or as an additional error term in image plane¹². Mitigation strategies typically depend on the algorithm-specific implementation and its assumptions. Analytical methods are, in essence, not subject to the discretisation bias.

Our method does not involve an adaptive gridding strategy for source plane coordinates. Hence the remaining degrees of freedom left for altering its alignment with image plane coordinates are constant offsets along both main axes. We introduce two additional non-linear parameters, $\delta_{s,x}$ and $\delta_{s,y}$, that characterise deviations from perfect alignment between the image-plane and source-plane coordinate grids. These two parameters are included in the set of non-linear parameters defined by other model components. Best-fit values and posterior distributions are therefore consistently estimated during optimisation or sampling.

4.6. Point source modelling

In the context of lensed point sources, the multiple images of background quasars are often seen as just the imprint of the PSF superimposed on the light of their host galaxy. There are two ways to model these features. One way is to increase the image plane resolution to the point that it is possible to

¹² We also performed source reconstructions with the “regridding error” term of Suyu et al. (2009), but found it leads to prominent features in the residuals due to overfitting.

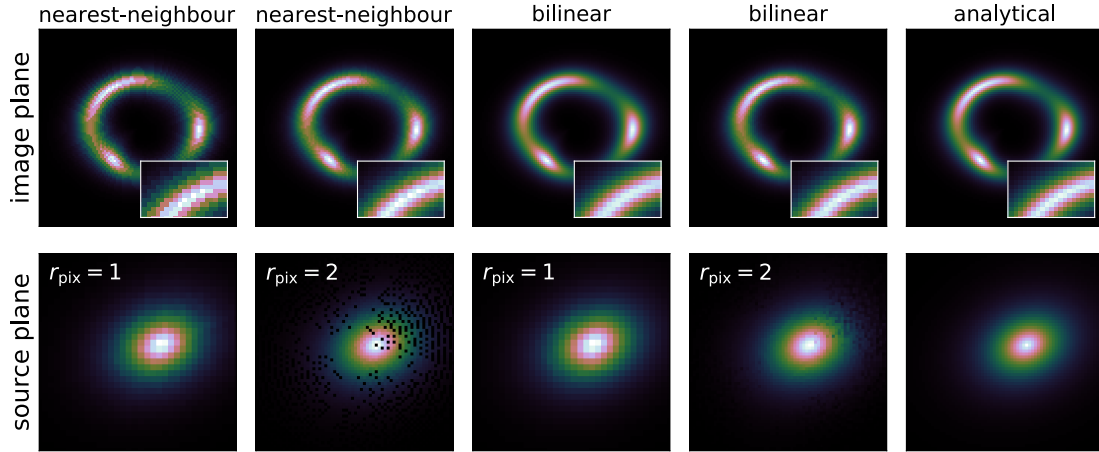


Fig. 3. Illustration of a pixelated mapping between source and image planes, as implemented in SLITRONOMY. *Top row, left to right:* nearest-neighbour interpolation of source surface brightness for two grid resolutions (same as image plane or doubled), bilinear interpolation of surface brightness for the same resolutions, and parametric “groundtruth” using LENSTRONOMY. *Bottom row, left to right:* source plane corresponding to the top row panels, for different data pixel size to source pixel ratio r_{pix} . Dark isolated pixels in source plane are not mapped to any image plane pixel, hence not constrained by imaging data. Our reconstruction technique is able to fill these “missing pixels”, through sparse regularisation and multiresolution property of wavelets (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/visualize_lensing_operator.ipynb).

resolve each quasar image, but this leads to extremely large numbers of source-plane pixels and quickly becomes intractable. Another simpler, faster and more effective method is to shift and scale – that is (de)magnify – the pixelated PSF at the position of each image. This is the solution implemented in LENSTRONOMY¹³, and other methods use the same procedure as well (Suyu & Halkola 2010; Auger et al. 2011). Our model of Eq. (6) can be straightforwardly extended to support point-source modelling:

$$y_{\text{ps}} = HFs + g_H + \sum_i A_i K + n, \quad (9)$$

where $A_i K$ represents the interpolated and amplitude-scaled PSF kernel at the location of the i th image. An iterative approach can be employed to solve for the three unknown components s , g_H and $\{K_i\}$, removing from the data two out of the three components to estimate the third one in sub-iterations, until convergence.

5. Current data with HST

In this section, we focus on imaging data corresponding to typical images obtained with the *Hubble* Space Telescope (HST). We first simulate realistic HST data to perform a detailed comparison between our pixel-based reconstruction technique and state-of-the-art analytical reconstructions. We also propose an automated process for refining the source model using both methods. We then reconstruct the source galaxy in a subset of Sloan Lens Advanced Camera for Surveys (SLACS) strong lens samples.

5.1. Source reconstruction on simulated data

5.1.1. Simulation setup

We use as sources the preprocessed high-resolution galaxies that were employed to generate the simulations of Time Delay Lens Modelling Challenge (TDLMC, see Ding et al. 2018, 2020). These sources consist of high-resolution cropped images of

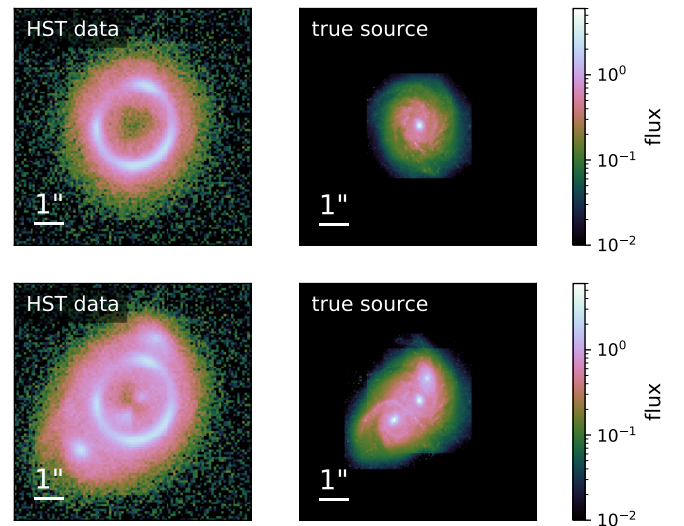


Fig. 4. Simulated HST data. The lens light is assumed to be perfectly subtracted. The resolution and noise covariance mimic those of real observations with the WCF3/F160W instrument. *First row:* simulation using a single galaxy. *Second row:* more complex configuration with a composite of three close-by galaxies in the source plane. Source objects are projected on a finer grid, to match the resolution of our reconstruction technique (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb).

nearby spiral galaxies from the *Hubble* Legacy Archive¹⁴, on which isolated stars and objects in the field have been removed and the background has been subtracted. Doing so ensures that a negligible amount of noise is propagated to the image plane, the presence of which can potentially lead to biases in the source reconstruction and recovered lens model parameters. We create two types of source: either a single galaxy or two close-by galaxies simulating a merging pair, that we place at redshift $z_s = 1.2$. Source galaxy images are shown in the right column of Fig. 4.

¹³ Given an accurate model for the source surface brightness, the pixelated PSF can further be refined iteratively.

¹⁴ <https://hla.stsci.edu/hlaview.html>

Once the source galaxy has been selected, we use LENTRONOMY and a built-in light profile that performs spline interpolation of the (high-resolution) source image. We then simulate the lensing effect through ray-tracing on the interpolated image, based on an analytical model for the lens galaxy mass. The main deflector is described by a singular isothermal ellipsoid (SIE) placed at redshift $z_d = 0.3$ with a central velocity dispersion of $\sigma_v = 260 \text{ km s}^{-1}$ and axis ratio $q_m = 0.8$. In addition, it is embedded in an external shear with moderate strength $\gamma_{\text{ext}} = 0.03$, misaligned by $\pi/4$ rad upwards, counter-clockwise, with respect to the lens galaxy, to simulate the influence of nearby perturbers. After ray-tracing to compute extended lens arcs, we convolve the image with the HST PSF kernel for the near-infrared *F160W* band, simulated with the TinyTim software (Krist et al. 2011). We choose the *F160W* band, because it usually gives a higher signal-to-noise for the lensed arcs, as well as better contrast between source and lens galaxies (although we ignore the lens light in this case). Typical measurement noise is then added in the form of a background Gaussian component and a Poisson component based on instrumental properties and exposure time, combined per pixel i as

$$\sigma_i^2 = \sigma_{\text{bkg}}^2 + \sigma_{\text{Poisson}}^2 \quad (10)$$

We assume uncorrelated noise in the above equation, as this approximation is common to CCD data. In general, correlated noise can be incorporated as well, as long as the correlations are known or properly estimated (the propagation of noise to wavelet space is left unchanged). Other instrumental settings (e.g. zero-point magnitude, read noise, etc.) are based on real HST observations. Final simulated images are shown in Fig. 4. We summarise in Table A.1 the instrumental settings and in Table A.2 the values of the lens model parameters used in our simulations.

5.1.2. Starlet source reconstruction

We demonstrate the source reconstruction capabilities of our starlet regularisation on the simulated data described above. As detailed in Sect. 3, the regularisation strength is defined by the detection threshold we wish to obtain in the source plane, in units of the noise. We set $\lambda' = 3$ (i.e. 3σ significance for the source reconstruction) for all our models. We find that this permissive value leads to a good fit to the data for all applications in this work. We show in Appendix E that the number of scales has small impact on the source modelling at the level required for this work, but has substantial impact on the computation time. We fix the number of decomposition scales to $J_s = 6$, giving reasonable computation time. We set the ratio of data pixel size to source pixel size to $r_{\text{pix}} = 3$. As we focus on the source reconstruction, lens model parameters are held fixed to their input values.

We show in the first row of Figs. 5 and 6 results of our sparse reconstruction technique using starlets. We show the image model along with image plane normalised residuals, and the source model along with source plane residuals. The reduced chi-square values, χ_v^2 , are computed considering the number of (un-masked) pixels i as the number of effective degrees of freedom v :

$$\chi_v^2 = \frac{1}{v} \sum_i \frac{(f_{\text{model},i} - f_{\text{data},i})^2}{\sigma_i^2}, \quad (11)$$

where σ_i^2 is noise variance of pixel i (Eq. (10)).

In the source plane, we assess the quality of the reconstruction by computing the source distortion ratio (SDR, as defined

in J19):

$$\text{SDR} = 10 \log_{10} \left(\frac{\|\mathbf{s}_{\text{true}}\|}{\|\mathbf{s} - \mathbf{s}_{\text{true}}\|} \right) \quad (12)$$

where \mathbf{s}_{true} is the true light distribution of the source. With such a definition, the higher the SDR, the better the reconstruction.

Both image plane and source metrics are shown in Figs. 5 and 6. The imaging data is modelled almost to the noise level, as one can expect from the assumed knowledge of the lens model, noise covariance and PSF. Despite small artefacts incorrectly filtered out by the starlet regularisation (similar features were noticed in J19), the global shape of the reconstructed source galaxy is in good agreement with the truth. One can distinguish smaller scale features such as the most prominent spiral arms, even though the deconvolution of the source (jointly performed by the algorithm) is limited by the fairly large FWHM of the *F160W* PSF.

It is worth pointing out that although we use parts of the same software to generate and to model the data, the actual algorithms used in each process are importantly different. The lensed sources were simulated using analytical ray-tracing from a spline-interpolated image with much higher resolution than that of the final data. In contrast, the source modelling was performed by a different approach, both when using starlets (via the lensing operator and bilinear interpolation) and shapelets (via analytical evaluation of basis functions). We therefore avoid the logical inconsistency of attempting to model the simulated data using the same algorithm that produced it. This also means that we cannot expect a perfect reconstruction, even assuming the true lens model.

5.1.3. Comparison with analytical shapelets

We next perform the same exercise with the analytical, yet flexible, modelling method using shapelet basis functions (Bernstein & Jarvis 2002; Refregier 2003; Birrer et al. 2015). While shapelets can potentially produce biases for shape measurements (e.g. Melchior et al. 2010, in the context of weak lensing), they have been successfully employed in strong lens modelling. Shapelets are Gauss-Hermite polynomials that form a complete and orthonormal basis set. To ensure a tractable optimisation of shapelet coefficients, only a limited number n_{max} of polynomials are considered to model the light distribution. A shapelet basis set is thus defined by three free non-linear parameters aside the fixed n_{max} : the two coordinates of the basis centre and the reference scale β . The total number of basis functions (linear amplitudes) is $n_{\text{tot}} = (n_{\text{max}} + 1)(n_{\text{max}} + 2)/2$. The polynomial order and the reference scale set the minimal and maximal scales that can be resolved $\beta_{\text{min}} = \beta(1 + n_{\text{max}})^{-1/2}$ and $\beta_{\text{max}} = \beta(1 + n_{\text{max}})^{1/2}$, respectively.

There are two ways of modelling the source with shapelet basis functions. They can on one hand be employed as a standalone light profile (e.g. Birrer et al. 2015, 2017; Tagore & Jackson 2016). Alternatively, they can be considered as small-scale corrections to an underlying large-scale light profile, commonly taken to be an elliptical Sérsic profile whose centroid coincides with the shapelets centroid (Birrer et al. 2019; Shajib et al. 2019, 2020a). For completeness we compare our reconstruction technique with both approaches, which we term “shapelets” and “shapelets+Sérsic”, respectively. We set the maximal polynomial order n_{max} as the lowest value required to obtain image plane residuals down to the noise, based on visual inspection.

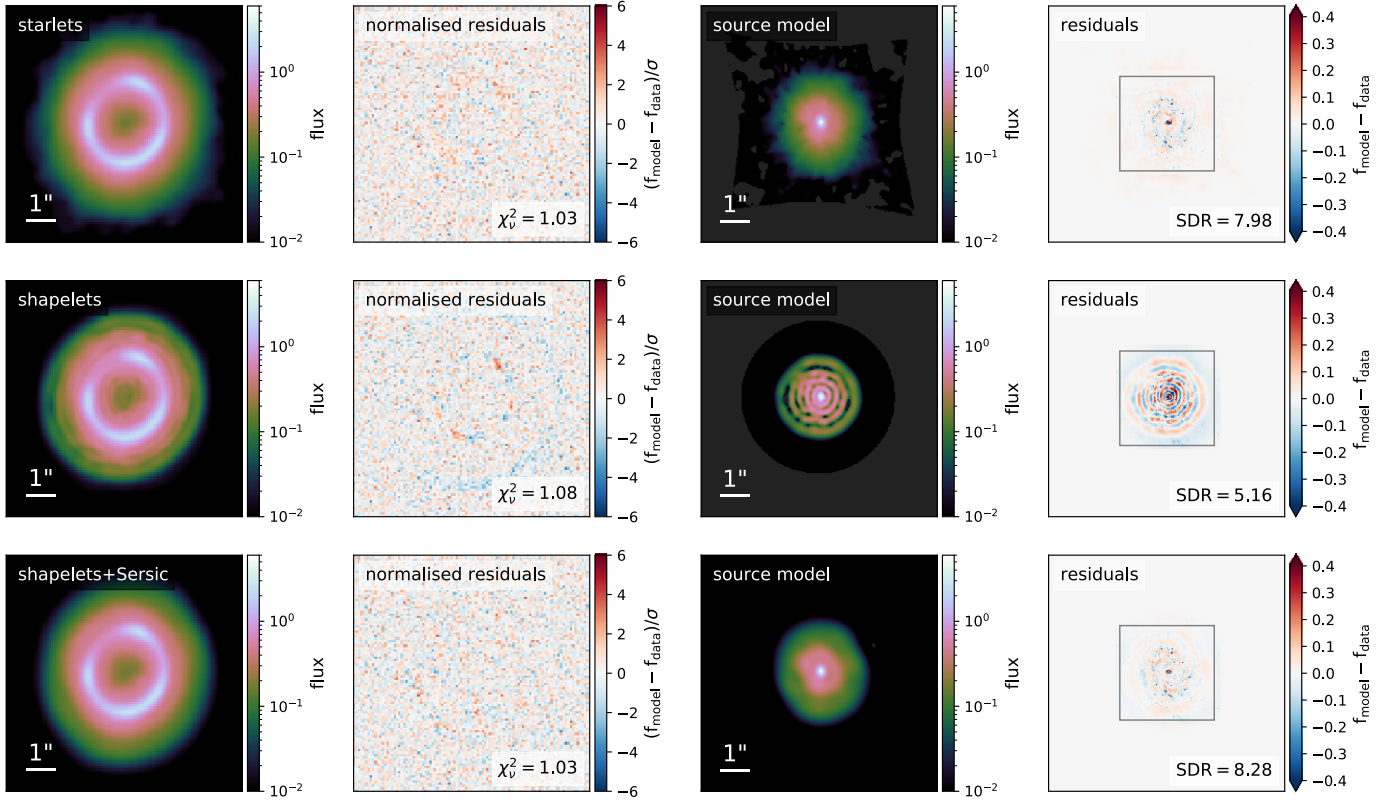


Fig. 5. Source modelling of the simulated data with a single source galaxy (Fig. 4). *Top to bottom*: sparse model with starlets ($r_{\text{pix}} = 3$), analytical model with shapelets ($n_{\text{max}} = 18$), analytical model with shapelets+Sérsic profile ($n_{\text{max}} = 8$). *Left to right*: image model, image normalised residuals and reduced chi-square, source model, source residuals and SDR. The SDR is computed only in the region indicated by the grey box, to avoid pixels with no flux to bias its value. All source reconstructions are shown at the supersampled resolution corresponding to the chosen r_{pix} . Dark grey areas in model panels correspond to non-positive pixel values (which can be negative with shapelets, https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb).

We optimise the non-linear parameters of the shapelet basis (scale and basis centre) using the Particle Swarm Optimiser (PSO, Kennedy & Eberhart 2001) implemented in LENSTRONOMY, which is particularly suited for finding global minima in large parameter spaces¹⁵.

We show in Fig. 5 both the “shapelets” and “shapelets+Sérsic” reconstructions, for direct comparison with the “starlets” reconstruction. One notices that both analytical and pixel-based methods display artefacts in source plane, although not at the same level. While starlets only introduce low significance false detections due to noise propagation from image plane, shapelets introduce prominent concentric rings that attempt to compensate small scale features in the source structure, such as the cuspy bright core of the spiral source.

We emphasise that starlets are able to represent the large dynamic range of source features on their own, while shapelets are not. By adding a Sérsic profile – or any other smooth profile – to the shapelet functions, one can split the modelling of small- and large-scale features. Indeed, the Sérsic profile acts as a low-pass filter, with shapelet functions capturing the remaining high frequencies and requiring a lower maximal order compared to shapelets alone. The global shape of the source is retained with shapelets+Sérsic, even though the boundaries of the light distribution are sharper than with starlets (the truth seems to lie in between). For small-

scale features, starlets show more fidelity to the true underlying surface brightness. Such features, when revealed with shapelets, need to be interpreted with care, since they may not be real, but rather artefacts from their limited flexibility.

While the starlet model provides slightly better normalised residuals and a smoother, more realistic source, the resulting χ^2_{ν} is very close to the one obtained with the shapelets+Sérsic model. To reach this value with the analytical model, additional constraints through the Sérsic profile are required, which are not accounted for in the χ^2_{ν} . We therefore believe our model performs better, but this is not well captured by the small difference in χ^2_{ν} . In contrast to just the likelihood value, a metric based on the posterior value – which has to be carefully designed so that it takes into account intrinsic differences between the methods – should be preferred, but this is beyond the scope of this work.

Similar observations can be made for the case of a source composed of multiple complex components. For the shapelets+Sérsic reconstruction, a single shapelet basis is not sufficient to obtain residuals down to the noise, requiring an n_{max} so large that it negatively affects the convergence efficiency of the optimisation and introduces significant artefacts in source plane (similar to the middle panel in Fig. 5). Only after carefully optimising multiple shapelets+Sérsic pairs at the location of each individual sub-component in source plane can the residuals be reduced to a satisfactory level. The initial location of these analytical profiles is crucial to prevent strong degeneracies from appearing during optimisation. If one is interested in modelling only a few strong lens systems, the initialisation can be

¹⁵ We also used the Nelder-Mead simplex algorithm (from the SCIPY library), but found that it quickly fails at exploring the parameter space once more complex light profiles such as shapelets+Sérsic or large shapelet basis are being optimised, given the constraints of HST images.

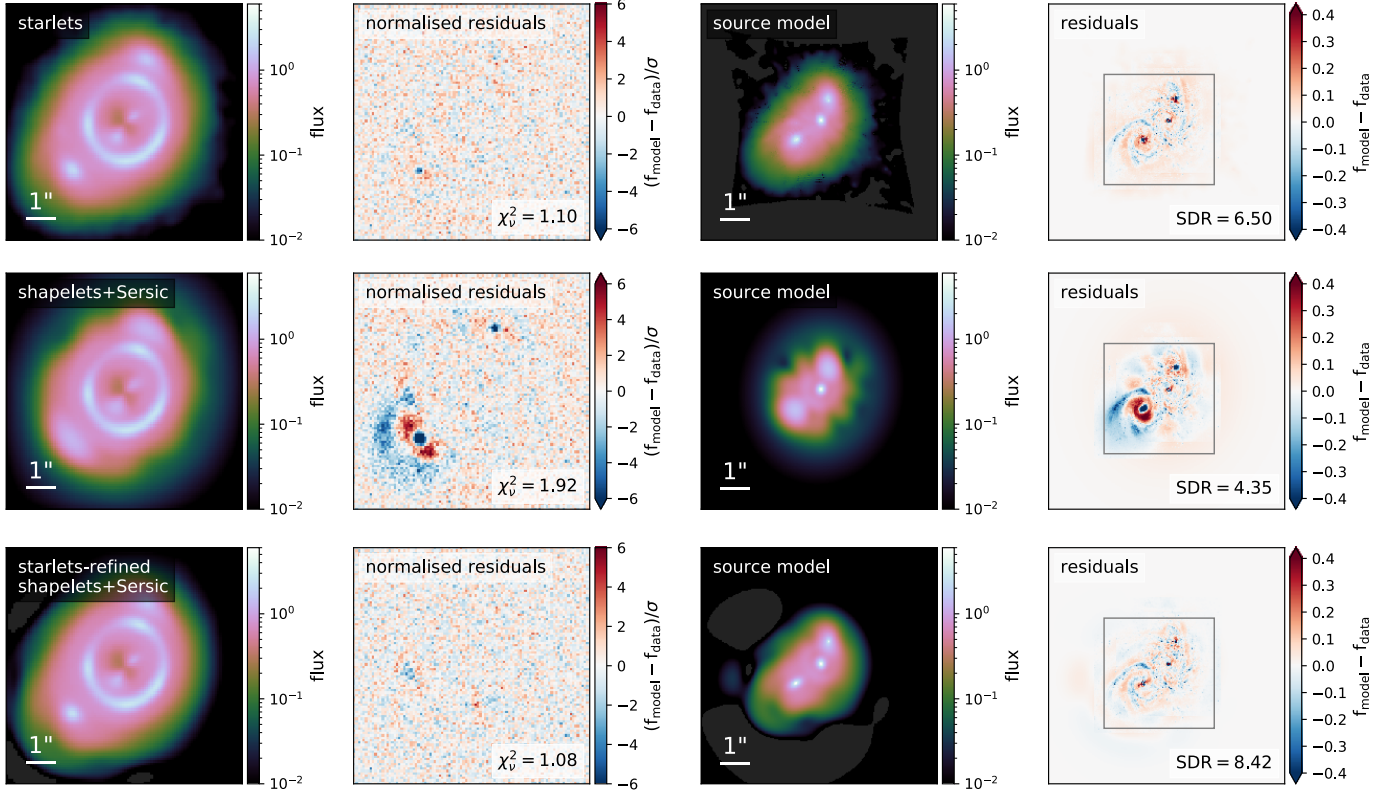


Fig. 6. Source modelling of the simulated lensed galaxy group (bottom panel of Fig. 4). *Top to bottom*: sparse model with starlets ($r_{\text{pix}} = 3$), analytical model shapelets+Sérsic ($n_{\text{max}} = 8$), and starlets-refined shapelets+Sérsic model ($n_{\text{max}} = \{3, 3, 9\}$). *Left to right*: image model, image normalised residuals and reduced chi-square, source model, source residuals and SDR. The SDR is computed only in the region indicated by the grey box. All source reconstructions are shown at the supersampled resolution corresponding to the chosen r_{pix} . Dark grey areas in model panels correspond to non-positive pixel values (which can be negative with shapelets). We emphasise that the iterative refinement process, uses the model in first row to automatically setup the low-complexity model in second row, and further refine it until the reconstruction shown in last row (see Fig. 7 and Appendix B), https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb.

set manually by the investigator, based on model residuals. For batch modelling of large datasets, however, this is not a viable solution.

To tackle this limitation, we design a workflow motivated by the advantages of both analytical and pixel-based reconstruction methods. While the former offers a considerable gain in speed – mostly when solving for a large number of non-linear parameters from lens and source models jointly – the latter does not require any choice of light profiles, nor a good set of initial parameter values for initialisation. Hence, analytical methods are better suited to rapidly explore the initial, potentially large, parameter space, and to provide a starting point for further refinement steps. These refinement steps can then be exclusively pixel-based or, as we propose here, use an intermediate solution between pixel-based and analytical. The goal of such a hybrid approach is to retain timing efficiency, while substantially reducing human choices by including pixel-based flexibility.

The proposed workflow can be summarised in four steps, illustrated in Fig. 7: based on a pre-optimised lens model obtained through a faster, fully-analytical modelling step, (1) the sparse starlet technique is used to obtain a high-resolution model of the source; (2) the number and location of each sub-component of the source are automatically detected on the starlet model; (3) individual pairs of shapelets+Sérsic are assigned to sub-components and their centres properly initialised with low polynomial order (typically $n_{\text{max}} = 2$); (4) the polynomial order of each shapelet basis is iteratively refined based on

image plane residuals that are specific to the individual sub-components. Finally, a full lens model optimisation can be performed, either with the updated analytical or pixel-based source model, depending on the specific goal and requirements (e.g. flexibility, computation time). We refer to Appendix B for more details. We note that although the initial lens model estimation needs to be good enough to discard or detect multiple source components, the pixel-based starlet reconstruction can be used to quickly assess of the reliability of the lens model (in addition to model residuals).

The result of this automated refinement is shown in the bottom row of Fig. 6. The final source model consists of three shapelets+Sérsic pairs with joint centroids and maximal polynomial orders of $\{3, 3, 9\}$. The higher order of the third shapelet basis is necessary to model the most prominent spiral arm. As expected, residuals are drastically improved by this automated procedure, reaching a similar level as the starlet reconstruction in image plane. In source plane, the SDR is slightly better than with the pixel-based approach, showing visually similar residuals when compared to the true source. The flexibility and simplicity of sparse modelling is directly used to improve the large set of analytical profiles, still driven by the imaging data itself. Some works have already attempted to minimise investigator time for batch modelling of strong lens systems (e.g. Nightingale et al. 2018; Shajib et al. 2019, 2020b). One can imagine the procedure developed here to be part of a larger automated modelling pipeline in the future.

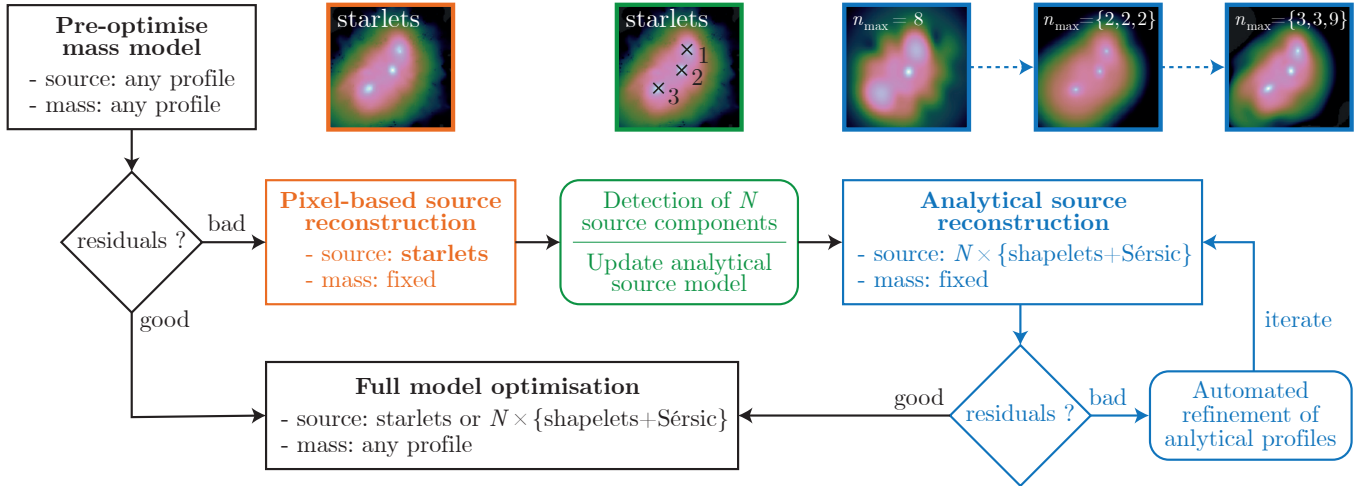


Fig. 7. Workflow for using the starlet model as an automatic refinement of the analytical model. From a first estimation of the lens model, model residuals are assessed through visual inspection and/or goodness-of-fit statistics. If the current source model is insufficient, the lens model is fixed and the source is reconstructed using sparse optimisation with starlets. Potential sub-components of the source are located in source plane through special filtering followed by local maxima detection. A pair of profiles shapelets+Sérsic is assigned to each sub-components, with low complexity (shapelet order $n_{\max} = 2$). An iterative process is engaged for refining individual shapelet orders based on residuals in pixels corresponding to their sub-component (see Appendix B for details). Once refined orders are sufficient, a full model optimisation can be performed.

5.2. Source reconstruction of SLACS lenses

We now apply SLITRONOMY to real data, focusing on a subset of lenses from the Sloan Lens Advanced Camera for Surveys sample (SLACS, Bolton et al. 2006). This sample of 80 galaxy-galaxy lenses has been well studied and used, in combination with spectroscopy, to infer statistical properties of lens galaxies such as their dark matter and baryonic mass distributions, their location in the fundamental plane, as well as their selection functions (e.g. Gavazzi et al. 2007; Auger et al. 2010). However, little focus has been given to the quality of the source reconstructions. It has also been shown that some properties of source galaxies can be recovered to sufficient precision without using pixel-level modelling (e.g. for deriving AGN mass-luminosity relation, Ding et al. 2021). However, for certain applications that require insights on small-scale features in source galaxies, such as the study of star-forming regions (e.g. James et al. 2018) or galaxy morphologies, more flexible methods are required to reconstruct complex sources.

In our analysis, we assume that the lens mass model has been optimised beforehand. We use the recent work of Shajib et al. (2020b), which uniformly modelled 23 of the SLACS lenses to put constraints on the mass profile and content of large elliptical galaxies. These models were also used as an external dataset in the hierarchical Bayesian analysis for time-delay cosmography introduced by Birrer et al. (2020). Here we briefly summarise the modelling choices and fitting procedure used to optimise the lens model parameters (implemented in the DOLPHIN package, a wrapper around LENSTRONOMY¹⁶), and refer the reader to Shajib et al. (2020b) for more details. The mass profile of the lens is a power-law ellipsoidal mass distribution (PEMD), embedded in an external shear. The lens light (although masked for some systems) is modelled with a double elliptical Sérsic profile. The source light is modelled with a shapelet function basis plus an elliptical Sérsic profile. For an individual SLACS system, the adopted fitting recipe consists in a series of particle swarm optimisation (PSO) runs, each time altering specific degrees of freedom in the model: (1) the PEMD slope and exter-

nal shear are fixed; (2) the flux from lensed arcs is masked for fitting of lens light; (3) all mass and light parameters for the deflector are fixed for fitting the source light, with fixed shapelet scale; (4) the PEMD parameters are relaxed for jointly optimising the source and mass profile parameters; (5) the shapelet scale is relaxed for further refinement of the source; (6) steps 2–5 are repeated with current best-fit values; (7) the PEMD slope and external shear are relaxed for a final MCMC sampling of all parameters together.

The modelling of Shajib et al. (2020b) has been performed with LENSTRONOMY, hence lens models are given in a format usable by our SLITRONOMY implementation with minimal pre-processing. Among the 23 lenses in the modelled sample, we first select one “simple” system, SDSS J1627–0053, that exhibits no particularly complex features (it is an almost perfect Einstein ring), and from which lens light can be removed properly. Three additional systems, intentionally selected for their higher complexity, are added to our subset of lenses, based on inspection of residuals from the analytical model and source reconstruction (shapelets+Sérsic). They are SDSS J1250+0523, SDSS J1630–4520 and SDSS J0959–0410. We use HST data obtained with the ACS instrument in the $F555W$ band for all four systems. We note that all of these systems have been considered to have good quality models by Shajib et al. (2020b), although the main goal of their work was to obtain reliable lens model posterior distributions. We then take their best-fit lens model parameters¹⁷ and apply our starlet reconstruction technique directly on the same imaging data cutouts, using the same PSF kernel as Shajib et al. (2020b) for the joint deconvolution. We use $r_{\text{pix}} = 3$ for source plane resolution.

Figure 8 shows source reconstructions of the four systems. From left to right are the lens-subtracted imaging data, the residuals and shapelets+Sérsic model of Shajib et al. (2020b), followed by the residuals and model obtained with starlets. Reduced chi-square metrics are also indicated. Similarly to the simulated data of Sect. 5.1, we notice that shapelet reconstructions for these systems display ring-like or wave-like features that do not likely represent the true source surface brightness.

¹⁶ <https://github.com/ajshajib/dolphin>

¹⁷ Private communication.

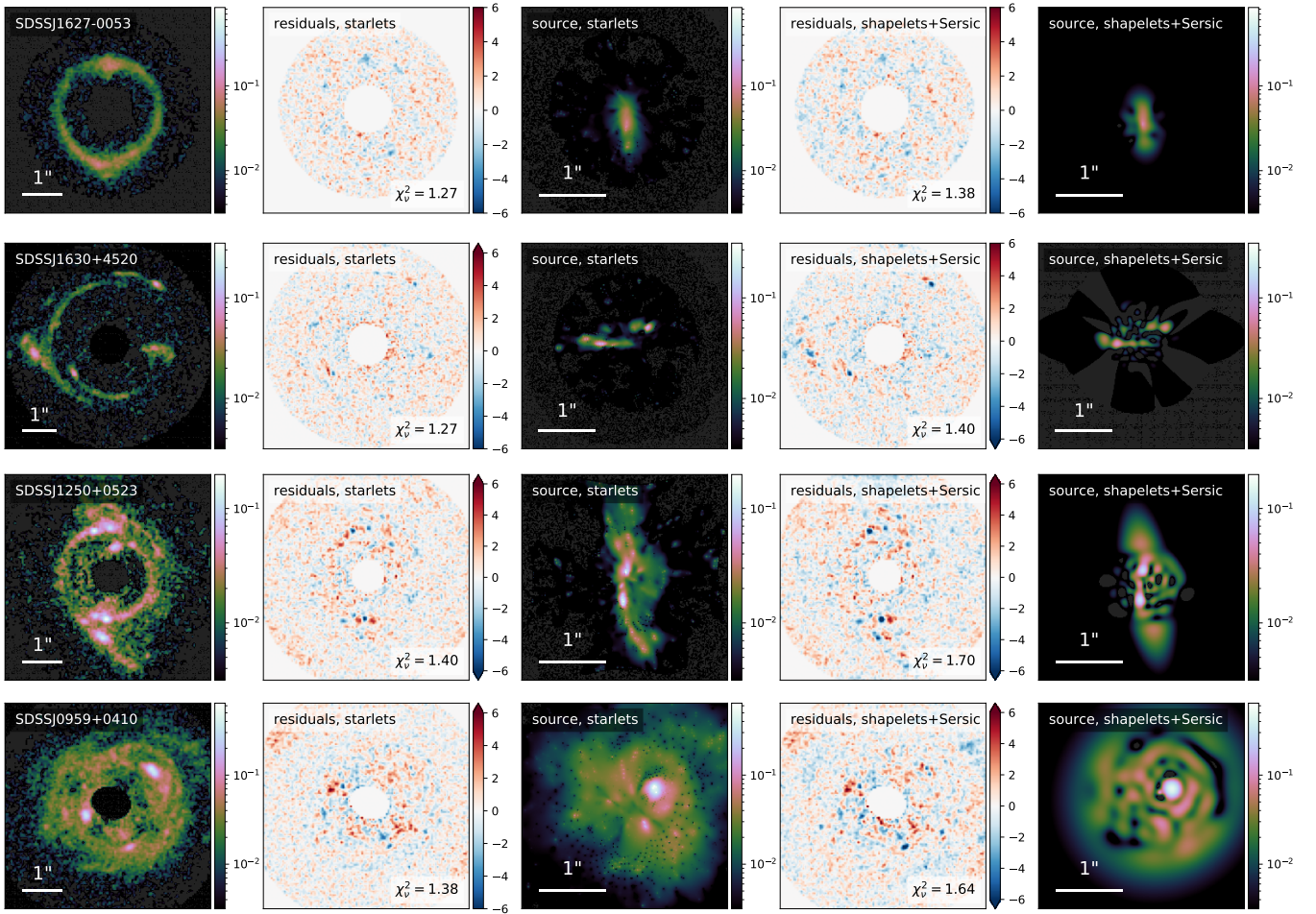


Fig. 8. Source reconstruction of SLACS lenses, assuming analytical mass and lens light models from Shajib et al. (2020b). *Left to right*: imaging data with lens-light subtracted, normalised residuals with shapelet model, shapelet source model, normalised residuals with starlet model, starlet source model ($r_{\text{pix}} = 3$). The maximal polynomial order n_{max} used for shapelet models are, *from top to bottom*: 6, 15, 12, 15. We refer to the text for more details, specifically concerning the case of SDSS J0959–0410. Dark grey areas in model panels correspond to non-positive pixel values (which can be negative with shapelets). All source reconstructions are shown at the supersampled resolution corresponding to the chosen r_{pix} . Colour maps for flux and residuals retain their meaning from Fig. 6 (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/SLACS_source_reconstruction.ipynb).

Additionally, the modelled light profile contains unphysical negative values, especially in the case of SDSS J1630–4520.

We observe a significant decrease in the residuals when using the sparse source reconstruction, compared to the original residuals obtained with the analytical model. While this may seem expected at first sight – as the former model is much more flexible – it should not be overlooked. Since we perform the pixel-based reconstruction at fixed mass in this case, it hints that lens model parameters that would be favoured by the sparse reconstruction are compatible with those obtained from analytical optimisation. If this were not the case, residuals would, on the contrary, have been worse than previously obtained, because the sparse reconstruction would have favoured another location in parameter space. Under the assumption that shapelet source reconstruction does not lead to biased lens model parameters (which is supported by previous works Birrer et al. 2015; Tagore & Jackson 2016), this is a reassuring result.

Specifically in the case of SDSS J0959–0410, we observe small-scale artefacts in the form of an array of low-amplitude spots that are clearly non-physical. They are a consequence of the pixel-based nature of our modelling technique, for which we identify three reasons. First, the source object – likely to be two or more nearby galaxies – extends over a region much

larger than the caustics where the lens mapping tends to suffer more from the pixelated approximation. Second, the analytical lens light model is particularly inaccurate due to prominent overlapping of the lens and lensed arcs, in addition to a likely dust absorption lane (visible before lens light subtraction) that cannot be modelled well with smooth profiles only. Third, the mask defined to exclude pixels affected by the imperfect lens light subtraction causes an artificial discontinuity between the lensed arcs and the masked region, which is propagated to the source plane at the location of the artefacts. These three effects lead to inaccuracies in the noise estimation and cause source pixels at these locations to not be properly regularised when applying the sparsity constraint. We investigate these points further in Appendix D.

Pixel-based methods tend to be subject to over-fitting, which can artificially improve residuals. However, we argue this is not the case here, as one can easily identify features in the residuals that still persist after the starlet reconstruction of the source. Overall, no overfitting features clearly stand out from the residual maps. Our method’s careful handling of noise levels, namely their propagation to starlet space and the subsequent application of sparsity constraints, all help to avoid overfitting.

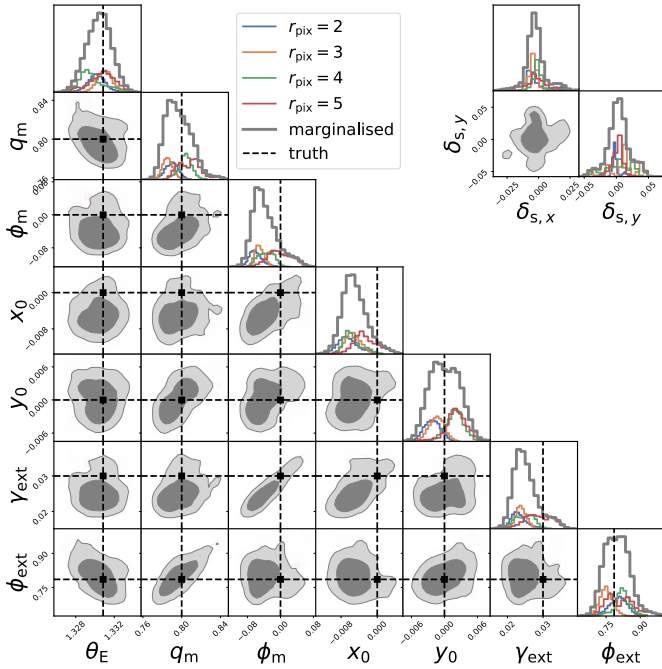


Fig. 9. Posterior distributions for lens model (SIE + external shear) and source grid offset parameters (see Sect. 4.5 for details), for the simulated HST data shown in Fig. 4 (first row). We show individual posteriors for different values of r_{pix} (thin colored contours), and the marginalised posteriors (thick grey contours) have been obtained by combination with equal weights. Two-dimensional contours show 68% and 95% of the distribution volume, and true values are indicated by dashed lines (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mass_sampling/corner_plot_mass_sampling.ipynb).

5.3. Optimisation of the analytical mass model

Although our primary focus is on the problem of source reconstruction, estimating deflector mass parameters is also of interest for many strong lensing applications. As an example of how this can be done in our framework, we use MCMC¹⁸ to sample the lens model parameter space of our simulated HST data in the single source galaxy case (top row of Fig. 4). The resulting posterior distributions for SIE and external shear parameters are shown in Fig. 9. We have checked the MCMC chains for convergence and verified that a sufficient number of samples has been discarded (we kept the last 35 000 samples each chain). By comparing the maximum a posteriori values to the true input values, we observe that the source pixel size (controlled through r_{pix}) acts as the main source of systematic errors. This is consistent with what has been observed from other pixelated source reconstruction methods (see e.g. Fig. 3 from Suyu et al. 2013).

In order to ensure robustness of the posterior and prevent underestimating error bars, we also marginalise over the different choices of r_{pix} . As there is no natural choice for this parameter, we apply equal weights to the individual distributions to obtain marginalised posteriors. The resulting posteriors are all statistically consistent with the true parameter values. This demonstrates that our source regularisation method is suited to lens model parameter estimation as long as we properly account for the choice of source pixel size. We note that using a source pixel size equal to that of the data (i.e. $r_{\text{pix}} = 1$) induces large pixela-

tion effects and does not lead to an acceptable fit to the data (see Fig. 3).

We have shown that a hybrid procedure of estimating the source on a pixelated grid while describing the lens mass through analytical functions does provide consistent results on simulated data. This can be extrapolated to real data only if real lens galaxies are indeed described accurately by simple analytical functions. However, if lens galaxies contain substantial complexity, such as non-elliptical mass distributions and dark substructure within the lens or along the line-of-sight, mass models must incorporate more flexibility to account for potential deviations from simple analytical distributions. Moreover, if the source model has significantly more flexibility than the lens model, part of mass complexity can be absorbed in the source reconstruction, leading to biased lens model parameters (Vernardos & Koopmans, in prep.).

6. Data in the era of ELTs

Current observing facilities, such as HST and ground-based instruments with adaptive optics (e.g. VLT and Keck), allow us to constrain many galaxy properties due to their high-resolution imaging. Future thirty-meter-class telescopes will push image quality even higher, and our modelling techniques must be robust enough to accommodate such improved data. In this section, we simulate strong lensing images from a thirty-meter-class telescope and, similar to Sect. 5, we compare our sparse source reconstruction technique to an analytical strategy.

6.1. Simulated E-ELT data

Data such as those to be obtained with the future 39 m European Extremely Large Telescope (E-ELT) are an obvious target for SLITRONOMY. The E-ELT will be equipped with, among other instruments, the near-infrared Imaging Camera for Deep Observations (MICADO, Davies et al. 2016). It is designed to observe at the diffraction limit, either in Single Conjugate Adaptive Optics (SCAO) mode, or in Multi-conjugate Adaptive Optics mode (MCAO) powered by the MAORY facility (Diolaiti 2010). This instrument is representative of the future generation of ultra-sharp imagers mounted on 30 m class telescopes.

In the following we use the same high-resolution source galaxy as in Sect. 5.1.1 (top right panel of Fig. 4), and the same lensing configuration (see Table A.2). We choose to render the image in the H band of the MICADO imager, as it is the closest to the WCF3/F160W band of HST. The pixel scale is 4 mas, or twenty times smaller than the HST pixel size in the corresponding filter. We use the Exposure Time Calculator for the E-ELT imager to obtain typical settings for the MICADO/MAORY instrument in the H band. We found that a single exposure of 1200 s corresponds to a typical setting for the type of object we are interested in. After ray-tracing, the seeing is simulated by convolving the image with the H -band MCAO PSF provided by the MAORY consortium through the package SIMCADO¹⁹ (Leschinski et al. 2016). Finally, a mix of Gaussian and Poisson noise is added, consistently with the chosen setup. All simulation settings are summarised in Table A.1. The lens and source properties are identical to the HST simulation of Sect. 5.1, and are summarised in Table A.2.

We show in Fig. 10 the simulated E-ELT strong lens system. The final square cutout is 2000×2000 pixels, which is a challenging size for many of the modelling tools currently in use. This is true not only in terms of computation time, but also in

¹⁸ The MCMC is performed using LENSTRONOMY, which runs routines from the python package EMCEE (<https://github.com/dfm/emcee>).

¹⁹ <https://github.com/astronomyk/SimCADO>

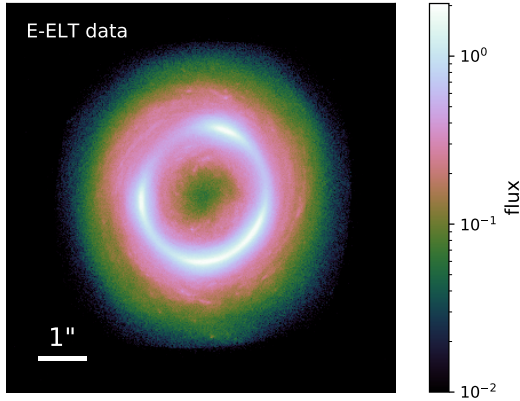


Fig. 10. Simulated E-ELT data. The lens light is assumed to be perfectly subtracted. The seeing conditions and noise covariance are based on simulated observations in the *H*-band of the MICADO/MOARY instrument. The source galaxy is the same as the one in the top right panel of Fig. 4.

terms of numerical pixel-level accuracy. One could reduce the effective data size by extracting smaller cutouts from the original image and modelling each of them individually along with a joint mass model. However, modelling the entire system at once is simpler, requires potentially less human interaction, and is expected to give much more constraining power as every pixel will contribute to the model. It also avoids the edge effects inherent to some lens inversion codes.

6.2. Source reconstruction with E-ELT data

We follow Sect. 5 and apply the SLITRONOMY algorithm to our simulated E-ELT data. We set the pixel scale of the source to half that of the data pixel, that is $r_{\text{pix}} = 2$, which differs from the HST case. With E-ELT data, we find that the resolution is already high enough that going to smaller scales only increases the computation time considerably without improving the quality of the source reconstruction. This is due to the smaller data pixel size combined with the E-ELT PSF that is much better sampled than the HST PSF. We also exclude a large fraction of pixels that contain no flux from the lensed source by defining a circular mask of 6.4 arcsec in diameter, centred on the deflector. This is to avoid the large number of pixels that contain no flux from biasing the reduced chi-squared metric by artificially driving it to low values.

We show source reconstructions in Fig. 11. In the first row is the starlet model. In the image plane the reconstruction is down to the noise, although a few isolated pixels (not visible on the plot) appear above 6σ level. The resolution and signal-to-noise of the E-ELT data, coupled with the multi-resolution property of our regularisation method, allows us to model the source galaxy at high fidelity. We are able to recover most of the features at all spatial scales simultaneously, which emphasises the advantage of using a genuine multi-scale reconstruction technique. As a result, small star-forming regions, larger-scale spiral arms and bulges, and even larger-scale luminous halos are equally well reconstructed.

Unlike with HST data, we expect an analytical model composed of shapelet basis functions superimposed on a smooth Sérsic profile to show limitations in terms of source reconstruction, because shapelets cannot capture a broad range of spatial scales simultaneously. In the second row of Fig. 11 we show such a reconstruction with a shapelets+Sérsic model obtained with

$n_{\text{max}} = 20$. Even with very high number of shapelet functions ($n_{\text{tot}} = 231$), the source cannot be modelled to the level allowed by the imaging data, both in the image and source planes.

Because of the high number of pixels in the source models, it can be difficult to visualise details. For this reason we show in Fig. 12 a zoomed-in view of both reconstructions, along with the ground truth. The two reconstructions differ considerably, and we observe how the limited number of shapelet basis functions affects the minimal and maximal scales that can be recovered (which are fixed by the pair of parameters $\{n_{\text{max}}, \beta\}$). Star-forming regions and smaller spiral arms cannot be modelled, nor the larger-scale halo (despite the Sérsic profile included). However, comparing the starlet reconstruction to the true source shows the level of detail that can be recovered thanks to sparsity and starlet constraints.

It is interesting to compare the computation time required by both methods, in this particular setting. In order to limit the number of calls to the loss function, we use the Nelder-Mead simplex minimiser to optimise the analytical model (with maximum number of iterations fixed to 50). While PSO is more efficient and better at finding the global minimum, it runs longer on average²⁰. By using the former, one therefore gets an approximate lower bound on the computation time. The SLITRONOMY algorithm takes about 7 min to perform the reconstruction with $r_{\text{pix}} = 2$, whereas the analytical reconstruction requires about 2.8 h to optimise all shapelet basis functions and Sérsic parameters. As the number of shapelet functions in the basis scales as $O(n_{\text{max}}^2)$, further increasing the maximal polynomial order drastically increases computation time.

We have noticed that going to a higher number of analytical basis functions – in combination to such a large number of data pixels – implies to allocate and invert a very large matrix, that caused memory issues. This limitation can be partially addressed by performing the source reconstruction from a down-scaled version of the original data. This trade-off between data resolution and number of basis functions in the model constitutes an alternative choice to the standard shapelets+Sérsic model discussed above. We also note that for a large number of basis functions, the computation time of the linear inversion step starts to dominate over the optimisation of non-linear parameters, such that in practice, suitable estimates of non-linear parameters are required.

While analytical methods have so far been efficient and accurate enough for many strong-lensing applications, they also have limitations. Where galaxies (lens or source) display small-scale and complex features, as can already be seen in current imaging data and will become the standard with upcoming thirty-meter-class telescopes, these techniques lack the flexibility to provide accurate source reconstructions. Moreover, their need for fine-tuning at the level of individual systems (in order to, for example, accurately recover sub-components of the lens) means increasingly more human time spent on modelling. Our model-independent sparse reconstruction technique alleviates these problems by being both highly flexible and efficient enough to handle high-resolution and morphologically complex data.

7. Conclusion

In this work we have introduced a novel implementation of the Sparse Linear Inversion Technique (SLIT) of J19, improved in

²⁰ We note that in LENSTRONOMY, other non-linear (e.g. lens model) are optimised jointly with shapelet parameters, so the computation time does not scale linearly in such a case.

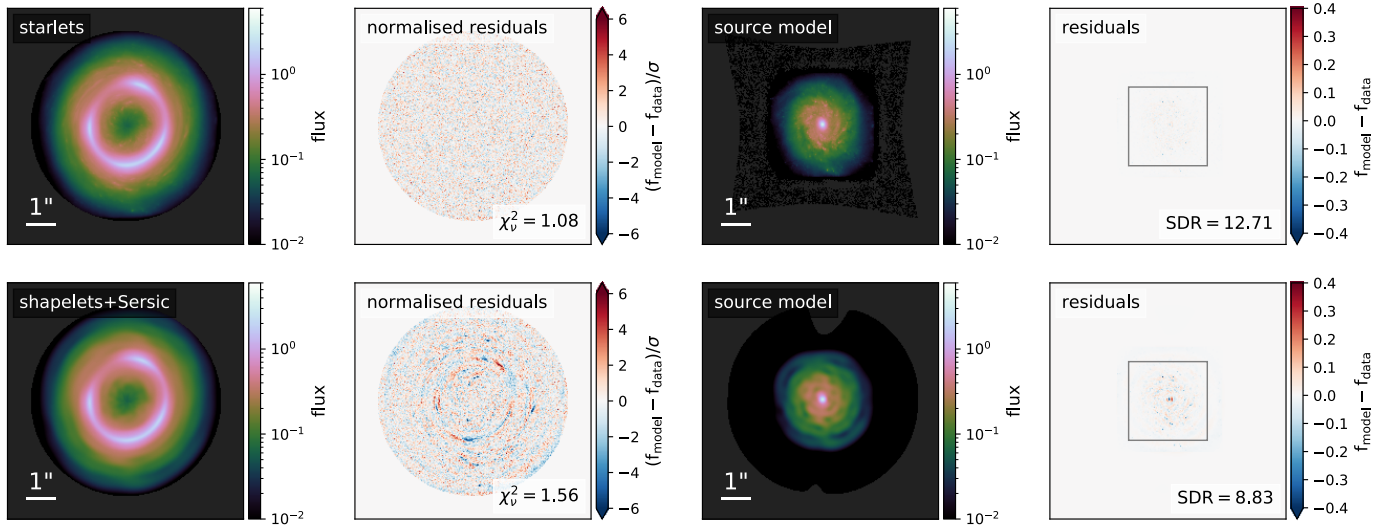


Fig. 11. Source modelling of the simulated E-ELT data of Fig. 10. *Top to bottom:* sparse model with starlets ($r_{\text{pix}} = 2$), analytical model with shapelets ($n_{\text{max}} = 20$). *Left to right:* image model, image-normalised residuals and reduced chi-square, source model, source residuals and SDR. The SDR is computed only in the region indicated by the grey box, to avoid pixels with no flux to bias its value (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb).

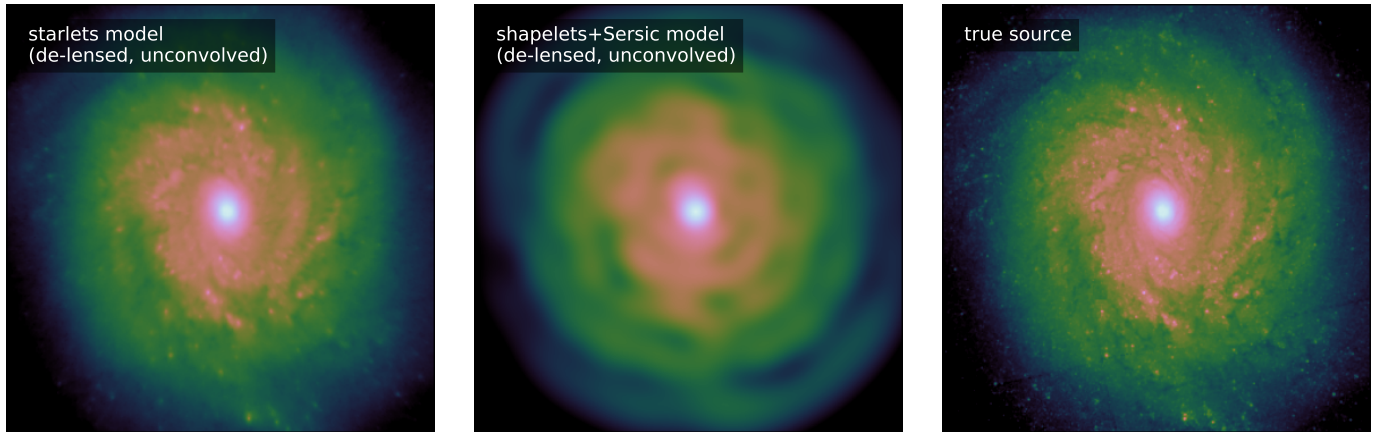


Fig. 12. Zoom of the source reconstructions shown in Fig. 11, from the modelling of simulated E-ELT data. *From left to right:* source reconstruction with starlets ($r_{\text{pix}} = 2$), source reconstruction with shapelets+Sersic profile ($n_{\text{max}} = 20$), true source. We note that images are 2000 pixels on-a-side. The colour scaling is the same across to the panels. All source reconstructions are shown at the supersampled resolution corresponding to the chosen r_{pix} (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb).

many aspects. Our SLITRONOMY plugin is fully compatible with the modelling software LENSTRONOMY (Birrer & Amara 2018) and adds efficient pixel-based reconstruction capabilities relying on sparsity and wavelets to be used jointly with analytical techniques. We demonstrated the quality of source reconstruction on various cases using mock HST data, real strong lenses from the SLACS sample, and simulated E-ELT imaging data.

We can summarise this paper as follows. Firstly, we use starlets, a specific type of wavelets, to represent the surface brightness of galaxies. Starlets are well suited to model complex sources without any need to choose an arbitrary number of analytical profiles to be stacked at arbitrary positions. No human intervention is required with wavelets, which is a genuine multi-scale decomposition of the data. Secondly, the multi-scale property of the starlet transform allows us to use smaller source pixels without a large decrease in the quality of the source reconstruction. Sparse regularisation enforces the trade-off between the reconstruction of high frequency signals in the source and noise removal. Thirdly, for future high-resolution observations, like

those obtained with thirty-meter-class telescopes, pixel-based methods may be the only viable option, as illustrated with simulated but realistic E-ELT data. Fourthly, pixel-based methods may be limited by their higher computation time compared to analytical methods, as long as the number of pixels used to represent the source remains modest, like with HST data. However, with increasing numbers of pixels per resolution element, as is the case with E-ELT data, the trend is in fact inverted, and pixel-based methods such as SLITRONOMY become faster than analytical methods.

We see the present work as a first step towards building a full wavelet-based strong lensing method where all the elements to be modelled are described with wavelets. The current SLITRONOMY implementation allows us to optimise for the deflector mass within the modelling software LENSTRONOMY, jointly reconstructing the source using starlets. However, this implementation is currently limited by the use of analytical lens mass profiles, potentially lacking of flexibility to capture complex mass distributions such as large scale twists in the mass

profile. Sparsity, combined with wavelets, is a powerful tool to reconstruct mass distributions in a model-independent way. It has been applied successfully in the context of weak lensing. Our future work in strong lensing will be focused on applying sparsity constraints for the deflector mass and light, hence leading to similar levels of flexibility in the modelling of both the source light and the lens mass and light. In doing so, it will be crucial to link the lens mass reconstruction to its kinematics, as the latter plays a vital role in alleviating the degeneracies inherent to the lensing data alone. Consequently, a full sparse modelling tool should also support the description of 2D or even 3D stellar kinematics in any complex potential well.

Acknowledgements. AG would like to thank the referee for their comments and suggestions that improved the content of the paper. AG warmly thanks S. Birrer for useful feedback related to the integration of SLITRONOMY into the LENSTRONOMY interface, and feedback and ideas on the paper. AG thanks G. Varnardos and M. Millon for insightful discussions. AG thanks X. Ding for providing the pre-processed HST images used for simulated source galaxies (used in [Ding et al. 2020](#)). AG thanks A.J. Shajib for providing the lens models for SLACS lenses (see [Shajib et al. 2020b](#)). This programme is supported by the Swiss National Science Foundation (SNSF) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (COSMICLENS: grant agreement No 787886). This research made use of LENSTRONOMY ([Birrer et al. 2015](#); [Birrer & Amara 2018](#)), SCIPY ([Virtanen et al. 2020](#)), NUMPY ([Oliphant 2006](#); [Van Der Walt et al. 2011](#)), MATPLOTLIB ([Hunter 2007](#)), SCIKIT-IMAGE ([van der Walt et al. 2014](#)), ASTROPY ([Astropy Collaboration 2013, 2018](#)) and JUPYTER ([Kluyver et al. 2016](#)).

References

- Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, *MNRAS*, **488**, 4440
- Amara, A., Metcalf, R. B., Cox, T. J., & Ostriker, J. P. 2006, *MNRAS*, **367**, 1367
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, *ApJ*, **724**, 511
- Auger, M. W., Treu, T., Brewer, B. J., & Marshall, P. J. 2011, *MNRAS*, **411**, L6
- Bernstein, G. M., & Jarvis, M. 2002, *AJ*, **123**, 583
- Bhattacharya, S., Habib, S., Heitmann, K., & Vikhlinin, A. 2013, *ApJ*, **766**, 32
- Birrer, S., & Amara, A. 2018, *Phys. Dark Univ.*, **22**, 189
- Birrer, S., Amara, A., & Refregier, A. 2015, *ApJ*, **813**, 102
- Birrer, S., Amara, A., & Refregier, A. 2017, *J. Cosmol. Astropart. Phys.*, **2017**, 037
- Birrer, S., Treu, T., Rusu, C. E., et al. 2019, *MNRAS*, **484**, 4726
- Birrer, S., Shajib, A. J., Galan, A., et al. 2020, *A&A*, **643**, A165
- Blum, K., Castorina, E., & Simonović, M. 2020, *ApJ*, **892**, L27
- Bobin, J., Moudén, Y., Starck, J. L., Fadili, J., & Aghanim, N. 2008, *Stat. Method.*, **5**, 307
- Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, *ApJ*, **638**, 703
- Brewer, B. J., & Lewis, G. F. 2006, *ApJ*, **651**, 8
- Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, *ApJ*, **744**, 41
- Candes, E. J., Wakin, M. B., & Boyd, S. P. 2007, ArXiv e-prints [arXiv:0711.1612]
- Chianese, M., Coogan, A., Hofma, P., Otten, S., & Weniger, C. 2020, *MNRAS*, **496**, 381
- Combettes, P. L., & Wajs, V. R. 2005, *Multiscale Modeling & Simulation*, **4**, 1168
- Cong Vu, B. 2011, ArXiv e-prints [arXiv:1110.1697]
- Davies, R., Schubert, J., Hartl, M., et al. 2016, *Ground-based and Airborne Instrumentation for Astronomy VI*
- Deep, A., Fiorentino, G., Tolstoy, E., et al. 2011, *A&A*, **531**, A151
- Ding, X., Treu, T., Shajib, A. J., et al. 2018, ArXiv e-prints [arXiv:1801.01506]
- Ding, X., Treu, T., Birrer, S., et al. 2020, *MNRAS*, **503**, 1096
- Ding, X., Treu, T., Birrer, S., et al. 2021, *MNRAS*, **501**, 269
- Diolaiti, E. 2010, *The Messenger*, **140**, 28
- Dye, S., & Warren, S. J. 2005, *ApJ*, **623**, 31
- Dye, S., Evans, N. W., Belokurov, V., Warren, S. J., & Hewett, P. 2008, *MNRAS*, **388**, 384
- Farrens, S., Grigis, A., El Gueddari, L., et al. 2020, *Astron. Comput.*, **32**, 100402
- Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, *ApJ*, **667**, 176
- Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., & Ruff, A. 2012, *ApJ*, **761**, 170
- Gilman, D., Birrer, S., & Treu, T. 2020, *A&A*, **642**, A194
- Han, J., Eke, V. R., Frenk, C. S., et al. 2015, *MNRAS*, **446**, 1356
- Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, *ApJ*, **823**, 37
- Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, *Nature*, **548**, 555
- Högbom, J. A. 1974, *A&AS*, **15**, 417
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- James, B. L., Auger, M., Pettini, M., et al. 2018, *MNRAS*, **476**, 1726
- Jiang, M., Bobin, J., & Starck, J. L. 2017, *SIAM J. Imaging Sci.*, **10**, 1997
- Joseph, R., Courbin, F., & Starck, J. L. 2016, *A&A*, **589**, A2
- Joseph, R., Courbin, F., Starck, J. L., & Birrer, S. 2019, *A&A*, **623**, A14
- Keeton, C. R. 2001, ArXiv e-prints [arXiv:astro-ph/0102340]
- Kennedy, J., & Eberhart, R. C. 2001, *Swarm Intelligence* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.)
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, eds. F. Loizides, & B. Schmidt (IOS Press), 87
- Kneib, J. P., Bonnet, H., Golse, G., et al. 2011, *LENSTOOL: A Gravitational Lensing Software for Modeling Mass Distribution of Galaxies and Clusters (strong and weak regime)*
- Kochanek, C. S., & Narayan, R. 1992, *ApJ*, **401**, 461
- Krist, J. E., Hook, R. N., & Stoehr, F. 2011, *Proc. SPIE*, **8127**, 81270J
- Lanusse, F., Starck, J. L., Leonard, A., & Pires, S. 2016, *A&A*, **591**, A2
- Leschinski, K., Czoske, O., Köhler, R., et al. 2016, *Proc. SPIE*, **9911**, 991124
- Meinhardt, T., Moeller, M., Hazirbas, C., & Cremers, D. 2017, 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 1799
- Melchior, P., Böhnert, A., Lombardi, M., & Bartelmann, M. 2010, *A&A*, **510**, A75
- Meng, X.-L., Treu, T., Agnello, A., et al. 2015, *J. Cosmol. Astropart. Phys.*, **2015**, 059
- Morningstar, W. R., Hezaveh, Y. D., Perreault Levasseur, L., et al. 2018, ArXiv e-prints [arXiv:1808.00011]
- Morningstar, W. R., Perreault Levasseur, L., Hezaveh, Y. D., et al. 2019, *ApJ*, **883**, 14
- Natarajan, P., De Lucia, G., & Springel, V. 2007, *MNRAS*, **376**, 180
- Nightingale, J. W., & Dye, S. 2015, *MNRAS*, **452**, 2940
- Nightingale, J. W., Dye, S., & Massey, R. J. 2018, *MNRAS*, **478**, 4738
- Oguri, M. 2010, *glafic: Software Package for Analyzing Gravitational Lensing*
- Oliphant, T. E. 2006, *A guide to NumPy* (USA: Trelgol Publishing), 1
- Pearson, J., Li, N., & Dye, S. 2019, *MNRAS*, **488**, 991
- Peel, A., Lanusse, F., & Starck, J.-L. 2017, *ApJ*, **847**, 23
- Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. 2017, *ApJ*, **850**, L7
- Powell, D., Vegetti, S., McKean, J. P., et al. 2021, *MNRAS*, **501**, 515
- Pratley, L., McEwen, J. D., d'Avezac, M., et al. 2018, *MNRAS*, **473**, 1038
- Refregier, A. 2003, *MNRAS*, **338**, 35
- Ritondale, E., Vegetti, S., Despali, G., et al. 2019, *MNRAS*, **485**, 2179
- Rizzo, F., Vegetti, S., Fraternali, F., & Di Teodoro, E. 2018, *MNRAS*, **481**, 5606
- Saha, P. 2000, *AJ*, **120**, 1654
- Schneider, P., & Sluse, D. 2013, *A&A*, **559**, A37
- Shajib, A. J., Birrer, S., Treu, T., et al. 2019, *MNRAS*, **483**, 5649
- Shajib, A. J., Birrer, S., Treu, T., et al. 2020a, *MNRAS*, **494**, 6072
- Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2020b, *MNRAS*, submitted [arXiv:2008.11724]
- Sonnenfeld, A. 2018, *MNRAS*, **474**, 4648
- Starck, J. L., Moudén, Y., Bobin, J., Elad, M., & Donoho, D. L. 2005, *SPIE Conf. Ser.*, **5914**, 209
- Starck, J.-L., Fadili, J., & Murtagh, F. 2007, *IEEE Trans. Image Proc.*, **16**, 297
- Sureau, F., Lechat, A., & Starck, J.-L. 2020, *A&A*, **641**, A67
- Suyu, S. H., & Halkola, A. 2010, *A&A*, **524**, A94
- Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, *MNRAS*, **371**, 983
- Suyu, S. H., Marshall, P. J., Blandford, R. D., et al. 2009, *ApJ*, **691**, 277
- Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, *ApJ*, **766**, 70
- Tagore, A. S., & Jackson, N. 2016, *MNRAS*, **457**, 3066
- Tagore, A. S., & Keeton, C. R. 2014, *MNRAS*, **445**, 694
- Treu, T., & Koopmans, L. V. E. 2004, *ApJ*, **611**, 739
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comput. Sci. Eng.*, **13**, 22
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, ArXiv e-prints [arXiv:1407.6245]
- 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
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Wallington, S., Kochanek, C. S., & Narayan, R. 1996, *ApJ*, **465**, 64
- Warren, S. J., & Dye, S. 2003, *ApJ*, **590**, 673
- Williams, L. L. R., & Saha, P. 2004, *AJ*, **128**, 2631
- Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, *MNRAS*, **498**, 1420
- Xu, D., Sluse, D., Schneider, P., et al. 2016, *MNRAS*, **456**, 739

Appendix A: Properties of simulated HST and E-ELT data

Table A.1. Instrumental and observational settings used for HST and E-ELT simulated data (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/fixed_param.py).

	HST	E-ELT
Instrument/filter	WCF3/F160W	MICADO/H-band
Pixel size [mas]	80	4
Single-exposure time [s]	500	1200
Number of exposures	4	1
Zero-point [mag]	27	34
Sky brightness [mag arcsec ⁻²]	/	25
Read noise [e ⁻]	/	3
CCD gain [e ⁻ /ADU]	2.5	1
Background noise [e ⁻]	0.05	/
PSF kernel (simulated)	TINYTIM	SIMCADO

Table A.2. Model choices and parameters used for both HST and E-ELT mock data simulations.

Deflector, $z_d = 0.3$	
Singular isothermal ellipsoid (SIE)	
Velocity dispersion, σ_v	260 km s ⁻¹
Ellipticity, q_m	0.8
Orientation, ϕ_m	0
Position, (RA = x_0 , Dec = y_0) [arcsec]	(0, 0)
External shear	
Strength, γ_{ext}	0.03
Orientation, ϕ_{ext}	$\pi/4$
Source, $z_s = 1.2$	
Source 1: NGC 3982	
Position, (RA, Dec) [arcsec]	(0.05, 0.05)
Magnitude	22
Source 2: ESO498G5	
Position, (RA, Dec) [arcsec]	(-0.8, -0.6)
Magnitude	22
Source 3: NGC 3259	
Position, (RA, Dec) [arcsec]	(0.3, 0.8)
Magnitude	23

Notes. The coordinates are oriented such as north is up and east is right. The orientation is zero when aligned with the horizontal axis and increases anti-clockwise. Images of the source galaxies were obtained from the HST archival database. “Source 1” is the main source component in all simulations in this paper, whereas “Source 2” and “Source 3” are only used as member of the simulated galaxy group of Fig. 6 (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/fixed_param.py).

In this section we summarise all fixed parameters that define the simulations used for source reconstructions in Sects. 5 and 6. They can be separated into two categories. Instrumental settings that control the resolution, seeing and signal-to-noise of the data are shown in Table A.1. Model parameters including redshifts, mass distributions of the deflector and positions of the source galaxy(-ies) are specified in Table A.2.

HST instrumental settings are inspired by typical near-infrared imaging data required for time delay cosmography after drizzling. E-ELT data settings are inspired by Meng et al. (2015, although for the Thirty Meter Telescope) and Deep et al. (2011,

limiting magnitudes), and the E-ELT Exposure Time Calculator²¹.

Lens model parameters were picked to be fairly representative of an average strong lens system. The mass distribution is described by the singular isothermal ellipsoid (SIE), sustained in a moderate external shear field, and slightly misaligned with the SIE orientation.

We use the SimulationAPI module of LENSTRONOMY to generate realistic imaging data based on instrumental and model parameters above.

Appendix B: Hybrid approach for automated modelling

In Sect. 5.1.3 we compare source reconstructions obtained with starlet and shapelet models on simulated HST data. For a complex source, the flexibility of a pixel-based approach outperformed the default analytical profile. However, analytical modelling has the advantage of being faster, and thus is better suited to batch automated modelling. Furthermore, for lower resolution large datasets like future surveys will deliver, the quality of analytical source reconstructions is often sufficient for reliable population-level lens model optimisation.

For these reasons, we propose a workflow using both pixel-based and analytical methods to refine in an automated manner an initially poor source model. Figure 7 illustrates the decision flow. Here we give a description of each step, assuming an approximation of the lens model has been previously obtained, for example through optimisation with a simple source profile:

1. if current model residuals display features that are likely to come from the limiting source profile (e.g. after visual inspection or chi-square) we perform a pixel-based source reconstruction using sparse modelling;
2. the reconstructed map of source surface brightness is smoothed using a Laplacian-of-Gaussian filter, to enhance individual source sub-components. A local maxima detection algorithm is applied on the filtered map to find the location of each sub-component;
3. an analytical model is initialised as one joint shapelets+Sérsic pair centred on each sub-component i . Shapelet maximal orders are set to a low $n_{\text{max},i} = 2$;
4. the source is re-optimised, and the quality of image plane residuals at the location of each lensed sub-component are used to iteratively refine the maximal order of each shapelet basis independently;
5. once localised residuals are below a given threshold, a final full optimisation (mass+source) can be performed.

The step refining individual shapelet polynomial orders requires establishing a suitable update rule to allow a smooth increase in model complexity without introducing too many degrees of freedom that may not be supported by the data. We found quite successful the following update rule: given a reduced $\chi^2_{v,i}$ based on residuals restricted to the location of the lensed sub-component i , the shapelet order is updated according to

$$n_{\text{max},i}^{\text{new}} = \begin{cases} n_{\text{max},i}^{\text{old}} + \max(1, 10 \times [\chi_{v,i}^2 - 1.1]) & \text{if } \chi_{v,i}^2 > 1.1, \\ n_{\text{max},i}^{\text{old}} & \text{if } \chi_{v,i}^2 \leq 1.1. \end{cases} \quad (\text{B.1})$$

²¹ <http://www.eso.org/observing/etc/bin/gen/form?INS.NAME=ELT+INS.MODE=swimaging>

Additionally, we set an overall maximal value of $n_{\max} = 14$ to prevent the procedure above from leading to an intractable number of basis functions.

We see the procedure described in Fig. 7 as a potential component of a larger pipeline for automatised modelling of strong lens systems. Various pipelines are currently in development based on recent works like the uniform modelling strategies proposed in Shajib et al. (2019, 2020b), the automatic differentiable pipeline of Chianese et al. (2020), and workflows developed by the teams that took part in the Time Delay Lens Modelling Challenge (Ding et al. 2020).

Appendix C: Threshold refinement for sparsity constraint

As shown in Eq. (8) we enforce sparsity on reconstructed surface brightness maps by minimising their ℓ_1 norm in starlet space. Sparsity can also be enforced through the ℓ_0 norm instead, but it is not convex like the ℓ_1 norm and tends to produce in practice more artefacts in the presence of noise. However, enforcing minimal ℓ_1 norm in the presence of noise has a greater dependence on the amplitude of each pixel, which can potentially bias the reconstruction. We employ a similar solution as proposed by Candes et al. (2007) by iteratively correcting for this bias through additional “reweighting” steps. Once the solution of the original unweighted problem has been found ($W_x^0 = 1$), the minimisation is performed a second time with threshold per pixel per starlet scale updated according to the following rule, for a given iteration n

$$W_x^n = \begin{cases} 1 & \text{if } n = 0, \\ \frac{1}{1 + \exp[10 \times (\Phi^\top x^{n-1} - \lambda)]} & \text{if } n > 0, \end{cases} \quad (\text{C.1})$$

where x^{n-1} stands either for the source or lens light at the previous iteration, and λ is the regularisation strength in units of noise per pixel in starlet space, $\lambda \equiv \lambda' \sigma_\Phi$, after proper propagation through appropriate operators of Eq. (6). In practice, a single additional optimisation loop with the weights updated according to the above equation already improve the source reconstruction significantly, as better residuals and lower chi-squared are obtained. Analogous to J19, we found that the update rule of Eq. (C.1) is suitable for our lens modelling application. The specific form of the reweighting scheme is not unique, however it has to decrease the effective threshold for specific starlet coefficients that are significant enough, from previous iterations, relative to the noise.

Appendix D: Artefacts in source model of SDSS J0959–0410

In Sect. 8, we briefly discuss reasons for the artefacts visible in the SDSS J0959–0410 source reconstruction with starlets. These artefacts are in part due to the pixelated approximation inherent to pixel-based modelling techniques, but mostly due to the complexity of the system and the difficulty of correctly modelling both the source and lens light distributions. To illustrate this, we create a realistic mock of the system SDSS J0959–0410 by simulating lensed arcs based on our starlets source reconstruction, after downsampling it to the data resolution for removing the artefacts. We take the same PSF, exposure time, and background noise levels as in the real data. By construction, this also ensures no contamination by the lens light. We then perform a

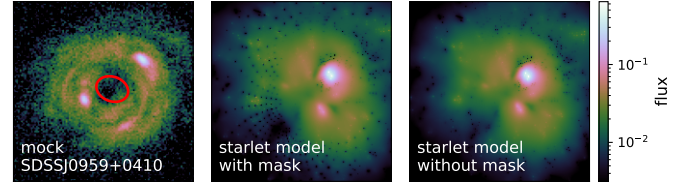


Fig. D.1. Origin of the artefacts seen in the starlet model of the SLACS system SDSS J0959–0410. *Left*: mock imaging data, with mask region covering the real lens light indicated in red. *Middle*: starlet reconstruction obtained with a mask on the central region (same as in Fig. 8), where artefacts from inaccurate lens light modelling are clearly visible. *Right*: starlet reconstruction without masking the central region, which suppresses the artefacts (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/SLACS_source_reconstruction.ipynb).

starlet reconstruction from the mock, with and without the central region at the location of the lens. We show both reconstructions side by side in Fig. D.1. We observe no artefacts in the reconstruction when not masking the central region, which confirms that these artefacts are primarily due to limitations of the analytical lens light modelling, followed by an inaccurate masking of the residual light.

Appendix E: Effect of the number of starlet scales

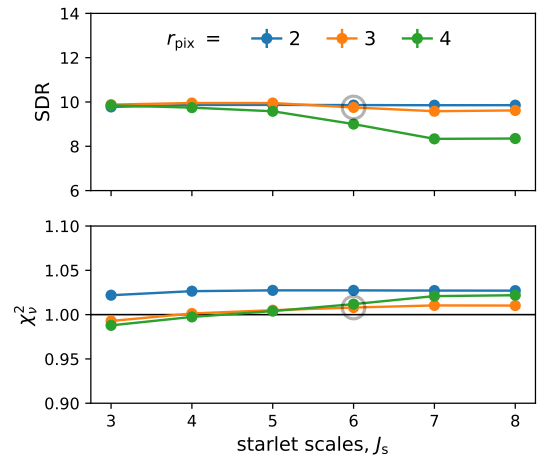


Fig. E.1. Effect of the choice of starlet decomposition scales, J_s , on the modelling quality, assessed by the two metrics introduced in Sect. 5.1.2. The value $J_s = 6$ chosen throughout this work is indicated by the grey circle (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/mock_source_reconstruction.ipynb).

In the implementation of our source reconstruction algorithm, which represents the surface brightness on a pixelated grid, we are subject to the choice of the source pixel size. In this work we introduce r_{pix} as the ratio between image and source pixel sizes. It is known that the choice of r_{pix} can impact the accuracy of lens model parameters (see Sect. 5.3). Here we assume a lens model that has been estimated accurately enough, for instance, using the scheme shown on Fig. 7.

The additional parameter specifically introduced by the starlet transform is the number of decomposition scales J_s and J_l for source and lens light, respectively. We show in Fig. E.1 image plane (χ_v^2) and source plane (SDR) metrics to assess the modelling quality for different J_s values. We take the source reconstruction of the single-component simulated HST data (top row

of Fig. 4) as a test case. The number of decomposition scales has a small impact on the quality of the reconstruction, and we choose $J_s = 6$ for reconstructions performed in this work. The number of decomposition scales can also simply be set to its maximal value, defined by the size of the region being regularised. For instance, if the source plane is defined on a grid of n_{pix} pixels on a side, the maximum number of scales allowed by the wavelet transform is $J_{s,\text{max}} = \lfloor \log_2 n_{\text{pix}} \rfloor$.

The choice of number of decomposition scale also affects the computation time considerably, as we show in Appendix F below.

Appendix F: SLITRONOMY benchmarks

In the current implementation of SLITRONOMY, lens model parameters can be optimised through different engines supported by LENSTRONOMY. This requires solving the sparse reconstruction problem for a given set of lens model parameters, computing the likelihood based on comparing the model with the imaging data, finally proposing a new set of lens model parameters, and so on. The total computation time required to find the best-fit lens model obviously depends heavily on the amount time necessary to solve Eq. (6), as it will be solved at each iteration. To have a rough approximation of the computation time necessary to solve the sparse source reconstruction (S in Eq. (6)), we give in Fig. F.1 computation times for various settings of the algorithm and data sizes.

When not explicitly mentioned in the panels, the baseline choice is similar to our HST reconstructions in Sect. 5: the simulated image is 100×100 pixels, with an image to source resolution $r_{\text{pix}} = 3$ and starlet scales $J_s = 6$.

In the top row we show timings for a subset of steps required for our algorithm to start, hence they are computed only once for a given lens model. They are shown in units of milliseconds. The top left panel shows the computation of the noise per pixel, propagated from image to source plane, including PSF effects and transformed in starlet space. This allows for a proper setting of the regularisation strength (see Sect. 3 for details). The top central and right panels show the computation of the lensing operator F for various numbers of image pixels and their corresponding number of source pixels through the choice of r_{pix} .

In the bottom row we consider the total time required to optimise the source light given a lens model. The first panel shows the runtime as a function of the number of source pixels, which can be automatically set to a minimal value when, for instance, using a mask in the image plane (e.g. in Fig. 11). The masked region, once mapped back to source plane, defines a much smaller angular region on which the pixelated source light is defined, hence reducing the computation time considerably for both the lensing operator and subsequent mapping operations. The two following panels show the influence of the resolution factor r_{pix} , the number of image pixels, and the number of decomposition scales J_s on the computation time.

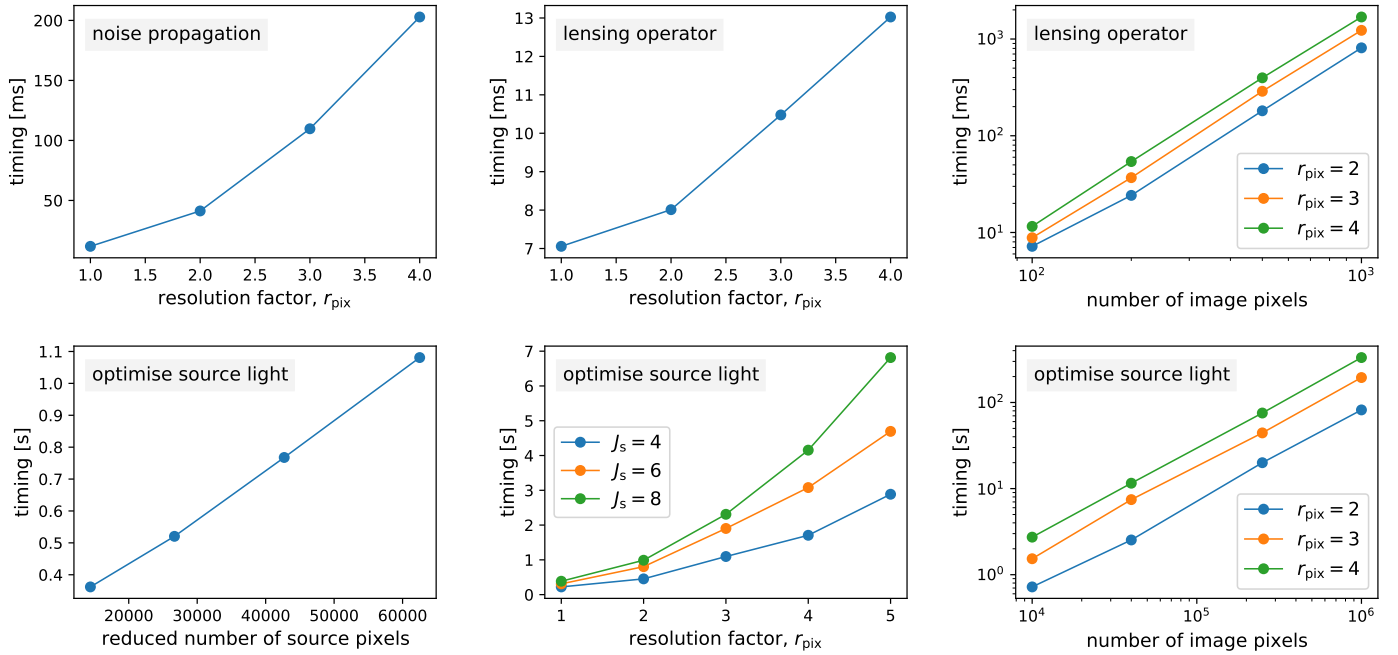


Fig. F.1. Runtimes for various settings of the sparse source solver of SLITRONOMY. *Top left:* propagation of the noise to transformed starlet space, for different source to image resolution r_{pix} . *Top centre:* computation of the lensing operator F for different r_{pix} . *Top right:* computation of F for total number of data pixels. *Bottom left:* optimisation of the source light \mathbf{s} for different choices of the minimum allowed total number of source pixels. *Bottom centre:* optimisation of \mathbf{s} for different r_{pix} and number of source starlet scales J_s . *Bottom right:* optimisation of \mathbf{s} for varying total number of data pixels. We refer to the text for more details (https://github.com/aymgal/SLITronomy-papers/blob/master/paper_I/benchmarks.ipynb).