Université de Montréal

Horseshoe Regularization for Wavelet-based Lensing Inversion

par

Hasti Nafisi

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Discipline

March 18, 2024

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Horseshoe Regularization for Wavelet-based Lensing Inversion

présenté par

Hasti Nafisi

a été évalué par un jury composé des personnes suivantes :

Guillaume Lajoie
(président-rapporteur)

Guy Wolf
(directeur de recherche)

Yashar Hezaveh
(codirecteur)

Laurence Perreault Levasseur
(membre du jury)

Résumé

Les lentilles gravitationnelles se produisent lorsque le champ gravitationnel d'un objet massif dévie la trajectoire de la lumière provenant d'un objet lointain, entraînant une distorsion ou une amplification de l'image de l'objet lointain.

La transformation Starlet fournit une méthode robuste pour obtenir une représentation éparse des images de galaxies, capturant efficacement leurs caractéristiques essentielles avec un minimum de données. Cette représentation réduit les besoins de stockage et de calcul, et facilite des tâches telles que le débruitage, la compression et l'extraction de caractéristiques.

La distribution a priori de fer à cheval est une technique bayésienne efficace pour promouvoir la sparsité et la régularisation dans la modélisation statistique. Elle réduit de manière agressive les valeurs négligeables tout en préservant les caractéristiques importantes, ce qui la rend particulièrement utile dans les situations où la reconstruction d'une image originale à partir d'observations bruitées est difficile.

Étant donné la nature mal posée de la reconstruction des images de galaxies à partir de données bruitées, l'utilisation de la distribution a priori devient cruciale pour résoudre les ambiguïtés. Les techniques utilisant une distribution a priori favorisant la sparsité ont été efficaces pour relever des défis similaires dans divers domaines.

L'objectif principal de cette thèse est d'appliquer des techniques de régularisation favorisant la sparsité, en particulier la distribution a priori de fer à cheval, pour reconstruire les galaxies d'arrière-plan à partir d'images de lentilles gravitationnelles.

Notre méthodologie proposée consiste à appliquer la distribution a priori de fer à cheval aux coefficients d'ondelettes des images de galaxies lentillées. En exploitant la sparsité de la représentation en ondelettes et le comportement de suppression du bruit de la distribution a priori de fer à cheval, nous obtenons des reconstructions bien régularisées qui réduisent le bruit et les artefacts tout en préservant les détails structurels. Des expériences menées sur des images simulées de galaxies lentillées montrent une erreur quadratique moyenne inférieure et une similarité structurelle plus élevée avec la distribution a priori de fer à cheval par rapport à d'autres méthodes, validant son efficacité.

Mots clés: Prior de fer à cheval, Technique bayésienne, Transformée Starlet, Lentille gravitationnelle, Régularisation, Problème inverse

Abstract

Gravitational lensing, a phenomenon in astronomy, occurs when the gravitational field of a massive object, such as a galaxy or a black hole, bends the path of light from a distant object behind it. This bending results in a distortion or magnification of the distant object's image, often seen as arcs or rings surrounding the foreground object. The Starlet wavelet transform offers a robust approach to representing galaxy images sparsely. This technique breaks down an image into wavelet coefficients at various scales and orientations, effectively capturing both large-scale structures and fine details.

The Starlet wavelet transform offers a robust approach to representing galaxy images sparsely. This technique breaks down an image into wavelet coefficients at various scales and orientations, effectively capturing both large-scale structures and fine details.

The horseshoe prior has emerged as a highly effective Bayesian technique for promoting sparsity and regularization in statistical modeling. It aggressively shrinks negligible values while preserving important features, making it particularly useful in situations where the reconstruction of an original image from limited noisy observations is inherently challenging.

The main objective of this thesis is to apply sparse regularization techniques, particularly the horseshoe prior, to reconstruct the background source galaxy from gravitationally lensed images. By demonstrating the effectiveness of the horseshoe prior in this context, this thesis tackles the challenging inverse problem of reconstructing lensed galaxy images.

Our proposed methodology involves applying the horseshoe prior to the wavelet coefficients of lensed galaxy images. By exploiting the sparsity of the wavelet representation and the noise-suppressing behavior of the horseshoe prior, we achieve well-regularized reconstructions that reduce noise and artifacts while preserving structural details. Experiments conducted on simulated lensed galaxy images demonstrate lower mean squared error and higher structural similarity with the horseshoe prior compared to alternative methods, validating its efficacy as an efficient sparse modeling technique.

Key words: Horseshoe prior, Bayesian technique, Starlet transform, Gravitational lensing, Regularization, Inverse problem

Contents

Résumé 5

Abstract 7

List of tables 13

List of figures 15

List of Symbols and Abbreviations 17

Remerciements 19

Introduction 21

Overview of The Chapters 22

Chapter 1. Background 25

Chapter 2. Statistical context 27

- 2.1. Inverse problems 27
  - 2.1.1. Concept of an inverse problem 27
  - 2.1.2. Introduction to inverse problems 28
  - 2.1.3. Ill-posedness and well-conditioning 29
- 2.2. Probability theory and Bayesian Inference 30
  - 2.2.1. Bayes' Theorem 30
  - 2.2.2. Likelihood 31
    - 2.2.2.1. The Multivariate Gaussian Likelihood 31
    - 2.2.2.2. The Maximum Likelihood Estimation 32
- 2.3. Regularization and Prior specification 33
  - 2.3.1. Regularization 33
  - 2.3.2. The Priors 34
    - 2.3.2.1. The Laplace prior (ℓ_1 Regularization) 34
    - 2.3.2.2. The Gaussian prior (ℓ_2 Regularization) 35

2.3.2.3. The Horseshoe Prior ... 36
2.3.2.4. Advantages of using the horseshoe prior for sparsity ... 37

Chapter 3. Cosmological context ... 39

3.1. Gravitational lensing ... 39
3.1.1. The basics of gravitational lensing ... 39
3.1.2. Deflection angle ... 40
3.1.3. Lens equation ... 43
3.1.4. Exploring Gravitational Lensing ... 45

Chapter 4. Starlet Wavelet transform ... 47

4.1. Wavelet Introduction ... 47
4.1.1. Parameterizing the Background Source ... 49
4.1.2. Introduction to the Starlet basis ... 50
4.1.2.1. The starlet basis promotes sparsity in different scales ... 52

Chapter 5. Methodology ... 53

5.1. Methodology ... 53
5.1.1. Data Preprocessing ... 53
5.1.1.1. Constructing the Lensing Matrix ... 53
5.1.1.2. Image Conversion and Standardization ... 54
5.1.2. Lensed Image Simulation ... 55
5.1.3. Simulating Noise ... 55

5.2. Approaches for Solving the Inverse Problem ... 56
5.2.1. Bayesian Computation Using MCMC ... 56
5.2.1.1. Markov Chain Monte Carlo ... 56
5.2.1.2. Hamiltonian Monte Carlo ... 58
5.2.1.3. No-U-Turn Sampler ... 60
5.2.2. The proposed approach for galaxy image reconstruction using HMC ... 61
5.2.3. Methodological Foundations: Probabilistic Reconstruction of Lensed Images 62

5.3. Error Measures ... 63
5.3.1. Mean Squared Error (MSE) ... 63
5.3.2. Structural Similarity Index (SSIM) ... 63

Chapter 6. Results ... 65

6.1. Results 65
6.1.1. Description of the simulated data 65
6.1.2. Implementation Details and Probabilistic Framework 68
6.1.3. Result Visualization 68
6.1.3.1. Image Reconstruction with Horseshoe Prior 68
6.1.3.2. Comparing Priors in Galaxy Image Reconstruction 74

Chapter 7. Discussion and conclusion 79

7.1. Discussion 79

7.2. Conclusion 80
7.2.1. Future Work 80

References 83

Appendix A. Appendix A 89

A.1. Gradient Descent-Based Regularization Framework for Galaxy Image Reconstruction 89
A.1.1. Regularization Techniques and Implementation Details 90

Appendix B. Appendix B 93

B.1. Detailed Analysis of Gravitational Lensing Inference 93

B.2. Example and Practical Implementation 94

List of tables

6.1 Comparison of Mean Squared Errors (MSE) for Horseshoe, Gaussian, and Laplace priors across different noise levels. 76

6.2 Structural Similarity Index (SSIM) values for Horseshoe, Gaussian, and Laplace priors under different noise conditions, indicating the robustness of each prior against noise-induced degradation in image quality. 77

List of figures

2.1 Comparing of different priors, adapted from [16] 37

2.2 Comparing of different priors' tail, adapted from [16] 37

3.1 Gravitational lensing effect, adapted from esahubble.org by NASA, ESA & L. Calçada 40

3.2 Strong and Week Gravitational lensing, adapted from frontierfields.org by A. Feild (STScI) 41

3.3 Typical configuration of a lensing system, adapted from [51] 43

4.1 decomposition of the original lensed galaxy and its Coefficients from scales j=1 to 6 52

5.1 Lensing function 54

5.2 Simulated noisy data of a lensed galaxy 56

5.3 Flowchart of the Gravitational Lensing Data Generation Process 57

5.4 Illustratin of the Proposed Galaxy Image Reconstruction Methodology. This figure outlines the step by step procedure for reconstructing galaxy images using the novel approach described in this thesis. The methodology includes the incorporation of a lensing operator, wavelet transform, likelihood and prior models, and Hamiltonian Monte Carlo sampling and the lensed image 62

6.1 Galaxy NGC 4414, Credit: The Hubble Heritage Team 66

6.2 (a) Preprocessed image of NGC 4441, illustrating the distribution of pixel intensities (b) Noisy, lensed observation of NGC 4441 66

6.3 Illustration of multiscale decomposition of NGC 4441 via the Starlet wavelet transform: The original galaxy image is presented in the left panel, the middle panels exhibit the coefficient images at each of the 6 wavelet scales utilized, and the right panel depicts the image reconstructed by summing all scales of the decomposition 67

6.4 Reconstructed image of Galaxy NGC 4414 with Horseshoe prior. 69

6.5 Coefficients of the reconstructed image with Horseshoe prior, demonstrating the hierarchical scale decomposition. 69

6.6 Comparative analysis of the Reconstructed Image (with Horseshoe Regularization) versus the Ground Truth Image of Galaxy NGC 4414, accompanied by histograms of pixel values. 70

6.7 Histograms showcasing the distribution of mean coefficient values across different scales in the reconstructed image with Horseshoe regularization 71

6.8 Galaxy Image reconstruction using the Starlet transform with a uniform prior . . . 72

6.9 Galaxy Image reconstruction with horseshoe prior and different noise levels 72

6.10 (a) Galaxy 3147, Credit:ESA/Hubble and NASA, (b) Observed noisy image 73

6.11 A comparison between the reconstructed galaxy 3147 (left) and the unaltered original image (right). The histograms below each image display the distribution of pixel intensity values, showcasing the success of the reconstruction process in preserving the structural details of the galaxy. 74

6.12 Galaxy Image reconstruction with different priors and different noise levels 75

6.13 Graphical representation of Mean Squared Errors (MSE) for Horseshoe, Gaussian, and Laplace priors, illustrating the impact of increasing noise levels on estimation accuracy. 76

6.14 Visual comparison of the Structural Similarity Index (SSIM) across varying noise levels for Horseshoe, Gaussian, and Laplace priors, demonstrating the preservation of image structural integrity under noise influence. 77

A.1 Galaxy 4441 Reconstructed image employing the horseshoe prior with optimization. 91

B.1 (a) Reconstructed Galaxy Image using optimized parameters with $\lambda = 0.01$ and a diagonal C_n, (b) Corresponding residuals illustrating the variance from the original data 95

B.2 (a) depicts the reconstructed galaxy image starting from a random matrix, (b) the reconstruction beginning with a zero matrix 96

List of Symbols and Abbreviations

MCMC	Markov Chain Monte Carlo
HMC	Hamiltonian Monte Carlo
NUTS	No-U-Turn Sampler
MSE	Mean Squared Error
SSIM	Structural Similarity Index
HS	Horseshoe Prior
L1	Laplace Prior (ℓ_1 Regularization)
L2	Gaussian Prior (ℓ_2 Regularization)
SSP	Significant Structural Patterns
HST	Hubble Space Telescope

Remerciements

I would like to express my heartfelt appreciation to my supervisor, Prof. Guy Wolf, and co-supervisor, Prof. Yashar Hezaveh, for their invaluable support and guidance throughout my master's thesis. Their expertise has been crucial in shaping the success of my research. Additionally, I am grateful to my parents for their unwavering encouragement and love, which provided the motivation and strength needed to complete this academic journey. None of this would have been achievable without the support and continuous encouragement that my sister provided throughout my graduate studies.

Introduction

Gravitational lensing, has the remarkable effect of bending and distorting the light from distant galaxies as it traverses the gravitational fields of intervening objects. This cosmic phenomenon presents astronomers with a tantalizing opportunity: the chance to peer into the deep cosmos through this natural lens and observe galaxies that would otherwise remain hidden. This opportunity comes with a formidable challenge—reconstructing the original, undistorted images of these galaxies.

Reconstructing the original undistorted galaxy images is a challenging ill-posed inverse problem. The goal is to uncover the true galactic morphology from distorted, noisy observations with missing information.

Representing the galaxy image using the Starlet wavelet basis provides an effective approach. The wavelet decomposition yields a sparse representation that compresses the image content into few significant coefficients. This sparsity makes wavelets well-suited for modeling complex galaxy images. However, the reconstruction still requires overcoming ill-posedness and noise.

In recent years, statistical techniques that promote sparsity have emerged as a powerful tool for addressing this challenge. Among these, the horseshoe prior, a Bayesian approach to sparsity-inducing regression and image analysis, has shown remarkable efficacy. Its ability to shrink insignificant values while retaining salient features aligns well with promoting sparsity. In sparse signal processing, the horseshoe prior has been extensively studied for its effectiveness in estimating sparse signals, particularly in signal processing, machine learning, and image processing. This prior is advantageous for its high accuracy and efficiency in applications like signal reconstruction and has been applied in various fields including physics for experimental measurements [44]. By applying the horseshoe prior to wavelet coefficients of the lensed galaxy image, we can achieve effective regularization. The horseshoe prior complements the sparsity of the wavelet basis, suppressing noise while preserving important galactic substructures.

Moreover, the wavelet transform, particularly the Starlet basis, offers an optimal framework for sparsely representing the hierarchical structures frequently observed in astronomy images. The coefficients generated by the wavelet transformation tend to exhibit sparsity, allowing us to compactly represent complex visual information. When combined with horseshoe regularization, this approach offers a comprehensive strategy for addressing the ill-posed nature of galaxy image reconstruction.

This work seeks to explore the synergistic combination of the Starlet transform's sparse representation and the horseshoe prior's shrinkage behavior to enable accurate reconstructions from noisy, distorted observations. The horseshoe prior applied to wavelet coefficients is expected to outperform common alternatives like L1 and L2 regularization that lack adaptive shrinkage and do not sufficiently promote sparsity.

This framework serves as an effective means to promote sparsity, resulting in well-regularized reconstructions. By jointly utilizing the sparsity inherent in wavelet transforms and the noise-reducing capabilities of the horseshoe prior, we aim to overcome the challenges posed by distorted and noisy observations. This motivates our approach of using horseshoe regularization on Starlet wavelet representations of gravitationally lensed galaxy images. The aim is to demonstrate the effectiveness of this framework for ill-posed inverse problems where key features must be recovered from incomplete, ambiguous data. Reconstructing obscured distant galaxies is an important testbed for this statistical technique.

Overview of The Chapters

Chapter 1 provides the background on this ill-posed inverse problem and the need for regularization techniques like sparsity-inducing priors.

Chapter 2 explores the statistical foundations of this work, including Bayesian inference, priors, regularization, and sparse modeling. The horseshoe prior and its sparsity-inducing properties are explained in detail, highlighting its noise suppression and feature retention abilities.

Chapter 3 covers the physics of gravitational lensing, explaining how galaxy images are warped by foreground masses. Key concepts like the lens equation and image distortions are introduced.

Chapter 4 introduces the Starlet wavelet transform and its effectiveness as a sparsifying dictionary for galaxy images. The wavelet decomposition and its multiscale analysis are described, emphasizing the sparsity induced.

Chapter 5 presents the full methodology for galaxy image reconstruction using the horseshoe-Starlet framework. The probabilistic model, Hamiltonian Monte Carlo inference, and the implementations are covered.

Chapter 6 applies the approach to simulated galaxy images, demonstrating its performance both visually and quantitatively using metrics like mean squared error. Comparisons to alternative regularization techniques validate the benefits of the horseshoe-Starlet framework. It concludes by discussing future potential extensions leveraging ever-advancing computational methods.

Chapter 1

Background

Gravitational lensing is a phenomenon in which the path of light rays from a distant galaxy is bent by the gravitational field of a foreground galaxy or cluster of galaxies. This effect was first predicted by Albert Einstein's theory of general relativity in 1915 and has since been observed and studied extensively in astrophysics. One important application of gravitational lensing is in the study of dark matter, a form of matter that does not emit, absorb, or reflect light, but can be detected through its gravitational effects on visible matter. Furthermore, Gravitational lensing is a tool that can yield accurate calculations of the Universe's expansion speed, also known as the Hubble constant. This is a crucial metric, especially in light of the ongoing cosmology crisis, which refers to a significant discrepancy in the measurements of this parameter obtained from different sources [71, 43]. This is also a potent instrument capable of charting the internal arrangement of matter within individual lens galaxies, providing priceless insights into the nature of the elusive dark matter particle [31].

The first application of gravitational lensing is akin to using a natural telescope. Massive cosmic structures like galaxies or clusters of galaxies bend the light from distant objects, magnifying and intensifying it. This gravitational magnification allows astronomers to peer further into the universe than conventional telescopes would permit, observing background galaxies that are otherwise beyond the reach of current observational technology. By acting as a natural magnifying lens, gravitational lensing enables the detailed study of the morphology, luminosity, and spectral signatures of remote celestial sources, thus extending the frontiers of our cosmic exploration. The second utility of gravitational lensing lies in its ability to infer the mass of the lensing objects, such as a foreground galaxy. The pattern and extent of the light bending are contingent upon the mass of the galaxy that is causing the lensing effect. By scrutinizing these light distortions, astronomers can deduce the mass distribution of the lensing galaxy. This technique is particularly insightful because it is independent of the traditional observational methods that rely on the light emitted by the objects themselves. Gravitational lensing thereby provides a unique and indispensable tool for estimating the distribution of both visible and dark matter within galaxies, which is crucial for advancing our understanding of the structure and composition of the universe.

By observing the distortion of light from background galaxies caused by gravitational lensing, astronomers can infer the distribution of dark matter in the intervening foreground galaxies or galaxy clusters. Furthermore, gravitational lensing can also provide insights into the large-scale structure of the Universe, as the distribution of foreground galaxies or galaxy clusters can reveal the underlying cosmic web of filaments and voids. A method for analyzing gravitational lens images when the source light distribution is pixelized. The method is suitable for high-resolution of a multiply-imaged extended source. The authors show that the step of inverting the image to obtain the deconvolved pixelized source light distribution, and the uncertainties, is a linear one. This means that the only parameters of the non-linear problem are those required to model the mass distribution. The method is extended in a straightforward way to include linear regularization. Applying the method to simulated Einstein ring images and demonstrate the effectiveness of the inversion for both the unregularized and regularized cases [70].

Chapter 2

Statistical context

Chapter 2 provides the statistical context for the inverse problem approach taken in this thesis. The chapter begins with an introduction to inverse problems, including the concepts of ill-posed and well-conditioned inverse problems. It then presents relevant probability theory and Bayesian inference methods, with a focus on Bayes' theorem, likelihood functions, and maximum likelihood estimation. The chapter concludes with a discussion of regularization techniques and prior specification, highlighting the use of Laplace, Gaussian, and horseshoe priors. Throughout the chapter, the aim is to provide a statistical foundation for the inverse problem methodology employed in later analyses. Key concepts are succinctly explained and directly linked to their relevance to the thesis.

2.1. Inverse problems

This section delves into the fundamental concept of inverse problems, essential in various scientific disciplines, involving the inference of hidden variables from observed outcomes. It discusses how Bayesian inference helps address the inherent uncertainties and complexities, presenting a versatile tool for handling ambiguous situations where conventional solutions may not suffice.

2.1.1. Concept of an inverse problem

The concept of an inverse problem is a fundamental aspect of various scientific disciplines. It involves deducing the underlying hidden variables that control observed outcomes. Inverse problems aim to uncover causes based on effects. Mathematically, this means reversing an equation or system to find the inputs that correspond to observed results. However, these problems are often complex and may lack unique solutions.

For example, in medical imaging, inverse problems focus on reconstructing internal structures from external measurements [55]. In seismology, researchers decipher Earth's properties from recorded seismic waves. Solving these problems demands advanced math, regularization methods, and prior knowledge. Bayesian inference is a useful approach to handle uncertainty in both inputs and outputs.

2.1.2. Introduction to inverse problems

In many scientific problems, accurate simulators can produce synthetic observations from a set of hidden variables that define the unobserved properties of the system under study. For example, given the mass, air resistance, shape, and length of the string of a pendulum (which are all hidden variables), a simulator can predict the position of the pendulum at time t. If measurements of these positions have been recorded (observed data) for a specific pendulum, then the simulator could be used to infer the hidden variables (mass, air resistance, shape, and length of the string).

These problems are referred to as "inverse" problems in the sense that by having the noisy output of a simulator (data) we would like to infer its input (latent variables). For a specific value of these hidden variables, it is trivial to run the simulator to predict the observed variables ("forward" model). But because of noise and other effects that result in a loss of data, there is often no inverse simulation algorithm, making this a challenging task. Solving inverse problems involves determining which hidden parameters, when propagated through the forward model, best reproduce the data.

For example, in astrophysics, the observed brightness and distorted shape of a background galaxy due to gravitational lensing depends on its true undistorted morphology. By modeling the process of light propagating through the gravitational potential, the aim is to infer the true image of the background source that has given rise to the observed distorted images.

In practice, noisy observations introduce uncertainty in the underlying model parameters. So even with an accurate forward model, we may only determine a likely range of values rather than a single solution. Assumptions about the physical processes also bring inherent uncertainties. Drawing meaningful inferences requires grappling with these issues through a formal probabilistic framework.

Inverse problems are inherently probabilistic, requiring statistical approaches like Bayesian inference to handle the uncertainties. There are, however, two general classes of inverse problems: well-posed and ill-posed problems. In summary, If the observations contain enough information to highly constrain the model parameters they are considered well-posed. But if there is simply not enough information in the data to meaningfully constrain a large number of hidden variables, the problem is considered ill-posed. In effect, the analysis of an ill-posed problem with only the given data and without any additional assumptions results in a large range of solutions, including solutions that are highly unphysical and unrealistic (e.g., a negative mass).

2.1.3. Ill-posedness and well-conditioning

A *well-posed* problem is a mathematical problem that satisfies three criteria:

(1) Existence: A solution exists for the problem.
(2) Uniqueness: The solution is unique and not ambiguous.
(3) Stability: The solution should vary continuously with changes in the input data or parameters.

Consider the canonical case where we have observations y belonging to Hilbert space $\mathcal{H}_1$. These observations can be related to parameters of interest x residing in another Hilbert space $\mathcal{H}_2$ through a well-defined forward model $F : \mathcal{D}(F) \subseteq \mathcal{H}_2 \to \mathcal{H}_1$, where $\mathcal{D}(\cdot)$ denotes the domain of F. Additionally, we define the range of F as $\mathcal{R}(\mathcal{H}_1)$. Considering this setup, the analysis of a well-posed problem within this framework can provide valuable insights into the intrinsic relationships between the observations and the parameters, helping to elucidate the underlying structure of the system being studied. This, in turn, facilitates a deeper understanding and potentially more accurate predictions and interventions in various applied settings.

An *ill-posed* problem (or *ill-posedness*) is the opposite. It fails to satisfy one or more of the criteria for a *well-posed* problem. This means that the problem might not have a solution, or if it does, the solution might not be unique or stable with respect to variations in the data or parameters.

Here, we consider bounded linear operators F, for which (i) requires F is surjective, (ii) requires F is injective, and (iii) requires the inverse of F is bounded. We refer to ill-posedness in the sense described above, though other definitions exist.

Another key concept is conditioning, which considers how a small variation in the input relates to the output variation. For an operator F, the relative condition number is:

$$\lim_{\delta y \to 0} \sup \frac{||F(y+\delta y) - F(y)||}{||F(y)||} \Big/ \frac{||\delta y||}{||y||} \tag{2.1.1}$$

Large condition numbers imply small input perturbations δy yield large output perturbations, making F ill-conditioned. Small condition numbers indicate well-posed F. Note ill-conditioning and ill-posedness are often related but distinct concepts.

While there's a strong correlation between ill-posedness and ill-conditioning, they aren't always synonymous. An *ill-posed* problem lacks a unique, stable solution, often due to the insufficiency or noisy nature of the available data. This can lead to multiple valid solutions or a solution that's highly sensitive to input variations. On the other hand, *ill-conditioning* relates to the numerical stability of solving a problem. It's about how susceptible a problem is to computational errors caused by limited precision or rounding in numerical calculations. An ill-conditioned problem might have a unique and stable solution, but the calculations to find that solution can magnify even minor errors.

While $ill-posed$ problems often arise from $ill-conditioned$ situations, and vice versa, it's entirely possible to encounter situations where an $ill-conditioned$ problem still possesses a unique and meaningful solution. Similarly, an *ill-posed* problem might involve well-conditioned computations but still lack a definitive solution due to the nature of the problem or data.

2.2. Probability theory and Bayesian Inference

Probability theory provides the mathematical foundation for logical reasoning and underlies much of scientific analysis today, including Bayesian inference and frequentist statistics [34]. This becomes evident when recognizing that many scientific problems are ill-posed and lead to the inference of complex latent spaces. As such, a probabilistic approach is well-suited to tackle these challenges. In a sense, the theory of probability is isomorphic to the consistent axioms of plausible reasoning [19].

Bayesian methodology is a prominent subdomain of probability theory widely used in astrophysics and other sciences. Bayesian inference offers a principled statistical framework for quantifying uncertainty in scientific measurements based on observed data and priors. In this paradigm, the resulting probabilities directly describe the degree of belief in a proposition. Such inferences are principled and interpretable.

In contrast, frequentist approaches rely fundamentally on repeatable outcomes, only recovering direct plausibility statements in the limit of infinite realizations. In cosmology and other sciences where few realizations may be observed, frequentist reasoning is more limited.

2.2.1. Bayes' Theorem

The prototypical inverse problem takes the following form: given observations y of an underlying quantity of interest x, which may be corrupted, for example, by instrumental noise; provided with such data, under an assumed forward model M that relates y to x, and incorporating certain prior assumptions about the nature of x, what inferences can be drawn about x, ?. This kind of inverse problem can be succinctly expressed using Bayes' theorem:

$$P(x|y,M) = \frac{P(y|x,M) \cdot P(x)}{P(y|M)} \tag{2.2.1}$$

Here, $P(A|B)$ represents the conditional probability of event A occurring given that B is known. Equation 2.2.1 is derived from fundamental rules of probabilities, where $P(A) + P(\neg A) = 1$ and $P(A,B) = P(A|B) \cdot P(B)$. The left-hand side of the equation is colloquially known as the posterior distribution, a pivotal concept in many aspects of Bayesian inference. It is the product of the likelihood distribution $L(x) = P(y|x,M)$ which encodes data fidelity, the prior distribution $P(x)$ which encodes the *a priori* assumptions about x, and the a renormalization factor, $z = P(y|M) = \int dx\, P(y|x,M)$, known as Bayesian evidence, which is useful for model comparison [2].

So far, we have discussed how Bayes' theorem enables us to reinterpret the conditional probability, representing the plausibility of parameter configuration x given model M, in view of observable data y. This process, often termed parameter inference, deals with probabilistic entities like $P(x|y,M)$. Alternatively, one might be fundamentally interested in the probabilistic quantity $P(M|y)$, directly assessing the plausibility of a model M in light of observations y [19]; [50].

2.2.2. Likelihood

In this section, we delve into the likelihood distribution, denoted as $L(x) = P(y|x,M)$. This component is a conditional probability that quantifies the likelihood of observing data y under a given parameter configuration x within model M.

The reason that observed data should be described by a probability distribution is the presence of stochastic noise in the data. This noise is often additive, giving rise to the following data-generating process

$$D = M(x) + n.$$

Here D is the observed data, M is the forward model which is a function of the latent variable of interest x, and n is a vector of additive noise.

The probability of a given value of x that is under consideration is then equal to the probability that $D - M(x)$ is generated by noise. This means that the probability density of a specific realization of noise, $N(n)$, also known as the noise model, should be known to quantify the likelihood. Therefore in the case of dealing with additive noise and when the probability density function of noise, N, is known, the likelihood could be evaluated as

$$P(D|x) = N(D - M(x)).$$

For many problems, instrumental noise is the main stochastic source of uncertainty in the data. Often, this noise is additive. The data generating process could then be written as

$$D = y + N$$

2.2.2.1. The Multivariate Gaussian Likelihood

An argument rooted in central limit theory often justifies the Gaussian nature of $n \sim N(\mu, \Sigma)$, leading to the formulation of the likelihood distribution as the exponential of the Mahalanobis distance (Mahalanobis, 1936):

$$L(x) = N(y - \tilde{y}, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}[y - \tilde{y} - \mu]^T \Sigma^{-1} [y - \tilde{y} - \mu]\right),$$

where y is a k-dimensional vector of observations, μ is the mean vector, and Σ is the covariance matrix. The covariance matrix Σ captures both the variance and inter-dependence among individual observations y_i for index $i \in \{1, 2, \ldots, k\}$.

For computational simplicity, it's often more convenient to work with the natural logarithm of the likelihood. In this case, the natural logarithm of $L(x)$ is given by:

$$\ln L(x) = -\frac{k}{2}\ln(2\pi) - \frac{1}{4}\ln|\Sigma| - \frac{1}{2}[y - \tilde{y} - \mu]^T \Sigma^{-1} [y - \tilde{y} - \mu],$$

When the covariance matrix Σ is diagonal (i.e., the observations y are taken to be independent), with diagonal elements denoted as $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$, the logarithm of the likelihood simplifies to:

$$\ln L(x) = -\frac{k}{2}\ln(2\pi) - \frac{1}{4}\sum_{i=1}^{k}\ln(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - \tilde{y}_i - \mu_i)^2}{\sigma_i^2},$$

Moreover, for independent variables (diagonal covariance matrix), the log-likelihood becomes analogous to the squared loss. By minimizing the squared loss (ordinary least squares), one implicitly retrieves the maximum likelihood solution, given the implicit uniform prior.

The Gauss-Markov theorem establishes that both ordinary least squares (OLS) and maximum likelihood estimation are the minimum variance estimators within the class of linear unbiased estimators, rendering them 'optimal' in a basic sense. However, for seriously ill-posed or ill-conditioned inverse problems, adopting a uniform prior (resulting in constant regularization over a potentially infinite domain) is suboptimal. Such problems often have insufficient data quality to effectively constrain the posterior, necessitating the introduction of weakly informative priors to recover more desirable estimators.

2.2.2.2. The Maximum Likelihood Estimation

In the context of Bayesian statistics, maximum likelihood estimation (MLE) is a method used to determine the parameter configuration x that maximizes the likelihood function $L(x)$. This estimation approach aims to find the parameter values that make the observed data y most probable under the given model M. Mathematically, MLE can be represented as:

$$\hat{x}_{\text{MLE}} = \arg\max_x L(x) = \arg\max_x P(y \mid x, M)$$

Here, $\hat{x}_{\text{MLE}}$ represents the estimated parameter configuration that maximizes the likelihood.

For the multivariate Gaussian likelihood, where the likelihood function is given by:

$$L(x) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}[y - \tilde{y} - \mu]^T \Sigma^{-1} [y - \tilde{y} - \mu]\right)$$

the maximum likelihood estimation involves finding the parameter values x that lead to the highest possible value of $L(x)$. This corresponds to finding the values of x that best explain the observed data y within the model M.

In the case of independent variables with a diagonal covariance matrix Σ, the log-likelihood simplifies to:

$$\ln L(x) = -\frac{k}{2}\ln(2\pi) - \frac{1}{4}\sum_{i=1}^{k}\ln(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - \tilde{y}_i - \mu_i)^2}{\sigma_i^2}$$

The MLE process involves finding the parameter values x that maximize this log-likelihood expression. This entails adjusting the parameters x in such a way that the terms in the equation contribute to a larger overall value of $\ln L(x)$, indicating a better fit between the model's predictions and the observed data.

2.3. Regularization and Prior specification

In the framework of Bayesian statistics, regularization takes the form of prior distributions that encapsulate our beliefs about the underlying solution.

2.3.1. Regularization

In many cases, the inverse problem we encounter is ill-posed or ill-conditioned. To tackle such challenges, we incorporate prior knowledge about the problem, effectively stabilizing the process of inversion. These stabilizing terms are commonly referred to as 'regularization functionals' in the field of signal processing, analogous to the prior functions in Bayesian theory [2, 57]. In this thesis, we often use these terms interchangeably, although they originate from different perspectives. It's important to note that our approach to regularization in this thesis is application-oriented. Beyond the scope of this work, regularization is a broader mathematical concept.

In general terms, regularization for an ill-posed or ill-conditioned inverse problem aims to introduce superior regularity properties. For instance, regularization may lead to a lower condition number in the problem (as defined in section 2.1). It may also constrain the solution space to a smaller subset, denoted as $H_{01} \subset H_1$, to promote convergence, prevent discontinuities, or eliminate redundant degeneracies in the solution space. In this thesis, we mainly focus on regularization functions that restrict or adjust the solution space H_1 to favor solutions exhibiting a known or assumed structure in x. Our primary approach involves using regularization functions that encourage sparsity, a recurring theme throughout this work.

Returning to the context of astronomical image reconstruction: suppose we possess prior knowledge that the intensity values in an astronomical image must be positive, and we observe neighboring regions of the sky, finding that they exhibit an average intensity I with an approximate distribution (e.g., Gaussian). In this scenario, it's reasonable to utilize this prior knowledge to adjust the degenerate solution space, favoring intensity maps with values akin to those observed in neighboring regions. This regularization is further enhanced by considering that astronomical sources are unlikely to exhibit intensity values drastically different from their surroundings. Hence, we include additional constraints on the feasible set of solutions, further reducing the degeneracy within the solution space.

A commonly adopted form of regularization in the literature is the use of sparsity-promoting functions. These functions bias or restrict the solution space towards solutions that can be represented with the fewest coefficients in a given dictionary [25, 29]. A dictionary where the solution is assumed to be sparse is known as a sparsifying dictionary. This use of sparsity to enhance regularity aligns with the mathematical interpretation of Occam's Razor or the principle of parsimony [7]. In practice, dictionaries localized solely in the temporal or real domains, such as the Fourier or Dirac bases, respectively, are often ineffective at capturing sharp, abrupt features. Throughout this thesis, we will employ wavelet dictionaries due to their inherent localization in both the temporal and real domains [35]; [36]. Wavelets are highly effective for a wide range of physical signals as they allow sparse representations [21, 4, 56, 5, 63, 14, 37, 72, 47, 39].

2.3.2. The Priors

2.3.2.1. The Laplace prior (ℓ_1 Regularization)

The Laplace prior, also known as $\ell_1$egularization, is a continuous probability distribution commonly used as a sparsity-inducing prior in Bayesian analysis [46, 26, 30]. Its probability density function is given by:

$$p(\beta|\lambda) = \frac{\lambda}{2}\exp(-\lambda|\beta|) \tag{2.3.1}$$

where β is the parameter or coefficient being estimated, and $\lambda > 0$ is a regularization hyperparameter controlling the spread of the distribution. For a regression model with p predictors, independent Laplace priors can be placed on each coefficient β_j:

$$p(\beta|\lambda) = \prod_{j=1}^{p} \frac{\lambda}{2}\exp(-\lambda|\beta_j|) \tag{2.3.2}$$

The logarithm of this prior density is then:

$$\log p(\beta|\lambda) = -\lambda \sum_{j=1}^{p} |\beta_j| + C \tag{2.3.3}$$

where C is a normalization constant. Combining this log prior with the log-likelihood $\log p(\mathbf{y}|\mathbf{X},\beta)$ via Bayes' theorem yields the posterior distribution:

$$\log p(\beta|\mathbf{y},\mathbf{X},\lambda) \propto \log p(\mathbf{y}|\mathbf{X},\beta) - \lambda \sum_{j=1}^{p} |\beta_j| + C \tag{2.3.4}$$

The Laplace prior's sharp peak at zero and heavy tails promote sparsity by shrinking small coefficient values to zero while retaining a few large coefficients [15, 30]. This enables variable selection and sparse regression in a Bayesian manner. The regularization strength λ controls the sparsity level. Various techniques exist for learning λ, such as cross-validation or maximum marginal likelihood estimation [26]. Overall, the Laplace prior provides a principled Bayesian approach for achieving sparsity. The Laplace prior's sharp peak at zero and heavy tails promote sparsity by shrinking small coefficient values to zero while retaining a few large coefficients [15, 30]. This enables variable selection and sparse regression in a Bayesian manner. The regularization strength λ controls the sparsity level. Various techniques exist for learning λ, such as cross-validation or maximum marginal likelihood estimation [67, 26]. Overall, the Laplace prior provides a principled Bayesian approach for achieving sparsity.

2.3.2.2. The Gaussian prior (ℓ_2 Regularization)

The Gaussian or normal prior also known as ℓ_2 Regularization, is another common continuous prior distribution used in Bayesian analysis:

$$p(\beta|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\beta^2}{2\sigma^2}\right) \tag{2.3.5}$$

β is the parameter being estimated and $\sigma > 0$ is a scale parameter controlling the variance [67, 23]. For a model with p predictors, independent Gaussian priors can be placed on each coefficient β_j:

$$p(\beta|\sigma) = \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\beta_j^2}{2\sigma^2}\right) \tag{2.3.6}$$

The log prior density is:

$$\log p(\beta|\sigma) = -\sum_{j=1}^{p} \frac{\beta_j^2}{2\sigma^2} + C \tag{2.3.7}$$

Combining this with the log-likelihood via Bayes' theorem gives:

$$\log p(\beta|\mathbf{y},\mathbf{X},\sigma) \propto \log p(\mathbf{y}|\mathbf{X},\beta) - \sum_{j=1}^{p} \frac{\beta_j^2}{2\sigma^2} + C \tag{2.3.8}$$

Unlike the Laplace prior, the Gaussian prior does not induce rigorous sparsity. Its Gaussian shape assumes parameters are centered at zero with normally distributed deviations

[42]. The variance σ^2 controls the spread of the distribution. Various techniques like cross-validation or empirical Bayes can be used to learn σ [15, 13]. Overall, the Gaussian prior is a convenient choice when parameters are expected to be smooth and lack sparsity.

2.3.2.3. The Horseshoe Prior

Horseshoe prior was first introduced by Carvalho, Polson, and Scott in their 2009 paper [15]. This prior is known for its ability to produce sparse estimates, such as estimates that set many of the regression coefficients to zero. This is achieved by placing a heavy-tailed distribution on the coefficients, that allow for large variability in their magnitudes, while also shrinking many of them towards zero. It would be done by assuming a global scale parameter on the regression coefficients to control their shrinkage towards zero. The horseshoe estimator assumes that the coefficients are drawn from a normal distribution with a mean zero and a scale parameter that is allowed to vary across different coefficients. The scale parameter is itself drawn from a half-Cauchy distribution, which has heavy tails and allows for the possibility of large-scale parameters.

The horseshoe prior can be represented as follows. Let Y be a vector of observations, X be a matrix of predictors, and β be a vector of regression coefficients. The linear regression model can then be written as:

$$Y = X\beta + \epsilon, \tag{2.3.9}$$

where ϵ is a vector of noise (error) [16]. The horseshoe estimator places a prior distribution on $\boldsymbol{\beta}$

$$\begin{aligned}
&\beta_i|\lambda_i, \tau \sim \mathcal{N}(0, \tau^2\lambda_i^2),\\
&\lambda_i \sim C^+(0,\gamma),\\
&\gamma \sim C^+(0, \sigma),
\end{aligned}$$

where τ is a global shrinkage parameter that controls the overall amount of shrinkage applied to the coefficient and it does not depend on index i, λ_i is a local shrinkage parameter that the tail of the distribution exhibits a slower decay compared to an exponential rate, which prevents significant shrinkage of β_i. The local shrinkage parameter ,λ_i, is drawn from a half-Cauchy distribution with scale parameter γ.

Besides, the half-Cauchy distribution has the following density function:

$$p(\lambda_i) = \frac{2\gamma}{\pi(\lambda_i^2 + \gamma^2)}, \tag{2.3.10}$$

when γ is the scale parameter.

There is no closed-form density function for the horseshoe prior since the distribution is defined in terms of the scale and global shrinkage parameters. However, there are upper and lower bounds for the horseshoe estimator that are given by

$$If \lim_{\beta \to 0} p(\beta) = \infty$$

$$If \beta \neq 0$$

Then,

$$\frac{K}{2}\log(1+\frac{4}{\beta^2}) < p(\beta) < Klog(1+\frac{2}{\beta^2}) \tag{2.3.11}$$

when $K = 1/(2\pi^3)^{\frac{1}{2}}$. Here are Figure 2.1 and Figure 2.2 which shows different priors and its tails [16] .

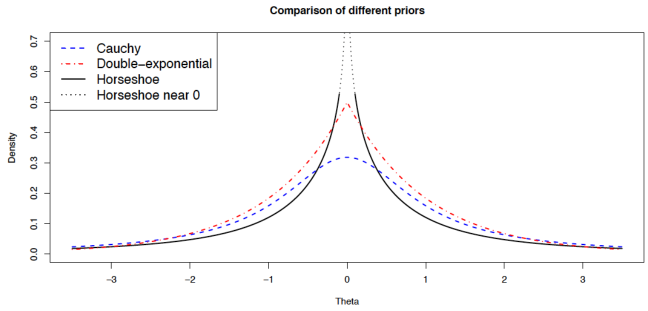


Fig. 2.1. Comparing of different priors, adapted from [16]

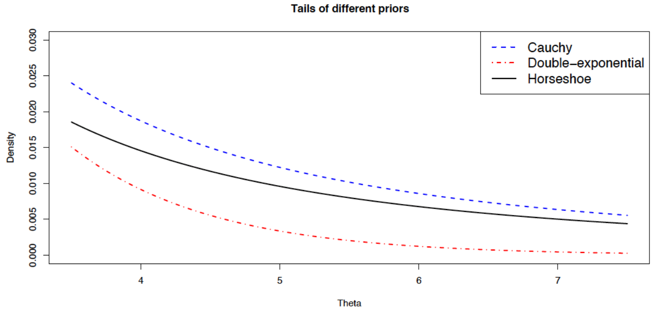


Fig. 2.2. Comparing of different priors' tail, adapted from [16]

2.3.2.4. Advantages of using the horseshoe prior for sparsity

The horseshoe prior [16]; offers several notable advantages that make it well-suited for inducing sparsity in statistical models [10].

- The horseshoe prior exhibits a sharper peak at zero and heavier tails compared to alternatives like the Laplace prior. Therefore, it is able to shrink small signals to zero while leaving large signals relatively untouched This is due to its "global-local" scale mixture structure, where the global parameter adapts to the overall sparsity and the local parameters identify signals.

- the horseshoe adaptively adjusts the degree of sparsity via its local-global shrinkage. The global parameter modulates the overall sparsity, while the local parameters tune the sparsity specific to each data point. This flexibility suits the horseshoe well to data where sparsity levels vary.
- the horseshoe prior has theoretical optimality properties related to minimax concave penalization that make it robust for sparse estimation tasks.
- The horseshoe estimator has been shown to have good performance for prediction and estimation in sparse linear regression models. It provides robustness against signals of varying magnitudes.
- The horseshoe prior has heavy tails which makes it robust to outliers compared to methods like the Lasso. The Lasso shrinks large coefficients which can lead to bias. In cases where outliers or extreme data points are present, the horseshoe prior demonstrates robustness. It can effectively mitigate the influence of outliers on the estimation process, contributing to more robust and reliable results.
- Various extensions of the horseshoe have been developed for generalized linear models and other likelihood models beyond Gaussian data. This provides flexibility compared to methods like Lasso which were designed for linear models. The horseshoe is computationally efficient compared to Bayesian spike-and-slab models which require exploring a huge model space.

Chapter 3

Cosmological context

In Section three of this thesis, we explore gravitational lensing. This phenomenon offers crucial insights into the structure of the Universe. By understanding the theoretical foundations of gravitational lensing and its significance in astrophysics, we lay the groundwork for our proposed sparse Bayesian approach to reconstructing background galaxies.

3.1. Gravitational lensing

Within the framework of General Relativity, gravity is interpreted as an outcome of space-time curvature caused by the presence of massive objects. A consequential prediction of this theory is the deflection of light as it propagates through curved space-time. This prediction was experimentally verified for the first time in 1919 during Eddington's solar eclipse expedition, where the bending of starlight around the sun confirmed Einstein's theory [24]. Since then, gravitational lensing has matured into a well-established cosmological tool, widely employed to explore and comprehend the structure and dynamics of the universe. One approach to analyzing gravitational lens images involves pixelizing the source light distribution, particularly suitable for high-resolution, high signal-to-noise ratio data of multiply imaged extended sources. Demonstrating linearity in the step of inverting the image to obtain the deconvolved pixelized source light distribution, along with its associated uncertainties, significantly simplifies parameter estimation [70]. Another pixelated method is employed to model both the lens potential and the source-intensity distribution concurrently [65].

3.1.1. The basics of gravitational lensing

Gravitational lensing is a phenomenon in which the path of light from a distant galaxy is bent by the gravitational field of a foreground galaxy or cluster of galaxies. In the realm of gravitational lensing, the influential entity responsible for bending light is referred to as the "lens." This lens can take two forms: either point-like, exemplified by a star, or extended, as observed in the case of a galaxy or a galaxy cluster. Conversely, the luminous entity whose light undergoes deflection is termed the "source". The source may be a background galaxy or a distant quasar. Gravitational lensing causes the appearance of background galaxies to be distorted and can result in arcs or multiple images of the same galaxy as shown in Figure 3.2.

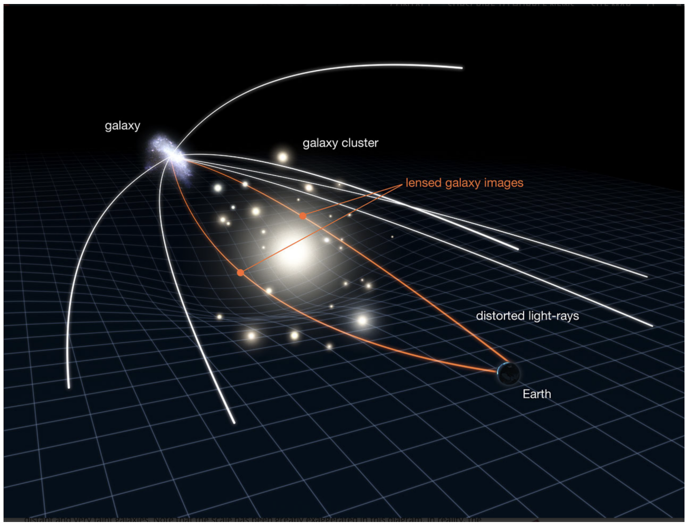


Fig. 3.1. Gravitational lensing effect, adapted from esahubble.org by NASA, ESA & L. Calçada

There are two categories of gravitational lensing as we could see in Figure 3.1:

1) Strong lensing, which results in multiple images of background sources, often in forms of extended arcs or Einstein rings.

2) Weak lensing, which causes subtle distortions (e.g. minor stretching) that require a substantial amount of observational data and multiple source to extract meaningful information through complex statistical frameworks [6].

In this work, we only study the case of strong gravitational lensing and focus on the inference of the morphology of background sources given the true foregound lensing distortion.

Redshift is another important concept in gravitational lensing. Redshift can also occur due to the gravitational influence of massive objects (gravitational redshift) and the expansion of space (cosmological redshift). It is a measure of the change in wavelength of light from a distant galaxy as it travels through space and is affected by the expanding Universe. The redshift of a galaxy can be used to determine its distance from Earth and its velocity relative to the observer.

3.1.2. Deflection angle

The idea that photons may be deflected by a mass located along their path was originally expressed by Newton in the context of the Corpuscular Theory of Light (Newton, I., 1704),

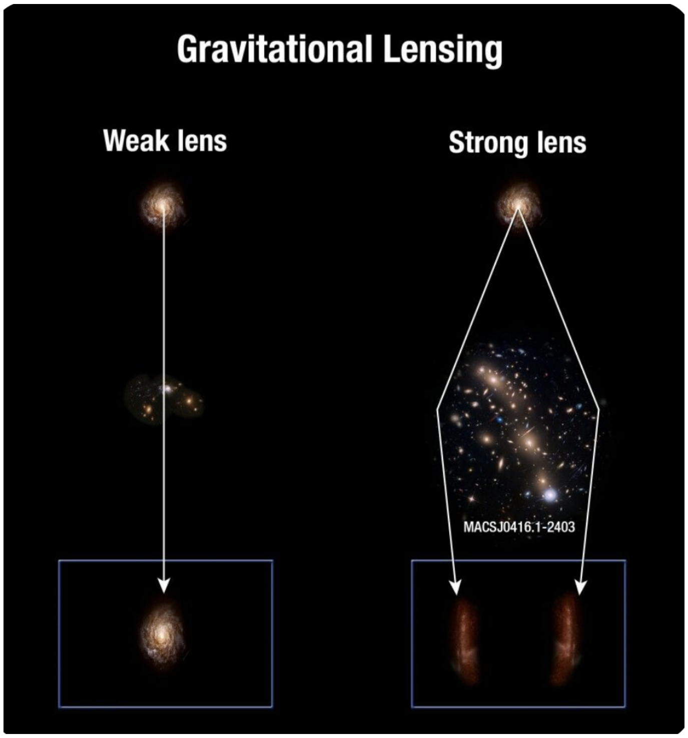


Fig. 3.2. Strong and Week Gravitational lensing, adapted from frontierfields.org by A. Feild (STScI)

and it was later confirmed and explained by Einstein's theory of General Relativity. As per this theory, the characteristics of space-time are contingent on its energy-matter distribution. Consequently, the gravitational influence of masses positioned between the source of light and the observer leads to the bending of light paths. To calculate the angle of deflection for the light path, we employ an assumption that the deflection angle is very small, a valid approximation in many astrophysical scenarios. This approximation necessitates that the Newtonian gravitational potential of the lens, denoted as Φ, satisfies the condition $\Phi/c^2 \ll 1$. Alternatively, an equivalent requirement is that the size of the lens should be considerably smaller compared to the dimensions of the optical system, encompassing the observer, the lens, and the source. Within this approximation, the Minkowski metric, which characterizes the undisturbed space-time, undergoes slight modifications due to a small perturbation. Consequently, the line element can be expressed as follows:

$$ds^2 = \left(1 + \frac{2\Phi}{c^2}\right) c^2 dt^2 - \left(1 - \frac{2\Phi}{c^2}\right) (d\vec{x})^2 \tag{3.1.1}$$

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

$$c' = \frac{d\vec{x}}{dt} = c\sqrt{\frac{1+2\Phi/c^2}{1-2\Phi/c^2}} \approx c\left(1+\frac{2\Phi}{c^2}\right) \tag{3.1.2}$$

Since $\Phi \leq 0$, we have $c' \leq c$. Then, we can describe the space-time as a medium with effective refraction index

$$n = \frac{c}{c'} = \frac{1}{1+2\Phi/c^2} \approx 1 - \frac{2\Phi}{c^2} \geq 1. \tag{3.1.3}$$

Employing Fermat's principle it can be shown [52] that the deflection angle is

$$\widehat{\vec{\alpha}}(b) = \frac{2}{c^2}\int_{-\infty}^{+\infty} \nabla_{\perp}\Phi dz \tag{3.1.4}$$

Under the provided assumption, light rays their path in the direction of $\vec{e}_z$ and encounter the lens at a point with a specific impact parameter b when passing through the plane located at $z = 0$. This outcome remains applicable when the spatial scales under consideration are smaller than the distances separating the source, lens, and observer, and if the time scale is sufficiently brief to render the universe's expansion insignificant. In case of a point mass equation (3.1.4) reads

$$|\widehat{\vec{\alpha}}|(b) = \frac{4GL}{c^2 b} \tag{3.1.5}$$

where L is the mass of the lens.

Since the deflection angle of a mass L exhibits a linear dependence, when dealing with multiple lenses represented by L_i where $1 \leq i \leq N$, their collective deflection angle can be determined by adding together their respective individual deflection angles.

$$\widehat{\vec{\alpha}}(\vec{\xi}) = \sum_i \widehat{\vec{\alpha}}_i(\vec{\xi} - \vec{\xi}_i) = \frac{4G}{c^2}\sum_i L_i \frac{\vec{\xi} - \vec{\xi}_i}{|\vec{\xi} - \vec{\xi}_i|^2}. \tag{3.1.6}$$

Here $\vec{\xi}_i$ are the positions of the lenses, $\vec{\xi}$ represents the position where the deflection angle is calculated.

In the field of astrophysics, employing this *thin lens approximation* is common and applicable to most scenarios. This approximation represents the lens as a surface mass density, enabling a simplified yet effective description of the gravitational lensing phenomenon.

$$\Sigma(\vec{\xi}) = \frac{4G}{c^2}\int_{-\infty}^{+\infty} \rho(\vec{\xi}, z) dz \tag{3.1.7}$$

ρ is the three-dimensional density of the lens, $\vec{\xi}$ defines a position on the lens plane and $\vec{e}_z$ is the direction perpendicular to it. The *thin lens approximation* is applicable due to the substantial differences in distances between the observer, lens, and source compared to the physical size of the lens. Figure 3.3 illustrates the geometrical arrangement of a gravitational lensing system, where the sources are presumed to lie on a plane. This approximation allows for a simplified representation of the lens as a surface mass density, facilitating the study of gravitational lensing phenomena while maintaining the logical integrity of the system.

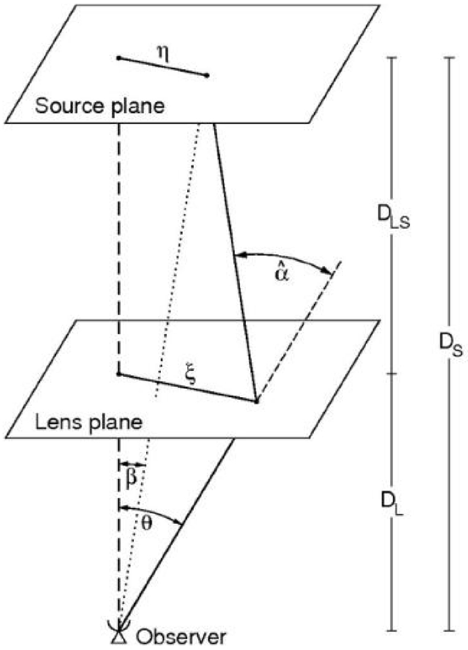


Fig. 3.3. Typical configuration of a lensing system, adapted from [51]

Under this approximation, the calculation of the deflection angle is given by:

$$\widehat{\alpha}(\vec{\xi}) = \frac{4G}{c^2} \int \frac{(\vec{\xi} - \vec{\xi'})\Sigma(\vec{\xi'})}{|\vec{\xi} - \vec{\xi'}|^2} d^2\xi' \tag{3.1.8}$$

3.1.3. Lens equation

The lens equation establishes a coordinate transformation, mapping the true position of a source in the universe (vector β) to its observed position (vector θ) in a two-dimensional plane. This transformation is pivotal in gravitational lensing, as it allows astronomers to trace the apparent distorted image back to its original alignment.

The influence of gravitational lensing relies on the relative positions and distances among the observer, lens, and source. To measure the angular diameter distances in astronomy, redshifts are an observational tool. The angular diameter distance is a way of expressing how large an object appears (its angular size) and how far away it is. We will examine a lens situated at an angular distance D_L (or equivalently, redshift z_L), which alters the path of light emitted by a source at an angular distance D_S (or redshift z_S). To measure the angular diameter distances in astronomy, redshifts are a fundamental observational tool. The angular diameter distance is a way of expressing how large an object appears (its angular size) and how far away it is.

If the true position of the source is $\vec{\beta}$, it will appear as if it was located in $\vec{\theta}$ because of the deflection angle $\widehat{\vec{\alpha}}$, as shown in Figure 3.3.

If $\vec{\theta}$, $\vec{\beta}$, $\widehat{\vec{\alpha}}$ are small, the lens equation relates the true and apparent positions of the source:

$$\vec{\theta} D_S = \vec{\beta} D_S + \widehat{\vec{\alpha}} D_{LS}. \tag{3.1.9}$$

Here D_{LS} represents the distance between the lens and the source.

Moreover, introducing the *reduced deflection angle*

$$\vec{\alpha}(\vec{\theta}) \equiv \frac{D_{LS}}{D_S} \widehat{\vec{\alpha}}(\vec{\theta}) \tag{3.1.10}$$

equation (3.1.9) would be:

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}). \tag{3.1.11}$$

Equation (3.1.9) is usually written in a dimensionless form. Considering the length scale ξ_0 on the lens plane and the corresponding length scale $\eta_0 = \xi_0 D_S / D_L$ on the source plane, we can define the two vectors

$$\vec{x} \equiv \frac{\vec{\xi}}{\xi_0}, \qquad \vec{y} \equiv \frac{\vec{\eta}}{\eta_0}$$

and the lens equation writes

$$\vec{y} = \vec{x} - \vec{\alpha}(\vec{x}) \tag{3.1.12}$$

where

$$\vec{\alpha}(\vec{x}) = \frac{D_L D_{LS}}{\xi_0 D_S} \widehat{\vec{\alpha}}(\xi_0 \vec{x}). \tag{3.1.13}$$

Gravitational lensing conserves surface brightness. In other words, it can be seen as the values of a scalar field under a coordinate transformation. The scalar field here represents the intensity of the image of the background source. A region of the source at position β now appears at a different position θ (or possibly multiple positions), but each image has the same intensity as the true source. This means that our model of gravitational lenses that have two sets of parameters (parameters η describing the mass morphology in the foreground lens causing spacetime distortions and parameters S describing the true image of background sources) are linear in the surface brightness parameters. In a simulator, if a pixel representing the true light intensity in the background is doubled, then all the intensity of all its images (in the distorted and lensed image) will be also doubled. This means that given the foreground lens parameters, μ, the physical model could be written as a linear equation

$$\vec{y}_{\text{pixels}} = L(\eta) \, \vec{S}_{\text{brightness}} \tag{3.1.14}$$

where L is a distortion matrix that needs to be calculated using a ray-tracing simulation code for the foreground parameters η.

3.1.4. Exploring Gravitational Lensing

Gravitational lensing analysis involves the intricate task of understanding the impact of massive objects on the paths of light rays, resulting in the distortion of observed images. The goal of such analysis is two-fold. Gravitational lensing analysis seeks to achieve two primary goals:

1- Inferring the undistorted image of the background source galaxy. This involves reconstructing the original appearance of the galaxy before lensing effects distorted its shape and brightness.

2- Determining the form of the distortion caused by lensing. This requires modeling the mass distribution of the foreground lens and how it deflects light rays from the background galaxy.

In this work, we aim to address the intricate task of understanding the impact of massive objects on the paths of light rays, resulting in the distortion of observed images. Our primary focus lies in the reconstruction of the distorted images of the background galaxy. By delving into the methodologies and techniques associated with this endeavor, we navigate through the challenges posed by lensing-induced distortion. It's noteworthy that throughout the thesis, we assume the correct mass distribution for the lens. This assumption provides a foundational framework for our exploration, particularly in the context of background image reconstruction.

In conclusion, Chapter 3 has provided a thorough exploration of the linear modeling framework for gravitational lensing, culminating in the formulation $\vec{y}_{\text{pixels}} = L(\eta)\,\vec{S}_{\text{brightness}}$. This model serves as a cornerstone for our subsequent analyses, allowing us to predict and understand the distortion effects on observed light with greater precision.

As we move to Chapter 4, we will introduce the starlet transform as a tool for decomposing astronomical images into their constituent layers. Furthermore, the incorporation of the horseshoe prior in our methodology will be discussed, illustrating its role in reinforcing the sparsity of our model.

Chapter 4

Starlet Wavelet transform

Recovering the original source image is an ill-posed inverse problem due to the information loss induced by lensing, limited resolution, and noise. To address this challenge, we need to infer the properties of the unlensed source based on the available distorted image. This requires parameterizing the source using a suitable model or basis representation. Common approaches include representing the source on a pixel grid or expanding it in terms of shapelet basis functions. Utilizing multiscale methodologies like the wavelet transform enables the segmentation of an image into distinct components across varying scales, making it particularly suited for the examination of astronomical data. The wavelet transform stands as a potent time-frequency analysis technique devised to address the Fourier transform's limitation regarding local capability in the time domain. It possesses crucial attributes, including strong localization characteristics in both time and frequency domains, and the ability to furnish frequency details for each sub-band of a signal. In broader terms, the most effective data breakdown results in the sparsest representation, where only a few coefficients are significantly large, while the majority are nearly zero.

4.1. Wavelet Introduction

Wavelets are a broad category of functions, characterized by their localization in both time and frequency domains, enabling efficient representation of non-stationary signals [38].

Analyzing a signal through its wavelet coefficients allows for the identification of features across different scales and positions, in contrast to Fourier analysis, which is restricted to scale.

In the context of sparse regularization for inverse problems, our interest in wavelets arises from their capacity to provide sparse representations for most natural signals. However, the choice of a specific wavelet type can significantly impact the quality of the result in sparse recovery applications. Hence, careful consideration should be given to selecting an appropriate wavelet for a given application [33, 61, 56].

In astronomy, wavelets have found broad applications spanning data filtering, deconvolution, and the detection of celestial objects like stars and galaxies, as well as the removal of artifacts like cosmic rays. Recent sparse representations such as ridgelets and curvelets have emerged for detecting anisotropic features like cosmic strings in the cosmic microwave background[60, 41].

SLITronomy, a sparsity-based method employing wavelets to characterize lensed sources and optimize parameters within an analytical lens mass profile. It demonstrates wavelets' effectiveness in reconstructing highly detailed substructures in lensed sources, especially in high-resolution images expected from future telescopes[27]. A wavelet-based method for modeling galaxy-scale strong gravitational lenses, Demonstrated on simulated Hubble Space Telescope (HST) data. It accurately captures diverse mass substructures, including dark matter subhalos and galaxy-scale multipoles [28].

The à trous wavelet transform, another type of wavelet, offers the benefit of providing a stable, uniformly distributed, and translationally invariant conversion, making it particularly suitable for applications in the analysis of astronomical data [33, 56]. Various scaling functions can be employed with the à trous wavelet transform, allowing for flexibility in adapting to different data characteristics and analysis requirements. Selection of the scaling function is determined by the unique properties of the image and the specific information sought to be extracted from it [69, 1].

In order to analyze the structure and its changes over time in an astronomical object, it is necessary to decompose the imaged object's structure into a collection of significant structural patterns (SSP) that can be effectively monitored across a series of images. This is commonly achieved by matching the structure with predetermined templates (like two-dimensional Gaussians, disks, rings, or other shapes) [66, 66]. Furthermore, permitting their parameters to fluctuate is essential. However, it is evident that to achieve a robust structural decomposition without relying on preconceived notions, the generic shape of these patterns must also be allowed to vary [41]. To enhance the method's robustness even further, the multiscale approach is expanded to include object detection, following a similar methodology to that developed for the multiscale vision model [49, 58]. In related research on object and structure detection [40, 54].

The discrete wavelet transform (DWT) is another type of the wavelets. It refines this by discretizing the scale and translation of the fundamental wavelet [3, 18]. The wavelet transform also offers frequency information of an image. The wavelet coefficients undergo rearrangement in a newly structured manner dictated by the parent-child relationship among the sub-bands. This paper [73] proposes new method to devise scanning modes that leverage direction information from high-frequency sub-bands and introduce an optimized measurement matrix with a double allocation of measurement rate.

4.1.1. Parameterizing the Background Source

We begin our exploration by discussing the parameterization of background sources and inference techniques. The choice of parameterization provides flexibility in modeling the intrinsic galactic morphology and substructures.

One common approach is to use mathematical functions that describe the intrinsic morphology of galaxies such as Gaussian or Sersic light profiles. The Sersic profile stands out as a widely used model . In this case, the intensity of the background source at location β can be computed as a simple function of the variables describing the system [68, 17, 53].

The Sersic profile provides a functional form for the intensity distribution of a galaxy, characterized by a single parameter called the Sersic index, denoted by n. This index controls the shape of the intensity profile, ranging from compact, centrally concentrated distributions for small n values to more extended, diffuse profiles for larger n values. Mathematically, the Sersic profile can be expressed as:

$$I(\beta) = I_0 \exp\left[-b_n\left(\left(\frac{r}{r_e}\right)^{1/n} - 1\right)\right] \tag{4.1.1}$$

where $I(\beta)$ represents the intensity of the background source at location β,I_0 is the intensity at the center of the galaxy, r is the distance from the center of the galaxy, r_e is the effective radius, defining the radius containing half of the total luminosity of the galaxy, and b_n is a constant that depends only on the Sersic index n.

This formulation allows us to model the brightness distribution of galaxies with varying degrees of complexity, capturing both the central concentration and outer extent of their light profiles. Such parameterization facilitates the inference process, enabling us to estimate the intrinsic properties of the background sources from their observed, lensed counterparts.

The advantage of this approach is that the highly smooth surface brightness resulting from simple functional forms acts as implicit regularizers. Therefore, regularization is effectively achieved by the choice of the functional form to represent surface brightness. Additionally, the low number of latent variables describing the system allows for easy inference in low-dimensional spaces.

However, these approaches can only represent relatively simple galaxy morphologies. Clumpy and complex structures with spiral arms or other complexities can not be represented by such low dimensional variables and simplistic functional forms.

To achieve more detailed representations, higher dimensional basis functions are used. For example, one approach is to represent the source on a pixel grid, where each pixel intensity is a parameter to be inferred. This can be computationally expensive for high-resolution images and requires well-defined priors (i.e. regularization) is high dimensional space of pixels.

An alternative is to expand the source in another basis of functions, such as shapelets, wavelets, or sparsity-promoting dictionaries. The coefficients of the basis functions become the parameters to infer. This compresses the image into fewer parameters [8, 48, 59, 20].

The advantage of these alternative high dimensional representations is that the basis functions can be somewhat optimally designed for the signal of interest so that only the combination of only a small number of them could reproduce complex signals. In this way, the Bayesian prior, instead of explicitly defining a probability density function in the high dimensional space of parameters, can be only enforced by requiring sparsity, meaning that we desire that the signal be reconstructed with only a small number of the available basis functions.

For wavelet expansion, the wavelet coefficients are the parameters of interest. The forward process applies the wavelet transform to the source image to obtain the coefficients. Fitting the model involves inferring these coefficients based on the observed lensed image.

The wavelet basis provides localization and sparsity, facilitating reconstruction. We will focus specifically on the starlet wavelet transform for its effectiveness in analyzing astronomical images. The starlet coefficients will form the parameters for our source model.

In the wavelet domain, we can apply regularization techniques and priors to solve this ill-posed inversion problem. The choice of basis functions plays a key role in the source inversion methodology.

4.1.2. Introduction to the Starlet basis

In general, a signal S, is considered sparse if most of its values are zero. Alternatively, a signal can be considered weakly sparse or compressible if only a few of its values have a large magnitude, while most of the other entries are in proximity to zero. In cases where a signal does not demonstrate sparsity, it can be converted into a sparse signal through the application of a designated data representation method. It is feasible to express a matrix signal (here called S) as a linear combination of T basic waveforms, known as elementary waveforms. It could be shown as follows

$$S = \Phi_{\alpha} = \sum_{i=1}^{T} \alpha[i]\phi_i, \tag{4.1.2}$$

where $\Phi = [\phi_1, ..., \phi_S]$, and $\alpha[i] = [S,\phi_i]$ are the decomposition coefficients of S. The Isotropic Undecimated Wavelet Transform (IUWT)[62, 64, 58], known as the starlet wavelet transform, is renowned in the field of astronomy for its excellent applicability to astronomical data. Employing a multi-scale pixel representation, wherein each pixel in the input image corresponds to a set of pixels in the multi-scale transform. This wavelet transformation is particularly adept at identifying isotropic features, explaining its efficacy in astronomical image processing, consisting mainly of isotropic or quasi-isotropic objects such as stars, galaxies, or galaxy clusters [60].

The starlet transform has several advantages over other wavelet transforms, including its fast and efficient computation, its ability to separate the noise from the image structure, and its ability to preserve the isotropy and rotational symmetry of the image features. These properties make it a popular choice for a variety of image processing tasks, including denoising, deconvolution, and feature extraction. The Starlet wavelet transform decomposes the galaxy image, S_0 of the size $N * N$ into a set of coefficients as follows:

$$S_0[k,l] = S_J[k,l] + \sum_{j=1}^{J} w_j[k,l], \tag{4.1.3}$$

where $W = \{w_1,...,w_1, S_0\}$ is a coefficient set while w_j represents the details of S_0 at scale 2^{-j} and S_J is a coarse version of the S_0. When $J = 1$ represents the finest scale and there are $J + 1$ sub-band arrays with the same dimensions as S. The process of decomposition is accomplished through the utilization of a filter bank. When δ is a Dirac function and h_{1D} is 1D filter and the tensor product of two h_{1D} is h_{2D}. The transition from one level to the subsequent level is achieved through the utilization of the algorithm as followes:

$$\begin{aligned}
&C_{j+1}[k,l] = \sum_m \sum_n h_{1D}[m] h_{1D}[n] c_j[k + 2^j m, l + 2^j n],\\
&w_{j+1}[k,l] = C_j[k,l] - C_{j+1}[k,l],
\end{aligned}$$

Assuming a B_3-spline for the scaling function,

$$\begin{aligned}
&\phi(x) = B_3(x) =\\
&(\frac{1}{12}|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3),
\end{aligned}$$

when

$$\begin{aligned}
&h_{1D} = (\frac{1}{16}\frac{1}{4}\frac{3}{8}\frac{1}{4}\frac{1}{16}),\\
&h_{2D} = (\frac{1}{16}\frac{1}{4}\frac{3}{8}\frac{1}{4}\frac{1}{16}) \begin{pmatrix} \frac{1}{16} \\ \frac{1}{4} \\ \frac{3}{8} \\ \frac{1}{4} \\ \frac{1}{16} \end{pmatrix} = \begin{pmatrix} \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{3}{256} & \frac{3}{32} & \frac{9}{64} & \frac{3}{32} & \frac{3}{128} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \end{pmatrix}
\end{aligned}$$

Regarding the Reconstruction of the starlet, the original map can be accurately replicated by combining all wavelet scales through a straightforward co-addition process. In Figure 4.1, we demonstrate the decomposition of the original lensed galaxy and its coefficients across scales $j = 1$ to 6, providing insight into the hierarchical structure and features of the galaxy.

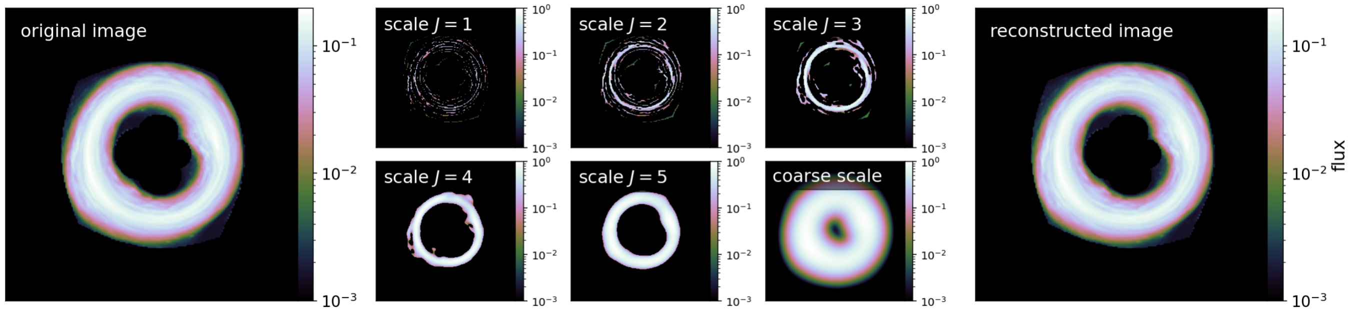


Fig. 4.1. decomposition of the original lensed galaxy and its Coefficients from scales j=1 to 6

4.1.2.1. The starlet basis promotes sparsity in different scales

The starlet transform [58] provides a multiscale wavelet decomposition well-suited for analyzing astronomical images. It returns J new images $\mathbf{x}_j$ of equal size to the original image $\mathbf{x}$, with each $\mathbf{x}_j$ corresponding to convolution with a filter amplifying features at scale 2^j pixels:

$$\mathbf{x}_j = h_j * \mathbf{x} \tag{4.1.4}$$

where h_j is the wavelet kernel at scale j. The final smoothed image represents coarse-grained averaging. The number of scales depends on the image size n_{pix}:

$$J_{\max} = \lfloor \log_2 n_{\mathrm{pix}} \rfloor \tag{4.1.5}$$

The transform operator Φ^T produces coefficients $\boldsymbol{\alpha}_{\mathbf{x}} = \Phi^T \mathbf{x}$. The inverse Φ reconstructs $\mathbf{x} = \Phi \boldsymbol{\alpha}_{\mathbf{x}}$.

In Chapter 4, we have examined the role of the transform operator and its coefficients in image reconstruction. Building on this foundation, Chapter 5 will introduce the practical application of these principles to astronomical imaging. We will specifically discuss how the starlet transform can be used to model the background of astronomical images, thereby allowing for a clearer analysis of lensed galaxies. Additionally, we will describe the use of the horseshoe prior in Bayesian inference to promote sparsity, which is expected to improve the accuracy of our image reconstructions.

Chapter 5

Methodology

This chapter provides a comprehensive overview of the key methodologies central to our study - Bayesian computation using Markov Chain Monte Carlo (MCMC) sampling and the synergistic combination of the Horseshoe prior with the Starlet wavelet transform. MCMC methods empower us to effectively sample from complex posterior distributions. This enables rigorous Bayesian analysis when deriving posterior estimates and uncertainties for model parameters is intractable. By constructing Markov chains that converge to the target posterior, MCMC generates representative samples that can estimate expectations of interest.

A major contribution of this work is showcasing the effectiveness of combining the Horseshoe prior with the Starlet wavelet basis for regularized sparse modeling of galaxy images. The Horseshoe prior complements this wavelet sparsity through its adaptive shrinkage behavior that aggressively penalizes insignificant values while retaining important features. Together, the joint sparsity induced by the Starlet basis and horseshoe regularization constrain the ill-posed inverse problem of reconstructing lensed galaxy images.

By leveraging Bayesian computational tools like MCMC, this chapter establishes a methodological framework to perform accurate and efficient analysis of distorted astronomical images.

5.1. Methodology

5.1.1. Data Preprocessing

5.1.1.1. Constructing the Lensing Matrix

In the context of gravitational lensing, lensing matrix, **L**, operates as a critical intermediary, facilitating the conversion of pixel clusters from the true (unknown) image to the (observed) distorted arcs. This matrix effectively encodes a coordinate transformation and distorts a scalar field represented by an image. To illustrate, consider a scenario where four adjacent pixels within the observed image collectively originate from a single pixel within the original galaxy image. **L** acts as a mathematical operator, mapping the intensity of the original pixel onto these four observed pixels. In this study, L is a sparse matrix with a size of 40,000 × 10,000.

To construct L, a ray tracing simulation with parameters η is performed and the source-plane and image-plane coordinates of pixels are used to determine the mapping between pixels.

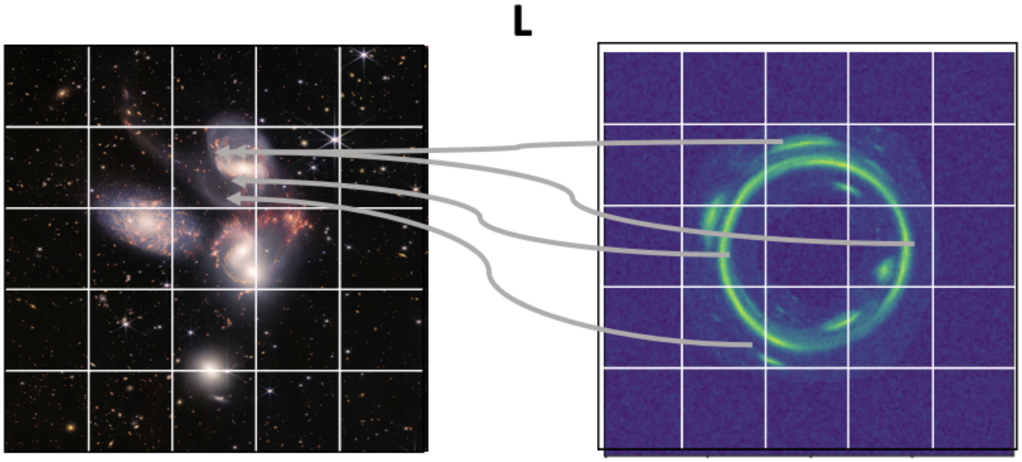


Fig. 5.1. Lensing function

5.1.1.2. Image Conversion and Standardization

To develop our methodology, we need to simulate the effect of lensing and data collection and then analyze the mock data with our inference methods, trying to recover the known ground truth used in the simulations. Of course, these simulations require an undistorted image for the background galaxies. These ground truth images are standardized in our pipeline by resizing the true image, $\mathbf{x_{true}}$, to achieve a predetermined dimension of 100×100 pixels. This step not only ensures that the image has the same resolution for analysis (to match the size of delensing matrix L) but also reduces computational complexity during further processing.

When dealing with color images in the form of a three-dimensional array, each pixel is characterized by its red, green, and blue color components. Since gravitational lensing is achromatic, these color channels could be treated independently and separately. Therefore, in our work we only focus on grayscale images, noting that more color information can be treated with the same methodology applied multiple times.

To convert RGB images to grayscale, we calculate the average of the red, green, and blue pixel values for each pixel. By doing so, the individual contributions of each color channel's lightness or luminance are blended into a cohesive grayscale approximation.

The resized image is converted into a format suitable for analysis by reshaping the 2D image matrix, 100×100 into a $1D$ vector of $10{,}000 \times 1$.

The normalization process is critical in preparing the image data for subsequent analysis, as it alleviates the impact of varying intensity scales that can result from distinct imaging conditions, sensor characteristics, or exposure settings. By transforming the data into a standardized range. The normalized pixel intensities become directly comparable across different images, making it possible to focus on the gravitational lensing effects and underlying astrophysical features. Each element in the vector corresponds to the grayscale intensity of a specific pixel in the image. The pixel intensity values are originally represented as grayscale values ranging from 0 to255, where 0 corresponds to black and 255 corresponds to white. To achieve normalization, the pixel intensity values are divided by the maximum possible intensity value, i.e., 255. This division operation effectively scales down all pixel values to lie within the unit interval $[0, 1]$.

5.1.2. Lensed Image Simulation

To produce mock data for analysis (where the ground truth is known), we use the matrix L calculated with a ray-tracing code, to produce distorted images of our galaxy images, yielding a synthetic "lensed" image. This image, stored as matrix D, mirrors what would be observed through the gravitational lens. The matrix D is produced through a linear transformation of our true image. This process involves the multiplication of two key constituents: the transpose of matrix L with a size of 40,000 times 10,000, and the normalized galaxy image data, with a size of $10{,}000 \times 1$. Figure 5.2 shows a lensed galaxy with a simulated noisy data.

5.1.3. Simulating Noise

After generating the simulated observation based on the lensing matrix and standardized data, the next step is to add noise. Real data is always noisy, due to various physical processes (photon shot noise, read-out noise, thermal noise, etc.). Although the true statistics of noise could be potentially complex, thanks to the Central Limit Theorem, noise is often very well approximated by a normal distribution. This is because many (possibly non=Gaussian) stochastic effects contribute in an additive manner to the noise, resulting in an almost-Gaussian behaviour.

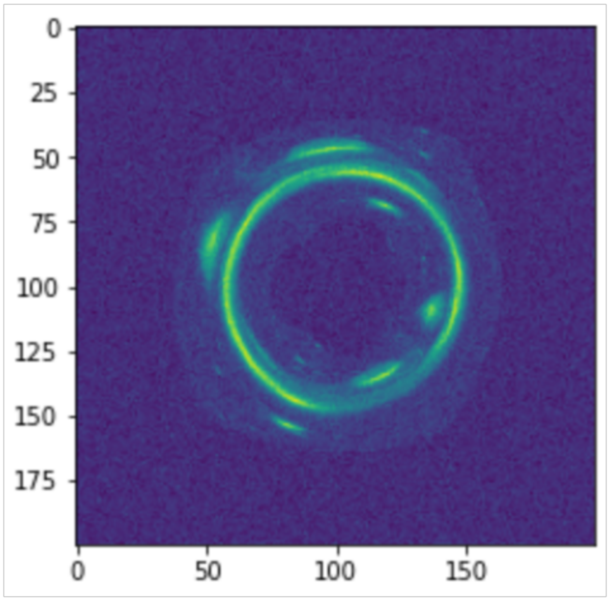


Fig. 5.2. Simulated noisy data of a lensed galaxy

To generate simulated noise, we produce random samples from a standard normal distribution. The noise is subsequently added to the previously simulated lensed image resulting in a noisy version of the observation.

This process of noise addition effectively simulates the various sources of noise encountered in astronomical observations, such as background noise, instrument noise, etc.

These steps yield two images: the S_{true} image, which encompasses the standardized grayscale groud truth image of the background source, and y, the simulated noisy observation. These images constitute the data for exploring, building and testing our analysis methods.

Figure 5.3 below showcases the flowchart corresponding subsection 5.1.1, which delineates the sequence of steps involved in the data generation process for gravitational lensing, from the construction of the lensing matrix to the simulation of noisy observational data.

5.2. Approaches for Solving the Inverse Problem

5.2.1. Bayesian Computation Using MCMC

5.2.1.1. Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are algorithms for sampling from complex multidimensional probability distributions [12]. The key idea behind the MCMC algorithm is to construct a Markov chain that converges to the target distribution as its stationary distribution. After running the chain for a sufficient number of steps, samples can be drawn to estimate expectations with respect to the target distribution.

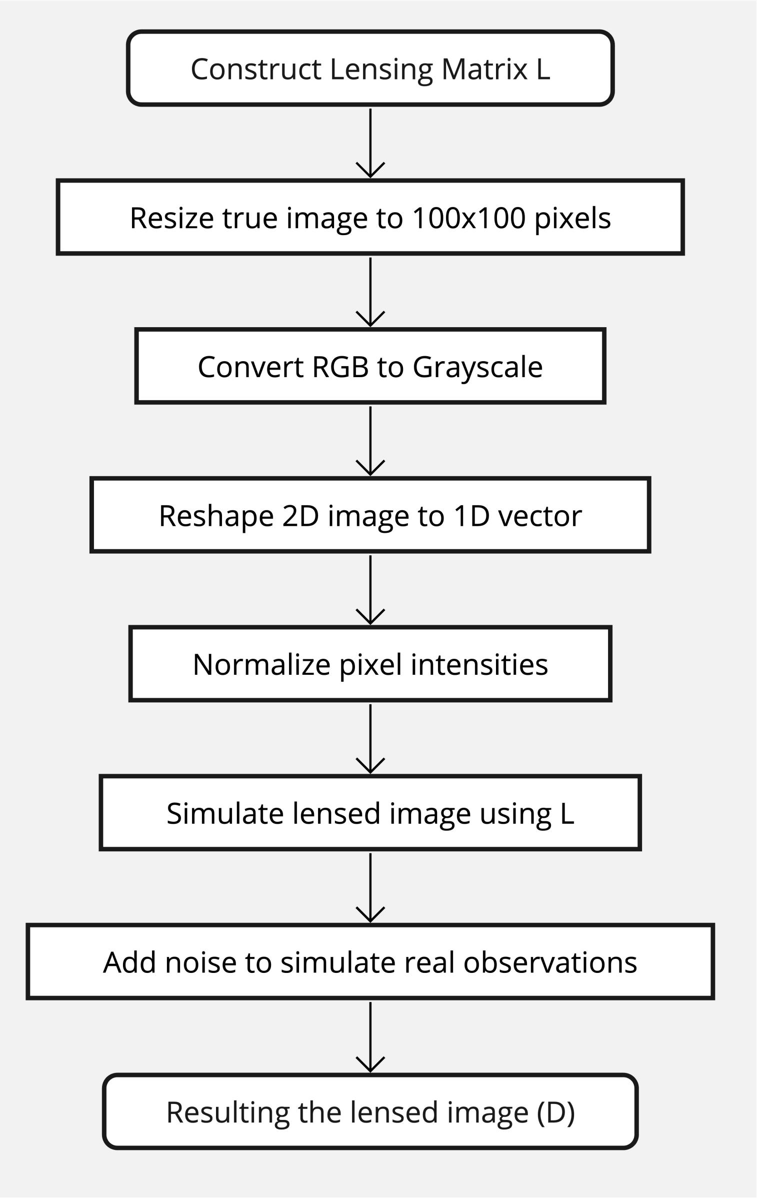


Fig. 5.3. Flowchart of the Gravitational Lensing Data Generation Process

Consider a probability distribution $\pi(x)$ that is difficult to sample from directly, with $x \in \mathcal{X}$ representing the state space. An MCMC algorithm constructs a Markov chain $\{X_t\}_{t=0}^{\infty}$ with state space $\mathcal{X}$ and transition kernel $P(x'|x)$ that satisfies the stationary distribution equation:

$$\pi(x) = \int_{\mathcal{X}} P(x|x')\pi(x')\,dx' \tag{5.2.1}$$

This equation states that sampling x' from the transition kernel $P(\cdot|x)$ and then sampling x from the distribution $\pi(x)$ yields the same joint distribution as sampling x from $\pi(x)$ and then sampling x' from $P(\cdot|x')$ [12]. Therefore, if the chain is started in the stationary distribution $\pi(x)$, then each subsequent sample will also be drawn from $\pi(x)$.

The transition kernel $P(x'|x)$ defines the stochastic dynamics of the Markov chain and is designed to satisfy detailed balance:

$$\pi(x)P(x'|x) = \pi(x')P(x|x') \tag{5.2.2}$$

This condition, along with irreducibility and aperiodicity, ensures the Markov chain is ergodic and has $\pi(x)$ as its unique stationary distribution according to the Markov chain convergence theorems [22].

After constructing the Markov chain, it is initialized at some state $X_0 = x_0$ and run for N iterations until convergence to the target distribution $\pi(x)$. Convergence can be assessed using potential scale reduction factor, effective sample size, and other diagnostics [11]. The samples $\{X_t\}_{t=1}^{N}$ can then be used to estimate expectations with respect to $\pi(x)$ via ergodic averages:

$$E[f(x)] = \int_{\mathcal{X}} f(x)\pi(x)\,dx \approx \frac{1}{N}\sum_{t=1}^{N} f(X_t) \tag{5.2.3}$$

The central challenge in MCMC is designing the transition kernel $P(x'|x)$ such that detailed balance is satisfied while still allowing efficient exploration of the state space $\mathcal{X}$. Different MCMC algorithms use various strategies to construct suitable Markov chain dynamics.

5.2.1.2. Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is an MCMC algorithm that uses Hamiltonian dynamics to efficiently sample from complex multidimensional target distributions [45, 9]. It introduces momentum variables and integrates Hamilton's equations to propose transitions between states.

As with all MCMC algorithms, HMC must satisfy detailed balance to ensure the correct stationary distribution. This is achieved using the Metropolis acceptance criterion: given a current state θ, a new proposed state θ^* is accepted with probability

$$P_{acc}(\theta, \theta^*) = \min\left(1, \frac{\pi(\theta^*)}{\pi(\theta)}\right) \tag{5.2.4}$$

where $\pi(\theta)$ is the target density. This accept/reject step compensates for any errors in the proposals to recover the desired stationary distribution.

Consider a target probability distribution $\pi(\theta)$ that is difficult to sample from directly, where $\theta \in \mathbb{R}^d$ represents the parameters. HMC augments the parameter space with momentum variables $\rho \in \mathbb{R}^d$ and defines a Hamiltonian function:

$$H(\theta, \rho) = U(\theta) + K(\rho) \tag{5.2.5}$$

where $U(\theta) = -\log \pi(\theta)$ is the potential energy and $K(\rho) = \frac{1}{2}\rho^T M^{-1}\rho$ is the kinetic energy with mass matrix M.

The dynamics of the Hamiltonian system are governed by Hamilton's equations:

$$\frac{d\theta}{dt} = \nabla_\rho H(\theta, \rho) = M^{-1}\rho \tag{5.2.6}$$

$$\frac{d\rho}{dt} = -\nabla_\theta H(\theta, \rho) = -\nabla U(\theta) \tag{5.2.7}$$

These equations describe a deterministic proposal mapping $(\theta, \rho) \to (\theta^*, \rho^*)$ that preserves volume in phase space and follows the contours of the target density [45]. This avoids the random walk behavior of simpler MCMC algorithms.

Discretizing Hamilton's equations with step size ϵ gives the leapfrog integrator which approximates the true dynamics:

$$\rho^{t+1/2} = \rho^t - \frac{\epsilon}{2}\nabla U(\theta^t) \tag{5.2.8}$$

$$\theta^{t+1} = \theta^t + \epsilon M^{-1}\rho^{t+1/2} \tag{5.2.9}$$

$$\rho^{t+1} = \rho^{t+1/2} - \frac{\epsilon}{2}\nabla U(\theta^{t+1}) \tag{5.2.10}$$

After sampling the momentum ρ from the Gaussian distribution $\mathcal{N}(0, M)$, the algorithm proceeds with L leapfrog steps to simulate the Hamiltonian dynamics. Each leapfrog step updates both the position θ and momentum ρ variables using the leapfrog integration equations. After simulating the dynamics for L steps, a Metropolis accept/reject step is applied to satisfy detailed balance. The acceptance probability is calculated using the formula:

$$\min\left(1, \exp\left[H(\theta, \rho) - H(\theta^*, \rho^*)\right]\right)$$

where (θ^*, ρ^*) represents the proposed state. This completes the description of the Hamiltonian Monte Carlo algorithm.

Algorithm 1 Hamiltonian Monte Carlo

```
Sample momentum $\rho \sim \mathcal{N}(0, M)$
Evolve $(\theta, \rho) \to (\theta, \rho)$ using $L$ leapfrog steps
Accept $(\theta, \rho)$ with probability:
    $\min\left(1, \exp\left[H(\theta, \rho) - H(\theta^*, \rho^*)\right]\right)$
```

HMC, by approximating Hamiltonian dynamics, offers a more efficient exploratory mechanism in the parameter space than basic MCMC algorithms, primarily due to its avoidance of random walk behavior. The parameters L and ϵ can be tuned to trade off accuracy vs. computational cost.

5.2.1.3. No-U-Turn Sampler

The No-U-Turn Sampler (NUTS) builds upon Hamiltonian Monte Carlo (HMC) by adapting the trajectory length to efficiently sample from complex multi-modal posteriors [32].

Whereas HMC requires manually tuning parameters like step size and number of leapfrog steps, NUTS enhances efficiency by adaptively selecting the number of steps, recursively building a candidate point set without the need for manual parameter tuning that trace out a trajectory through the target distribution.

Starting from an initial state (θ, ρ), the NUTS algorithm uses a recursive build tree procedure to extend the trajectory by applying leapfrog steps until some stopping criteria is reached. This allows the path length to adapt based on the curvature of the target density.

The build tree algorithm is as follows:

Algorithm 2 BuildTree(j, θ, ρ, $\bar{\theta}$, $\bar{\rho}$, ϵ, L)

```
if $j = L$ then
    return $\theta, \rho, s = j - 1$
end if
Apply leapfrog update to get $(\theta', \rho')$
if $U(\theta') > U(\bar{\theta})$ then
    return $\theta, \rho, s = j - 1$
else
    $(\bar{\theta}, \bar{\rho}, s) = \text{BuildTree}(j + 1, \bar{\theta}, \bar{\rho}, \theta', \rho', \epsilon, L)$
end if
Apply leapfrog update to get $(\theta'', \rho'')$
return $\bar{\theta}, \bar{\rho}, s$
```

By recursively applying leapfrog steps in this way, NUTS adapts the trajectory length while avoiding wasteful random walks or retracing of steps. This enables efficient exploration of complex posteriors without extensive parameter tuning.

Algorithm 3 Hamiltonian Monte Carlo Overview

```
1: Initialization: Start with an initial state $x^{(0)}$ and momentum $p^{(0)}$
2: Simulate Hamiltonian dynamics: Propose a trajectory by numerically solving Hamilton's equations using leapfrog integration.
3: Termination Criterion: Detect U-turns in the trajectory to determine whether the chain should continue or terminate.
4: Metropolis Acceptance: Accept the proposed state with a probability that balances exploration and exploitation.
5: Repeat: Generate subsequent states by iteratively applying the algorithm.
```

Advantages of using NUTS is that the NUTS eliminates the need to manually set tuning parameters and adaptively adjusts the trajectory length, leading to faster convergence and reduced autocorrelation among samples. Its ability to explore complex distributions efficiently makes it a valuable tool in Bayesian inference.

5.2.2. The proposed approach for galaxy image reconstruction using HMC

This subsection outlines our proposed method for reconstructing galaxy images using the Hamiltonian Monte Carlo (HMC) technique. We present an algorithm that systematically outlines the key steps involved in this approach. Beginning with the acquisition of a noisy lensed image , y, and the initialization of a wavelet transform, we subsequently define the likelihood function connecting the observed image to the image coefficients. Incorporating a prior distribution on these coefficients, such as the horseshoe prior, we employ HMC to sample from the posterior distribution. This leads to the reconstruction of the source image, $D_{estimated}$, and finally, the estimation of the source image through posterior mean calculations. The following algorithm encapsulates these steps, providing a comprehensive overview of our proposed methodology.

Algorithm 4 The proposed method for galaxy image reconstruction based on HMC

```
1: For galaxy image $\mathbf{D_{true}}$ and lensing operator $\mathbf{L}$ , obtain noisy lensed image $\mathbf{y}$
2: Initialize wavelet transform with $J$ scales
3: Define likelihood $p(\mathbf{y}|\mathbf{c})$ relating image to coefficients $\mathbf{c}$
4: Set prior $p(\mathbf{c})$ on coefficients (e.g. horseshoe)
5: Run HMC to sample from posterior $p(\mathbf{c}|\mathbf{y})$
6: Reconstruct source images $\mathbf{D_{estimated}}$ from posterior samples of $\mathbf{c}$
7: Estimate source image as posterior mean $\hat{\mathbf{D}} = \mathbb{E}[\mathbf{S_{estimated}}|\mathbf{y}]$
```

5.2.3. Methodological Foundations: Probabilistic Reconstruction of Lensed Images

In the reconstruction of lensed galaxy images, a pivotal role is played by the integration of Hamiltonian Monte Carlo (HMC) and the starlet transform, a combination that facilitates a robust and efficient inference process as shown in Figure 5.4. The starlet transform, renowned for its ability to provide a sparse representation of galaxy images across different scales, is utilized to decompose the background source image into a set of coefficients. These coefficients are then treated as the parameters of our model, encapsulating the intricate details of the galaxy's structure. The HMC algorithm, with its proficiency in navigating through complex, high-dimensional parameter spaces, is employed to infer these starlet coefficients. By generating samples from the posterior distribution of the coefficients, HMC ensures a thorough exploration of the parameter space, capturing the uncertainty and variability inherent in the reconstruction process. The proposed values of the starlet coefficients are utilized to reconstruct the background source image through the inverse starlet transform. Subsequently, this reconstructed image undergoes a lensing transformation, aligning it with the observational data (denoted as S). The likelihood of observing D, given our model parameters, is then computed, providing a quantitative measure of the goodness-of-fit between the lensed, reconstructed image and the actual observed data. This is supplemented by a prior model, which encapsulates our prior beliefs and knowledge about the distribution of the starlet coefficients, serving to regularize the solution and impose constraints that are consistent with the physical nature of galaxy images.

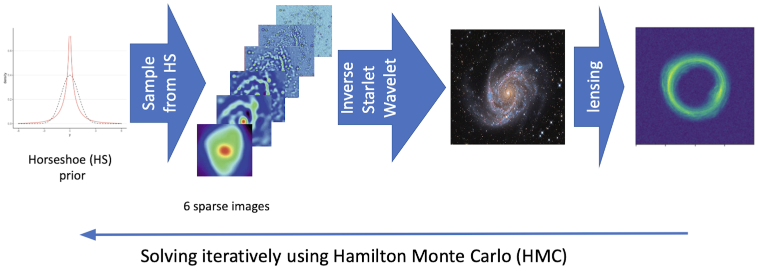


Fig. 5.4. Illustratin of the Proposed Galaxy Image Reconstruction Methodology. This figure outlines the step by step procedure for reconstructing galaxy images using the novel approach described in this thesis. The methodology includes the incorporation of a lensing operator, wavelet transform, likelihood and prior models, and Hamiltonian Monte Carlo sampling and the lensed image

5.3. Error Measures

In in all inference tasks, various error measures are employed to assess the quality of reconstructed or processed data. In this section, we will discuss two commonly used error measures: Mean Squared Error (MSE) and Structural Similarity Index (SSIM). We use these error estimates to quantify the similarity of our reconstructed images of background sources to the ground truths.

5.3.1. Mean Squared Error (MSE)

The Mean Squared Error (MSE) is a used error measure that quantifies the average squared difference between the pixels of a reference image and a reconstructed image. It provides a measure of the overall discrepancy between the two images.

Given a reference image I and a reconstructed image $\hat{I}$, the MSE is calculated as follows:

$$MSE(I, \hat{I}) = \frac{1}{N}\sum_{i=1}^{N}(I_i - \hat{I}_i)^2 \tag{5.3.1}$$

In Equation 5.3.1, N is the total number of pixels in the images, and I_i and $\hat{I}_i$ represent the intensity values of the corresponding pixels in the reference and reconstructed images, respectively.

A lower MSE value indicates a better resemblance between the reference and reconstructed images, with a value of zero indicating a perfect match. However, MSE alone may not always capture perceptual differences between images, and other factors such as noise or distortion may affect its value.

5.3.2. Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is an error measure that takes into account not only the pixel-wise differences but also the structural information and perceptual aspects of the images. It aims to assess the similarity of the local patterns and structures between the reference and reconstructed images.

The SSIM index is calculated as

$$SSIM(I, \hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + c_1 \cdot 2\sigma_{I\hat{I}} + c_2}{\mu_I^2 + \mu_{\hat{I}}^2 + c_1 \cdot (\sigma_I^2 + \sigma_{\hat{I}}^2 + c_2)},$$

where μ_I and $\mu_{\hat{I}}$ are the average intensities of the reference and reconstructed images, respectively, σ_I and $\sigma_{\hat{I}}$ are their respective standard deviations, and $\sigma_{I\hat{I}}$ is the covariance between them. The constants c_1 and c_2 are small values added to avoid division by zero.

The SSIM index ranges between -1 and 1, with a value of 1 indicating perfect similarity. A higher SSIM value suggests a better match between the images in terms of their structure and perception. SSIM is known to correlate better with human perception than MSE, making it a valuable tool in image quality assessment.

Both MSE and SSIM provide useful information for evaluating image quality, but they focus on different aspects. MSE primarily measures pixel-wise differences, while SSIM incorporates perceptual and structural similarities. It is often recommended to use a combination of these measures to obtain a more comprehensive assessment of image fidelity and quality.

Chapter 6

Results

6.1. Results

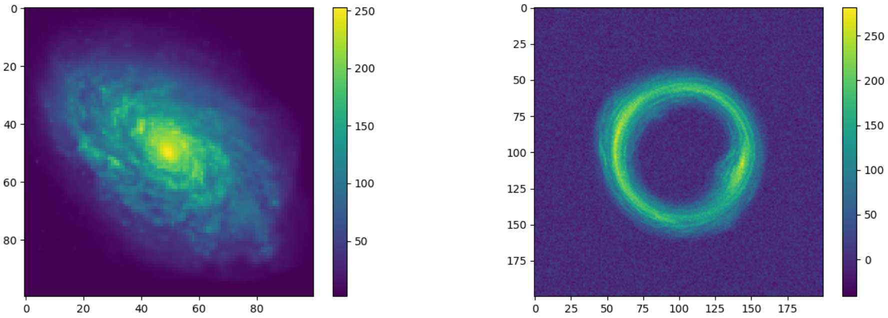
In this section, we present the results of applying the Bayesian framework presented above with the Horseshoe prior and Hamilton Monte Carlo sampling to the reconstruction of the true images of lensed galaxies. The goal is to infer the underlying structure of the galaxy and its various scales of features from noisy observed data.

6.1.1. Description of the simulated data

In this section, we showcase the results derived from applying our proposed methodology, which integrates the Horseshoe prior, Starlet wavelet transformation, and Hamilton Monte Carlo (HMC) sampling to the tasks of reconstructing background galaxy images. To develop and test our method, we use the image of a nearby galaxy that is *not* gravitationally lensed, $NGC4414$, as a background source in strong lensing simulations and try to recover its undistorted image from simulated datasets. The image of $NGC4414$ is depicted in Figure 6.1. This galaxy is a dusty barred spiral galaxy located in the Coma Berenices constellation. The galaxy is characterized by a central bar, elegant spiral arms, and is situated approximately 60 million light-years away. Through this work, we aim to demonstrate the effectiveness of our methodology in mitigating the effects of gravitational lensing and noise on the appearance of the galaxy, ultimately yielding a clearer and more accurate representation of the structure of lensed background galaxies.

Essential transformations were conducted to ensure that the galaxy images were represented appropriately within a numerical scale compatible with the proposed method described in Sections 5 and 5.1.1. Following the data preparation steps outlined in Section 5.1.1, we transformed the image in Figure 6.1 into an observed lensed galaxy image. The left subplot in Figure 6.2 presents the original $NGC4441$ galaxy image after grayscale conversion and pixel intensity standardization. The right subplot in Figure 6.2 depicts the simulated

Fig. 6.1. Galaxy NGC 4414, Credit: The Hubble Heritage Team



(a) (b)

Fig. 6.2. (a) Preprocessed image of NGC 4441, illustrating the distribution of pixel intensities (b) Noisy, lensed observation of NGC 4441

observed image, obtained by applying a lensing effect, multiplication by the matrix L, to the original image, and adding noise, $\mathcal{N}(0,10)$, as detailed in Sections 5.1.2 and 5.1.3 respectively. Figure 6.2 illustrates the data simulation pipeline utilized to generate realistic, noisy, lensed images from original galaxy images, serving as a testbed for evaluating the reconstruction methodology.

In the subsequent stage of our analysis, we employ the starlet transform, which facilitates an effective multiscale wavelet analysis particularly apt for examining spiral galaxies such as NGC 4414. This approach enables us to succinctly capture the galaxy's overall shape as well as its localized substructures, providing a comprehensive representation of its morphology.

This multi-resolution sparsification contributes to regularizing the ill-posed inverse problems associated with lensing reconstruction. In the case of NGC 4414, Figure 6.3 illustrates the multiscale Starlet wavelet decomposition of the original galaxy image. The leftmost panel shows the input galaxy image. The middle panels showcase the coefficient images obtained at each wavelet scale. Brighter pixels indicate larger coefficient magnitudes. There are $J = 6$ scales displayed corresponding to different levels of image details. The coarsest scale coefficient image reflects the smooth overall morphology, while finer scales capture increasingly localized features.

The decomposition process hierarchically isolates features, ranging from the overall morphology to the finest details. As we progress to finer scales, the coefficients become increasingly sparse, highlighting only the most striking patterns. This multiresolution analysis is instrumental in detecting and extracting pertinent structures within the galaxy. The sparsity of the coefficients is a notable observation, with pixel brightness concentrating at specific scales and spatial locations. This characteristic enables the compression of image content into salient coefficients that represent the hierarchical structure of distinct image features. At the finest scales, we observe the isolation of the most abrupt brightness variations, identifying features such as the core and spiral arms of the galaxy. Conversely, the coarsest scale captures the extended shape of the galaxy, illustrating its overall structure. In essence, this multiresolution analysis decomposes the galaxy image, transitioning from large-scale brightness patterns to the minutiae of small-scale details.

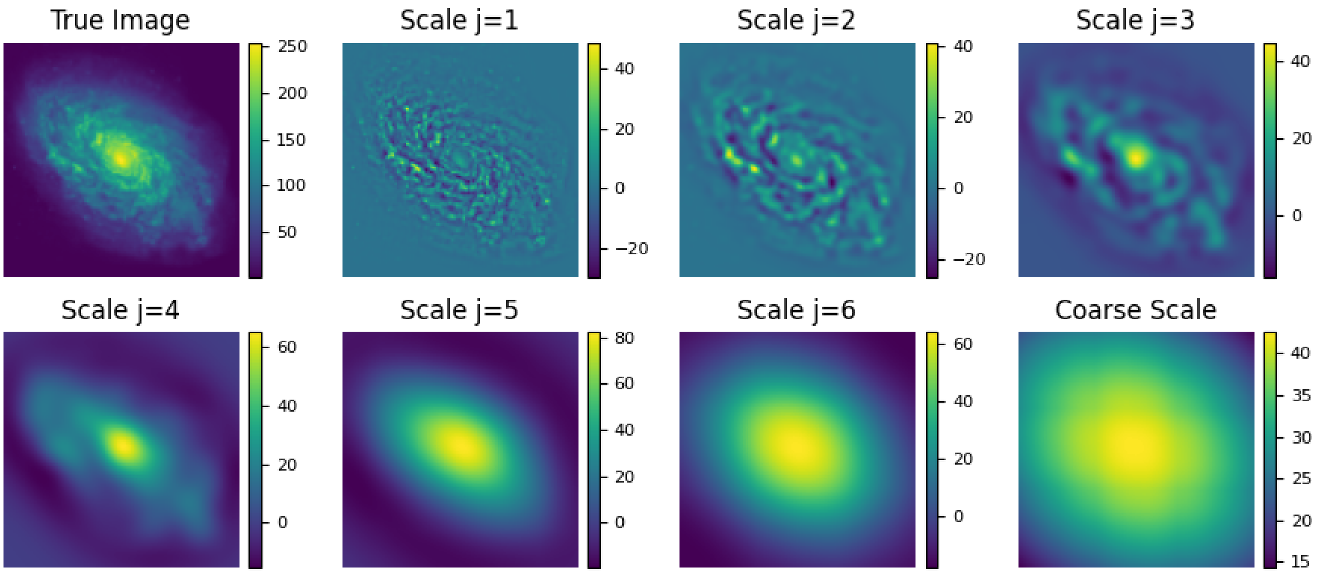


Fig. 6.3. Illustration of multiscale decomposition of NGC 4441 via the Starlet wavelet transform: The original galaxy image is presented in the left panel, the middle panels exhibit the coefficient images at each of the 6 wavelet scales utilized, and the right panel depicts the image reconstructed by summing all scales of the decomposition.

6.1.2. Implementation Details and Probabilistic Framework

Our framework is implemented in `Python`, utilizing `Pytorch` and `Pyro`, and supplemented by the integration of JAX for specific computational tasks. `Pyro`, as a flexible probabilistic programming language, leverages the capabilities of `PyTorch`, such as automatic differentiation and GPU acceleration.

In our probabilistic model, which focuses on reconstructing lensed galaxy images, we treat the starlet coefficients, as illustrated in Figure 6.3, as parameters within a Bayesian framework. The inference of the posterior distribution of these coefficients is carried out using the Hamiltonian Monte Carlo (HMC) method, coupled with the No-U-Turn Sampler (NUTS) for efficient sampling. JAX, in our framework, is employed for optimizing select numerical computations.

As highlighted in subsection 5.2.3, the Starlet transform's coefficients, are the parameters we aim to infer using Hamiltonian Monte Carlo. After determining these coefficients, as depicted in Figure 6.3, we apply the inverse starlet transform, utilizing functions such as `reconstruct_second_generation`, to transform these coefficients back into the image space, resulting in a reconstructed image of the background source. This reconstructed source is then lensed using a lensing operator, and the resulting lensed image,Figure 6.4, is compared to the observed data $\mathcal{D}$, as shown in panel (b) of Figure 6.2. This comparison is conducted within a likelihood framework, and together with an appropriately defined prior, it guides the HMC algorithm in the inference process, ensuring the accurate reconstruction of the background source image.

In the subsequent section, we will explore visual representations of our results, highlighting the effectiveness of our proposed methodology and the accuracy of our probabilistic approach in reconstructing images of lensed galaxies.

6.1.3. Result Visualization

Here, we show the results of our inference using the Horseshoe prior and compare them against the ground truth when using different priors such as the Laplace and Normal priors.

6.1.3.1. Image Reconstruction with Horseshoe Prior

The reconstructed image of NGC 4414, facilitated by the Bayesian inference model employing Horseshoe regularization, is exhibited in Figure 6.4. We have assumed that noise follows a normal distribution $\mathcal{N}(0, 10)$.

Figure 6.5 illustrates the scales of the coefficients obtained through the Horseshoe regularization. In the eight subplots, the hierarchical decomposition of the image reconstruction is revealed. The first subplot, titled "Reconstructed Image," showcases the complete reconstructed image. The following six subplots illustrate different scales of coefficients, with Scale

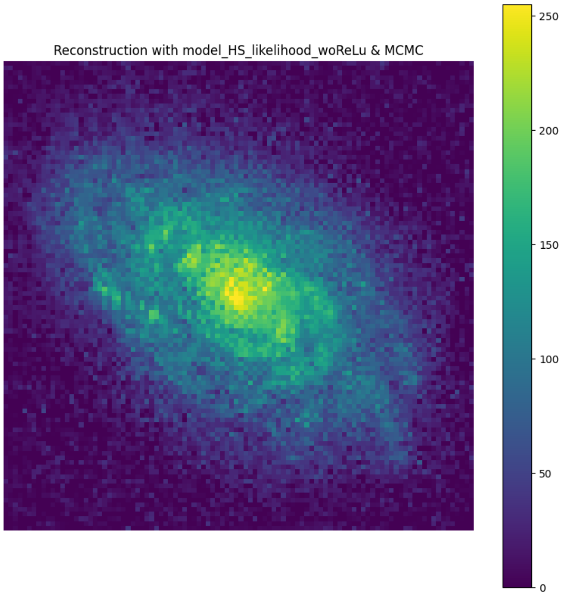


Fig. 6.4. Reconstructed image of Galaxy NGC 4414 with Horseshoe prior.

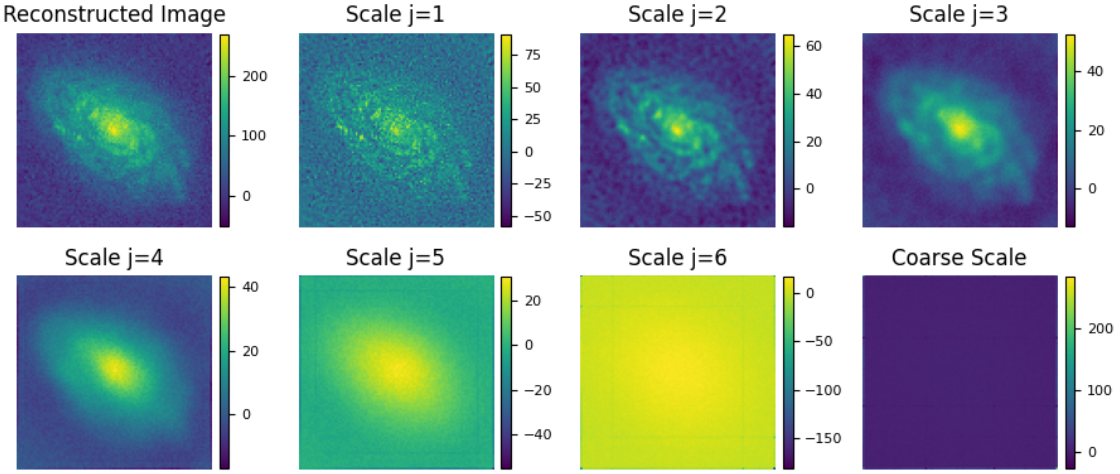


Fig. 6.5. Coefficients of the reconstructed image with Horseshoe prior, demonstrating the hierarchical scale decomposition.

$j = 1$ representing the finest scale, and "Coarse Scale" showing the coarsest scale. The color bars beside each subplot indicate the intensity values and their sizes are adjusted for clarity. This figure provides valuable insights into the hierarchical structure of the reconstructed image, revealing the contributions of different scales to the overall reconstruction.

The reconstructed image underscores the effectiveness of our approach in extracting meaningful information from noisy astronomical observations. In order to evaluate the effectiveness of our Bayesian inference model with Horseshoe regularization, we conducted a comparative analysis between the reconstructed image and the ground truth image of NGC 4414.

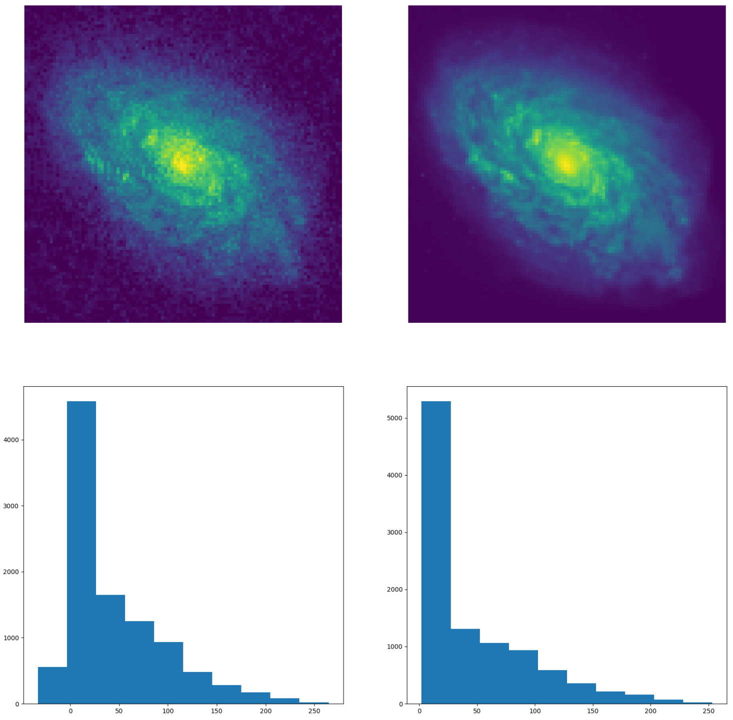


Fig. 6.6. Comparative analysis of the Reconstructed Image (with Horseshoe Regularization) versus the Ground Truth Image of Galaxy NGC 4414, accompanied by histograms of pixel values.

Figure 6.6 presents a comprehensive comparison between the reconstructed image obtained through our approach and the ground truth image. In the top-left subplot, the reconstructed image is displayed, showcasing the galaxy's features and structures that have been extracted from the noisy data. On the top-right subplot, the ground truth image is shown, providing a reference for the actual features present in NGC 4414.

Furthermore, the lower-left histogram illustrates the distribution of pixel values in the ground truth image. In contrast, the lower-right histogram represents the distribution of pixel values in the reconstructed image. By comparing these histograms, we can assess how well our methodology captures the intensity distribution of the galaxy.

The histograms in figure Figure 6.7 provide a detailed analysis of the coefficients' distribution for various scales in the reconstructed image achieved using Horseshoe regularization. Each histogram corresponds to a specific scale, denoted by j. By examining these histograms, we can gain insights into the statistical properties of the coefficients at different scales. The histograms showcase the frequency distribution of the mean coefficient values.

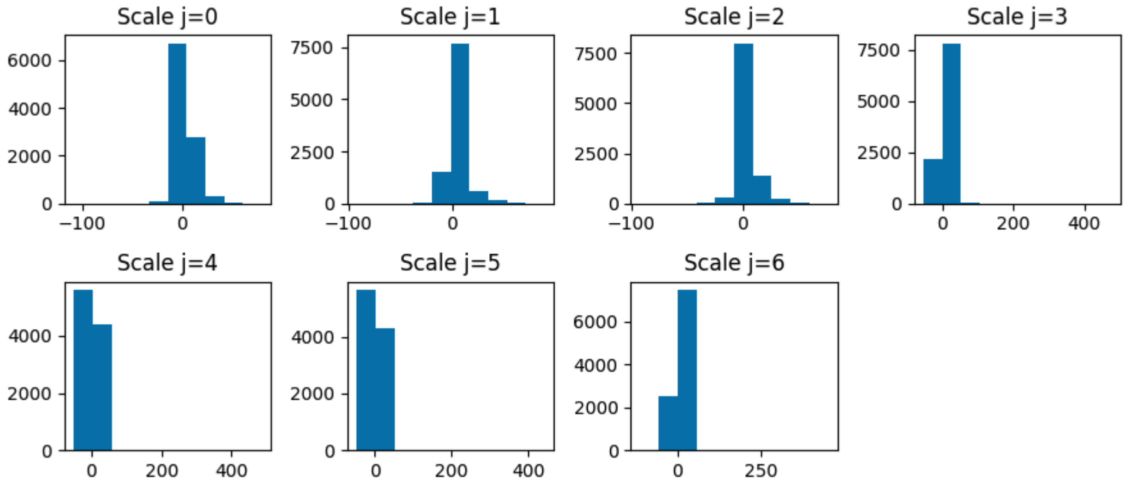


Fig. 6.7. Histograms showcasing the distribution of mean coefficient values across different scales in the reconstructed image with Horseshoe regularization

In order to provide a comprehensive analysis, we have considered a scenario in which no regularization is applied to the source reconstruction. For this comparison, we assumed a uniform regularization—effectively a scenario with minimal prior constraints. The resulting image, as presented in Figure 6.8, serves to highlight the stark differences when regularization is excluded. The contrasts are particularly evident; without the guiding influence of a regularization term, the reconstruction tends to exhibit increased noise and less coherent structures.

The reconstruction of the galaxy image of NGC4144 with horseshoe prior and different levels of noise is shown in the following Figure 6.9. With the gradual escalation of noise levels from 0 to 40, an observable trend emerges. The lensed and galaxy images progressively manifest heightened degrees of blurriness. Throughout this noise-induced degradation, the reconstructed image with the horseshoe prior, maintains a high quality, underscoring the appropriateness of this prior for reconstructing the signal.

In addition to the analysis of NGC 4414 as a prime example, the effectiveness of the proposed methodology was further demonstrated through the reconstruction of an additional galaxy images. Galaxy 3147 is a spiral galaxy, which means it has swirling arms coming out from its center. It's in the constellation of Ursa Major, which is a pattern of stars in the sky. This galaxy is interesting because it has a big bulge in the middle and arms that wrap around it. When scientists look at this galaxy, they find a lot of hydrogen and helium, which are

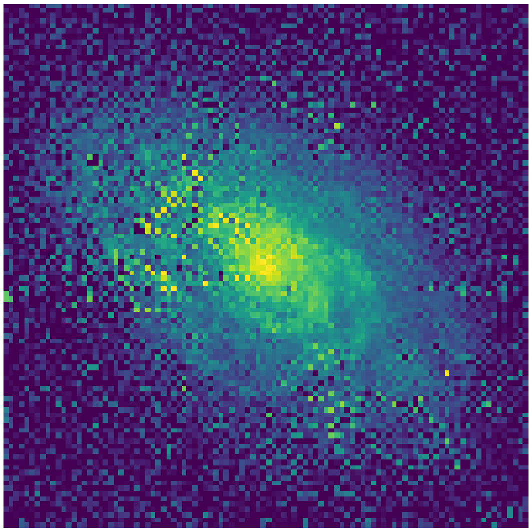


Fig. 6.8. Galaxy Image reconstruction using the Starlet transform with a uniform prior

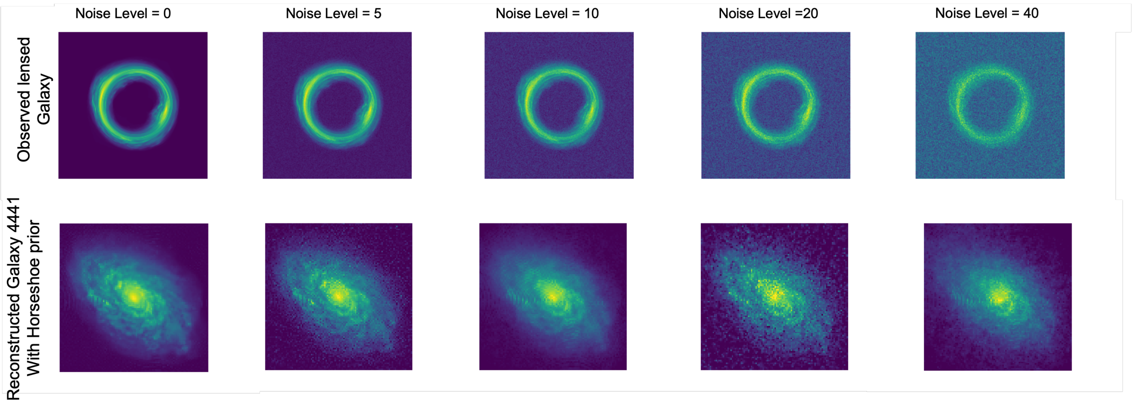
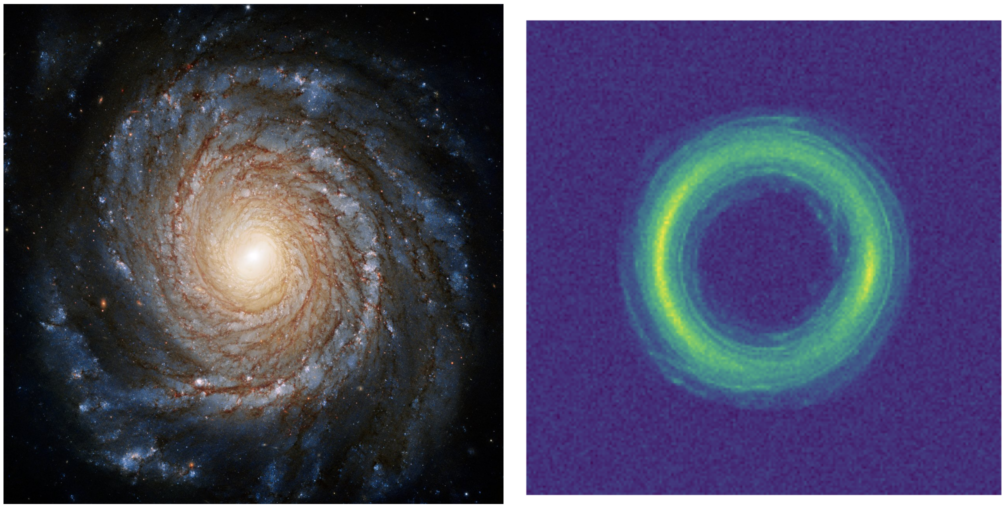


Fig. 6.9. Galaxy Image reconstruction with horseshoe prior and different noise levels

the most common elements in the universe. One of the most exciting things about Galaxy 3147 is that it has a supermassive black hole in its center. Studying this black hole helps scientists learn more about how galaxies form and change over time. Figure 6.10 provides a visual representation of it. On the left, the original image of Galaxy 3147, showcasing its spiral structure and central bulge. On the right, a lensed and noisy observation of the same galaxy, illustrating the effects of gravitational lensing and observational noise on galactic imagery.

In Figure 6.11, we compare the reconstructed galaxy image on top left, with the original on top right. Additionally, the histogram of pixel intensity values accompanying these images



(a)

(b)

Fig. 6.10. (a) Galaxy 3147, Credit:ESA/Hubble and NASA, (b) Observed noisy image

offers insight into the distribution of luminance across the reconstructed image, providing a quantitative perspective on the reconstruction's fidelity.

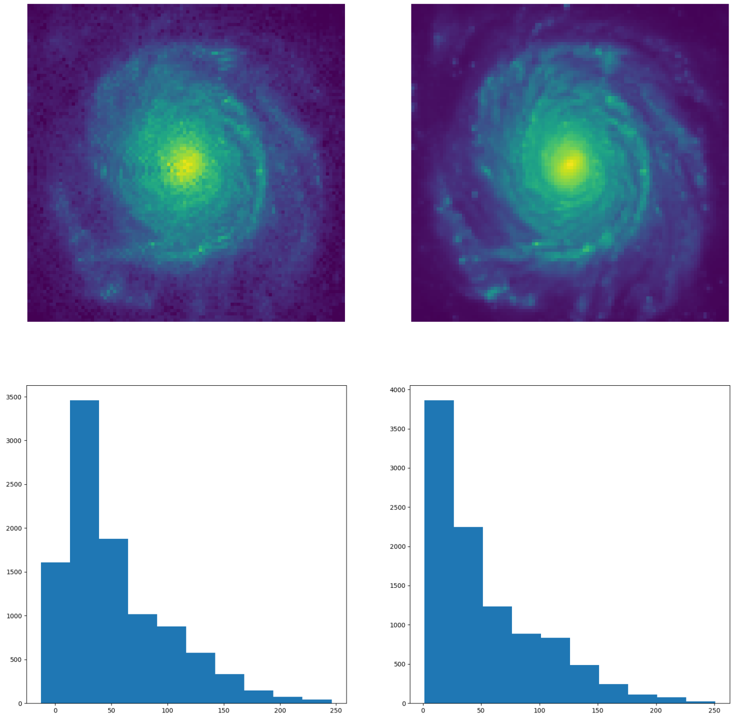


Fig. 6.11. A comparison between the reconstructed galaxy 3147 (left) and the unaltered original image (right). The histograms below each image display the distribution of pixel intensity values, showcasing the success of the reconstruction process in preserving the structural details of the galaxy.

6.1.3.2. Comparing Priors in Galaxy Image Reconstruction

In this section, we explore the impact of different regularization priors on the process of galaxy image reconstruction. We consider three distinct priors: Laplace (L1) regularization, Gaussian (L2) regularization, and the horseshoe prior with the image of Figure 6.1. Each of these priors brings its unique characteristics to the reconstruction process, influencing the final results. Our experiments involve applying each of these regularization priors to the galaxy image reconstruction process and evaluating the results. Specifically, we examine factors such as the Mean Squared Error (MSE), the Structural Similarity Index (SSIM). These metrics provide insights into the accuracy and structural fidelity of the reconstructed images.

In Figure 6.12, We compare three potential priors for wavelet-based galaxy image reconstruction: L1 prior (Laplace distribution), L2 prior (Gaussian distribution) and Horseshoe

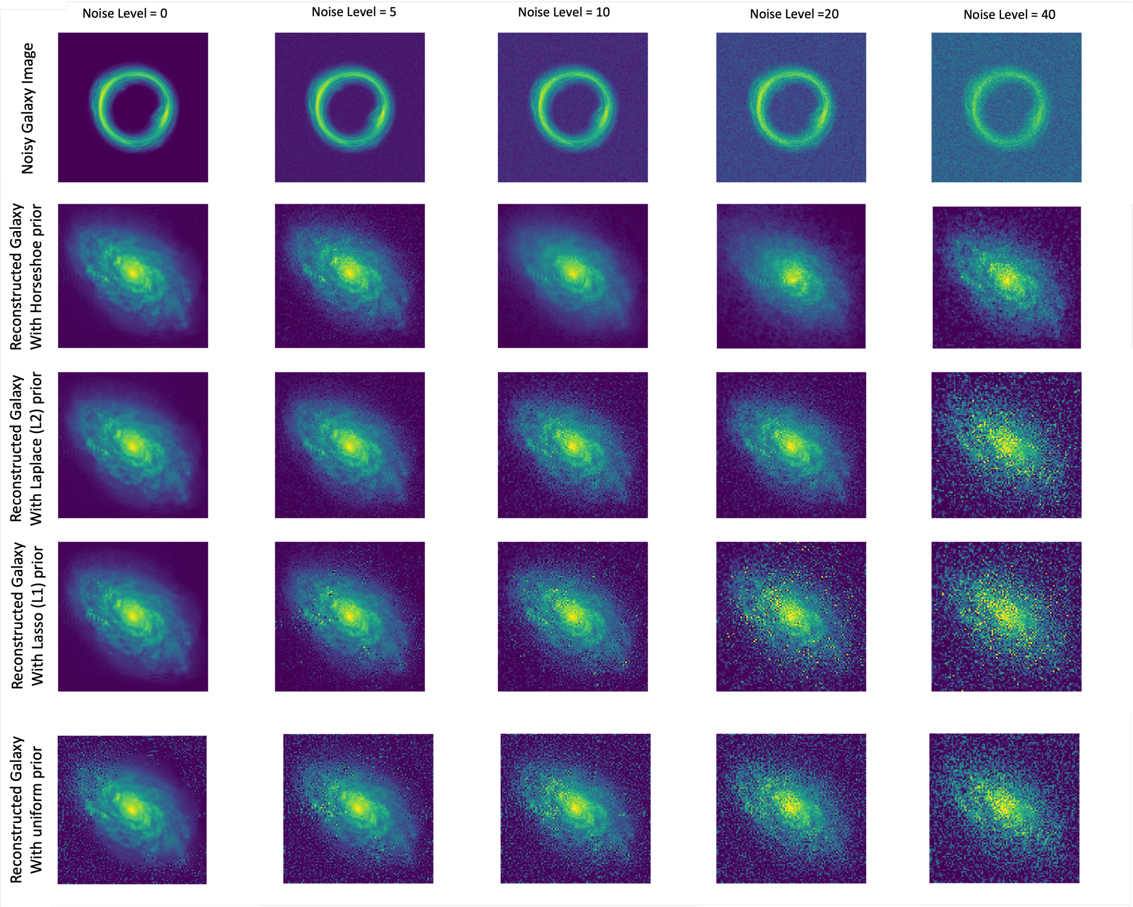


Fig. 6.12. Galaxy Image reconstruction with different priors and different noise levels

prior. These priors exhibit different attributes regarding sparsity and handling of small coefficients that influence the reconstructed galaxy image. The Laplace prior strongly promotes sparsity, while the Gaussian prior encourages smoothness. The horseshoe offers a balance, shrinking insignificant coefficients while retaining important values. By performing Bayesian inference using each prior in turn, we can compare the resulting galaxy image reconstructions. Metrics such as mean squared error and structural similarity index will quantify the reconstruction quality.

Mean Squared Error (MSE): Figure 6.13 presents a visual depiction of the MSE outcomes, offering insight into the accuracy of estimations for Horseshoe, Gaussian, and Laplace priors as the noise level escalates. Complementing this graphical analysis, Table 6.1 enumerates the precise MSE values corresponding to each prior across the spectrum of considered noise intensities. These illustrations and quantitative assessments elucidate the efficacy of each prior in noise-robust statistical inference. To further elucidate, the following points elaborate on the significance of MSE in our analysis.

- MSE quantifies the pixel-wise dissimilarity between the reconstructed and original images. Lower MSE values indicate better reconstruction quality.
- The Horseshoe Prior (HS) consistently achieved the lowest MSE, signifying the highest fidelity in terms of pixel-wise accuracy.
- Gaussian and Laplace regularizations, while effective, yielded slightly higher MSE values compared to Horseshoe.

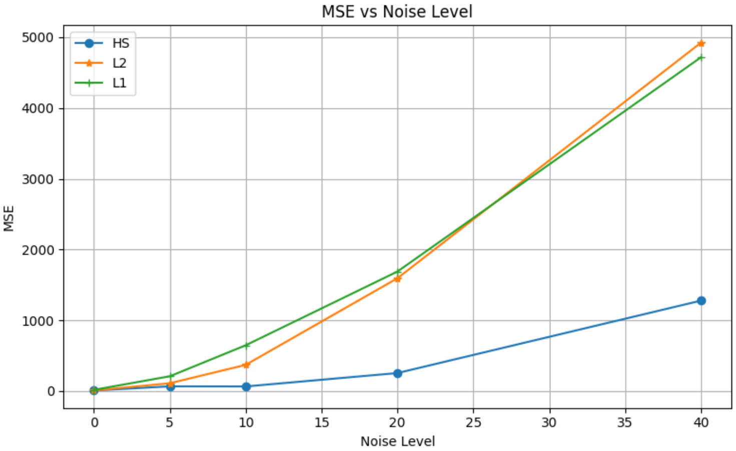


Fig. 6.13. Graphical representation of Mean Squared Errors (MSE) for Horseshoe, Gaussian, and Laplace priors, illustrating the impact of increasing noise levels on estimation accuracy.

Prior / Noise Level	**0**	**5**	**10**	**20**	**40**
Horseshoe Prior	7.8714	66.2966	64.6709	254.2778	1277.5113
Gaussian Prior	8.1593	110.1288	369.8086	1594.1022	4922.5447
Laplace Prior	16.8017	208.5986	646.7087	1688.8453	4716.0925

Table 6.1. Comparison of Mean Squared Errors (MSE) for Horseshoe, Gaussian, and Laplace priors across different noise levels.

Structural Similarity Index (SSIM):

Complementary to the Mean Squared Error analysis, the Structural Similarity Index (SSIM) offers a perspective on perceptual image quality. In Figure 6.14, we provide a graphical representation that details the SSIM values across different noise levels for the Horseshoe, Gaussian, and Laplace priors, showcasing their impact on perceived image integrity. Table 6.2 presents a tabulated summary of these results, allowing for a precise comparison of SSIM metrics under varying conditions of noise interference.

- SSIM evaluates the structural similarity between the reconstructed and original images, with values closer to 1 indicating better structural preservation.

- HS outperformed Gaussian and Laplace in terms of SSIM across all experiments, demonstrating its superior ability to capture fine structural details.

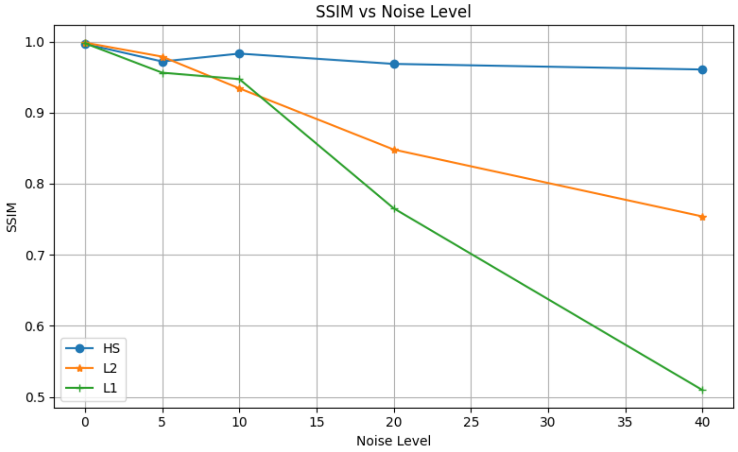


Fig. 6.14. Visual comparison of the Structural Similarity Index (SSIM) across varying noise levels for Horseshoe, Gaussian, and Laplace priors, demonstrating the preservation of image structural integrity under noise influence.

Prior / Noise Level	**0**	**5**	**10**	**20**	**40**
Horseshoe Prior	0.9985	0.9785	0.9453	0.9637	0.9655
Gaussian Prior	0.9984	0.9792	0.9151	0.7517	0.6937
Laplace Prior	0.9968	0.9243	0.9044	0.6820	0.7179

Table 6.2. Structural Similarity Index (SSIM) values for Horseshoe, Gaussian, and Laplace priors under different noise conditions, indicating the robustness of each prior against noise-induced degradation in image quality.

Chapter 7

Discussion and conclusion

7.1. Discussion

The results demonstrate the effectiveness of the proposed Bayesian framework with horseshoe regularization for reconstructing gravitationally lensed galaxy images. Leveraging the sparsity-promoting properties of the horseshoe prior coupled with the multiscale sparse representation of the Starlet transform proved highly advantageous.

Quantitative assessments using MSE and SSIM showed the horseshoe prior outperformed alternatives like Laplace and Gaussian regularization across different noise levels. This highlights its capabilities in suppressing noise while retaining important image features due to its ability to adaptively shrink insignificant coefficients.

Visually, the horseshoe prior also better preserved the structural details of the galactic morphology in the reconstructed images. The hierarchical nature of the Starlet decomposition further allowed isolating features at different scales, from overall galactic shape to finer details. This multiresolution analysis provided interpretable insights into the contributing scales.

The proposed approach overcomes the ill-posed inverse problem induced by lensing by effectively regularizing through exploiting sparsity.

A key advantage demonstrated is the robustness of the horseshoe-Starlet approach in high noise settings. As evident in the experiments, the reconstruction quality remained relatively stable even as noise levels increased substantially. The horseshoe prior's adaptive shrinkage prevents overfitting to noise while retaining important signals. This noise resilience, coupled with the Starlet transform's multiscale sparse representation, provides a robust framework for galaxy image analysis. The ability to recover fine details from noisy, low-resolution observations makes this technique well-suited for real-world gravitational lensing data where noise is ubiquitous. The proposed method's stability in high noise suggests it generalizes well to noisy astronomical images, a vital characteristic lacking in some alternative techniques.

7.2. Conclusion

This thesis presented a novel Bayesian framework leveraging the horseshoe prior and the Starlet wavelet transform for reconstructing gravitationally lensed galaxy images. The ill-posed inverse problem induced by gravitational lensing poses significant challenges in recovering the original appearance of distant galaxies. To address this, we proposed using the sparsity-promoting horseshoe prior applied to wavelet coefficients of the source galaxy image. The horseshoe prior's ability to shrink insignificant values while retaining salient features, coupled with the Starlet transform's multiscale sparse representation of galactic morphology, enables effective regularization of the underdetermined inversion. Through probabilistic modeling and Markov chain Monte Carlo posterior inference, we implemented this approach to analyze simulated observations of lensed galaxies with added noise. Our methodology reliably extracted meaningful information from the noisy data, reconstructing the key structural details of the original galaxy image. Quantitative assessments using metrics like mean squared error and structural similarity index demonstrated the superiority of the horseshoe-Starlet method compared to alternatives like L1 and L2 regularization. The proposed framework provides a principled Bayesian technique for galaxy image reconstruction that overcomes the challenges induced by gravitational lensing. By promoting sparsity in a mathematically and computationally efficient wavelet domain representation, the horseshoe prior imposed on Starlet coefficients proves highly effective at image denoising and deconvolution.

7.2.1. Future Work

While the presented methodology has demonstrated promising performance for gravitational lens inversion, there remain several worthwhile avenues for further improving and extending this approach. Some potential directions for future work include:

- Incorporating more complex noise models beyond the Gaussian assumption made here, to better capture noise characteristics in real astronomical images. Possible options include Poisson or mixed Poisson-Gaussian models.
- Exploring the use of learned priors like those based on variational autoencoders, which can potentially capture complex galactic morphologies more accurately than the handcrafted horseshoe prior.
- Applying the framework to analyze real observed gravitational lens data, which poses additional challenges like unknown point spread functions and other imperfections absent in simulated data. This can help validate effectiveness on real-world observations.
- Combining the lens inversion method with techniques for modeling the mass distribution of the foreground lensing galaxies. This allows reconstructing the lensing

potential and underlying total mass profile jointly with inferring the background source galaxy.
- Adapting the framework for other generalized inverse problems in astronomical imaging, such as deconvolution, superresolution, and inpainting, which exhibit similar mathematical structure.

Pursuing these directions can help improve reconstruction accuracy, enhance applicability to real data, and expand the scope and impact of the proposed methodology on astronomical imaging problems involving ill-posed inversions. With its robustness and computational efficiency, this approach shows potential for becoming a valuable tool for analyzing gravitational lensing observations to study distant galaxies.

References

---

[1] N Ahuja, S Lertrattanapanich, and NK Bose. “Properties determining choice of mother wavelet”. In: *IEE Proceedings-Vision, Image and Signal Processing* 152.5 (2005), pp. 659–664.

[2] C Andrieu, A Doucet, and CP Robert. “Computational advances for and from Bayesian analysis”. In: *Statistical science* (2004), pp. 118–127.

[3] M Antonini. “M. Barlaud, P”. In: *Mathieu, and l. Daubechies." Image Coding USing Wavelet TranSiOT", IEEE Transon Image Processing* 1 (1992), pp. 205–220.

[4] Marc Antonini et al. “Image coding using wavelet transform.” In: *IEEE Trans. Image Processing* 1 (1992), pp. 20–5.

[5] Paolo Baldi et al. “Asymptotics for spherical needlets”. In: (2009).

[6] Matthias Bartelmann and Peter Schneider. “Weak gravitational lensing”. In: *Physics Reports* 340.4-5 (2001), pp. 291–472.

[7] James O Berger and William H Jefferys. “The application of robust Bayesian analysis to hypothesis testing and Occam’s razor”. In: *Journal of the Italian Statistical Society* 1.1 (1992), pp. 17–32.

[8] RH Berry, Michael P Hobson, and S Withington. “Modal decomposition of astronomical images with application to shapelets”. In: *Monthly Notices of the Royal Astronomical Society* 354.1 (2004), pp. 199–211.

[9] Michael Betancourt. “A conceptual introduction to Hamiltonian Monte Carlo”. In: *arXiv preprint arXiv:1701.02434* (2017).

[10] Anindya Bhadra et al. “Lasso meets horseshoe”. In: *Statistical Science* 34.3 (2019), pp. 405–427.

[11] Steve Brooks et al. *Handbook of markov chain monte carlo*. CRC press, 2011.

[12] Olivier Cappé et al. “Population monte carlo”. In: *Journal of Computational and Graphical Statistics* 13.4 (2004), pp. 907–929.

[13] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC press, 2008.

[14] Rafael E Carrillo, Jason D McEwen, and Yves Wiaux. "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging". In: *Monthly Notices of the Royal Astronomical Society* 426.2 (2012), pp. 1223–1234.

[15] Carlos M Carvalho, Nicholas G Polson, and James G Scott. "Handling sparsity via the horseshoe". In: *Artificial intelligence and statistics*. PMLR. 2009, pp. 73–80.

[16] Carlos M Carvalho, Nicholas G Polson, and James G Scott. "The horseshoe estimator for sparse signals". In: *Biometrika* 97.2 (2010), pp. 465–480.

[17] Renyue Cen. "Gaussian random field: physical origin of Sersic profiles". In: *The Astrophysical Journal Letters* 790.2 (2014), p. L24.

[18] Hyunho Choi and Jechang Jeong. "Speckle noise reduction technique for SAR images using statistical characteristics of speckle noise and discrete wavelet transform". In: *Remote Sensing* 11.10 (2019), p. 1184.

[19] Richard T Cox. "Probability, frequency and reasonable expectation". In: *American journal of physics* 14.1 (1946), pp. 1–13.

[20] Arwa Dabbech, David Mary, and Chiara Ferrari. "Astronomical image deconvolution using sparse priors: An analysis-by-synthesis approach". In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2012, pp. 3665–3668.

[21] Ingrid Daubechies. "Orthonormal bases of compactly supported wavelets". In: *Communications on pure and applied mathematics* 41.7 (1988), pp. 909–996.

[22] Kun Deng et al. "An information-theoretic framework to aggregate a Markov chain". In: *2009 American Control Conference*. IEEE. 2009, pp. 731–736.

[23] Jan Drugowitsch. "Variational Bayesian inference for linear and logistic regression". In: *arXiv preprint arXiv:1310.5438* (2013).

[24] Frank Watson Dyson, Arthur Stanley Eddington, and Charles Davidson. "IX. A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of May 29, 1919". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 220.571-581 (1920), pp. 291–333.

[25] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Vol. 2. 1. Springer, 2010.

[26] Mário AT Figueiredo. "Adaptive sparseness for supervised learning". In: *IEEE transactions on pattern analysis and machine intelligence* 25.9 (2003), pp. 1150–1159.

[27] A Galan et al. "SLITRONOMY: Towards a fully wavelet-based strong lensing inversion technique". In: *Astronomy & Astrophysics* 647 (2021), A176.

[28] Aymeric Galan et al. "Using wavelets to capture deviations from smoothness in galaxy-scale strong lenses". In: *Astronomy & Astrophysics* 668 (2022), A155.

[29] Ioannis Gkioulekas, Georgios Evangelopoulos, and Petros Maragos. "Spatial Bayesian surprise for image saliency and quality assessment". In: *2010 IEEE international conference on image processing*. IEEE. 2010, pp. 1081–1084.

[30] Chris Hans. "Model uncertainty and variable selection in Bayesian lasso regression". In: *Statistics and Computing* 20 (2010), pp. 221–229.

[31] Yashar D Hezaveh et al. "Detection of lensing substructure using ALMA observations of the dusty galaxy SDP. 81". In: *The Astrophysical Journal* 823.1 (2016), p. 37.

[32] Matthew D Hoffman, Andrew Gelman, et al. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623.

[33] Matthias Holschneider et al. "A real-time algorithm for signal analysis with the help of the wavelet transform". In: *Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987*. Springer. 1990, pp. 286–297.

[34] Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

[35] Daniel TL Lee and Akio Yamamoto. "Wavelet analysis: theory and applications". In: *Hewlett Packard journal* 45 (1994), pp. 44–44.

[36] Boris Leistedt et al. "3D weak lensing with spin wavelets on the ball". In: *Physical Review D* 92.12 (2015), p. 123010.

[37] Boris Leistedt et al. "S2LET: A code to perform fast wavelet analysis on the sphere". In: *Astronomy & Astrophysics* 558 (2013), A128.

[38] Stephane G Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". In: *IEEE transactions on pattern analysis and machine intelligence* 11.7 (1989), pp. 674–693.

[39] Jason D McEwen and Matthew A Price. "Scale-discretised ridgelet transform on the sphere". In: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE. 2019, pp. 1–5.

[40] A Men'Shchikov et al. "A multi-scale, multi-wavelength source extraction method: getsources". In: *Astronomy & Astrophysics* 542 (2012), A81.

[41] Florent Mertens and Andrei Lobanov. "Wavelet-based decomposition and analysis of structural patterns in astronomical images". In: *Astronomy & Astrophysics* 574 (2015), A67.

[42] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[43] Steven T Myers. "Scaling the universe: Gravitational lenses and the Hubble constant". In: *Proceedings of the National Academy of Sciences* 96.8 (1999), pp. 4236–4239.

[44] Yasushi Nagano and Koji Hukushima. "Effect of global shrinkage parameter of horseshoe prior in compressed sensing". In: *arXiv preprint arXiv:2306.02607* (2023).

[45] Radford M Neal et al. “MCMC using Hamiltonian dynamics”. In: *Handbook of markov chain monte carlo* 2.11 (2011), p. 2.

[46] Trevor Park and George Casella. “The bayesian lasso”. In: *Journal of the American Statistical Association* 103.482 (2008), pp. 681–686.

[47] Matthew A Price et al. “Sparse Bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere”. In: *Monthly Notices of the Royal Astronomical Society* 500.4 (2021), pp. 5436–5452.

[48] Alexandre Refregier. “Shapelets—I. A method for image analysis”. In: *Monthly Notices of the Royal Astronomical Society* 338.1 (2003), pp. 35–47.

[49] Frédéric Rué and Albert Bijaoui. “A multiscale vision model to analyse field astronomical images”. In: *Experimental Astronomy* 7 (1997), pp. 129–160.

[50] Leonard J Savage et al. “On the foundations of statistical inference: Discussion”. In: *Journal of the American Statistical Association* 57.298 (1962), pp. 307–326.

[51] Peter Schneider, Jürgen Ehlers, and Emilio Falco. “Gravitational Lenses XIV”. In: Vol. XIV (Jan. 1992). DOI: `10.1007/978-1-4612-2756-4`.

[52] Peter Schneider et al. “Derivation of the lens equation”. In: *Gravitational Lenses* (1992), pp. 119–155.

[53] JL Sérsic. “Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy”. In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6 (1963), pp. 41–43.

[54] Michael D Seymour and Lawrence M Widrow. “Multiresolution analysis of substructure in dark matter halos”. In: *The Astrophysical Journal* 578.2 (2002), p. 689.

[55] Amirmohammad Shamaei, Jana Starcukova, and Zenon Starcuk Jr. “Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data”. In: *Computers in Biology and Medicine* 158 (2023), p. 106837.

[56] Mark J Shensa et al. “The discrete wavelet transform: wedding the a trous and Mallat algorithms”. In: *IEEE Transactions on signal processing* 40.10 (1992), pp. 2464–2482.

[57] Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial.* OUP Oxford, 2006.

[58] J-L Starck and Fionn Murtagh. “Astronomical image and data analysis”. In: (2007).

[59] J-L Starck et al. “Sparsity and the Bayesian perspective”. In: *Astronomy & Astrophysics* 552 (2013), A133.

[60] Jean-Luc Starck and Jérôme Bobin. “Astronomical Data Analysis and Sparsity: From Wavelets to Compressed Sensing”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 1021–1030. DOI: `10.1109/JPROC.2009.2025663`.

[61] Jean-Luc Starck, Jalal Fadili, and Fionn Murtagh. "The Undecimated Wavelet Decomposition and its Reconstruction". In: *IEEE Transactions on Image Processing* 16.2 (2007), pp. 297–309. DOI: `10.1109/TIP.2006.887733`.

[62] Jean-Luc Starck, Jalal Fadili, and Fionn Murtagh. "The undecimated wavelet decomposition and its reconstruction". In: *IEEE transactions on image processing* 16.2 (2007), pp. 297–309.

[63] Jean-Luc Starck, Fionn Murtagh, and Jalal M Fadili. *Sparse image and signal processing: wavelets, curvelets, morphological diversity.* Cambridge university press, 2010.

[64] Jean-Luc Starck, Fionn D Murtagh, and Albert Bijaoui. *Image processing and data analysis: the multiscale approach.* Cambridge University Press, 1998.

[65] SH Suyu et al. "Dissecting the gravitational lens B1608+ 656. I. Lens potential reconstruction". In: *The Astrophysical Journal* 691.1 (2009), p. 277.

[66] Greg B Taylor, Chris Luke Carilli, and Richard A Perley. "Synthesis imaging in radio astronomy II". In: *Synthesis Imaging in Radio Astronomy II* 180 (1999).

[67] Michael E Tipping. "Sparse Bayesian learning and the relevance vector machine". In: *Journal of machine learning research* 1.Jun (2001), pp. 211–244.

[68] Ignacio Trujillo, Alister W Graham, and Nicola Caon. "On the estimation of galaxy structural parameters: the Sérsic model". In: *Monthly Notices of the Royal Astronomical Society* 326.3 (2001), pp. 869–876.

[69] Michael Unser. "Splines: A perfect fit for signal and image processing". In: *IEEE Signal processing magazine* 16.6 (1999), pp. 22–38.

[70] SJ Warren and Simon Dye. "Semilinear gravitational lens inversion". In: *The Astrophysical Journal* 590.2 (2003), p. 673.

[71] Kenneth C Wong et al. "H0LiCOW–XIII. A 2.4 per cent measurement of H 0 from lensed quasars: 5.3 σ tension between early-and late-Universe probes". In: *Monthly Notices of the Royal Astronomical Society* 498.1 (2020), pp. 1420–1439.

[72] Jingxiang Yang et al. "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution". In: *Remote sensing* 11.13 (2019), p. 1557.

[73] Yong Zhang, Jie Jiang, and Guangjun Zhang. "Compression of remotely sensed astronomical image using wavelet-based compressed sensing in deep space exploration". In: *Remote Sensing* 13.2 (2021), p. 288.

Appendix A

A.1. Gradient Descent-Based Regularization Framework for Galaxy Image Reconstruction

In the context of the optimization framework delineated within the code, the objective is to identify an optimal solution $\mathbf{s}$ subject to regularization constraints, which are encapsulated within distinct prior models. These priors are predicated on various mathematical norms and functions that are intrinsically linked to the inherent characteristics of the galexy image being reconstructed.

The optimization procedure is instantiated via an iterative algorithm designed to minimize a cost function composed of a data fidelity term and a regularization term. The fidelity term, denoted as $\mathcal{L}(\mathbf{D}, \mathbf{s})$, measures the consistency of the solution with the observed data (the lensed image, $\mathbf{D}$), whereas the regularization term, denoted as $\mathcal{R}(\mathbf{s})$, imposes a priori constraints on the solution to encode desirable properties such as sparsity or smoothness.

The optimization algorithm operationalizes the Stochastic Gradient Descent (SGD) method, parameterized by a learning rate and momentum. The algorithm proceeds by iteratively updating the solution $\mathbf{s}$ in the direction that most steeply decreases the cost function. At each iteration i, the cost function $\mathcal{C}$ is evaluated, which is a weighted sum of the data fidelity term and the regularization term, formally represented as:

$$\mathcal{C}_i = \mathcal{L}(\mathbf{D}, \mathbf{s}_i) + \lambda \mathcal{R}(\mathbf{s}_i)$$

Here, λ is a regularization factor that controls the trade-off between the fidelity to the data and the regularization constraints. This iterative process is governed by a predefined number of iterations.

Upon completion of the iterations, the resultant solution $\mathbf{s}$ is transformed into a visual representation and saved for further qualitative analysis. The success of the reconstruction is quantified by a metric such as the mean squared error (MSE) between the reconstructed image and the ground truth image, which provides a statistical measure of the reconstruction accuracy.

The specific choice of regularization priors and the optimization methodology are reflective of the underlying assumptions about the image characteristics and the nature of the noise or uncertainties in the observed data.

A.1.1. Regularization Techniques and Implementation Details

The regularization functions embedded within the optimization framework play a pivotal role in encoding prior knowledge about the solution's characteristics. Each regularization function is designed to promote certain features in the solution $\mathbf{s}$, based on the underlying physical or statistical model of the data.

Horseshoe Regularization: The horseshoe prior is a heavy-tailed distribution that is particularly suited for problems where the signal-to-noise ratio is low. It effectively shrinks small coefficients towards zero while leaving large coefficients relatively unchanged, acting as a bridge between the L1 and L2 norms. This makes it appropriate for scenarios with signals that have a few significant coefficients amidst many insignificant ones.

The optimization algorithm for image reconstruction is implemented in a Python environment leveraging the PyTorch library, which provides a dynamic computational graph that facilitates automatic differentiation. The key components of the implementation encompass the regularization functions, the optimization procedure, and the convergence monitoring mechanisms. The starlet transform is carried out through [27] which is designed for efficient multi-resolution analysis.

The algorithm employs Stochastic Gradient Descent (SGD) with momentum of 0.9 to optimize the solution. The solution $\mathbf{s}$ is initialized as a trainable parameter with Gaussian random values. The optimization loop runs for a 5000 number of iterations. Figure A.1 demonstrates the effectiveness of the minimization method in image reconstruction for galaxy 4441, utilizing the horseshoe prior for enhanced handling of sparsity and noise reduction when the MSE of it is 71.41.

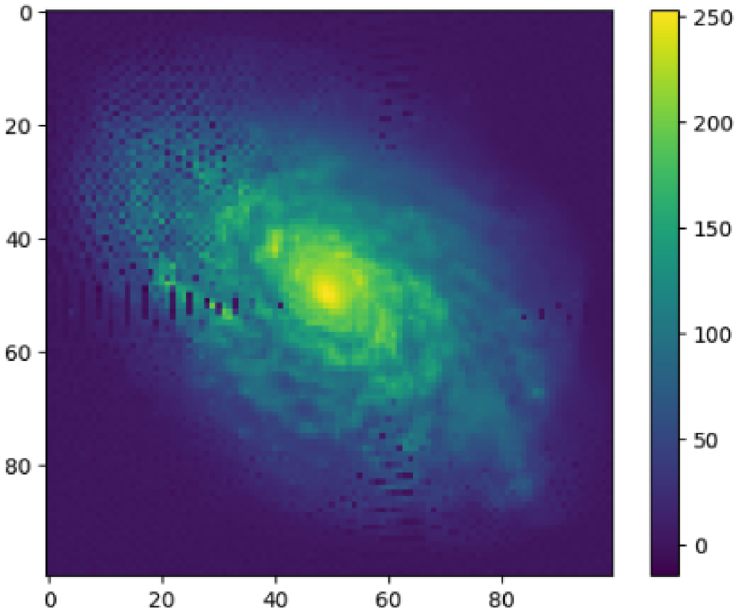


Fig. A.1. Galaxy 4441 Reconstructed image employing the horseshoe prior with optimization.

Appendix B

Appendix B

B.1. Detailed Analysis of Gravitational Lensing Inference

In this thesis, as mentioned in subsection 2.2.2, the data generating equation is the foundational equation in this process as follows:

$$D = LS + N \tag{B.1.1}$$

Here, L is a known gravitational lensing distortion matrix, S represents the value of the pixels of the background source we aim to infer, and N is a vector of random noise from a known normal distribution. The ground-truth background source is denoted as S_{True}.

In a Bayesian framework, the posterior of S given observations D is expressed as:

$$P(S|D) = \frac{P(D|S)P(S)}{P(D)} \tag{B.1.2}$$

From this formulation, we derive that:

$$P(S|D) \propto \frac{1}{P(D)}P(D|S) + P(S) \tag{B.1.3}$$

The likelihood $P(D|S)$ is Gaussian due to Gaussian noise in the data, and similarly, the prior $P(S)$ is Gaussian. Thus, the posterior probability distribution is also Gaussian. The log posterior can be expressed as:

$$\log P(S|D) \propto \log P(D|S) + \log P(S) \tag{B.1.4}$$

Further, we calculate $\log P(D|S)$ and $\log P(S)$ separately:

$$\log P(D|S) = \log\left(\det(2\pi C_N)^{-1/2}\exp\left(-\frac{1}{2}(D - LS)^T C_N^{-1}(D - LS)\right)\right) \tag{B.1.5}$$

Assuming $\det(2\pi C_N)^{-1/2}$ as a constant:

$$\log P(D|S) = -\frac{1}{2}(D - LS)^T C_N^{-1}(D - LS) \tag{B.1.6}$$

For the Gaussian prior $P(S)$ from Equation B.1.4:

$$\log P(S) = \log\left(\det(2\pi C_S)^{-1/2} \exp\left(-\frac{1}{2}S^T C_S^{-1} S\right)\right) \tag{B.1.7}$$

Simplified to:

$$\log P(S) = -\frac{1}{2}S^T C_S^{-1} S \tag{B.1.8}$$

In determining S, we maximize the log posterior by differentiating it with respect to S and solving for S:

$$P(S|D) = -e^{-\frac{1}{2}}\left[(D - LS)^T C_N^{-1}(D - LS)\right] - e^{-\frac{1}{2}}\left(S^T C_S^{-1} S\right) \tag{B.1.9}$$

Setting this derivative to zero provides the equation for S_{map}:

$$(L^T C_N^{-1} L + C_S^{-1})S = L^T C_N^{-1} D \tag{B.1.10}$$

Defining A and B for simplification:

$$A = L^T C_N^{-1} L + C_S^{-1} \tag{B.1.11}$$

$$B = L^T C_N^{-1} D \tag{B.1.12}$$

Thus, we obtain:

$$S = A^{-1} B \tag{B.1.13}$$

B.2. Example and Practical Implementation

This section includes a practical example demonstrating the reconstruction process of a background galaxy 4441 image as shown in Figure B.1, showcasing the transformation of S and D, the addition of Gaussian noise, and the creation of D_{noisy}. Various optimization techniques, including JAX for GPU acceleration and PyTorch for gradient descent, are employed to achieve accurate reconstructions.

We plan to reconstruct the galaxy image by applying the equation from the last section, in particular Equation B.1.13. The scaling of the C_s matrix will be varied using different values of λ, defined as $C_s = \lambda \times \text{eye}()$. In Figure B.1, we set λ to 0.01 and consider C_n as a diagonal matrix where each element is one. Our aim is to reduce the residuals derived from the subsequent equation:

$$Residuals = D_{\text{noisy}} - L \times S_{\text{map}}$$

Following this, we present in Figure B.1 the outcomes of our reconstruction process. The left panel of the image showcases the reconstructed galaxy 4441, achieved through the application of the discussed techniques and parameters. The right panel, in contrast, displays the residuals, providing a visual representation of the difference between the reconstructed image and the original data. The mean squared error (MSE) for this reconstruction is calculated to be 0.9354.

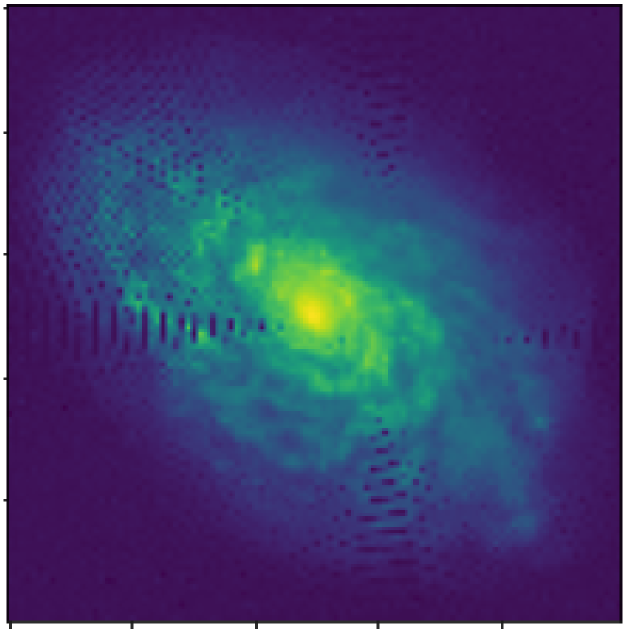
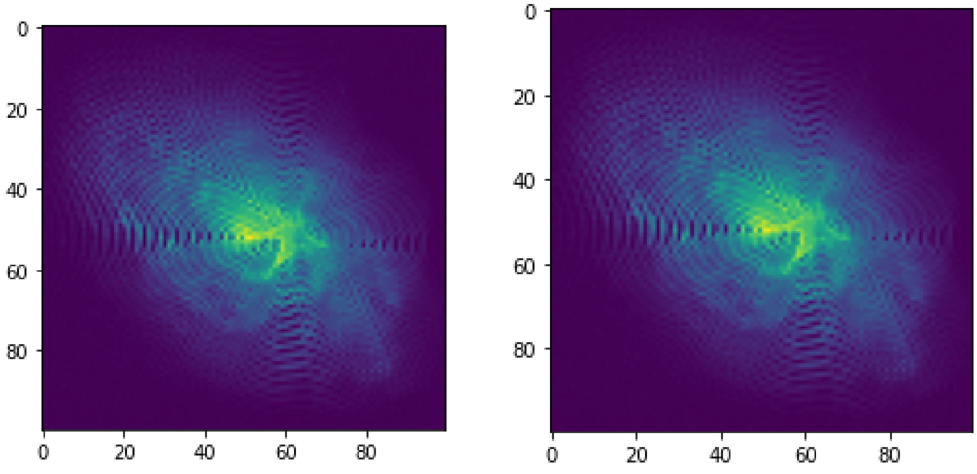


Fig. B.1. (a) Reconstructed Galaxy Image using optimized parameters with $\lambda = 0.01$ and a diagonal C_n, (b) Corresponding residuals illustrating the variance from the original data

To reduce the residuals, we implemented function optimization using a downhill optimizer and calculated the mean squared error (MSE) to improve it. This process was time-intensive, exceeding 3 hours. To enhance efficiency, I utilized JAX for GPU-based optimization of the posterior and transformed the matrices L and S into JAX sparse matrices (BCOO format). However, the multiplication of these matrices could not be performed due to the lack of implementation for this operation in JAX.

Furthermore, rather than employing linear algebraic methods to determine S_{map}, a numerical approach was adopted. Utilizing the gradient descent method in PyTorch, L and S_{map} were converted into sparse matrices, enabling their multiplication. This code was then adapted for GPU execution using Cuda. In Figure B.2, the left panel is labeled as (a), where the initial point was a random matrix S_{True}, whereas in figure (b), it started from a matrix of zeros. These processes were iterated with varying values of lambda and iteration count,



(a) (b)

Fig. B.2. (a) depicts the reconstructed galaxy image starting from a random matrix, (b) the reconstruction beginning with a zero matrix

yielding the most effective results under normal noise conditions at 8000 iterations and a lambda value of 0.01.