Full-Field cosmological inference with weak lensing : from automatic differentiation to neural density estimation.

Denise Lanzieri

▶ **To cite this version:**

Denise Lanzieri. Full-Field cosmological inference with weak lensing : from automatic differentiation to neural density estimation.. Astrophysics [astro-ph]. Université Paris Cité, 2023. English. NNT : 2023UNIP7277 . tel-04736107

HAL Id: tel-04736107

https://theses.hal.science/tel-04736107v1

Submitted on 14 Oct 2024

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

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

Université Paris Cité

Université Paris cité

Ecole doctorale:
Astronomie et Astrophysique d'Ile de France n. 127
Laboratoire:
Laboratoire de Cosmologie et Statistiques
Commissariat à l'Energie Atomique et aux Energies Alternatives

Thèse de doctorat en Astronomie et Astrophysique

Full-Field cosmological inference with weak lensing: from automatic differentiation to neural density estimation

Par **Denise Lanzieri**

Présentée et soutenue publiquement le 6 Octobre 2023

Directeur de thèse: Jean-Luc Starck, Directeur de Recherche, CEA

Composition du Jury:

Alan Heavens	Professeur, Imperial College London	Président
Roberto Trotta	Professeur, Imperial College London, SISSA	Rapporteur
Anne Ealet	Directrice de Recherche, CNRS, IP2I Lyon	Examinatrice
Vincenzo Fabrizio Cardone	Chercheur, Astronomical Observatory of Rome	Examinateur
François Lanusse	Chercheur, CNRS, CEA Paris-Saclay	Membre invité

Considerate la vostra semenza:

fatti non foste a viver come bruti ma per seguir virtute e canoscenza.

— Dante Alighieri,
Inferno, Canto XXVI, vv.118-120

Abstract

Full-Field cosmological inference with weak lensing: from automatic differentiation to neural density estimation

The upcoming stage-IV Dark Energy surveys, such as Euclid and LSST, will observe the Universe with unprecedented accuracy, allowing us to investigate fundamental problems in cosmology. These surveys will use weak gravitational lensing as one of the main probes to investigate the origin of the accelerated expansion of the Universe and the properties of its dark matter component. However, traditional cosmological inference for weak gravitational lensing has two important limitations: First, the two-point statistics do not fully extract the non-Gaussian information from cosmological data. Second, even for the two-point statistics, writing down an accurate model for the likelihood function can be very arduous (small-scale uncertainties, non-Gaussian signals, etc.). In recent years, it has been shown that statistics of order higher than a second can help to access non-Gaussian information, nevertheless, these approaches are characterized by the absence of analytical models to describe the observed signal and require calibrating the cosmology inference from weak lensing simulations. One possible way to circumvent an explicit likelihood consists in using Likelihood-free inference methods. These methods estimate posterior distributions through forward modeling of simulated data. Alternatively, Bayesian forward-modeling methods can be used, which integrate observations into a forward model, enabling the exact reconstruction of the likelihood.

All the methodologies and developments in this thesis work towards two major goals: making fast approximated simulations suitable for the data analysis pipeline of upcoming cosmological surveys and investigating forward modeling techniques to exploit the potential of next-generation weak lensing data.
In this context, we have developed and validated the Differentiable Lensing Lightcone (DLL) package within the LSST framework. DLL is a fully automatically differentiable

physical model designed for fast inference, aiming to achieve high accuracy with low computational costs. The DLL tool is designed to be used as a forward model in Bayesian inference algorithms requiring access to the derivatives of the likelihood of the model. We have also developed a new correction scheme to enhance the accuracy of quasi-N-body simulations, aiming to replicate the precision of high-resolution N-body simulations. Finally, we investigate the performance of different procedures to optimally extract informative summaries obtained from mock weak lensing mass maps compressed using Convolution Neural Networks.

Keywords: Weak gravitational lensing; cosmology; higher order statistics; cosmological simulations; Dark energy; automatic differentiation;

Résumé de la thèse en français

Inférence cosmologique plein champ avec lentillage faible: de la différenciation automatique à l'estimation neuronale de la densité

Les prochaines études sur l'énergie noire de phase IV, telles que Euclid et LSST, observeront l'Univers avec une précision sans précédent, ce qui nous permettra d'étudier des questions non résolues dans la cosmologie actuelle. Ces études utiliseront l'effet de lentille gravitationnelle faible comme l'une des principales sondes pour étudier l'origine de l'expansion accélérée de l'Univers et les propriétés de sa composante de matière noire. Cependant, l'inférence cosmologique traditionnelle pour l'effet de lentille gravitationnelle faible présente deux limites importantes : Premièrement, les statistiques en deux points ne permettent pas d'extraire pleinement les informations non gaussiennes des données cosmologiques. Deuxièmement, même pour les statistiques en deux points, l'élaboration d'un modèle précis pour la fonction de vraisemblance peut être très ardue (incertitudes à petite échelle, signaux non gaussiens, etc.). Ces dernières années, il a été démontré que les statistiques d'ordre supérieur à deux peuvent aider à accéder à l'information non gaussienne. Néanmoins, ces approches sont caractérisées par l'absence de modèles analytiques pour décrire le signal observé et nécessitent de calibrer l'inférence cosmologique à partir de simulations de lentillage faible. Une façon de contourner une vraisemblance explicite consiste à utiliser des méthodes d'inférence sans vraisemblance. Ces méthodes estiment les distributions postérieures par le biais d'une modélisation à partir des données simulées. Il est également possible d'utiliser des méthodes bayésiennes de modélisation, qui intègrent les observations dans un modèle, permettant la reconstruction exacte de la vraisemblance.

Toutes les méthodologies et les développements de cette thèse visent deux objectifs principaux : faire des simulations rapides et approximatives adaptées au chaîne de traitement et d'analyse des données des prochaines études cosmologiques et étudier les techniques de modélisation pour exploiter le potentiel des données de lentillage faible

de la prochaine génération.
Dans ce contexte, nous avons développé et validé le paquet Differentiable Lensing Lightcone (DLL) dans le cadre du LSST. DLL est un modèle physique différentiable entièrement automatique conçu pour de l'inférence rapide, visant à atteindre une grande précision avec de faibles coûts de calcul. L'outil DLL est conçu pour être utilisé comme modèle direct dans les algorithmes d'inférence bayésienne nécessitant l'accès aux dérivées de la vraisemblance du modèle. Nous avons également développé un nouveau schéma de correction pour améliorer la précision des simulations d'un quasi-système à N corps, visant à reproduire la précision des simulations d'un système à N corps à haute résolution. Enfin, nous étudions les performances de différentes procédures pour extraire de manière optimale les résumés informatifs obtenus à partir de cartes de masse de faibles lentilles fictives comprimées à l'aide de réseaux neuronaux de convolution.

Mots clés: Lentillage gravitationnel faible; Statistiques d'ordre supérieur; Paramètres cosmologiques; Grandes structures de l'Univers; Simulations cosmologiques; Énergie noire; Différenciation automatique;

Introduction

The typical description of the Universe's evolution encompasses its transformation from a homogeneous state following the Big Bang to the currently observed hierarchical structure, including galaxies, galaxy clusters, and superclusters. This description is typically based on our current understanding of the initial condition of the Universe and the nature of gravity, as well as on the assumption of the existence of two components: *dark energy* and *dark matter*. According to current cosmology, dark energy represents 68% of the energy density of the Universe and is the cause of the accelerating expansion of the Universe, while dark matter constitutes 27% of the energy density. However, to date, there are no incontrovertible theoretical explanations or undeniable evidence for the nature of dark energy and dark matter.

In the absence of a trusted theoretical guide, observational exploration becomes the primary tool to understand the composition of the Universe. The upcoming Stage-IV galaxy experiments, such as the Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory, the Euclid mission by the European Space Agency, and the Roman Space Telescope, will try to answer some of these questions, primarily focusing on observing and studying the effects of dark energy and dark matter. Among the key probes used in these future galaxy surveys is weak gravitational lensing. Gravitational lensing refers to the phenomenon where massive objects located between an observer and background sources cause a curvature in space-time, resulting in a deviation of light coming from these background sources. These deviations subsequently distort the original image of the source, providing a powerful means to investigate the distribution of total matter responsible for the distortion. When these deformations are very small, it is referred to as the *weak lensing* regime.

We need to develop a statistical strategy to extract cosmological information from weak lensing observations. This is typically done by selecting specific summary statistics and comparing them with the predictions of a given model. Traditional weak lensing summary statistics rely on two-point statistics, namely the two-point correlation function of the data or its Fourier transform, known as the power spectrum. Although widely used in past analyses, we are aware of their limitations. Indeed, it is known that second-order statistics fail to fully extract the non-Gaussian informa-

tion from cosmological data. This has led to the introduction of several higher-order statistics, such as the weak lensing one-point PDF, lensing peak counts, Minkowski functionals, moments of mass maps, wavelet and scattering transforms, and 3-point statistics. Recently, machine learning-based methods and Bayesian forward-modeling frameworks have also been introduced to attempt to fully account for the non-Gaussian content in the weak lensing signal.
We will discuss some of these statistics in more detail in the following chapters, but we might already be wondering, which of these options is the best, enabling us to maximize the amount of information extracted from this data. Next-generation surveys will provide measurements of billions of galaxy shapes with unprecedented accuracy, accessing highly non-linear scales and generating vast amounts of data. In general, the process of extracting cosmological information from input data can be seen as comprising several tasks, each with its associated challenges. Specifically, throughout this thesis, we will ask: Which statistics most effectively constrain our cosmological parameters? What is the optimal inference strategy? How can we compress our input data without losing sensitive information? How can we address the fundamental limits of computation?

The primary goal of this thesis is to attempt to answer some of these questions and to develop a physical inference framework suitable for a stage-IV dark energy survey, with the aim of fully exploiting the potential of next-generation data.
The content of this thesis is structured as follows: in the first part, composed of chapter 1, chapter 2, and chapter 3, I provide the theoretical tools required to understand the rest of the thesis. Specifically, in chapter 1, I introduce the standard model of cosmology, the homogeneous Universe, the inhomogeneous cosmic web, the consequences of expansion, and the principles of formation and growth of structures. chapter 2 outlines the formalism of numerical simulations, starting by describing in more detail the growth of structures from initial conditions to the observed matter distribution that we observe today and concluding by presenting the state of the art of current cosmological simulations. chapter 3 delves into the weak lensing formalism, second-order summary statistics, examples of statistics beyond the second order, and various systematic effects related to weak lensing observables. In chapter 4, I introduce some useful Bayesian statistical tools necessary for chapter 7.

In chapter 5 and chapter 6 I describe the contributions of my PhD work. Specifically, in chapter 5, I present the results of a work in which we developed a new correction scheme for quasi-N-body simulations aimed at mimicking the precision of high-resolution N-body simulations. Indeed, despite being fast and having low computational costs, quasi-N-body schemes lack force resolution on small scales and provide halo profiles that are less sharp than their full N-body counterparts. To overcome this issue, we developed a new Hybrid Physical-Neural approach (HPN) that introduces an additional force to the PM gravitational potential, parameterized by a Fourier Neural Network. We presented a proof of concept of this method in Lanzieri et al. (2022) at the machine learning for astrophysics workshop during the 39th International Conference on Machine Learning (ICML 2022).

In chapter 6, I presented the Differentiable Lensing Lightcone (DLL) package, a differentiable weak lensing simulator designed to efficiently compute gradients for the development of new inference algorithms in weak lensing surveys. To achieve this, we extended the framework of the FlowPM package (Modi et al., 2021) by implementing the Born approximation and simulating lensing lightcones using the TensorFlow framework. The weak lensing package enables the inclusion of systematics, such as intrinsic alignment, and allows for the computation of nonlinear lensing convergence maps at different source redshifts. This work was presented in the paper: Lanzieri et al. (2023).

In chapter 7 illustrates my current work. Specifically, in chapter 7, I present a comparison of the performance of various neural network-based compression schemes within the Likelihood-Free Inference (LFI) framework. These methods differ in terms of the loss functions used to train the neural network, but they share the same inference strategy based on neural density estimation. To address the need for result validation in the LFI context, we employ a Bayesian forward modeling approach that enables inference on the joint posterior of the cosmological parameters and the convergence field.

In chapter 8, I summarize our results and conclusions.

Résumé étendu en français

Nous avons l'habitude de décrire l'évolution de l'univers depuis un état homogène après le Big Bang qui va ensuite former des structures hiérarchiques et créer les objets que l'on observe aujourd'hui, avec notament les galaxies, les amas de galaxies et les superamas. Cette description est typiquement basée sur notre compréhension actuelle de l'état initial de l'univers et de la nature de la gravité, ainsi que sur l'hypothèse de l'existence de deux composantes: *l'énergie noire* et *la matière noire*. Selon le modèle standard actuel de la cosmologie (ΛCDM), l'énergie noire représente 68% de la densité d'énergie de l'Univers et est la cause de l'expansion de l'Univers, tandis que la matière noire constitue 27% de la densité d'énergie. Cependant, à ce jour, il n'existe pas d'explications théoriques incontestables ni de preuves indéniables de la nature de l'énergie noire et de la matière noire.

En l'absence d'un guide théorique fiable, l'exploration observationnelle devient l'outil principal pour comprendre la composition de l'Univers. Les prochaines expériences sur les galaxies de la phase IV, telles que le *Legacy Survey of Space and Time* (LSST) à l'Observatoire Vera C. Rubin, la mission *Euclid* de l'Agence spatiale européenne et le télescope spatial *Roman*, tenteront de répondre à certaines de ces questions, en se concentrant principalement sur l'observation et l'étude des effets de l'énergie noire et de la matière noire. Parmi les sondes clés utilisées dans ces futurs relevés de galaxies figure l'effet de lentille gravitationnelle faible. L'effet de lentille gravitationnelle désigne le phénomène par lequel des objets massifs situés entre un observateur et des sources d'arrière-plan provoquent une courbure de l'espace-temps, ce qui entraîne une déviation de la lumière provenant de ces sources d'arrière-plan. Ces déviations déforment ensuite l'image originale de la source, ce qui constitue un moyen puissant d'étudier la distribution de la matière totale responsable de la déformation. Lorsque ces déformations sont très faibles, on parle du régime *de lentille faible*.

Une fois que nous disposons des observations, nous devons développer une stratégie statistique pour en extraire des informations cosmologiques. Cela se fait généralement en sélectionnant des statistiques spécifiques et en les comparant aux prédictions d'un modèle donné. Les statistiques traditionnelles de lentille gravitationelle reposent sur des fonctions de corrélations à deux points ou sa transformée de Fourier, connue sous

le nom de spectre de puissance. Cependant, les statistiques à deux points ne sont optimales que pour les champs de densité gaussiens et ne rendent pas pleinement compte de l'information non gausienne du signal de lentille graviationelle faible aux échelles auxquelles les futures études pourront accéder (par exemple, l'information encodée dans les pics et dans les filaments de la distribution de la matière). Cela a conduit à l'introduction d'un certain nombre de statistiques d'ordre supérieur pour accéder à l'information non gaussienne des données de lentille gravitationnelle faible: fonction de distribution de probabilité (PDF en anglais) en un point de l'effet de lentille gravitationnelle faible, le nombre de pics dans les cartes de convergence, les fonctionnelles de Minkowski, les moments des cartes de masse, les transformations et diffusions en ondelettes et les fonctions de corrélations à trois points. Récemment, des cadres de modélisation bayésiens et méthodes basés sur l'apprentissage automatique ont également été introduits pour tenter de prendre pleinement en compte le contenu non gaussien du signal de lentille gravitationelle faible. Contrairement aux méthodes décrites ci-dessus, ces approches sont conçues pour accéder à l'ensemble du contenu informatif compris dans le champ de densité. Même si ces techniques sont théoriquement asymptotiquement optimales en termes de récupération de l'information, elles souffrent encore de limitations significatives.

Nous discuterons de certaines de ces statistiques plus en détail tout au long de cette thèse, mais nous pouvons déjà nous demander laquelle de ces options est la meilleure, c'est à dire celle nous permettant de maximiser la quantité d'informations extraites de ces donnée? Les grands relevés de stage IV fourniront des mesures de milliards de formes de galaxies avec une précision sans précédent, générant d'énormes quantités de données et permettant d'inclure des échelles hautement non linéaires dans l'analyse. En général, le processus d'extraction d'informations cosmologiques à partir d'images du ciel peut être considéré comme comprenant plusieurs tâches, chacune avec ses défis associés. Plus précisément, tout au long de cette thèse, nous nous poserons les questions suivantes: Quelles statistiques contraignent le plus efficacement nos paramètres cosmologiques? Quelle est la stratégie d'inférence optimale? Comment pouvons-nous comprimer nos données d'entrée sans perdre d'informations essentielles? Comment pouvons-nous aborder les limites de ressources compuationelles?

L'objectif principal de cette thèse est de tenter de répondre à certaines de ces questions et de développer un cadre d'inférence physique adapté à une étude de l'énergie noire de phase IV, dans le but d'exploiter pleinement le potentiel des données de la prochaine génération de grands relevés.
Le contenu de cette thèse est structuré comme suit: dans la première partie, composée de chapitre 1, chapitre 2, chapitre 3, je fournirai les outils théoriques nécessaires à la compréhension de la thèse. Plus précisément, dans le chapitre 1, je présenterai le modèle standard de la cosmologie, l'Univers homogène, la toile cosmique inhomogène, les conséquences de l'expansion et les principes de formation et de croissance des structures. Le chapitre 2 expose les formalismes des simulations numériques, en commençant par décrire plus en détail la croissance des structures depuis les conditions initiales jusqu'à

la distribution de matière que nous observons aujourd'hui, et en concluant par une présentation des méthodes de référence des simulations cosmologiques actuelles. Le chapitre 3 approfondit le formalisme de lentille graviationelle faible, les statistiques de second ordre, des exemples de statistiques au-delà du second ordre, et divers effets systématiques liés aux observations de lentille gravitationelle faible.

Dans le chapitre 5 et dans le chapitre 6 je décris les contributions de mon travail de doctorat. En particulier, dans le chapitre 5, je présente les résultats d'un travail dans lequel nous avons développé un nouveau schéma de correction pour les simulations à approximatives à N corps visant à imiter la précision des simulations à N corps à haute résolution. Malgré leur rapidité et leur faible coût de calcul, les schémas d'approximation ài N-corps manquent de résolution aux petites échelles et fournissent des profils de halo moins nets que leurs équivalents N-corps complets. Pour résoudre ce problème, nous avons développé une nouvelle approche Hybride Physique-Neurale (HPN) qui introduit une force supplémentaire au potentiel gravitationnel, paramétrée par un réseau neuronal de Fourier. Cette approche combine la description physique de la formation des structures avec un réseau neuronal plus simple pour corriger les approximations faites dans les algorithmes pseudo-N-corps. Notre méthode présente des similitudes avec les travaux de Dai et al. (2018), qui a démontré comment améliorer la modélisation de la distribution de la matière dans les halos dans les simulations quasi-N-corps à l'aide d'un schéma basé sur le gradient. L'un des avantages de ces méthodes par rapport aux techniques d'apprentissage automatique complètes est la plus faible dimensionnalité de l'espace des paramètres à optimiser. Pour illustrer les mérites de notre approche, nous avons comparé les résultats des simulations pseudo-N-corps corrigées avec les simulations CAMELS à haute résolution. Nous avons comparé notre modèle au schéma PGD (Dai et al., 2018)), montrant que, pour la cosmologie de référence, les deux méthodes produisent des améliorations similaires dans le spectre de puissance à petite échelle, mais diffèrent de manière significative en termes de coefficients de corrélation. Nous avons également présenté les résultats d'expériences impliquant des changements de résolution et de paramètres cosmologiques. Ces expériences ont démontré que notre méthode est plus performante que le schéma PGD en termes de spectre de puissance et de coefficients de corrélation croisée, et qu'elle est moins sensible aux paramètres des simulations utilisées pour l'entraînemen du réseau de neurone. Nous avons présenté une preuve de concept de cette méthode dans Lanzieri et al. (2022) à l'atelier sur l'apprentissage automatique pour l'astrophysique pendant la *39e International Conference on Machine Learning (ICML 2022)*.

Dans le chapitre 6, je présente le modèle *Differentiable Lensing Lightcone* (DLL), un simulateur rapide de cône lumineux de lentille gravitationnelle faible donnant accès au gradient. Nous avons intégré au simulateur à N-corps publique FlowPM l'approximation de Born dans le cadre de Tensorflow pour créer des cartes de convergence non gaussiennes de l'effet de lentille gravitationnelle faible. Pour permettre à DLL de fonctionner à basse résolution sans affecter de manière significative la précision, nous complétons le code à N-corps FlowPM avec le schéma Neural Hybride-Physique. Nous validons notre

outil en comparant les statistiques de C_ℓ et du nombre de pics avec les prédictions des simulations κTNG-Dark. Pour ce faire, nous effectuons des simulations suivant l'évolution des particules 128^3 et nous produisons des cartes de convergence de lentille gravitationnelle faible pour plusieurs des sources à différents redshift. Nous montrons que, même avec un faible coût de calcul, nous obtenons des résultats comparables pour des redshifts égaux ou supérieurs à $z = 0.91$. Pour démontrer le potentiel de notre outil, dans un premier cas d'utilisation, nous exploitons la différentiabilité automatique des simulations pour faire des prévisions de Fisher. Grâce à la rétro-propagation, l'accès à la dérivée à travers les simulations en ce qui concerne les paramètres cosmologiques et le paramètre A_{IA} est possible sans coût de calcul supplémentaire. Dans un cadre similaire à celui de LSST, nous simulons des cartes de convergence de lentille gravitationnelle faible pour une source unique à redshift $z = 0.91$ et d'extension angulaire de $5°$, dans une boîte périodique de volume comobile égal à 205 h^{-1}Mpc. Nous calculons les contraintes sur les cartes de convergence résultantes avec les pics de comptage des étoiles et utilisons un spectre de puissance de lentille filtré par ondelettes comme point de repère pour la comparaison. Dans les limites des choix d'analyse faits dans cette étude, nous obtenons les résultats suivants:

- confirmons que les statistiques de comptage de pics surpassent les statistiques à deux points telles que cela a été montré dans Ajani et al. (2020), même dans un espace de paramètres cosmologiques de haute dimension avec de nombreux paramètres de nuisance.

- trouvons que les comptages de pics fournissent les contraintes les plus strictes sur l'amplitude de l'alignement intrinsèque des galaxies A_{IA}.

Le cadre décrit ici peut offrir de nombreux avantages dans le contexte de l'inférence des paramètres cosmologiques: c'est la première étape dans le développement d'algorithmes d'inférence entièrement différentiables pour l'effet de lentille gravitationnelle faible, c'est un outil rapide pour explorer davantage l'impact des erreurs systématiques sur les statistiques d'ordre supérieur. Ce travail a été présenté dans l'article: Lanzieri et al. (2023).

Dans le chapitre 7 je illustre mon travail actuel. Plus précisément, dans le chapitre 7, j'ai présenté des travaux en cours visant à comparer deux stratégies différentes basées sur des cartes de convergence pour déduire les paramètres cosmologiques: la stratégie explicite de champ complet, également connue sous le nom d'inférence hiérarchique bayésienne, basée sur un échantillonneur Hamiltonien Monte-Carlo (HMC), et la stratégie d'inférence implicite, également connue sous le nom d'inférence sans vraisemblance, basée sur un estimateur de densité neuronale.
Nous commençons par un constat: Les approches d'apprentissage profond pour l'inférence implicite comportent généralement deux étapes. La première étape est l'apprentissage automatique d'une statistique optimale à petite dimension, et la deuxième étape est l'utilisation d'un estimateur de densité neuronal en basse dimension pour construire une estimation $P_{\boldsymbol{\varphi}}$ de la fonction de vraisemblance $p(\boldsymbol{x}|\boldsymbol{\theta})$ (estimation neuronale de la

vraisemblance) ou de construire une estimation P_{φ} de la distribution de probabilité à posteriori $p(\boldsymbol{\theta}|\boldsymbol{x})$ (estimation postérieure neuronale).
On peut maintenant comprendre que ces deux étapes peuvent potentiellement avoir un impact sur les contraintes finales des paramètres d'intérêt.
Cela dit, la principale motivation de ce travail est d'évaluer l'impact d'une stratégie de compression données sur la distribution postérieure finale et, par conséquent, de déterminer s'il existe une stratégie de compression optimale. En outre, l'objectif est de démontrer qu'en utilisant cette stratégie, les méthodes implicites et explicites obtiennent la même distribution distribution de probabilité à posteriori.
Nous construisonsle modèle pour l'inférence explicite et simuler les données fictives nécessaires pour entraîner le modèle implicite, nous avons développé `SbiLens`, un simulateur de lentille gravitationnelle faible basé sur Jax et optimisé pour les applications d'inférence qui ont besoin d'accéder aux dérivés du modèle. Notre analyse est basée sur des données synthétiques de lentille gravitationnelle faible avec cinq bins tomographiques, imitant une étude de type LSST-Y10. Après avoir donné un aperçu des différentes stratégies de compression adoptées dans la littérature pour l'inférence sans vraisemblance et les stratégies d'inférence basées sur la vraisemblance, nous avons comparé l'impact de certaines d'entre elles sur les contraintes finales des paramètres cosmologiques pour un modèle ΛCDM. Nous avons obtenu les résultats préliminaires suivants :

1. Le *Mean Square Error* (MSE) et le *Maximum Absolute Error* (MAE) conduisent à des résultats en très bon accord, tandis que le *Variational Mutual Information Maximization* (VMIM) conduit à des contours sensiblement différents, en particulier pour le paramètre w_0. Nous avons quantifié ces résultats en examinant la figure de mérite et les statistiques marginalisées et avons constaté que:

2. nous utilisons le VMIM pour comprimer les données originales à haute dimension, nous comparons les distributions de probabilité à posteriori obtenus dans le cadre de l'inférence implicite avec ceux obtenus à partir de la modélisation hiérarchique bayésienne et du spectre de puissance. Nous démontrons que les deux approches fondées sur les cartes conduisent à une amélioration significative de la contrainte de Ω_c, w_0, σ_8 par rapport aux statistiques à 2 points. Cependant, h, n_s, Ω_b ne sont contraints par aucune des deux approches et sont dominés par la distribution de probabilité à priori.

3. Lorsque l'on utilise le VMIM pour comprimer les données originales à haute dimension, dans le cadre de notre application, les deux méthodes, c'est-à-dire l'inférence hiérarchique bayésienne et l'inférence sans vraisemblance, conduisent aux mêmes distributions de probabilité à posteriori.

Dans le chapitre 8, je résume nos résultats et nos conclusions.

Acknowledgments

I would like to express my gratitude to Alan Heavens, Roberto Trotta, Vincenzo Fabrizio Cardone, and Anne Ealet for accepting the invitation to join my thesis committee. Having you as part of my jury was a great honor. I would particularly like to express my gratitude to my reviewers Alan Heavens, and Roberto Trotta, for dedicating time to thoroughly read my manuscript and providing valuable comments to enhance it.

A Ph.D. journey is a milestone in a person's life. For some, it's a real trauma, while for others, it's the most beautiful period of their life. For me, it was both —a continuous alternation between peaks of joy and satisfaction and moments of anxiety and frustration. I would like to thank all the people who have been the source of "beauty", often the same ones who helped me survive the period of "trauma".

My deepest thanks go to my advisors, Jean-Luc and François. Not just for their careful and constant supervision over the three years, but also for their invaluable advice and guidance regarding my professional future. In particular, I want to thank Jean-Luc, because, over the past three years, I have witnessed a positive evolution in our relationship. Your mentorship has been invaluable, and I appreciate the trust and esteem you have shown me. Your encouragement has played a significant role in my academic and personal growth, and I am truly grateful for the opportunities you've provided.

I want to express my sincere gratitude for François' support and guidance during my Ph.D. journey. Your patience, understanding, and empathy have been instrumental in my growth as a researcher and as a person. I am truly thankful for the time and effort you've invested in me. Everything I know and have learned is a testament to your exceptional mentorship. Thank you for being an incredible advisor!

My Ph.D. was carried out in an amazing laboratory, CosmoStat. Saying goodbye to a place where one has spent a significant period of time is always traumatic. The fact that I let tears flow freely during my farewell says much more than I could ever express. I would like to thank all the people who have been part of this magical place. Special thanks to Sam, Valeria, and Martin. Martin, thank you for the scientific discussions. Being able to knock on your door and ask random scientific questions added value to my Ph.D. Thanks to Sam for all your advice on software, your efforts in organizing

learning sessions, and your guidance on applying for job positions. Most of all, thank you for your empathy and for the human support you gave me over the years, thank you for all your "sei brava". Thank you, Valeria, for being, on more than one occasion, an 'unofficial' mentor.
Thanks to all the Ph.D. students, post-docs, and interns I met: Fadi, Tobias, Virginia, Benjamin, Lisa, Vilasini, Fabian, Ezequiel, Lucie, André, Sergio, Santi. Thanks to my first office-mate, Fadi, for all the jokes, Schoko-Bons, and patience towards my infinite questions. Thanks to my Ph.D. twin, Benjamin, for the human support, the chats, the walks around our buildings, and the time spent together. Thanks to Vilasini, the sweetest office-mate ever, for your support and admiration. I promise to make an effort to always be a source of inspiration for women researchers. Thank you, Justine, for being a travel and adventure (not always positive) companion. Thanks for the nice time spent together, your sweetness, and sensitivity. Thanks to my wonderful and powerful office mates, Vilasini and Lisa, for making every day in the office a pleasant one. I would also like to thank the JC organizers throughout the years: Virginia, Tob, Vilasini, and Lisa.

Being far from home can sometimes feel lonely, but my Italian gang in Paris has made this city feel like a second home. Your kindness, laughter, and shared moments have turned what could have been a challenging experience into one of the best chapters of my life. In particular, I have to thank all the characters of *Paris Season 1*: Federico C., Francesco, Luca, Federica, Virginia & Fil, and Matteo; of *Paris Season 2*: Federica, Pippo, Luca, Michi, Paolo, Federico, Francesco. A special thank you to Federica for being an incredible friend and always supporting me (especially during shopping).
I want to express my deepest gratitude to my flatmates of Rue Ampere, Vicio, Federico, and Sap (adopted), for your support and presence during all stages of my Ph.D. journey, especially navigating through the challenges of the pandemic. Your constant support, whether it was a listening ear, a shared laugh, or a helping hand (not with housework), has meant the world to me. Having flatmates like you has made the Ph.D. experience not only manageable but also memorable.

I would like to extend my thanks to my parents for their trust in my abilities and potential, as well as for their unwavering support. Additionally, I want to express my gratitude to my dearest and closest friends—Livia, Lidia, Roberta, Corrado, Eros, Francesco, Luigi and Enrica. Your support has been indispensable, especially considering the distance from home. Finally, I want to thank Domenico for being my pillar of strength, my sounding board, and my cheerleader during both the highs and lows of this academic adventure. Your belief in my abilities has been a constant source of inspiration. Distance may physically separate us, but your support has bridged any emotional gaps, making this challenging journey more bearable and, in many ways, more meaningful.

Contents

List of Figures

List of Tables

Chapter 1

The Cosmological Background

The aim of this first chapter is to introduce the general cosmological context and provide the theoretical background necessary to understand chapter 2 and chapter 3. In particular, I will introduce the standard cosmological model and the equations of Friedmann-Lemaître, along with the main consequences of the Friedmann Expansion and a brief introduction to structure formation. Finally, I will conclude by briefly describing the primary goals of the present and future cosmological survey aimed to address the mystery of dark energy.

1.1 Standard cosmological model

In this section, I introduce in brief the standard model of the Universe. Depending on their proprieties and the values of the cosmological parameters, different cosmological models can be assumed. All the studies presented in this thesis are based on the ΛCDM model, often referred to as the *standard cosmological model*. The ΛCDM paradigm assumes that on scales larger than $\sim$ 100Mpc, the Universe is isotropic and homogeneous. It starts from a hot and dense initial state, known as the Big Bang about 13.7 billion years ago. This event was followed by a phase of cosmic inflation of about 10^{-33} seconds. Shortly after the Big Bang, the Universe was extremely hot and dense, fully ionized, with a plasma of coupled photons and baryons, and opaque to radiation. Only about 370,000 years later, the temperature reached 3000 K, creating the condition for charged electrons and protons to become bound to form neutral hydrogen atoms. During this epoch, known as *recombination*, photons decoupled from matter, becoming free to travel through the space without interacting with matter and giving rise to the Cosmic Microwave Background (CMB). According to this model, the Universe is consisting mostly of three components. **First**, there is the cosmological constant Λ, associated with the finite vacuum energy density and accounting for $\sim$ 68% of the total energy density. **Second**, there is the Cold Dark Matter (CDM), consisting of $\sim$ 27% of the total energy density. **Third**, there is baryonic matter, accounting for less than 5%, mostly composed of hydrogen and helium gas. Moreover, we observe electromagnetic radiation, due to CMB photons and cosmological neutrinos.

The "cold" nature of the Dark matter is suggested by the observations. Indeed, in order to clump efficiently, the dark matter must not be hot, meaning it should have low velocities. If the dark matter particles were dominated by high velocities, what we observe today would not be consistent with the large-scale structure (Dodelson and Schmidt, 2020). Dark matter, by its very definition, does not interact (significantly) with other ordinary forms of matter, except through gravitational forces. Table 1.1 summarizes the 68% confidence limits of the cosmological parameters for the ΛCDM model. Our knowledge about the history of the Universe is summarised in Figure 1.1.

1.2 Homogeneous Isotropic Universe

To a first approximation, the Universe can be described as homogeneous and isotropic with respect to our position. This assumption may seem to contradict observations since the world around us is highly inhomogeneous and structured. However, while homogeneity is not valid on small scales, observations of the distribution of galaxies and the isotropy of the CMB confirm that the universe is isotropic when averaged over large spatial scales. Therefore, if we assume that our position in the universe is not peculiar, the isotropy we observe around us implies isotropy around every location, leading to the homogeneity of the universe as an immediate consequence of isotropy

Description	Parameter	Value
Hubble parameter	H_0	$(67.66 \pm 0.42)\ \mathrm{km\,s^{-1}Mpc^{-1}}$
Total matter density	Ω_m	0.3111 ± 0.0056
Dark matter density	$\Omega_\mathrm{c}h^2$	0.11933 ± 0.00091
Baryon density	$\Omega_\mathrm{b}h^2$	0.02242 ± 0.00014
Dark energy density	Ω_Λ	0.6889 ± 0.0056
Power spectrum normalisation	σ_8	0.8102 ± 0.0060
Spectral index	n_s	0.9665 ± 0.0038
Reionisation optical depth	τ	0.0561 ± 0.0071
Sum of neutrino masses	M_ν	< 0.12 eV

Table 1.1: Summary of current best estimates of the main cosmological parameters of the ΛCDM model. Values derived from Planck Collaboration et al. (2020).

everywhere. This assumption is known as the cosmological principle.
As we will see in the next section, a homogeneous and isotropic universe is described by the so-called Friedmann-Lemaître-Robertson-Walker model.

1.2.1 Friedmann-Lemaître-Robertson-Walker model

Before introducing the Friedmann-Lemaître-Robertson-Walker (FLRW) model, we can introduce the basic concepts of General Relativity and the definition of the metric. According to Special and General Relativity theories, space and time are reference-frame dependent and paired through the metric. The metric allows us to compute the line element ds^2, which defines the distance between two points, as:

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu, \tag{1.1}$$

with $\mu, \nu \in [0, 3]$, being 0 the time by convention and $1, 2, 3$ the spatial Cartesian coordinates. The metric tensor $g_{\mu\nu}$ describes the time evolution of the metric. For example, according to General Relativity, the Universe can be dynamic. If we consider a flat Universe expanding with time, after introducing the scale factor $a(t)$, describing how the distance between two objects changes with the contraction and expansion of the Universe (Hubble, 1929), the metric tensor becomes:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}. \tag{1.2}$$

This expression defines the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. In particular, for the FLRW model, assuming that the speed of light in the vacuum $c = 1$,

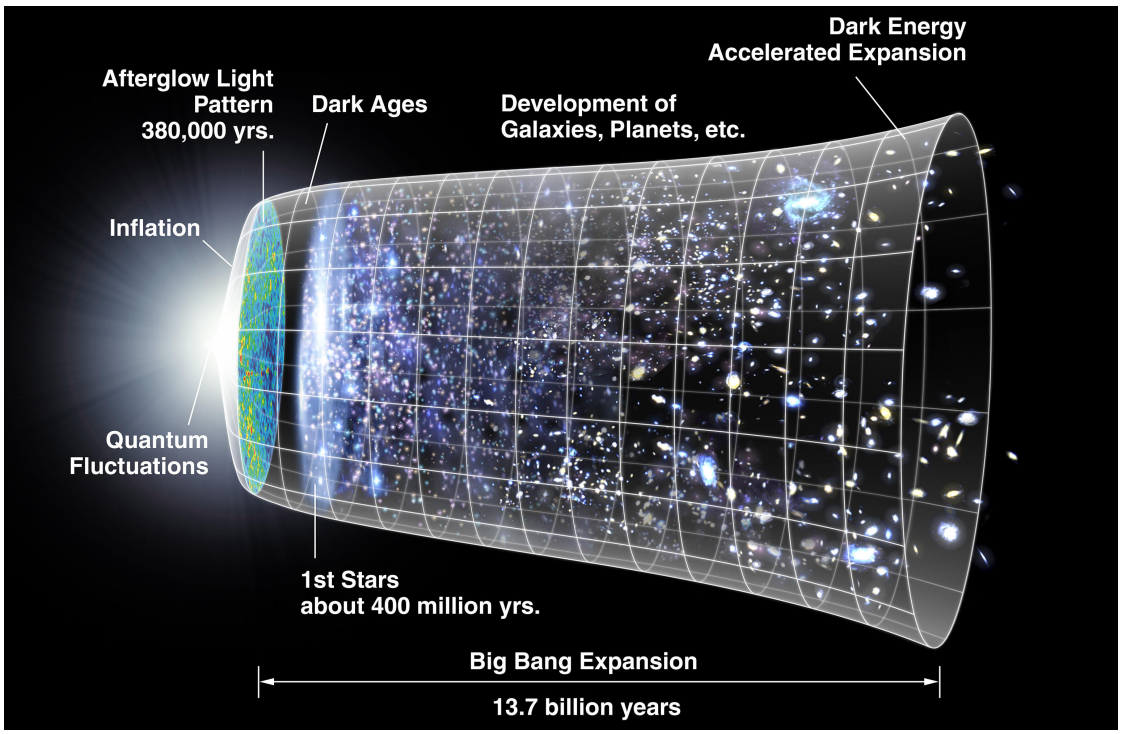

Figure 1.1: The universe's expansion over time. Credit: NASA/WMAP Science Team.

we can write Equation 1.1 as:

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr}{1-Kr^2} + r^2 d\theta^2 + r^2 sin^2\theta d\phi^2\right] \tag{1.3}$$

$$= dt^2 - a^2(t)\left[d\chi^2 + f_K(\chi)^2\left(d\theta^2 + sin^2\theta d\phi^2\right)\right], \tag{1.4}$$

where we introduced the spherical coordinate (r, θ, ϕ), the comoving radial distance χ, the transverse distance $f_k(\chi)$[1] and curvature parameter K. The Friedmann-Lemaître-Robertson-Walker (FLRW) metric is derived using invariants to satisfy the homogeneity and isotropy condition required by cosmological principle criteria. The choice to work in a comoving reference frame with spherical coordinates is well motivated by the cosmological principle and the isotropy properties. The parameter K is interpreted as the curvature parameter of the Universe today. To be more precise, we should say that the homogeneous and isotropic space at $t = t_0$ is characterized by a curvature K. According to its sign or value, we can distinguish the following cases:

$$K = \begin{cases} 1 & \text{Closed shape} \quad (\boldsymbol{S}^3) \\ 0 & \text{Flat shape} \quad (\boldsymbol{R}^3) \\ -1 & \text{Open shape} \quad (\boldsymbol{H}^3) \end{cases} \tag{1.5}$$

[1] defined in Equation 1.31 and Equation 1.33.

namely, for any fixed time t, if $K = 0$ the three-dimensional space is Euclidean, i.e. flat ($\boldsymbol{R}^3$), if $K = 1$ the space is spherical, with the quantity $1/\sqrt{K}$ defining the curvature radius of the spherical 3-dimensional space ($\boldsymbol{S}^3$), if $K = -1$ the space is hyperbolic ($\boldsymbol{H}^3$).

1.2.2 Geodesic equation and Energy-Momentum tensor

General Relativity (GR) is an essential tool to derive some of the fundamental findings that we have about Cosmology. In the context of GR, instead of describing gravity as an external force, we can incorporate it in the metric, this allows us to describe particles as moving in curved space-time, whose geometry is entirely described by the metric tensor $g_{\mu\nu}$ introduced in Equation 1.1.

In generic space-time, the notion of a straight line of a particle is generalized with the concept of *geodesic*, which represents the path of a particle when no external forces are applied (Weinberg, 1972):

$$\frac{d^2x^\mu}{d\lambda^2} = \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}, \tag{1.6}$$

where λ indicates the affine parameter, the scalar which parameterizes the particle along its trajectory, and $\Gamma^\mu_{\alpha\beta}$ is the Christoffel symbol, obtained from the metric as:

$$\Gamma^\mu_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]. \tag{1.7}$$

The Equation 1.6 is particularly relevant since we will use it to describe the propagation of light bundles in the presence of massive bodies and to introduce the fundamental equations of gravitational lensing. One fundamental result of General Relativity is how it relates the metric to the matter and the energy of the universe. Specifically, the Einstein field equations relate the geometry of space-time with the distribution of matter within it:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} = 8\pi G T_{\mu\nu}, \tag{1.8}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor; $\mathcal{R} = R_{\mu\nu}R^{\mu\nu}$ is the Ricci scalar curvature; $g_{\mu\nu}$ is the metric tensor; G is Newton's gravitational constant; $T_{\mu\nu}$ is the stress–energy tensor. Equation 1.8 can be easily interpreted in this way: on the left, we can find all the physical descriptors of the geometry and the curvature of the Universe, whereas on the right, we can find all the information about the source of the curvature, described by the energy-stress tensor. This expression is crucially important to obtain information about the evolution of the Universe and its content. For example, another consequence of the cosmological principle is that if we assume a comoving inertial frame, we can consider the energy-stress tensor a perfect fluid and describe it by its rest-frame energy density ρ and the momentum p such that:

$$T^{\mu}{}_{\nu} = T^{\mu\alpha} g_{\alpha\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.9}$$

Alternately, it can be described in a non-comoving reference frame by (Hobson et al., 2006):

$$T^{\mu\nu} = (\rho + p) U^{\mu} U^{\nu} - p g^{\mu\nu}. \tag{1.10}$$

1.3 The dynamics of the universe

After introducing the metric and the energy-momentum tensor, thanks to simplification made possible by the cosmological principle, the dynamical evolution of the Universe can be derived. Indeed, the latter can be described by the cosmological equations, which in turn can be derived from the Friedmann-Robertson-Walker metric, after computing Ricci's and Einstein's tensors, and the Christoffel symbols from the Euler–Lagrange equations. The derivation of these equations goes beyond the purposes of this work, however, in view of their importance, their role and consequences are presented in the following sections.

1.3.1 Friedmann–Lemaître Expansion Equations

Before the discovery of the expansion of the Universe, static cosmological models were proposed. Einstein modified the equations of fields 1.8, by introducing a cosmological constant Λ:

$$R_{\mu\nu} - \Lambda g_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi G T_{\mu\nu}. \tag{1.11}$$

with the purpose of obtaining a static solution for the resulting expansion equations. The zeroth-order component of Equation 1.11 yields to the equations of motion, or *first Friedmann equation* (Friedmann, 1922):

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2} - \frac{\Lambda}{3} = \frac{8\pi G \rho}{3}, \tag{1.12}$$

with $\dot{a}$ indicates the derivatives with respect to time t. From the ij^{th} components of the Equation 1.11 we obtain the *second Friedmann equation*:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}, \tag{1.13}$$

describing whether the expansion is accelerating or decelerating. To get the entire set of equations describing the dynamics of the equation we still need to compute

the equation expressing the conservation of energy. Specifically, by combining the Friedmann equations we obtain the following relation describing the evolution of ρ:

$$\frac{\partial(a^3\rho)}{\partial t} + p\frac{\partial}{\partial t}a^3 = 0. \tag{1.14}$$

The set of Equation 1.12, 1.13, 1.14 constitutes the set of cosmological equations we were looking for. The solution of this system consists in computing the evolution of the scale factor $a(t)$ as a function of the cosmological time, and the evolution of p and ρ. This dynamic is fully determined by imposing an equation of state for the matter that establishes a relationship between the pressure p and the energy density ρ.

Before deriving the equation of state, it is interesting to say a couple of words about the Λ term. As mentioned in the introduction of this section, it was introduced to find a static solution for the expansion equations. However, there was no physical interpretation for this constant, and after the discovery of the expansion of the Universe Einstein discarded this term again. Today, observations seem to suggest a cosmological constant $\Lambda \neq 0$, although the physical interpretation is profoundly different. In quantum mechanics, the *vacuum*, defined as a totally empty space, is characterized by a finite amount of energy, namely the vacuum energy density. For physical measurements involving gravity, this energy becomes important, since $E = mc^2$ corresponds to a mass. The new interpretation of GR equations consists of a cosmological constant equivalent to a finite vacuum energy density.

1.3.2 Equation of state

A solution of Equation 1.14 requires specifying the properties of each energy component in the form of an *equation of state*, $p = p(\rho)$; we specify this relation through a parameter w:

$$w \equiv \frac{p}{\rho}. \tag{1.15}$$

The Equation 1.15 may depend on the redshift; however, in the case it is time-independent and is inserted into Equation 1.14, it leads to the following expression (Blanchard et al., 2020):

$$\dot{\rho} = -3(1+w)\frac{\dot{a}}{a}. \tag{1.16}$$

Thus the equation of state, solution of Equation 1.16:

$$\rho_i(a) \propto a^{-3(1+w_i)}, \tag{1.17}$$

characterizes each i component of the matter.
Once the relations $\rho_i(a)$ are known, those can be used in Equation 1.12 to find a solution for $a(t)$. For example, the matter-energy density at late times is mainly in the form of baryons and cold dark matter particles, which are described by $w_b = w_c = 0$, whereas

the photon radiation density is characterized by $w_\gamma = 1/3$. For these two quantities the Equation 1.17 becomes:

$$\rho_\gamma(a) = \rho_{\gamma,0} a^{-4} \quad ; \quad \rho_{b,c}(a) = \rho_{b,c,0} a^{-3}. \tag{1.18}$$

As proved from Equation 1.13 the acceleration of the Universe requires a fluid, dubbed Dark Energy (DE), whose equation of state is $w_{DE} < -1/3$. The ΛCDM model, as mentioned in the previous section, assumes that this cosmic acceleration is due to the presence of a constant term, usually referred to as Λ, characterized by the equation of state $w_\Lambda = -1$ which, looking at Equation 1.16, corresponds to an energy density ρ_Λ independent from time.

A more generalized scenario postulates that the Dark Energy is dynamic, with a time−dependent equation of state parameter $w_\Lambda(a)$. The most commonly used parameterization for time dependence is (Linder and Jenkins, 2003):

$$w_\Lambda(a) = w_0 + w_a(1 - a) \tag{1.19}$$

where w_0 is the present ($t = t_0$) value of the equation of state and w_a is a measure of its time variation. In this scenario, the DE density evolution obeys the following expression:

$$\rho_\Lambda(a) = \rho_{0,\Lambda} a^{-3(1+w_0+w_a)} \exp\left[-3w_a(1 - a)\right]. \tag{1.20}$$

We can express the total density energy $\rho(a)$ as:

$$\rho(a) = \rho_m(a) + \rho_\gamma(a) + \rho_\Lambda(a). \tag{1.21}$$

As one can see from Equation 1.12, there is a value of the density ρ, dubbed *critical density*, which leads to a spatially flat geometry with $K = 0$. This is defined as:

$$\rho_{crit}(a) = \frac{3H^2(a)}{8\pi G}, \tag{1.22}$$

where we introduce the Hubble parameter $H^2(t)$:

$$H^2(t) \equiv \left[\frac{\dot{a}(t)}{a(t)}\right]^2 = -\frac{Kc^2}{a^2} + \frac{8\pi G\rho}{3}. \tag{1.23}$$

Commonly, for a generic matter component i, the dimensionless density parameter is introduced:

$$\Omega_i(a) \equiv \frac{\rho_i(a)}{\rho_{crit}(a)}, \tag{1.24}$$

which combined with Equation 1.21 leads to:

$$\Omega(a) = \Omega_{\rm m} a^{-3} + \Omega_{\rm r} a^{-4} + \Omega_\Lambda a^{f(a)}, \tag{1.25}$$

where we summarize in $f(a)$ all the scale-factor dependent terms of Equation 1.20. We can now define an effective curvature density parameter

$$\Omega_k(z) = \frac{-Kc^2}{[a^2(z)H^2(z)]} \tag{1.26}$$

such that :

$$\sum_{i=0}^{N} \Omega_i(z) + \Omega_k(z) = 1, \tag{1.27}$$

with the sum running over all the matter species N in the model. Combining the Equation 1.23 and Equation 1.27 we obtain:

$$H^2(a) = H_0^2 \left[\Omega_{\rm m} a^{-3} + \Omega_{\rm r} a^{-4} + \Omega_{\rm k} a^{-2} + \Omega_{\rm de} a^{f(a)}\right], \tag{1.28}$$

where $H_0 \equiv H(t = t_0)$ represents the Hubble parameter today, which is commonly written as

$$H_0 = 100\ h\ \text{km s}^{-1}\text{Mpc}^{-1}, \tag{1.29}$$

where h is the dimensionless Hubble parameter.

1.4 Consequences of the Friedmann Expansion

1.4.1 Proper and Comoving distance

One can consider two observers with no peculiar velocity, only moving apart because of the expansion of space. Their *comoving distance* is defined as:

$$\chi = \frac{r(t)}{a(t)}, \tag{1.30}$$

with $r(t)$ the proper distance. As a result, the scale factor $a(t)$ can be interpreted as the measurement of how large the Universe is at a given time t. It is often normalized such that $a(t_0) = 1$, where t_0 is the time today.

In particular, we refer to the *line-of-sight* comoving distance as the distance traveled by the light emitted by an observed source. Since the physical distance covered in time dt is $dr = cdt$, the equivalent comoving distance is $d\chi = -dt/a$. The total distance traveled by the light since it was emitted at time t_S is:

$$\chi_s = \int_0^{\chi_s} d\chi = \int_{t_s}^{t_0} \frac{dt}{a(t)}. \tag{1.31}$$

Finally, we can define the peculiar velocity in terms of the rate of change of proper distance $r(t)$, such that:

$$v_r \equiv \frac{dr}{dt} = \frac{d(a(t)\chi)}{dt} = \dot{a}\chi = \frac{\dot{a}}{a} r \equiv Hr, \tag{1.32}$$

with dots denoting the derivatives w.r.t. time. This result, known as *Hubble's law* (Hubble, 1929), asserts that, if H is constant, the peculiar velocity is proportional to the distance. This result implies that all galaxies are moving apart due to the expansion of the Universe. More generally, as we have seen in Equation 1.23, the Hubble parameter is a function of time, the value of which is constant only today. The Hubble law is therefore valid only for nearby objects, where $H \sim H_0$.

1.4.2 Comoving transverse distance

To compute distances in the Friedmann–Lemaître framework, we need to define the function $f_K(\chi)$:

$$f_K(\chi) = \begin{cases} \frac{1}{\sqrt{K}} \sin\left(\sqrt{K}\chi\right) & \text{for } K > 0 \\ \chi & \text{for } K = 0 \\ \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K}\chi\right) & \text{for } K < 0. \end{cases} \tag{1.33}$$

known as *comoving transverse distance*, which depends only on the curvature K of the space and on the comoving distance χ.

1.4.3 The cosmological redshift

Consider the incoming light from a source at fixed radial comoving distance χ, such that:

$$\chi = \int_0^{\chi} d\chi' = -\int_{t_2}^{t_1} \frac{dt}{a(t)} = \int_{t_1}^{t_2} \frac{dt}{a(t)} \tag{1.34}$$

where t_1 and t_2 are the time at which a wave crest is emitted and received, respectively. If a new wave crest is emitted at time $t_1 + \delta t_1$ with $\delta t_1 = 1/\nu_1$, it will be observed at time $t_2 + \delta t_2$. If we consider the radial comoving distance fixed, it has to be

$$\chi = \int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{dt}{a(t)} = \int_{t_1}^{t_2} \frac{dt}{a(t)}, \tag{1.35}$$

which implies:

$$\frac{\delta t_2}{a(t_2)} - \frac{\delta t_1}{a(t_1)} = 0 \quad \Rightarrow \frac{\nu_1}{\nu_2} = \frac{\delta t_2}{\delta t_1} = \frac{a(t_2)}{a(t_1)}. \tag{1.36}$$

Hence, one of the consequences of the expanding universe is that the frequency of radiation that propagates through the Universe is shifted by an amount that can be quantified as:

$$\frac{\lambda_2}{\lambda_1} = \frac{a(t_2)}{a(t_1)}. \tag{1.37}$$

If $a(t_2) = a(t_1)$, i.e., if there is no expansion, the frequency of the photons is not changed as they travel the Universe. However, the Universe expands, the scale factor $a(t_2) > a(t_1)$, hence the $\lambda_2 > \lambda_1$. This implies that the frequency of the radiation is

shifted towards a lower frequency. The shift is usually quantified by introducing the cosmological redshift z, such that:

$$1 + z \equiv \frac{\lambda_2}{\lambda_1} = \frac{\nu_1}{\nu_2} = \frac{a(t_2)}{a(t_1)}. \tag{1.38}$$

The relation between the scale factor and the redshift is crucial. Considering that $a(t_0) = 1$, we obtain:

$$a = \frac{1}{1+z}. \tag{1.39}$$

The redshift is the only available information about the distance we have for most astronomical sources. Indeed, in our Universe, the scale factor a is a monotonic function of time, hence, z is a monotonic function of time as well. This implies that a, t, and z can be adopted as measures of the distance of a source.

1.4.4 Angular diameter distance and Luminosity distances

Since the light from a source we observe today was emitted at an earlier time, its comoving and proper distance χ and $a(t)\chi$, can not be observed directly. However, two other distances can be computed: *the angular diameter* and *luminosity distance*. Consider r the proper distance of an object of physical size D and luminosity L, observed with the angular size θ and with the flux F. We can define the angular-diameter distance d_A and the luminosity distance d_L as:

$$d_A(a) = \frac{D}{\theta}, \tag{1.40}$$

and

$$d_L^2(a) = \frac{L}{4\pi F}. \tag{1.41}$$

In the absence of expansion, $d_A = d_L = r$. However, when the distances are considered in the expanding Universe, d_A, d_L, and r may be different from each other. In particular, Equation 1.40 and Equation 1.41 can be expressed in terms of transverse comoving distance $f_K(\chi)$:

$$d_A = a(t) f_K(\chi) \tag{1.42}$$

and

$$d_L = \frac{1}{a} f_K(\chi) = \frac{1}{a^2} d_A = (1+z)^2 d_A(z). \tag{1.43}$$

1.5 Linear growth of large scale structure

Observations suggest that the Universe around us is isotropic only on large scales. On smaller scales, the universe contains a great deal of structure (consider for example galaxies, superclusters of galaxies, and voids). To describe this "lumpiness", we should look at the history of structure formation.

1.5.1 Primordial perturbation

The large-scale structure observable today origins from the gravitational collapse of inhomogeneities in the early Universe. Inflation (Guth, 1981; Baumann, 2012), is the most widely-supported theory today regarding the physics of the nascent Universe. According to this model, quantum fluctuations in the very early Universe grow during a period of exceptionally fast expansion until reaching the macroscopic scales of the large-scale structures we observe today. The growth of these initial perturbations is described by the primordial power spectrum $\mathcal{P}_\zeta(k)$:

$$\mathcal{P}_\zeta(k) = A_\mathrm{s} \left(\frac{k}{k_p}\right)^{n_\mathrm{s}-1}, \tag{1.44}$$

parametrized as a power law, of which A_s is the amplitude and k_p is the pivot scale. The term n_s is known as the *spectral index*. The current observations indicate that $n_s = 0.9603 \pm 0.0073$. This percent-level deviation from the value $n_s = 1$ (corresponding to a scale-invariant power spectrum) is the direct measurement of time dependence in the inflationary dynamics (Baumann, 2012). Moreover, it is the convention to characterize the amplitude of the primordial power spectrum in terms of the RMS (Root Mean Square) variance of the spheres of radius R =8 h^{-1}Mpc, known as σ_8.

1.5.2 Linear growth

The primordial density perturbations produced during inflation can be considered the seeds of the structure growth. The fluctuations of the density around the background mean density are described by the *matter density contrast*:

$$\delta(\boldsymbol{r}, t) = \frac{\delta\rho_\mathrm{m}(\boldsymbol{r}, t)}{\bar{\rho}_\mathrm{m}(t)} = \frac{\rho_\mathrm{m}(\boldsymbol{r}, t) - \bar{\rho}_\mathrm{m}(t)}{\bar{\rho}_\mathrm{m}(t)}. \tag{1.45}$$

In the regime of small density fluctuations ($\delta \ll 1$), the density contrast evolution can be described by the linear perturbation theory. This condition occurs at early times and on large scales. In the chapter 2 we will derive in detail the evolution of the density contrast under the assumption of pressure-less fluid. For completeness of this chapter, I will outline the principal outcomes of this derivation here. It will be demonstrated that, in the linear regime, this evolution is described by an ordinary second-order differential equation (Dodelson and Schmidt, 2020):

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\bar{\rho}_\mathrm{m}\delta, \tag{1.46}$$

admitting the general solution (Peebles, 1980; Peacock, 1999):

$$\delta(t, \boldsymbol{r}) = D_+(t)\delta_+(\boldsymbol{r}) + D_-(t)\delta_-(\boldsymbol{r}), \tag{1.47}$$

with $D_+(t)$ and $D_-(t)$ growing and decaying mode. Since $D_-(t)$ vanishes as time goes on and so does not impact the formation of structures, the density contrast evolution

can then be described only by the growing mode, which is expressed in terms of scale factor as:

$$D_+(a) = \frac{5\Omega_{\rm m}H(a)}{2H_0}\int_0^a \frac{da'}{a'^3}\left(\frac{H_0}{H(a')}\right)^3. \tag{1.48}$$

The growing mode can be then normalized such that $D_+(a) = a$ when the universe is dominated by matter. It is conventional to normalize this factor, usually referred to as *growth factor*, as:

$$D(a) = \frac{D_+(a)}{D_+(a=1)}, \tag{1.49}$$

in order to have $D_+(a=1) = D_+(z=0) = 1$, and then express the density contrast in the linear regime as:

$$\delta(\boldsymbol{r}, z) = D(z)\delta(\boldsymbol{r}). \tag{1.50}$$

1.5.3 Matter power spectrum

In order to compare the theory of structure formation with the observation, we need to perform statistical measurements of the density field. The initial density fluctuations of the field were random but correlated, this means that according to their separation $\boldsymbol{r}$, the fluctuations in two points in the space are statistically related to each other. We can introduce the *two-point correlation function* (2PCF), defined as:

$$\langle\delta(\boldsymbol{r}')\delta(\boldsymbol{r}+\boldsymbol{r}')\rangle_{\boldsymbol{r}'} = \int_{\mathbb{R}^3}\delta(\boldsymbol{r}+\boldsymbol{r}')\delta(\boldsymbol{r}')d\boldsymbol{r}', \tag{1.51}$$

or express the two-point correlation function of the Fourier transform of the perturbation density $\tilde{\delta}(\boldsymbol{k})$ as:

$$\langle\tilde{\delta}(\boldsymbol{k})\tilde{\delta}^\star(\boldsymbol{k}')\rangle = (2\pi)^3 P_\delta(k)\delta_{\rm D}^{(3)}(\boldsymbol{k}-\boldsymbol{k}'), \tag{1.52}$$

with $\tilde{\delta}^\star(\boldsymbol{k})$ complex conjugate, $\delta_{\rm D}^{(3)}$ Dirac delta and $P_\delta(k)$ matter power spectrum.

When the universe is dominated by radiation, perturbation with comoving scale smaller than the horizon grows as $\delta \propto a^2$, after the 'horizon entry' and before matter starts to dominate $\delta = constant$, when the universe is matter dominated $\delta \propto a$. Since different scales cross the horizon at different times, the different Fourier modes of the density contrast will evolve differently. This means that the matter power spectrum $P_\delta(k)$ today differs from the simple power law of the primordial power spectrum we introduced in the previous section. This difference is quantified by the linear *Transfer function*, expressed in the Fourier space as:

$$\mathcal{T}(k) = \frac{\tilde{\delta}(k, a=1)\tilde{\delta}(k=0, a=0)}{\tilde{\delta}(k, a=0)\tilde{\delta}(k=0, a=1)}. \tag{1.53}$$

To compute the transfer function Boltzmann solvers such as `CAMB` (Lewis et al., 2000) or `CLASS` (Blas et al., 2011) are required. However, in several applications is

possible to use fitting formulae or approximation methods to derive analytical expressions for the transfer function. For example, Eisenstein and Hu (1998)[2] developed a well-motivated fitting formula for CDM-cosmology accounting for baryon effects in the matter transfer function. This formula allows us to express the relation between the power spectrum today and the primordial power spectrum as:

$$P_\delta(k) = \mathcal{T}^2(k) P_\zeta(k). \tag{1.54}$$

However, this linear transfer function is only valid in the linear regime. At small scales and late times, when $\delta \sim 1$, perturbation theory is no longer valid, and N-body simulations are required. In particular, N-body simulations can be used to derive fitting formulae for the nonlinear matter power spectrum. Within this context, Takahashi et al. (2012)[3] revisited the halofit model from Smith et al. (2003) to derive a more accurate fitting formula for the non-linear gravitational clustering. They run high-resolution N-body simulations for 16 different cosmological models, including models with a dark energy constant equation of state. The results were used to re-calibrate the parameters of the halofit model in order to better reproduce small-scale power spectra from the N-body simulations while preserving the precision at larger scales.

In particular, in the halofit model, the power spectrum is described by two terms:

$$P_{NL}(k) = P_Q(k) + P_H(k) \tag{1.55}$$

where the first term $P_Q(k)$ is the quasi-linear term, encoding for the power generated by the large-scale placement of haloes, whereas the second term $P_H(k)$ represents the power due to the self-correlation of haloes. This revisited fitting formulae reproduce with accuracy the predictions of the nonlinear matter power spectrum in the range of $0 \leq z \leq 10$ and $k \leq 30h$ Mpc^{-1} (up to 5% of accuracy for $k \leq 1h$ Mpc^{-1}).

1.6 Dark Energy science

Observations suggest that Dark energy, causing the acceleration of the Universe's expansion, is the dominant component of the Universe. Understanding the origin of dark energy is currently one of the most compelling challenges in physical science, however, to date, we do not have a convincing theoretical explanation for its existence. Without theoretical guidance, focusing on observational explorations is the only way to understand what the Universe is composed of. Toward this goal ambitious observational programs have to be implemented. In particular, according to the Dark Energy Task Force (DETF, Albrecht et al. (2006)) these scientific surveys should address the following tasks:

[2] The Eisenstein and Hu (1998) formula will be used to compute the initial condition of the simulations described in chapter 6.

[3] The Halofit model will be used several times throughout this thesis, for example, to validate the simulations described in chapter 5 and chapter 6, or to compute the theoretical predictions in chapter 7.

- Assess whether the accelerating expansion is due to a cosmological constant.
- If the cosmic acceleration is not consistent with being due to a cosmological constant, measure the time evolution of the dark energy equation of state $w(a)$
- Compare the effect of dark energy on cosmic expansion and on the growth of galaxies and galaxy clusters in order to search for the potential failure of General Relativity (GR).

In order to attain the objectives set by dark energy observational programs, we should gauge both the expansion history of the Universe and the rate of structure growth. These measurements of dark energy proprieties can then be expressed in terms of w_0, the value of the dark energy today, and its evolution w_a. If a failure of GR causes accelerating expansion, this can be revealed from discrepancies between the values of $w(a)$ constrained from these two types of data. The DETF (Albrecht et al., 2006), developed a quantitative "figure of merit"[4] to evaluate dark energy studies in the areas of Baryon Acoustic Oscillations, Galaxy Clusters, Supernovas, and Weak Lensing. They concluded that there is no single observational technique able to answer questions about dark energy or to produce a satisfying FoM. On the contrary, a combination of different cosmological probes is necessary. In this regard, The Legacy Survey of Space and Time (LSST) has predicted the constraining power on the dark energy parameters with the 10-year LSST data set, as reported in the LSST DESC Science Requirements Document (Mandelbaum et al., 2018). The contours, presented in Figure 1.2, show the constraints from all five probes individually and the joint forecast from the combined analysis.

1.6.1 Cosmological probes

Several astrophysical observations store the history of cosmic expansion. Different cosmological probes have different strengths and systematics, being sensitive in different ways to the properties of dark energy and to different cosmological parameters. For example, the DETF (Albrecht et al., 2006) and the LSST Dark Energy Science Collaboration (Mandelbaum et al., 2018) identified the following cosmological probes as the most promising in constraining dark energy parameters:

- Weak Lensing (WL): The measurements focus on the distortions of the images of background galaxies due to the deflection of light caused by the presence of foreground matter between the source and the observer. The WL probe is sensitive to dark energy through the structure growth rate and through the effect of dark energy on the relation redshift vs. angular-diameter distance.

[4]This is defined as the inverse of the area of the error ellipse which contains the 95% confidence limit in the $w_0 - w_a$ plane. A larger FOM is indicative of greater accuracy.

- Large Scale Structure (LSS): The measurements focus on the galaxy positions (galaxy clustering only). The Baryonic Acoustic Oscillations (BAO) information is implicitly included in Large scale structure probes. The LSS probe is sensitive to dark energy through structure growth and the expansion history of the Universe.

- Galaxy Clusters (GL): The measurements focus on the spatial distribution, density, and masses of galaxy clusters. Similar to WL, the GL probe is sensitive to dark energy due to its effect on the growth rate of structure and the relationship between angular-diameter distance and redshift. Additionally, it is sensitive to dark energy through its impact on the time evolution of the expansion rate.

- Type Ia supernova (SN): The measurements focus on the relationship between the redshift and the distance derived from light curves analysis from Type Ia supernovae. This cosmological probe is sensitive to dark energy because of its impact on this relation.

- Strong Lensing: The measurements focus on the distortions, time delay, and displacement and multiple observations of the images of the background galaxies due to the deflection of light caused by the presence of massive foreground objects between the source and the observer.

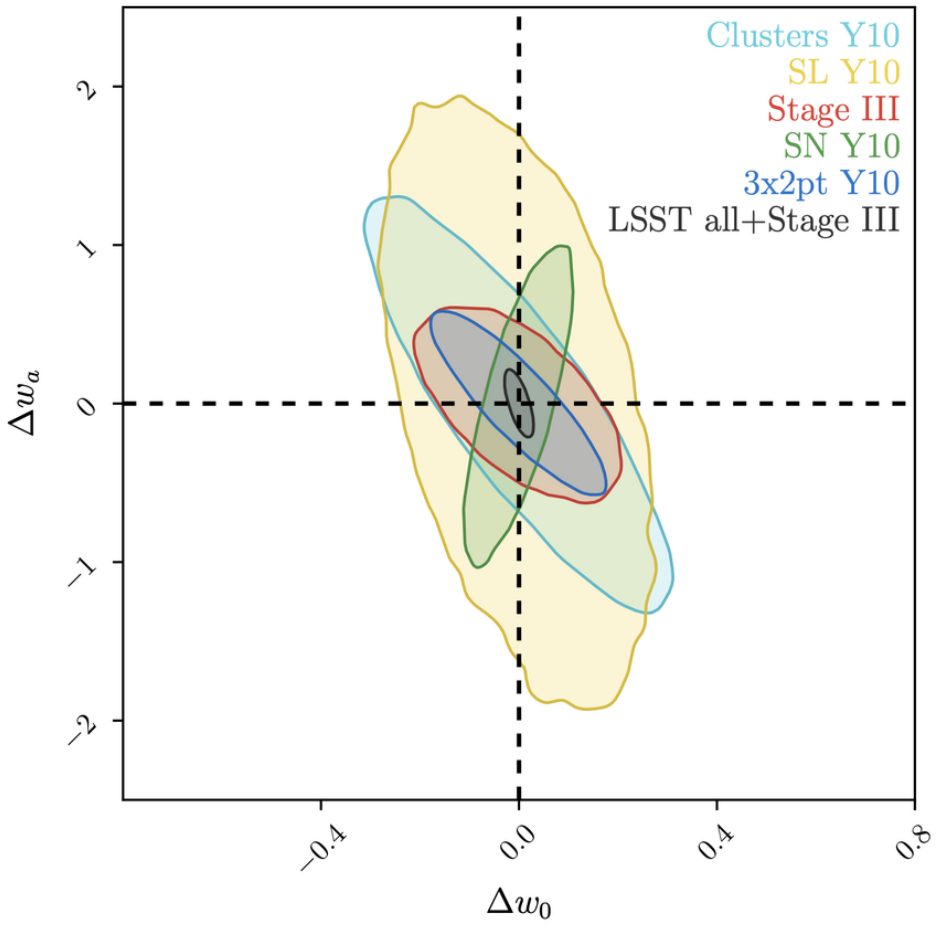

Figure 1.2: Forecast dark energy constraints for the Year 10 (Y10) LSST survey. The contours show the 68% confidence constraints and are computed for each of the five cosmological probes described in subsection 1.6.1 individually and combined. Specifically are shown: galaxy clusters (Clusters), strong lensing (SL), expected precision from Stage III surveys, Type Ia supernova (SN), weak lensing and large-scale structure (3×2 pt) and the joint analysis forecast, which includes Stage III priors. The quantities Δw_0 and Δw_a refer to the deviation of w_0 and w_a from their fiducial values of -1 and 0.
Figure from The LSST Dark Energy Science Collaboration (Mandelbaum et al., 2018).

Chapter 2

Gravity and dynamics of matter

This chapter delves into the formalism of numerical simulations for studying the large-scale structure of the Universe. I will begin by introducing the equations solved by these simulations, along with a brief introduction to the Eulerian and Lagrangian Perturbation theory. Subsequently, I will present the most common techniques found in the literature for the main steps of the numerical simulations: the implementation of the initial condition, the calculation of gravitational forces and the solution of the Poisson equation, and the time integration. Finally, I will provide an overview of the role of machine learning within the context of numerical simulations.

2.1 Gravitational instability

The following provides a summary of the key findings regarding the dynamics of gravitational instability that are relevant to the theory of N-body numerical simulations. All calculations are carried out within the framework of Newtonian gravity. Indeed, since Cold Dark Matter consists of non-relativistic particles, at scales smaller than the Hubble radius, the equations of motion simplify to the Newtonian gravity equation (Bernardeau et al., 2002).
These results are based on the work of Bernardeau et al. (2002); Jenkins (2010); Angulo and Hahn (2022).

2.1.1 The Vlasov Equation

The first step to describe the evolution of an ensemble of particles under their mutual gravitational interactions is considering their equation of motion:

$$\frac{d\mathbf{v}}{dt} = Gm \sum_i \frac{\mathbf{r}_i - \mathbf{r}}{|\mathbf{r}_i - \mathbf{r}|^3}, \tag{2.1}$$

here m indicates the mass of the particle; $\mathbf{v}$ is the velocity; $\mathbf{r}$ is the position; G is the Newton's gravitational constant. In the context of gravitational instabilities, the number of particles is large enough to rewrite the equation in terms of a smooth gravitational potential ϕ:

$$\frac{d\mathbf{v}}{dt} = -\frac{\partial \phi}{\partial \mathbf{r}} \tag{2.2}$$

induced by the local mass density $\rho(\mathbf{r})$, such that:

$$\phi(\mathbf{r}) = \mathrm{G} \int \frac{\rho(\mathbf{r}' - \mathbf{r})}{|(\mathbf{r}' - \mathbf{r})|} \mathrm{d}^3\mathbf{r}'. \tag{2.3}$$

The equation of motion becomes:

$$\frac{d\mathbf{v}}{dt} = -\nabla_{\mathbf{r}}\phi = \mathrm{G} \int \frac{\rho(\mathbf{r}' - \mathbf{r})(\mathbf{r}' - \mathbf{r})}{|(\mathbf{r}' - \mathbf{r})|^3} \mathrm{d}^3\mathbf{r}'. \tag{2.4}$$

In the scenario of an expanding universe, the description of equations of motion in terms of comoving distance and conformal time is preferred. We indicate with $\mathbf{x}$, the comoving distance such that $\boldsymbol{r} = a\boldsymbol{x}$, with a the cosmological scale factor, and τ the conformal time such that $dt = a(\tau)d\tau$. Remembering the definition of density contrast δ(**x**,t):

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{2.5}$$

and that $\bar{\rho}(t) \propto 1/a^3$ (Equation 1.17 with $w = 0$), we can rewrite Equation 2.4 as:

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= \mathrm{G}\int \frac{\rho(\mathbf{x}'-\mathbf{x})a(\mathbf{x}'-\mathbf{x})}{a^3|(\mathbf{x}'-\mathbf{x})|^3}a^3\mathrm{d}^3\mathbf{x}' \\ &= \mathrm{G}a\bar{\rho}\int \frac{(\mathbf{x}'-\mathbf{x})}{|(\mathbf{x}'-\mathbf{x})|^3}\mathrm{d}^3\mathbf{x}' + \mathrm{G}a\bar{\rho}\int \delta(\mathbf{x}'-\mathbf{x})\frac{(\mathbf{x}'-\mathbf{x})}{|(\mathbf{x}'-\mathbf{x})|^3}\mathrm{d}^3\mathbf{x}'.\end{aligned} \tag{2.6}$$

The velocity $\mathbf{v}$ can be expressed as:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{a}\mathbf{x} + a\frac{d\mathbf{x}}{dt}, \tag{2.7}$$

from which follows:

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= \ddot{a}\mathbf{x} + \dot{a}\frac{d\mathbf{x}}{dt} + \frac{d\mathbf{u}}{dt} \\ &= \ddot{a}\mathbf{x} + \frac{\dot{a}}{a}\mathbf{u} + \frac{d\mathbf{u}}{dt},\end{aligned} \tag{2.8}$$

where we indicated with $\mathbf{u}$ the peculiar velocities of the particles such that $\mathbf{u} \equiv a\frac{d\mathbf{x}}{dt}$. Using the transformation $\nabla_{\mathbf{r}} = \nabla_{\mathbf{x}}/a$, we can now express the potential in Equation 2.4 as the sum of the cosmological Gravitational potential $\boldsymbol{\Phi}$ and the background Gravitational potential $\boldsymbol{\phi}$, such that:

$$\phi \equiv \boldsymbol{\phi} + \boldsymbol{\Phi}, \tag{2.9}$$

and write:

$$\ddot{a}\mathbf{x} \equiv -\frac{1}{a}\nabla_x\boldsymbol{\phi}, \tag{2.10}$$

and

$$\frac{\dot{a}}{a}\mathbf{u} + \frac{d\mathbf{u}}{dt} \equiv -\frac{1}{a}\nabla_x\boldsymbol{\Phi}. \tag{2.11}$$

We aim to find the equations of motion valid in a homogeneous and isotropic background evolving according to the Friedmann equations. Combining the second of the Friedmann equation:

$$\ddot{a} = -\frac{4\pi\mathrm{G}}{3}a\bar{\rho}, \tag{2.12}$$

with Gauss's theorem:

$$\frac{4\pi}{3}\mathbf{x} = -\int \frac{(\mathbf{x}'-\mathbf{x})}{|\mathbf{x}'-\mathbf{x}|^3}\mathrm{d}^3\mathbf{x}', \tag{2.13}$$

we can write:

$$\ddot{a}\mathbf{x} = \mathrm{G}a\bar{\rho}\int \frac{(\mathbf{x}'-\mathbf{x})}{|\mathbf{x}'-\mathbf{x}|^3}\mathrm{d}^3\mathbf{x}'. \tag{2.14}$$

By substituting Equation 2.14 in Equation 2.8, we obtain the following equation of motion for the peculiar velocities:

$$\frac{\dot{a}}{a}\mathbf{u} + \frac{d\mathbf{u}}{dt} = \frac{d\mathbf{v}}{dt} + \ddot{a}\mathbf{u} = \mathrm{G}a\bar{\rho}\int \delta(\mathbf{x}'-\mathbf{x})\frac{(\mathbf{x}'-\mathbf{x})}{|\mathbf{x}'-\mathbf{x}|^3}\mathrm{d}^3\mathbf{x}'. \tag{2.15}$$

By integrating w.r.t. $\mathbf{x}$ the Equation 2.10 and combining it with the Equation 2.12, we obtain the following expression for the gravitational potential $\boldsymbol{\phi}$:

$$\boldsymbol{\phi}(\mathbf{x}) = \frac{4\pi \mathrm{G}}{3} a^2 \bar{\rho} \left(\frac{1}{2} |\mathbf{x}|^2 \right) = -\frac{\partial \mathcal{H}}{\partial \tau} \left(\frac{1}{2} |\mathbf{x}|^2 \right) \tag{2.16}$$

with $\mathcal{H} = \mathrm{d}\ln a / \mathrm{d}\tau = Ha$. By applying the Laplacian to the Equation 2.9 and remembering that:

$$\nabla^2_{\mathbf{r}} \phi = \frac{\nabla^2_{\mathbf{x}} \phi}{a^2} = 4\pi \mathrm{G} \bar{\rho} (1 + \delta) \tag{2.17}$$

we obtain the Poisson equation for the cosmological gravitational potential:

$$\nabla^2_{\mathbf{x}} \Phi = 4\pi \mathrm{G} a^2 \bar{\rho} \delta = \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta. \tag{2.18}$$

Given the Equation 2.15, we can express the momentum of a single particle of mass m as:

$$\mathbf{p} = ma\mathbf{u}, \tag{2.19}$$

and then from the equation of motion, we get:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau} = -ma \nabla_x \boldsymbol{\Phi}. \tag{2.20}$$

By defining the particle number density in the phase space as $f(\mathbf{x},\mathbf{p},\tau)$, from Liouville's theorem and phase-space conservation, we get the Vlasov equation:

$$\frac{\mathrm{d}f}{\mathrm{d}\tau} = \frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \nabla f - ma \nabla \Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \tag{2.21}$$

2.1.2 Eulerian Dynamics

Let us consider the zeroth order moment of the distribution functions:

$$\int f(\mathbf{x},\mathbf{p},\tau) \mathrm{d}^3\mathbf{p} \equiv \rho(\mathbf{x},\tau), \tag{2.22}$$

as we can see, this equation links the phase space density $f(\mathbf{x},\mathbf{p},\tau)\mathrm{d}^3\mathbf{p}$ to the local mass density field $\rho(\mathbf{x},\tau)$. The consequence is nonlinearity induced in Equation 2.21 since the potential $\nabla\Phi$ depends on $\rho(\mathbf{x},\tau)$. Due to the nonlinearity and the presence of multiple variables in this partial differential equation, it is very difficult to solve. Fortunately, we only aim to solve the evolution of the spatial distribution, which can be achieved by considering successive moments of the distribution function. From the first and second order moment, we can define the *peculiar velocity flow* $\mathbf{u}(\mathbf{x},\tau)$ and the *stress tensor* $\sigma_{ij}(\mathbf{x},\tau)$:

$$\int \frac{\mathbf{p}}{ma} f(\mathbf{x},\mathbf{p},\tau) \mathrm{d}^3\mathbf{p} \equiv \rho(\mathbf{x},\tau) \mathbf{u}(\mathbf{x},\tau), \tag{2.23}$$

$$\int \frac{\mathbf{p}_i \mathbf{p}_j}{m^2 a^2} f(\mathbf{x},\mathbf{p},\tau) \mathrm{d}^3\mathbf{p} \equiv \rho(\mathbf{x},\tau)\mathbf{u}_i(\mathbf{x},\tau)\mathbf{u}_j(\mathbf{x},\tau) + \sigma_{ij}(\mathbf{x},\tau). \tag{2.24}$$

We can compute the equation for $\mathbf{u}(\mathbf{x},\tau)$ and $\sigma_{ij}(\mathbf{x},\tau)$ from the successive moments of the Vlasov equations. Specifically, we obtain the continuity equation by taking the zeroth moment:

$$\frac{\partial \delta(\mathbf{x},\tau)}{\tau} + \nabla \cdot \{[1+\delta(\mathbf{x},\tau)]\mathbf{u}(\mathbf{x},\tau)\} = 0, \tag{2.25}$$

and the Euler equation by subtracting $\bar{\rho}\mathbf{u}(\mathbf{x},\tau)$ multiplied by the continuity equation from the first momentum equation:

$$\begin{aligned} &\frac{\partial \mathbf{u}_i(\mathbf{x},\tau)}{\partial \tau} + \mathcal{H}(\tau)\mathbf{u}_i(\mathbf{x},\tau) + \mathbf{u}_j(\mathbf{x},\tau)\cdot\nabla_j \mathbf{u}_i(\mathbf{x},\tau) = \\ &-\nabla_i \Phi(\mathbf{x},\tau) - \frac{1}{\rho(\mathbf{x},\tau)}\nabla_j(\sigma_{ij}(\mathbf{x},\tau)). \end{aligned} \tag{2.26}$$

The equations Equation 2.25 and Equation 2.26 describe the conservation of mass and momentum, respectively. Once the phase-space information has been integrated out, several phenomenological models allow us to close the system by assuming an ansatz for the stress tensor σ_{ij}. The ansatz for the stress tensor represents the equation of state for the cosmological fluid and assumes that the formation of matter structures involves matter with negligible pressure and velocity dispersion (*single-stream approximation*). The physical interpretation of the stress tensor is the deviation of the particles' motion from a single coherent flow (Bernardeau et al., 2002). Therefore, during the first stage of structure formation, when they are not yet collapsed and virialized, we can assume that $\sigma_{ij} \sim 0$. This approximation will no longer be valid at progressively larger scales, and the breakdown of $\sigma_{ij} \sim 0$, known as *shell crossing*, depicts the generation of velocity dispersion caused by several streams.

2.1.3 Eulerian perturbation theory

At very large scales and at the initial stages of gravitational structure formation, we expect the matter distribution in the Universe to be smooth. In other words, we can assume that:

$$|\delta(\mathbf{x},\tau)| \ll 1, \tag{2.27}$$

$$|\nabla_j \mathbf{x}_i(\mathbf{x},\tau)| \ll \mathcal{H}(\tau) \tag{2.28}$$

Hence, it is possible to linearize the set of equations 2.23, 2.25, 2.26 and obtain the Ordinary Differential Equation (ODE) describing the motion in the linear regime:

$$\frac{\partial \delta(\mathbf{x},\tau)}{\partial \tau} + \theta(\mathbf{x},\tau) = 0, \tag{2.29}$$

$$\frac{\partial \mathbf{u}_i(\mathbf{x},\tau)}{\partial \tau} + \mathcal{H}(\tau)\mathbf{u}_i(\mathbf{x},\tau) = -\nabla_i \Phi(\mathbf{x},\tau). \tag{2.30}$$

Here, we indicate $\theta(\mathbf{x},\tau) \equiv \nabla \cdot \mathbf{u}(\mathbf{x},\tau)$ as the divergence of the velocity field, whose motion is described by:

$$\frac{\partial\theta(\mathbf{x},\tau)}{\partial\tau} + \mathcal{H}(\tau)\theta(\mathbf{x},\tau) + 4\pi \mathrm{G}a^2(\tau)\bar{\rho}(\tau)\delta(\mathbf{x},\tau). \tag{2.31}$$

Similarly, we can introduce the curl of the velocity field $\mathbf{w}(\mathbf{x},\tau) \equiv \nabla \times \mathbf{u}(\mathbf{x},\tau)$, whose equation is:

$$\frac{\partial\mathbf{w}(\mathbf{x},\tau)}{\partial\tau} + \mathcal{H}(\tau)\mathbf{w}(\mathbf{x},\tau) = 0. \tag{2.32}$$

Note that these equations follow directly by applying the divergence and the curl to Equation 2.29 and Equation 2.30.

By replacing the time derivative of Equation 2.31 in Equation 2.29, we can write the ODE describing the linear instability of density fluctuations as:

$$\frac{\partial^2\delta(\mathbf{x},\tau)}{\partial\tau^2} + \mathcal{H}(\tau)\frac{\partial\delta(\mathbf{x},\tau)}{\partial\tau} = +4\pi \mathrm{G}a^2(\tau)\bar{\rho}(\tau)\delta(\mathbf{x},\tau) = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(\mathbf{x},\tau), \tag{2.33}$$

that yields exactly to Equation 1.47. This equation can be written in terms of the *linear growth factor* $\mathrm{D}(\tau)$ such that: $\delta(\mathbf{x},\tau) = \mathrm{D}(\tau)\delta(\mathbf{x},0)$ and Equation 2.33 becomes:

$$\frac{d^2\mathrm{D}(\tau)}{d\tau^2} + \mathcal{H}(\tau)\frac{d\mathrm{D}(\tau)}{d\tau} = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\mathrm{D}(\tau)'. \tag{2.34}$$

The two independent solutions of this second-order differential equation can be written as : $\delta(\mathbf{x},\tau) = \mathrm{D}_+(\tau)\delta_+(\mathbf{x}) + \mathrm{D}_-(\tau)\delta_-(\mathbf{x})$, referred to as the fastest-growing mode and the slowest-growing mode, respectively. Equivalently, we can write Equation 2.34 in terms of the scale factor as:

$$a^2\frac{\mathrm{d}^2\mathrm{D}}{\mathrm{da}^2} + \left(\Omega_\Lambda(a) - \frac{\Omega_m(a)}{2} + 2\right)a\frac{\mathrm{dD}}{\mathrm{da}} = \frac{3}{2}\Omega_m(a)D. \tag{2.35}$$

While analytic solutions exist for some cosmological models (see Chernin et al. (2003); Demianski et al. (2005)), in general, this equation needs to be solved numerically. Going beyond the linear order in Equation 2.26, a recursion relation to all orders in perturbation theory can be obtained (Goroff et al., 1986) establishing the framework for standard Eulerian cosmological perturbation theory (SPT). For further details, refer to Bernardeau et al. (2002).

2.1.4 Lagrangian perturbation theory

Alternatively, it is possible to describe the dynamic in the Lagrangian framework. In this case, we are interested to consider the displacement field $\boldsymbol{\Psi}(\mathbf{q})$ mapping the initial Lagrangian coordinate $\mathbf{q}$ to the Eulerian particles trajectories $\mathbf{x}$:

$$\mathbf{x}(\tau) \equiv \mathbf{q} + \boldsymbol{\Psi}(\mathbf{q},\boldsymbol{\tau}). \tag{2.36}$$

In this context, the distribution function introduced in subsection 2.1.2, becomes:

$$f(\mathbf{x},\mathbf{p},\tau) = \delta_D(\mathbf{x}-\mathbf{q}-\mathbf{\Psi}(\mathbf{q},\tau))\delta_D(\mathbf{p}-ma\mathbf{\Psi'}(\mathbf{q},t)). \tag{2.37}$$

From mass conservation follows that the density

$$\bar{\rho}(1+\delta(\mathbf{x}))d^3x = \bar{\rho}d^3q \tag{2.38}$$

can be obtained from the jacobian of the transformation between Eulerian and Lagrangian space $J := \det J_{ij} := \det \partial x_i/\partial q_j$ as:

$$1+\delta(\mathbf{x}) = \frac{1}{\mathrm{Det}(\delta_{ij}+\Psi_{i,j})} \equiv \frac{1}{J(\mathbf{q},\tau)}. \tag{2.39}$$

The equation of motion in conformal time reads then:

$$\frac{d^2\mathbf{x}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\mathbf{x}}{d\tau} = -\nabla_{\mathbf{x}}\Phi(\mathbf{x}), \tag{2.40}$$

and applying the divergence to this equation we obtain:

$$J(\mathbf{q},\tau)\nabla\cdot\left[\frac{d^2\mathbf{\Psi}}{d^2\tau} + \mathcal{H}\tau\frac{d\mathbf{\Psi}}{d\tau}\right] = \frac{3}{2}\Omega_m\mathcal{H}(J-1). \tag{2.41}$$

This last equation can be expressed as an equation including only derivative with respect to the Lagrangian coordinates, considering that $\nabla_i = (\delta_{i,j}+\Psi_{i,j})^{-1}\partial/\partial\boldsymbol{q}$:

$$J(\delta_{ij}+\Psi_{i,j})^{-1}(\Psi''_{i,j}+\mathcal{H}\Psi'_{i,j}) = \frac{3}{2}\Omega_m\mathcal{H}(J-1), \tag{2.42}$$

This expression is then solved perturbatively truncating in time the Taylor expansion of the form $\mathbf{\Psi}(\mathbf{q},\tau) = \sum_{n=1}^{\infty} D(\tau)^n\mathbf{\Psi}^{(n)}(\mathbf{q})$ (Buchert, 1989, 1994; Bouchet et al., 1994; Catelan, 1995).

Particularly important is the first order of this expression leading to the *Zel'dovich* approximation (Zel'Dovich, 1970)

$$\mathbf{x}(\mathbf{q},\tau) = \mathbf{q} + D_+(\tau)\boldsymbol{\nabla}_q\nabla_q^{-2}\delta_+(\boldsymbol{q}) \tag{2.43}$$

with $\delta_+(\boldsymbol{q})$ the growing mode spatial fluctuation part of the standard Eulerian perturbation Theory (SPT). The Lagrangian Perturbation Theory (LPT) solutions are of particular interest even considering that they can be used to define the initial condition in the simulations.

2.2 Numerical simulations

We have seen in the previous sections that the current understanding of large-scale structure formation relies on the assumption that the gravitational amplification of small perturbations gives rise to these structures. Numerical simulations are the only available tool for investigating the formation and the evolution of these structures in the non-linear regime, when analytical methods for calculating relevant quantities are missing. Furthermore, cosmological simulations are essential for a variety of other purposes. For example, they provide a good way to test theories and hypotheses by comparing mock data with observations. Additionally, in the context of data-driven approaches and emulators, they are extremely important for creating input training data. They also help us to generate synthetic data from next-generation surveys and forecast results from those, allowing us to test different statistical inference techniques and quantify systematic and statistical errors, among other things. However, developing N-body cosmological codes is not a straightforward task, and it is crucial to keep in mind that, since we are dealing with real physical processes, there are certain physical requirements that need to be satisfied. As an example, we know that the Universe consists of a significantly larger amount of matter than we could ever simulate in our N-body code and the average density over progressively larger scales approaches constant values. This suggests that the simulation volume V can not be considered isolated with respect to the regions outside it, and these regions must be filled appropriately. This is why periodic boundary conditions are necessary. Without them, most of the matter in the box would collapse towards the center of the box (Bagla and Padmanabhan, 1997). Moreover, it is essential to ensure that specific boundary conditions do not affect the evolution of perturbations and that the average density inside and outside the box is the same. Lastly, the interaction of the N-body particles must be purely collisionless.

In the following section, we will see that the main steps characterizing an N-body simulation are: setting the initial conditions; evaluating the gravitational force; the time integration. These results are based on the work of Bernardeau et al. (2002); Bagla and Padmanabhan (1997); Bagla (2002); Bagla and Ray (2003); Springel (2005); Dolag et al. (2008); Angulo and Hahn (2022).

2.2.1 Initial condition

A key ingredient in every numerical simulation is the use of correct and justified initial conditions. The initial stages of the universe are described by physical processes that are more complex than just the collisionless interaction between dark matter particles. In the early universe, radiation and hot plasma were tightly coupled and linked to neutrinos and dark matter particles via gravity. The large-scale structures we observe today are the result of the gravitational evolution of these species from initial tiny perturbations, and this evolution is described by coupled non-linear Einstein-Boltzmann

equations. These equations have been solved using Einstein-Boltzmann solvers, which solve the equations not in full non-linearity, but by using approximated methods from cosmological perturbation theory (Fidler et al., 2017). Examples of the most well-known Boltzmann codes are `Camb` (Lewis et al., 2000) and `Class` (Blas et al., 2011), which are used to solve the coupled multi-physics systems to the first order. Although Boltzmann codes are able to solve the set of equations with high accuracy, they are limited to cases where the perturbation of matter and radiation is very small, i.e. they can only be used in the early phases of gravitational evolution. After that, matter perturbations become larger and larger, leading to the non-linear gravitational collapse of matter. At this stage, cosmological (N-body) simulations are required to solve the fully non-linear problem.

However, we need to consider that N-body simulations are performed using the Newtonian approximation and by considering only collisionless matter. This approximation is only valid at late times and on small scales, where we can neglect the effect of radiation. Matching relativistic solutions that involve radiation and matter in the early universe with Newtonian physics at a later time presents a significant challenge.

To create the initial conditions for N-body simulations, we can use the output of the Einstein-Boltzmann solvers. One possible approach adopted to do this is the *forwards approach*, which involves computing the matter density at a high redshift using the Boltzmann code and then computing the initial particle displacements and velocities that will be used as the starting point of the N-body simulation. However, at early times, there is a residual component of thermal radiation from the CMB and cosmic neutrino background, and the N-body simulations cannot describe the effect of inhomogeneous radiation, which leads to the suppression of the growth structure (Fidler et al., 2017).

One possible alternative for setting up the initial conditions is using the *Backscaling method*. This method involves generating the linear matter power spectrum at redshift $z = 0$ (today) using the Boltzmann codes. Then, the amplitude of the matter power spectrum is rescaled back using the linear growth function D_+ to create an initial density field at the time when the N-body simulation is initialized. In other words, the rescaling procedure provides an artificial universe at the time the N-body simulation is initialized with today's radiation content. The simulations initialized with this approach will have, by construction, the correct matter power spectrum on the largest scales, where the linear growth function is reproduced by the N-body simulations.

The first step in computing the initial condition is to obtain the initial gravitational potential. We should note that the density contrast is connected to the gravitational potential through the Poisson equation. Therefore, the gravitational potential can be derived from the matter density contrast using the following formula:

$$\phi = \frac{\nabla^{-2}\delta}{a}. \tag{2.44}$$

Using the linear Boltzmann solvers and considering that for the rescaling procedure, a pure growing mode is assumed (Fidler et al., 2017), we can obtain the density contrast

δ^{BC} as:

$$\delta^{BC}(\mathbf{x}, a) = C_+(\mathbf{x}) D_+(a) \tag{2.45}$$

with $C_+(\mathbf{x})$ spatial constant fixed by the boundary conditions and $D_+(a)$ the growth function. If we compute the Equation 2.44 at $a \to 0$ and express the matter density via Equation 2.45, we obtain:

$$\lim_{a\to 0} \phi = \nabla^{-2} C_+ \lim_{a\to 0} \frac{D_+(a)}{a} \equiv \phi_{ini}. \tag{2.46}$$

Since $D_+(a)$ is analytical around $a = 0$ and for a ΛCDM universe $D_+(a) \propto a + \mathcal{O}(a^4)$, by expressing C_+ in term of the matter density at time a_0, we can write:

$$\phi_{ini} = \frac{\nabla^{-2}\delta^{code}(\mathbf{x}, a_0)}{D_+(a_0)} \lim_{a\to 0} \frac{D_+(a)}{a}. \tag{2.47}$$

After obtaining the initial gravitational field, the next step is to establish the initial conditions for fully non-linear cosmological simulations. One way to compute the initial displacement and velocities of the particles is through Lagrangian perturbation theory. An example of this is applying the Zel'dovich approximation to determine the initial positions:

$$\mathbf{x} = \mathbf{q} - D_+(z)\phi_{ini}(\mathbf{q}) \tag{2.48}$$

and velocities:

$$\mathbf{v} = \dot{D}_+(z)\nabla\phi_{ini}(\mathbf{q}). \tag{2.49}$$

Similarly, higher-order perturbation theory beyond the Zel'dovich approximation can be used to compute the nLPT displacement from ϕ_{ini}. A more detailed description can be found in e.g. Angulo and Hahn (2022); Fidler et al. (2017).

2.2.2 Calculating Force

Direct Summation

The most direct method for computing the gravitational force is the particle-particle (PP) method (Aarseth et al., 1979). This involves summing the contributions of all the individual particles to the gravitational potential:

$$\Phi(\mathbf{r}) = -G \sum_j \frac{m_j}{\sqrt{|\mathbf{r} - \mathbf{r}_j|^2 + \epsilon^2}}, \tag{2.50}$$

where m_j and $\mathbf{r}_j$ indicate the mass and position of the jth particle, respectively. The gravitational softening parameter ϵ is used to smooth the force between two particles and to minimize the two-body relaxation effect. This effect occurs when the number of particles is small enough to break the collisionless fluid assumption (Dolag et al., 2008). Indeed the number of dark particles we can simulate is always limited compared

to the number of dark matter particles characterizing real physical systems. Although the PP is very robust, it can only be implemented with a small number of particles, due to its high computation cost. Specifically, the computing time is $\propto N^2$, where N is the total number of particles.

The Tree Algorithm

As mentioned above, the limitation of the PP method lies in its computational cost, mainly arising from the direct pairwise summation over all particles. The most natural extension of the PP algorithm is represented by the *tree code algorithm*.
The key idea is that the contribution from distant particles to each particle can be evaluated by a multipole expansion of groups of particles that, if selected correctly, allows truncating the expansion at a low order (Bernardeau et al., 2002). In other words, we can approximate the contribution from distant particles as the force coming from a single particle lying at the center of mass of the group of particles. Therefore, the interaction of each particle is reduced compared to the PM method and is of the order of $\sim \log N$, changing the number of calculations to $N \log N$ (Bagla, 2005). Most commonly, the practical implementation of the multipole expansion consists of the recursive division of particles into groups using a tree structure. The entire volume of the simulation V is considered a cube. It is hierarchically partitioned into smaller cubes, with a volume $v - 1/8V$, until one particle (or a few) is left in the smallest cells. In contrast, the set of particles in the larger cell is used for fast rapid calculation. A key role in the algorithm is the cell acceptance criterion, which is the determining factor for classifying a group of particles as either close or far apart (Barnes and Hut, 1986; Bagla, 2005).
Although the final result of the algorithm is an approximate version of the real force, the error can be reduced by adapting the cell acceptance criterion, and finding a balance between the computational resource invested and the correct force.
Despite several advantages, such as the fact that the degree system clustering does not impact the CPU time for each step, its high level of parallelizability, and the ability to look at the detail of specific regions of interest, such as voids, clusters, or galaxy halos, due to the use of individual masses for the particle, one of the most challenging aspects of the tree code is the computation of periodic boundary conditions, which demands a lot of memory (Bagla, 2005; Bernardeau et al., 2002). Nevertheless, tree codes have been efficiently used for several cosmological N-Body simulations, including PKDGRAV (Potter et al., 2017), GADGET (Springel et al., 2001), and `CHANGA`.

The PM Algorithm

The Particle-Mesh (PM) codes are ideal for cosmological N-Body simulations that aim to study the large-scale structure of the universe. The key feature of these methods is that, instead of attempting to directly solve the N-body problem, they approximate gravitational forces by estimating the densities of the dark matter particles on a grid.

The PM algorithm can be described as follows: First, the density at the mesh points is computed from the particle positions. Then, the density field is transformed into Fourier space to obtain the gravitational potential using Green's function and Poisson's equation. Finally, the force on each particle is determined by interpolating the potential's derivative back to the particle's position. In many cases, the number of particles used in the simulation corresponds to the number of cells in the implemented mesh (Dolag et al., 2008). In other words, in this scheme, the gravitational force is treated as a field quantity evaluated on a grid. The power of the PM algorithm lies in the fact that Poisson's equation in Fourier space is a simple algebraic equation, which can be solved using Fast Fourier Transform (FFT) (Bagla, 2005). The calculation of the Fourier transform via FFT is extremely fast, requiring only order $N_g \log N_g$ operations, where N_g is the number of cells in the grid (Dolag et al., 2008). However, although the use of mesh introduces a natural way to soften the force at small scales, it also implies that the structure at scales smaller than the mesh size, can not be accurately resolved. This softening parameter is $\sim L/N_g$ with L box size of the simulation. To have a resolution similar to the one of a tree code, high values of N_g have to be used, leading to high memory and CPU cost (Bernardeau et al., 2002). Nevertheless, PM codes are generally cheap in terms of CPU and memory, scaling as $\mathcal{O}(N, N_g^3 \log N_g)$. Other noteworthy advantages are that the Fourier method allows easy computation of the periodic boundary and the implementation of this kind of code is thus rather simple and parallelizable.

Hybrid Methods

Several Hybrid methods have been developed to increase the resolution of the PM scheme, specifically adding a correction to the short-range force computed using the PM method. A widely used Hybrid code is the P^3M, computing the force between close neighbors using the PP method. Although the implementation of this code has enabled the generation of very large and accurate cosmological N-body simulations, it has several disadvantages. For example, the correction for the force assumes isotropy, but the force computed at a large scale has anisotropies due to the mesh structure's anisotropic nature. The correction is applied up to only $\sim 1.5 - 2$ grid lengths, which is far below the actual distance where the force is underestimated by the PM method (Bagla, 2005). Finally, as the system evolves and dark matter structures become highly clustered, short-range interactions become more predominant compared to long-range forces. Consequently, the amount of time required to compute the PP interactions becomes progressively more significant.

Another method that employs a hybrid approach is the TreePM code (Bagla, 2002; Xu, 1994; Bode et al., 2000; Bagla and Ray, 2003), which partitions the gravitational potential into a long-range and a short-range component in Fourier space as follows:

$$\Phi_{\mathbf{k}}^{long} = \Phi_{\mathbf{k}} \exp(-\mathbf{k}^2 r_s^2) \tag{2.51}$$

$$\Phi_{\mathbf{k}}^{short}(\mathbf{x}) = -G \sum_i \frac{m_i}{\mathbf{r}_i} \mathrm{erfc}\left(\frac{\mathbf{r}_i}{2r_s}\right). \tag{2.52}$$

Here, $\mathbf{r}_i$ is the distance vector between the ith particle and the point $\mathbf{x}$, r_s describes the spatial scale of the force-split, and erfc is the complementary error function (Bagla, 2002). While the long-range potential is calculated using the mesh-based method, the short-range potential is computed using the tree algorithm and is modified by a long-range cut-off factor. This method has several advantages. If r_s is appropriately chosen (i.e., slightly larger than the mesh scale), it helps to reduce the force anisotropies characterizing the PM methods. Moreover, it allows for preserving several advantages of the classical tree algorithm, such as the control of the softening scale and the unlimited dynamical range (Bagla, 2002).

2.2.3 Time integration

After computing the gravitational forces for each particle, we must integrate the equations of motion over time. The accuracy of the final simulations depends on the integrator scheme we use. To solve the ordinary differential equations (ODEs) in the form of $\dot{\mathbf{y}} = f(\mathbf{y})$ with appropriate initial conditions, we can use implicit or explicit methods to evolve the system from state n to state $n+1$. Although implicit methods have better proprieties, they require resolving the system iteratively, which involves inverting a sparsely sampled matrix with the dimension of the total number of data points.

Euler's method is the simplest way to integrate ODEs. It requires multiplying the derivatives with the time step:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_n)\Delta t, \tag{2.53}$$

and assumes that the derivatives are constant during the time step.

Another possibility is to use the so-called *Runge-Kutta method* (RK). This method can be as accurate as the Taylor series approach, without the need for higher-order derivatives. As an example, the most widely used Runge–Kutta formula is the four-order RK scheme (or RK4), which can be written as:

$$\mathbf{k}_1 = \mathbf{f}(\mathbf{y}_n, t_n) \tag{2.54}$$

$$\mathbf{k}_2 = \mathbf{f}(\mathbf{y}_n + \mathbf{k}_1 \Delta t/2, t_n + \Delta t/2) \tag{2.55}$$

$$\mathbf{k}_3 = \mathbf{f}(\mathbf{y}_n + \mathbf{k}_2 \Delta t/2, t_n + \Delta t/2) \tag{2.56}$$

$$\mathbf{k}_4 = \mathbf{f}(\mathbf{y}_n + \mathbf{k}_3 \Delta t/2, t_n + \Delta t/2) \tag{2.57}$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \left(\frac{\mathbf{k}_1}{6} + \frac{\mathbf{k}_2}{3} + \frac{\mathbf{k}_3}{3} + \frac{\mathbf{k}_4}{6}\right)\Delta t. \tag{2.58}$$

As we can see, this method implies that the interval Δt is subdivided into subsets, providing the points where the derivatives have to be calculated. Runge-Kutta scheme of any order can be constructed, although the derivation of an order higher than four become progressively complicated.

One of the most commonly employed integrators is the so-called *Leap-frog* method, where velocities and position are shifted in time by half a time step. One can distinguish a *kick-drift-kick* (KDK) and *drift-kick-drift* (DKD) method for evolving the system from step n to step $n+1$. The DKD version of the Leap-frog method can be written as:

$$\mathbf{x}_{n+1/2} = \mathbf{x}_n + \mathbf{v}_n \Delta t/2 \tag{2.59}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mathbf{f}(\mathbf{x}_{n+1/2})\Delta t \tag{2.60}$$

$$\mathbf{x}_{n+1} = \mathbf{x}_{n+1/2} + \mathbf{v}_{n+1}\Delta t/2, \tag{2.61}$$

whereas the KDK version is written as:

$$\mathbf{v}_{n+1/2} = \mathbf{v}_n + \mathbf{f}(\mathbf{x}_n)\Delta t/2 \tag{2.62}$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{v}_{n+1/2}\Delta t \tag{2.63}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} + \mathbf{f}(\mathbf{x}_{n+1})\Delta t/2. \tag{2.64}$$

As we can note, the two versions differ on whether one begins by shifting the velocities by half a time step using the forces at the current time step (Kick), and then updating positions for the full-time step(drift), and recomputing forces with new positions before shifting velocities for the full step (Kick), or viceversa. The biggest advantage of the leap-frog scheme lies in its symplectic properties, i.e. they preserve the Hamiltonian structure of the dynamical system. Indeed, all Hamiltonian systems are not structurally stable against non-Hamiltonian perturbations. When an ordinary numerical integrator such as Runge–Kutta is adopted, it can introduce non-Hamiltonian perturbations, and consequently, introduce significant changes in the long-term behavior of the system. To prevent this and preserve the Hamiltonian structure of the system during the time of integration, a symplectic integration scheme must be used, i.e. each step of integration has to be formulated as a canonical transformation. A more detailed description can be found in e.g. Springel (2005).

2.2.4 Resolution

Both the number of particles used and the size of the simulated box affect the properties of the non-linear structures formed. The appropriate setup of the box size and the number of particles is the key to achieving one percent agreement in the power spectrum. Low resolution prevents the capturing of the very non-linear scales, while small simulated volumes lead to the leakage of large modes.

If we assume that L is the side length of the simulation box, only the Fourier modes above the fundamental $k > k_0 = 2\pi/L$ are non-zero. This implies that the structures are formed in regions that are biased representations of a real finite volume in the universe. Power and Knebe (2006) investigated the effect of finite simulation volume of cosmological simulations on the internal properties of cold dark matter halos and their kinematics. They found that the missing large modes lead to a suppression of the strength of clustering and a decrease in the number of most massive haloes. However, they showed that the suppression of long wavelengths does not affect the internal structures of halos. The reduction in the number of most massive haloes also impacts the non-linear power spectrum, whose amplitude is reduced at quasi-linear scales. For example, Schneider et al. (2016) quantified the potential errors due to box size using a suite of simulations performed with the `Pkdgrav3` code, demonstrating that the simulation box size should be larger than $L = 0.5h^{-1}Gpc$ to avoid systematic non-linear finite-volume effects. Moreover, they showed that a box size larger than $L = 2.5h^{-1}Gpc$ is needed to reduce the Gaussian sample variance at the percent level for $k > 0.1hMpc^{-1}$.
On the other hand, the mass resolution of the simulation determines the minimum size of a halo that can be resolved in a given box. This means that if the haloes that make the relevant contribution to the non-linear power spectrum are not accurately resolved, the power spectrum will feature a drop in power. Schneider et al. (2016) computed the minimal mass resolution necessary to achieve percent convergence in the matter power spectrum. Specifically, they found that a mass resolution of $M_p = 10^9 M_\odot$ provides a percent-converged power spectrum up to $k = 10h$ Mpc^{-1} at redshift zero. However, a mass resolution of $M_p = 10^8 M_\odot$ can already guarantee a percent-level convergence at $k \sim 1h$ Mpc^{-1}.

2.3 Machine learning methods

Cosmological simulations play a crucial role in understanding and interpreting observed data. However, a large number of simulations are required for this purpose. As we know, simulating the growth and formation of large-scale structures in the Universe is not an easy task. It involves evolving billions of particles over billions of years and across a vast volume. This can be significantly challenging.

In the last few decades, machine learning techniques have had a considerable impact on cosmology. In particular, within the context of cosmological simulations, there are two important applications we want to focus on: the use of machine learning to generate mock data, and the use of machine learning to enhance and accelerate cosmological simulations. In the first case, using a set of cosmological simulations as a training set, deep learning algorithms can be used to create mock realizations of the universe that are consistent with observed data. As an example, in their study, He et al. (2019) proposed D^{3}M, a deep neural network that learns the nonlinear mapping from first-order perturbation theory linear displacements (Zel'dovich approximation) to the

displacement field of FastPM simulations. They demonstrated that D^3M can more accurately reproduce the cosmic structure in the non-linear regime than Second-order LPT. Furthermore, they demonstrated that the model can be generalized and applied to simulations with cosmological parameters Ω_m and A_s that differ from those used to create the training set. Recently, Li et al. (2021) demonstrated that a generative model can be trained to super-resolve the particle displacement field, allowing for the enhancement of the resolution of a low-cost approximate N-body simulation. Their approach uses high-resolution (HR) image data as the training set and simulates super-resolution (SR) versions of low-resolution (LR) images. They showed that the model was able to simulate SR simulations with a matter power spectrum matching the HR results up to $k \sim 16h^{-1}$ Mpc at all redshifts to the percent level by using only 14 pairs of small-volume LR-HR simulations. Additionally, the model was able to reproduce the HR halo mass function within 10% down to $10^{11}M_\odot$.

Machine learning algorithms can also be used to improve the results of numerical simulations. For example, Giusarma et al. (2019) employed U-net architectures to generate cosmological simulations mimicking the presence of massive neutrino starting from standard ΛCDM simulations. They demonstrated that the method was able to reproduce the 3-D matter power spectrum down to non-linear scales $k < 0.7h$Mpc at redshift $z = 0$. Additionally, the model enables the creation of massive neutrino simulations five orders of magnitude faster than standard methods.

These methods can also be applied to recreate the distribution of baryons in simulations that include only dark matter. In this regard, Tröster et al. (2019) presented an example of using deep generative models and variational auto-encoders trained on full hydrodynamical simulations, to generate the gas pressure distribution in only-dark simulations. They showed the effectiveness of the model in reproducing the power and cross-power spectra from BAHAMAS, a set of full hydrodynamical simulations. Additionally, they demonstrated that once the model is trained, it enables the generation of maps of the tSZ effect much faster compared to standard hydrodynamical simulations. One drawback of these approaches is their use of deep learning as a black box. Consequently, there is no guarantee that the models will perform well outside the training regime or accurately capture the dependence on cosmology.

Another class of examples includes data-driven methods that have been developed to enhance the results of existing numerical simulations. For example, Dai et al. (2018) introduces the Potential Gradient Descent scheme (PGD), a gradient-based method to correct the PM approximation in FastPM and improve the modeling of the matter distribution within halos. PGD models the effect of short-range interactions as an additional particle displacement term, moving the particles towards a minimum of the gravitational potential after band-pass filtering:

$$\mathbf{S} = \frac{4\pi G\bar{\rho}\alpha_{PM}}{H_0^2}\nabla\hat{\mathbf{O}}_l(k)\hat{\mathbf{O}}_s(k)\nabla^{-2}\delta, \tag{2.65}$$

where $\hat{\mathbf{O}}_l(k) = \exp\left(-\frac{k^2}{k_l^4}\right)$ and $\hat{\mathbf{O}}_s(k) = \exp\left(-\frac{k^4}{k_s^4}\right)$ are the low and high pass filter

introduced to remove the long-range force and to reduce the numerical effects induced by the mesh resolution. PGD introduces 3 nuisance parameters fitted on training simulations: α_{PM}, defining the amplitude of the filter and the long and the short-range scale parameters k_l and k_s. Following the PGD model, Dai and Seljak (2021) proposed the Lagrangian Deep Learning (LDL) model as a novel approach to learning the physics controlling baryons in hydrodynamic simulations. Using both Quasi-N-body simulations (FastPM) and full N-body simulations, they demonstrated that the LDL model can generate maps of tSZ, X-ray, stellar mass, and kSZ at several redshifts from the linear density fields. The primary motivation behind the LDL model is to provide an effective description of the underlying physics. The authors argued that such a description should obey symmetries such as rotation and translation invariance, which, when implemented, enable the learning of the effective description of the physical laws instead of the data distribution.

Chapter 3

Weak gravitational lensing

This chapter aims to introduce in detail the weak gravitational lensing formalism. After deriving the main lensing equation from light deflection, I will introduce the convergence and shear observables, explain the connection between these observables and cosmology, and discuss some of the systematic effects in weak lensing analysis. Finally, I will present an overview of the techniques used in weak lensing that are related to this thesis and are used to constrain cosmological parameters.

3.1 Principles of Weak Gravitational Lensing

This section summarizes the weak lensing effects and equations most relevant to the purpose of this thesis. The derivation of these quantities can be found in Seitz et al. (1994); Bartelmann (2010); Kilbinger (2015).

3.1.1 Gravitational Light Deflection

To derive the lens equation, which describes the propagation of light bundles in the presence of massive bodies, we will make three relevant assumptions: we will work within the framework of General Relativity (GR), we will consider the weak field case, assuming that the Newtonian gravitational potential is small ($\Phi \ll c^2$), and we will assume that the massive bodies, which act as sources of the potential, have peculiar velocities smaller than the speed of light.

We may remember that in Equation 1.6 of subsection 1.2.2, we have introduced the deflection of light from the geodesic equation as:

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}, \tag{3.1}$$

where we indicated with $-\Gamma^\mu_{\alpha\beta}$ the Christoffel symbols for a given metric and with λ the affine parameter, the monotonically increasing parametrization of the path. As Seitz et al. (1994), let us consider $\gamma^\mu_0(\lambda)$ an arbitrary fiducial light ray propagating along a null geodesic at the center of an infinitesimally thin light beam. Consider now $\gamma^\mu(\lambda, \boldsymbol{\theta})$, the light ray separated from the fiducial ray by an angle $\boldsymbol{\theta} = (\theta_1, \theta_2)$, such that we can express the separation vector as $\xi^\mu(\lambda, \boldsymbol{\theta}) = \gamma^\mu(\lambda, \boldsymbol{\theta}) - \gamma^\mu_0(\lambda)$. Specifically, we are interested in the transverse components of this separation which can be evaluated by introducing a two-dimensional screen in the 3-dimensional space of the observer perpendicular to the light ray. The four-dimensional vector $\xi^\mu(\lambda, \boldsymbol{\theta})$ can be projected on this screen by introducing its component ξ_1 and ξ_2 on an orthogonal basis of the screen. With this formalism, the physical separation $\xi^\mu(\lambda, \boldsymbol{\theta})$ evolves through the following geodesic deviation equation:

$$\frac{d^2\xi^\mu(\lambda, \boldsymbol{\theta})}{d\lambda^2} = \mathcal{T}(\lambda)\xi^\mu(\lambda, \theta), \tag{3.2}$$

where we introduce the *optical tidal matrix* $\mathcal{T}(\lambda)$, accounting for the effect of the curvature of the local space-time on the light ray trajectory. To express the tidal matrix $\mathcal{T}$ in terms of the matter content of the Universe in the case of a weakly inhomogeneous Universe, we assume a slightly perturbed metric of the Universe, defined as (Schneider, 2006):

$$ds^2 = a^2(\eta)\left(-\left(1+\frac{2\Phi}{c^2}\right)d\eta^2 + \left(1-\frac{2\Phi}{c^2}\right)(dr^2 + f_K^2(r)d\Omega^2)\right). \tag{3.3}$$

This metric is defined as the combination of an FLRW metric, describing a homogeneous and isotropic Universe, and a Newtonian potential produced by local density fluctuations of scales much smaller than the Hubble length. Under this assumption, the tidal matrix $\mathcal{T}$ can be described as sum of two contributions:

$$\mathcal{T} = \mathcal{T}_{bg} + \mathcal{T}_{cl}, \tag{3.4}$$

with a *background* contribution $\mathcal{T}_{bg}$ describing the lensing effect due to the smooth homogeneous Universe, and a *clump* contribute $\mathcal{T}_{cl}$, encoding the tidal effect caused by the inhomogeneities of the density distribution. Seitz et al. (1994) derived the explicit expression for these two terms:

$$(\mathcal{T}_{bg})_{ij} = \frac{4\pi G}{c^2}\frac{\rho_0}{a^5}\delta_{ij}, \tag{3.5}$$

$$(\mathcal{T}_{cl})_{ij} = -\frac{1}{c^2 a^2}\left(2\frac{\partial^2}{\partial\xi_i\partial\xi_j} + \delta_{ij}\frac{\partial}{\partial\xi_3}\right)\Phi, \tag{3.6}$$

with $i, j \in [1, 2]$. If we assume that the source of inhomogeneities is geometrically thin and the variation along ξ can be neglected (also known as *thin-lens* approximation), by injecting the explicit expression of the tidal matrix in the Equation 3.2, we obtain:

$$\frac{d^2\mathbf{x}}{d\chi^2} + K\mathbf{x} = -\frac{2}{c^2}(\nabla_\perp\Phi(\mathbf{x}, \chi) - \nabla_\perp\Phi(0, \chi)), \tag{3.7}$$

which describes how the comoving separation vector $\mathbf{x} = \frac{1}{a}\xi$ evolves w.r.t the comoving distance. In Equation 3.7 we indicate with K the spatial curvature of the Universe, with χ the radial comoving distance, and with $\nabla_\perp = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2})$ the comoving gradient in the 2-dimensional screen perpendicular to the light ray. With the following boundary condition:

$$\mathbf{x}(\chi = 0); \quad \frac{d\mathbf{x}}{dr}(0) = \boldsymbol{\theta} \tag{3.8}$$

the solution of Equation 3.7 for the radial comoving distance is given by:

$$\mathbf{x}(\chi) = f_k(\chi)\boldsymbol{\theta} - \frac{2}{c^2}\int_0^\chi d\chi' f_k(\chi - \chi')[\boldsymbol{\nabla}_\perp\Phi(\mathbf{x}, \chi') - \boldsymbol{\nabla}_\perp\Phi^{(0)}(\chi')], \tag{3.9}$$

with $f_k(\chi)$ the angular comoving distance and Φ the Newtonian potential integrated along the perturbed light path $\mathbf{x}(\chi')$. Note that in Equation 3.9, in the absence of lensing, i.e. in the absence of density fluctuations, the second term vanishes and the observer sees the distance $\mathbf{x}$ under an angle $\boldsymbol{\theta} = \boldsymbol{\beta} = \mathbf{x}(\chi)/f_k(\chi)$. In the presence of lensing, we can introduce the angular deflection $\boldsymbol{\alpha}$, defined as:

$$\boldsymbol{\alpha} = \frac{2}{c^2}\int_0^\chi d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi)}[\boldsymbol{\nabla}_\perp\Phi(\mathbf{x}, \chi') - \boldsymbol{\nabla}_\perp\Phi^{(0)}(\chi')], \tag{3.10}$$

so that, the difference between the apparent angle $\boldsymbol{\theta}$ and the source position $\boldsymbol{\beta}$ defines the lens equation:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\beta}. \tag{3.11}$$

The Born approximation The expression in Equation 3.9 can be simplified by integrating the potential gradient along the *unperturbed* ray, assuming a small deviation between the potential evaluated on the perturbed light path and the one on the unperturbed line of sight. In this assumption, known as *Born approximation* (Schneider, 2006), we can consider a series expansion in powers of Φ. Since we have assumed that Φ is small, we can focus on the first-order approximation. This can be obtained by substituting the zeroth-order solution $x_0(\chi) = f_K(\chi)\boldsymbol{\theta}$ into the comoving separation $x(\chi)$. Thus, the gravitational potential can be expressed as:

$$\Phi(f_k(\chi')\boldsymbol{\theta}, \chi'). \tag{3.12}$$

Equation 3.10 becomes:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}, \chi) = \boldsymbol{\theta} - \boldsymbol{\beta}(\boldsymbol{\theta}, \chi) = \frac{2}{c^2}\int_0^{\chi} d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi)} [\boldsymbol{\nabla}_{\perp}\Phi(f_k(\chi')\boldsymbol{\theta}, \chi') - \boldsymbol{\nabla}_{\perp}\Phi^{(0)}(\chi')], \tag{3.13}$$

representing the *lens equation* under the *Born approximation*.

3.1.2 Shear and Convergence

Formally, the effect of the lens mapping is described by the Jacobian matrix:

$$\mathcal{A}(\boldsymbol{\theta}, \chi) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{1}{f_k(\chi)} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}, \tag{3.14}$$

describing the linear mapping from the lensed image to the unlensed source: Under the Born approximation assumption, the Jacobian matrix can be written as:

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, \chi) = \delta_{ij} - \frac{2}{c^2}\int_0^{\chi} d\chi' \frac{f_k(\chi - \chi') f_k(\chi')}{f_k(\chi)} \partial_i \partial_j \Phi(f_k(\chi')\boldsymbol{\theta}, \chi'). \tag{3.15}$$

If we define the 2D potential, the *lensing potential* as:

$$\psi(\boldsymbol{\theta}, \chi) \equiv -\frac{2}{c^2}\int_0^{\chi} d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi) f_k(\chi')} \Phi(f_k(\chi')\boldsymbol{\theta}, \chi'), \tag{3.16}$$

the Jacobian matrix can be written as:

$$\mathcal{A}_{ij} = \delta_{ij} - \partial_i \partial_j \psi(\boldsymbol{\theta}, \chi). \tag{3.17}$$

From the parametrization of the symmetrical matrix $\mathcal{A}$, we can define the spin-two *shear* field $\gamma = (\gamma_1, \gamma_2)$ and the scalar *convergence* field, κ:

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}; \tag{3.18}$$

hence, the convergence and the shear are defined as the second derivatives of the potential:

$$\begin{aligned}\kappa =&\frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\psi;\\ \gamma_1 =&\frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\psi;\\ \gamma_2 =&\partial_1\partial_2\psi.\end{aligned} \tag{3.19}$$

The two fields γ and κ represent the distortion in the image shape, and the change in the angular size, respectively. The last equations explain the meaning of both convergence and shear. The distortion induced by the convergence is isotropic, so the images are only rescaled by a constant factor in all directions. Setting only κ to be nonzero in the distortion tensor leads to a dilated image, made larger by the convergence, and so magnified. On the other hand, the shear stretches the intrinsic shape of the source along one privileged direction.

3.2 Gravitational lensing by the large scale structure

The previous section introduced the formalism of gravitational lensing by a generic matter distribution. In this thesis, we will focus on cosmological application related to the so-called *cosmic shear*. Light bundles propagating through the Universe are continuously deflected and distorted by the gravitational field of the in-homogeneous mass distribution, the large-scale structures (LSS). The cosmic shear is defined as the coherent distortion of images of distant galaxies due to the tidal field of these structures.
From the study of this quantity, we can infer information about the evolution of structures and the geometry of the Universe. To understand the source of this cosmological information, in this section, I will explore the relationship between the weak lensing convergence and the projected matter distribution, how the weak lensing convergence is connected to the shear, and how the shear is connected to the galaxy ellipticities. Finally, I will describe how systematic effects can affect cosmic shear studies and observations.

3.2.1 Relation between convergence and projected overdensity

The convergence κ will be the pivotal quantity for the purposes of this thesis. It can be seen as the source term of a 2D Poisson equation, hence directly related to the lensing potential ψ:

$$\kappa =\nabla^2\frac{1}{2}\psi \tag{3.20}$$

$$=\frac{1}{c^2}\int_0^\chi d\chi' \frac{f_k(\chi-\chi')}{f_k(\chi)f_k(\chi')}\nabla^2\Phi(f_k(\chi')\boldsymbol{\theta},\chi'). \tag{3.21}$$

and be interpreted as a projected surface density. We can add an additional second-order term along the comoving coordinate, $\partial^2 x/\partial^2 \chi^2$, in order to turn the original 2-dimensional Laplacian into a 3-dimensional Laplacian. This additional term is expected to vanish, considering that positive and negative contributions cancel out when integrating along the line-of-sight under homogeneity arguments. Substituting in Equation 3.20 the 3-D Poisson equation:

$$\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \delta \tag{3.22}$$

we obtain the convergence κ in terms of the 3D density contrast δ:

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^{\chi} d\chi' \frac{f_k(\chi - \chi') f_k(\chi')}{f_k(\chi)} \frac{\delta(f_k(\chi')\boldsymbol{\theta}, \chi')}{a(\chi')} \tag{3.23}$$

with the usual definitions of $\delta = (\rho - \bar{\rho})/\bar{\rho}$, and $\bar{\rho} \propto a^{-3}$. This expression can be interpreted as the weighted projection of the density along comoving coordinates, where the weights are the quantities characterizing the weak lensing optical system, i.e. the source, the deflector, and the observer. If we consider a population of source galaxies with redshift distribution $n(\chi)$, the mean convergence can be computed as:

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi_s} n(\chi) \kappa(\boldsymbol{\theta}, \chi) d\chi, \tag{3.24}$$

with χ_s the limiting comoving distance of the population. By combining Equation 3.23 and Equation 3.24, we obtain:

$$\kappa_{born}(\boldsymbol{\theta}) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^{\chi_s} \frac{d\chi}{a(\chi)} g(\chi) f_k(\chi) \delta(f_k(\chi)\boldsymbol{\theta}, \chi), \tag{3.25}$$

where we define the *lensing efficiency*:

$$g(\chi) \equiv \int_{\chi}^{\chi_s} \mathrm{d}\chi' n(\chi') \frac{f_k(\chi' - \chi)}{f_k(\chi')}. \tag{3.26}$$

Thus, the Born–approximated convergence can be interpreted as the integrated total matter density along the line of sight, weighted by the distance ratios and the normalized source galaxy distribution $n(\chi)\mathrm{d}\chi = n(z)\mathrm{d}z$.

3.2.2 From galaxy ellipticities to cosmic shear

When we use cosmic shear as a cosmological probe, the quantity that is actually measured from observation is the galaxy ellipticities. Subsequently, the shear is measured from the observed ellipticities and the convergence is estimated by using the theoretical relations encoded in Equation 3.19.

In the context of cosmic shear, the distortion due to the inhomogeneities is very weak, hence the images are very weakly distorted and the values of γ and κ are of

the order of a few percent. In this regime, the lens maps the original source onto one unique image, we can not observe multiple images, and the matrix $\mathcal{A}$ is invertible.
At this point, we can introduce a new quantity, namely the *reduced shear*, defined as:

$$g = \frac{\gamma}{1-\kappa}, \tag{3.27}$$

with the same spin-two transformation properties defined for the shear. The reduced shear is the only quantity based on the measurement of galaxy shapes and therefore observable. Indeed, if we define ϵ_{int} the intrinsic complex ellipticity of a source galaxy, the reduced shear g is related to the observed galaxy ellipticity ϵ_{obs} via (Seitz and Schneider, 1997):

$$\epsilon_{obs} = \frac{\epsilon_{int} + g}{1 + g^* \epsilon_{int}}, \tag{3.28}$$

with $|g| \leq 1$ and g^* complex-conjugate of $g = g_1 + ig_2$. In the regime of weak lensing ($|\kappa| \ll 1$), we can approximate this relation to $\epsilon_{obs} \approx \epsilon_{int} + \gamma$ and, taking into account the fact the intrinsic ellipticity of galaxy has no preferred orientation[1], the expectation values of this quantity vanishes $\langle \epsilon_{int} \rangle = 0$. It follows that the expected value of the observed ellipticity of a galaxy becomes:

$$\langle \epsilon_{obs} \rangle \approx \gamma \approx g. \tag{3.29}$$

As a result, the averaged observed ellipticity can be considered an unbiased estimator of the reduced shear. The variance of the observed ellipticity is given by $\sigma^2(\epsilon_{obs}) \approx \sigma^2(\epsilon_{int}) + \sigma^2(\gamma)$, where the variance of the source, referred as *shape noise*, is dominant. In the weak-lensing regime, to reduce the shape noise and detect the lensing signal, the latter cannot be measured from one single galaxy.

3.2.3 From cosmic shear to convergence maps

We have observed in preceding sections that there exists a relationship between the estimated shear γ and convergence κ as shown in Equation 3.19. These expressions can be rephrased in the Fourier domain as follows:

$$\hat{\gamma_1} = \frac{(k_1^2 - k_2^2)}{2}\hat{\psi}; \quad \hat{\gamma_2} = k_1 k_2 \hat{\psi} \quad \hat{\kappa} = \frac{(k_1^2 + k_2^2)}{2}\hat{\psi}; \tag{3.30}$$

from the combinations of these equations, Kaiser and Squires (1993) arrived to the following expression of the reconstructed convergence κ:

$$\hat{\kappa}_\gamma = \frac{(k_1^2 - k_2^2)}{k^2}\hat{\gamma_1} + \frac{2k_1 k_2}{k^2}\hat{\gamma_2}; \tag{3.31}$$

[1] We will see that this assumption is not entirely accurate due to a phenomenon known as intrinsic galaxy alignments (see subsubsection 3.2.4)

with $k^2 = k_1^2 + k_2^2$. It is important to note that the formula lacks definition when $k_1 = k_2 = 0$. As a result, the convergence can only be obtained up to a constant term, leading to what is known as the *mass-sheet degeneracy* phenomenon. The expression in Equation 3.31 can be alternatively expressed in the direct domain as:

$$\kappa(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta'})\gamma(\boldsymbol{\theta'})d\boldsymbol{\theta'} + \kappa_0. \tag{3.32}$$

Here the kernel $\mathcal{D}$ is the Fourier transform of the kernel in Equation 3.31 and κ_0 is the constant offset due to the mass sheet degeneracy. It is evident from this equation that a linear relationship exists between the shear and the convergence and that the latter can be derived through convolution from the shear field. Additionally, this linearity enables us to compute the noise on the convergence by considering the noise on the shear. Under the assumption of uncorrelated shear components, each having a variance of σ_ϵ, the variance of κ can be expressed as:

$$\left\langle \hat{\tilde{\kappa}}^* \hat{\tilde{\kappa}} \right\rangle = \frac{(k_1^2 - k_2^2)^2}{k^2}\sigma_\epsilon^2 + \frac{4k_1^2 k_2^2}{k^4}\sigma_\epsilon^2 = \sigma_\epsilon^2. \tag{3.33}$$

This implies that the estimated κ obtained through the Kaiser-Squires method shares the same noise variance as the input shear. However, there are several limitations to the Kaiser-Squires inversion method. For instance, the estimator is defined in terms of the shear γ, whereas the cosmological observable is the reduced shear $g = 1/(1 + \kappa)$. Although the method can map the large-scale matter distribution effectively since the lensing signal on large scales is well within the weak regime, the convergence may become significant on small scales up to the point that the reduced shear can no longer serve as a direct estimator of the shear. This non-linearity implies that the inversion problem cannot be solved through simple convolution.

This has led to the introduction of alternative mass-mapping techniques. Examples of these approaches include the Seitz and Schneider (1995) inversion methods, which attempt to solve the inversion problem using local differential relations between the convergence and the derivatives of the shear; the method proposed by Pires et al. (2009), which aims to mitigate the impact of the mask using sparse inpainting; the method proposed by Bacon et al. (2006), which uses a minimum variance filter to estimate the convergence from the gradient of the convergence; and methods such as GLIMPSE (Lanusse et al., 2016), where the mass mapping problem, addressed as an ill-posed inverse problem, relies on a robust sparsity-based regularization scheme. In recent years, with the rise of deep learning, deep neural network approaches have also been developed for the mass-mapping problem. For example, the DeepMass method (Jeffrey et al., 2020) estimates the mean posterior of the convergence map by training a Convolutional Neural Network (U-net architecture) on a set of synthetic simulations characterized by non-Gaussian shape noise with cosmological parameters drawn from a broad prior. Additionally, Shirasaki et al. (2021) proposed a new model to denoise weak lensing mass maps based on a Generative Adversarial Network (GAN), and Remy et al.

(2022) proposed an approach that combines elements of deep learning and Bayesian inference to provide a tractable way to sample from the full high-dimensional posterior distribution of convergence maps.

3.2.4 Modelling of systematic effects in weak lensing

Several systematic effects can affect cosmic shear studies, potentially mimicking the shear signal and leading to biased constraints on cosmological parameters. Moreover, modeling of systematic effect is a key ingredient of numerical simulations for weak lensing. For this reason, in this section, I provide a short overview of the most significant systematic effects, including galaxy intrinsic alignment, baryonic feedback, photometric redshift uncertainty, and multiplicative shear bias.

Intrinsic alignment of galaxies

In subsection 3.2.2, we depicted the expectation value of the observed ellipticity as the unbiased estimator of the reduced shear. This outcome arises directly from $\langle \epsilon_{int} \rangle = 0$. In reality, galaxies undergo interactions with the surrounding matter distribution during their formation and evolution, leading to a coherent alignment of their intrinsic shapes. As a result, the intrinsic ellipticity of the source ϵ_{int} can be broken down into two components: the *Intrinsic Alignment* term ϵ^{IA} and the random component ϵ_{ran}.
Dealing with Intrinsic Alignment (IA) is extremely complicated, as its effect cannot be simply removed or predicted, and its nature is strongly connected with the formation of galaxies. The first studies on the intrinsic alignment relied on analytical calculations (Crittenden et al., 2001; Catelan et al., 2001b; Mackey et al., 2002), on dark matter simulations to predict IA from halos (Croft and Metzler, 2000; Heavens et al., 2000a), or on semi-analytic models to describe shape correlations of galaxies and populate halos with them (Heavens et al., 2000a). However, all these methods led to different predictions. Recently, thanks to advancements and increased resolution in hydrodynamical simulations, it has been possible to make progress toward the understanding of this phenomenon. As of now, as highlighted by Kilbinger (2015), we believe that the two most likely mechanisms contributing to the correlation of the intrinsic shape of galaxies are: *Stretching* of halo shapes of galaxies due to the tidal field; *Tidal torquing*, which occurs during the galaxy formation process when the axes of galaxies become aligned. This effect is explained by the exertion of the torquing moment caused by the tidal field.
However, even today, different theoretical models exist to describe the physics of intrinsic alignments. Among them, for example, we have the Non-Linear tidal Alignment model (NLA) (e.g. in Bridle and King (2007)), the tidal torquing model (Hirata and Seljak, 2004; Catelan et al., 2001a), or the combination of both the Tidal Alignment and Tidal Torquing model (TATT) (Blazek et al., 2019). As the work presented in chapter 6 is primarily centered on the NLA model, I will provide a brief overview of this model.

Following Harnois-Déraps et al. (2021), the IA effect can be modelled by a linear relationship between the intrinsic shapes of galaxies and the non-linear projected tidal field s_{ij}:

$$\epsilon_1^{IA} = -\frac{A_{IA}\bar{C}_1\bar{\rho}(z=0)}{D(z)}(s_{xx}-s_{yy}), \quad \epsilon_2^{IA} = -\frac{2A_{IA}\bar{C}_1\bar{\rho}(z=0)}{D(z)}s_{xy}, \tag{3.34}$$

from which the observed ellipticities are computed as:

$$\boldsymbol{\epsilon}_{obs} = \frac{\boldsymbol{\epsilon}_{int}+\mathbf{g}}{1+\boldsymbol{\epsilon}^*_{int}\mathbf{g}}, \quad \text{with} \quad \boldsymbol{\epsilon}_{int} = \frac{\boldsymbol{\epsilon}^{IA}+\boldsymbol{\epsilon}_{\boldsymbol{ran}}}{1+\boldsymbol{\epsilon}^{*,IA}\boldsymbol{\epsilon}_{\boldsymbol{ran}}}. \tag{3.35}$$

The A_{IA} term in Equation 3.34 defines the strength of the tidal coupling, $\bar{C}_1$ is a constant calibrated in Brown et al. (2002), $D(z)$ is the linear growth function introduced in chapter 2 and $\bar{\rho}$ is the average matter density. The IA signal adds an excess correlation to the two-point shear correlation function (also known as cosmic shear GG or shear-shear correlation) with two terms: 1) the intrinsic-intrinsic (II) term, tracing the correlation of the intrinsic shape of two galaxies and 2) and the intrinsic-shear coupling (GI) term, describing the correlation between the intrinsic ellipticity of one galaxy with the shear of another galaxy (Kilbinger, 2015).

$$\begin{aligned} P_{II}(k,z) &= \left(\frac{A_{IA}\bar{C}_1\bar{\rho}(z)}{\bar{D}(z)}\right)^2 a^4(z)P_\delta(k,z), \\ P_{GI}(k,z) &= \frac{A_{IA}\bar{C}_1\bar{\rho}(z)}{\bar{D}(z)}a^2(z)P_\delta(k,z), \end{aligned} \tag{3.36}$$

where $\bar{D}(z) \equiv D(1+z)$ (Harnois-Déraps et al., 2021). Note that the *non-linear* nature of this model arises from the power spectra in Equation 3.36. Indeed, the NLA model is a modification of the Linear Alignment (LA) (Catelan et al., 2001b), involving the replacement of the linear power spectrum with the non-linear one. Under the Limber approximation, the projected angular power spectra for the IA terms become:

$$C_{II} = \int_0^{\chi_{lim}} \mathrm{d}\chi \frac{n^2(\chi)}{a^2(\chi)}P_{II}(k,\chi), \tag{3.37}$$

$$C_{GI} = \frac{3\Omega_m H_0^2}{2c^2}\int_0^{\chi_{lim}} \mathrm{d}\chi \frac{g(\chi)n(\chi)}{a(\chi)}P_{GI}(k,\chi). \tag{3.38}$$

Photometric redshift uncertainty

More stringent constraints on the dark energy parameters are obtained when a tomographic analysis of the weak lensing signal is performed, i.e. when the source galaxies are binned according to their redshift. This means that the cosmic shear signal relies on the redshift of all galaxies used in the analysis, and errors in redshift estimation

can impact the final constraints on the cosmological parameters. However, spectroscopic determination of the redshift is not feasible for such a large quantity of images. Consequently, photometric redshift measurements (also known as "photo-z") are used instead. This means that, among the various sources of error contributing to the dark energy error, we have also to consider errors that arise from using an approximate redshift. This uncertainty typically affects both the central value and the width of the redshift distribution of each tomographic bin. For instance, Hildebrandt et al. (2016) demonstrated that a 1σ uncertainty in the measured mean redshift of each tomographic bin could lead to a reduction of approximately 25% in the precision of the cosmological parameter estimates.

Nevertheless, it is possible to mitigate the impact of statistical errors in photometric redshifts on the overall error budget through accurate calibration and error analysis of the distribution of photometric redshifts (Hildebrandt et al., 2016). The specific accuracy requirements usually depend on the survey.

Multiplicative shear bias

The observed galaxy ellipticities can be affected by processes different from gravitational lensing, such as blurring due to the atmosphere or instrument response, effects from telescope optics, and inaccuracies in the noise model. For example, pixels can suffer from Poisson noise coming from galaxy photons and other noise contributions, such as sky background. All these systematics can impact our statistics and propagate to cosmological parameter estimates. Hence, it is essential to understand the nature of these biases and reduce their impacts.

Some of the most studied sources of bias are: the *noise bias*, which is commonly described as a consequence of the nonlinear relationship between the image pixel and the true shape of the galaxy (Bridle et al., 2009, 2010; Refregier et al., 2012; Melchior and Viola, 2012; Kacprzak et al., 2014); the *model bias*, referring to the result of utilizing an inadequate model to depict the morphologies of the observed galaxy (Refregier et al., 2012; Kacprzak et al., 2014); the *PSF bias*, if the Point Spread Function is not modeled properly (Jarvis et al., 2016; Liaudat, 2022); the *selection bias* pertaining to the effect that the various systematics have on different sets of galaxies, resulting in uneven impacts across galaxy samples (Kacprzak et al., 2012, 2014). (See Pujol et al. (2020) for a full description of different sources of shear bias.)

The shear bias is usually modeled as a multiplicative m and an additive bias c:

$$\langle \boldsymbol{e} \rangle = (1 + m) \langle \boldsymbol{g} \rangle + c \tag{3.39}$$

where $\mathbf{g}$ and $\mathbf{e}$ are the averaged shear and ellipticity. Usually, c is interpreted as an incomplete correction for the point-spread function, introducing a preferred direction in the plane of the image. On the contrary, m encodes in part the propriety of the unobserved galaxies. Hence, if the former can be removed with sufficient knowledge of the experiment, the latter is impossible to know *a priori* (Huff and Mandelbaum, 2017).

Moreover, the algorithms used to measure the ellipticity **e** can introduce a nonlinear response that leads to multiplicative and additive biases interacting with the unknown true ensemble galaxy properties (Massey et al., 2007; Zhang and Komatsu, 2011). One possible way to handle the shear estimation biases consists of using simulations. The shear error can be calibrated by applying a given pipeline for real data to simulation with known input shear. However, this kind of approach requires simulations whose proprieties have to match with the ones of the given real data. An alternative approach has been proposed by Huff and Mandelbaum (2017), dubbed *metacalibration*, which implemented the first method for self-calibration of shear measurements not relying on simulation or deeper calibration fields.

Baryonic effects

Weak lensing observations in current and future cosmological surveys are strongly affected by baryonic feedback. During galaxy formation, processes such as star formation, supernova feedback, radiative cooling, and feedback from black hole accretion, can alter the gravitational potential of a given halo by redistributing the gas inside it. These astrophysical processes can mock signals similar to those expected from varying the cosmological parameters leading to systematic errors in the theoretical predictions (Osato et al., 2021; Rudd et al., 2008; Semboloni et al., 2011a; Zentner et al., 2013; Mohammed et al., 2014; Osato et al., 2015; Harnois-Déraps et al., 2015). Numerical simulations of only dark matter particles provide accurate predictions for the matter power spectrum, achieving a precision of 5 percent over the full relevant range in physical scales and redshift (Schneider et al., 2019). Nonetheless, van Daalen et al. (2011) demonstrated that the assumption implicit in these simulations, that baryons do not influence the large-scale structure, is incorrect. For instance, past studies have shown that the energy injection of active galactic nuclei can heat up or push out large amounts of gas, affecting the matter clustering signal at cosmological scales (Schneider et al., 2019). In theory, complete hydrodynamical simulations are the appropriate method for forecasting the weak lensing signal. However, these simulations are computationally demanding and involve complex physical models that are often approximated using semi-analytical methods, which can produce discrepant results from other hydrodynamical simulations (Schneider et al., 2019). For example, in terms of the matter power spectrum, some hydrodynamical simulations show no effect until $k \sim 1$ h/Mpc (Hellwing et al., 2016; Springel et al., 2018; Chisari et al., 2018), while others predict very strong effects starting at $k \sim 0.1$ h/Mpc (van Daalen et al., 2011; Semboloni et al., 2011a; Huang et al., 2019; Parimbelli et al., 2019; Harnois-Déraps et al., 2015). The angular power spectrum is significantly altered at the angular scale of $\ell \leq 1000$ or $\ell \leq 3000$ according to the assumed statistical uncertainty (Kilbinger, 2015). That is why a significant effort has been devoted to quantifying the impact of baryons on the weak-lensing signal using analytical and semi-analytical methods. For instance, some work (Cooray and Sheth, 2002; Yang et al., 2013; Mead et al., 2015) include baryonic effects by introducing physically motivated free parameters and adjusting

the halo mass-concentration relation based on the halo-model formalism, while others (Schneider et al., 2019; Aricò et al., 2020) consider the impact of the gas, stellar, and dark matter density components individually. Additionally, Eifler et al. (2015); Huang et al. (2019) quantify the changes in the full power spectrum due to baryonic feedback as PCAs, and Dai et al. (2018) introduce a gradient-based method to improve the modeling of matter distribution within halos in only dark matter N-body simulations. All these methods can be calibrated against hydrodynamic simulations or real data. In the past and present, some cosmological results have been achieved by excluding small, highly non-linear scales. Nonetheless, it is evident that for the upcoming weak lensing surveys, accounting for the baryonic effect in calculations is crucial to fully utilize their potential.

3.3 Extracting cosmological information from data

In the previous section, we explored the connections between galaxy ellipticities, shear, convergence, and cosmology. In the following section, I will describe several methods to constrain cosmological parameters from cosmic shear observables.

3.3.1 Second-order statistics

The shear correlation function

Second-order statistics, both in the form of shear 2-point correlation function $\xi_{\pm}(\theta)$, or its counterpart in Fourier space, the angular power spectrum C_ℓ, have been widely used to extract the cosmological information from weak lensing surveys. For example, the real space shear 2-point correlation function (2PCF) can be computed as the averaged multiplication of the ellipticity of galaxy pairs, therefore can be considered the most basic weak lensing statistic. By convention, the two components of the shear are decomposed into *tangential-component* γ_t, and a *cross-component* γ_x. Given ϕ the polar angle of a given direction $\boldsymbol{\theta}$, the two shear components become:

$$\gamma_t = \Re(\gamma e^{-2i\phi}); \quad \gamma_x = \Im(\gamma e^{-2i\phi}). \tag{3.40}$$

From these quantities, we can form three different two-point correlators, $\langle\gamma_t\gamma_t\rangle$, $\langle\gamma_x\gamma_x\rangle$, $\langle\gamma_t\gamma_x\rangle$. If we assume the Universe to be invariant under parity transformation, the $\langle\gamma_t\gamma_x\rangle$ term vanishes, while γ_t remains unchanged and the sign of γ_x is changed. We can combine the remaining correlators to form the two components of the shear 2PCF :

$$\xi_+ = \langle\gamma\gamma^*\rangle(\theta) = \langle\gamma_t\gamma_t\rangle(\theta) + \langle\gamma_x\gamma_x\rangle(\theta); \tag{3.41}$$

$$\xi_- = \Re[\langle\gamma\gamma\rangle(\theta)e^{-4i\phi}] = \langle\gamma_t\gamma_t\rangle(\theta) - \langle\gamma_x\gamma_x\rangle(\theta). \tag{3.42}$$

Using the insight from Equation 3.29, which states that in the absence of intrinsic galaxy alignment, the observed ellipticity is an unbiased estimator of the shear, we can write an estimator of the 2PCF as found in Schneider et al. (2002):

$$\hat{\xi}_{\pm}(\theta) = \frac{\sum_{ij} w_i w_j (\epsilon_{t,i}\epsilon_{t,j} \pm \epsilon_{x,i}\epsilon_{x,j})}{\sum_{ij} w_i w_j}, \tag{3.43}$$

with ϵ_x and ϵ_t the cross and the tangential component of the *observed* shear (the subscript *obs* has been omitted for easy reading). To each galaxy, a weight w has been attributed to encode the measurement uncertainty of the ellipticity. The sum in Equation 3.43 is computed considering every combination of galaxies (i, j), at the position on the sky $\boldsymbol{\theta_i}$ and $\boldsymbol{\theta_j}$, whose separation $|\boldsymbol{\theta_i} - \boldsymbol{\theta_j}|$ falls within a specific angular range centered around $\boldsymbol{\theta}$.

Convergence power spectrum

Analogously to the shear 2PCF, if we assume that the density field δ is invariant under translation and rotation, the convergence 2PCF becomes a function of the modulus of the separation vector $\boldsymbol{\theta}$. Its expression in Fourier space, known as convergence power spectrum P_κ, has been widely used to extract cosmological information from weak lensing surveys:

$$\langle \tilde{\kappa}(\boldsymbol{\ell})\tilde{\kappa}^*(\boldsymbol{\ell'})\rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell'}) C_\kappa(\ell). \tag{3.44}$$

with δ_D Dirac delta function and $\tilde{\kappa}$ the Fourier transform of κ:

$$\tilde{\kappa}(\boldsymbol{\ell}) = \int d^2\theta \kappa(\boldsymbol{\theta}) e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \tag{3.45}$$

function of the 2D wave vector $\boldsymbol{\ell}$, Fourier-conjugate of $\boldsymbol{\theta}$. By computing the square of the Fourier transform of the expression for the convergence κ defined in Equation 3.25, we obtain the quantity $C_\kappa(\ell)$. Inserting this expression in Equation 3.44, the angular power spectrum of the convergence field can be computed as:

$$C_\kappa(\ell) = \frac{9\Omega_m^2 H_0^4}{4c^4} \int_0^{\chi_{lim}} \mathrm{d}\chi \frac{g^2(\chi)}{a^2(\chi)} P_\delta\left(k = \frac{\ell}{f_K(\chi)}, \chi\right), \tag{3.46}$$

where P_δ defines the matter power spectrum of the density contrast.
However, note that the Equation 3.46 has limited applicability as it is based on a number of approximations that are only accurate under certain conditions. For example, the computation of the full power spectrum consists of a triple integral for each ℓ. This computation is simplified in the *Limber approximation* (Limber, 1953; Kaiser, 1992; Simon, 2007), i.e. considering only the mode lying in the plane of the sky and disregarding the correlation in the direction of the observer's line of sight. Furthermore, in the flat-sky limit, spherical harmonics are replaced by Fourier transforms. Finally, the small-angle approximation is applied by utilizing only the first order of the expansion of the trigonometric function.

3.3.2 Non-Gaussian information from weak lensing data

Traditional cosmological analyses rely on measurements of the two-point statistics, either the shear two-point correlation functions or its Fourier transform, the lensing power spectrum. However, the two-point statistics are only optimal for Gaussian fields and do not fully capture the non-Gaussian information imprinted in the lensing signal at the scales that future surveys will be able to access (e.g. information encoded in the peaks and in the filamentary features of the matter distribution). This has led to the introduction of a number of higher-order statistics to access the non-Gaussian information from weak lensing data: the weak lensing one point PDF (Liu and Madhavacheril, 2019; Uhlemann et al., 2020; Boyle et al., 2021), lensing peak counts (Liu et al., 2015a,b; Lin and Kilbinger, 2015; Kacprzak et al., 2016b; Peel et al., 2017b; Shan et al., 2018; Martinet et al., 2018; Ajani et al., 2020; Harnois-Déraps et al., 2021; Zürcher et al., 2022), Minkowski functionals (Kratochvil et al., 2012; Petri et al., 2013), moments of mass maps (Gatti et al., 2021), wavelet and scattering transform (Ajani et al., 2021; Cheng and Ménard, 2021), and 3 point statistics (Takada and Jain, 2004; Semboloni et al., 2011b; Rizzato et al., 2019; Halder et al., 2021).

Recently, machine learning-based methods (Fluri et al., 2019; Kacprzak and Fluri, 2022; Fluri et al., 2021; Jeffrey et al., 2021; Fluri et al., 2022), and Bayesian forward-modeling frameworks (Alsing et al., 2016, 2017; Porqueres et al., 2021; Boruah et al., 2022b) have also been introduced to attempt to fully account for the non-Gaussian content in the weak lensing signal. Unlike the methods described above, these approaches are designed to access the full field-level information content.

In this section, I will give an overview of some of the most promising methodologies capable of extracting the cosmological information encrypted in the non-Gaussian part of the weak lensing field. Specifically, I will focus on the methodologies related to the work presented in this thesis.

Lensing peak counts

One of the most widely used higher-order statistics for weak lensing is undoubtedly the peak-count statistics. In its simplest application, it can be computed by identifying the local maxima in the shear or convergence maps, specifically by considering the pixels with values higher than their eight neighbors. While obtaining the peak count statistic from a data-processing perspective is relatively straightforward, predicting it for a specific parameter set or cosmological model is highly challenging. Like many other higher-order statistics, the issue with peak counts is that they cannot be directly linked to an analytical expression through a straightforward formula. Accurately modeling peak counts remains an unresolved question in current cosmology. Thus far, three distinct methods have been investigated: analytical approaches, modeling approaches based on N-body simulations, and stochastic forward approaches.

Examples of analytical models for peak counts include those proposed by Maturi et al. (2010, 2011) and Fan et al. (2010). These approaches are based on the peak

theory of the Gaussian random field, which involves considering the galaxy-shape noise as a Gaussian random field and the weak lensing signal as a foreground field. According to random field theory, the probability of finding a local maximum is a function of the noise level and the foreground, allowing for the construction of a peak function defined as a density function of signal-to-noise ratio. However, there are several challenges associated with the analytic approach. For instance, when accounting for realistic effects such as masked regions in the data, bias from photometric redshifts, and errors in shape measurements, the performance of these models significantly deteriorates. Additionally, incorporating additional cosmological effects, such as intrinsic alignment, poses an extremely challenging task.

Another approach involves utilizing N-body simulations. The methodology for models based on N-body simulations is as follows. First, the structure formation of the universe is simulated in a densely sampled cosmological parameter space. Subsequently, weak lensing maps are generated from the output of these N-body simulations using either a full ray-tracing algorithm or the Born approximation. The peak count statistic is then computed for the sampled cosmological parameters and subsequently interpolated to other points within the cosmological parameter space. A vast number of papers can be found in the literature that employs this strategy (Dietrich and Hartlap, 2010; Liu J. et al., 2015; Kacprzak et al., 2016a; Li et al., 2019; Ajani et al., 2020). These works differ in terms of the varied cosmological parameters, the number of sampled points in the cosmological parameter space, and the adopted interpolation strategy (such as bilinear smoothing splines, multidimensional Radial Basis Function, Gaussian Processes, etc). It is important to note that these approaches, which depend on N-body simulations, have the significant drawback of being computationally expensive and time-consuming. However, unlike analytical methods, they have the advantage of easily incorporating the exact survey mask and shape noise when performing analyses on real data.

Finally, an example of stochastic forward approaches was presented by Lin, Chieh-An and Kilbinger, Martin (2015) with their algorithm `CAMELUS`. This approach can be viewed as a hybrid between an analytical and a simulation-based approach. It employs a probabilistic approach to model peak counts by generating *fast simulations*, that require much less computational effort than standard N-body simulations. It is implemented as follows: halos are sampled from a given mass function, density profiles are assigned to the halos, and the halos are randomly distributed within the field of view. Subsequently, convergence maps are created through ray-tracing, and the analysis of peak counts is performed. Despite being less computationally expensive compared to simulation-based approaches and being flexible (in the sense that is straightforward to include effects such as masking, photo-z errors, IA, and baryonic feedback), this method does have limitations. In a study by Peel et al. (2017a), it was employed to investigate the constraining power of peak counts on cosmological parameters $(\Omega_\mathrm{m}, \sigma_8, w_0)$ in a *Euclid*-like scenario. Their findings revealed that while the method produced tight contours, they were also significantly biased in the $(\Omega_\mathrm{m}, \sigma_8)$ plane.

Despite the challenges associated with them, several studies have shown that the weak-lensing peak counts provide a way to capture information from non-linear structures that is complementary to the information extracted by the power spectrum. Following I report some important results on real data constraints.

The first peak counts analysis on observations corresponds to two independent analyses performed by Liu X. et al. (2015) and Liu et al. (2015a).
Liu X. et al. (2015) presented constraints from a peak-counts analysis, as measured from the Canada-France-Hawaii Telescope Stripe 82 Survey (CS82). They showed that for a flat ΛCDM model, the cosmological constraints on (Ω_m, σ_8) are consistent with those derived from the cosmic shear correlation analysis, while the degeneracy direction of the two parameters is flatter compared to the one from the correlation studies. Additionally, they proved that their results from peak analyses were fully consistent with those from both WMAP9 and Planck.
Liu et al. (2015a) conducted a study to assess the constraining power of weak lensing peak counts on cosmological parameters $(\Omega_\mathrm{m}, \sigma_8, w)$ using CFHTLenS data covering an area of approximately ~ 154 deg^2. They employed a set of ray-tracing N-body simulations for a grid of 91 cosmological models, reproducing the CFHTLenS observations in terms of galaxy sky positions, redshifts, and shape noise. Their findings revealed that the results obtained from peak counts were comparable to those derived from the power spectrum. However, when combined, the contour areas of the $(\Omega_\mathrm{m}, \sigma_8)$ constraints were reduced by a factor of ~ 2. However, both statistics were unable to constrain the parameter w without the use of external data.

Another study by Kacprzak et al. (2016a) performed a peak counts analysis on the Dark Energy Survey Science Verification data set using aperture mass maps from the set of N-body simulations generated by Dietrich and Hartlap (2010) spanning the $(\Omega_\mathrm{m}, \sigma_8)$ plane. They only considered peaks in bins of low and medium S/N, since the correction of systematic effects was more uncertain for high S/N peaks. They also investigate the impact of uncertainties on shear multiplicative bias and the mean of the redshift distribution, finding that those could affect the observed peak up to $\sim 10\%$ and $\sim 16\%$, respectively. However, their finding showed that the overall impact on the cosmological constraints was comparable to the 2-pt statistics.

In another study by Harnois-Déraps et al. (2021) cosmological parameters were constrained through a joint analysis of tomographic two-point shear correlation functions and peak counts on Dark Energy Survey (DES-Y1) data. To model the peak counts dependence from cosmological parameters, they adopted the set of *cosmo*-SLICS simulations spanning the $(\Omega_\mathrm{m}, \sigma_8, h, w_0)$ parameter space. Additionally, they assessed the impact of baryons, intrinsic alignment, and source-lens clustering. Their finding showed an improvement of $\sim 20\%$ on the S_8 constraints in the joint analysis compared to the correlation function analysis.

Considering the above studies, we can conclude that, despite the difficulties and the limitation related to the absence of an analytical model capable of predicting the observed

signals, peak counts statistics have proven to be a powerful method to extract non-Gaussian cosmological information from weak lensing data. Nevertheless, it is evident that further development in the modeling of baryonic feedback and intrinsic alignment is necessary for future analyses. Harnois-Déraps et al. (2021) argued that an improvement of $\sim 20\%$ in the constraint on S_8 can be achieved by modeling and marginalizing over the impact of baryons, as demonstrated in previous studies on the DES-Y1 3×2 points. This improvement was made possible by including additional small-scale elements in the $\xi_{\pm}$ vector. A similar improvement could be achieved with peak statistics by increasing the signal-to-noise upper limit, but this would only be feasible with accurate modeling of baryons at the smallest scales. Furthermore, Harnois-Déraps et al. (2021) identified intrinsic alignment as the primary limiting factor in their analysis. Their findings indicated that it can impact cosmological constraints at a level similar to baryons and potentially up to the 1σ level.

Machine learning methods for cosmological inference

Over the past few years, machine learning approaches have enhanced our ability to interpret and take the most from cosmological data. In this section, I will focus on how machine learning techniques allow us to infer cosmological parameters, density fields, or simply discern between different cosmological models from weak gravitational lensing maps. The first application of machine learning within the weak lensing scenario can be traced back to the work of Schmelzle et al. (2017). In their paper, the authors demonstrated that a classification model based on a Deep Convolutional Neural Network (DCNN) can be used to discriminate between different cosmological models. Specifically, the DCNN was trained to learn the relation between five distinct cosmological models in the $\sigma_8 - \Omega_m$ plane and the convergence maps they generate. By comparing the results with higher-order statistics such as skewness and kurtosis, they showed that the network outperforms these statistics in terms of mean discrimination efficiency (85% vs 70%). Peel et al. (2019) addressed a similar classification problem by using a CNN model to discriminate between the standard ΛCDM model and Modified Gravity models (MG), that exhibit the same Gaussian weak-lensing statistics but include massive neutrinos. By including the full redshift information, the CNN successfully distinguished between the ΛCDM model and the MG models with an accuracy of 80% percent on noise-free data, outperforming peak count statistics even though its correct classification rate decreased on noisy data.

Another example is the work by Merten et al. (2019), which investigated the capacity of three different machine learning techniques in discriminating nine different cosmological models, which include the effect of massive neutrinos and modified gravity. The study revealed that the CNN provided the most discriminatory representation compared to summary statistics such as the power spectrum, peak counts, and Minkowski functionals when employed on noise-free maps.

A different class of example consists of Machine learning approaches used to infer the cosmological parameters. Within this class, we can make a further distinction:

The methods aimed to find a low-dimensional summary statistic and those trained to directly provide unbiased estimates of the cosmological parameters. In the former scenario, the output of the neural network is employed in a standard inference pipeline, in a procedure similar to other higher-order statistics.

An example falling into this category is the work by Gupta et al. (2018), which demonstrated that a 2D CNN could generate confidence contours in the $\sigma_8 - \Omega_m$ plane $\sim 5\times$ tighter than the power spectrum and $\sim 4\times$ tighter than the lensing peak counts. The CNN was trained on simulated noiseless convergence maps covering 96 different cosmological models.

A step further was taken by Ribli et al. (2018) who proposed an improved CNN capable of providing more accurate estimates of σ_8 and Ω_m. Following the approach and using the same simulations of Gupta et al. (2018), they employed modifications to the CNN such as additional activations, an expanded number of filters, and a block structure. The final results proved that the CNN was capable to estimate with no bias both σ_8 and Ω_m with $\sim 2\times$ smaller errors than peak counts.

In a study conducted by Matilla et al. (2020), the Neural Network developed in Ribli et al. (2018), was used to constrain the cosmological pair σ_8 and Ω_m. The results were compared to those obtained from the power spectrum, peak counts, and Minkowski functionals, single and combined. The network improved the results by $\sim$ 20% compared to this combination, even in the presence of noise. Additionally, several saliency methods were employed to identify the most relevant features of the convergence maps for inferring the cosmological parameters. The results showed that the most relevant pixels were those with extreme values. Specifically, in the noise-free setting the most relevant regions corresponded to negative convergence values, while in the presence of realistic noisy maps, the regions containing positive convergence values held the majority of the features relevant to the inference task.

Finally, a study by Fluri et al. (2018a) investigated the performance of the CNN for different levels of shape noise and smoothing scales. The results indicated an improvement in the estimation of σ_8 and Ω_m compared to the standard power spectrum, but also a comparatively smaller enhancement with increasing noise level and smoothing scale.

The first example of a Machine learning approach applied to real observations dates back to the work of Fluri et al. (2019). They presented the results obtained with a CNN from the KiDS-450 tomographic dataset, aiming to constrain Ω_m, σ_8, and the intrinsic alignment amplitude A_{IA}. A 30% improvement was observed compared to the power spectrum analysis. This work was further extended in Fluri et al. (2022), where a full wCDM analysis of the KiDS-1000 weak lensing maps was performed using a Graph-Convolutional Neural Network (GCNN). In this study, a different approach was adopted for the inference procedure, employing a likelihood-free inference method, the Gaussian Process Approximate Bayesian computation (GPABC), to constrain the cosmological parameters. The GCNN analysis yielded a 16% improvement in constraining the degeneracy parameter$S_8 \equiv \sigma_8\sqrt{\Omega_m/0.3}$ compared to the power spectrum analysis.

An example of a machine learning approach that directly provides unbiased estimates of cosmological parameters is presented in the work by Jeffrey et al. (2021). Although this is another example of Likelihood free inference analysis on observed weak lensing data, it differs significantly from the work by Fluri et al. (2022). Jeffrey et al. (2021) compared the posterior distribution for different compressed summary statistics, including the power spectrum, peak counts, the combination of these two statistics, and a summary statistic obtained by compressing the convergence maps directly using a Deep Convolutional Neural Network. These compressed statistics were then compared to the (compressed) DES Science Verification data and used to estimate the posterior distribution through a Neural Likelihood Estimator, specifically the `PyDelfi` package (Alsing et al., 2019). The results indicated that this approach was consistent and achieved slightly tighter constraints compared to the power spectrum.

In conclusion, machine learning methods have immense potential to extract additional information beyond second-order statistics. However, despite their potential, several questions still need to be addressed. For example, understanding how the introduction of systematic effects such as baryonic feedback can impact the final constraining power, how to control and check systematic errors during the training phase when applying these methods to real data, and finally, how to interpret the learned features from a physical perspective since they are no longer directly associated with physical properties.

Bayesian hierarchical modeling of weak lensing

A different way to extract information encoded at non-linear scales is to adopt the Bayesian hierarchical inference approach. A significant difference between this approach and other higher-order statistics is that the former does not rely on summary statistics. Instead, it aims to capture the full information content of the data through a full-field approach.

The Bayesian forward modeling approach involves using a given model to predict observations and then comparing these predictions with real observations to infer the parameters of the model. The *hierarchical* nature of this methodology comes from the fact that a complex inference task, such as the weak lensing inference task, can be broken down into several hierarchies of elements, which can be understood and from which we can sample from (Heavens et al., 2018). Indeed, inferring cosmological parameters from weak lensing observations is not an easy task. We have seen that key aspects of weak lensing measurements are the measurement of the galaxy shapes and the measurement of their redshift. However, to obtain those from pixelized images or photometric measurements, we need to properly model point spread functions, considering seeing effects, pixel noise, and other instrumental factors that can affect the measurements. Then, once we have measurements of the observed shape galaxies, we require a model that relates cosmic shear to cosmology in order to infer the cosmological parameters. Constructing a global hierarchical model would, in principle, allow us to account for all the statistical dependencies between the model components. This

approach has the advantage of easily incorporating systematic effects such as intrinsic alignments, baryon feedback, and uncertain redshift distributions or easily combining multiple cosmological probes by joint simulations. Furthermore, it is advantageous to directly extract information from the raw pixel data without relying on summary statistics, as it preserves all the available information.

Although this procedure offers advantages, such as the ease of analyzing sub-elements, it complicates the propagation of uncertainties through the pipeline and the accounting of the model inter-dependencies. Therefore, it requires the correction of potential biases introduced by these factors (see Schneider et al. (2015); Alsing et al. (2016) for a more detailed discussion).

Although several forward modeling approaches for cosmic shear can be found in the literature (Schneider et al., 2015; Alsing et al., 2016, 2017; Böhm et al., 2017; Porqueres et al., 2021, 2022, 2023), the first application of this approach to observed data dates back to the work of Alsing et al. (2017). In their paper, the authors applied two different Bayesian hierarchical inference procedures to infer shear power spectra, shear maps and cosmological parameters from Canada–France–Hawaii Telescope (CFHTLenS). They found cosmological parameter constraints consistent with previous CFHTLenS results, as well as constraints on (σ_8, Ω_m) that showed a 2σ tension with the Plank 2015 study, consistent with previous analysis. Additionally, they extended the ΛCDM model to include massive neutrinos and showed that they were capable of inferring the total neutrino mass to $\sum m_\nu < 4.6\text{eV}$ from the CFHTLenS data alone. They also constrained the parameters $p_1 = -0.25^{+0.53}_{-0.60}$ and $p_2 = -0.15^{+0.17}_{-0.15}$ when considering the linear redshift-dependent photo-z bias $\Delta z = p_2(z - p_1)$. Moreover, they investigated the computational cost of the Bayesian inference schemes and argued that for future stage IV surveys, the computational cost would be comparable to that of standard estimator-likelihood sampling methods. However, it is worth noting that the methods implemented assumed a Gaussian lensing field, which, while appropriate at large scales, is sub-optimal at smaller scales.

This limitation was addressed by Porqueres et al. (2021) who presented a Bayesian hierarchical scheme complemented with a gravity model capable of inferring the matter density field, the lensing, and the matter power spectra from cosmic shear data. Specifically, the model is an extension of the Bayesian Origin Reconstruction from Galaxies (BORG, Jasche and Wandelt (2013); Jasche and Wandelt (2013)) framework incorporating weak lensing. BORG includes a physical description of the dynamics of dark matter, enabling a more accurate and realistic representation of the data and thus the ability to exploit information beyond second-order statistics. The model was further developed in Porqueres et al. (2022) and Porqueres et al. (2023).

Another example is the work of Boruah et al. (2022b), who extended the *KaRMMa* algorithm to infer cosmological parameters through a map-based approach. *KaRMMa* was capable of generating convergence maps that accurately reproduced the 1-point and 2-point statistics of the convergence field, as well as the peak and void statistics, by forward-modeling the convergence maps as a log-normal field conditioned on cosmic

shear data. In particular, Boruah et al. (2022b) addressed several limitations of the original *KaRMMa* implementation, including computational and memory issues. They also improved upon the algorithm by considering correlations between different tomographic bins and allowing for sampling of the cosmological parameters of the log-normal prior. In an LSST-Y10 scenario, the authors provided a forecast on the constraints in the $\sigma_8 - \Omega_8$ plane, demonstrating an improvement of $\sim 30\%$ compared to traditional cosmic shear analysis.

Despite their enormous potential, this class of approaches has a significant limitation: they often lead to high-dimensional problems and therefore require advanced statistical sampling techniques.

Chapter 4

Statistical inference techniques for cosmology

The purpose of this chapter is to provide an introduction to several Bayesian statistical tools directly or indirectly related to chapter 7. Specifically, after a brief introduction to Bayesian statistics, I will discuss tools such as the Markov Chain Monte Carlo (MCMC) and the Hamiltonian Monte Carlo (HMC). Additionally, I will provide a concise overview of the main Likelihood Free Inference methods along with the primary linear dimension reduction methods proposed for cosmological applications. The following results are based on Verde (2010); Heavens (2009); Neal et al. (2011)

4.1 Bayesian inference

Let us consider a dataset $\mathbf{d}$ and assume a cosmological paradigm, such as the ΛCDM, as our hypothesis or model, characterized by the cosmological parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_N\}$. From the definition of joint $\mathcal{P}(x_1, x_2)$ and condition probability $\mathcal{P}(x_2|x_1)$, it can be demonstrated the following identity:

$$\mathcal{P}(x_1, x_2) = \mathcal{P}(x_1)\mathcal{P}(x_2|x_1) = \mathcal{P}(x_2)\mathcal{P}(x_1|x_2). \tag{4.1}$$

This identity can be written in terms of the conditional probability of the cosmological parameters, given the data $\mathbf{d}$:

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{d}, \mathcal{M}) = \frac{\mathcal{P}(\boldsymbol{\theta}|\mathcal{M})\mathcal{P}(\boldsymbol{d}|\boldsymbol{\theta}, \mathcal{M})}{\mathcal{P}(\boldsymbol{d}|\mathcal{M})}. \tag{4.2}$$

This last expression is known as **Bayes theorem** and represents the basis of Bayesian statistical inference. The quantity $\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{d}, \mathcal{M})$ is the *posterior* probability distribution, $\mathcal{P}(\boldsymbol{\theta}|\mathcal{M})$ the *prior*, $\mathcal{P}(\boldsymbol{d}|\boldsymbol{\theta}, \mathcal{M})$ the *likelihood*, and $\mathcal{P}(\boldsymbol{d}|\mathcal{M})$ the *evidence*.

Prior In the Bayesian approach, we aim to find the probability distribution for the cosmological parameters associated with our model given the dataset, i.e. the quantity $\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{d})$. A standard inference approach, consists in defining the evidence, assuming a given functional form for the likelihood, and defining a prior. The prior describes what we already know about the parameters and can come from the posterior of a previous experiment or theoretical knowledge (for example, we know that the total matter density Ω_m has to be positive). In other cases, when no information is available, the principle of indifference can be adopted. The easiest way to represent the state of ignorance is to assume that the prior is constant. This choice is usually referred to as *flat prior* and implies that all parameter values have an equal probability. However, when we assign a uniform prior and then perform a re-parametrization, it may happen that this re-parameterization leads to a non-uniform distribution and hence results in a *not uninformative* prior. To handle this problem, Jeffreys (1946) proposed a prior consistent under re-parametrization. Jeffreys prior is equivalent to a flat prior after a variance-stabilizing transformation and is defined as follows:

$$p(\boldsymbol{\theta}) \propto \sqrt{\det \boldsymbol{F}(\boldsymbol{\theta})} \tag{4.3}$$

with $\boldsymbol{F}(\boldsymbol{\theta})$ Fisher information matrix.
Moreover, other different choices from the uninformative prior are possible (Jaynes, 1968; Bernardo, 2005).
In particular, if the prior arises from theoretical consideration, it is important to determine if the results depend on the prior assumptions. If there is a significant dependence, it indicates that the data does not have sufficient statistical power to constrain the parameters.

Evidence The evidence $\mathcal{P}(\boldsymbol{d})$ plays a crucial role when we consider multiple theoretical models and need to choose the most probable one. However, if we are only interested in estimating cosmological parameters, the Equation 4.2 can be rewritten as:

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{\mathcal{P}(\boldsymbol{\theta})\mathcal{P}(\boldsymbol{d}|\boldsymbol{\theta})}{\mathcal{P}(\boldsymbol{d})}. \tag{4.4}$$

Here, the evidence $\mathcal{P}(\boldsymbol{d})$ is used only to normalize the probability:

$$\mathcal{P}(\boldsymbol{d}) = \int d\boldsymbol{\theta}\mathcal{P}(\boldsymbol{d}|\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta}), \tag{4.5}$$

and since the probabilities of the parameters being inferred do not depend on the evidence, it is often ignored.

Likelihood If the prior is assumed to be flat, estimating the cosmological parameter becomes a matter of computing the posterior $\mathcal{P}(\boldsymbol{\theta}|\mathbf{d})$ from the likelihood $\mathcal{P}(\mathbf{d}|\boldsymbol{\theta})$ (often indicated as $L(\mathbf{d}|\boldsymbol{\theta})$ or simply as $L(\boldsymbol{\theta})$), as shown in Equation 4.4, which becomes:

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) \propto L(\mathbf{d}|\boldsymbol{\theta}). \tag{4.6}$$

By maximizing the likelihood, we can determine the peak of the posterior distribution for the parameters. However, if a non-flat prior is selected, the maximum likelihood estimate may not correspond to the peak in the posterior distribution.

It is worth noting that the likelihood function is the only place where the data appear. Thus, given a model, the likelihood $L(\mathbf{d}|\boldsymbol{\theta})$ contains all the information about the cosmological parameters $\boldsymbol{\theta}$ that are contained in the data $\mathbf{d}$. In some cases, the likelihood is known and can be expressed as an analytical function. For example, if the data follow a Gaussian distribution, and the covariance matrix $\boldsymbol{C}$, which describes the covariance between the data, is independent of the parameter $\boldsymbol{\theta}$, the likelihood function can be written as a multivariate Gaussian function. Let $\mathcal{L}$ denote the negative logarithm of the likelihood; we can write:

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta}) = \frac{1}{2}\left[(\boldsymbol{d} - \mu(\boldsymbol{\theta}))^T \boldsymbol{C}^{-1} (\boldsymbol{d} - \mu(\boldsymbol{\theta}))\right], \tag{4.7}$$

where $\mu(\boldsymbol{\theta})$ represents the theoretical predictions as functions of the parameter $\boldsymbol{\theta}$. The Gaussian likelihood assumption is highly prevalent in cosmological analysis and is often justified by the principle of maximum entropy or the central limit theorem (Verde, 2010). As a simple example, consider situations where the data do not follow a Gaussian distribution but are still drawn from a distribution with finite variance. In such cases, the data can be binned, so that the resulting error distribution for each bin can be approximated by a Multivariate Gaussian for the central limit theorem. As I will describe in section 4.2, in other cases the likelihood can be estimated using density estimation approaches, within the context of Likelihood-free Inference methods. Alternatively, as described in subsubsection 3.3.2, Bayesian forward-modeling methods can be used, allowing the integration of observations into a forward model and the exact reconstruction of the likelihood.

4.1.1 Parameter inference

Once the likelihood is defined, we need to explore the parameter space to derive the posterior distribution defined in Equation 4.2. In principle, we could locate the peak of the likelihood by evaluating it on a sufficiently fine grid. However, when the number of parameters grows, this procedure easily becomes impossible because the number of grid points grows exponentially with the dimensionality of the parameter space. In general, this approach is suboptimal anyway because most of the parameter space is uninteresting, being characterized by a small likelihood. Several alternative approaches to sample the likelihood surface exist. Below, I will briefly introduce two of the most commonly used algorithms in cosmology: the Markov Chain Monte Carlo and the Hamiltonian Monte Carlo.

4.1.2 Markov Chain Monte Carlo (MCMC)

Let $p(\boldsymbol{\theta})$ be a generic probability distribution (e.g. the likelihood or the posterior), which we refer to as the *target distribution*. Starting from an initial point $\boldsymbol{\theta}_0$, the Markov Chain Monte Carlo (MCMC) algorithm samples the parameter space through a random walk. In practice, its goal is to generate a set of points whose distribution is an approximation of the target distribution. The target distribution can be expressed as:

$$p(\boldsymbol{\theta}) \simeq \frac{1}{N} \sum_{i=1}^{N} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i), \tag{4.8}$$

where δ is a delta function, N refers to the number of iterations or samples, and $\boldsymbol{\theta}$ refers to the set of n values of parameters of the $n-$dimensional parameter space. From this approximated target density, quantities of interest, such as the mean and variance, can be estimated.

One of the most popular MCMC algorithms is the Metropolis-Hastings (MH) algorithm, which works by sampling a new point $\boldsymbol{\theta}_{i+1}$ in the parameter space based on a *proposal* probability distribution function $q(\tilde{\boldsymbol{\theta}}_{i+1}|\boldsymbol{\theta}_i)$.

The MH algorithm evaluates the likelihood of the proposed sample by calculating the ratio of the posterior probabilities at the current sample and the proposed sample:

$$a = \min\left(1, r \equiv \frac{p(\tilde{\boldsymbol{\theta}}_{i+1}) q(\boldsymbol{\theta}_i|\tilde{\boldsymbol{\theta}}_{i+1})}{p(\boldsymbol{\theta}_i) q(\tilde{\boldsymbol{\theta}}_{i+1}|\boldsymbol{\theta}_i)}\right), \tag{4.9}$$

then the step to the proposed value $\tilde{\boldsymbol{\theta}}_{i+1}$ is accepted with probability equal to a if the acceptance ratio parameter $r < 1$, hence $\boldsymbol{\theta}_{i+1} = \tilde{\boldsymbol{\theta}}_{i+1}$. Otherwise, if $r > 1$ the step to the proposed value is rejected, and $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$. The process is repeated until the density of the point $\{\boldsymbol{\theta}_i\}_{i=1,..N}$ is proportional to the target distribution for $N \to \infty$. Regarding the proposal distribution, it must be finely tuned to explore the parameter space efficiently. It is important to note that if the proposal distribution is too small,

the distance between two points is also small, and the chain takes a long time to explore the entire parameter space, resulting in almost all proposed steps being accepted. On the other hand, if the proposal distribution is too large, the proposed sample can be far away from the current sample, leading to a low acceptance rate. As a result, the algorithm may get stuck in regions of low probability, failing to explore the parameter space effectively.

4.1.3 Hamiltonian Monte Carlo (HMC)

The MCMC algorithm is doomed to fail in high-dimensional and correlated spaces where a large number of steps are required to effectively explore the space. The Hamiltonian Monte Carlo (or Hybrid Monte Carlo, HMC) algorithm improves the sampling process by using a physical analogy to guide the sampling process. In practice, a new point is proposed by simulating the particle dynamics inside a potential energy framework induced by the target distribution. This dynamic is described using the Hamiltonian formulation. In Hamiltonian Dynamics (HD), the state of a physical system is described by a pair of d-dimensional vectors (q, p). The variable q defines the *position* vector, while p defines the *momentum* vector. The system's dynamics are expressed by the *Hamiltonian equations*:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{4.10}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \tag{4.11}$$

here $H(q, p, t)$ is the *Hamiltonian* of the system, which represents the total energy of the system, i.e. the sum of the *potential energy* $U(q)$ and *kinetic energy* $K(p)$. The Hamiltonian is time-independent due to the conservation of energy.
The HD formulation has three important properties that are advantageous for the HMC algorithm: reversibility, i.e. mapping the state $(q(t), p(t))$ at time t to the state at $t + s$ is a reversible process; volume preservation, i.e. the mapping preserves the volume in the (q, p) space; Hamiltonian conservation, i.e. $H(q, p)$ is time-independent with $dH/dt = 0$.

The first step in applying the HD to sample from the target distribution $p(\boldsymbol{\theta})$, is to derive appropriate energy functions. In statistical mechanics, an important result links the probability density $f_S(s)$ for a particle to be observed in a state s with the energy $E(s)$ of that state, via:

$$f_S(s) = \exp(-E(s)). \tag{4.12}$$

By inverting this relation and returning to our notation, we can obtain the energy for the target distribution $p(\boldsymbol{\theta})$ as:

$$U(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}), \tag{4.13}$$

which means that the potential energy favors motion towards the negative log-likelihood (NLL) of the target distribution.

In the HD formulation, the state of the system consists of the original variables of interest $\boldsymbol{\theta}$, plus extended momentum variables $\boldsymbol{u}$, forming a 2d-dimensional space $(\boldsymbol{\theta}, \boldsymbol{u})$. Considering the addictive nature of the Hamiltonian $H(\boldsymbol{\theta}, \boldsymbol{u}) = U(\boldsymbol{\theta}) + K(\boldsymbol{u})$, the joint canonical distribution becomes:

$$p(\boldsymbol{\theta}, \boldsymbol{u}) \propto \exp\left(-H(\boldsymbol{\theta}, \boldsymbol{u})\right) \tag{4.14}$$

$$\propto p(\boldsymbol{\theta}) \cdot \exp\left(-K(\boldsymbol{u})\right). \tag{4.15}$$

One degree of freedom in the HMC method is represented by the choice of the canonical momentum distribution used to generate the momentum, or equivalently, the choice of kinetic energy. One possible choice, comes from the request of having a canonical momentum distribution with a mean of zero, to preserve the dynamic and not make the computation of the acceptance probability unnecessarily complicated. Describing the kinetic energy in the quadratic form as in classical physics, we have:

$$K(\boldsymbol{u}) = \frac{1}{2}\mathbf{u}^T M^{-1}\mathbf{u}, \tag{4.16}$$

where M is a positive definite mass matrix. The canonical momentum distribution will be a multivariate Gaussian distribution with mean zero and covariance matrix M. Therefore, Equation 4.14 becomes:

$$p(\boldsymbol{\theta}, \boldsymbol{u}) \propto p(\boldsymbol{\theta})\mathcal{N}(\mathbf{u}). \tag{4.17}$$

Finally, if we marginalize over the variable $\mathbf{u}$, the resulting marginal distribution will sample the target distribution $p(\boldsymbol{\theta})$ we are looking for.

The HMC algorithm can be broken down into two main steps. The first step changes the momentum, which is randomly drawn from a Gaussian distribution while keeping the position unchanged. The second step changes both the position and the momentum, using the Metropolis-Hastings algorithm. Starting from the current state, the Hamiltonian dynamics are simulated for L steps of size ϵ using a reversible and volume-preserving integrator (for example, the Leapfrog method). At the end of the simulation, the momentum variables are negated, resulting in a proposed new state $(\boldsymbol{\theta}^*, \boldsymbol{u}^*)$. This proposed state is then evaluated to determine whether it is a good candidate to become the next state of the system. Specifically, it will be accepted with probability:

$$a = min\left(1, \exp\left[-H(\boldsymbol{\theta}^*, \boldsymbol{u}^*) + H(\boldsymbol{\theta}, \boldsymbol{u})\right]\right). \tag{4.18}$$

If the proposed state is rejected, the next state will be equal to the current one. The acceptance step is necessary because the numerical integrators cannot conserve the Hamiltonian exactly. Consequently, the probability a may differ from 1, necessitating an acceptance step. However, symplectic integrators, such as the leapfrog method, ensure that the numerical error remains bounded, thus keeping the rejection rate small.

4.2 Likelihood free inference

Although standard cosmological analyses rely on the Gaussian likelihood assumption, it has been shown that Non-Gaussianities exist within underlying Gaussian fields, such as the lensing field, and neglecting them can introduce biases in the parameter constraints (Sellentin and Heavens, 2017). To address the challenges that occur when the likelihood is unknown or intractable, alternative ways have been developed to perform inference. Sometimes referred to as Likelihood Free Inference (LFI) or Simulation-Based Inference (SBI) or Implicit Likelihood Inference (ILI) methods, those are characterized by the absence of an analytical model to describe the observed signal and instead rely on learning a likelihood from simulations. The first step in the LFI pipeline is to have a robust simulator to generate the data from which to learn the likelihood or the posterior distribution. Within this context, following **?**, a simulator can be defined as a computer program that, taken as input the parameter $\boldsymbol{\theta}$, samples some *latent variables* $\boldsymbol{z} \sim p(\boldsymbol{z}|\boldsymbol{\theta})$, and generates the output $\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{z})$. Such programs, characterized by random sampling and interpreted as statistical models, are usually referred to as *probabilistic programs*. Let us try to depict this formulation in the cosmological application. $\boldsymbol{\theta}$ are the cosmological parameters describing the mechanistic process, e.g. a cosmological simulation. The latent variables $\boldsymbol{z}$ arise during the data generation and can correspond to the physical state of the system in a direct or indirect way. Finally, the output $\boldsymbol{x}$ represents the observations. Both the parameters $\boldsymbol{\theta}$ and the latent variables $\boldsymbol{z}$ can be inferred.

However, if we want to only infer the cosmological parameters $\boldsymbol{\theta}$, the likelihood functions $p(\boldsymbol{x}|\boldsymbol{\theta})$ is typically intractable since it corresponds to an integral over all the possible values of the latent variables:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int d\boldsymbol{z} p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}), \tag{4.19}$$

and for real simulations where the latent space is very large, it is impossible to solve this integral explicitly.

4.2.1 Brief mention of traditional LFI methods

Approximate Bayesian Computation (ABC) One of the most well-known traditional algorithms used to bypass the evaluation of the likelihood is the *Approximate Bayesian computation* (ABC) (Pritchard et al., 1999; Marin et al., 2012; Ishida et al., 2015; Akeret et al., 2015; Lintusaari et al., 2017; Jennings and Madigan, 2017; Prangle, 2017; Leclercq, 2018). ABC-based methods consist of approximating the likelihood by comparing the outcome of simulations with observed data using *rejection criteria*, which define the level of discrepancy between them using a distance measure (based on a given metric). In the most basic form of their application, a set of parameters $\boldsymbol{\theta}$ is drawn from the prior distribution $p(\boldsymbol{\theta})$. Assuming a statistical model determined by

$\boldsymbol{\theta}$, the dataset $\boldsymbol{d}_{sim}$ is simulated. Finally, the mock data $\boldsymbol{d}_{sim}$ are compared with real data. If their distances are higher than a given threshold, the sampled parameters are discarded. Otherwise, the parameters are accepted with a tolerance ϵ. In essence, the probability that $\|\boldsymbol{d}_{obs} - \boldsymbol{d}_{sim}\| \leq \epsilon$, is an approximation of the likelihood:

$$p(\boldsymbol{d}_{sim} = \boldsymbol{d}_{obs}|\boldsymbol{\theta}) \approx p(\|\boldsymbol{d}_{obs} - \boldsymbol{d}_{sim}\| \leq \epsilon|\boldsymbol{\theta}). \tag{4.20}$$

Then, the outcome of the ABC algorithm is a set of sampled parameters, the distributions of which are approximately in accordance with the desired posterior.
Although this class of algorithms is mathematically well-founded, they are based on assumptions and approximations whose impact has to be evaluated (Sunnåker et al., 2013). Specifically, the inferred distribution becomes exact as $\epsilon \to 0$. However, very small values of ϵ require an excessively high number of simulations. On the contrary, if ϵ is too large, it degrades the quality of the inference. Moreover, since the data are involved in the rejection process, if new observations need to be evaluated, the selection of simulations satisfying Equation 4.20 must be repeated.

DELFI The Density Estimation Likelihood Free Inference (DELFI) represents another class of likelihood-free inference methods. Within these methods, the inference task is approached as a density estimation problem using a set of mock parameter-data pairs $\{\boldsymbol{\theta}, \boldsymbol{x}\}$ or $\{\boldsymbol{\theta}, \boldsymbol{t}\}$, where $\boldsymbol{t}$ represents some compressed data summaries.

Quoting the definition proposed by Papamakarios et al. (2019), a *conditional neural density estimator* can be described as a parametric model, denoted as $q_{\boldsymbol{\varphi}}$ (e.g., a neural network), defined by a set of parameters $\boldsymbol{\varphi}$. This model outputs a conditional probability density $q_{\boldsymbol{\varphi}}(\boldsymbol{u}|\boldsymbol{v})$ for a given pair of points $\{\boldsymbol{u}, \boldsymbol{v}\}$. Suppose we have a set of independent training data $\{\boldsymbol{u}_n, \boldsymbol{v}_n\}_{1:N}$ characterized by the joint probability density $p(\boldsymbol{u}, \boldsymbol{v})$. In that case, $q_{\boldsymbol{\varphi}}$ can be trained to maximize the log-probability $\sum_n \log q_{\boldsymbol{\varphi}}(\boldsymbol{u}_n|\boldsymbol{v}_n)$ w.r.t $\boldsymbol{\varphi}$. With a sufficient amount of data and a suitably flexible model, $q_{\boldsymbol{\varphi}}(\boldsymbol{u}|\boldsymbol{v})$, can learn to approximate $p(\boldsymbol{u}|\boldsymbol{v})$.

We can classify the algorithms into two primary classes, which aim to directly estimate either the likelihood or the posterior probability.

The *Synthetic Likelihood*, typically implemented as the inner part of an MCMC sampler, aims to estimate the intractable likelihood. The first versions of these methods were mainly based on some Gaussian approximations, while more recent variants use neural networks to approximate the density $p(\boldsymbol{x}|\boldsymbol{\theta})$. For example, Papamakarios et al. (2019) presented the *Sequential Neural Likelihood* (SNL) algorithm, characterized by a sequential training procedure able to guide the simulation and reduce the number of necessary mock data. The method uses an autoregressive flow to learn the likelihood function in the region of high posterior density.

Other methods, such as the one on which is based the analysis presented in the next chapter, rely on using the estimator $q_{\boldsymbol{\varphi}}(\boldsymbol{\theta}|\boldsymbol{x})$ to directly model the posterior function. First, a set of samples $\{\boldsymbol{\theta}_n, \boldsymbol{x}_n\}_{1:N}$ is obtained from the joint distribution $p(\boldsymbol{\theta}, \boldsymbol{x})$ by drawing $\boldsymbol{\theta}_n$ from $p(\boldsymbol{\theta})$ and $\boldsymbol{x}_n$ from $p(\boldsymbol{x}|\boldsymbol{\theta})$. Then, this set of samples is used to train

$q_{\boldsymbol{\varphi}}$ to obtain an approximation of the density posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$. Finally, $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ is computed from $q_{\boldsymbol{\varphi}}(\boldsymbol{\theta}|\boldsymbol{x}_0)$. Unfortunately, this strategy requires a large number of simulations. For this reason, several enhancements have been proposed to reduce the number of synthetic data required to train the conditional neural density estimator. Within this class methods, we can find for example the *Sequential Neural Posterior Estimation* (Papamakarios and Murray, 2016; Lueckmann et al., 2017a). Very schematically, instead of using the prior $p(\boldsymbol{\theta})$, the parameters $\boldsymbol{\theta}_n$ are sampled from a proposal distribution $\tilde{p}(\boldsymbol{\theta})$, which makes the data $\boldsymbol{x}_n$ more likely closer to the observed point $\boldsymbol{x}_0$. The algorithms find a good proposal adopting a sequential procedure, i.e. the training of $q_{\boldsymbol{\varphi}}$ is made in several rounds, in each of which the proposal is the approximate posterior obtained in the previous round.
One of the disadvantages of the SNPE is that the proposal introduces a bias in the approximation procedure. Indeed the neural density estimator approximate $p(\boldsymbol{x}_0|\boldsymbol{\theta})\tilde{p}(\boldsymbol{\theta})$ instead of $p(\boldsymbol{x}_0|\boldsymbol{\theta})p(\boldsymbol{\theta})$, hence some adjustments have to be made.
The LFI algorithms employing neural posterior estimators differ in the way they deal with this issue. For example, Papamakarios and Murray (2016) account for this effect multiplying the learned posterior $q_{\boldsymbol{\varphi}}(\boldsymbol{\theta}|\boldsymbol{x}_0)$ by $\tilde{p}(\boldsymbol{\theta})$ and dividing it by $p(\boldsymbol{\theta})$. This method is usually referred to as SNPE-A in literature. Lueckmann et al. (2017b), on the other hand, assigned a weight $w_n = p(\boldsymbol{\theta})/\tilde{p}(\boldsymbol{\theta})$ to adjust the parameters sample $\boldsymbol{\theta}_n$, and trained the NDE using the weighted log-likelihood $\sum_n w_n \log q_{\boldsymbol{\varphi}}(\boldsymbol{\theta}_n|\boldsymbol{x}_n)$. This method is usually referred to as SNPE-B in literature. Unfortunately, both methods have issues. SNPE-A restricts the proposal to be Gaussian, and in cases where the variance of the proposal is smaller than the variance of any of the components of $q_{\boldsymbol{\varphi}}(\boldsymbol{x}_0|\boldsymbol{\theta})$, the division returns a Gaussian with negative variance and the algorithm is interrupted. SNPE-B can have high variance weights and lead to instabilities during training.
An alternative solution has been proposed by Greenberg et al. (2019a) with the *Automatic Posterior Transformation* (APT), which combines the advantages of the posterior estimators, such as targeting directly the posterior, with the ones from likelihood estimation, such as flexible proposal and absence of weights or post-hoc corrections. In this approach, the network is trained to learn the map from the data to the true posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$ by maximizing the probability of the parameters of the simulation under the proposal posterior $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{x})$. Specifically, a *proposal* loss $\tilde{\mathcal{L}}(\boldsymbol{\varphi}) = -\sum_n \log \tilde{q}_{x,\boldsymbol{\varphi}}(\boldsymbol{\theta}_n)$ is minimized, with:

$$\tilde{q}_{x,\boldsymbol{\varphi}}(\boldsymbol{\theta}) = q_{F(\boldsymbol{x},\boldsymbol{\varphi})}(\boldsymbol{\theta})\frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\frac{1}{Z(\boldsymbol{x},\boldsymbol{\varphi})}, \tag{4.21}$$

where $Z(\boldsymbol{x},\boldsymbol{\varphi}) = \int_{\boldsymbol{\theta}} q_{F(\boldsymbol{x},\boldsymbol{\varphi})}(\boldsymbol{\theta})\frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}$ is a normalization constant and $\boldsymbol{\varphi}$ are the weights of a neural network F. APT can use arbitrary and dynamically updated proposals and is compatible with a wide range of density estimators, including mixture-density networks and normalizing flows.

4.3 The need for data compression

I conclude this chapter by introducing a concept that will play a crucial role in chapter 7: the need for data compression.
In weak lensing analyses, we have seen that the data are usually compressed into summary statistics, such as the 2-point correlation functions or the higher-order statistics. This typically represents the first compression step to reduce large cosmological data sets and to make the subsequent inference task tractable. However, there are situations where the space of summaries is still too large for inference, and further compressed statistics are required. In certain circumstances, it may be necessary to reduce the original N data points to a subset of N' points before conducting classical analysis based on likelihood.

Moreover, we have seen that, for some of these summary statistics, the likelihood function is unknown, necessitating assumptions or approximations that are not always valid. In the previous sections, we have discussed the LFI methods, that offer the advantage of facilitating the Bayesian analysis using only forward simulations, without requiring any assumptions about the likelihood. However, these methods are often subject to the curse of dimensionality, making it necessary to utilize compression methods to reduce the dimensionality of the data space and extract summary statistics.
This has motivated the introduction of several dimensionality reduction techniques aimed at reducing the representation of the data while preserving as much information from the original dataset as possible. Within these methods, we can make a distinction between *linear* compression techniques and *neural network-based* compression techniques.
The following section introduces some useful theorems needed to understand the next chapter. Additionally, it provides a brief overview of two widely used dimension reduction methods proposed in the literature for weak lensing applications. While in chapter 7, our focus will be exclusively on *neural* compression methods, for the sake of completeness on this topic, I will outline two other important compression strategies here: the *Moped* algorithm and the *Generalized massive optimal data compression* algorithm.

The Fisher information matrix

The accuracy with which we can constrain model parameters from a dataset was discussed for the first time in Fisher (1935). We can summarize the results as follows. We denote the true vector of the model parameters as $\boldsymbol{\theta_0}$ and the estimated vector from the data as $\boldsymbol{\theta}$. Since $\mathbf{x}$ is a random variable and $\boldsymbol{\theta}$ is a function of $\mathbf{x}$, it is also a random variable. For $\boldsymbol{\theta}$ to be a good estimate of $\boldsymbol{\theta_0}$, it must satisfy the following criteria:

$$\langle \boldsymbol{\theta} \rangle = \boldsymbol{\theta_0}, \tag{4.22}$$

and it should represent the Best Unbiased Estimator (BUE), meaning that it minimizes the standard deviations:

$$\Delta\theta_\alpha \equiv (\langle\theta_\alpha^2\rangle - \langle\theta_\alpha\rangle^2)^{1/2}. \tag{4.23}$$

We can now define the *maximum likelihood estimator* as the vector of model parameters $\boldsymbol{\theta}_{ML}$ maximizing the likelihood function $p(\mathbf{x}|\boldsymbol{\theta})$, and remembering the definition of the Fisher matrix:

$$\mathbf{F}_{\alpha\beta} \equiv \left\langle \frac{\partial^2\mathcal{L}}{\partial\theta_\alpha\partial\theta_\beta} \right\rangle,^{1} \tag{4.24}$$

we can report the useful set of theorems[2]:

- Being $\Delta\theta_\alpha$ any unbiased estimator, $\Delta\theta_\alpha \geq 1/\sqrt{\mathbf{F}_{\alpha\alpha}}$. In other words, there is a lower limit on the error bars we can get from our estimates, no matter what methods we use. If other parameters are also computed from the data, the expression becomes $\Delta\theta_\alpha \geq 1/(\mathbf{F}^{-1})^{1/2}_{\alpha\alpha}$. This result is known as the Cramér-Rao inequality.

- If a BUE $\boldsymbol{\theta}$ exists, then it is the Maximum Likelihood (ML) estimator.

- In the limit of a very large data set the ML-estimator is asymptotically the BUE.

As highlighted by Heavens et al. (2000b), if the full data set is used, the Fisher matrix is the best that can possibly be done via likelihood methods with the data, i.e considering the theorems described above, it is the minimum error in the constraints of the cosmological parameters.
In practice, some of the data could be poorly informative due to noise or their low sensitivity to the parameters. Consequently, we can potentially discard them keeping the amount of information partially unchanged.

MOPED Heavens et al. (2000b) presented a method for linear compression that is valid when the noise in the data is independent of the model parameters. They showed that it is possible to create M linear combinations of the data containing the same amount of information as the full data set. They demonstrated that the method is lossless by proving that the final Fisher information matrices are identical. The method involves creating several linear combinations of the data and discarding the less informative ones. The data vector $\boldsymbol{x}$ is represented as the sum of a signal $\boldsymbol{\mu}$ and a zero-mean noise $\boldsymbol{n}$. First, a single linear combination of the data is considered:

$$y \equiv \mathbf{b}^t\mathbf{x}, \tag{4.25}$$

where $\mathbf{b}$ is a weight vector and $\mathbf{b}^t$ its transpose; The goal is to find the weight vector $\mathbf{b}$ that encodes as much information about a particular parameter θ_α.

[1] with the usual notation $\mathcal{L} \equiv$ -ln$L$

[2] for a more detailed description see (Kenney and Keeping (1951); Kendall and Stuart (1969))

The compression of the dataset can potentially increase the error bars of the parameters we want to estimate. As discussed in the previous section, the error of our estimates is related to the Fisher matrix, which, in the case where a single parameter is estimated by the data, can be expressed as (Tegmark et al., 1997):

$$\boldsymbol{F}_{\alpha,\beta} = \frac{1}{2}\left(\frac{\boldsymbol{b}^t \boldsymbol{C}_{,\alpha}\boldsymbol{b}}{\boldsymbol{b}^t \boldsymbol{C}\boldsymbol{b}}\right)\left(\frac{\boldsymbol{b}^t \boldsymbol{C}_{,\beta}\boldsymbol{b}}{\boldsymbol{b}^t \boldsymbol{C}\boldsymbol{b}}\right) + \frac{(\boldsymbol{b}^t \boldsymbol{\mu}_{,\alpha})(\boldsymbol{b}^t \boldsymbol{\mu}_{,\beta})}{\boldsymbol{b}^t \boldsymbol{C}\boldsymbol{b}}. \tag{4.26}$$

In the most general case, both $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{C}$ are dependent on the parameters. Unfortunately, the resulting maximization leads to a non-linear eigenvalue problem in $\mathbf{b}$. To address this, Heavens et al. (2000b) focused on the scenario where $\mathbf{C}$ is known.
Specifically, they demonstrated that for estimating M parameters, there exist M orthogonal vectors $\mathbf{b}_m$ (with $m = 1, ..., M$) such that the summaries y_m encode as much information about parameter θ_m as possible, excluding what is already contained in y_q with $q < m$. The orthogonal vector $\mathbf{b}_m$ takes the following form:

$$\mathbf{b}_m = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}_{,m} - \sum_{q=1}^{m-1}(\boldsymbol{\mu}_{,m}^t \mathbf{b}_q)\mathbf{b}_q}{\sqrt{\boldsymbol{\mu}_{,m}\mathbf{C}^{-1}\boldsymbol{\mu}_{,m} - \sum_{q=1}^{m-1}(\boldsymbol{\mu}_{,m}^t \mathbf{b}_q)^2}}. \tag{4.27}$$

Moreover, they demonstrated the potential of the method in the generalized case where the noise is dependent on parameters. Specifically, they considered an application example in which parameters characterizing the galaxy star formation history needed to be constrained from a set of mock galaxy spectra. Their findings revealed that the compression was still almost lossless, resulting in an increase in the error on the parameters by a factor of $\sim \sqrt{1 + 1/(2s)}$, with s representing the mean number of photons in each channel of the spectrum.

Generalized massive optimal data compression Alsing and Wandelt (2018) provided a generalization of the results of Tegmark et al. (1997) and Heavens et al. (2000b), describing a more general procedure for optimal compression suitable for generic functions in a framework that is not restricted to linear statistics and Gaussian data. They also extended the work of Heavens et al. (2000b) to derive explicit compressed statistics for cases with Gaussian likelihoods where the mean and covariance are parameter-dependent. The compression proceeds as follows: by Taylor expanding the log-likelihood to the second order around some fiducial point $\boldsymbol{\theta}_*$, we obtain the expression:

$$\mathcal{L} = \mathcal{L}_* + \delta\boldsymbol{\theta}^T \nabla\mathcal{L}_* + \frac{1}{2}\delta\boldsymbol{\theta}^T(\nabla\nabla^T\mathcal{L})\delta\boldsymbol{\theta}. \tag{4.28}$$

In this expression, the derivative of the log-likelihood is commonly known as the *score function*. To the linear order in the model parameters, these parameters only couple to the data through the score function, which is a vector of length n, and a function of the data. This means that computing only the n combinations of the data appearing

in the score function, this latter provides an immediate data compression from N data down to n compressed numbers. Hence, we can write:

$$\mathbf{t} = \nabla \mathcal{L}_*. \tag{4.29}$$

In particular, they compute the explicit compressed statistic in the case where a Gaussian likelihood is assumed and both the mean and covariance depending on the parameters:

$$\mathbf{t} = \nabla \boldsymbol{\mu}_*^T \mathbf{C}_*^{-1}(\mathbf{x} - \boldsymbol{\mu}_*) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_*)^T \mathbf{C}_*^{-1} \nabla \mathbf{C}_* \mathbf{C}_*^{-1}(\mathbf{x} - \boldsymbol{\mu}_*). \tag{4.30}$$

As we have already mentioned there are several situations where the likelihood is unknown or non-Gaussian. However, in this specific application, the assumed approximate Gaussian likelihood function is used only for the purpose of performing the data compression. Once the compressed summaries are obtained, the subsequent inference is performed without this assumption, i.e. defining a non-Gaussian model for the likelihood-based inferences or with a likelihood-free procedure. In the worst-case scenario, the compression will be sub-optimal in the sense that the approximate likelihood function for the compression will be lossy, but the approximations made in the compression step will not bias the subsequent inference (Alsing and Wandelt, 2018; Alsing et al., 2018). Nevertheless, it is worth noting that the score function of a given likelihood is considered to be optimal because it retains as much Fisher information as possible from the data. If the likelihood is a highly non-Gaussian or a multimodal function of the parameters, the Fisher information may not accurately reflect the information in the data and other compression schemes may be more useful.

Chapter 5

Hybrid Physical-Neural ODEs for Fast N-body Simulations

The results presented in this chapter were presented at the 2022 Workshop on Machine Learning for Astrophysics at the International Conference on Machine Learning (ICML). My contributions included method development, implementation and testing, leading the analysis, and writing the paper. This work was conducted in collaboration with François Lanusse and Jean-Luc Starck.

In chapter 2, we discussed the computational challenges associated with solving the full N-body problem, which requires significant computational resources and time. To address these challenges, Quasi-N-body numerical schemes such as FastPM (Feng et al., 2016) and COLA (Tassev et al., 2013) have been proposed. These schemes aim to accelerate the computational time and generate cost-effective realizations of large-scale structures. However, the Particle-Mesh (PM) scheme used in these simulations has a limitation. It approximates the interactions between particles at short distances, making it unable to resolve structures with scales smaller than the mesh resolution. As a result, simulations using PM schemes often lack power on small scales, and the sharpness of halo profiles is reduced compared to full N-body simulations.

In this chapter, we propose a method to improve the accuracy of PM simulations.

We augment the physical differential equations of a PM N-body with a minimally-parametric neural network component that models a *residual effective force*. This component compensates for the approximations made in the PM scheme.

It is important to remember that the ultimate goal of cosmological simulation is to replicate the observed data, in order to gain a better understanding of the underlying physical processes. True data is generated from physical processes, which preserve certain symmetries and involve specific physical constraints. When generating mock data, it is crucial to incorporate these considerations. For instance, since there is no preferred position or direction in the Universe, we strive to preserve rotational and translational symmetries. Moreover, we aim to satisfy conservation laws, such as the conservation of dark and baryonic matter.

In our approach, we achieve rotational and translational invariance by implementing a neural correction in the form of a learned isotropic Fourier filter. This filter exclusively depends on the amplitude of the wave vector. Meanwhile, mass conservation is guaranteed as we only modify the forces acting on the particles.

To train the neural network through the simulator, we follow the scheme proposed by Chen et al. (2018), i.e. we treat the Ordinary Differential Equations (ODEs) solver running the simulation as a black box and compute gradients using the adjoint sensitivity method (Pontriagin, 1964). We train and compare the model to the CAMELS simulations (Villaescusa-Navarro et al., 2021) and test it against several changes in the simulation setting (resolution, volume, cosmological parameters). We also compare our method to the Potential Gradient Descent scheme (PGD, Dai et al. (2018)).

5.1 Hybrid Physical-Neural ODE

Cosmological simulations implement the integration of gravitational evolution starting from the Gaussian initial conditions of the Universe to the observed large-scale structures. We choose to compute the time integration starting from a system of Ordinary Differential Equations and treat the ODE solver as a black box:

$$\begin{cases} \frac{d\mathbf{x}}{da} & = \frac{1}{a^3 E(a)}\mathbf{v} \\ \frac{d\mathbf{v}}{da} & = \frac{1}{a^2 E(a)}\boldsymbol{F}_\theta(\mathbf{x}, a) \\ \boldsymbol{F}(\mathbf{x}, a) & = \frac{3\Omega_m}{2}\nabla\left[\phi_{PM}(\mathbf{x})\right]. \end{cases} \tag{5.1}$$

Here $\mathbf{x}$ and $\mathbf{v}$ define the position and the velocity of the particle, a is the cosmological scale factor, $E(a)$ the ratio between the Hubble expansion rate and the Hubble parameter, $\mathbf{F}$ the gravitational force experienced by the dark matter particles in the mesh, and ϕ_{PM} is the gravitational potential estimated by using the cloud-in-cell (CiC) interpolation scheme (Hockney and Eastwood, 2021). We can use this parametrization to complement the physical ODE with a neural network:

$$\boldsymbol{F}_\theta(\mathbf{x}, a) = \frac{3\Omega_m}{2}\nabla\left[\phi_{PM}(\mathbf{x}) * \mathcal{F}^{-1}(1 + f_\theta(a, |\mathbf{k}|))\right], \tag{5.2}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform and $f_\theta(a, |\mathbf{k}|)$ is the learned neural filter, aimed to model the residual effective force compensating for the PM approximations. We choose to define $f_\theta(a, |\mathbf{k}|)$ as a B-spline function, whose coefficients (knot points) are the output of a neural network of parameters θ.

B-spline functions are piecewise-defined polynomials that can be easily adapted to construct smooth curves and surfaces. If employed in neural networks, they can provide a flexible framework for modeling functions. Indeed, the control points of the splines can be adjusted to modify the shape and the behavior of the function, enabling a wide range of representations. Specifically, we choose a B-spline of order 3, defined over 16 knot points.

5.2 Simulation data

In this work, we rely on the CAMELS dataset (Villaescusa-Navarro et al., 2021), in particular, we use the suite of dark-matter-only N-body simulations implementing the N-body tree-particle-mesh approach of IllustrisTNG (Nelson et al., 2019). We have 34 snapshots, generated following the evolution of 256^3 dark matter particles in a periodic box of comoving volume equal to 25^3 $(h^{-1}$ Mpc$)^3$, with initial conditions generated at z=127 using second-order Lagrangian perturbation theory (2LPT). The following cosmological parameters are kept fixed in all simulations: $h = 0.6711$, $n_s = 0.9624$, $M_\nu = 0.0$ eV, $w = -1$, $\Omega_k = 0.$, while the values of Ω_m and σ_8 are varied across simulations. We further downsample the CAMELS particle data, down to 64^3 particles, to make the problem more manageable on a single GPU.

5.3 Training

5.3.1 Backpropagation of ODE solutions

To train our hybrid physical-neural ODE model and back-propagate through the ODE solver, we compute the gradient adopting the adjoint sensitivity method (Chen et al., 2018; Pontryagin et al., 1962), consisting in solving a second ODE backwards in time, and treat the ODE solver as a black box.

Consider an ODE solver that depends on the parameters $\mathbf{z}(t_0), f, t_0, t_1, \theta$, where $\boldsymbol{z}(t)$ is the state variable, f the function modeling the dynamics of the system, t_0 the start time, t_1 the stop time and θ the dynamic parameter. The function $\mathcal{L}$ which takes the results of this ODE solver as input

$$\mathcal{L}(\mathbf{z}(t_1)) = \mathcal{L}(\text{ODESolve}(\mathbf{z}(t_0), f, t_0, t_1, \theta)) \tag{5.3}$$

can be differentiated with respect to the parameter θ. First, we need to compute the *adjoint*:

$$\mathbf{a}(t) = \frac{\partial \mathcal{L}}{\partial \mathbf{z}(t)}, \tag{5.4}$$

i.e. the gradient of $\mathcal{L}$ respect to the hidden state $\mathbf{z}(t)$. Then, we can determine the dynamics of the adjoint through:

$$\frac{d\mathbf{a}(t)}{dt} = -\mathbf{a}(t)^\intercal \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}}. \tag{5.5}$$

Finally, we compute the gradients with respect to the parameter θ evaluating a third integral:

$$\frac{d\mathcal{L}}{d\theta} = \int_{t_1}^{t_0} \mathbf{a}(t)^\intercal \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \theta} dt. \tag{5.6}$$

In a framework like Jax (Bradbury et al., 2018) or TensorFlow (Abadi et al., 2015), where automatic differentiation is guaranteed, the products between the vector and the Jacobian $\mathbf{a}(t)^T \frac{\partial f}{\partial \mathbf{z}}$, and $\mathbf{a}(t)^T \frac{\partial f}{\partial \theta}$ can be computed at the exact computational cost of evaluating f. Moreover, in a single call to the ODE solver, the original state, the adjoint, and the derivatives are concatenated into a single vector, and the integrals for $\boldsymbol{z}$, $\boldsymbol{a}$, and $\frac{\partial \mathcal{L}}{\partial \theta}$ are solved.

Backpropagation through the ODE solver has the significant advantage of not requiring the storage of all intermediate steps from the forward simulations. This allows for lower memory costs during both the forward and backward passes of the simulation. Additionally, by defining a tolerance for the ODE solver, we can adjust the trade-off between accuracy and speed based on our specific requirements.

5.3.2 Train and validation loss

HPN To train the neural network in the Hybrid Physical-Neural (HPN) scheme, we adopt the following loss function, which penalizes both the particle positions and the overall matter power spectrum at different snapshot times s, compared to a reference full N-body simulation:

$$\mathcal{L} = \sum_s ||\mathbf{x}_s^{Nbody} - \mathbf{x}_s||_2^2 + \lambda||\frac{P_s(k)}{P_s^{Nbody}(k)} - 1||_2^2 \ . \tag{5.7}$$

In this equation, $P_s(k)$ denotes the power spectrum of the PM-simulation, and $P_s^{Nbody}(k)$ represents the power spectrum of the corresponding full N-body simulation. Additionally, $\boldsymbol{x}_s$ and $\boldsymbol{x}_s^{Nbody}$ are the position vectors of the particles, and λ is left as a hyper-parameter balancing the contributions of the two terms.

In contrast to He et al. (2019) who train the model minimizing the mean square error on particle displacements, we have empirically found that penalizing the overall matter power spectrum yields better improvement for the small-scale power spectrum, compared to penalizing the particle positions only. One possible reason for this different result may be the reduced number of degrees of freedom in our model compared to the one used by He et al. (2019). In our training process, the neural network is constrained by physical limitations and our simple parameterization, and therefore cannot approximate arbitrary non-linear functions.

However, it is worth noting that different displacement fields can produce identical density fields and, consequently, identical power spectra. If we were to penalize only the power spectrum term, we might increase the overall power spectrum on small scales without however accurately recovering the correct particle positions. For this reason, our final loss function includes both terms from Equation 5.7. We have empirically found that $\lambda = 0.1$ provides the optimal balance in terms of overall correction and overfit. The model is trained by gradient descent with the Adam optimizer (Kingma and Ba, 2014) and learning rates of 0.01. The code is implemented in the Python package jax (Bradbury et al., 2018) using the jax-based Dormand-Prince ODE solver.

PGD Following the same strategy adopted by Dai et al. (2018), we fit the PGD parameters only for the final snapshot (a=1) by penalizing the following loss function:

$$\mathcal{L} = \left\| \sigma(k) \left(\frac{P(k)}{P^{Nbody}(k)} - 1 \right) \right\|_2^2 . \tag{5.8}$$

Here $\sigma(k)$ is a weight added to the loss function to downweight small scales in comparison to the large scales. It is defined as:

$$\sigma(k) = \frac{1 - \frac{k}{\pi(\frac{\mathrm{M}_s}{\mathrm{V}_b})}}{\sum_k \left(1 - \frac{k}{\pi(\frac{\mathrm{M}_s}{\mathrm{V}_b})} \right)}, \tag{5.9}$$

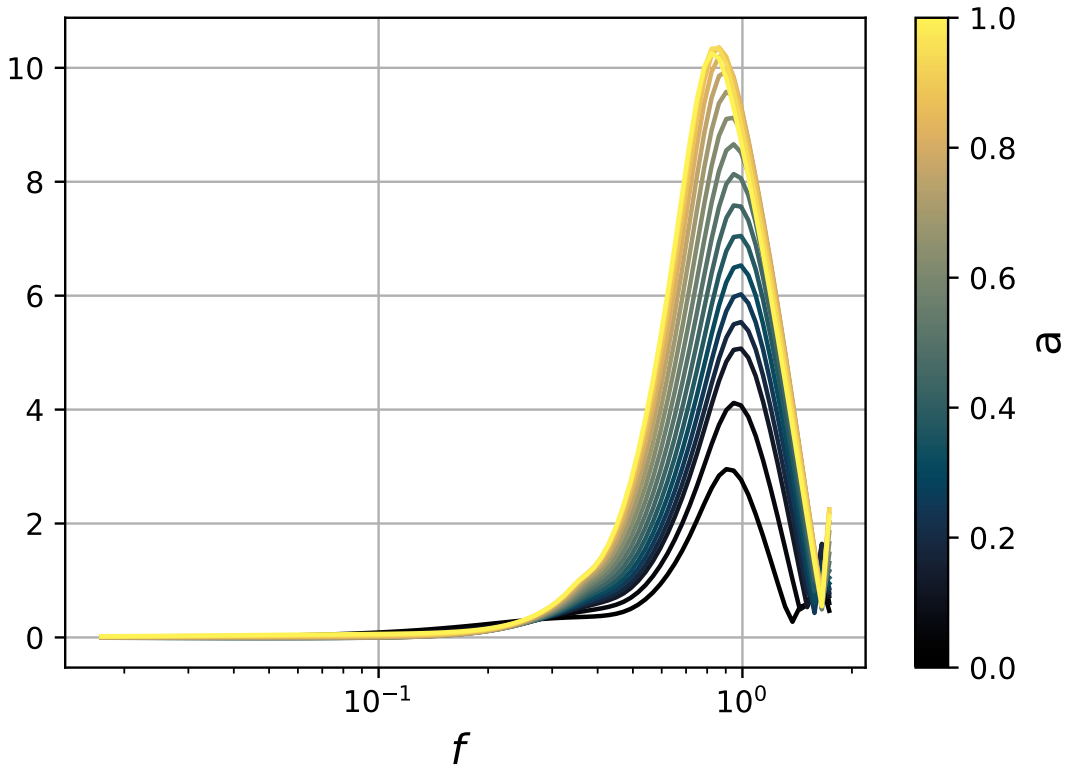

Figure 5.1: Learned neural filter $f_\theta(a, |\mathbf{k}|)$ as a function of normalized mesh frequency ($f = 1$ corresponds to the Nyquist frequency of the mesh), for different scale factor a.

Where M_s refers to the mesh size of the PM simulations, and V_b refers to the comoving volume of the periodic box in the PM simulations. This weight choice leads to a significant improvement in the correction provided by PGD for the small-scale power spectrum, as opposed to only considering the power spectrum. However, we have observed that this weight choice does not result in significant enhancements or variations in the performance of PGD when applied to simulations with different cosmologies or resolutions.

5.4 Results

In this section, we train and evaluate our hybrid neural ODE on CAMELS data and compare it to results obtained using the PGD approach by Dai et al. (2018). We train the neural ODE using the loss function described in Equation 5.7, while the PGD parameters are fitted only for the final snapshot ($a = 1$) by penalizing the weighted power spectrum ratio between the corrected and reference simulations (i.e. Equation 5.8).

For both approaches, we use a *single* CAMELS N-body simulation with the fiducial cosmology values of $\Omega_m = 0.3$ and $\sigma_8 = 0.8$ to fit the parameters of the models.

Figure 5.2 shows the matter overdensity fields obtained using the two different

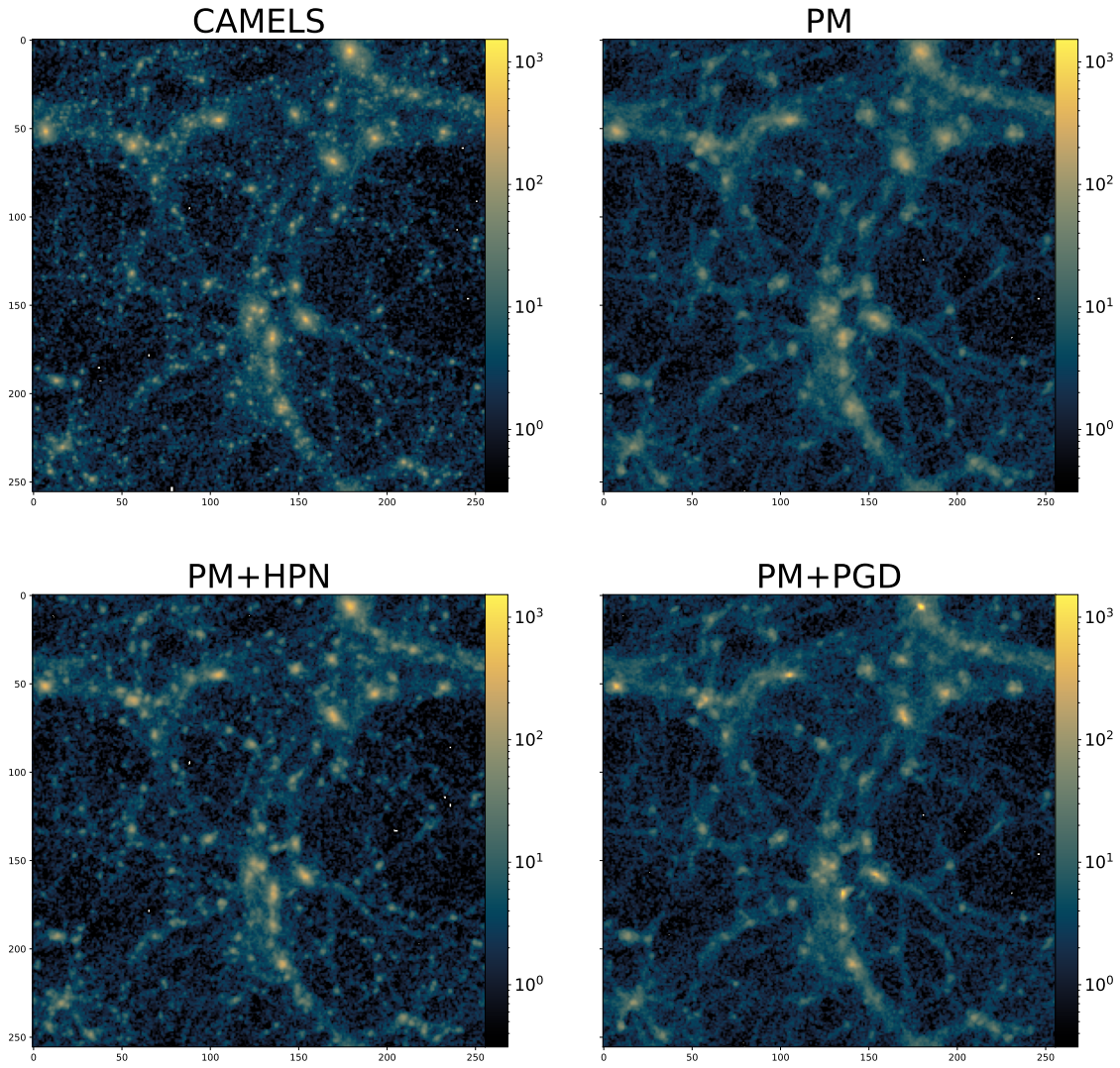

Figure 5.2: Top panel from left to right: Projections of final density field for CAMELS and PM only. Bottom panel from left to right: PM corrected by the HPN scheme and PM corrected by PGD. Although the PGD model improves the sharpness of the halos, it fails to fully recreate the dark matter structures, producing a smoother density profile compared to the HPN model, which appears closer to CAMELS.

correction schemes in comparison to the CAMELS simulation and the standalone PM (i.e. pure PM simulation). As we will quantify later, it is evident that the pure PM simulation (top panel on the right) appears smoother than the reference CAMELS simulation (top panel on the left). On the other hand, the neural correction (second panel on the left) successfully recovers most of the missing information. In comparison, the PGD correction (second panel on the right) does sharpen somewhat some structures in the field in order to increase the overall power spectrum on small scales, but is less effective in restoring the actual shape of the dark matter structures.

Learned Neural Filter We present the trained Fourier-space filter $f_\theta(a, |\mathbf{k}|)$ as a function of normalized mesh frequency in Figure 5.1. As expected, the correction only affects small scales, having no effect on the large-scale modes of the simulations, where the PM solver provides accurate results. An interesting feature to note is that the correction is scale-factor dependent. This suggests that the model does not simply apply a sharpening filter akin to a CiC compensation filter, but rather adapts the correction based on the density field properties at different cosmological times. However, one drawback of this observation is that if the neural correction relies on the specific dynamics of the dark matter density field, it may be sensitive to the cosmology employed.

Power spectrum comparison at fiducial cosmology In Figure 5.3, we compare the power spectrum ratio and the cross-correlation between CAMELS and PM simulations to illustrate the differences between the two correction schemes. As can be seen, both schemes successfully recover most of the missing power on small scales. However, the improvement in cross-correlation coefficients is significantly larger when utilizing our scheme.

Robustness to changes in resolution and cosmological parameters To evaluate the robustness of our correction scheme, we compute the following tests:

1. We investigate the effects of **varying the cosmological parameters**. We perform PM simulations at $\Omega_m = 0.10$ (instead of $\Omega_m = 0.3$ used during the training) and compare them to a set of CAMELS simulations generated with the same cosmology and initial conditions. The top panel of Figure 5.4 displays the results for the matter power spectrum in this test. Comparing the HPN and PGD correction schemes, we observe only a slight deviation in the HPN correction compared to the results obtained at the fiducial cosmology (shown in Figure 5.3). In contrast, PGD appears to be highly sensitive to Ω_m and no longer provides the desired correction. In the bottom panel of Figure 5.4, we present the results for the cross-correlation coefficients between CAMELS and the PM field. We note that while the HPN correction leads to an improvement in the cross-correlation coefficients, the improvement is less evident with the PGD correction.

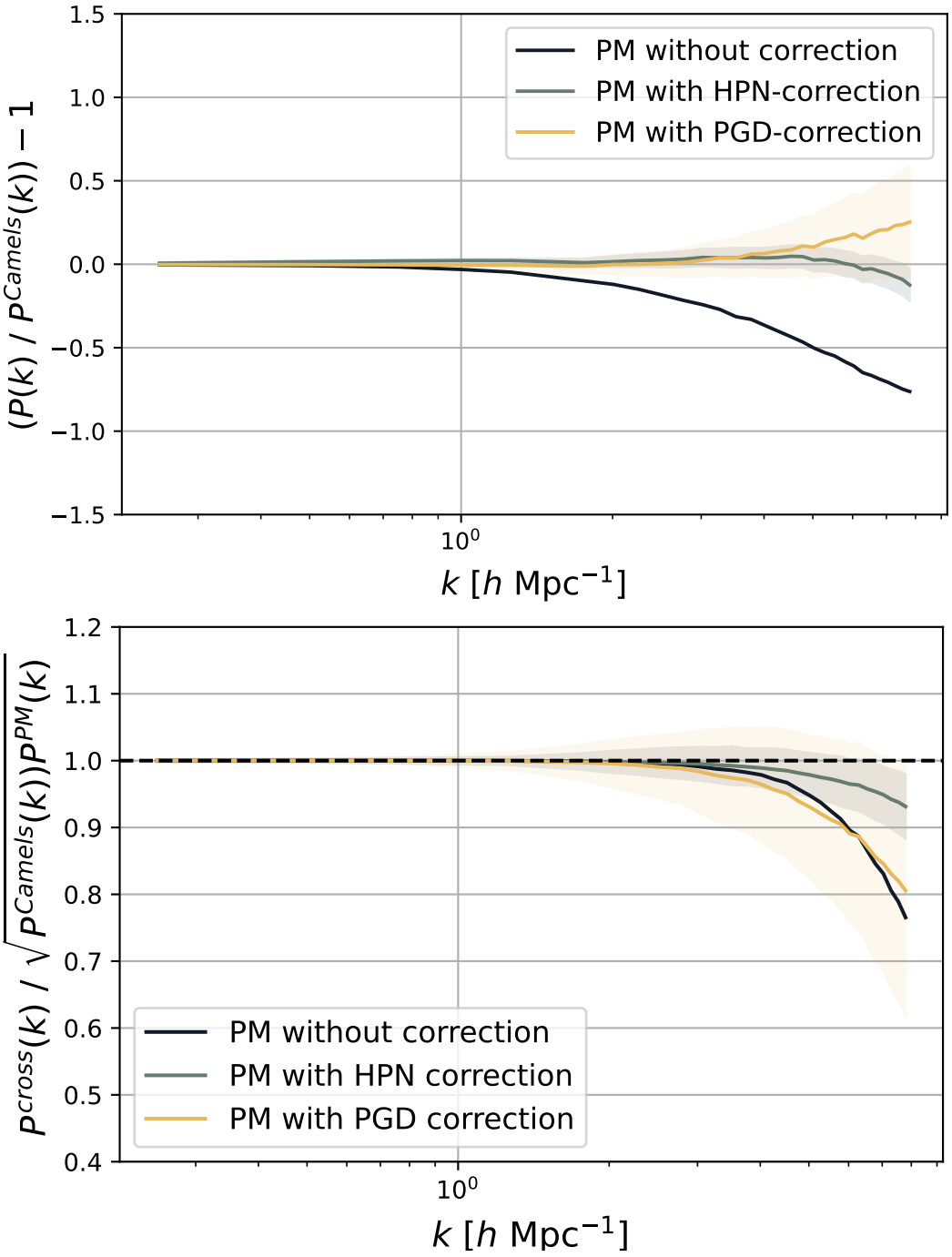

Figure 5.3: Top panel: Fractional matter power spectrum of PM simulations before and after using the correction models and CAMELS simulations. Bottom panel: cross-correlation coefficients with the reference simulation before and after the two different correction schemes. The shaded regions represent the standard deviation from the same realization corrected with 10 different parameters fitted from different initial conditions.

2. We investigate the effects of **increasing the simulation resolution**. We generate a PM simulation with 128^3 particles for the same 25 h^{-1} Mpc box size. The top panel of Figure 5.5 shows the results for the matter power spectrum. We observe that the HPN correction outperforms the PGD method up to $k \sim 7$, after which the method tends to overemphasize the small-scale power. The bottom panel of Figure 5.5 presents the results for the cross-correlation coefficients between the CAMELS and the PM field when the simulation resolution is increased.
   We highlight that the improvement achieved with our correction scheme is superior to the PGD method, although a significant portion of the correlation between the two fields is not recovered at small scales.

3. Finally, we evaluate the two schemes at a **lower resolution** by increasing both the volume of the simulation to 200 $(h^{-1}\text{Mpc})^3$ and the number of particles to 128^3. Since the CAMELS simulations are generated in a box of 25 $(h^{-1}\text{Mpc})^3$, we use the halofit power spectrum $P(k)/P(k)^{halofit}$ as the reference in this test. From Figure 5.6, we observe that even in this extreme case, the performance of the HPN correction remains very good, while the PGD method no longer provides the desired correction.

5.5 Conclusion and discussion

In section 2.3, we discussed various machine learning techniques available for emulating high-cost N-body simulations, including the approaches proposed by He et al. (2019) and Li et al. (2021). These methods use particle displacements as inputs and outputs for their models, relying on large Deep Convolutional Networks to learn an effective mapping that generates the desired outputs. In contrast, we have presented a correction scheme for quasi-N-body PM solvers, based on a Neural Network implemented as a Fourier-space filter. This approach combines the physical description of structure formation with a simpler neural network to correct the approximations made in pseudo-N-body solvers. Our method shares similarities with the work of Dai et al. (2018), who demonstrated how to enhance the modeling of matter distribution within halos in quasi-N-body simulations using a gradient-based scheme. One advantage of these methods compared to full ML techniques is the lower dimensionality of the parameter space that needs to be optimized. To illustrate the merits of our approach, we compared the results of corrected PM simulations with the high-resolution CAMELS simulations. We benchmarked our model against the PGD scheme, showing that, at the fiducial cosmology, both methods yield similar improvements in the small-scale power spectrum but significantly differ in terms of the correlation coefficients. We also presented the results of our experiments involving changes in resolution and cosmological parameters. These experiments demonstrated that our method outperforms the PGD scheme in terms of both the power spectrum and cross-correlation coefficients,

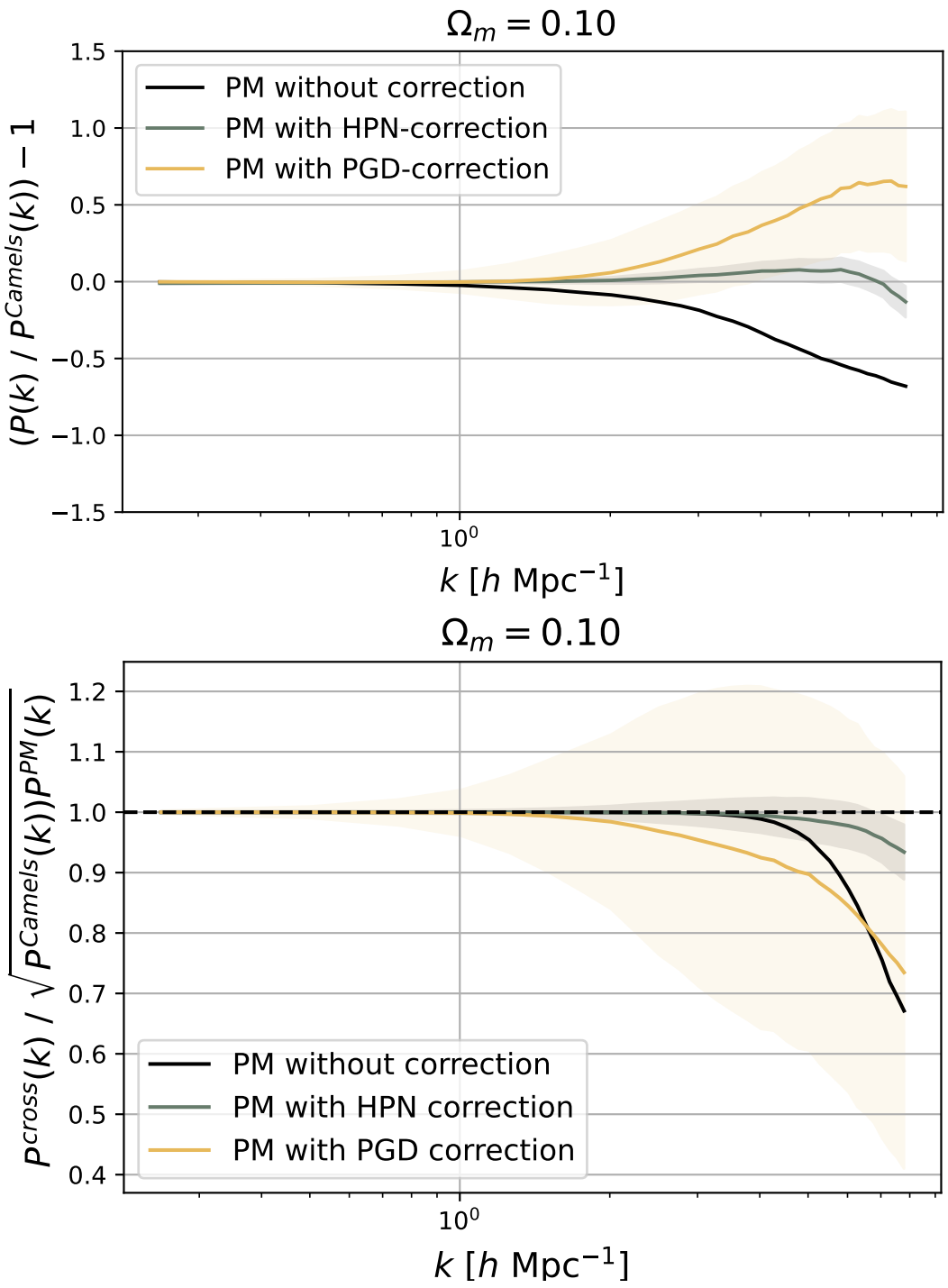

Figure 5.4: Benchmarks on the matter power spectrum and cross-correlation coefficients varying the cosmological parameters Ω_m employed in the simulation. Top panel: Fractional matter power spectrum of PM simulations before and after using the correction models and CAMELS simulations. Bottom panel: cross-correlation coefficients with the reference simulation before and after the two different correction schemes. The shaded regions represent the standard deviation from the same realization corrected with 10 different parameters fitted from different initial conditions.

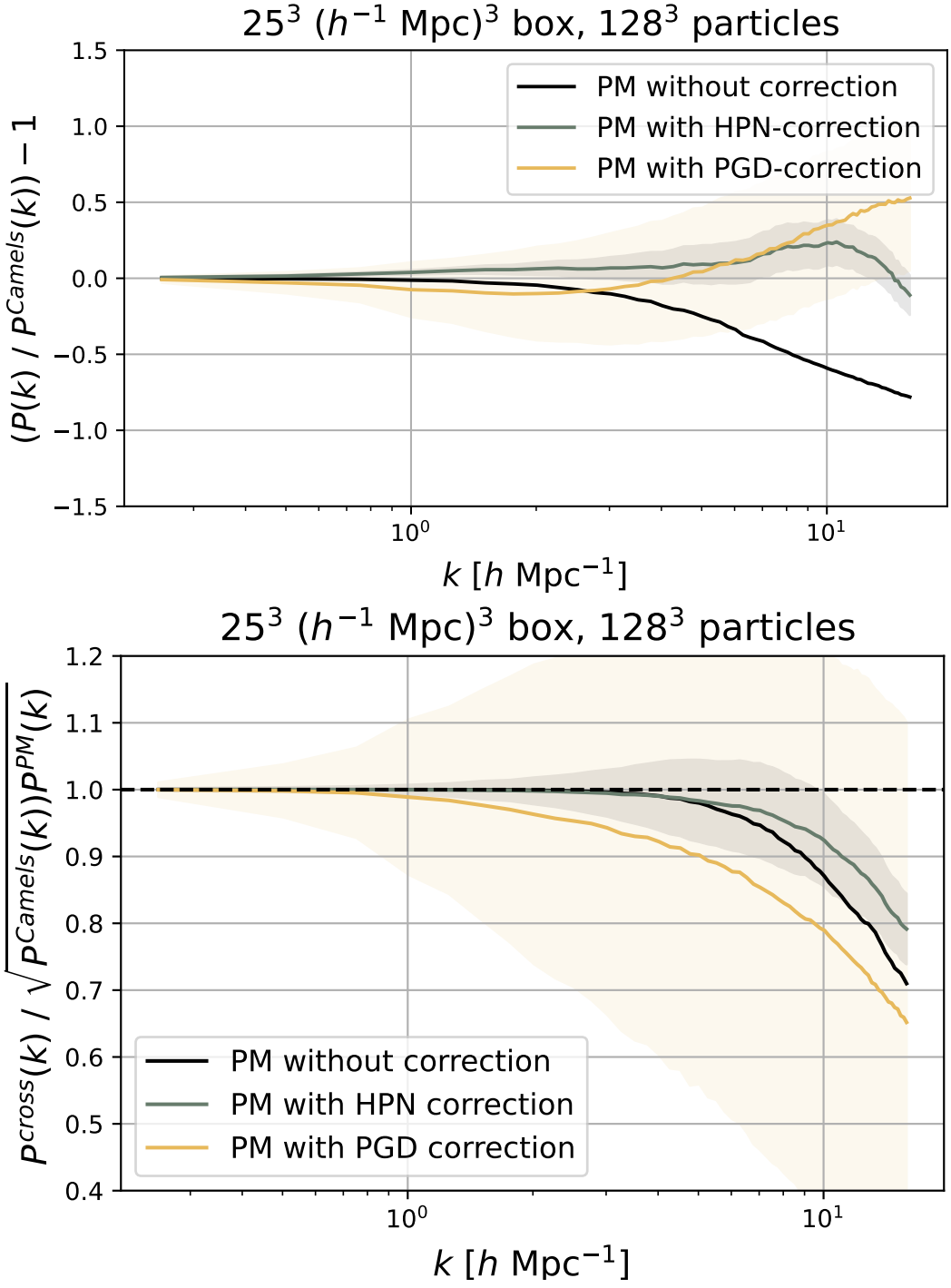

Figure 5.5: Benchmarks on the matter power spectrum and cross-correlation coefficients when the number of particles employed in the simulation is increased (to 128^3). Top panel: Fractional matter power spectrum of PM simulations before and after using the correction models and CAMELS simulations. Bottom panel: cross-correlation coefficients with the reference simulation before and after the two different correction schemes. The shaded regions represent the standard deviation from the same realization corrected with 10 different parameters fitted from different initial conditions.

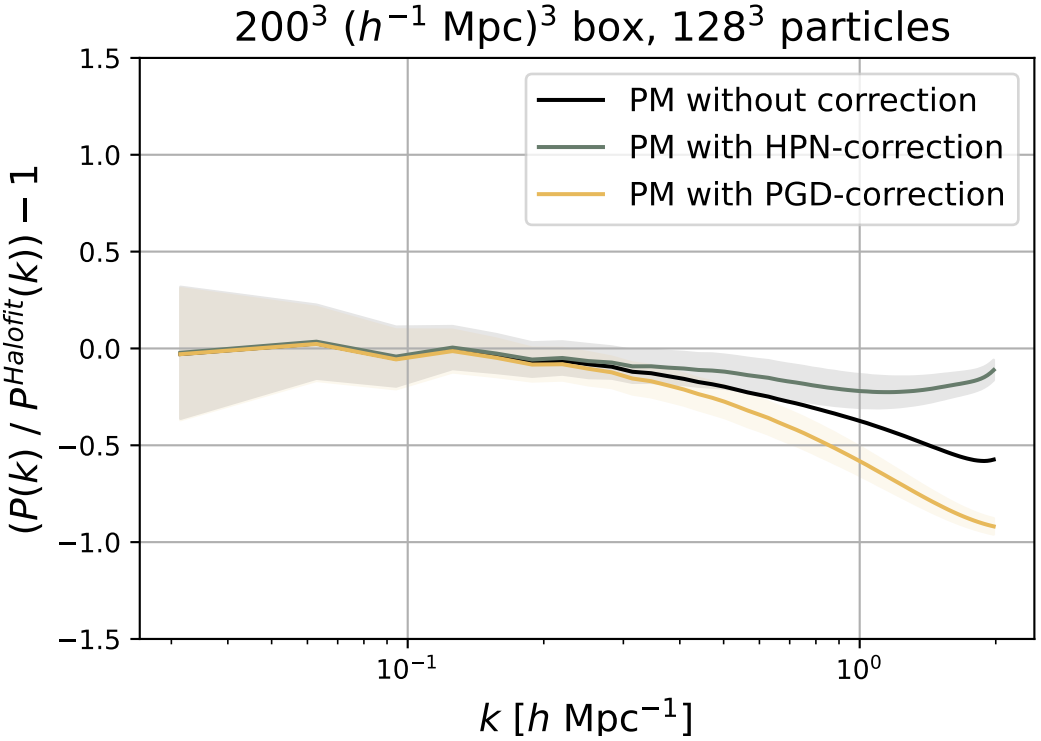

Figure 5.6: Benchmarks on the matter power spectrum, when the number of particles (128^3) and box size ($205\ h^{-3}\text{Mpc}^3$) employed in the simulation are increased. Note that in this last comparison, the theoretical HaloFit matter power spectrum is considered as a reference. The shaded regions represent the standard deviation from 1000 realizations with different initial conditions corrected with the same parameters.

and it is less sensitive to the settings of the simulations used for training. However, we observed that our scheme slightly overemphasizes the small-scale power for $k \sim 7$ and does not substantially improve the results for the cross-correlation coefficients when tested on simulations with higher resolutions.

Chapter 6

Forecasting the power of Higher Order Weak Lensing Statistics with automatically differentiable simulations

The results presented in this chapter have been published in Astronomy & Astrophysics in Lanzieri et al. (2023). My contributions consisted of the development, implementation, and testing of the simulation code, the generation and analysis of mock data, and the writing of the paper. The research project was done in collaboration with Francois Lanusse, Chirag Modi, Benjamin Horowitz, Joachim Harnois-Déraps, Jean-Luc Starck, and The LSST Dark Energy Science Collaboration (LSST DESC).

We present the Differentiable Lensing Lightcone (DLL), a fully differentiable physical model designed for being used as a forward model in Bayesian inference algorithms requiring access to derivatives of lensing observables with respect to cosmological parameters. We extend the public FlowPM N-body code, a particle-mesh N-body solver, simulating lensing lightcones and implementing the Born approximation in the Tensorflow framework. We validate our simulations in an LSST setting against high-resolution κTNG-Dark simulations by comparing both the lensing angular power spectrum and multiscale peak counts. As a first use case, we use this tool to investigate the relative constraining power of the angular power spectrum and peak counts statistic in an LSST setting. Such comparisons are typically very costly as they require a large number of simulations, and do not scale well with the increasing number of cosmological parameters. As opposed to forecasts based on finite differences, these statistics can be analytically differentiated with respect to cosmology, or any systematics included in the simulations at the same computational cost as the forward simulation.

This chapter is structured as follows: in section 6.1 we introduce the numerical simulations illustrating the numerical methods used to generate mock WL maps. In section 6.2 we validate the simulations by comparing the statistics from our simulations and κTNG-Dark ones. The Fisher forecast formalism and the survey and noise setting are shown in section 6.3. We finally discuss our results and present our conclusions afterward, in section 6.4 and section 6.5.

6.1 Fast and Differentiable Lensing Simulations

Analytical models with which to predict the observed signals do not exist for most higher-order summary statistics. To circumvent this issue, one approach is to rely on generating a suite of numerical simulations. In the following two sections, we introduce our procedure for simulating weak lensing maps, which includes a description of the N-body simulator and the construction of lightcones. As mentioned in the introduction, our goal is to efficiently compute gradients of the simulations, as this will benefit the development of new inference algorithms for weak lensing surveys. To achieve this, we extend the framework of the FlowPM package Modi et al. (2021) by implementing differentiability of the cosmological functions with respect to the cosmological parameters[1], which were not included in the original FlowPM code. Additionally, we implement the Born approximation and simulate lensing lightcones in the Tensorflow framework. TensorFlow is a tensor library that includes the ability to perform automatic differentiation. Automatic differentiation enables us to compute gradients exactly, as opposed to finite differences, which only provide approximate gradients. Specifically, TensorFlow implements the backpropagation algorithm to compute gradients, i.e. first it creates a graph (e.g. data structures representing units of computation), then it works backward through the graph by applying the chain rule at each node.

6.1.1 Differentiable Particle-Mesh N-body simulations

We adjust the FlowPM approach (Modi et al., 2021) by performing the time integration through a system of Ordinary Differential Equations, as depicted in Equation 5.1. As introduced in chapter 5, these ODEs describe the gravitational evolution of the particles in the simulation using a black-box ODE integrator. This is in contrast to the leapfrog integration method used in FlowPM. One reason for this change is that adaptive ODE solvers can automatically adjust the time step of the simulation based on the desired accuracy for the result. Another reason is that modern automatic differentiation frameworks like TensorFlow provide automatically differentiable solvers, which significantly reduce the memory footprint of the simulation when computing the gradients, as described below. Indeed, the decision to extend the FlowPM code with the ODE implementation is motivated by the fact that to compute the gradient of the forward model, the original algorithm needs to store all the intermediate steps of the simulations. This induces a memory overhead that scales with the number of time steps in the simulation. In the adjoint ODE approach, this is instead replaced by solving another ODE backward in time when evaluating the gradient. We illustrate the potential of differentiating through ODE solvers, highlighting the fact that the simulations and the gradients presented in this chapter are computing using one single GPU for 128^3 particles. Additionally, we have discussed that although quasi-N-body

[1]For example, implementing the cosmological functions introduced in subsection 1.3.1, subsection 1.3.2, section 1.4

Table 6.1: Approximate execution times of different parts of the simulations. The values represent the mean and the standard deviation of 7 runs, 1 loop each.

Model	Time
LPT	744 ms ±10.7 ms
N-Body (128^3)	16.5 s ±7.89 ms
N-Body ODE (128^3)	13 s ±120 ms
N-Body ODE+HPN (128^3)	1 min 24 s ± 573 ms
N-Body (256^3)	1 min 14 s ± 14.1 ms
N-Body ODE (256^3)	1 min 39 s ± 181 ms
Raytracing (Born Approximation)	320 ms ± 16.9 ms

schemes help bypass the computational effort in terms of time and CPU/GPU power required in fully N-body simulations, they lack resolution on small scales and are unable to resolve structures with scales smaller than the mesh resolution.

Therefore, we integrate the Hybrid Physical-Neural approach presented in chapter 5 into FlowPM code to address this limitation. The computational cost of the various steps of the simulations is listed in Table 6.1.
It should be noted that when computing the gradient, some instability may arise from the ODE solver due to numerical errors in discretizing the continuous dynamics. This can result in the adaptive solvers taking too many time steps and slowing down the gradient evaluation. However, this can be mitigated simply by limiting the number of time steps in adaptive ODE solvers. Another way to control these numerical errors is by using Leapfrog ODE solvers where one can exploit the reversibility of the Leapfrog dynamics in evaluating the gradients with adjoint methods, as done in Li et al. (2022).

6.1.2 Differentiable Lensing Simulations

To extract the lens planes and construct the lightcone, we export 11 intermediate states from the N-body simulation of a fixed interval of 205 h^{-1} Mpc in a redshift range between $z = 0.03 - 0.91$. To recover the redshift range of the lightcone, one unit box is replicated using periodic boundary conditions. First, we generate rotation matrices along the three axes, hence, each snapshot is rotated around each of the three axes, finally, all the particles are randomly shifted along the axes. To obtain the final density field, each snapshot is projected in a 2D plane by estimating its density with a cloud-in-cell (CiC) interpolation scheme (Hockney and Eastwood, 1988). After creating a Cartesian grid of coordinates, each slice is interpolated onto sky coordinates. This procedure differs from the one implemented in the MADLens package (Böhm et al., 2021). In MADLens the lightcone is built by translating the redshift of the particles into distances, then the particles are projected onto the convergence map at the proper evolution step corresponding to that distance.

Implementation of Born lensing

We generate the convergence map by integrating the lensing density along the unperturbed line of sight, i.e. applying the Born approximation (Schneider, 2006). In particular we discretize the Equation 3.23 such that it becomes:

$$\kappa_{born} = \frac{3H_0^2\Omega_m}{2c^2} \sum_i \bar{\delta}_i \left(1 - \frac{\chi_i}{\chi_s}\right) \left(\frac{\chi_i}{a_i}\right) \Delta\chi \tag{6.1}$$

where the i index runs over the different lens planes, the $\bar{\delta}_i$ indicates the matter overdensity projected into the lightcone, χ_s defines the comoving distance of the source and $\Delta\chi$ is the width of the lens plane.

Implementation of IA with NLA

We model the effect of IA on the convergence map level following the model proposed by Fluri et al. (2019). This allows us to create pure IA convergence maps to combine with shear convergence maps in order to generate a contaminated signal. Following Harnois-Déraps et al. (2021), the Fourier transform of the intrinsic ellipticities can be phrased as:

$$\begin{aligned} \tilde{\epsilon}_1^{IA}(\mathbf{k}_\perp) &\propto \left(\frac{k_x^2 - k_y^2}{k^2}\right) \tilde{\delta}_{2D}(\mathbf{k}_\perp)\mathcal{G}_{2D}(\sigma_g) \\ \tilde{\epsilon}_2^{IA}(\mathbf{k}_\perp) &\propto \left(\frac{k_x k_y}{k^2}\right) \tilde{\delta}_{2D}(\mathbf{k}_\perp)\mathcal{G}_{2D}(\sigma_g), \end{aligned} \tag{6.2}$$

where σ_g defines the smoothing scale of a two-dimensional smoothing kernel $\mathcal{G}_{2D}$, the tilde symbols $\sim$ refers to the Fourier transformed quantities, $\mathbf{k}_\perp$ denotes the two Fourier wave-vector components perpendicular to the line of sight. Combining Equation 3.34 and Equation 6.2 we can calculate the intrinsic alignment as part of the convergence map:

$$\kappa_{IA_i} = -A_{IA}\bar{C}_1\rho_c\Omega_m \int_{z_{min}}^{z_{max}} n_i(z)\delta_{s_i}\frac{dz}{D(z)} \tag{6.3}$$

where the index i refers to the i-th redshift bins.

Differentiable Wavelet peak counts

We conduct a multi-scale peak counts analysis by applying a starlet wavelet filter to the original noisy simulated convergence maps. In this section, we introduce the fundamentals of wavelet theory, our definition of a *peak*, and our approach to ensure the peak counts statistics are differentiable.

Wavelet Transform The wavelet transform has been widely used in analysing astronomical images due to its ability to decompose astronomical data into components at different scales. This multiscale approach is well-suited for the study of astronomical data since their complex hierarchical structure. A wavelet function $\psi(x)$ is a function that satisfies the admissibility condition:

$$\int_{\mathbb{R}^+} |\hat{\psi}(k)|^2 \frac{dk}{k} < \infty, \tag{6.4}$$

where we indicate with $\hat{\psi}(k)$ the Fourier transform of $\psi(x)$, with $\int \psi(x)dx = 0$ in order to satisfy the admissibility condition. A given signal is decomposed in a family of scaled and translated functions:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right), \tag{6.5}$$

where $\psi_{a,b}$ are the so-called *daughter wavelets*, scaled and translated version of the *mother* wavelet, with a and b scaling and translation parameters. The continuous wavelets transform is defined from the projections of a function $f \in L_2(\mathbb{R})$ onto the family of daughter wavelets. The coefficients of this projection represent the wavelet coefficient, obtained by :

$$W_f(a,b) = \int_{\mathbb{R}} f(x)\psi^*_{a,b}(x)dx = \frac{1}{\sqrt{a}}\int_{\mathbb{R}} f(x)\psi^*\left(\frac{x-b}{a}\right)dx, \tag{6.6}$$

with ψ^* the complex conjugate of ψ, and $\forall a \in \mathbb{R}^+$, $b \in \mathbb{R}$. In this work, we filter the original convergence maps with the starlet transform, an isotropic and undecimated (i.e. not down-sampled) wavelet transform, suited for astronomical applications where objects are mostly more or less isotropic (Starck et al., 2007).

It decomposes an image c_0 as the sum of all the wavelet scales and the coarse resolution image c_J:

$$c_0(x,y) = c_J(x,y) + \sum_{j=1}^{J_{max}} w_j(x,y) \tag{6.7}$$

where J_{max} is the maximum number of scales and w_j is the wavelet images showing the details of the original image at dyadic scales with a spatial size of 2^j pixels and $j = J_{max} + 1$.

The starlet wavelet function is a specific translational invariant wavelet transform:

$$\frac{1}{4}\Psi\left(\frac{x_1}{2}, \frac{x_2}{2}\right) = \phi(x_1, x_2) - \frac{1}{4}\phi\left(\frac{x_1}{2}, \frac{x_2}{2}\right) \tag{6.8}$$

specified by an isotropic scaling function ϕ, that, for astronomical application, can be defined as a B-spline of order 3:

$$\phi_{1D}(x) = \frac{1}{12}(|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3). \tag{6.9}$$

The N-dimensional scaling functions can be built starting from the separable product of N ϕ_{1D}: $\phi(x_1, x_2) = \phi_{1D}(x_1)\phi_{1D}(x_2)$. Each set of wavelet coefficients w_j is obtained as the convolution of the input map with the corresponding wavelet kernel. For a full description of the starlet transform function, see Starck et al. (2007) and Starck et al. (2010).

Differentiable Peak Counts The peaks identify regions of weak lensing map where the density value is higher, in this way they are particularly sensitive to massive structures. There are two different ways to record weak lensing peaks: as 1) local maxima of the signal-to-noise field or 2) local maxima of the convergence field. In both cases, they are defined as pixels of larger value than their eight neighbors in the image.

One of the difficulties in estimating derivatives of traditional peak count statistics is that it relies on building a histogram of peak intensities and, therefore, due to the discrete nature of the bins, histograms are not differentiable. However, the underlying idea of peak counting is just meant to build an estimate of the density distribution of the number of peaks as a function of their intensity. Histograms are one way to build such an estimate, and have been historically preferred, but for no particular reason. To circumvent the non-differentiability of histograms, here we prefer to estimate this density using an alternative method, namely, the Kernel Density Estimation (KDE). As a continuous equivalent to a histogram, KDEs are differentiable and can just as well be used to define the peak counts statistic. We defined the KDE for the peak counts as:

$$\mathrm{KDE} = \frac{1}{b_w\sqrt{2\pi}} \exp\left(-\frac{(X-x)^2}{2b_w^2}\right) \tag{6.10}$$

where b_w is the smoothing bandwidth parameter, X is the number of peaks in a given bin, and x is the center of each bin.

This procedure yields a peak count statistic that is smoothly differentiable with respect to the input map and thus can be used for applications such as Fisher forecasting as discussed later in this work.

6.2 Validating simulations for LSST

In this section, we compare the results from our simulation to other works, including the analytic models for the matter power spectrum *Halofit* (Smith et al., 2003; Takahashi et al., 2012) and the cosmological N-body simulations, κTNG (Osato et al., 2021).

The κTNG-Dark simulations is a suite of publicly available weak lensing mock maps based on the cosmological hydrodynamical simulations IllustrisTNG, generated with the moving mesh code AREPO (Springel, 2010). In particular, we use the κTNG-Dark suit of maps based on the corresponding dark matter-only TNG simulations. The simulations have a side length of the box equal to 205 Mpc/h and 2500^3 CDM particles. To model the propagation of light rays and simulate the weak lensing maps, a

multiple-lens plane approximation is employed. The simulation configuration consists of a map size of 5×5 deg^2, 1024×1024 pixels, and a resolution of 0.29 arcmin/pixel. For a complete description of the implementation see Osato et al. (2021).

To produce our simulations, we follow the evolution of 128^3 dark matter particles in a periodic box of comoving volume equal to 205^3 $(h^{-1}$ Mpc$)^3$, with initial conditions generated at z=6 using the linear matter power spectrum as implemented by Eisenstein and Hu (1998). In particular, we implement the Eisenstein-Hu transfer function in the Tensorflow framework, in order to compute its gradients automatically.
We assume the following cosmological parameters: $h = 0.6774$, $n_s = 0.9667$, $M_\nu = 0.0$ eV, $w = -1$, $\Omega_k = 0.$, $\Omega_m = 0.3075$, $\sigma_8 = 0.8159$, such that they match the results of Planck 2015 (Ade et al., 2016). We reproduce the same configuration of κTNG-Dark, i.e. each map is on a regular grid of 1024^2 pixels and 5×5 deg^2.
The actual choice of bins to include in the forecasting is made following the DESC data requirement for the angular power spectra (Mandelbaum et al., 2018), i.e. adopting $\ell_{max,shear} = 3000$ and $\ell_{min,shear} = 300$.

6.2.1 HPN validation

To compensate for the small-scale approximations resulting from PM, we applied the HPN approach presented in chapter 5. We show on Figure 6.1 the power spectrum and the fractional power spectrum of PM simulations before and after the HPN correction compared to analytic Halofit predictions (Smith et al., 2003; Takahashi et al., 2012) for redshift $z = 0.03$ and $z = 0.91$. We observe a bias between our measured power spectrum and the theoretical prediction at low k. This reduced power is explained by the small box size of our simulation and the associated reduced number of large-scale modes. At redshift $z = 0.91$ most of the missing power is recovered by the HPN correction up to $k \sim 1$, after which the method overemphasizes the small-scale power. In this work however, we can assume that this effect does not impact the results of the cosmological parameters forecast, since it concerns scales that are beyond the range of frequencies that are taken into account for the analysis. At redshift $z = 0.03$, the correction model does not improve significantly the results.

In Figure 6.2 we show an example of our convergence map at $z = 0.91$, from pure PM simulation (first panel) and the HPN corrected simulation (second panel). The HPN model sharpens structures in the lensing field without introducing any artifacts.

In the upper panel of Figure 6.3, we present the angular power spectrum computed from our Differentiable Lensing Lightcone (DLL hereafter) complemented by the HPN scheme and a conventional DLL simulation with the same resolution. Both the outputs are compared to the κTNG-Dark prediction. In the lower panel of Figure 6.3 the fractional differences between the convergence power spectra from the two maps and the κTNG-Dark are shown. Both the power spectra and ratios are averaged over $N = 100$ realisations. We can see that the HPN model reduces the relative deviations of the angular power spectra to within 30%. We also observe a perfect match at large

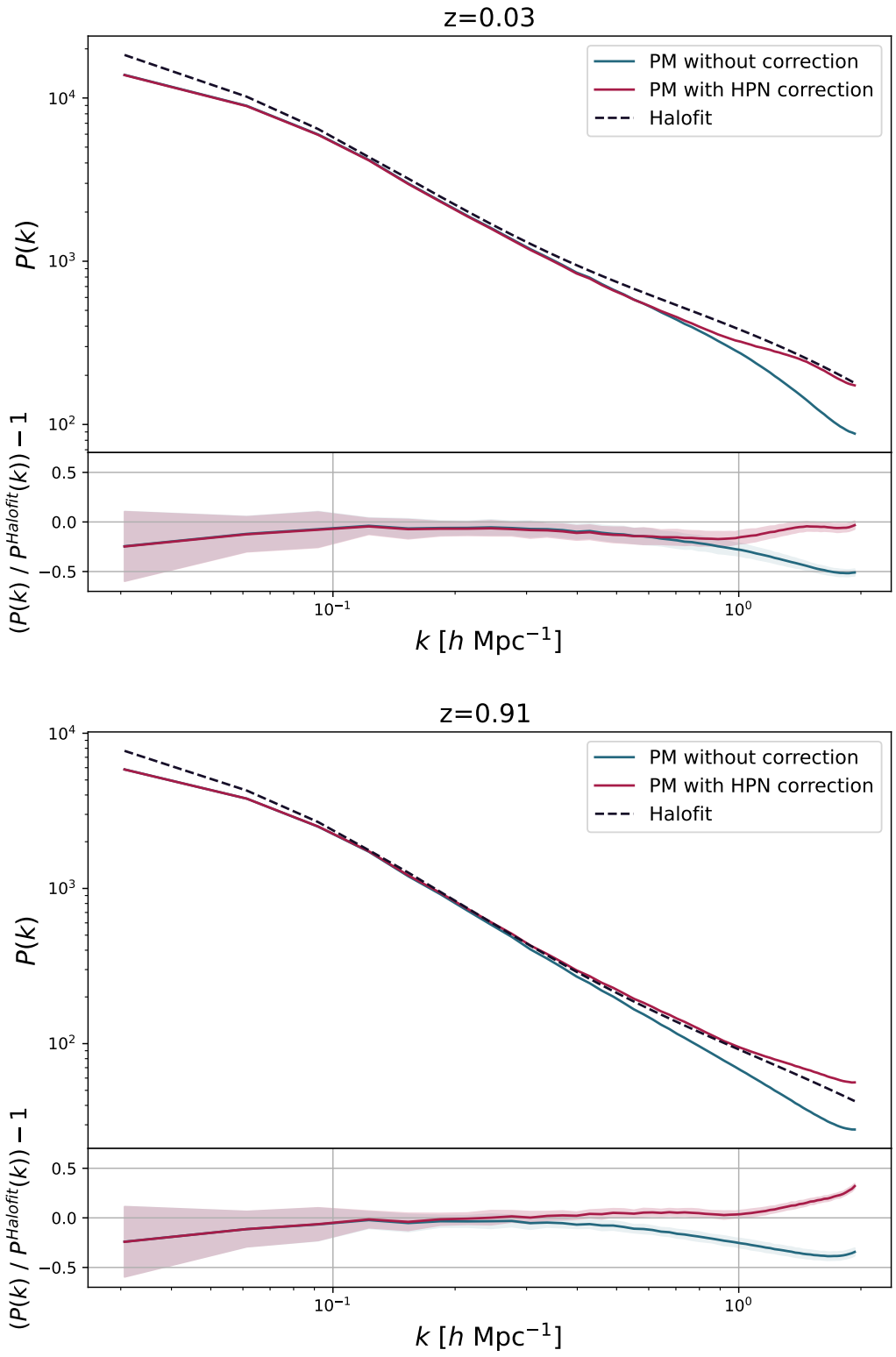

Figure 6.1: Matter power spectrum and fractional matter power spectrum of PM simulations before and after using the Hybrid Physical-Neural (HPN) correction model and the theoretical halofit model for redshift z=0.03 (upper panel) and redshift z=0.91 (lower panel). The power spectra and ratios are means over 100 independent map realizations. The shaded regions represent the standard deviation from 100 independent DLL realizations.

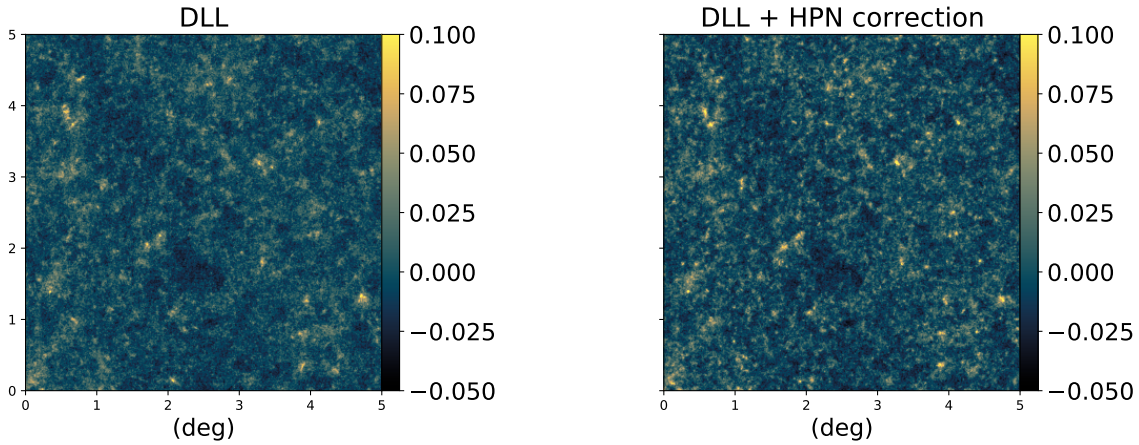

Figure 6.2: **Left panel**: Convergence map at source redshift $z = 0.91$ from DLL, PM only. **Right panel**: Same convergence map when the HPN correction is applied.

scales, since the κTNG-Dark and the DLL simulations have the same box size of 205 Mpc/h^3.

6.2.2 IA validation

In the upper panel of Figure 6.4, we present the C_ℓ^{II} and C_ℓ^{GG} contributes from our DLL simulations compared to theoretical Halofit predictions (Smith et al., 2003). In the lower panel of the same figure, we show the fractional differences between the mentioned contributions. To validate the IA infusion, only for this experiment, we run simulations keeping the term $A_{IA} = 1$. As we can see, the fractional difference for the C_ℓ^{II} term features uncertainty consistent with C_ℓ^{GG} term, validating our infusion process. The signal is computed for the source redshift $z_s = 0.91$ and is averaged over 100 realizations. The theoretical predictions are computed using the public Core Cosmological Library (CCL, Chisari et al. (2019)).

6.2.3 Lensing C_ℓ

To quantify the accuracy of the simulations we aim to reproduce the summary statistics from the Dark Matter Only κTNG simulations. We compare the results from the angular power spectrum for different source redshift, just investigating how well we can recover the power spectrum for a given source plane. The results of the angular power spectrum from the sources redshift z=[0.25,0.46,0.65,0.91,1.30] are shown in the upper panel of Figure 6.5, as well as the fractional differences between the κTNG-Dark and DLL maps in the lower panel. We observe that the differences for $z_s = 0.91$ and $z_s = 1.30$ curves are within 10% of accuracy for scales larger than $\ell = 1000$, within 25% for scales $1000 < \ell < 2000$ and within 25% and 45% for scales $2000 < \ell < 3000$. For lower source redshifts the deficit of power in our simulations becomes worse. This can

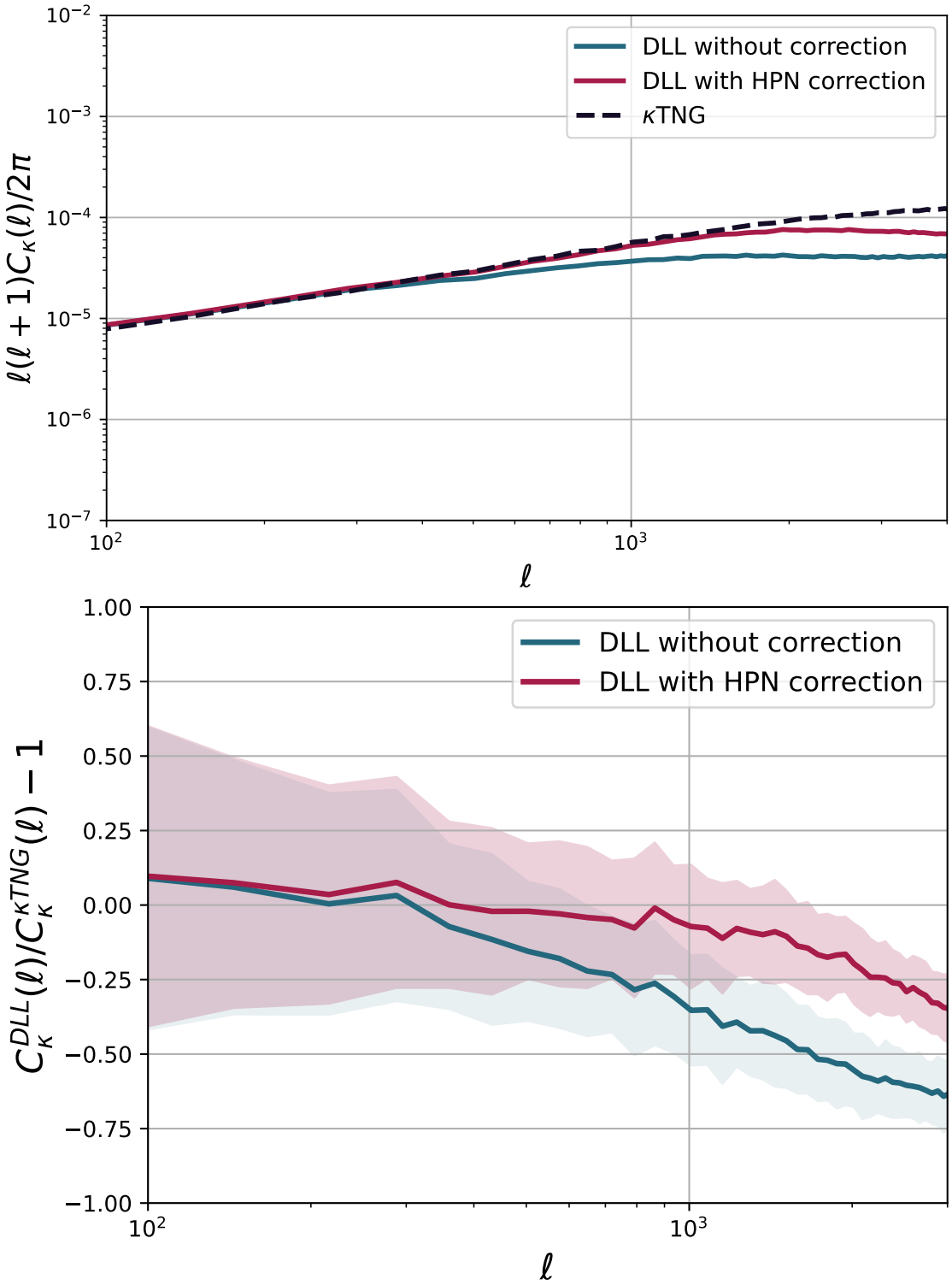

Figure 6.3: **Upper panel**: Angular power spectra of PM simulations before and after using the Hybrid Physical-Neural (HPN) correction model compared to the κTNG-Dark prediction. **Lower panel**: fractional angular power spectrum of PM simulations before and after using the Hybrid Physical-Neural (HPN) correction model and the κTNG-Dark prediction. The power spectra and ratios are means over 100 independent map realisations and the shaded regions represent the standard deviation from 100 DLL realisations. The spectra are computed for the source redshift $z_s = 0.91$.

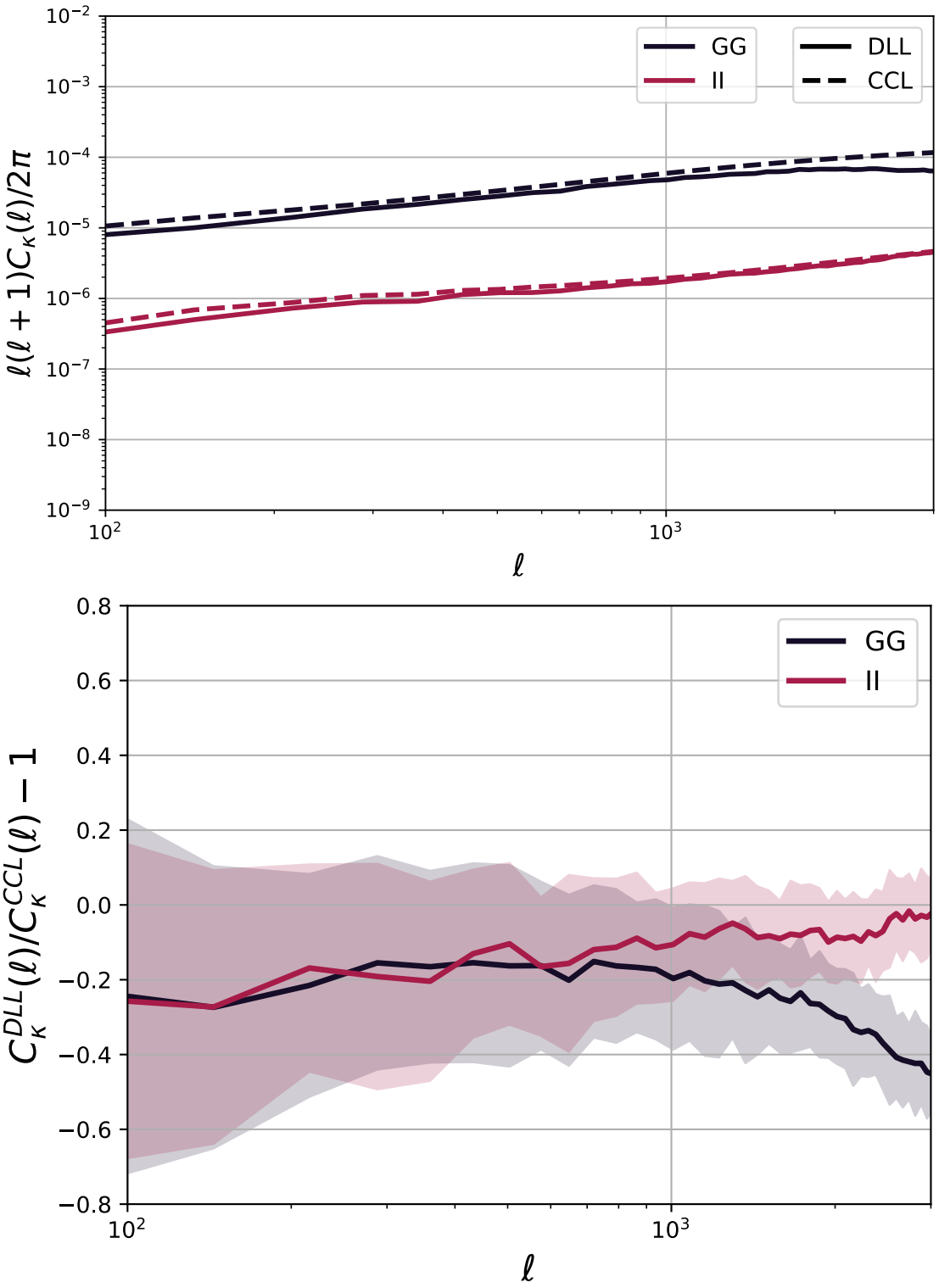

Figure 6.4: **Upper panel**: The C_ℓ^{II} and C_ℓ^{GG} contributions from theoretical predictions (dashed line) and DLL simulations.
Lower panel: The fractional difference between the theoretical and simulated C_ℓ^{II} and C_ℓ^{GG} contributes. We can see that we measure a reduced power spectrum at low ℓ compared to the theoretical predictions. This can be explained by the small volume of our simulation and the related low number of large-scale modes. The power spectra and ratios are means over 100 independent map realisations and the shaded regions represent the standard deviation from 100 realisations.

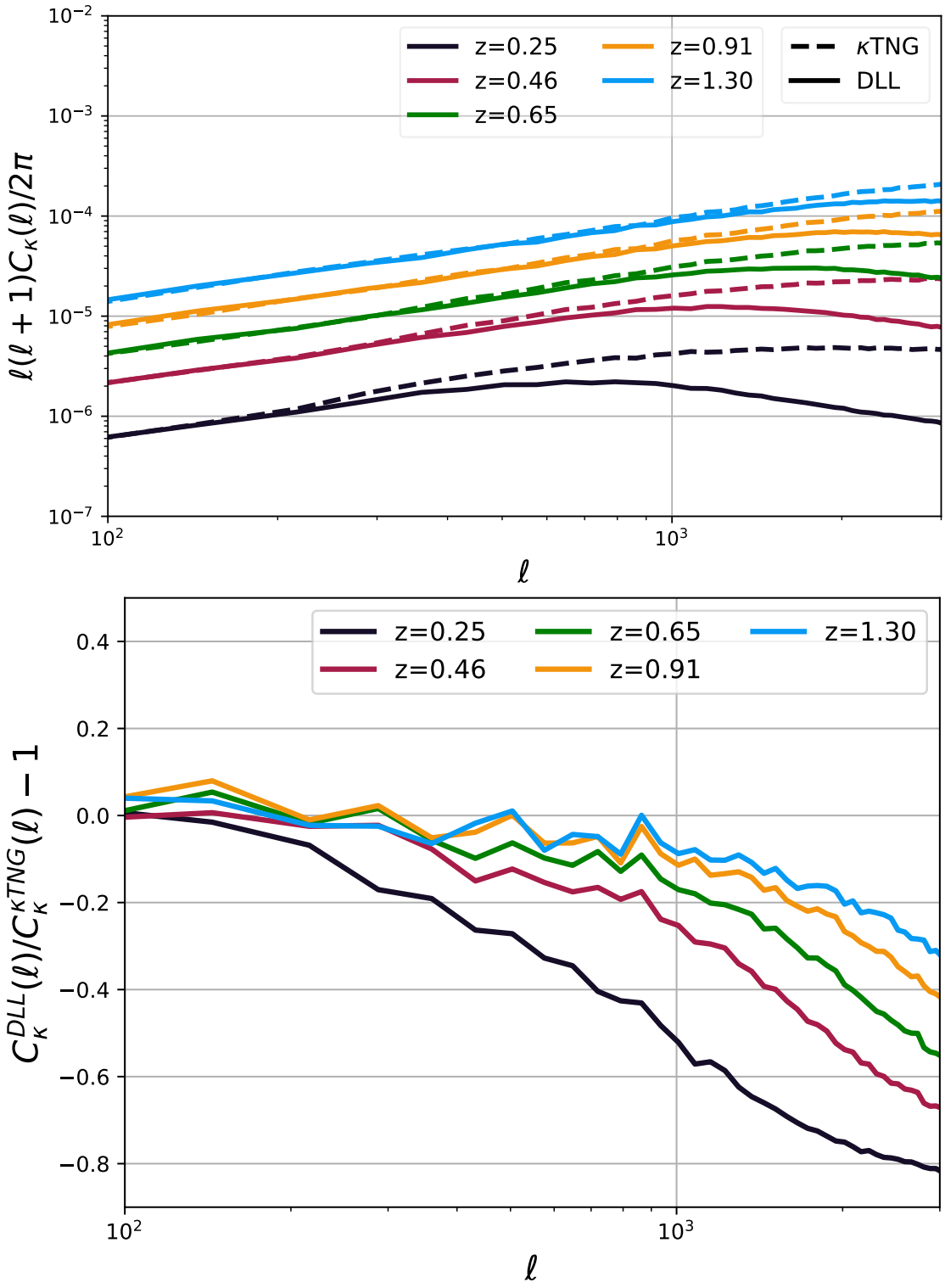

Figure 6.5: **Upper panel**: Angular power spectra for 5 source redshift from our DLL maps compared to the κTNG-Dark predictions. **Lower panel:** Fractional angular power spectra of DLL simulations and κTNG-Dark simulations for different source redshift. The power spectra mean over 100 independent map realisations and the shaded regions represent the standard deviation from 100 independent DLL realisations.

be explained considering that a given value of ℓ at lower redshift corresponds to smaller scales, some of those below the resolution of our simulations. We conclude that, if for z=0.91 and z=1.30 we have a general agreement with κTNG-Dark, with this specific setting of the model, we can not model correctly cases with sources redshift lower than z=0.91.

We want to highlight that the results shown are produced keeping the resolution of the simulations extremely low, and we do not expect to get the same precision as κTNG-Dark. The purpose of these tests, and the overall goal of this study, is to present a proof of concept of the DLL package and its potential. In practice, we will not work at this resolution.

Nevertheless, note that the simulations presented here already achieve a similar resolution to the MassiveNus simulations (Liu et al., 2018), despite being generated using one single GPU.

6.2.4 Peak counts

We compute the starlet peak counts as wavelet coefficients with values higher than their eight neighbors. We define $J_{max} = 7$ in Equation 6.7, this starlet filter applied to our map with a pixel size of 0.29 arcmin, corresponds to a decomposition in 7 maps of resolution [0.59, 1.17, 2.34, 4.68, 9.33, 18.79, 37.38] arcmin and a coarse map. To satisfy the survey requirement and keep the analysis centered in the range $\ell = [300, 3000]$, we consider only the scales corresponding to [9.33, 18.79, 37.38] arcmin. The peaks are counted for 8 linearly spaced bins within the range $(\kappa * \mathcal{W}) = [-0.1, 1.]$.

As with the power spectrum, we compare the peak counts statistic from our map to the one from the κTNG-Dark for different redshift bins. We present the results in Figure 6.6. These results are shown in terms of S/N, where the signal to noise is defined as the ratio between the amplitude of wavelet coefficients over the standard deviation of the noise expected for our survey choice in each wavelet band. At wavelet scale $\theta = 9.33$ arcmin the differences for the $z_s = 0.91$ curves are within the 20% up to $S/N = 3$, for $S/N > 3$ the accuracy is between the 20% and the 50%. At larger scale, $\theta = 18.79$ arcmin the accuracy is within the 20%. Finally, at $\theta = 37.38$ arcmin the accuracy is within the 15%, except $S/N < 1$ where the accuracy decreases up to 28%. The results slightly improve for z=1.30, showing an accuracy within the 35% for scale $\theta = 9.33$ arcmin, within the 10% for $\theta = 18.79$ arcmin and 25% for $\theta = 37.38$ arcmin. As for the power spectrum case, we observe higher discrepancies at lower redshift, hence we can conclude that, with the current setting of our simulation, we can not model correctly such redshift.

6.2.5 Validation with higher resolution simulations

In section 6.2, we presented a validation of our simulations by comparing the statistics from DLL and κTNG-Dark. In particular, we have seen a discrepancy on small scales

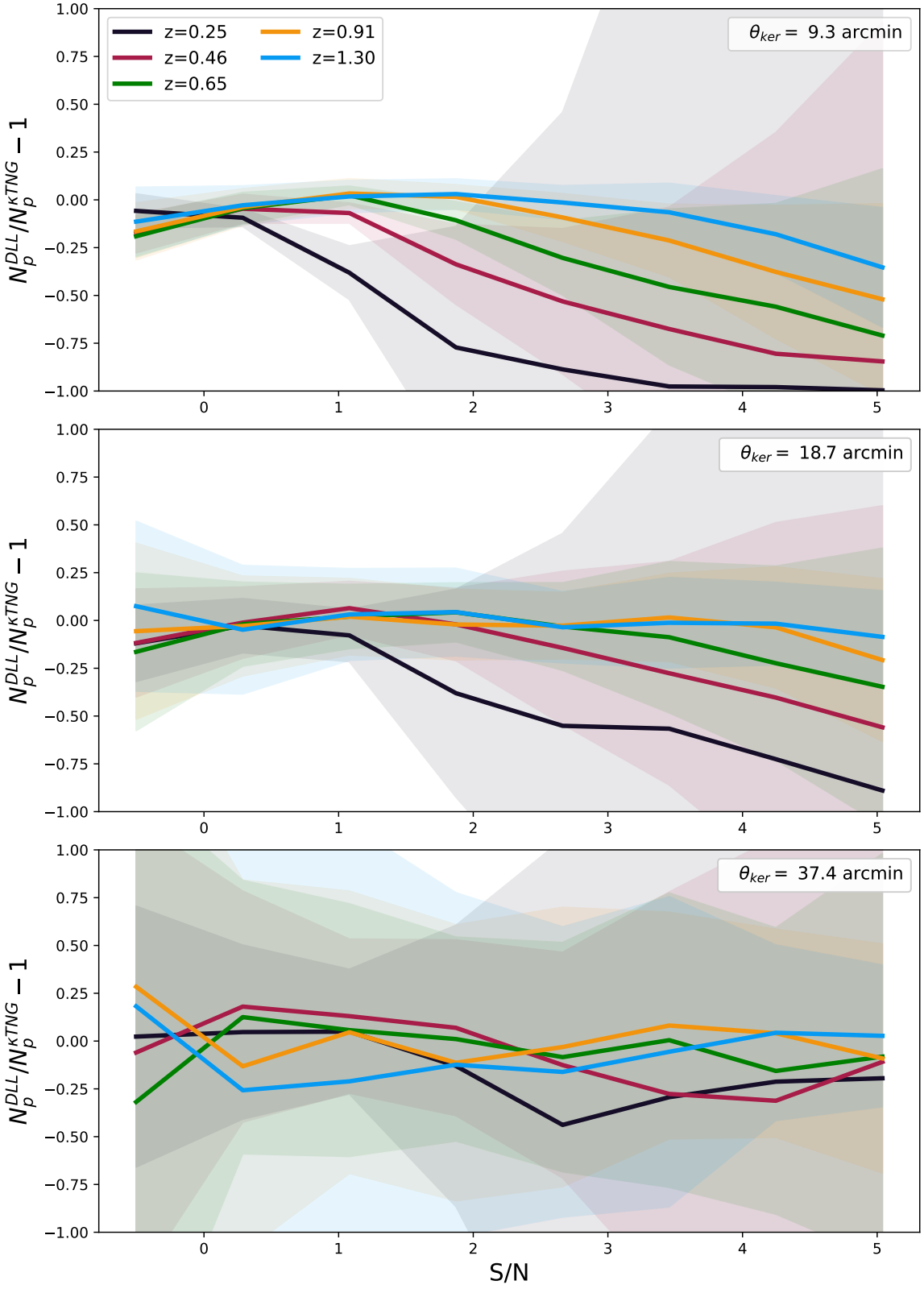

Figure 6.6: Fractional number of peaks of DLL simulations and κTNG-Dark simulations for different sources redshift. The peak counts distributions are shown for each starlet scales resolutions used: 9.34 (upper panel), 18.17 (center panel), 37.38 arcmins (lower panel). The results mean over 100 independent map realisations and the shaded regions represent the standard deviation from 100 independent DLL realisations.

for both the C_ℓ and the peak counts. We have attributed this bias to the low resolution of the simulations. Therefore, to justify this assumption, we simulate convergence maps of higher resolution, i.e., we raise the number of particles but keep the same box of $205^3(h^{-1}\text{Mpc})^3$ in the simulation. In the left panel of Figure 6.11, we present the angular power spectrum computed from our DLL with the original number count of particles (128^3) and the power spectrum computed from higher resolutions DLL simulations (212^3). The two outputs are compared to the κTNG-Dark predictions. In the right panel of Figure 6.11 the fractional differences between the convergence power spectra from the two maps and the κTNG-Dark are shown. We can see that, by increasing the number of particles, we can improve the accuracy of the lensing C_ℓ up to 20%. As for the power spectrum, we compare the peak counts statistic from our map to the one from the κTNG-Dark for different resolutions. In Figure 6.12, we compare the peak counts statistic from our map to the one from the κTNG-Dark for different resolutions. We use the same wavelet decomposition presented in subsection 6.2.4. As for the power spectrum, we note the same tendency to recover better accuracy when the resolution is increased.

6.3 Application: Fisher forecast

As an example of application of differentiable simulations, we aim to investigate the degeneracy between the cosmological parameters in high dimensional space and when systematics, such as the intrinsic alignment, are included in the analysis. Thanks to automatic differentiation, taking the derivative through the simulation with respect to the initial cosmological and nuisance parameters is now possible, thus allowing, among other things, for Fisher forecasts. In this section, we briefly introduce the Fisher forecast formalism. We also describe in detail the specific choices for the analysis we use throughout.

6.3.1 Forecast formalism

Fisher forecast is a widely used tool in cosmology for different purposes, e.g. investigate the impact of systematic sources or forecast the expected constraining power of the analysis (Tegmark et al., 1997). It can be thought as a tool to forecast error from a given experimental setup and quantify how much information we can extract from it. The Fisher matrix is the expectation value of the Hessian matrix of the negative log-likelihood $\mathcal{L}(C(\ell);\theta)$:

$$F_{\alpha\beta} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle , \tag{6.11}$$

where we indicate with θ the cosmological parameters or any systematics included in the simulation. If we assume a Gaussian likelihood and a Covariance matrix C_{ij}

independent from the parameters θ, the Fisher matrix can be computed as:

$$F_{\alpha\beta} = \sum_{ij} \frac{\partial \mu}{\partial \theta_\alpha} C_{ij}^{-1} \frac{\partial \mu}{\partial \theta_\beta} \tag{6.12}$$

were we indicate as $\frac{\partial \mu}{\partial \theta_\alpha}$ the derivatives of the summary statistics w.r.t the cosmological or nuisance parameters evaluated at the fiducial values. So, under the assumption of Gaussian likelihood, the Fisher information matrix provides a lower bound on the expected errors on cosmological parameters.

6.3.2 Analysis choices

To perform our study we use a single source redshift at z=0.91. Specifically, we generate 5000 independent map realisations to which we add Gaussian noise with mean zero and variance:

$$\sigma_n^2 = \frac{\sigma_e^2}{A_{pix} n_{gal}}, \tag{6.13}$$

where we set the shape noise $\sigma_e = 0.26$, the pixel area $A_{pix} = 0.086$ arcmin^2 and the galaxy number density $n_{gal} = 20$ arcmin^{-2}. We assume a parameter-independent covariance matrix computed as:

$$C_{ij} = \sum_{r=1}^{N} \frac{(x_i^r - \mu_i)(x_j^r - \mu_j)}{N-1} \tag{6.14}$$

where N is the number of independent realizations, x_i^r is the value of the summary statistics in the i^{th} bin for a given realization r, and μ is the mean of the summary statistics over all the realization in a given bin. In addition, we adopt the estimator introduced by Hartlap et al. (2007) to take into account the loss of information caused by the finite numbers of bins and realizations, i.e. we compute the inverse of the covariance matrix as :

$$C^{-1} = \frac{N - n_{bins} - 2}{N-1} C_*^{-1}, \tag{6.15}$$

where C_* is the covariance matrix defined in Equation 6.14. As mentioned before, we want to focus on a fair comparison between the power spectrum and the peak counts method. To be sure we are considering the same scales for both statistics, we apply a wavelet pass-band filter to the maps to isolate particular scales before measuring the power spectrum. We use the same scales used for the Peak counts, i.e. we decompose the noisy convergence map in seven images, we sum back only the three maps corresponding to [9.33, 18.79, 37.38] arcmin and compute the angular C_ℓ on the resulting image. An example of the C_ℓ computed for each individual starlet scale image and for the summed image is depicted in Figure 6.7.

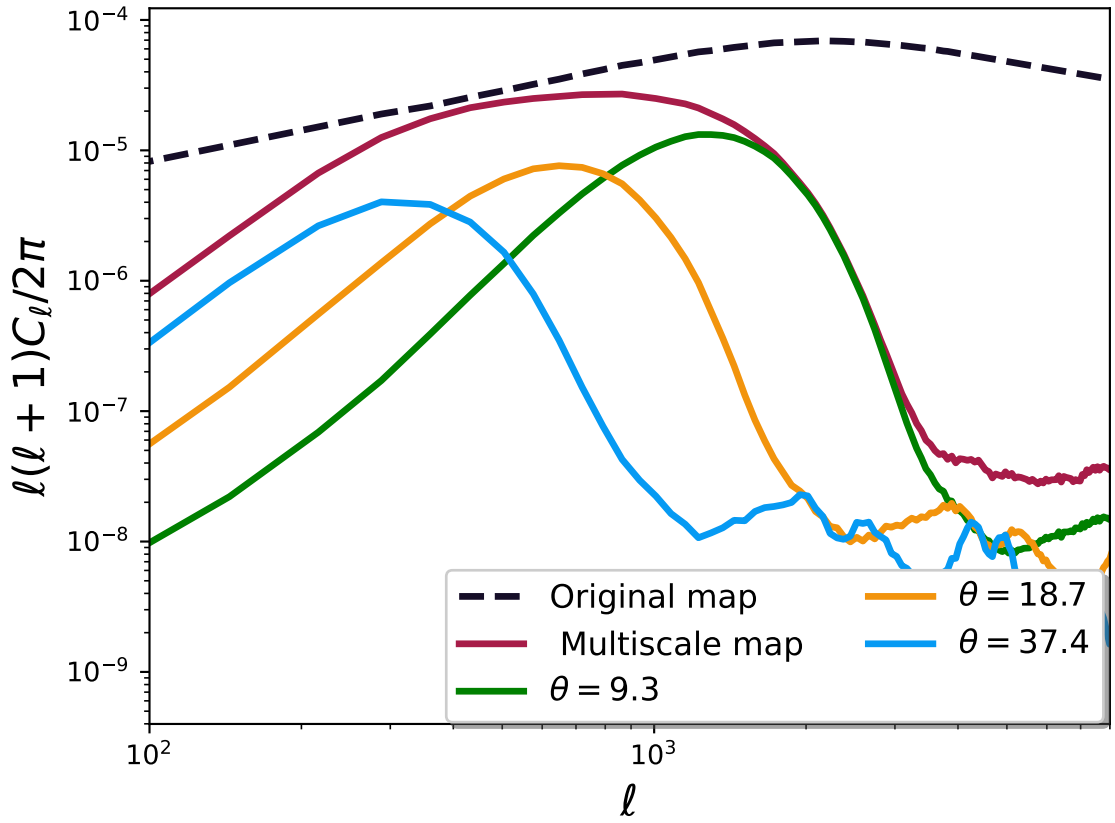

Figure 6.7: Example of the filtered C_ℓ used for the analysis. The colored lines show the C_ℓ computed on maps with a different resolution of the starlet decomposition. Specifically: the blue line (multiscale map) corresponds to C_ℓ computed on the summed image, the black dashed line (Original map) corresponds to the standard C_ℓ computed from a non-filtered map.

For each map, we compute the angular power spectrum and the peak counts by using our own differentiable code implemented in the TensorFlow framework. [2]

The derivatives with respect to all parameters are evaluated at the fiducial cosmology as the mean of 1500 and 2600 independent measurements for the peak counts and the C_ℓ respectively. Indeed, while the peak counts reach the convergence with ~ 1500 simulations, the C_ℓ proves to be more sensitive to noise and thus, requires more realizations to convergence. In Figure 6.10 in appendix B we test the stability of the Fisher contours by changing the number of simulated maps used to compute the Jacobian.

The priors used in the forecast process, are listed in Table 7.1 following Zhang et al. (2022). To take into account the partial coverage of the sky, we scale the Fisher matrix by the ratio $f_{\rm map}/f_{\rm survey}$, where $f_{\rm map}$ is the angular extend of our κmap $f_{\rm map} = 25$ deg^2 and $f_{\rm survey}$ corresponds to the size of the convergence maps for Stage IV-like survey $f_{\rm survey} = 15000$ deg^2.

6.3.3 Results

We now compare the relative constraining power of the lensing angular power spectrum and multiscale peak counts using the Fisher matrix formalism. As mentioned before, our interest is to investigate the sensitivity of the two weak-lensing statistics when systematic, such as the intrinsic alignment, and more cosmological parameters are included in the forecast. The results presented here are obtained from one single source redshift at z=0.91, assuming the fiducial cosmology and survey requirement presented in section 6.2 and subsection 6.3.2. The fiducial and priors ranges of the parameters are listed in Table 7.1.

Figure 6.8 shows the 2σ contours on the full ΛCDM parameter space and intrinsic alignment term considered for the two analyses. The contours obtained by the angular C_ℓ analysis are plotted in grey, and the ones for the peak counts in yellow. We find that in constraining Ω_c, σ_8 and A_{IA} peak counts outperform the power spectrum, while h, n_s and Ω_b parameters, within the limit of our setting, are not constrained by either and are prior dominated.

This is an interesting result, confirming the higher constraining power of weak-lensing peak counts as found in Ajani et al. (2020), especially considering that the two studies differ in multiple aspects. The most important difference between these two analyses is the parameters they include. Whereas Ajani et al. (2020) derive constraints on the sum of neutrino masses M_ν, the total matter density Ω_m, and the primordial power spectrum normalization A_s, we include the five cosmological parameters of the ΛCDM model and intrinsic alignment amplitude. The constraining power of the peak count statistic keeps being higher even in high dimensional cosmological parameter space and when the intrinsic alignment is included.

The chosen angular scales differ as well. Ajani et al. (2020) consider angular scales

[2] Code publicly available at:
https://github.com/LSSTDESC/DifferentiableHOS/statistics

Table 6.2: Prior and fiducial values used for the forecasting.

Parameter	Prior	Fiducial value
Ω_c	$\mathcal{N}$ (0.2589,0.2)	0.2589
Ω_b	$\mathcal{N}$ (0.0486,0.006)	0.0486
σ_8	$\mathcal{N}$ (0.8159,0.14)	0.8159
h	$\mathcal{N}$ (0.6774,0.063)	0.6774
n_s	$\mathcal{N}$ (0.9667,0.08)	0.9667
A_{IA}	$\mathcal{N}$ (0,3)	0.0

in the range $l = [300, 5000]$, while we focus, for both multiscale peak counts and C_ℓ, on scale approximately corresponding to the range $l = [300, 3000]$. Despite we are neglecting scales $\ell > 3000$, containing a larger amounts of non-Gaussian information, we find that for mildly non-linear scales we are considering, the peak counts statistic still constrains the cosmological parameters the most.

Finally, we find that the contours on the galaxy intrinsic alignment are significantly better constrained by the peak counts.

Validation against a theory prediction

We show the 2σ constraints obtained from our Fisher analysis of standard C_ℓ (orange contours) and the theoretical prediction from halofit (blue contours) in Figure 6.9. The dashed black contours define the prior used for the forecasting. The analysis is performed for one single source redshift at z=0.91 and the survey setup presented in section 6.2.

The constraints from the theoretical predictions are compatible with the ones obtained from the mock DLL maps for all cosmological parameters, except n_s. Indeed, despite sharing the same direction of degeneracy, the theoretical contours for n_s are narrower compared to the ones obtained in our analysis. In general, this translates into an increased uncertainty in constraining n_s, most likely due to the deficit in power observed for the C_ℓ at small scales.

The theoretical predictions are computed using the public library `jax-cosmo` (Campagne et al., 2023). We want to highlight that both the theoretical Fisher matrices and the ones from our analysis are obtained by automatic differentiation.

Validate the stability of the Fisher contours

To ensure that the shape of the ellipses and the direction of the degeneracies are not the results of stochasticity, we prove the stability of the Fisher analysis results by testing the convergence of the jacobians. In Figure 6.10 we present the results for Fisher constraints obtained when varying the number of independent simulations used to compute the jacobians. As we can see, the convergence seems to be reached for the

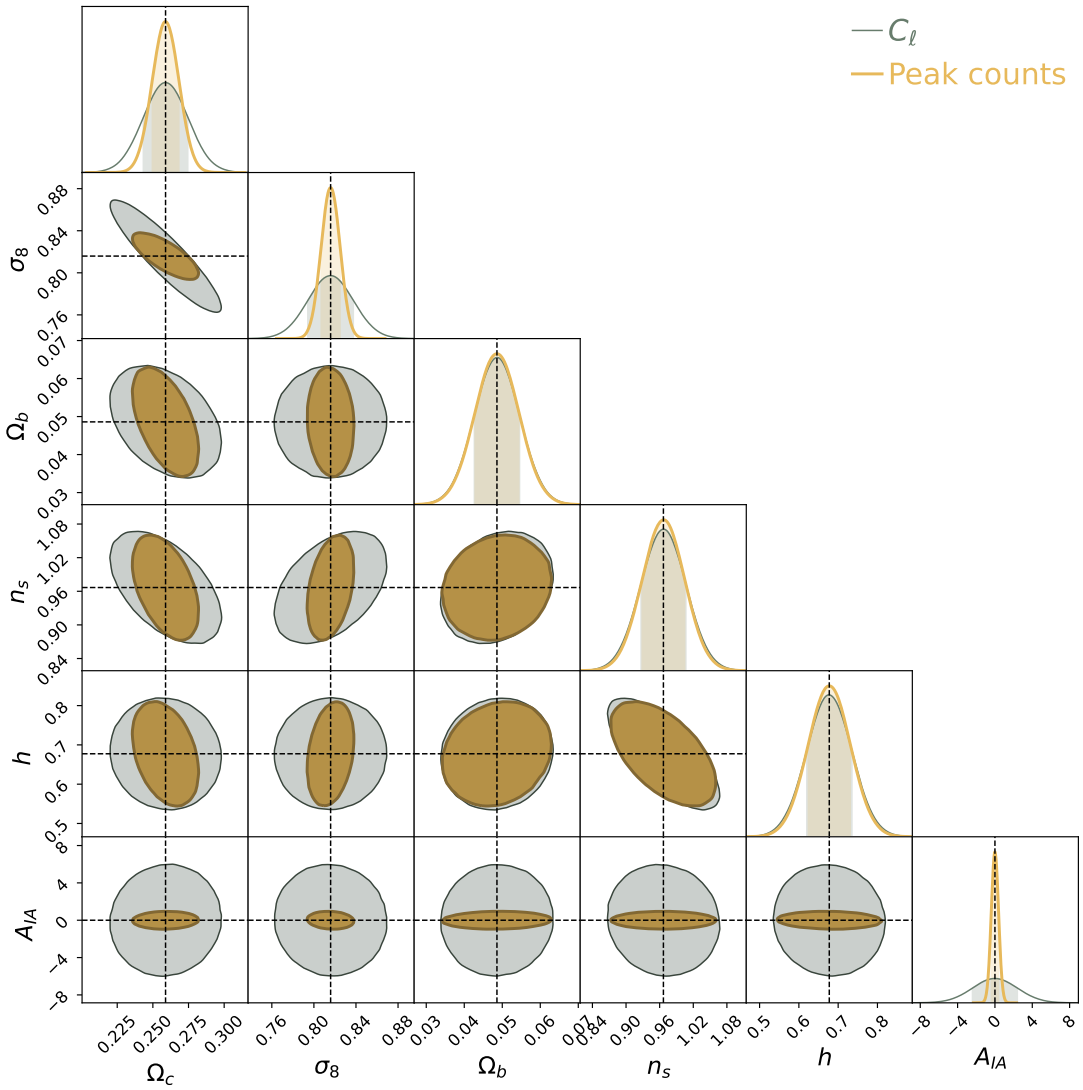

Figure 6.8: 2σ contours derived for one single source redshift at z=0.91 and the survey setup presented in section 6.2. The constraints are obtained by applying the starlet Peak counts (yellow contours) computed on noisy maps filtered using a starlet kernel of $[9.33, 18.70, 37.38]$ arcmin together and the wavelet pass-band filter for the C_ℓ statistics (grey contours) as described in section 6.2. The dashed black lines are located at the fiducial parameter values.

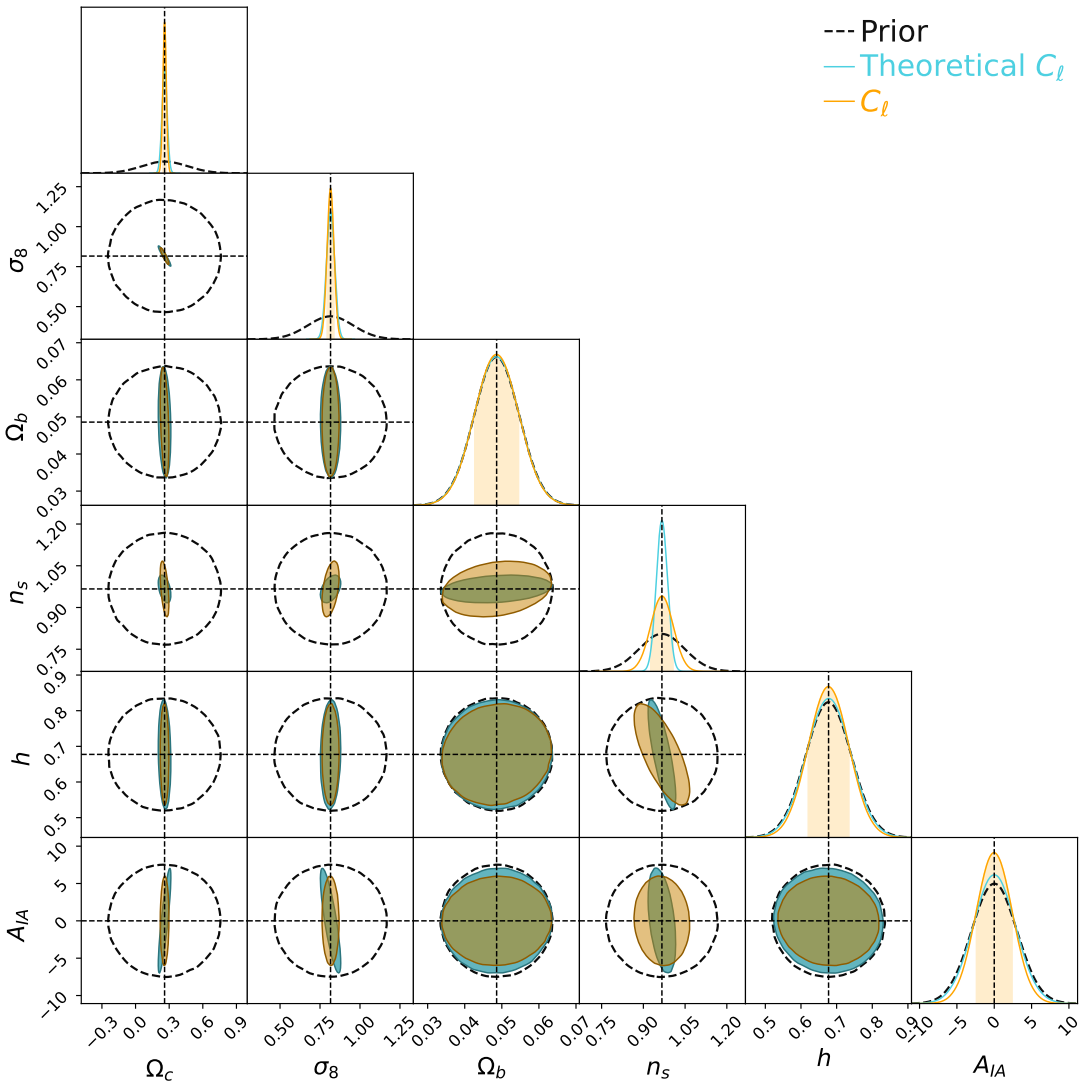

Figure 6.9: 2σ contours derived for one single source redshift at z=0.91 and the survey setup presented in section 6.2. We compare the Fisher matrix constraints on cosmological parameters and A_{IA} amplitude obtained with the C_ℓ from our mock maps (orange) and the theoretical C_ℓ (blue) obtained from the public library `jax-cosmo` (Campagne et al., 2023). In both cases, the constraints are obtained by applying the wavelet passband filter for the C_ℓ as described in section 6.2. The dashed black contours are the prior used for the forecasting.

peak counts with, at least, 1500 realizations. On the other hand, the angular C_ℓ proves to be more sensitive to noise and thus requires at least 2600 realizations.

Validation with higher resolutions simulations

Finally, we reproduce the results of the Fisher analysis for the intrinsic alignment parameter A_{IA} with higher-resolution simulations. We adopt the same Forecast criteria presented in section 6.3. However, for this specific test, we compute the derivatives numerically using finite differences. The step sizes used for these variations are $\Delta x_{IA} = 0.15$ for the C_ℓ and $\Delta x_{IA} = 1.2$ for the peak counts. In order to check the reliability of the numerical derivatives, we investigate the stability of the Fisher forecast against different step sizes used to compute them. The derivatives are computed as the mean of 3000 independent realizations for both C_ℓ and peak counts.

We confirm the peak counts provide the most stringent constraints on the galaxy intrinsic alignment amplitude A_{IA}. As for the full analysis, we tested the stability of the Fisher forecast by varying the number of simulated maps used to compute the derivatives. In Figure 6.13 we present the 1σ error on A_{IA} when varying the number of independent realizations used to compute the derivatives. It is interesting to note, that, even in this case, the derivatives of the C_ℓ can not be considered fully converged. However, as can be noted from the stability plots of Figure 6.13, the noise in the derivatives leads to tighter constraints in the Fisher forecast. Hence, the fully converged derivatives of the C_ℓ would result in even broader constraints, without changing the results we found.

6.4 Discussion

In this section, we discuss the limitations of the methodology and results obtained in the presented study highlighting in particular strategies for future extensions and applications.

In this work, we only used a single source plane in our Fisher forecast analysis, which does not allow us to evaluate the full impact that IA would have in a tomographic analysis. In particular, we do not have a contribution from the GI term. Many studies have demonstrated that the tomographic analysis can significantly improve constraints on cosmological and IA parameters. (King and Schneider, 2003; Heymans et al., 2004; Troxel and Ishak, 2015). Although it is straightforward to generalize all the results shown in this work to the tomographic case, this will require increasing the resolution of the simulation at lower redshifts (as illustrated by Figure 6.5) in order to model correctly low redshift bins. Since the maximum number of particles we can adopt in a simulation is closely limited to the GPU memory, we are building a distributed implementation of DLL, which will allow us to increase the resolutions of the simulations to the point of modeling correctly even the smaller scales at the lower redshifts. Indeed, current GPU-based simulations are capable of simulating 256^3 particles

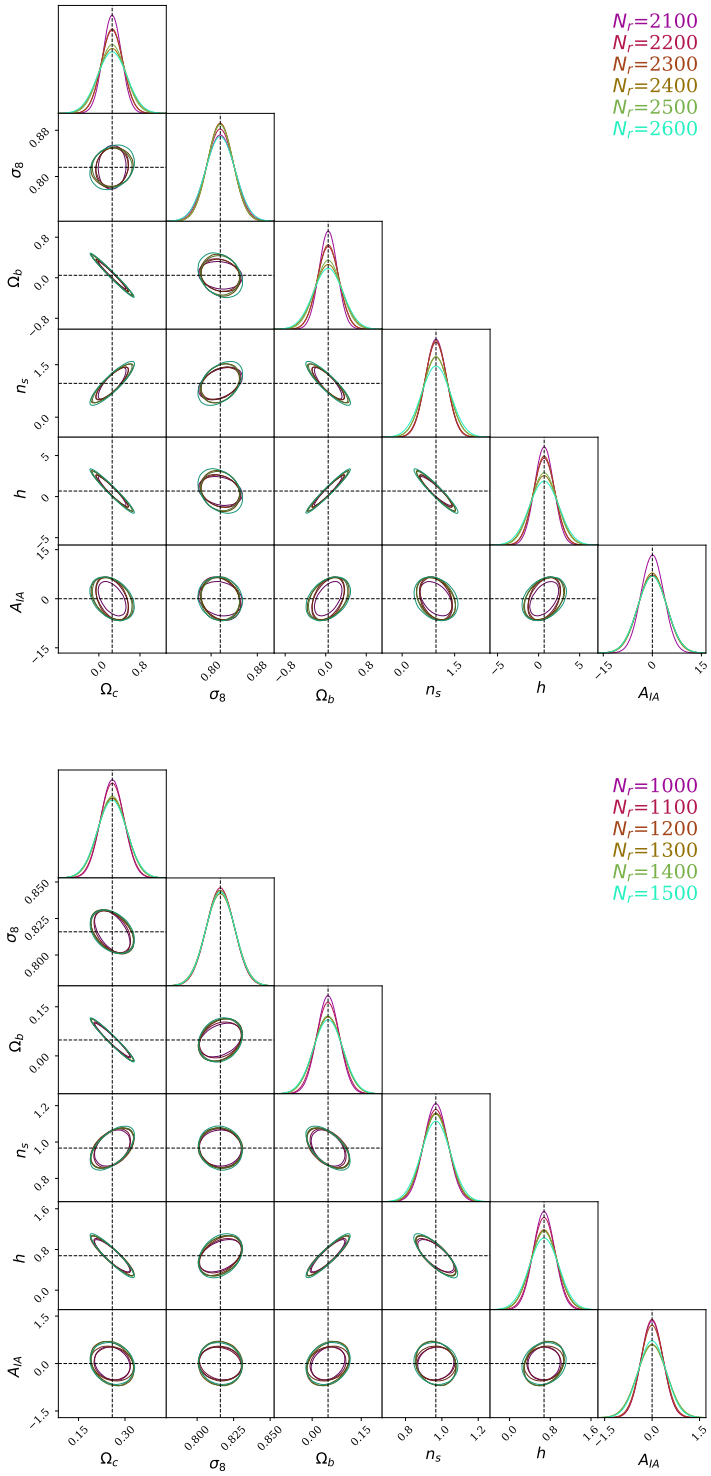

Figure 6.10: 1σ Fisher contours derived for one single source redshift at z=0.91 and the survey setup presented in section 6.2 for the C_ℓ (upper panel) and the Peak counts (lower panel). The different colors refer to the number of independent realizations used to mean the Jacobian in the Equation 6.12. The dashed black lines are located at the fiducial parameter values.

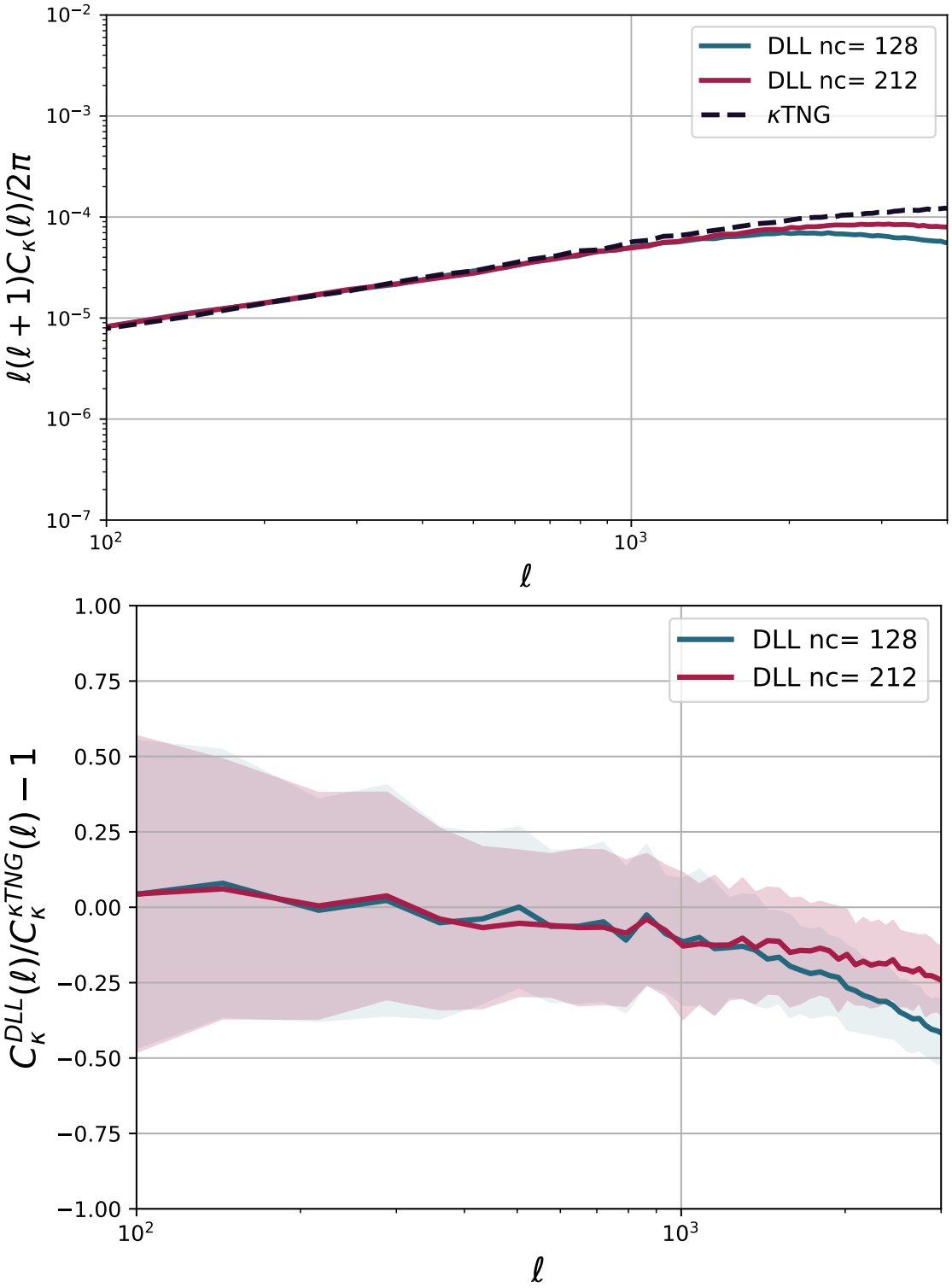

Figure 6.11: **Left panel**: Angular power spectra of PM simulations with original resolution (number count 128^3, blue line) and higher resolution (number count 212^3, red line) compared to the κTNG-Dark prediction. **Right panel**: fractional angular power spectrum of PM simulations with original and higher resolution and the κTNG-Dark prediction. The power spectra and ratios are means over 100 independent map realisations and the shaded regions represent the error on the mean. The spectra are computed for the source redshift $z_s = 0.91$.

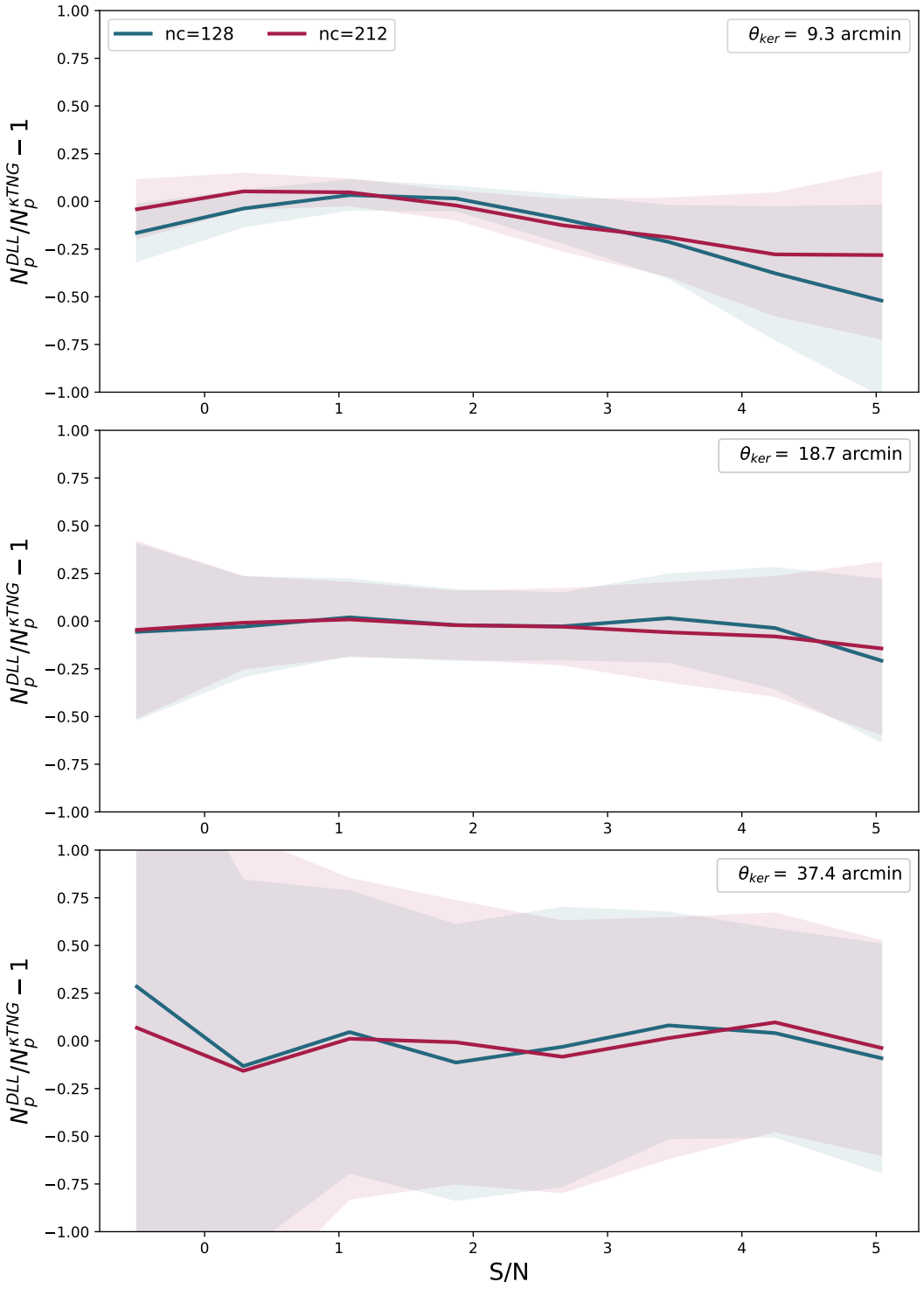

Figure 6.12: Fractional number of peaks of DLL simulations and κTNG-Dark simulations. The results are shown for the number counts 128^3 (blue lines) and 212^3 (orange lines). The peak counts distributions are shown for each starlet scales resolutions used: 9.34 (upper panel), 18.17 (center panel), 37.38 arcmins (lower panel). The results mean over 100 independent map realisations, the shaded regions represent the error on the mean. The statistics are computed for the source redshift $z_s = 0.91$.

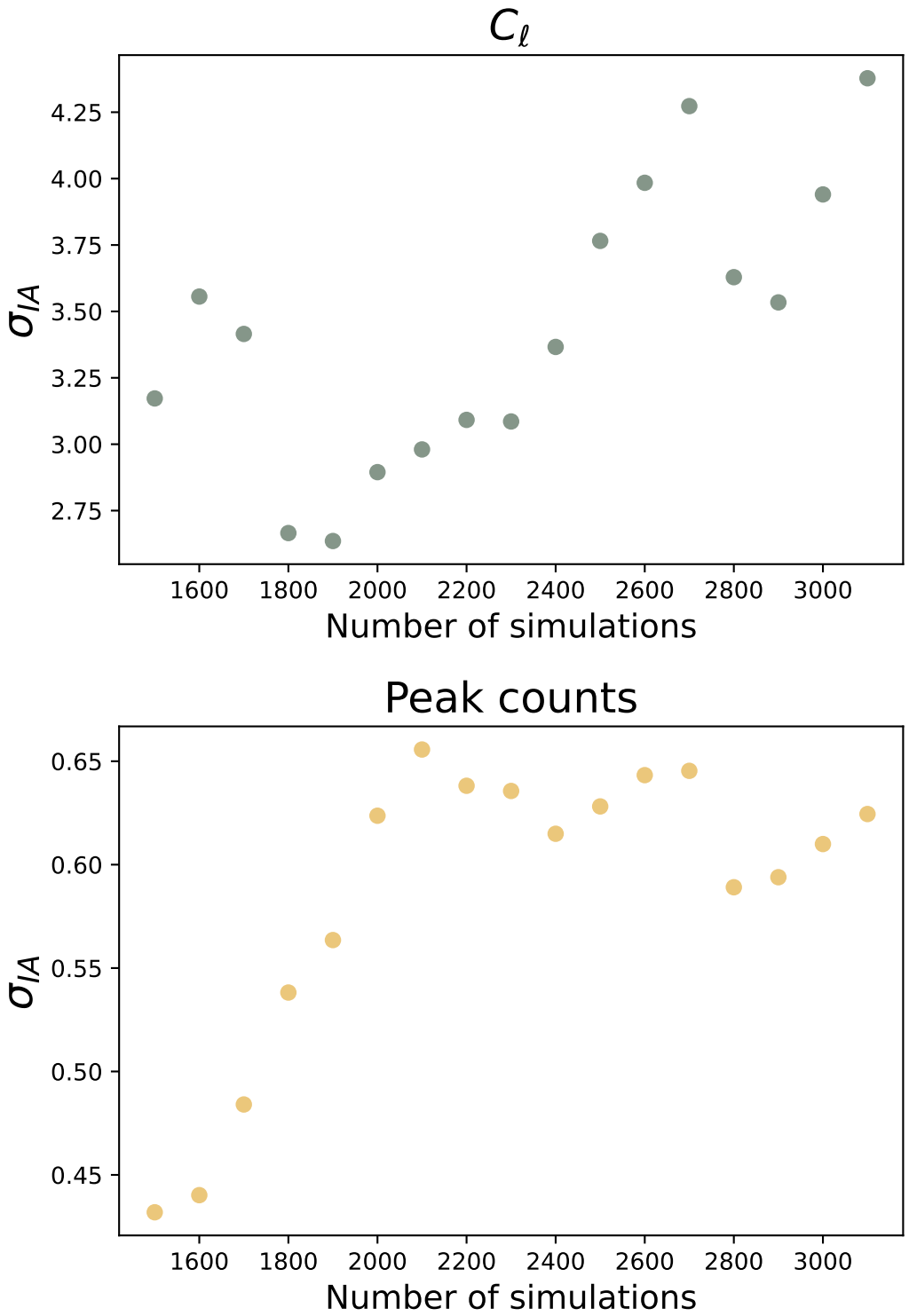

Figure 6.13: 1σ error on A_{IA} derived for one single source redshift at z=0.91 for different numbers of independent realizations used to mean the Jacobian in the Equation 6.12. The results are shown for the C_ℓ (left panel) and the peak counts (right panel).

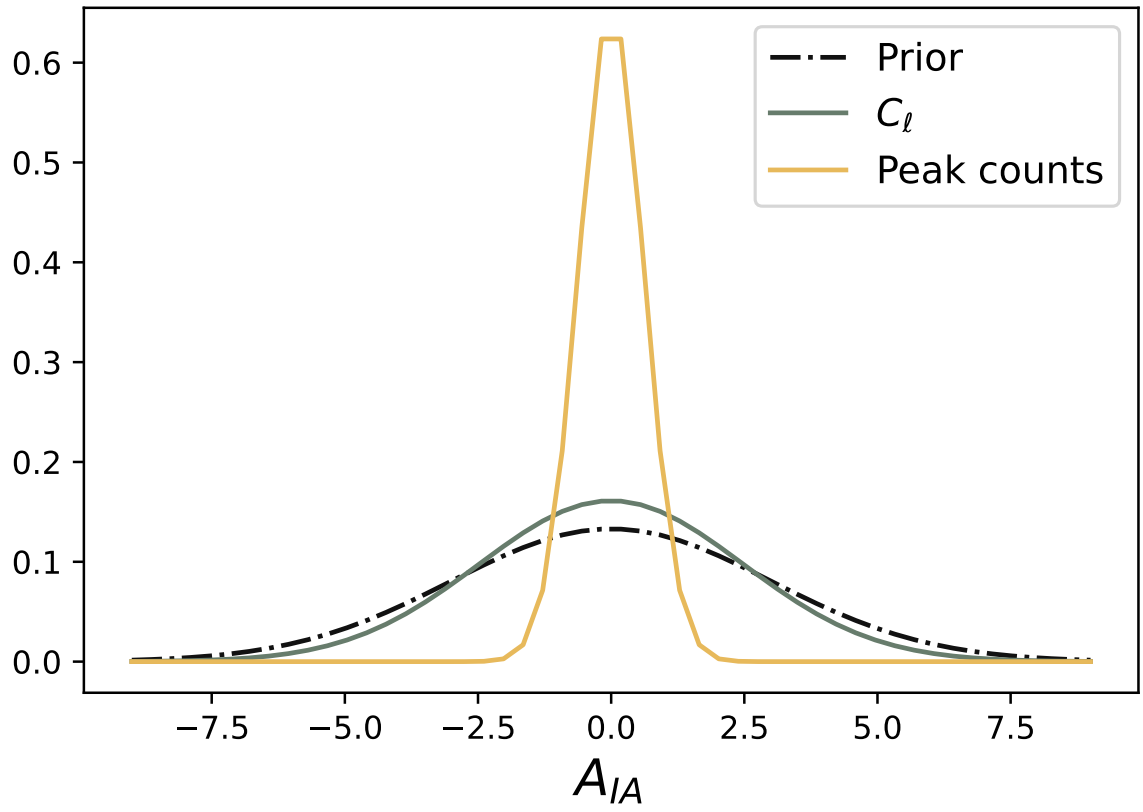

Figure 6.14: 1σ error on A_{IA} for one single source redshift at z=0.91 and the survey setup presented in section 6.2, from mock simulations with 212^3 particles. The results are obtained by applying the starlet Peak counts (yellow contours) computed on noisy maps filtered using a starlet kernel of $[9.33, 18.70, 37.38]$ arcmin together and the wavelet pass-band filter for the C_ℓ statistics (grey contours) as described in section 6.2.

on most available GPUs, and even 512^3 particles on the latest h-100 NVIDIA GPUs, which come with 80 GB of memory. This capability already allows us to run realistic simulations for developing pipelines for cosmological analysis. To surpass these limitations, we are actively developing distributed implementations. It is worth noting that this has already been achieved for previous generations of simulations, like FlowPM, which can run 2048^3 particles across 256 NVIDIA V100 GPUs. Additionally, we should mention that these simulations can also run on CPUs, which are less constrained by memory but lack GPU-based accelerations. A time-scaling comparison of GPU and CPU-based simulations is presented in Modi et al. (2021) and Li et al. (2022). The results clearly demonstrate the advantages of GPU-based simulations, even when not considering additional benefits such as automatic differentiation capabilities.

Another direction for further development is the ray tracing methodology. In our method, we construct the weak lensing maps assuming the Born approximation. However, Petri et al. (2017) shows that for an LSST-like survey, while the Born approximation leads to negligible parameter bias for the power spectrum, it can lead to significant parameter bias for higher-order statistics. Hence, the natural next step will be to implement a ray-tracing algorithm beyond the Born approximation in our pipeline. We aim to adopt the multiple-lens-plane approximation (Blandford and Narayan, 1986; Seitz et al., 1994; Jain et al., 2000; Vale and White, 2003; Hilbert, S. et al., 2009), i.e. by introducing lens planes perpendicular to the line-of-sight, the deflection experienced by the light rays due to the matter inhomogeneities will be approximated through multiple deflections at the lens planes. More specifically, we will implement the memory-efficient ray-tracing scheme proposed by Hilbert, S. et al. (2009) in the Tensorflow framework.

On the theoretical modeling side, we studied the impact of the intrinsic alignment of galaxies assuming a linear coupling between the intrinsic galaxy shapes and the non-linear projected tidal fields, i.e. adopting the NLA model. This physical description for the IA is only an approximation since it does not take into account the tidal torque field. In future work, we aim to extend the NLA model by implementing the extended δNLA model, described by Harnois-Déraps et al. (2021).

Finally, we presented a tool based on only Dark matter simulations. We note that this would force us to perform conservative scale cuts in the inference analysis to not include scales affected by baryonic effects. A future prospect is to include baryonic effects in the analysis. One possible way applicable to our methodology could be to extend the Hybrid Physical Neural ODE approach and apply more sophisticated models to learn the physics that controls the hydrodynamics simulations.

We expect that the methods illustrated in this study will be extended to different relevant use-cases. A particularly suitable example is related to the application of algorithms such as the Variational Inference and the Hamiltonian Monte Carlo that are widely used in the Bayesian inference context and were until now excluded due to the lack of derivatives. A further example is provided by Zeghal et al. (2022), which demonstrates that having access to the gradients of the forward model is beneficial to

constrain the posterior density estimates.

6.5 Conclusions

In this chapter, we have presented the Differentiable Lensing Lightcone (DLL) model, a fast lensing lightcone simulator providing access to the gradient. We extended the public FlowPM N-body code, implementing the Born approximation in the Tensorflow framework to create non-Gaussian convergence maps of weak gravitational lensing. To allow DLL to run at low resolution without affecting significantly the accuracy, we complement the FlowPM N-body code with the Hybrid-Physical Neural scheme, a new correction scheme for quasi N-body PM solver, based on Neural Network implemented as a Fourier-space filter. We validate our tool by comparing the C_ℓ and peak counts statistics against predictions from κTNG-Dark simulations. To do this, we run simulations following the evolution of 128^3 particles and we produce weak lensing convergence maps for several redshift sources. We show that, despite being generated at low computation costs, we recover a good match for redshift equal to or higher than $z = 0.91$. To demonstrate the potential of our tool, as a first use case, we exploit the automatic differentiability of the simulations to do Fisher forecast. Thanks to back-propagation, accessing the derivative through the simulations w.r.t. the cosmological parameters and A_{IA} parameter is possible at the same computational cost as the forward simulation. Assuming an LSST-like setting, we simulate weak lensing convergence maps for a single source redshift $z = 0.91$ and angular extend of $5°$, based on a periodic box of comoving volume equal to 205 h^{-1}Mpc. We compute the constraints on the resulting convergence maps with the starlet peak counts and use a wavelet-filtered lensing power spectrum as a benchmark for the comparison. Within the limits of the analysis choices made in this study, we obtain the following results:

- We confirm that the peak count statistics outperform the two-point statistics as found in Ajani et al. (2020), even in high dimensional cosmological and nuisance parameter space.
- We find the peak counts to provide the most stringent constraints on the galaxy intrinsic alignment amplitude A_{IA}.

To conclude, the framework outlined here can provide many advantages in the context of cosmological parameter inference: it is the first step in the development of fully differentiable inference pipelines for weak lensing, and it is a fast tool to further explore the sensitivity of higher-order statistics to systematics.

Chapter 7

Optimal Neural Summarisation for Full-Field Implicit Inference by Density Estimation

This chapter presents some preliminary results of an ongoing work. My contributions included building the physical model for the simulator, conducting the analysis, and preparing the material for publication. The research project is done in collaboration with Justine Zeghal, T. Lucas Makinen, Alexandre Boucaud, François Lanusse, and Jean-Luc Starck.

As mentioned several times throughout this thesis, in many cosmological applications the likelihood function is either unknown or intractable, limiting the use of traditional inference methods. Likelihood-Free Inference (LFI) approaches offer a promising solution by using forward-simulated synthetic data to approximate the posterior (or the

likelihood) distribution instead of evaluating an explicit likelihood function. However, these methods suffer from the curse of dimensionality, necessitating the development of compression schemes to reduce high-dimensional data into lower-dimensional summary statistics. The goal of this chapter is to investigate the performance of several neural-compression strategies for full field Likelihood free inference by density estimation. Moreover, we aim to demonstrate that, by using an optimal compression strategy, the posterior distribution is comparable to those derived from a Bayesian forward modeling approach. We present a comparative analysis of different loss functions employed during the training of the neural network. We evaluate their performance by measuring their impact on the constraints on the ΛCDM parameters expected from LSST-Y10. For both forward-modeling the convergence field and generating observed mock data, we develop the `SBILens` package. `SBILens` is a Jax-based differentiable physical model used to generate convergence maps as lognormal random fields. It is specifically designed for Likelihood Free Inference and Bayesian inference algorithms that require access to likelihood derivatives. Albeit with preliminary results, our experiments validate the effectiveness of LFI methods, as we demonstrate comparable posterior distributions between LFI and Bayesian forward modeling approaches.

7.1 Motivation

With the increased statistical power of stage IV surveys, our cosmological analysis should not rely on the measurement of sub-optimal summary statistics, which may not fully capture the non-Gaussian information present in the lensing field at the scales accessible to future surveys. In this chapter, we introduce a forward model that directly extracts information from the convergence maps, rather than relying on the analytical evaluation of summary statistics. By doing so, we aim to preserve all available information and facilitate the incorporation of systematic effects and the combination of multiple cosmological probes through joint simulations. To ensure a smooth reading, we will now re-establish the definition of the simulator as previously introduced in section 4.2. In a forward model context, the simulator of the observables serves as our physical model, where each component is tractable. These models, often referred to as *probabilistic programs*, as illustrated by **?**, can be described as follows: The models take as input a vector parameter $\boldsymbol{\theta}$, which describes the underlying deterministic model. Then, they sample internal states $\boldsymbol{z}$, dubbed *latent variables*, from the distribution $p(\boldsymbol{z}|\boldsymbol{\theta})$. These states can be directly or indirectly related to a physically meaningful state of the system. Finally, the models generate the output $\boldsymbol{x}$ from the distribution $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{z})$, where $\boldsymbol{x}$ represents the observations.

The ultimate goal of Bayesian inference in cosmology is to compute the posterior distribution:

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int d\boldsymbol{\theta'} p(\boldsymbol{x}|\boldsymbol{\theta'})p(\boldsymbol{\theta'})}, \tag{7.1}$$

however, a problems arises because the marginal likelihood $p(\boldsymbol{x}|\boldsymbol{\theta})$ is typically intractable:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z} = \int p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) p(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}, \tag{7.2}$$

since it involves integrating over all potential paths through the latent space. To overcome this limitation while still capturing the full information content of the data, two different approaches have been proposed in the literature. Although these approaches are often referred to by different names, hereinafter we will make the following distinction:

Explicit inference referring to all Likelihood-based inference approaches. This approach involves treating the simulator as a probabilistic model and performing inference over the joint posterior:

$$p(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) p(\boldsymbol{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{7.3}$$

In the context of the forward model and full field analysis, Bayesian Hierarchical Models (BHMs) fall into this category, as discussed in subsubsection 3.3.2. As a reminder, a Bayesian forward approach involves using a given physical model to predict observations and then comparing these predictions with observations to infer the parameters of the model.

Implicit inference referring to all the approaches that try to infer the marginal likelihood from simulations. This second class of approaches involves treating the simulator as a black box with only the ability to sample from the joint distribution:

$$(\boldsymbol{x}, \boldsymbol{\theta}) \sim p(\boldsymbol{x}, \boldsymbol{\theta}). \tag{7.4}$$

Within this class of methods, as discussed in chapter 7, we can differentiate between more traditional methods such as Approximate Bayesian Computation (ABC) and Density Estimation Likelihood-Free Inference (DELFI) methods. In this work, our focus will be on DELFI methods, which approach the inference task as a density estimation problem.
In particular, the standard deep learning approach for Likelihood-Free Inference can be described as separated into two distinct steps:

1. Automatically learning an optimal low-dimensional summary statistic.

2. Using Neural Density Estimation in low dimensions to infer the target distributions.

In the first step, we introduce a parametric function $F_{\boldsymbol{\varphi}}$ such that:

$$\boldsymbol{t} = F_{\boldsymbol{\varphi}}(\boldsymbol{\theta}), \tag{7.5}$$

which aims to reduce the dimensionality of the data while preserving information. Typically, the compressed statistics $\boldsymbol{t}$ is assumed to have the same dimension of $\boldsymbol{\theta}$.
In the second step, Neural Density Estimation can target either building an estimate $p_{\boldsymbol{\varphi}}$ of the likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta})$ (referred to as the Neural Likelihood Estimation (NLE) task), or targeting the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$, (known as Neural Posterior Estimation (NPE) task).

The main motivation for this work is to evaluate the impact of a given compression strategy on the final posterior distribution and determine whether an optimal compression strategy exists. Indeed, it is important to consider that different neural compression techniques may not be equivalent. We summarize the various neural compression strategies found in the literature in Table 7.3. Many of these papers have used neural compression techniques that rely on optimizing the Mean Square Error or the Mean Absolute Error.
As we will demonstrate in the following sections, this corresponds to training the model to estimate the mean and the median of the posterior distribution. Other papers rely on assuming proxy Gaussian likelihoods and estimate the mean and covariance of these likelihoods from simulations. Such compression of summaries could be sub-optimal in certain applications, resulting in a loss of information and hence larger contours.

Therefore, our primary focus will be to investigate the effectiveness and validity of these compression techniques in the context of weak lensing applications (within the limits of our setting).

Ultimately, we should keep in mind that, for a given simulation model, if an optimal compression statistic is used, the two approaches should converge to the same posterior. Therefore, the goals of this work will be:

1. Find a near-optimal compression strategy.
2. Demonstrate that by using this near-optimal compression strategy, both the implicit and explicit methods yield comparable results.

7.2 The SBILens framework

To investigate the questions above, we have developed the Python package `SBILens`, a weak lensing differentiable simulator based on a log-normal model. `SBILens` enables the sampling of convergence maps in a tomographic setting while considering the cross-correlation between different redshift bins.

7.2.1 Lognormal Modeling

For various cosmological applications, the non-Gaussian field can be modeled as a Lognormal field (Coles and Jones, 1991; Böhm et al., 2017). This model offers the advantage of generating the matter or convergence field rapidly while allowing the extraction of information beyond the two-point statistics. Although studies demonstrated that this model fails in describing the 3D field (Klypin et al., 2018), it properly describes the 2D convergence field (Clerkin et al., 2017; Xavier et al., 2016). Assuming a simulated Gaussian convergence map κ_g, whose statistical properties are fully described by its power spectrum C_ℓ we know that this model is not a suitable representation of late-time and more evolved structures. One potential solution is to find a transformation $f(\kappa_g)$ of this map that captures the non-Gaussian features in the convergence field. In doing so, it is crucial to ensure that the transformed map maintains the correct mean and variance, effectively recovering the correct two-point statistics. Denoting μ and σ_g^2 the mean and covariance matrix of κ_g respectively, we can define the transformed convergence κ_{ln} as a shifted lognormal random field:

$$\kappa_{ln} = e^{\kappa_g} - \lambda, \tag{7.6}$$

where λ is a free parameter that determines the shift of the lognormal distribution. The convergence κ in a given redshift bin is fully determined by the shift parameter λ, the mean μ of the associated Gaussian field κ_g, and its variance σ_g^2. The correlation of the lognormal field, denoted as ξ_{ln}, is also a function of these variables and is related to ξ_g^{ij} through the following equations:

$$\begin{aligned} \xi_{ln}^{ij}(\theta) &\equiv \lambda_i \lambda_j (e^{\xi_g^{ij}(\theta)} - 1) \\ \xi_g^{ij}(\theta) &= \log\left[\frac{\xi_{ln}^{ij}(\theta)}{\lambda_i \lambda_j} + 1\right]. \end{aligned} \tag{7.7}$$

Here i and j define a pair of redshift bins. The parameter λ, also known as *minimum convergence parameter*, defines the lowest values for all possible values of κ. The modeling of the shift parameter can be approached in various ways. For example, it can be determined by matching moments of the distribution (Xavier et al., 2016) or by treating it as a free parameter (Hilbert et al., 2011). In general, the value of λ depends on the redshift, cosmology, and the scale of the field at which smoothing is applied.

While it is straightforward to simulate a single map, if we want to constrain the convergence map in different redshift bins, an additional condition must be met. The covariance of the map should recover the correct angular power spectrum:

$$\left\langle \tilde{\kappa}_{ln}^{(i)}(\ell)\tilde{\kappa}_{ln}^{*(j)}(\ell')\right\rangle = C_{ln}^{ij}(\ell)\delta^K(\ell-\ell') \tag{7.8}$$

where $C_{ln}^{ij}(\ell)$ is the power spectrum of κ_{ln} in Fourier space, defined as:

$$C_{ln}^{ij}(\ell) = 2\pi\int_0^\pi d\theta \sin\theta P_\ell(\cos\theta)\xi_{ln}^{ij}(\theta) \tag{7.9}$$

and P_ℓ is the Legendre polynomial of order ℓ. Using the lognormal model, we can simultaneously constrain the convergence field in different redshift bins while considering the correlation between the bins, as described by Equation 7.7.

In the SBILens framework, the sampling of the convergence maps can be described as follows. First, we define the survey in terms of galaxy number density, redshifts, and shape noise. Then, we compute the theoretical auto-angular power spectrum $C^{ii}(\ell)$ and cross-angular power spectrum $C^{ij}(\ell)$ for each tomographic bin. These theoretical predictions are calculated using the public library `jax-cosmo`. Next, we project the one-dimensional $C(\ell)$ onto two-dimensional grids with the desired final convergence map size. Afterwards, we compute the Gaussian correlation functions $\xi_g^{ij}(\theta)$ using Equation 7.7. To sample the convergence field in a specific redshift bin while considering the correlation with other bins, we use Equation 7.8. We construct the covariance matrix $\mathbf{\Sigma}$ of the random field $\boldsymbol{\kappa}$, where $\boldsymbol{\kappa}$ represents the vector of convergence maps at different redshifts, as follows:

$$\mathbf{\Sigma} = \begin{pmatrix} C_\ell^{11} & C_\ell^{12} & \cdots & C_\ell^{1n} \\ C_\ell^{21} & C_\ell^{22} & \cdots & C_\ell^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_\ell^{n1} & C_\ell^{n2} & \cdots & C_\ell^{nn} \end{pmatrix}. \tag{7.10}$$

To sample more efficiently, we perform an eigenvalue decomposition of $\mathbf{\Sigma}$ to obtain a new matrix $\tilde{\mathbf{\Sigma}}$:

$$\tilde{\mathbf{\Sigma}} = \boldsymbol{Q}\mathbf{\Lambda}^{1/2}\boldsymbol{Q}^T, \tag{7.11}$$

where $\boldsymbol{Q}$ and $\mathbf{\Lambda}$ are the eigenvectors and eigenvalues of $\mathbf{\Sigma}$, respectively. Next, we sample the Gaussian random maps $\boldsymbol{\kappa_g}$ using the equation:

$$\boldsymbol{\kappa_g} = \hat{\boldsymbol{Z}} \cdot \tilde{\mathbf{\Sigma}} \tag{7.12}$$

where $\hat{\boldsymbol{Z}}$ represents the Fourier transform of the latent variables of the simulator. Finally, we transform the Gaussian map κ_g into a LogNormal map κ_{ln} using Equation 7.6. To ensure that we recover the correct auto- and cross-power spectra, we compare the results from our simulations to theoretical predictions for different tomographic bin combinations. We show the results in Figure 7.3.

7.2.2 Data generation

Our analysis is based on a standard flat ΛCDM cosmological model, which includes the following parameters: the baryonic density fraction Ω_b, the total matter density fraction Ω_m, the Hubble parameter h_0, the spectral index n_s, the amplitude of the primordial power spectrum σ_8 and the dark energy parameter w_0. The priors used in the simulations and in the inference process are listed in Table 7.1, following Zhang et al. (2022). To simulate our data, we develop the SBILens package, which employs a Lognormal model to represent the convergence maps, as explained in the previous section. Specifically, the package uses the public library `jax-cosmo` (Campagne et al., 2023) to compute the theoretical power- and cross-spectra. The computation of the lognormal shift parameter is performed using the `Cosmomentum` code (Friedrich et al., 2018, 2020), which utilizes perturbation theory to compute the cosmology-dependent shift parameters. In Cosmomentum the calculation of the shift parameters assumes a cylindrical window function, while our pixels are rectangular. Following Boruah et al. (2022a), we compute the shift parameters at a characteristic scale, $R = \Delta L/\pi$, where ΔL represents the pixel resolution.
For each redshift bin, we tested the dependency of the shift parameter λ on various cosmological parameters. Specifically, we investigated how the value of λ changed when varying a specific cosmological parameter while keeping the others fixed. Our findings revealed that the parameters Ω_b, h_0, and n_s had almost no significant impact on λ. As a result, we computed the shift parameters for each redshift using the fiducial cosmology values of Ω_b, h_0, and n_s. To account for the cosmology dependence of λ on Ω_c, σ_8, and w_0, we calculated the shift for various points in the cosmological parameter space and then interpolated the shift values for other points in the parameter space. Each map is reproduced on a regular grid with dimensions of 256×256 pixels and covers an area of 10×10 deg^2. An example of a tomographic convergence map simulated using the `SBILens` package is shown in Figure 7.2.

7.2.3 Noise and survey setting

We conduct a tomographic study to reproduce the redshift distribution and the expected noise for the LSST-Y10 data release. Following Zhang et al. (2022), we model the underlying redshift distribution using the parametrized Smail distribution (Smail et al., 1995):

$$n(z) =\propto z^2 \exp -(z/z_0)^\alpha, \tag{7.13}$$

Parameter	Prior	Fiducial value
Ω_c	$\mathcal{N}_T$ (0.2664, 0.2)	0.2664
Ω_b	$\mathcal{N}$ (0.0492, 0.006)	0.0492
σ_8	$\mathcal{N}$ (0.831, 0.14)	0.831
h	$\mathcal{N}$ (0.6727, 0.063)	0.6727
n_s	$\mathcal{N}$ (0.9645, 0.08)	0.9645
w_0	$\mathcal{N}_T$ (-1.0, 0.9)	-1.0

Table 7.1: Prior and fiducial values used for the analyses. The symbol $\mathcal{N}_T$ represents a Truncated Normal distribution. The lower bound of the support for the Ω_c distribution is set to zero, while the lower and upper bounds for the w_0 distribution are set to -2.0 and -0.3, respectively.

Redshift binning	5 bins
Redshift distribution (z_0, α)	(0.11, 0.68)
Number density n_s	27/arcmin^2
Shape noise σ_e	0.26
Redshift error σ_z	0.05(1+z)

Table 7.2: LSST Y10 source galaxy specifications in our analysis. All values are based on the LSST DESC SRD.

with $z_0 = 0.11$ and $\alpha = 0.68$. We also assume a photometric redshift error $\sigma_z = 0.05(1 + z)$ as defined in the LSST DESC Science Requirements Document (SRD, Mandelbaum et al. (2018)). The galaxy sources are divided into 5 tomographic bins, each containing an equal number of galaxies. For each redshift bin, we assume Gaussian noise with mean zero and variance given by:

$$\sigma_n^2 = \frac{\sigma_e^2}{A_{pix} n_{gal}}, \tag{7.14}$$

where we set the shape noise $\sigma_e = 0.26$ and the galaxy number density $n_{gal} = 27$ arcmin^{-2}. Both the shape noise and galaxy number density values are obtained from SRD. The pixel area is given by $A_{pix} \approx 5.49$ arcmin^2. Figure 7.1 illustrates the resulting source redshift distribution, and Table 7.2 provides a summary of the survey specifications.

7.3 Experiment

In the following section, we illustrate the inference strategy for the two approaches under investigation: the Bayesian forward modeling and the map-based inference based

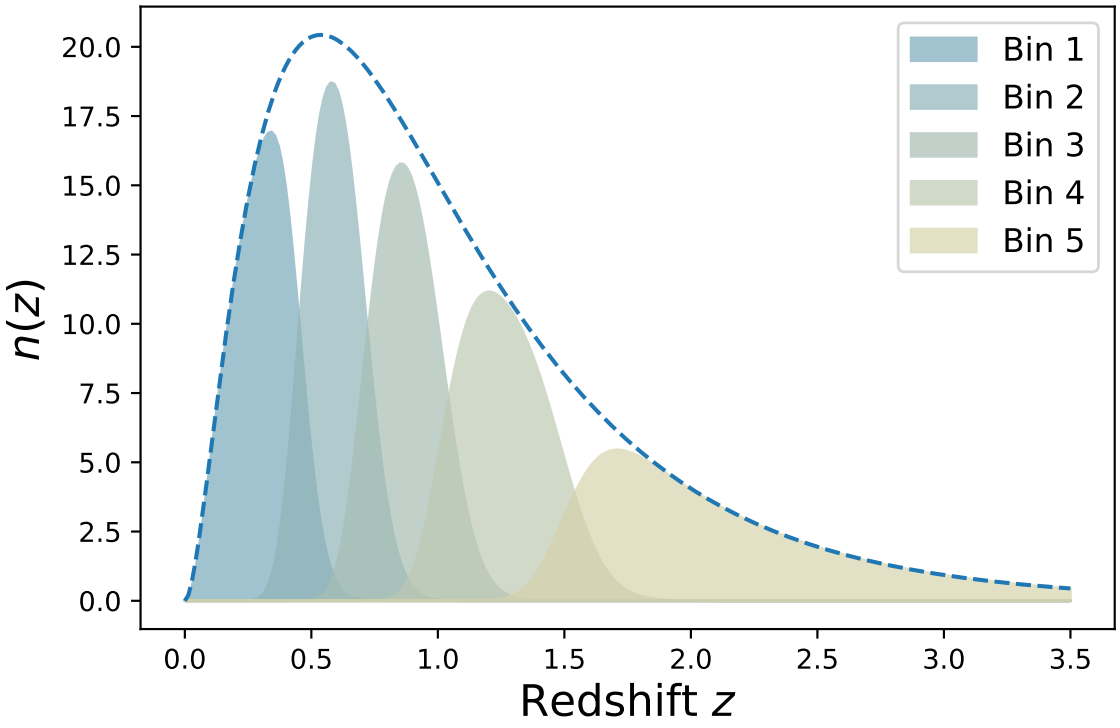

Figure 7.1: Source sample redshift distributions for each tomographic bin for LSST Y10. The number density on the y-axis is shown in arcmin^2.

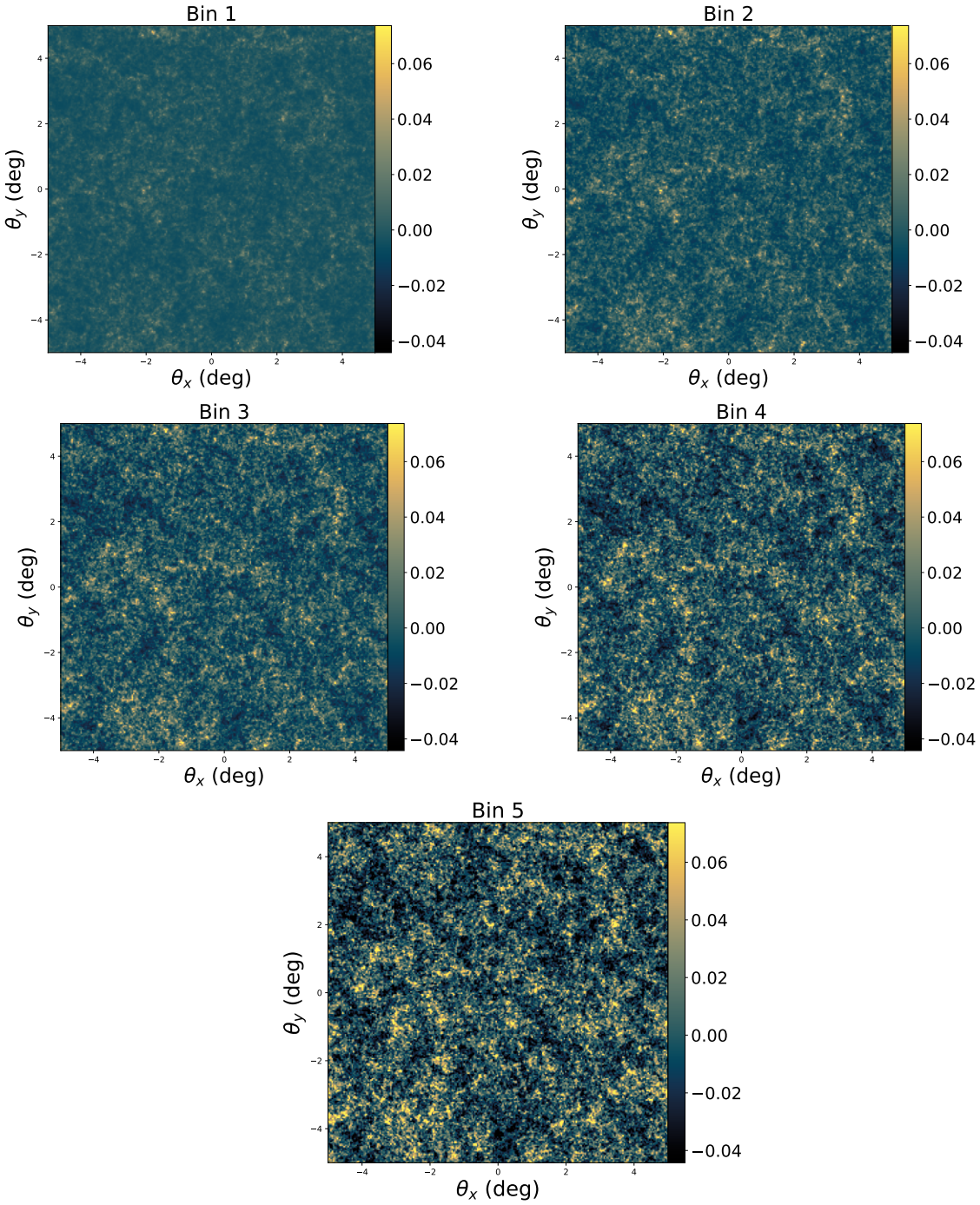

Figure 7.2: Example of convergence map simulated using the `SBILens` package.

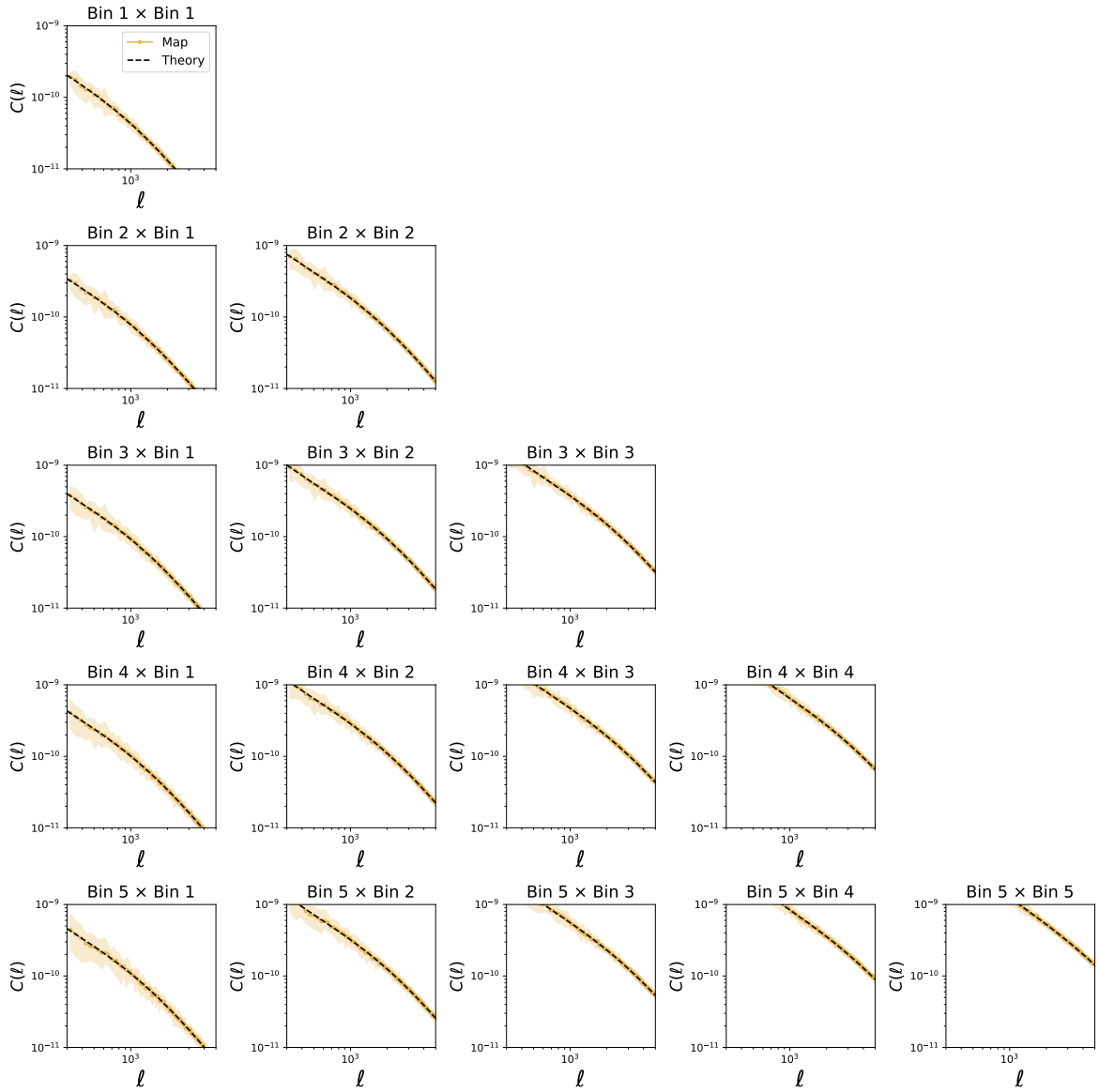

Figure 7.3: Convergence power spectra for different tomographic bin combinations. The solid yellow line shows the measurement from 20 simulated maps using the survey setting described in section 7.2, while the black dashed line shows the theoretical predictions computed using jax-cosmo. In this figure, the shaded regions represent the standard deviation from 20 independent map realizations.

on the LFI. Additionally, we conduct a power spectrum study. Indeed, as discussed in subsection 7.2.1, log-normal fields offer the advantage of rapidly generating convergent fields while accounting for non-Gaussianities. To emphasize this claim further, along with the full-field analysis, we include a power spectrum analysis. This analysis demonstrates that there is indeed a gain of information when adopting a full-field approach. Finally, we provide a detailed overview of various neural compression strategies that differ in terms of the loss functions used to train the neural network.

7.3.1 Explicit Inference

Full field with BHMs

To construct the explicit map-based inference strategy in the Hierarchical Bayesian framework, we built a likelihood based on the data model described in subsection 7.2.1. As mentioned before, this means that the simulator serves as the physical model capable of generating the non-linear representation of the convergence map.

However, in practical terms, the measurement of convergence for each pixel and bin will differ from real observations due to noise. This is taken into consideration in the likelihood. Specifically, for LSST Y10, the number of galaxies for each pixel should be sufficiently high so that, according to the central limit theorem, we can assume the observation is characterized by Gaussian noise, with $\sigma_n^2 = \sigma_e^2/N_s$, where N_s represents the total number of source galaxies per bin and pixel. Given σ_n^2 the variance of this Gaussian likelihood, its negative log-form can be expressed as:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_i^{N_{pix}} \sum_j^{N_{bins}} \log P(\kappa_{i,j}^{obs} | \kappa_{i,j}, \boldsymbol{\theta}) = -\sum_i^{N_{pix}} \sum_j^{N_{bins}} \frac{[\kappa_{i,j} - \kappa_{i,j}^{obs}]^2}{2\sigma_n^2}, \tag{7.15}$$

where κ^{obs} refers to the values of convergence from noise maps.
Since the full-field approach does not rely on any summary statistics, it typically leads to a high-dimensional problem, requiring more sophisticated statistical techniques. To sample the posterior distribution for $\boldsymbol{\theta}$, we use a Hamiltonian Monte Carlo (HMC) algorithm. Specifically, we employ the `NUTS` algorithm, an adaptive variant of HMC implemented in `NumPyro` (Phan et al., 2019; Bingham et al., 2019), which uses the No U-Turn Sampler (NUTS, Hoffman et al. (2014)).
As discussed in subsection 4.1.3, the HMC algorithm is particularly helpful in high-dimensional spaces where a large number of steps are required to effectively explore the space. It improves the sampling process by leveraging the information contained in the gradients to guide the sampling process. As the code is implemented in Jax, the gradients are accessible via automatic differentiation.

Power spectrum

To obtain the probability distribution of the cosmological parameters using a summary-statistics-methods, we focus on the 2-point statistics, specifically, on the angular power

spectra C_ℓ. We assume a Gaussian likelihood with a cosmology-independent covariance matrix:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}[\boldsymbol{d} - \boldsymbol{\mu}(\boldsymbol{\theta})]^T \boldsymbol{C}^{-1}[\boldsymbol{d} - \boldsymbol{\mu}(\boldsymbol{\theta})]. \tag{7.16}$$

To compute the expected theoretical predictions $\boldsymbol{\mu}(\boldsymbol{\theta})$ we use the public library `jax-cosmo` (Campagne et al., 2023). The Covariance matrix $\boldsymbol{C}$ of the observables is computed at the fiducial cosmology, presented in Table 7.1, using the same theoretical library. Specifically, in `jax-cosmo`, the Gaussian covariance matrix is defined as:

$$\mathrm{Cov}(C_\ell, C_{\ell'}) = \frac{1}{f_{sky}(2\ell+1)}\left(C_\ell + \frac{\sigma_\epsilon^2}{2n_s}\right)\delta^K(\ell - \ell') \tag{7.17}$$

where f_{sky} is the fraction of sky observed by the survey, and n_s is the number density of galaxies. To obtain the data vector $\boldsymbol{d}$, containing the auto- and the cross-power spectra for each tomographic bin, we use the `LensTools` package (Petri, 2016) on a single noisy simulated map (our fiducial). To constrain the cosmological parameters, we sample from the posterior distribution using an HMC algorithm with adaptive step size, implemented in the TensorFlow Probability MCMC python package.

7.3.2 Implicit Inference

Benchmark compression scheme

To ensure the scalability of density estimation LFI in cases where forward simulations are computationally expensive, it becomes necessary to employ compression techniques that reduce the dimensionality of the data space and extract summary statistics. Specifically, we try to find a function $\boldsymbol{t} = F(\boldsymbol{x})$, where $\boldsymbol{t}$ represents low-dimensional summaries of the original data vector $\boldsymbol{x}$. The objective is to achieve a compression function $F(\boldsymbol{x})$ that maximizes the information while minimizing dimensionality. Previous studies (Heavens et al., 2000b) have demonstrated that a compression scheme can be achieved where the dimension of the summaries $\dim(\boldsymbol{t})$ is equal to the dimension of the unknown parameters $\dim(\boldsymbol{\theta})$ without any loss of information at the Fisher level. Multiple approaches exist in an attempt to satisfy this condition. This section aims to provide an overview of the various neural compression-based methods employed in previous works.

Mean Square Error (MSE) One of the commonly used techniques for training a Neural Network is by minimizing the L_2 norm or Mean Square Error (MSE). This methods has been widely adopted in various previous studies (Ribli et al., 2018; Lu et al., 2022, 2023), where the loss function is typically formulated as follows:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N_\theta}\sum_{i=1}^{N_\theta}(t_i - \theta_i)^2. \tag{7.18}$$

Here N_θ represents the number of cosmological parameters, t denotes the summary statistics, and θ corresponds to the cosmological parameters. However, it is important to note that this approach does not guarantee the recovery of maximally informative summary statistics. Indeed, minimizing the L_2 norm is equivalent to training the model to estimate the mean of the posterior distribution. Namely:

$$\langle \boldsymbol{\theta} \rangle_{p(\boldsymbol{\theta}|\boldsymbol{x})} = \underset{F(\boldsymbol{x})}{\operatorname{argmin}} \, \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\|\boldsymbol{\theta} - F(\boldsymbol{x})\|_2^2], \tag{7.19}$$

where the posterior mean $\langle \boldsymbol{\theta} \rangle_{p(\boldsymbol{\theta}|\boldsymbol{x})}$, is calculated as follows:

$$\langle \boldsymbol{\theta} \rangle_{p(\boldsymbol{\theta}|\boldsymbol{x})} = \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}]. \tag{7.20}$$

To demonstrate this statement, we need to minimize the expected value of the L_2 norm with respect to $F(\boldsymbol{x})$. Let us consider its derivative:

$$\begin{aligned} &\frac{\partial}{\partial F(\boldsymbol{x})} \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x}))}[(\boldsymbol{\theta} - F(\boldsymbol{x}))^2] = \\ &\frac{\partial}{\partial F(\boldsymbol{x})} \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}^2 + F(\boldsymbol{x})^2 - 2\boldsymbol{\theta} F(\boldsymbol{x})] = \\ &\frac{\partial}{\partial F(\boldsymbol{x})} [\mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}^2] + F(\boldsymbol{x})^2 - 2F(\boldsymbol{x}) \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}]] = \\ &2F(\boldsymbol{x}) - 2\mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}]. \end{aligned} \tag{7.21}$$

Setting it equal to zero, we obtain the critical value:

$$F(\boldsymbol{x}) = \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}]. \tag{7.22}$$

Considering the second-order derivative:

$$\frac{\partial^2}{\partial^2 F(\boldsymbol{x})} \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[(\boldsymbol{\theta} - F(\boldsymbol{x}))^2] = 2 > 0, \tag{7.23}$$

we can assert that this critical value is also a minimum.
From Equation 7.22 and Equation 7.20, we obtain Equation 7.19.

Mean Absolute Error (MAE) Another commonly used approach involves minimizing the L_1 norm or Mean Absolute Error (MAE). In this approach, the loss function is defined as:

$$\mathcal{L}_{\text{MAE}} = \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} |t_i - \theta_i|. \tag{7.24}$$

where N_θ represents the number of cosmological parameters, t denotes the summary statistics, and θ corresponds to the cosmological parameters. We can demonstrate that

minimizing this loss function is equivalent to training the model to estimate the median of the posterior distribution $\boldsymbol{\theta}^{M}_{p(\boldsymbol{\theta}|\boldsymbol{x})}$. Namely:

$$\boldsymbol{\theta}^{M}_{p(\boldsymbol{\theta}|\boldsymbol{x})} = \underset{F(\boldsymbol{x})}{\operatorname{argmin}}\, \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[|\boldsymbol{\theta} - F(\boldsymbol{x})|]. \tag{7.25}$$

By definition, the median of a one-dimensional[1] probability density function $p(x)$ is a real number m that satisfies:

$$\int_{\infty}^{m} p(x)dx = \int_{m}^{\infty} p(x)dx = \frac{1}{2}. \tag{7.26}$$

The expectation value of the mean absolute error is defined as:

$$\mathbb{E}_{p(x)}[|x - m|] = \int_{\infty}^{\infty} p(x)|x - m|dx \tag{7.27}$$

which can be decomposed as

$$\int_{\infty}^{m} p(x)|x - m|dx + \int_{m}^{\infty} p(x)|x - m|dx. \tag{7.28}$$

To minimize this function with respect to m, we need to compute its derivative:

$$\frac{d\mathbb{E}[|x - m|]}{dm} = \frac{d}{dm}\int_{\infty}^{m} p(x)|x - m|dx + \frac{d}{dm}\int_{m}^{\infty} p(x)|x - m|dx. \tag{7.29}$$

Considering that $|x - m| = (x - m)$ for $m \leq x$ and $|x - m| = (m - x)$ $m \geq x$, we can write Equation 7.29 as:

$$\frac{d\mathbb{E}[|x - m|]}{dm} = \frac{d}{dm}\int_{\infty}^{m} p(x)(m - x)dx + \frac{d}{dm}\int_{m}^{\infty} p(x)(x - m)dx. \tag{7.30}$$

Using the Leibniz integral rule, we get:

$$\begin{aligned} &\frac{d\mathbb{E}[|x - m|]}{dm} = \\ &p(x)(m - m)\frac{dm}{dm} + \int_{\infty}^{m} \frac{\partial}{\partial m}[p(x)(m - x)]dx + \\ &- p(x)(m - m)\frac{dm}{dm} + \int_{m}^{\infty} \frac{\partial}{\partial m}[p(x)(m - x)]dx. \end{aligned} \tag{7.31}$$

Setting the derivative to zero, we obtain:

$$\frac{d\mathbb{E}[|x - m|]}{dm} = \int_{\infty}^{m} p(x)dx - \int_{m}^{\infty} p(x)dx = 0. \tag{7.32}$$

[1]For notational simplicity, we demonstrate this statement in one dimension; the generalization to N-dimensions is straightforward.

Thus,

$$\int_{\infty}^{m} p(x)dx = \int_{m}^{\infty} p(x)dx. \tag{7.33}$$

Considering that

$$\int_{\infty}^{m} p(x)dx + \int_{m}^{\infty} p(x)dx = 1, \tag{7.34}$$

we obtain Equation 7.26.

While extensively employed in various previous studies (Gupta et al., 2018; Fluri et al., 2018b; Ribli et al., 2019), similar to the previous consideration for the MSE, it is important to note that the effectiveness of this loss function in extracting sufficient statistics needs to be demonstrated.

Variational Mutual Information Maximization (VMIM) The Variational Mutual Information Maximization (VMIM) technique was first introduced for cosmological inference problems by Jeffrey et al. (2021). This approach aims to maximize the mutual information $I(\boldsymbol{t}, \boldsymbol{\theta})$ between the cosmological parameters $\boldsymbol{\theta}$ and the summary statistics $\boldsymbol{t}$. In the VMIM approach, the loss function is defined as:

$$\mathcal{L}_{\text{VMIM}} = -\log q(\boldsymbol{\theta}|F_{\boldsymbol{\varphi}}(\boldsymbol{x}); \boldsymbol{\varphi'}). \tag{7.35}$$

Here, $q(\boldsymbol{\theta}|F_{\boldsymbol{\varphi}}(\boldsymbol{x}); \boldsymbol{\varphi'})$ represents a variational conditional distribution, where $\boldsymbol{\theta}$ corresponds to the data vector of the cosmological parameters, and $\boldsymbol{\varphi'}$ to the parameters characterizing the variational conditional distribution itself. $F_{\boldsymbol{\varphi}}$ denotes the compression network of parameter $\boldsymbol{\varphi}$, used to extract the summary statistics $\boldsymbol{t}$ from the original high-dimensional data vector $\boldsymbol{x}$, such that $\boldsymbol{t} = F_{\boldsymbol{\varphi}}(\boldsymbol{x})$. In order to understand the significance of this loss function, it is necessary to start by considering the mathematical definition of mutual information $I(\boldsymbol{t}, \boldsymbol{\theta})$:

$$\begin{aligned}
I(\boldsymbol{t}, \boldsymbol{\theta}) &= D_{KL}(p(\boldsymbol{t}, \boldsymbol{\theta})||p(\boldsymbol{t})p(\boldsymbol{\theta})) \\
&= \int d\boldsymbol{\theta} d\boldsymbol{t} p(\boldsymbol{t}, \boldsymbol{\theta}) \log\left(\frac{p(\boldsymbol{t}, \boldsymbol{\theta})}{p(\boldsymbol{t})p(\boldsymbol{\theta})}\right) \\
&= \int d\boldsymbol{\theta} d\boldsymbol{t} p(\boldsymbol{t}, \boldsymbol{\theta}) \log\left(\frac{p(\boldsymbol{\theta}|\boldsymbol{t})}{p(\boldsymbol{\theta})}\right) \\
&= \int d\boldsymbol{\theta} d\boldsymbol{t} p(\boldsymbol{t}, \boldsymbol{\theta}) \log p(\boldsymbol{\theta}|\boldsymbol{t}) - \int d\boldsymbol{\theta} d\boldsymbol{t} p(\boldsymbol{t}, \boldsymbol{\theta}) \log p(\boldsymbol{\theta}) \\
&= \int d\boldsymbol{\theta} d\boldsymbol{t} p(\boldsymbol{t}, \boldsymbol{\theta}) \log p(\boldsymbol{\theta}|\boldsymbol{t}) - \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) \\
&= \mathbb{E}_{p(\boldsymbol{t},\boldsymbol{\theta})}[\log p(\boldsymbol{\theta}|\boldsymbol{t})] - \mathbb{E}_{p(\boldsymbol{\theta})}[\log p(\boldsymbol{\theta})] \\
&= \mathbb{E}_{p(\boldsymbol{t},\boldsymbol{\theta})}[\log p(\boldsymbol{\theta}|\boldsymbol{t})] - H(\boldsymbol{\theta});
\end{aligned} \tag{7.36}$$

in the above equation, D_{KL} is the Kullback-Leibler divergence (Kullback and Leibler, 1951), $p(\boldsymbol{t}, \boldsymbol{\theta})$ is the joint probability distribution of summary statistics and cosmological parameters, and $H(\boldsymbol{\theta})$ represents the *entropy* of the distribution of cosmological

parameters. Essentially, mutual information measures the amount of information contained in the summary statistics $\boldsymbol{t}$ about the cosmological parameters $\boldsymbol{\theta}$. The goal is to find the parameters of the network $\boldsymbol{\varphi}$ that maximize the mutual information between the summary and cosmological parameters:

$$\boldsymbol{\varphi}^* = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}}\, I(F_{\boldsymbol{\varphi}}(\boldsymbol{x}), \boldsymbol{\theta}). \tag{7.37}$$

However, the mutual information expressed in Equation 7.36 is not tractable. To overcome this limitation, various approaches have been developed that rely on tractable bounds, enabling the training of deep neural networks to optimize mutual information. In this study, we adopt the same strategy used by Jeffrey et al. (2021), which involves using the variational lower bound (Barber and Agakov, 2003):

$$I(\boldsymbol{t}, \boldsymbol{\theta}) \geq \mathbb{E}_{p(\boldsymbol{t},\boldsymbol{\theta})}[\log q(\boldsymbol{\theta}|\boldsymbol{t}; \boldsymbol{\varphi}')] - H(\boldsymbol{\theta}). \tag{7.38}$$

Here, the variational conditional distribution $\log q(\boldsymbol{\theta}|\boldsymbol{t}; \boldsymbol{\varphi}')$ is introduced to approximate the true posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{t})$. As the entropy of the cosmological parameters remains constant for a fixed training set, the optimization problem based on the lower bound in Equation 7.38 can be formulated as:

$$\underset{\boldsymbol{\varphi},\boldsymbol{\varphi}'}{\operatorname{argmax}}\, \mathbb{E}_{p(\boldsymbol{x},\boldsymbol{\theta})}[\log q(\boldsymbol{\theta}|F_{\boldsymbol{\varphi}}(\boldsymbol{x}); \boldsymbol{\varphi}')], \tag{7.39}$$

yielding Equation 7.35.

Gaussian Negative Log-Likelihood (GNLL) Recognizing that the uncertainty on different cosmological parameters will vary, a third class of inverse variance weighted MSE was proposed in Fluri et al. (2018b) with the idea of ensuring that each parameter contributes fairly to the overall loss by taking into account its uncertainty. The loss function typically takes the following form:

$$\mathcal{L}_{\text{GNLL}} = \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \frac{1}{2}(\boldsymbol{t} - \boldsymbol{\theta})^\top \Sigma^{-1}(\boldsymbol{t} - \boldsymbol{\theta}) \tag{7.40}$$

where $\boldsymbol{t}$ is the summary statistics and $\boldsymbol{\Sigma}$ is the covariance matrix representing the uncertainty on the cosmological parameters $\boldsymbol{\theta}$. Both $\boldsymbol{t}$ and $\boldsymbol{\Sigma}$ can be outputs of the compression network, i.e. $f_{\boldsymbol{\varphi}}(\boldsymbol{x}) = (\boldsymbol{t}, \boldsymbol{\Sigma})$.

One recognizes here the expression of a Gaussian probability function, and this expression can thus be straightforwardly related to the VMIM case by simply assuming a Gaussian distribution as the variational approximation for the posterior $q(\boldsymbol{\theta}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{t}, \boldsymbol{\Sigma})$.

This means that for this loss function:

- The summary $\boldsymbol{t}$ extracted by the neural network is, similarly to the MSE case, only an estimate of the mean of the posterior distribution, which is not guaranteed to be sufficient.

- Because of the Gaussian variational assumption, the variational posterior can potentially be biased with respect to the true posterior, and thus the mean of this Gaussian approximation may be biased with respect to the true posterior mean.

Information Maximizing Neural Networks A different approach has been proposed by Charnock et al. (2018) and further explored in Makinen et al. (2021, 2022). They implemented the Information Maximizing Neural Networks (IMNN), a neural network trained on forward simulations designed to learn optimal compression schemes, even in circumstances where the likelihood function is intractable or unknown. Specifically, they proposed a new scheme to find optimal non-linear data summaries by using the Fisher information to train a neural network. Inspired by the MOPED algorithm (introduced in paragraph 4.3), they aimed to find a transformation f that maps the data to compressed summaries: $f : \mathbf{x} \rightarrow \mathbf{t}$ while conserving the Fisher information. The loss function takes the following form:

$$\mathcal{L}_{\text{IMNN}} = -|\det(\mathbf{F})| + r_{\Sigma}, \tag{7.41}$$

where $\mathbf{F}$ is the Fisher matrix, and r_{Σ} is a regularization term typically dependent on the covariance matrix, introduced to control the magnitude of the summaries. Since computing the Fisher matrix requires a large number of simulations, they proceed as follows: a large number of simulations with the same fiducial cosmology but different initial random conditions are fed forwards through identical networks. The summaries from these simulations are combined to compute the covariance matrix. Additionally, the summaries from simulations created with different fixed cosmologies are used to calculate the derivative of the mean of the summary with respect to the parameter.[2] Finally, the covariance and the mean derivatives are combined to obtain the Fisher matrix.

An application of this method in the context of weak lensing can be traced back to the work of Fluri et al. (2021, 2022). Following Charnock et al. (2018), they presented a compression approach that relies on the optimization of the Cramér-Rao bound. However, their approach does not assume that the summary follows a Gaussian likelihood. The loss function takes the form:

$$\mathcal{L}_{\text{IMNN}} = \log\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{t})) - 2\log\left|\det\left(\frac{\partial\Psi_{\boldsymbol{\theta}}(\boldsymbol{t})}{\partial\boldsymbol{\theta}}\right)\right|, \tag{7.42}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix, and $\Psi_{\boldsymbol{\theta}}(\boldsymbol{t}) = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{\theta})}[\boldsymbol{t}]$. This loss function is equivalent to the log-determinant of the inverse Fisher matrix used in Equation 7.41.

[2] The method of finite differences is necessary when the framework in which the code is implemented does not support automatic differentiation.

Compression strategy

The different compression strategies differ in the loss function employed during the training phase. For all of them, we use the same convolutional neural network architecture to build the compressor, specifically the ResNet-18, an 18-layer ResNet architecture (He et al., 2016). The ResNet-18 is implemented using Haiku (Hennigan et al., 2020), a Python deep learning library built on top of Jax.

While the different compression strategies share the same architecture, the training strategies for VMIM differ significantly. To train the Neural compressor under VMIM, we jointly optimize the weights $\boldsymbol{\varphi}$ of the Neural Network $F_{\boldsymbol{\varphi}}$ (the compressor) and the parameters $\boldsymbol{\varphi}'$ of the variational distribution $q(\boldsymbol{\theta}|\boldsymbol{t};\boldsymbol{\varphi}')$ in Equation 7.35. The variational distribution is modeled using a Normalizing Flow. The Normalizing Flow used for the inference strategy is also adopted for the variational distribution required in the compression procedure. After training, we export the results of the neural compression $F_{\boldsymbol{\varphi}}$ but discard the results from the density estimator. Indeed, as mentioned before, we perform the Likelihood-Free Inference (LFI) as a two-step procedure. This choice is motivated by the fact that it is difficult to train a very precise conditional density estimator when the compressor can still change from iteration to iteration. Hence, the choice to split the problem into two steps: first, the dimensionality reduction, where the density estimation part does not need to be perfect; second, the density estimation itself, which needs to be done very carefully now, but it is much easier because we are in low dimension.

Inference strategy

By comparing the posteriors $p(\boldsymbol{\theta}|\boldsymbol{t})$ obtained with different compression strategies, we can assess the sensitivity of the results to the choice of the compression method. In particular, the more informative the summary statistics, the closer the posterior $p(\boldsymbol{\theta}|\boldsymbol{t})$ is to the true posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$.
As we have mentioned before, in the LFI context, the simulator is a black box that defines the likelihood distribution $p(\boldsymbol{x}|\boldsymbol{\theta})$ as an implicit distribution. We can sample from it, but we cannot directly evaluate it. The idea of Neural density Estimation is to introduce a parametric distribution model $\mathcal{P}_{\boldsymbol{\varphi}'}$ to approximate the implicit distribution $\mathcal{P}$.
In particular, we focus on Neural Posterior Estimation (Papamakarios and Murray, 2018; Lueckmann et al., 2017b; Greenberg et al., 2019b), i.e., we aim to directly approximate the posterior distribution. Specifically, we use a conditional Normalizing Flows (Rezende and Mohamed, 2016; Papamakarios et al., 2021) to model the parametric conditional distribution $q_{\boldsymbol{\varphi}'}(\boldsymbol{\theta}|\boldsymbol{t})$, and we optimize the parameters $\boldsymbol{\varphi}'$ according to the following negative log-likelihood (NLL) loss function:

$$\mathcal{L} = -\log q_{\boldsymbol{\varphi}'}(\boldsymbol{\theta}|\boldsymbol{t}). \tag{7.43}$$

Reference	Loss function	Inference strategy
Gupta et al. (2018)	MAE	Likelihood-based analysis
Fluri et al. (2018b)	GNLL	Likelihood-based analysis
Fluri et al. (2019)	GNLL	Likelihood-based analysis
Ribli et al. (2019)	MAE	Likelihood-based analysis
Matilla et al. (2020)	MAE	Likelihood-based analysis
Jeffrey et al. (2021)	MSE VMIM	Likelihood Free Inference (Py-Delfi)
Fluri et al. (2021)	IMNN	Likelihood Free Inference (GPABC)
Fluri et al. (2022)	IMNN	Likelihood Free Inference (GPABC)
Lu et al. (2022)	MSE	Likelihood-based analysis
Kacprzak and Fluri (2022)	GNLL	Likelihood-based analysis
Lu et al. (2023)	MSE	Likelihood-based analysis

Table 7.3: Table summarizing the different neural compression schemes used for weak lensing applications. The table includes all studies conducted within the context of implicit analysis (Likelihood-free) and standard Likelihood-based analysis.
Abbreviations used in the Table: MSE-Mean Square Error; MAE-Mean Absolute Error; GNLL- Gaussian Negative Log Likelihood; VMIM- Variational Mutual Information Maximization; IMNN- Information Maximizing Neural Networks; GPABC-Gaussian Processes Approximate Bayesian Computation.

In the limit of a large number of samples and sufficient flexibility, we obtain:

$$q_{\boldsymbol{\varphi}'^*}(\boldsymbol{\theta}|\boldsymbol{t}) \approx p(\boldsymbol{\theta}|\boldsymbol{t}) \tag{7.44}$$

where we indicate with $\boldsymbol{\varphi}'^*$ the values of $\boldsymbol{\varphi}'$ minimizing Equation 7.43.
Finally, the target posterior $p(\boldsymbol{\theta}|\boldsymbol{t} = \boldsymbol{t_0})$ is approximated by $q_{\boldsymbol{\varphi}'^*}(\boldsymbol{\theta}|\boldsymbol{t} = \boldsymbol{t_0})$, where $\boldsymbol{t_0} = F_{\boldsymbol{\varphi}}(\boldsymbol{x_0})$, i.e., the compressed statistics from the fiducial convergence map $\boldsymbol{x_0}$ for a given compression strategy.

For each summary statistic, we approximate the posterior distribution using the same NF architecture, namely a RealNVP (Dinh et al., 2017) with 4 coupling layers. The shift and the scale parameters are learned using a neural network with 2 layers of 128 neurons with Sigmoid-weighted Linear Units (SiLU) activation functions (Elfwing et al., 2017).

7.4 First results and discussion

We present the preliminary results of the constraints on the full ΛCDM parameter space expected for a survey like LSST-Y10. The results are obtained using the simulation procedure outlined in section 7.2 and the parameter inference strategy described in subsubsection 7.3.2. We begin by analyzing the impact of the different compression strategies on the final cosmological constraints. Subsequently, we compare the outcomes of three different inference procedures: the 2-point statistics, the explicit full-field statistics, and the implicit full-field statistics using CNN summaries.

7.4.1 Power spectrum and full-field statistics

We now compare the constraining power of the three different approaches described in section 7.3: the standard-two point statistics and two map-based approaches, the explicit inference and the implicit inference strategy. As outlined before, our interest is to prove that the two map-based approaches lead to very comparable posterior distributions. We present the 68.3% and 95.5% confident regions from one fiducial map for the full ΛCDM parameters in Figure 7.4. The contours obtained by the angular C_ℓ analysis are plotted in violet, the ones for the explicit inference (HMC) in orange, and those for the implicit inference (Normalizing Flow) in blue. We can see from the figure that the contours from the HMC-based inference and the Normalizing flow-based inference are remarkably similar. We quantify the results by computing the Figure of Merit defined as follows:

$$\mathrm{FoM}_{\alpha\beta} = \sqrt{\det(\tilde{F}_{\alpha\beta})}. \tag{7.45}$$

Here, α and β represent a pair of cosmological parameters, and $\tilde{F}_{\alpha\beta}$ refers to the marginalized Fisher matrix. We calculate $\tilde{F}_{\alpha\beta}$ as the inverse of the parameter space covariance matrix $C_{\alpha\beta}$, which is estimated from the Normalizing flow sampling. Under the assumption of a Gaussian covariance, the FoM defined in Equation 7.45 is proportional to the inverse of the $2-\sigma$ contours in the 2-dimensional marginalized parameter space of the α and β pair. The results are presented in Table 7.4. The remarkably strong agreement between the two posteriors confirms that the two map-based cosmology inference methods yield the same results. The ratio of their FoM, corresponding to 1.00, 1.04, 1.03, in the $(\Omega_c - \sigma_8); (\Omega_c - w_0); (\sigma_8 - w_0)$ plane, establishes the validity of the implicit inference strategy.
Confirming previous work, we note that h, Ω_b, and n_s result prior dominated and hence are constrained by none of the three approaches. Additionally, for the full-field strategies we find that the size of the contours is significantly smaller than the size for the prior distributions adopted, suggesting that the posterior on these parameters are not affected by the specific choice of prior. Moreover, we find that the two-point statistic is suboptimal in constraining Ω_c, σ_8, and w_0, while the two map-based approaches yield

much tighter constraints on these parameters. We find that the map-based explicit and implicit strategies lead to an improvement in the figure of merit of 2.06×, 1.98×, 2.28× and 2.06×, 1.90×, 2.22×, respectively, in the $(\Omega_c - \sigma_8); (\Omega_c - w_0); (\sigma_8 - w_0)$ plane.

7.4.2 Optimal compression strategy

Figure 7.5 shows our 68.3% and 95.5% constraints from one fiducial map using different compressed summaries. We compare the constraints obtained from MSE (black contours), MAE (magenta contours), and VMIM (blue contours). We note that the results obtained using different summaries are generally in agreement with each other, and there are no tensions present for any of the cosmological parameters. We report the marginalized summary constraints in Table 7.5. The results concern the cosmological parameters that are better constrained from weak lensing: Ω_c, σ_8, w_0. We note that the VMIM compressed summary statistics prefer values of Ω_c, σ_8, and w_0 that are closer to our fiducial cosmology than those inferred by MSE and MAE.
To further quantify these outcomes, we consider the Figure of Merit described in Equation 7.45. The results are presented in Table 7.4. We can see that VMIM yields more precise measurements than MSE and MAE for all considered parameters. In particular, the FoM of (Ω_c, σ_8) is improved by 1.24× and 1.1× compared to MSE and MAE, respectively; the FoM of (Ω_c, w_0) is improved by 1.25× and 1.17× from the MSE and MAE; the FoM of (σ_8, w_0) is improved by 1.23× and 1.15× from the MSE and MAE.

7.5 Summary, ongoing work and future prospects

In this chapter, I presented ongoing work focused on comparing two different map-based strategies for inferring cosmological parameters: the explicit full-field strategy, also known as Bayesian Hierarchical inference, based on an HMC sampler, and the implicit inference strategy, also known as Likelihood-Free Inference, based on a neural density estimator.

We start with a consideration: Deep learning approaches for implicit inference typically involve two steps. The first step is the automatic learning of an optimal low-dimensional summary statistic, and the second step is the use of a Neural Density Estimator in low dimensions to either build an estimate P_{φ} of the likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta})$ (Neural Likelihood Estimation) or build an estimate P_{φ} of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$ (Neural Posterior Estimation).
Now, one can understand that both of these steps may potentially impact the final constraints on the parameters of interest.
Having said that, the main motivation for this work is to evaluate the impact of a given compression strategy on the final posterior distribution and, consequently, determine whether an optimal compression strategy exists. Furthermore, the aim is to demonstrate that by using this strategy, both implicit and explicit methods yield the same posterior.

To construct the forward model for explicit inference and simulate the mock data required to train the implicit model, we developed `SBILens`, a Jax-based weak lensing simulator optimized for inference applications that need access to the model's derivatives. Our analysis is based on synthetic weak lensing data with five tomographic bins, mimicking a survey like LSST-Y10.

After providing an overview of the different compression strategies adopted in the literature for both Likelihood-Free Inference and likelihood-based inference strategies, we compared the impact of some of those on the final constraints on the cosmological parameters for a ΛCDM model. We found the following results:

1. The marginalized summary statistics indicate that VMIM produces better results for Ω_c, w_0, and σ_8 in terms of agreement with the fiducial value. However, it is important to note that the results from MSE and MAE are not in tension with the fiducial parameters. Furthermore, we quantified the outcomes by examining the figure of merit and found that VMIM provides more precise measurements compared to MSE and MAE.

2. When using the VMIM to compress the original high-dimensional data, we compared the posterior obtained in the implicit inference framework with those obtained from Bayesian hierarchical modeling and the power spectrum. We demonstrate that both map-based approaches lead to a significant improvement in constraining Ω_c, w_0, σ_8 compared to the 2-point statistics. However, h, n_s, Ω_b are not constrained by either and are prior-dominated.

3. When using the VMIM to compress the original high-dimensional data the two methods, i.e., Bayesian hierarchical inference and Likelihood-free inference, lead to the same posterior distributions.

The results shown in this section are very preliminary, especially those related to the impact of different compression schemes. In practice, the next steps of this project involve evaluating the impact of two other loss functions: the Gaussian negative log-likelihood, as defined in paragraph 7.3.2, and the Information Maximizing Neural Networks (IMNN), discussed in paragraph 7.3.2.

I think it is interesting even at this preliminary stage to consider the potential limitations of the current implementation and highlight particular strategies for future extensions and applications of this project. In this work, we employed a physical model based on a lognormal prior, which is notably faster than simulation-based methods. Although we have shown that this description accounts for additional non-Gaussian information, as evidenced by the fact that we obtain different posteriors from the full-field and power spectrum methods, it is important to note that this is a good approximation for the convergence at intermediate scales but may not be appropriate for analyzing small scales.

FoM	C_ℓ	Full Field (HMC)	VMIM	MSE	MAE
$\Omega_c - \sigma_8$	1222	2520	2526	2043	2316
$\Omega_c - w_0$	100	198	190	152	162
$\sigma_8 - w_0$	77	176	171	139	149

Table 7.4: Figure of Merit (FoM) for different inference strategies: the convergence power spectrum C_ℓ, the HMC, the CNN map compressed statistics with the MSE, the VMIM, and the MAE loss functions. The values of the figure of merit are inversely proportional to the area of the contours; the larger the FoM, the higher the constraining power.

Furthermore, the lognormal shift parameters are computed using the `Cosmomentum` code (Friedrich et al., 2018, 2020), which employs perturbation theory. However, as mentioned by Boruah et al. (2022a), the perturbation theory-based approach may not provide accurate results at small scales.[3]

Additionally, we did not include any systemic in the current application, although previous studies demonstrated that the map-based approaches help to dramatically improve the constraints on systematic and cosmological parameters in the presence of these last. The main reason for this absence is mainly related to the difficulty of modeling systematic effects, like for example intrinsic alignment, in the lognormal description.

Hence, the natural next step will be to implement an N-body model as the physical model for the `SBILens` package. Specifically, we aim to work with the differentiable simulation we have presented in this thesis. As mentioned in section 6.4 of chapter 5, the simulations can be improved and further developed by including additional systematics such as redshift uncertainties, baryonic feedback, and a more complicated intrinsic alignment model. With these improvements, our model will be better suited to handle real cosmic shear data, allowing us to fully maximize the information gained from next-generation surveys.

[3]However, as the main objective of this project is to compare the different inference strategies, we are not very concerned with the potential implications of this approximation at this stage.

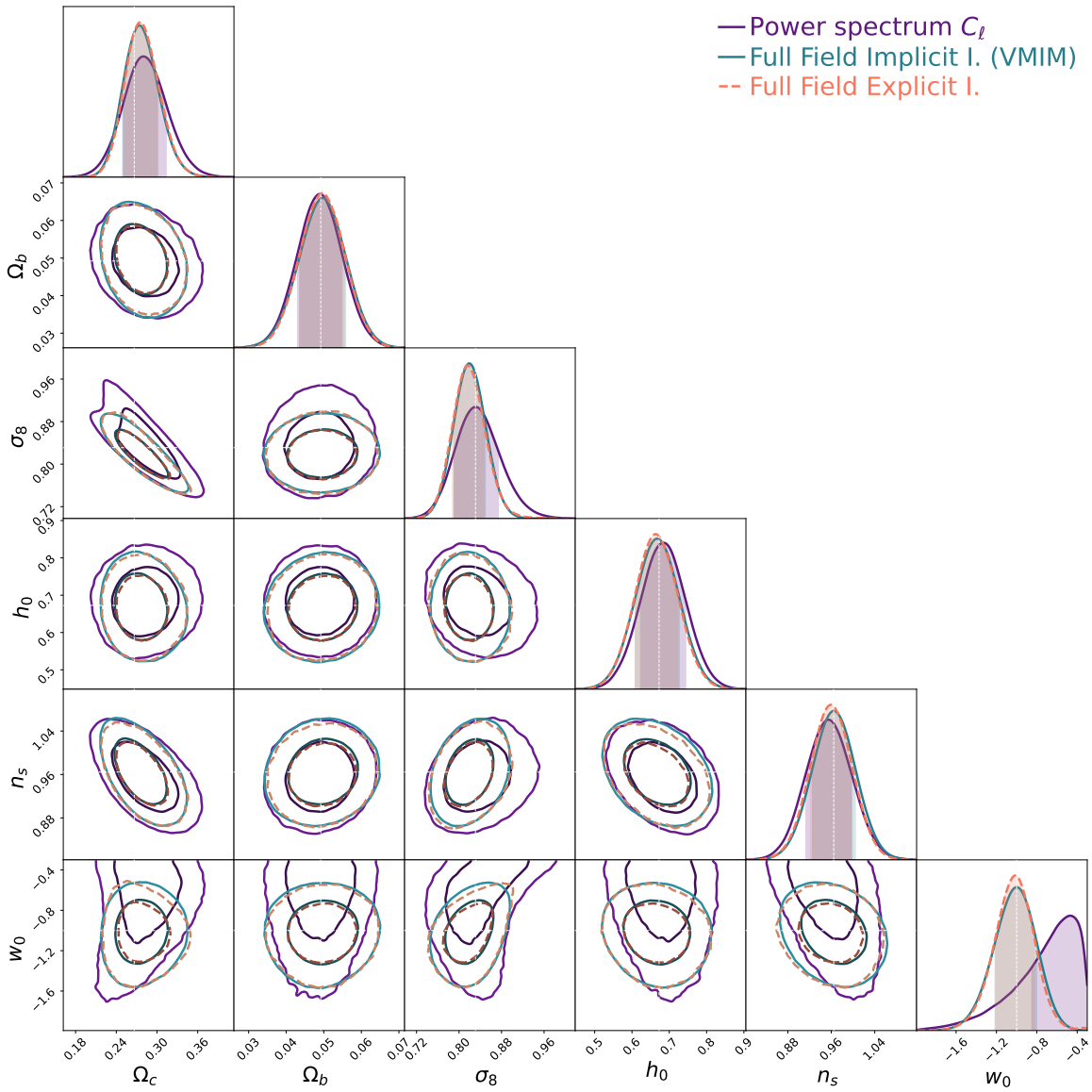

Figure 7.4: Constraints on the ΛCDM parameter space as found in the LSST Y10 survey setup. The constraints are obtained by applying the C_ℓ (violet contours), the full field explicit inference (orange contours), and the full field implicit inference strategy (blue dashed contours), described in section 7.3. The contours show the 68% and the 95% confidence regions. The dashed white lines define the true parameter values.

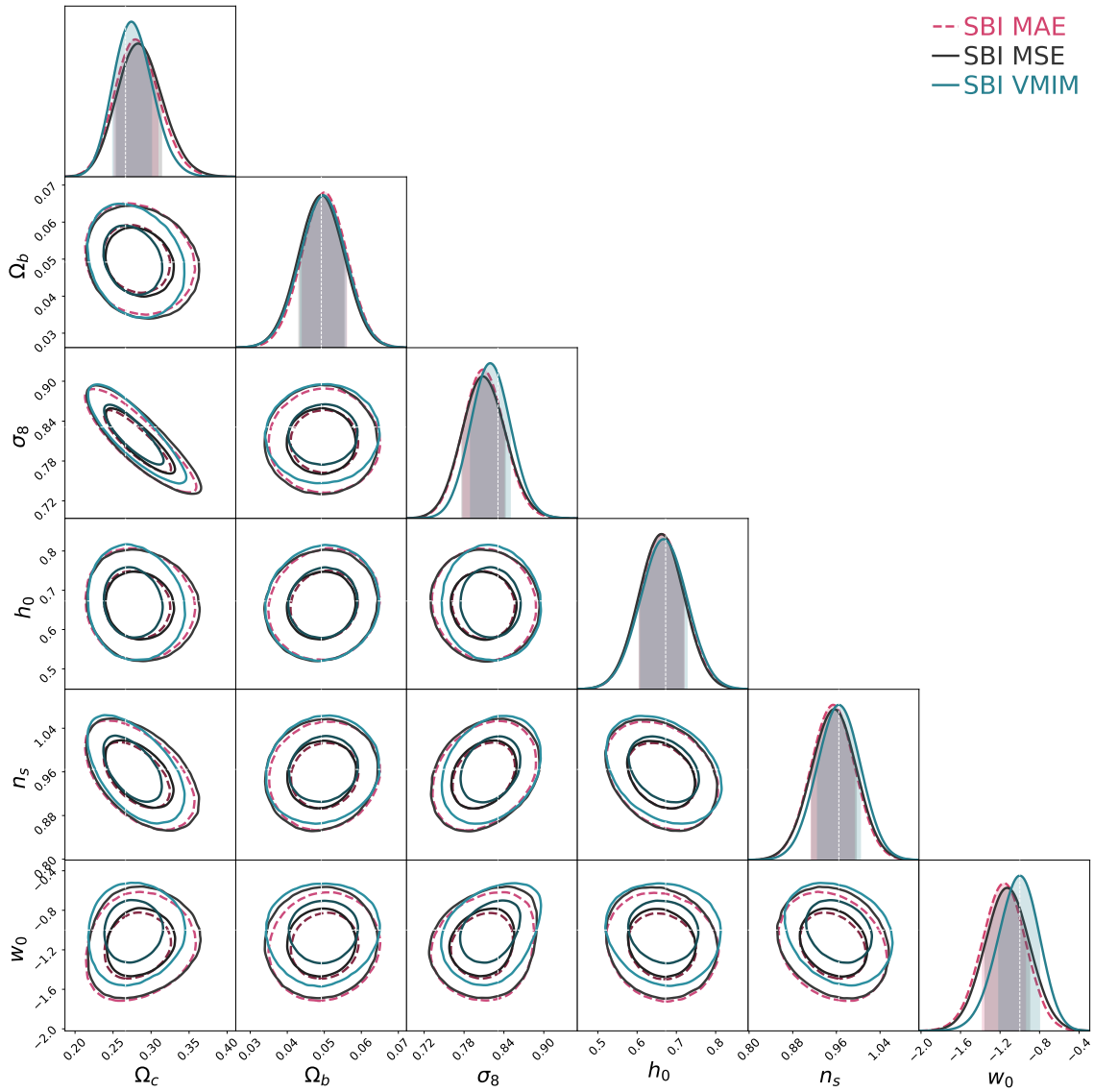

Figure 7.5: Constraints on the ΛCDM parameter space as found in the LSST Y10 survey setup. The constraints are obtained from three CNN map compressed statistics: the MSE (black dashed contours), the MAE (magenta dashed contours), and VMIM (blue contours), described in section 7.3. The contours show the 68% and the 95% confidence regions. The dashed white lines define the true parameter values.

	VMIM	MSE	MAE
Ω_c	$0.274^{+0.026}_{-0.025}$	$0.283^{+0.031}_{-0.029}$	$0.279^{+0.030}_{-0.028}$
Ω_b	$\left(49.9^{+6.0}_{-6.4}\right) \times 10^{-3}$	$(49.2 \pm 6.1) \times 10^{-3}$	$\left(50.1^{+5.9}_{-6.1}\right) \times 10^{-3}$
σ_8	$0.819^{+0.031}_{-0.029}$	$0.808^{+0.034}_{-0.032}$	$0.808^{+0.032}_{-0.030}$
w_0	$-1.00^{+0.20}_{-0.22}$	$-1.13^{+0.23}_{-0.22}$	-1.16 ± 0.22
h_0	$0.666^{+0.061}_{-0.057}$	$0.662^{+0.056}_{-0.057}$	$0.664^{+0.056}_{-0.058}$
n_s	$0.963^{+0.041}_{-0.038}$	0.956 ± 0.041	$0.955^{+0.039}_{-0.041}$

Table 7.5: Summary of the marginalized parameter distributions.

Chapter 8

Conclusion and Perspectives

During my Ph.D., I initiated a highly successful collaboration to build the next generation of cosmological inference and simulation tools based on *automatic differentiation*. Numerical simulations play a key role in investigating and testing our cosmological model. Many current cosmic shear analyses use methods (e.g., Hamiltonian Monte-Carlo, Variational Inference, dimensionality reduction by Fisher-Information Maximization) that require fast, accurate, and differentiable forward simulations to perform efficient inference in a large number of dimensions. Moreover, *fast* and *differentiable* simulations can significantly enhance performance compared to conventional techniques. For instance, preliminary results presented by Zeghal et al. (2022); Zeghal et al. (in prep.) have demonstrated that having access to the gradients of the forward model is beneficial for constraining the posterior density estimates.

In this final chapter, I will describe the contributions of this thesis and conclude with its perspectives.

8.1 Summary and Contributions

In the first part of this manuscript, I focused on my initial work, whose goal was to develop differentiable weak lensing simulations capable of including an innovative gravity-solving scheme, a GPU framework, a small-scale correction approach, and systematics such as intrinsic alignment.

Specifically, I extended and contributed to the development of the N-body simulator *FlowPM* (Modi et al., 2021). First, I implemented the cosmological functions necessary for the N-Body solver in the TensorFlow framework. Then, I extended the PM-code by adding an alternative integration scheme based on a system of Ordinary Differential Equations. In addition, I worked on the Hybrid Physical-Neural (HPN) scheme, a correction scheme aimed at calibrating the PM simulations to mimic the accuracy of high-resolution N-body simulations. The scheme was based on a minimally-parametric neural network component, modeling a residual effective force compensating for the PM

approximations. I showed that the HPN scheme, incorporated as part of the FlowPM code, yielded good results for the matter power spectrum and the cross-correlation coefficient. I benchmarked the model against the Potential Gradient Descent (PGD) scheme (Dai et al., 2018), demonstrating that, if the two methods were comparable in terms of improving the small-scale power spectrum, the HPN approach showed larger improvements in terms of cross-correlation coefficients. Moreover, unlike the PGD method, the HPN approach proved to be less sensitive to the settings of the simulations used for training.

Meanwhile, I implemented the Differentiable Lensing Lightcone (DLL), a simulator for Weak Lensing maps based on the cosmological N-body simulations realized with FlowPM. The weak lensing package allows the inclusion of systematics, such as intrinsic alignment, and enables computing the nonlinear lensing convergence maps at different source redshifts.
I validated our simulations against the weak lensing simulations κTNG-Dark (Osato et al., 2021) by comparing both the lensing angular power spectrum and multiscale peak counts. As a first application, I demonstrated how the differentiability of numerical simulations can be exploited to evaluate the Fisher Matrix. Indeed, I compared the constraining power of two weak lensing statistics: the lensing power spectrum and peak counts, and investigated the degeneracy in high-dimensional cosmological and nuisance parameter space through Fisher forecasts. This work was part of a project I led in the LSST Dark Energy Science Collaboration (DESC), presented in Lanzieri et al. (2023). As mentioned several times in this thesis, this novel framework has been thought of as a powerful tool for cosmological analysis, mainly to exploit the advantages of map-based forward modeling approaches for cosmological inference with upcoming stage-IV survey data.

This has motivated the second part of this thesis, I finally made use of our results for weak lensing analysis within the Likelihood-Free Inference (LFI) and Bayesian Hierarchical Method (BHM). In particular, starting from the consideration that LFI approaches are more efficient, the lower the dimensionality of the summary statistics, I am currently working on benchmarking different procedures to optimally extract informative summaries of the weak lensing mass map, i.e., finding the compressed data vector as informative as possible while being as low-dimensional as possible. Specifically, I am comparing the posterior probability distributions obtained from mock weak lensing mass maps compressed using Convolutional Neural Networks (CNNs) trained with different loss functions. The preliminary results presented in this thesis show results from neural networks trained using the Mean Square Error, the Mean Absolute Error, and the Variational Mutual Information Maximization procedure (Jeffrey et al., 2021). However, we are currently working to obtain posterior distributions using the Gaussian Negative Log-likelihood and the Information Maximizing Neural Network. In order to validate the results in the LFI context, we also perform a Bayesian forward modeling analysis. To forward model the convergence maps and create the simulations required to train the LFI models, I contributed to the development of `SBILens`, a

Python package designed for weak lensing inference with a differentiable simulator implemented in Jax. `SBILens` allows the sampling of convergence maps in a tomographic setting, accounting for the cross-correlation between different redshift bins. Although the results obtained so far are very preliminary, and the entire benchmark between different compression schemes is not yet completed, we can draw an important conclusion. Within the context of our setting and considering the current progress, we can start to validate the results obtained with the LFI to those obtained with a Bayesian hierarchical strategy. As expected, we find the LFI analysis to lead to the same posterior distributions as the BHM, and both are much more powerful than a standard analysis based on the 2-point function.

8.2 Perspectives

The results presented in this work represent only the first step towards constructing a physical inference framework suitable for a stage-IV dark energy survey. In addition to the potential developments and improvements we have already discussed throughout this thesis at several points, there are a few specific additional aspects that I would like to delve into.

Graph Networks with ODE Integrators for modeling baryons To exploit the full potential of next-generation weak lensing surveys, we must increase our knowledge of baryonic physics, that, if not taken into account, may cause biases in our cosmological parameter inference. Although different methods to incorporate baryonic feedback in post-processing steps, based on semi-analytical or data-driven models, have already been proposed, a straightforward improvement of our simulator involves extending the Hybrid Physical-Neural ODEs I have developed in Lanzieri et al. (2022). Specifically, one could explore the integration of PM-codes with graph neural networks that model short-range particle-particle interactions, learned from high-resolution full hydrodynamical simulations. Inspired by Sanchez-Gonzalez et al. (2019), the idea is to combine graph networks with differentiable ordinary differential equations (ODEs), which define the evolution of the system, as a mechanism for predicting future states. In other words, training the neural network to learn the ODEs. The result would be a hybrid physical-data-driven approach that, due to the nature of its architecture, should be more suitable for density fields with different features than the ones used for training.

Differentiable particle-mesh simulations with massive neutrinos Massive neutrinos have a significant impact on the background evolution of the Universe, suppressing the growth of cosmic structure on small scales. Modeling this effect and inferring the sum of the three neutrino masses pose crucial challenges in modern cosmology. Upcoming cosmological surveys, such as LSST and Euclid, are expected to provide

precise measurements of neutrino mass.

In this context, the Differentiable lensing lightcone package I have developed could be extended to include a model beyond standard ΛCDM, capable of generating simulations with massive neutrinos for stage-IV-like surveys. The neutrino particles could be incorporated into the FlowPM scheme following the methodology presented by Bayer et al. (2021), i.e., by including neutrinos as an extra set of particles in the simulation. This approach would allow us to preserve the full physical description of the simulations.

On the contrary, approaches relying on using deep learning as a black-box to learn neutrino effects would not guarantee to work outside the training regime and might not capture cosmology dependence correctly.

3x2pt analyses with Simulation-Based Inference The last decade has shown a notable improvement in both the precision and the multitude of cosmological probes, allowing us to test the ΛCDM paradigm with unprecedented accuracy. However, recent works have revealed possible tensions between cosmological parameters as measured by different cosmological probes. These tensions lead us to believe that a multi-probe analysis is necessary to enhance control over systematics and uncertainties, as well as to investigate models beyond the ΛCDM. Unfortunately, multi-probe analyses require the ability to compute joint covariance matrices that properly describe the cross-correlation of different observables. In this context, evaluating the accurate likelihood becomes even more challenging.

This may motivate a new approach to explore the combinations of various neural compressed summaries (e.g., weak lensing power spectrum, galaxy clustering power spectrum, weak lensing peaks, galaxy clustering peaks) to infer the cosmological parameters of the ΛCDM model in a likelihood-free analysis. A first natural step would be extending the simulations I have developed, with a view to reproducing a 3x2 point analysis for stage IV surveys. After that, with the results from our studies on LFI approaches, the summary statistics will be compressed down to low-dimensional summaries, and the posterior distributions on the parameters of interest will be obtained using a density estimator.

The result of combining the likelihood-free inference approach and the multi-probe analysis will be a new cosmological inference pipeline, capable of breaking degeneracies and controlling systematics.

Bibliography

Sverre J Aarseth, J Richard Gott III, and Edwin L Turner. N-body simulations of galaxy clustering. i-initial conditions and galaxy collapse times. *The Astrophysical Journal*, 228:664–683, 1979.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow, Large-scale machine learning on heterogeneous systems, November 2015.

Peter AR Ade, N Aghanim, M Arnaud, Mark Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, JG Bartlett, N Bartolo, et al. Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.

Virginia Ajani, Austin Peel, Valeria Pettorino, Jean-Luc Starck, Zack Li, and Jia Liu. Constraining neutrino masses with weak-lensing multiscale peak counts. *Physical Review D*, 102(10):103531, 2020.

Virginia Ajani, Jean-Luc Starck, and Valeria Pettorino. Starlet 1-norm for weak lensing cosmology. *Astronomy & Astrophysics*, 645:L11, 2021.

Joël Akeret, Alexandre Refregier, Adam Amara, Sebastian Seehars, and Caspar Hasner. Approximate bayesian computation for forward modeling in cosmology. *Journal of Cosmology and Astroparticle Physics*, 2015(08):043, 2015.

Andreas Albrecht, Gary Bernstein, Robert Cahn, Wendy L Freedman, Jacqueline Hewitt, Wayne Hu, John Huth, Marc Kamionkowski, Edward W Kolb, Lloyd Knox, et al. Report of the dark energy task force. *arXiv preprint astro-ph/0609591*, 2006.

Justin Alsing and Benjamin Wandelt. Generalized massive optimal data compression. *Monthly Notices of the Royal Astronomical Society: Letters*, 476(1):L60–L64, 2018.

Justin Alsing, Alan Heavens, Andrew H Jaffe, Alina Kiessling, Benjamin Wandelt, and Till Hoffmann. Hierarchical cosmic shear power spectrum inference. *Monthly Notices of the Royal Astronomical Society*, 455(4):4452–4466, 2016.

Justin Alsing, Alan Heavens, and Andrew H Jaffe. Cosmological parameters, shear maps and power spectra from cfhtlens using bayesian hierarchical inference. *Monthly Notices of the Royal Astronomical Society*, 466(3):3272–3292, 2017.

Justin Alsing, Benjamin Wandelt, and Stephen Feeney. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *Monthly Notices of the Royal Astronomical Society*, 477(3):2874–2885, 2018.

Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, 2019.

Raul E Angulo and Oliver Hahn. Large-scale dark matter simulations. *Living Reviews in Computational Astrophysics*, 8(1):1, 2022.

Giovanni Aricò, Raul E Angulo, Carlos Hernández-Monteagudo, Sergio Contreras, Matteo Zennaro, Marcos Pellejero-Ibañez, and Yetli Rosas-Guevara. Modelling the large-scale mass density field of the universe as a function of cosmology and baryonic physics. *Monthly Notices of the Royal Astronomical Society*, 495(4):4800–4819, 2020.

David J Bacon, DM Goldberg, BTP Rowe, and AN Taylor. Weak gravitational flexion. *Monthly Notices of the Royal Astronomical Society*, 365(2):414–428, 2006.

J. S. Bagla. Cosmological n-body simulation: Techniques, scope and status. *Current Science*, 88(7):1088–1100, 2005. ISSN 00113891. URL http://www.jstor.org/stable/24110531.

Jasjeet S Bagla. Treepm: A code for cosmological n-body simulations. *Journal of Astrophysics and Astronomy*, 23:185–196, 2002.

Jasjeet Singh Bagla and T Padmanabhan. Cosmological n-body simulations. *Pramana*, 49:161–192, 1997.

JS Bagla and Suryadeep Ray. Performance characteristics of treepm codes. *New Astronomy*, 8(7):665–677, 2003.

David Barber and Felix Agakov. Information maximization in noisy channels: A variational approach. *Advances in Neural Information Processing Systems*, 16, 2003.

Josh Barnes and Piet Hut. A hierarchical o (n log n) force-calculation algorithm. *nature*, 324(6096):446–449, 1986.

Matthias Bartelmann. Gravitational lensing. *Classical and Quantum Gravity*, 27: 233001, 2010.

Daniel Baumann. Tasi lectures on inflation, 2012.

Adrian E Bayer, Arka Banerjee, and Yu Feng. A fast particle-mesh simulation of non-linear cosmological structure formation with massive neutrinos. *Journal of Cosmology and Astroparticle Physics*, 2021(01):016, 2021.

F Bernardeau, S Colombi, E Gaztañaga, and R Scoccimarro. Large-scale structure of the universe and cosmological perturbation theory. *Physics Reports*, 367(1-3):1–248, 2002.

José M Bernardo. Reference analysis. *Handbook of statistics*, 25:17–90, 2005.

Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul A. Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.*, 20:28:1–28:6, 2019. URL http://jmlr.org/papers/v20/18-403.html.

Alain Blanchard, S Camera, Carmelita Carbone, VF Cardone, S Casas, Sébastien Clesse, S Ilić, M Kilbinger, T Kitching, Martin Kunz, et al. Euclid preparation-vii. forecast validation for euclid cosmological probes. *Astronomy & Astrophysics*, 642: A191, 2020.

Roger Blandford and Ramesh Narayan. Fermat's principle, caustics, and the classification of gravitational lens images. *The Astrophysical Journal*, 310:568–582, 1986.

Diego Blas, Julien Lesgourgues, and Thomas Tram. The cosmic linear anisotropy solving system (class). part ii: approximation schemes. *Journal of Cosmology and Astroparticle Physics*, 2011(07):034, 2011.

Jonathan A. Blazek, Niall MacCrann, M. A. Troxel, and Xiao Fang. Beyond linear galaxy alignments. *Phys. Rev. D*, 100:103506, Nov 2019. doi: 10.1103/PhysRevD.100.103506. URL https://link.aps.org/doi/10.1103/PhysRevD.100.103506.

Paul Bode, Jeremiah P Ostriker, and Guohong Xu. The tree particle-mesh n-body gravity solver. *The Astrophysical Journal Supplement Series*, 128(2):561, 2000.

Vanessa Böhm, Stefan Hilbert, Maksim Greiner, and Torsten A Enßlin. Bayesian weak lensing tomography: Reconstructing the 3d large-scale distribution of matter with a lognormal prior. *Physical Review D*, 96(12):123510, 2017.

Vanessa Böhm, Yu Feng, Max E Lee, and Biwei Dai. Madlens, a python package for fast and differentiable non-gaussian lensing simulations. *Astronomy and Computing*, 36:100490, 2021.

Supranta S Boruah, Eduardo Rozo, and Pier Fiedorowicz. Map-based cosmology inference with lognormal cosmic shear maps. *Monthly Notices of the Royal Astronomical Society*, 516(3):4111–4122, 2022a.

Supranta S Boruah, Eduardo Rozo, and Pier Fiedorowicz. Map-based cosmology inference with lognormal cosmic shear maps. *Monthly Notices of the Royal Astronomical Society*, 516(3):4111–4122, 2022b.

FR Bouchet, S Colombi, E Hivon, and R Juszkiewicz. Perturbative lagrangian approach to gravitational instability. *arXiv preprint astro-ph/9406013*, 1994.

Aoife Boyle, Cora Uhlemann, Oliver Friedrich, Alexandre Barthelemy, Sandrine Codis, Francis Bernardeau, Carlo Giocoli, and Marco Baldi. Nuw cdm cosmology from the weak-lensing convergence pdf. *Monthly Notices of the Royal Astronomical Society*, 505(2):2886–2902, 2021.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Sarah Bridle and Lindsay King. Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *New Journal of Physics*, 9(12):444–444, dec 2007. doi: 10.1088/1367-2630/9/12/444. URL https://doi.org/10.1088/1367-2630/9/12/444.

Sarah Bridle, John Shawe-Taylor, Adam Amara, Douglas Applegate, Sreekumar T Balan, Joel Berge, Gary Bernstein, Hakon Dahle, Thomas Erben, Mandeep Gill, et al. Handbook for the great08 challenge: An image analysis competition for cosmological lensing. *The Annals of Applied Statistics*, pages 6–37, 2009.

Sarah Bridle, Sreekumar T Balan, Matthias Bethge, Marc Gentile, Stefan Harmeling, Catherine Heymans, Michael Hirsch, Reshad Hosseini, Mike Jarvis, Donnacha Kirk, et al. Results of the great08 challenge: an image analysis competition for cosmological lensing. *Monthly Notices of the Royal Astronomical Society*, 405(3):2044–2061, 2010.

ML Brown, AN Taylor, NC Hambly, and S Dye. Measurement of intrinsic alignments in galaxy ellipticities. *Monthly Notices of the Royal Astronomical Society*, 333(3): 501–509, 2002.

T Buchert. A class of solutions in newtonian cosmology and the pancake theory. *Astronomy and Astrophysics*, 223:9–24, 1989.

Thomas Buchert. Lagrangian theory of gravitational instability of friedman–lemaitre cosmologies–a generic third-order model for non-linear clustering. *Monthly Notices of the Royal Astronomical Society*, 267(4):811–820, 1994.

Jean-Eric Campagne, Franç ois Lanusse, Joe Zuntz, Alexandre Boucaud, Santiago Casas, Minas Karamanis, David Kirkby, Denise Lanzieri, Austin Peel, and Yin Li. JAX-COSMO: An end-to-end differentiable and GPU accelerated cosmology library. *The Open Journal of Astrophysics*, 6, apr 2023. doi: 10.21105/astro.2302.05163. URL https://doi.org/10.21105%2Fastro.2302.05163.

Paolo Catelan. Lagrangian dynamics in non-flat universes and non-linear gravitational evolution. *Monthly Notices of the Royal Astronomical Society*, 276(1):115–124, 1995.

Paolo Catelan, Marc Kamionkowski, and Roger D. Blandford. Intrinsic and extrinsic galaxy alignment. *Monthly Notices of the Royal Astronomical Society*, 320(1):L7–L13, 01 2001a. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2001.04105.x. URL https://doi.org/10.1046/j.1365-8711.2001.04105.x.

Paolo Catelan, Marc Kamionkowski, and Roger D Blandford. Intrinsic and extrinsic galaxy alignment. *Monthly Notices of the Royal Astronomical Society*, 320(1):L7–L13, 2001b.

Tom Charnock, Guilhem Lavaux, and Benjamin D Wandelt. Automatic physical inference with information maximizing neural networks. *Physical Review D*, 97(8): 083004, 2018.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Sihao Cheng and Brice Ménard. Weak lensing scattering transform: dark energy and neutrino mass sensitivity. *Monthly Notices of the Royal Astronomical Society*, 507 (1):1012–1020, 2021.

AD Chernin, DI Nagirner, and SV Starikova. Growth rate of cosmological perturbations in standard model: Explicit analytical solution. *Astronomy & Astrophysics*, 399(1): 19–21, 2003.

Nora Elisa Chisari, Mark LA Richardson, Julien Devriendt, Yohan Dubois, Aurel Schneider, Amandine MC Le Brun, Ricarda S Beckmann, Sebastien Peirani, Adrianne Slyz, and Christophe Pichon. The impact of baryons on the matter power spectrum from the horizon-agn cosmological hydrodynamical simulation. *Monthly Notices of the Royal Astronomical Society*, 480(3):3962–3977, 2018.

Nora Elisa Chisari, David Alonso, Elisabeth Krause, C Danielle Leonard, Philip Bull, Jérémy Neveu, Antonio Villarreal, Sukhdeep Singh, Thomas McClintock, John Ellison, et al. Core cosmology library: Precision cosmological predictions for lsst. *The Astrophysical Journal Supplement Series*, 242(1):2, 2019.

Lucinda Clerkin, Donnacha Kirk, M Manera, O Lahav, F Abdalla, Adam Amara, D Bacon, C Chang, E Gaztanaga, A Hawken, et al. Testing the lognormality of the galaxy and weak lensing convergence distributions from dark energy survey maps. *Monthly Notices of the Royal Astronomical Society*, 466(2):1444–1461, 2017.

Peter Coles and Bernard Jones. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society*, 248(1):1–13, 1991.

Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Physics reports*, 372(1):1–129, 2002.

Robert G Crittenden, Priyamvada Natarajan, Ue-Li Pen, and Tom Theuns. Spin-induced galaxy alignments and their implications for weak-lensing measurements. *The Astrophysical Journal*, 559(2):552, 2001.

Rupert AC Croft and Christopher A Metzler. Weak-lensing surveys and the intrinsic correlation of galaxy ellipticities. *The Astrophysical Journal*, 545(2):561, 2000.

Biwei Dai and Uroš Seljak. Learning effective physical laws for generating cosmological hydrodynamics with lagrangian deep learning. *Proceedings of the National Academy of Sciences*, 118(16):e2020324118, 2021.

Biwei Dai, Yu Feng, and Uroš Seljak. A gradient based method for modeling baryons and matter in halos of fast simulations. *Journal of Cosmology and Astroparticle Physics*, 2018(11):009, 2018.

Marek Demianski, Zdzisław A Golda, and Andrzej Woszczyna. Evolution of density perturbations in a realistic universe. *General Relativity and Gravitation*, 37:2063–2082, 2005.

J. P. Dietrich and J. Hartlap. Cosmology with the shear-peak statistics. *monthly Notices of the Royal Astronomical Society*, 402(2):1049–1058, 02 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.15948.x. URL https://doi.org/10.1111/j.1365-2966.2009.15948.x.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp, 2017.

Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Academic press, 2020.

Klaus Dolag, Stefano Borgani, Sabine Schindler, Antonio Diaferio, and Andrei M Bykov. Simulation techniques for cosmological simulations. *Space science reviews*, 134:229–268, 2008.

Tim Eifler, Elisabeth Krause, Scott Dodelson, Andrew R Zentner, Andrew P Hearin, and Nickolay Y Gnedin. Accounting for baryonic effects in cosmic shear tomography: determining a minimal set of nuisance parameters using pca. *Monthly Notices of the Royal Astronomical Society*, 454(3):2451–2471, 2015.

Daniel J Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605, 1998.

Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, 2017.

Zuhui Fan, Huanyuan Shan, and Jiayi Liu. NOISY WEAK-LENSING CONVERGENCE PEAK STATISTICS NEAR CLUSTERS OF GALAXIES AND BEYOND. *The Astrophysical Journal*, 719(2):1408–1420, 2010. doi: 10.1088/0004-637x/719/2/1408. URL https://doi.org/10.1088/0004-637x/719/2/1408.

Yu Feng, Man-Yat Chu, Uroš Seljak, and Patrick McDonald. Fastpm: a new scheme for fast simulations of dark matter and haloes. *Monthly Notices of the Royal Astronomical Society*, 463(3):2273–2286, 2016.

Christian Fidler, Thomas Tram, Cornelius Rampf, Robert Crittenden, Kazuya Koyama, and David Wands. Relativistic initial conditions for n-body simulations. *Journal of Cosmology and Astroparticle Physics*, 2017(06):043, 2017.

Ronald A Fisher. The logic of inductive inference. *Journal of the royal statistical society*, 98(1):39–82, 1935.

Janis Fluri, Tomasz Kacprzak, Alexandre Refregier, Adam Amara, Aurelien Lucchi, and Thomas Hofmann. Cosmological constraints from noisy convergence maps through deep learning. *Phys. Rev. D*, 98:123518, 2018a. doi: 10.1103/PhysRevD.98.123518. URL https://link.aps.org/doi/10.1103/PhysRevD.98.123518.

Janis Fluri, Tomasz Kacprzak, Alexandre Refregier, Adam Amara, Aurelien Lucchi, and Thomas Hofmann. Cosmological constraints from noisy convergence maps through deep learning. *Physical Review D*, 98(12):123518, 2018b.

Janis Fluri, Tomasz Kacprzak, Aurelien Lucchi, Alexandre Refregier, Adam Amara, Thomas Hofmann, and Aurel Schneider. Cosmological constraints with deep learning from kids-450 weak lensing maps. *Physical Review D*, 100(6):063514, 2019.

Janis Fluri, Tomasz Kacprzak, Alexandre Refregier, Aurelien Lucchi, and Thomas Hofmann. Cosmological parameter estimation and inference using deep summaries. *Physical Review D*, 104(12):123526, 2021.

Janis Fluri, Tomasz Kacprzak, Aurelien Lucchi, Aurel Schneider, Alexandre Refregier, and Thomas Hofmann. Full w cdm analysis of kids-1000 weak lensing maps using deep learning. *Physical Review D*, 105(8):083518, 2022.

A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, January 1922. doi: 10.1007/BF01332580.

Oliver Friedrich, Daniel Gruen, J DeRose, D Kirk, E Krause, T McClintock, ES Rykoff, S Seitz, RH Wechsler, GM Bernstein, et al. Density split statistics: Joint model of counts and lensing in cells. *Physical Review D*, 98(2):023508, 2018.

Oliver Friedrich, Cora Uhlemann, Francisco Villaescusa-Navarro, Tobias Baldauf, Marc Manera, and Takahiro Nishimichi. Primordial non-gaussianity without tails–how to measure f nl with the bulk of the density pdf. *Monthly Notices of the Royal Astronomical Society*, 498(1):464–483, 2020.

M Gatti, B Jain, C Chang, M Raveri, D Zürcher, L Secco, L Whiteway, N Jeffrey, C Doux, T Kacprzak, et al. Dark energy survey year 3 results: cosmology with moments of weak lensing mass maps. *arXiv preprint arXiv:2110.10141*, 2021.

Elena Giusarma, Mauricio Reyes Hurtado, Francisco Villaescusa-Navarro, Siyu He, Shirley Ho, and ChangHoon Hahn. Learning neutrino effects in cosmology with convolutional neural networks. *arXiv preprint arXiv:1910.04255*, 2019.

MH Goroff, Benjamin Grinstein, S-J Rey, and Mark B Wise. Coupling of modes of cosmological mass density fluctuations. *The Astrophysical Journal*, 311:6–14, 1986.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019a.

David S. Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. Automatic posterior transformation for likelihood-free inference, 2019b.

Arushi Gupta, José Manuel Zorrilla Matilla, Daniel Hsu, et al. Non-Gaussian information from weak lensing data via deep learning. 97(10):103515, 2018. doi: 10.1103/PhysRevD.97.103515.

Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, 1981. doi: 10.1103/PhysRevD.23.347. URL https://link.aps.org/doi/10.1103/PhysRevD.23.347.

Anik Halder, Oliver Friedrich, Stella Seitz, and Tamas N Varga. The integrated three-point correlation function of cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 506(2):2780–2803, 2021.

Joachim Harnois-Déraps, Ludovic van Waerbeke, Massimo Viola, and Catherine Heymans. Baryons, neutrinos, feedback and weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 450(2):1212–1223, 2015.

Joachim Harnois-Déraps, Nicolas Martinet, and Robert Reischke. Cosmic shear beyond 2-point statistics: Accounting for galaxy intrinsic alignment with projected tidal fields. *Monthly Notices of the Royal Astronomical Society*, 2021.

J Hartlap, Patrick Simon, and P Schneider. Why your model parameter confidences might be too optimistic. unbiased estimation of the inverse covariance matrix. *Astronomy & Astrophysics*, 464(1):399–404, 2007.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabás Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019.

Alan Heavens. Statistical techniques in cosmology. *arXiv e-prints*, art. arXiv:0906.0664, June 2009.

Alan Heavens, Alexandre Refregier, and Catherine Heymans. Intrinsic correlation of galaxy shapes: implications for weak lensing measurements. *Monthly Notices of the Royal Astronomical Society*, 319(2):649–656, 2000a.

Alan Heavens, Justin Alsing, Andrew Jaffe, Till Hoffmann, Alina Kiessling, and Benjamin Wandelt. Bayesian hierarchical modelling of weak lensing: The golden goal. In *The Fourteenth Marcel Grossmann Meeting On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories: Proceedings of the MG14 Meeting on General Relativity, University of Rome "La Sapienza", Italy, 12–18 July 2015*, pages 3005–3010. World Scientific, 2018.

Alan F Heavens, Raul Jimenez, and Ofer Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 317(4):965–972, 2000b.

Wojciech A Hellwing, Matthieu Schaller, Carlos S Frenk, Tom Theuns, Joop Schaye, Richard G Bower, and Robert A Crain. The effect of baryons on redshift space distortions and cosmic density and velocity fields in the eagle simulation. *Monthly Notices of the Royal Astronomical Society: Letters*, 461(1):L11–L15, 2016.

Tom Hennigan, Trevor Cai, Tamara Norman, and Igor Babuschkin. Haiku: Sonnet for JAX, 2020. URL http://github.com/deepmind/dm-haiku.

Catherine Heymans, Michael Brown, Alan Heavens, Klaus Meisenheimer, Andy Taylor, and Christian Wolf. Weak lensing with combo-17: estimation and removal of intrinsic alignments. *Monthly Notices of the Royal Astronomical Society*, 347(3):895–908, 2004.

Stefan Hilbert, Jan Hartlap, and Peter Schneider. Cosmic shear covariance: the log-normal approximation. *Astronomy & Astrophysics*, 536:A85, 2011.

Hilbert, S., Hartlap, J., White, S. D. M., and Schneider, P. Ray-tracing through the millennium simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *A&A*, 499(1):31–43, 2009. doi: 10.1051/0004-6361/200811054. URL https://doi.org/10.1051/0004-6361/200811054.

H Hildebrandt, M Viola, C Heymans, S Joudaki, K Kuijken, C Blake, T Erben, B Joachimi, D Klaes, L Miller, C B Morrison, R Nakajima, G Verdoes Kleijn, A Amon, A Choi, G Covone, J T A de Jong, A Dvornik, I Fenech Conti, A Grado, J Harnois-Déraps, R Herbonnet, H Hoekstra, F Köhlinger, J McFarland, A Mead, J Merten, N Napolitano, J A Peacock, M Radovich, P Schneider, P Simon, E A Valentijn, J L van den Busch, E van Uitert, and L Van Waerbeke. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 465(2):1454–1498, 10 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw2805. URL https://doi.org/10.1093/mnras/stw2805.

Christopher M. Hirata and Uro š Seljak. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Phys. Rev. D*, 70:063526, Sep 2004. doi: 10.1103/PhysRevD.70.063526. URL https://link.aps.org/doi/10.1103/PhysRevD.70.063526.

M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2006. doi: 10.1017/CBO9780511790904.

Roger W Hockney and James W Eastwood. *Computer simulation using particles*. crc Press, 2021.

RW Hockney and JW Eastwood. Computer simulation using particles taylor & francis. *Inc., USA*, 1988.

Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

Hung-Jin Huang, Tim Eifler, Rachel Mandelbaum, and Scott Dodelson. Modelling baryonic physics in future weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 488(2):1652–1678, 2019.

Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. ISSN 0027-8424. doi: 10.1073/pnas.15.3.168. URL https://www.pnas.org/content/15/3/168.

Eric Huff and Rachel Mandelbaum. Metacalibration: Direct self-calibration of biases in shear measurement. *arXiv e-prints*, pages arXiv–1702, 2017.

Emille EO Ishida, Sandro DP Vitenti, Mariana Penna-Lima, Jessi Cisewski, Rafael S de Souza, Arlindo MM Trindade, Ewan Cameron, Vinicius C Busti, COIN collaboration, et al. Cosmoabc: likelihood-free inference via population monte carlo approximate bayesian computation. *Astronomy and Computing*, 13:1–11, 2015.

Bhuvnesh Jain, Uroš Seljak, and Simon White. Ray-tracing simulations of weak lensing by large-scale structure. *The Astrophysical Journal*, 530(2):547, 2000.

M Jarvis, E Sheldon, J Zuntz, T Kacprzak, SL Bridle, A Amara, R Armstrong, MR Becker, GM Bernstein, C Bonnett, et al. The des science verification weak lensing shear catalogues. *Monthly Notices of the Royal Astronomical Society*, 460(2): 2245–2281, 2016.

Jens Jasche and Benjamin D. Wandelt. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. 432(2):894–913, June 2013. doi: 10.1093/mnras/stt449.

Jens Jasche and Benjamin D Wandelt. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *Monthly Notices of the Royal Astronomical Society*, 432(2):894–913, 2013.

Edwin T Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241, 1968.

Niall Jeffrey, François Lanusse, Ofer Lahav, and Jean-Luc Starck. Deep learning dark matter map reconstructions from des sv weak lensing data. *Monthly Notices of the Royal Astronomical Society*, 492(4):5023–5029, 2020.

Niall Jeffrey, Justin Alsing, and François Lanusse. Likelihood-free inference with neural compression of des sv weak lensing map statistics. *Monthly Notices of the Royal Astronomical Society*, 501(1):954–969, 2021.

Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

Adrian Jenkins. Second-order lagrangian perturbation theory initial conditions for resimulations. *Monthly Notices of the Royal Astronomical Society*, 403(4):1859–1872, 2010.

Elise Jennings and Maeve Madigan. astroabc: an approximate bayesian computation sequential monte carlo sampler for cosmological parameter estimation. *Astronomy and computing*, 19:16–22, 2017.

T. Kacprzak, D. Kirk, O. Friedrich, A. Amara, A. Refregier, (The DES Collaboration), et al. Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data. *monthly Notices of the Royal Astronomical Society*, 463(4): 3653–3673, 08 2016a. ISSN 0035-8711. doi: 10.1093/mnras/stw2070. URL https://doi.org/10.1093/mnras/stw2070.

Tomasz Kacprzak and Janis Fluri. Deeplss: Breaking parameter degeneracies in large-scale structure with deep-learning analysis of combined probes. *Physical Review X*, 12(3):031029, 2022.

Tomasz Kacprzak, Joe Zuntz, Barnaby Rowe, Sarah Bridle, Alexandre Refregier, Adam Amara, Lisa Voigt, and Michael Hirsch. Measurement and calibration of noise bias in weak lensing galaxy shape estimation. *Monthly Notices of the Royal Astronomical Society*, 427(4):2711–2722, 2012.

Tomasz Kacprzak, Sarah Bridle, Barnaby Rowe, Lisa Voigt, Joe Zuntz, Michael Hirsch, and Niall MacCrann. Sérsic galaxy models in weak lensing shape measurement: model bias, noise bias and their interaction. *Monthly Notices of the Royal Astronomical Society*, 441(3):2528–2538, 2014.

Tomasz Kacprzak, D Kirk, O Friedrich, A Amara, A Refregier, L Marian, JP Dietrich, E Suchyta, J Aleksić, D Bacon, et al. Cosmology constraints from shear peak statistics in dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 463(4):3653–3673, 2016b.

Nick Kaiser. Weak gravitational lensing of distant galaxies. *The Astrophysical Journal*, 388:272–286, 1992.

Nick Kaiser and Gordon Squires. Mapping the dark matter with weak gravitational lensing. *The Astrophysical Journal*, 404:441–450, 1993.

Maurice George Kendall and Alan Stuart. The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed), 1969.

John F Kenney and ES Keeping. Mathematics of statistics, vol. ii, 1951.

Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, 2015.

Lindsay J King and Peter Schneider. Separating cosmic shear from intrinsic galaxy alignments: Correlation function tomography. *Astronomy & Astrophysics*, 398(1): 23–30, 2003.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Anatoly Klypin, Francisco Prada, Juan Betancort-Rijo, and Franco D Albareti. Density distribution of the cosmological matter field. *Monthly Notices of the Royal Astronomical Society*, 481(4):4588–4601, 2018.

Jan M Kratochvil, Eugene A Lim, Sheng Wang, Zoltán Haiman, Morgan May, and Kevin Huffenberger. Probing cosmology with weak lensing minkowski functionals. *Physical Review D*, 85(10):103513, 2012.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Francois Lanusse, J-L Starck, Adrienne Leonard, and Sandrine Pires. High resolution weak lensing mass mapping combining shear and flexion. *Astronomy & Astrophysics*, 591:A2, 2016.

Lanzieri, Lanusse, François, Modi, Chirag, Horowitz, Benjamin, Harnois-Déraps, Joachim, Starck, Jean-Luc, and The LSST Dark Energy Science Collaboration (LSST DESC). Forecasting the power of higher order weak-lensing statistics with automatically differentiable simulations. *A&A*, 679:A61, 2023. doi: 10.1051/0004-6361/202346888. URL https://doi.org/10.1051/0004-6361/202346888.

Denise Lanzieri, François Lanusse, and Jean-Luc Starck. Hybrid physical-neural odes for fast n-body simulations. *arXiv preprint arXiv:2207.05509*, 2022.

Florent Leclercq. Bayesian optimization for likelihood-free cosmological inference. *Physical Review D*, 98(6):063511, 2018.

Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473, 2000.

Yin Li, Yueying Ni, Rupert AC Croft, Tiziana Di Matteo, Simeon Bird, and Yu Feng. Ai-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Sciences*, 118(19):e2022038118, 2021.

Yin Li, Chirag Modi, Drew Jamieson, Yucheng Zhang, Libin Lu, Yu Feng, François Lanusse, and Leslie Greengard. Differentiable cosmological simulation with adjoint method. *arXiv e-prints*, pages arXiv–2211, 2022.

Zack Li, Jia Liu, José Manuel Zorrilla Matilla, et al. Constraining neutrino mass with tomographic weak lensing peak counts. 99(6):063527, 2019. doi: 10.1103/PhysRevD.99.063527.

Tobias Ignacio Liaudat. *Data-driven modelling of ground-based and space-based telescope's point spread functions*. Theses, Université Paris-Saclay, October 2022. URL https://theses.hal.science/tel-03944690.

D Nelson Limber. The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field. *The Astrophysical Journal*, 117:134, 1953.

Chieh-An Lin and Martin Kilbinger. A new model to predict weak-lensing peak counts-ii. parameter constraint strategies. *Astronomy & Astrophysics*, 583:A70, 2015.

Lin, Chieh-An and Kilbinger, Martin. A new model to predict weak-lensing peak counts - i. comparison with n-body simulations. *A&A*, 576:A24, 2015. doi: 10.1051/0004-6361/201425188. URL https://doi.org/10.1051/0004-6361/201425188.

E. V. Linder and A. Jenkins. Cosmic structure growth and dark energy. *Monthly Notices of the Royal Astronomical Society*, 346(2):573–583, 2003. doi: https://doi.org/10.1046/j.1365-2966.2003.07112.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2966.2003.07112.x.

Jarno Lintusaari, Michael U Gutmann, Ritabrata Dutta, Samuel Kaski, and Jukka Corander. Fundamentals and recent developments in approximate bayesian computation. *Systematic biology*, 66(1):e66–e82, 2017.

Jia Liu and Mathew S Madhavacheril. Constraining neutrino mass with the tomographic weak lensing one-point probability distribution function and power spectrum. *Physical Review D*, 99(8):083508, 2019.

Jia Liu, Andrea Petri, Zoltán Haiman, Lam Hui, Jan M Kratochvil, and Morgan May. Cosmology constraints from the weak lensing peak counts and the power spectrum in cfhtlens data. *Physical Review D*, 91(6):063507, 2015a.

Jia Liu, Simeon Bird, José Manuel Zorrilla Matilla, J Colin Hill, Zoltán Haiman, Mathew S Madhavacheril, Andrea Petri, and David N Spergel. Massivenus: cosmological massive neutrino simulations. *Journal of Cosmology and Astroparticle Physics*, 2018(03):049, 2018.

Xiangkun Liu, Chuzhong Pan, Ran Li, Huanyuan Shan, Qiao Wang, Liping Fu, Zuhui Fan, Jean-Paul Kneib, Alexie Leauthaud, Ludovic Van Waerbeke, et al. Cosmological constraints from weak lensing peak statistics with canada–france–hawaii telescope stripe 82 survey. *Monthly Notices of the Royal Astronomical Society*, 450(3):2888–2902, 2015b.

Jia Liu J., Andrea Petri, Zoltán Haiman, Lam Hui, Jan M. Kratochvil, and Morgan May. Cosmology constraints from the weak lensing peak counts and the power spectrum in cfhtlens data. *Phys. Rev. D*, 91:063507, 2015. doi: 10.1103/PhysRevD.91.063507. URL https://link.aps.org/doi/10.1103/PhysRevD.91.063507.

Xiangkun Liu X., Chuzhong Pan, Ran Li, Huanyuan Shan, Qiao Wang, Liping Fu, Zuhui Fan, Jean-Paul Kneib, Alexie Leauthaud, Ludovic Van Waerbeke, Martin Makler, Bruno Moraes, Thomas Erben, and Aldée Charbonnier. Cosmological constraints from weak lensing peak statistics with Canada–France–Hawaii Telescope Stripe 82 Survey. *monthly Notices of the Royal Astronomical Society*, 450 (3):2888–2902, 05 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv784. URL https://doi.org/10.1093/mnras/stv784.

Tianhuan Lu, Zoltán Haiman, and José Manuel Zorrilla Matilla. Simultaneously constraining cosmology and baryonic physics via deep learning from weak lensing. *Monthly Notices of the Royal Astronomical Society*, 511(1):1518–1528, 2022.

Tianhuan Lu, Zoltán Haiman, and Xiangchong Li. Cosmological constraints from hsc survey first-year data using deep learning. *arXiv preprint arXiv:2301.01354*, 2023.

Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, 30, 2017a.

Jan-Matthis Lueckmann, Pedro J. Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Macke. Flexible statistical inference for mechanistic models of neural dynamics, 2017b.

Jonathan Mackey, Martin White, and Marc Kamionkowski. Theoretical estimates of intrinsic galaxy alignment. *Monthly Notices of the Royal Astronomical Society*, 332 (4):788–798, 2002.

T. Lucas Makinen, Tom Charnock, Justin Alsing, and Benjamin D. Wandelt. Lossless, scalable implicit likelihood inference for cosmological fields. *Journal of Cosmology and Astroparticle Physics*, 2021(11):049, nov 2021. doi: 10.1088/1475-7516/2021/11/049. URL https://doi.org/10.1088%2F1475-7516%2F2021%2F11%2F049.

T Lucas Makinen, Tom Charnock, Pablo Lemos, Natalie Porqueres, Alan F Heavens, and Benjamin D Wandelt. The cosmic graph: Optimal information extraction from large-scale structure using catalogues. *The Open Journal of Astrophysics*, 5(1), dec 2022. doi: 10.21105/astro.2207.05202. URL https://doi.org/10.21105%2Fastro.2207.05202.

Rachel Mandelbaum, Tim Eifler, Renée Hložek, Thomas Collett, Eric Gawiser, Daniel Scolnic, David Alonso, Humna Awan, Rahul Biswas, Jonathan Blazek, et al. The lsst dark energy science collaboration (desc) science requirements document. *arXiv preprint arXiv:1809.01669*, 2018.

Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate bayesian computational methods. *Statistics and computing*, 22(6):1167–1180, 2012.

Nicolas Martinet, Peter Schneider, Hendrik Hildebrandt, HuanYuan Shan, Marika Asgari, Jörg P Dietrich, Joachim Harnois-Déraps, Thomas Erben, Aniello Grado, Catherine Heymans, et al. Kids-450: cosmological constraints from weak-lensing peak statistics–ii: Inference from shear peaks using n-body simulations. *Monthly Notices of the Royal Astronomical Society*, 474(1):712–730, 2018.

Richard Massey, Barnaby Rowe, Alexandre Refregier, David J Bacon, and Joel Bergé. Weak gravitational shear and flexion with polar shapelets. *Monthly Notices of the Royal Astronomical Society*, 380(1):229–245, 2007.

José Manuel Zorrilla Matilla, Manasi Sharma, Daniel Hsu, and Zoltán Haiman. Interpreting deep learning models for weak lensing. *Phys. Rev. D*, 102:123506, 2020. doi: 10.1103/PhysRevD.102.123506. URL https://link.aps.org/doi/10.1103/PhysRevD.102.123506.

M Maturi, C Fedeli, and L Moscardini. Imprints of primordial non-gaussianity on the number counts of cosmic shear peaks. *Monthly Notices of the Royal Astronomical Society*, 416(4):2527–2538, 2011.

Matteo Maturi, Christian Angrick, F Pace, and M Bartelmann. An analytic approach to number counts of weak-lensing peak detections. *Astronomy & Astrophysics*, 519: A23, 2010.

Alexander J Mead, John A Peacock, Catherine Heymans, Shahab Joudaki, and Alan F Heavens. An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. *Monthly Notices of the Royal Astronomical Society*, 454(2):1958–1975, 2015.

Peter Melchior and Massimo Viola. Means of confusion: how pixel noise affects shear estimates for weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 424(4):2757–2769, 2012.

Julian Merten, Carlo Giocoli, Marco Baldi, Massimo Meneghetti, Austin Peel, Florian Lalande, Jean-Luc Starck, and Valeria Pettorino. On the dissection of degenerate cosmologies with machine learning. *Monthly Notices of the Royal Astronomical Society*, 487(1):104–122, 04 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz972. URL https://doi.org/10.1093/mnras/stz972.

Chirag Modi, Francois Lanusse, and Uros Seljak. Flowpm: Distributed tensorflow implementation of the fastpm cosmological n-body solver. *Astronomy and Computing*, 37:100505, 2021.

Irshad Mohammed, Davide Martizzi, Romain Teyssier, and Adam Amara. Baryonic effects on weak-lensing two-point statistics and its cosmological implications. *arXiv preprint arXiv:1410.6826*, 2014.

Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, et al. The illustristng simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):1–29, 2019.

Ken Osato, Masato Shirasaki, and Naoki Yoshida. Impact of baryonic processes on weak-lensing cosmology: Power spectrum, nonlocal statistics, and parameter bias. *The Astrophysical Journal*, 806(2):186, 2015.

Ken Osato, Jia Liu, and Zoltán Haiman. κtng: effect of baryonic processes on weak lensing with illustristng simulations. *Monthly Notices of the Royal Astronomical Society*, 502(4):5593–5602, 2021.

George Papamakarios and Iain Murray. Fast ε-free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.

George Papamakarios and Iain Murray. Fast ϵ-free inference of simulation models with bayesian conditional density estimation, 2018.

George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 837–848. PMLR, 2019.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference, 2021.

Gabriele Parimbelli, Matteo Viel, and Emiliano Sefusatti. On the degeneracy between baryon feedback and massive neutrinos as probed by matter clustering and weak lensing. *Journal of Cosmology and Astroparticle Physics*, 2019(01):010, 2019.

John A. Peacock. *Cosmological Physics*. 1999.

P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

A. Peel et al. Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *A&A*, 599:A79, 2017a. doi: 10.1051/0004-6361/201629928. URL https://doi.org/10.1051/0004-6361/201629928.

Austin Peel, Chieh-An Lin, François Lanusse, Adrienne Leonard, Jean-Luc Starck, and Martin Kilbinger. Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *Astronomy & Astrophysics*, 599:A79, 2017b.

Austin Peel, Florian Lalande, Jean-Luc Starck, Valeria Pettorino, Julian Merten, Carlo Giocoli, Massimo Meneghetti, and Marco Baldi. Distinguishing standard and modified gravity cosmologies with machine learning. *Physical Review D*, 100(2):023508, 2019.

A. Petri. Mocking the weak lensing universe: The LensTools Python computing package. *Astronomy and Computing*, 17:73–79, October 2016. doi: 10.1016/j.ascom.2016.06.001.

Andrea Petri, Zoltán Haiman, Lam Hui, Morgan May, and Jan M Kratochvil. Cosmology with minkowski functionals and moments of the weak lensing convergence field. *Physical Review D*, 88(12):123002, 2013.

Andrea Petri, Zoltán Haiman, and Morgan May. Validity of the born approximation for beyond gaussian weak lensing observables. *Physical Review D*, 95(12):123503, 2017.

Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554*, 2019.

Sandrine Pires, J-L Starck, Adam Amara, Romain Teyssier, Alexandre Réfrégier, and Jalal Fadili. Fast statistics for weak lensing (fastlens): fast method for weak lensing statistics and map making. *Monthly Notices of the Royal Astronomical Society*, 395 (3):1265–1279, 2009.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., et al. Planck 2018 results - vi. cosmological parameters. *A&A*, 641:A6, 2020. doi: 10.1051/0004-6361/201833910. URL https://doi.org/10.1051/0004-6361/201833910.

L.S. Pontriagin. *The Mathematical Theory of Optimal Processes*. International series of monographs in pure and applied mathematics. Pergamon Press; [distributed in the Western Hemisphere by Macmillan, New York], 1964. ISBN 9780080101767. URL https://books.google.fr/books?id=aakrAAAAYAAJ.

LS Pontryagin, VG Boltyanski, RV Gamkrelidze, and EV Mishchenko. The mathematical theory of optimal processes. 1962. interscience. *New York*, 1962.

Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Bayesian forward modelling of cosmic shear data. *Monthly Notices of the Royal Astronomical Society*, 502(2):3035–3044, 2021.

Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Lifting weak lensing degeneracies with a field-based likelihood. *Monthly Notices of the Royal Astronomical Society*, 509(3):3194–3202, 2022.

Natalia Porqueres, Alan Heavens, Daniel Mortlock, Guilhem Lavaux, and T Lucas Makinen. Field-level inference of cosmic shear with intrinsic alignments and baryons. *arXiv preprint arXiv:2304.04785*, 2023.

Douglas Potter, Joachim Stadel, and Romain Teyssier. Pkdgrav3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1):2, 2017.

Chris Power and Alexander Knebe. The impact of box size on the properties of dark matter haloes in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 370(2):691–701, 2006.

Dennis Prangle. Adapting the abc distance function. 2017.

Jonathan K Pritchard, Mark T Seielstad, Anna Perez-Lezaun, and Marcus W Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798, 1999.

Arnau Pujol, Florent Sureau, Jerome Bobin, Frederic Courbin, Marc Gentile, and Martin Kilbinger. Shear measurement bias-i. dependencies on methods, simulation parameters, and measured parameters. *Astronomy & Astrophysics*, 641:A164, 2020.

Alexandre Refregier, Tomasz Kacprzak, Adam Amara, Sarah Bridle, and Barnaby Rowe. Noise bias in weak lensing shape measurements. *Monthly Notices of the Royal Astronomical Society*, 425(3):1951–1957, 2012.

Benjamin Remy, Francois Lanusse, Niall Jeffrey, Jean-Luc Starck, Ken Osato, and Tim Schrabback. Probabilistic mass mapping with neural score estimation. *arXiv e-prints*, pages arXiv–2201, 2022.

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows, 2016.

Dezső Ribli, Bálint Ármin Pataki, José Manuel Zorrilla Matilla, Daniel Hsu, Zoltán Haiman, and István Csabai. Weak lensing cosmology with convolutional neural networks on noisy data. *Monthly Notices of the Royal Astronomical Society*, 490(2): 1843–1860, 2019.

Dezső Ribli, Bálint Ármin Pataki, and István Csabai. An improved cosmological parameter inference scheme motivated by deep learning, 2018.

Matteo Rizzato, Karim Benabed, Francis Bernardeau, and Fabien Lacasa. Tomographic weak lensing bispectrum: a thorough analysis towards the next generation of galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 490(4):4688–4714, 2019.

Douglas H Rudd, Andrew R Zentner, and Andrey V Kravtsov. Effects of baryons and dissipation on the matter power spectrum. *The Astrophysical Journal*, 672(1):19, 2008.

Alvaro Sanchez-Gonzalez, Victor Bapst, Kyle Cranmer, and Peter Battaglia. Hamiltonian graph networks with ode integrators. *arXiv preprint arXiv:1909.12790*, 2019.

Jorit Schmelzle, Aurelien Lucchi, Tomasz Kacprzak, Adam Amara, Raphael Sgier, Alexandre Réfrégier, and Thomas Hofmann. Cosmological model discrimination with deep learning, 2017.

Aurel Schneider, Romain Teyssier, Doug Potter, Joachim Stadel, Julian Onions, Darren S Reed, Robert E Smith, Volker Springel, Frazer R Pearce, and Roman Scoccimarro. Matter power spectrum and the challenge of percent accuracy. *Journal of Cosmology and Astroparticle Physics*, 2016(04):047, 2016.

Aurel Schneider, Romain Teyssier, Joachim Stadel, Nora Elisa Chisari, Amandine MC Le Brun, Adam Amara, and Alexandre Refregier. Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation. *Journal of Cosmology and Astroparticle Physics*, 2019(03):020, 2019.

Michael D Schneider, David W Hogg, Philip J Marshall, William A Dawson, Joshua Meyers, Deborah J Bard, and Dustin Lang. Hierarchical probabilistic inference of cosmic shear. *The Astrophysical Journal*, 807(1):87, 2015.

Peter Schneider. Weak gravitational lensing. In *Gravitational lensing: strong, weak and micro*, pages 269–451. Springer, 2006.

Peter Schneider, Ludovic van Waerbeke, Martin Kilbinger, and Yannick Mellier. Analysis of two-point statistics of cosmic shear-i. estimators and covariances. *Astronomy & Astrophysics*, 396(1):1–19, 2002.

C Seitz and P Schneider. Steps towards nonlinear cluster inversion through gravitational distortions. iii. including a redshift distribution of the sources. *Astronomy and Astrophysics*, 318:687–699, 1997.

Stella Seitz and Peter Schneider. Cluster lens reconstruction using only observed local data–an improved finite-field inversion technique. *arXiv preprint astro-ph/9503096*, 1995.

Stella Seitz, Peter Schneider, and Jürgen Ehlers. Light propagation in arbitrary spacetimes and the gravitational lens approximation. *Classical and Quantum Gravity*, 11 (9):2345–2373, 1994. doi: 10.1088/0264-9381/11/9/016.

Stella Seitz, Peter Schneider, and Jürgen Ehlers. Light propagation in arbitrary spacetimes and the gravitational lens approximation. *Classical and Quantum Gravity*, 11 (9):2345, sep 1994. doi: 10.1088/0264-9381/11/9/016. URL https://dx.doi.org/10.1088/0264-9381/11/9/016.

Elena Sellentin and Alan F. Heavens. On the insufficiency of arbitrarily precise covariance matrices: non-Gaussian weak-lensing likelihoods. *Monthly Notices of the Royal Astronomical Society*, 473(2):2355–2363, 09 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx2491. URL https://doi.org/10.1093/mnras/stx2491.

Elisabetta Semboloni, Henk Hoekstra, Joop Schaye, Marcel P van Daalen, and Ian G McCarthy. Quantifying the effect of baryon physics on weak lensing tomography. *Monthly Notices of the Royal Astronomical Society*, 417(3):2020–2035, 2011a.

Elisabetta Semboloni, Tim Schrabback, Ludovic van Waerbeke, Sanaz Vafaei, Jan Hartlap, and Stefan Hilbert. Weak lensing from space: first cosmological constraints from three-point shear statistics. *Monthly Notices of the Royal Astronomical Society*, 410(1):143–160, 2011b.

HuanYuan Shan, Xiangkun Liu, Hendrik Hildebrandt, Chuzhong Pan, Nicolas Martinet, Zuhui Fan, Peter Schneider, Marika Asgari, Joachim Harnois-Déraps, Henk Hoekstra, et al. Kids-450: cosmological constraints from weak lensing peak statistics–i. inference from analytical prediction of high signal-to-noise ratio convergence peaks. *Monthly Notices of the Royal Astronomical Society*, 474(1):1116–1134, 2018.

Masato Shirasaki, Kana Moriwaki, Taira Oogi, Naoki Yoshida, Shiro Ikeda, and Takahiro Nishimichi. Noise reduction for weak lensing mass mapping: an application of generative adversarial networks to subaru hyper suprime-cam first-year data. *Monthly Notices of the Royal Astronomical Society*, 504(2):1825–1839, 2021.

Patrick Simon. How accurate is limber's equation? *Astronomy & Astrophysics*, 473 (3):711–714, 2007.

Ian Smail, David W Hogg, Lin Yan, and Judith G Cohen. Deep optical galaxy counts with the keck telescope. *The Astrophysical Journal*, 449(2):L105, 1995.

R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 341(4):1311–1332, 06 2003. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2003.06503.x. URL https://doi.org/10.1046/j.1365-8711.2003.06503.x.

Volker Springel. The cosmological simulation code gadget-2. *Monthly notices of the royal astronomical society*, 364(4):1105–1134, 2005.

Volker Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Monthly Notices of the Royal Astronomical Society*, 401 (2):791–851, 2010.

Volker Springel, Naoki Yoshida, and Simon DM White. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001.

Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, et al. First results from the illustristng simulations: matter and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 475(1):676–698, 2018.

Jean-Luc Starck, Jalal Fadili, and Fionn Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16(2): 297–309, 2007. doi: 10.1109/TIP.2006.887733.

Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, USA, 2010. ISBN 0521119138.

Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLOS Computational Biology*, 9(1):1–10, 01 2013. doi: 10.1371/journal.pcbi.1002803. URL https://doi.org/10.1371/journal.pcbi.1002803.

Masahiro Takada and Bhuvnesh Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. *Monthly Notices of the Royal Astronomical Society*, 348(3):897–915, 2004.

Ryuichi Takahashi, Masanori Sato, Takahiro Nishimichi, Atsushi Taruya, and Masamune Oguri. Revising the halofit model for the nonlinear matter power spectrum. *The Astrophysical Journal*, 761(2):152, 2012.

Svetlin Tassev, Matias Zaldarriaga, and Daniel J Eisenstein. Solving large scale structure in ten easy steps with cola. *Journal of Cosmology and Astroparticle Physics*, 2013(06):036, 2013.

Max Tegmark, Andy N Taylor, and Alan F Heavens. Karhunen-loeve eigenvalue problems in cosmology: How should we tackle large data sets? *The Astrophysical Journal*, 480(1):22, 1997.

Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G McCarthy. Painting with baryons: augmenting n-body simulations with gas using deep generative models. *Monthly Notices of the Royal Astronomical Society: Letters*, 487(1): L24–L29, 2019.

MA Troxel and Mustapha Ishak. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Physics Reports*, 558: 1–59, 2015.

Cora Uhlemann, Oliver Friedrich, Francisco Villaescusa-Navarro, Arka Banerjee, and Sandrine Codis. Fisher for complements: extracting cosmology and neutrino mass from the counts-in-cells pdf. *Monthly Notices of the Royal Astronomical Society*, 495 (4):4006–4027, 2020.

Chris Vale and Martin White. Simulating weak lensing by large-scale structure. *The Astrophysical Journal*, 592(2):699, aug 2003. doi: 10.1086/375867. URL https://dx.doi.org/10.1086/375867.

Marcel P van Daalen, Joop Schaye, CM Booth, and Claudio Dalla Vecchia. The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 415(4):3649–3665, 2011.

Licia Verde. Statistical methods in cosmology. In *Lectures on Cosmology: Accelerated Expansion of the Universe*, pages 147–177. Springer, 2010.

Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N Spergel, Rachel S Somerville, Romeel Dave, Annalisa Pillepich, Lars Hernquist, Dylan Nelson, Paul Torrey, et al. The camels project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, 915(1):71, 2021.

Steven Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. 1972.

Henrique S Xavier, Filipe B Abdalla, and Benjamin Joachimi. Improving lognormal models for cosmological fields. *Monthly Notices of the Royal Astronomical Society*, 459(4):3693–3710, 2016.

Guohong Xu. A new parallel n-body gravity solver: Tpm. *arXiv preprint astro-ph/9409021*, 1994.

Xiuyuan Yang, Jan M Kratochvil, Kevin Huffenberger, Zoltán Haiman, and Morgan May. Baryon impact on weak lensing peaks and power spectrum: Low-bias statistics and self-calibration in future surveys. *Physical Review D*, 87(2):023511, 2013.

J. Zeghal, D. Lanzieri, A. Boucaud, F. Lanusse, and Aubourg E. Simulation-efficient implicit inference for weak lensing with differentiable simulators. in prep.

Justine Zeghal, François Lanusse, Alexandre Boucaud, Benjamin Remy, and Eric Aubourg. Neural posterior estimation with differentiable simulators. *arXiv preprint arXiv:2207.05636*, 2022.

Ya B Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.

Andrew R Zentner, Elisabetta Semboloni, Scott Dodelson, Tim Eifler, Elisabeth Krause, and Andrew P Hearin. Accounting for baryons in cosmological constraints from cosmic shear. *Physical Review D*, 87(4):043509, 2013.

Jun Zhang and Eiichiro Komatsu. Cosmic shears should not be measured in conventional ways. *Monthly Notices of the Royal Astronomical Society*, 414(2):1047–1058, 2011.

Zhuoqi Zhang, Chihway Chang, Patricia Larsen, Lucas F Secco, Joe Zuntz, and LSST Dark Energy Science Collaboration. Transitioning from stage-iii to stage-iv: cosmology from galaxy× cmb lensing and shear× cmb lensing. *Monthly Notices of the Royal Astronomical Society*, 514(2):2181–2197, 2022.

Dominik Zürcher, J Fluri, R Sgier, T Kacprzak, M Gatti, C Doux, L Whiteway, A Réfrégier, C Chang, N Jeffrey, et al. Dark energy survey year 3 results: Cosmology with peaks using an emulator approach. *Monthly Notices of the Royal Astronomical Society*, 511(2):2075–2104, 2022.