Weak Gravitational Lensing Cosmology with Novel Analytical and Machine Learning Frameworks

Zhengyangguang Gong

München 2025

Weak Gravitational Lensing Cosmology with Novel Analytical and Machine Learning Frameworks

Zhengyangguang Gong

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Zhengyangguang Gong
aus Hefei, China

München, den 15.05.2025

Erstgutachter: Prof. Dr. Ralf Bender
Zweitgutachter: Prof. Dr. Eiichiro Komatsu
Tag der mündlichen Prüfung: 15.07.2025

Contents

Zusammenfassung **vii**

Abstract **ix**

1 Introduction to cosmology **1**
1.1 Foundations of the cosmological dynamics 2
1.2 Cosmological distance measurement 6
1.3 Standard ΛCDM parametrization 7
1.4 Cosmological experiments 9

2 Statistics and perturbation theory of cosmic density fields **15**
2.1 Random fields 16
2.1.1 Homogeneity, isotropy and ergodicity 16
2.1.2 PDF, moments and cumulants 17
2.2 Density perturbation and its correlation functions 19
2.2.1 Power spectrum 21
2.2.2 Gravitational dynamics of the density perturbation field 23
2.2.3 Standard perturbation theory 24
2.3 Weak lensing cosmology 27
2.3.1 Weak lensing basics 27
2.3.2 Weak lensing measurement 29
2.3.3 Weak lensing shear 2PCF and integrated 3PCF 32
2.4 Bias theory in large-scale structure 34
2.4.1 Galaxy bias 34
2.4.2 Extrema bias in weak lensing field 36
2.5 Statistical analysis with multiprobe cosmology 39

3 Basic machine learning concepts and their applications to cosmology **41**
3.1 Multilayer perceptron (MLP) 42
3.1.1 Training, validation, optimization and testing 42
3.1.2 Emulators for the shear integrated 3PCF 44
3.2 Convolutional neural network (CNN) 46
3.2.1 Fundamental principles of CNNs 46

3.2.2 Interpretable C3NN 48
3.3 Simulation-based inference (SBI) 49
3.3.1 SBI with C3NN 51

4 Intermezzo: integrating analytical methods and machine learning in cosmology 55

5 Clustering of the extreme: A theoretical description of weak lensing critical points power spectra in the mildly nonlinear regime 59

6 Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators 89

7 C3NN: Cosmological Correlator Convolutional Neural Network an Interpretable Machine-learning Framework for Cosmological Analyses 123

8 Making the leap. Part I. Modelling the reconstructed lensing convergence PDF from cosmic shear with survey masks and systematics 141

9 Summary and conclusions 183

Acknowledgments 194

Zusammenfassung

Moderne kosmologische Experimente wie Euclid, Vera Rubin LSST und DESI werden enorme Datenmengen liefern, deren Analyse eine zentrale Herausforderung darstellt. Maschinelles Lernen (ML) hat sich in Bereichen wie Systemklassifikation, synthetische Datengenerierung und Parameterinferenz etabliert und hat erheblich zur Datenanalyse beigetragen. Dennoch bestehen Bedenken hinsichtlich Genauigkeit, Robustheit und Interpretierbarkeit von ML-Ansätzen. Zudem bleibt die Kombination analytischer Methoden mit ML-Techniken weitgehend unerforscht.

In dieser Arbeit werden statistische Analysemethoden für schwache Gravitationslinsenfelder sowohl aus analytischer als auch aus maschineller Lernperspektive entwickelt, mit dem Ziel, unser Verständnis von kosmologischen Modellen, Strukturbildung und Entwicklung zu verbessern. Auf der analytischen Seite leite ich explizite Formeln für die Leistungsspektren und Zweipunkt-Korrelationsfunktionen (2PCFs) von 2D-kritischen Punkten ab, einschließlich Peaks (Maxima), Voids (Minima) und Sattelpunkten, in schwachen Gravitationslinsen-Konvergenzfeldern mit geringer Nicht-Gaussianität. Mithilfe einer störungstheoretischen Bias-Expansion modelliere ich deren Clustering und leite das Leistungsspektrum von kritischen Punkten der schwachen Gravitationslinsen bis zur nächst-nächstführenden Ordnung (NNLO) in der gravitativen Störungstheorie her, wobei Trispektrum-Konfigurationen berücksichtigt werden. Dies dient als Benchmark-Test für N-Body-Simulationen, um sicherzustellen, dass Statistiken wie das Clustering von Linsenpeaks und Voids nicht durch systematische Simulationseffekte verzerrt werden.

Für die ML-Anwendung auf die Analyse von Survey-Daten arbeitete ich an der Entwicklung einer Likelihood-Analyse-Pipeline für kosmologische Einschränkungen unter Verwendung der integrierten Scherungs-Dreipunkt-Korrelationsfunktion $\zeta_{\pm}$. Insbesondere entwickelte ich einen hochpräzisen neuronalen Netzwerk-Emulator für schnelle theoretische Vorhersagen in der Markov-Chain-Monte-Carlo (MCMC)-Parameterinferenz. Mithilfe simulierter Daten, die die Dark Energy Survey (DES) Year-3-Footprint, Masken und tomographische Quellen-Bins nachbilden, zeige ich, dass die Einbeziehung von $\zeta_{\pm}$ zusätzlich zur herkömmlichen Scherungs-Zweipunkt-Korrelationsfunktion $\xi_{\pm}$ die Einschränkungen für ze-ntrale kosmologische Parameter wie $A_{\rm s}$ (oder σ_8) und w_0 um etwa $10-25\%$ verbessert.

Um die Interpretierbarkeit von ML zu verbessern und es besser mit analytischen Ansätzen zu verbinden, stelle ich das Cosmological Correlator Convolutional Neural Network (C3NN) vor, ein hybrides Framework, das Convolutional Neural Networks (CNNs) mit kosmologischen N-Punkt-Korrelationsfunktionen (NPCFs) kombiniert. Ich zeige, dass die Ausgaben

von C3NN explizit in Form analytisch behandelbarer NPCFs ausgedrückt werden können, wodurch es möglich wird, die “Black Box” von ML zu öffnen. Zusammen mit zusätzlichen Algorithmen ermöglicht dieser Ansatz eine quantitative Rangordnung verschiedener NPCF-Ordnungen basierend auf ihrem Beitrag zu Klassifikationsaufgaben und liefert so tiefere Einblicke in die gelernten Merkmale.

Abstract

Modern cosmological experiments such as Euclid, Vera Rubin's LSST, and the Dark Energy Spectroscopic Instrument (DESI) will generate an unprecedented volume of data in the coming years. Effectively analyzing this vast dataset to deepen our understanding of the Universe is an urgent challenge. In response, machine learning (ML) techniques have emerged across various areas of cosmological research, including image classification, synthetic data generation, and parameter inference, significantly enhanced real data analyses. However, concerns remain about the accuracy, robustness, and interpretability of ML approaches. Furthermore, the integration of conventional analytical methods with cutting-edge ML techniques is an underexplored avenue.

This thesis develops statistical analysis methods for weak gravitational lensing fields from both analytical and machine learning perspectives, aiming to improve our understanding of cosmological models, structure formation, and evolution. On the analytical side, I derive explicit formulae for the power spectra and two-point correlation functions (2PCFs) of 2D critical points, including peaks (maxima), voids (minima), and saddle points, in weak gravitational lensing convergence field with mild non-Gaussianity. Using a perturbative bias expansion, I model their clustering and derive the power spectrum of weak lensing critical points up to next-to-next-to-leading order (NNLO), incorporating trispectrum configurations. This serves as a benchmark test for N-body simulations, ensuring that statistics such as lensing peak and void clustering are not biased by simulation systematics.

For the ML application to survey data analysis, I collaborated on the development of a likelihood analysis pipeline for cosmological constraints using the integrated shear three-point correlation function $\zeta_{\pm}$. Specifically, I built a high-precision neural network emulator for fast theoretical predictions in parameter inference using Markov Chain Monte Carlo (MCMC). With simulated data that mimics the Dark Energy Survey (DES) Year-3 footprint, mask, and source tomographic bins, we demonstrate that incorporating $\zeta_{\pm}$ alongside the conventional shear 2PCF $\xi_{\pm}$ improves constraints on key cosmological parameters, such as $A_{\rm s}$ (or σ_8) and w_0, by approximately $10-25\%$.

To address the challenge of ML interpretability and better integrate it with analytical approaches, I introduce the Cosmological Correlator Convolutional Neural Network (C3NN), a hybrid framework that merges convolutional neural networks (CNNs) with cosmological N-point correlation functions (NPCFs). We show that the C3NN output can be explicitly expressed in terms of analytically tractable NPCFs, allowing us to open the "black box" of ML. Along with auxiliary algorithms, this approach enables a quantitative ranking of dif-

ferent orders of NPCF based on their contribution to classification tasks, providing deeper insights from the learned features into cosmological physics.

Chapter 1

Introduction to cosmology

Currently the most widely accepted theory explaining the evolution of the observed Universe from its earliest known period is the Big Bang paradigm. Within its framework, our Universe was generated approximately 13.8 billion years ago with a homogeneously and isotropically distributed energy density at high temperature and pressure. After a tiny fraction of the first second since the initial Big Bang, the Universe would expand exponentially to encompass our whole observable Universe. This so called inflation paradigm is conjectured to be driven by a scalar field and the microscopic quantum fluctuations of this scalar field will be stretched and amplified to a macroscopic scale during the inflationary period. These fluctuations will become primordial perturbations which would seed the Cosmic Microwave Background (CMB) anisotropies and the later structure growth.

At the end of inflation, elementary particles in the Standard Model of particle physics are produced. These particles interact with each other and reach a thermal equilibrium while the Universe continues to expand and cool down. The Universe would then enter an era in which the dominating energy component is radiation (Mukhanov, 2005).

As the expansion and cooling of the Universe persist, those early generated elementary particles of quarks and gluons would combine to form composite subatomic particles such as baryons (e.g. protons and neutrons) and mesons. This process is called Hadronization. Baryons would further combine with each other to form nuclei and they together with electrons generate light elements such as hydrogen and helium. This process is called Big Bang Nucleosynthesis (BBN) (Dodelson & Schmidt, 2020).

The energy density of radiation decreases at a faster rate than the non-relativistic matter along with the expansion of the Universe. The matter domination epoch takes over the energy evolution during which the initial perturbations seen in CMB anisotropies (Aghanim et al., 2020) are significantly enhanced by nonlinear gravitational evolution and grow into Large Scale Structures (LSS) such as galaxies, galaxy clusters, cosmic voids and filaments (Dodelson & Schmidt, 2020).

Additionally there is one piece of critical information embedded in the evolution of the Universe summarized above: *Cosmic neutrinos*, which are generated as elementary particles and obtain their masses at early Universe, evolve along with the expansion of the

Universe. They first travel through the Universe as relativistic particles, and then transit to non-relativistic around redshift $z \sim 100 - 200$ with a variation depending on their exact mass. They interact with other constituents of matter through which they leave an imprint on LSS. In modern cosmology, a widely used statistical tool that can characterize the *matter distribution of LSS* is the 2-point correlation function (2PCF). Therefore the imprint cosmic neutrinos impose on LSS can be reflected in the quantification of this statistic. By properly modelling this statistic, one would obtain constraints on neutrinos' properties which in this case is the sum of their masses (Lesgourgues et al., 2013).

In this chapter, we first discuss some basic cosmological concepts of an expanding universe in Sec. 1.1. Then in Sec. 1.2 we introduce different distance measurements in cosmology. In Sec. 1.3 we present the standard parametrization of Big Bang paradigm−the ΛCDM model with its associated energy density components. Finally in Sec. 1.4 we briefly introduce some major cosmological experiments and some unresolved puzzles from their observations.

The content of this chapter can be found in references such as Peacock (1999), Mukhanov (2005), Lesgourgues et al. (2013) and Dodelson & Schmidt (2020).

1.1 Foundations of the cosmological dynamics

We have concrete evidence to show that our Universe is expanding (Riess et al., 1998). To describe the increasing distance between two points in the Universe, it is convenient to introduce the scale factor $a(t)$. The contemporary value of it is set to one and it becomes smaller until zero when we trace back in time to the beginning of Big Bang (Dodelson & Schmidt, 2020). To be in accordance with this concept, one also introduces the comoving distance χ which is the distance between two points in the comoving coordinates. This coordinate system expands uniformly as the Universe and thus the comoving distance remains constant with respect to time. The actual physical distance d between two points is then proportional to the multiplication of the scale factor and the comoving distance (Dodelson & Schmidt, 2020):

$$d \propto a(t)\chi \ . \tag{1.1}$$

Besides the scale factor, there is another commonly used quantity for the expansion: the redshift z. It is related to the fact that the physical wavelength of light emitted from a distant object is stretched proportional to the scale factor as it propagates towards the observer. It is defined as (Dodelson & Schmidt, 2020):

$$1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm emit}} = \frac{a(t_0)}{a(t_{\rm emit})} = \frac{1}{a(t_{\rm emit})} \ , \tag{1.2}$$

where $\lambda_{\rm obs}$ is the observed wavelength, $\lambda_{\rm emit}$ is the emission wavelength. In conclusion, we can write:

$$1 + z = \frac{1}{a(t)} \ . \tag{1.3}$$

In order to measure distances in the Universe, we also need to know its geometry. There are three possibilities which are parametrized by the curvature k (Dodelson & Schmidt, 2020):

- Euclidean (Flat): The Euclidean space with $k = 0$ and the Universe is of infinite volume.
- Open: A hyperbolic universe which can be imagined locally as an infinitely extended saddle shape with $k < 0$.
- Closed: Analogous to the surface of a sphere with constant positive curvature $k > 0$.

Before writing the distance between two points in an expanding curved space explicitly, we first have to declare two properties for the distribution of mass in our Universe as premises which together are generally called the cosmological principle (Peacock, 1999):

- Homogeneous: At large scales (> 100 Mpc), the matter has a constant density.
- Isotropic: At large scales (> 100 Mpc), the matter distribution looks the same in all directions.

With the above ingredients, we can write the differential distance in an expanding curved universe in a spherical coordinate system as (Mukhanov, 2005):

$$\mathrm{d}l_{3d}^2 = a^2(t)\Big(\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)\Big) \ , \tag{1.4}$$

where r is the comoving distance along the radial direction, and θ, ϕ are the polar and azimuth angles respectively. Using Einstein summation convention we can reformulate Eq. (1.4) as:

$$\mathrm{d}l_{3d}^2 = \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j \ , \tag{1.5}$$

where we should carry out a summation over indices i and j. Both of them can go from 1 to 3, referring to three spatial components in the spherical coordinate. γ is a diagonal matrix called metric tensor and it defines the rule to calculate the distance between two points in the space. In this case, γ reads:

$$\gamma_{ij} = \begin{pmatrix} \frac{a^2(t)}{1-kr^2} & 0 & 0 \\ 0 & a^2(t)r^2 & 0 \\ 0 & 0 & a^2(t)r^2\sin^2\theta \end{pmatrix} \ . \tag{1.6}$$

However, the gravitational interaction in cosmology is described by the theory of general relativity and in relativistic theory the spatial distance is not an invariant quantity with respect to coordinate transformations (Mukhanov, 2005). In order to compute such an

invariant in cosmology, we have to upgrade Eq. (1.4) to a four-dimensional spacetime in which the differential interval is calculated as (Dodelson & Schmidt, 2020):

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \tag{1.7}$$

$$= -\mathrm{d}t^2 + a^2(t)\Big(\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)\Big) \ , \tag{1.8}$$

where we use natural units assuming the speed of light $c = 1$, a convention we stick to in the following discussion unless mentioned otherwise. Indices μ and ν go from 0 to 3 with 0 indicating the time coordinate t and 1 to 3 representing the spatial coordinates. The metric tensor g in the four-dimensional spacetime is the well known Friedmann–Lemaître–Robertson–Walker (FLRW) metric expressed in spherical coordinates. Contemporary experiments such as Planck support the claim that our Universe has a flat spatial geometry (Efstathiou & Gratton, 2020). Therefore for most computation purposes, we can take $k = 0$. Similar to Eq. (1.6), the metric tensor g can be written as:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & \frac{a^2(t)}{1-kr^2} & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2\sin^2\theta \end{pmatrix} \ . \tag{1.9}$$

Another aspect of general relativity is that it relates the metric to the constituents of the Universe. This relation is contained in the Einstein field equations which can be neatly written down as a collection of tensor equations (Dodelson & Schmidt, 2020):

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{1.10}$$

The first term on the left hand side of the equation is the Einstein tensor defined as (Dodelson & Schmidt, 2020):

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \ , \tag{1.11}$$

where $R_{\mu\nu}$ is the Ricci tensor and can be computed from the metric and its derivatives. The scalar R is the Ricci scalar and is the contraction of the Ricci tensor $R = g^{\mu\nu}R_{\mu\nu}$ (Dodelson & Schmidt, 2020). Back to Eq. (1.10), Λ is the cosmological constant and it serves as one standard form of interpreting the dark energy density in the Universe. G on the right hand side of the equation is the gravitational constant and T is the energy-momentum tensor which characterizes the constituents of the Universe such as matter, radiation and so on. In the case of a universe described by the FLRW metric in Eq. (1.9), the energy-momentum tensor is diagonal and can be expressed using the density and pressure of the constituents such that each of them can be treated as a perfect fluid (Dodelson & Schmidt, 2020):

$$T^\mu{}_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \ , \tag{1.12}$$

where ρ is the energy density and p is the pressure.

Based on Eq. (1.10), one can derive multiple equations governing the evolution of the Universe. Detailed derivations can be found in the previously mentioned references and here we just show some important results for a homogeneous and isotropic expanding universe. When $\mu = \nu = 0$, the metric only considers the time-time component and Eq. (1.10) becomes

$$G_{00} + \Lambda g_{00} = 8\pi G T_{00} \ . \tag{1.13}$$

If we substitute the tensor entries on both sides of the equation with their corresponding expressions, we obtain the first Friedmann equation (Dodelson & Schmidt, 2020):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{a^2} + \frac{\Lambda}{3} \ , \tag{1.14}$$

where $\dot{a}$ represents the derivative of the scale factor with respect to time and ρ is the energy density of the constituents of the Universe such as matter and radiation. Next when considering the spatial part of Eq. (1.10) with $\mu = \nu = i$ where $i = 1, 2, 3$, we obtain the same equation for each index:

$$G_{ii} + \Lambda g_{ii} = 8\pi G T_{ii} \ , \tag{1.15}$$

from which the second Friedmann equation can be derived (Dodelson & Schmidt, 2020):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \ , \tag{1.16}$$

where $\ddot{a}$ is the acceleration of the scale factor and p is the pressure of those constituents of the Universe. It is related to the corresponding energy density through the equation of state:

$$p = \omega\rho \ , \tag{1.17}$$

where ω is the coefficient and has different values for different constituents in the Universe. We have $\omega_{\rm m} = 0$ for non-relativistic matter, $\omega_{\rm r} = 1/3$ for radiation and $w_\Lambda = -1$ for the cosmological constant (Dodelson & Schmidt, 2020).

If we combine Eq. (1.14) and Eq. (1.16) we can derive the equation describing the energy density evolution (Mukhanov, 2005):

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) \ , \tag{1.18}$$

where one can replace the pressure with the equation of state and solve the differential equation to obtain the energy density of a specific constituent as a function of the scale factor.

1.2 Cosmological distance measurement

In modern cosmology, distance between the observer and a distant object is a critical quantity. In this section, we introduce two types of distance: One is the *comoving distance* and the other is the *angular diameter distance*. For simplicity, we only consider the scenario when the Universe has a flat geometry ($k = 0$) which is relevant to our theoretical modelling later. Readers can find expressions for distances in a universe with non-zero curvature in the previously mentioned references.

Imagine a light ray is emitted from a distant object at time t and reaches the observer at the current time t_0. It travels along the null-geodesic which means $\mathrm{d}s$ in Eq. (1.8) is always equal to zero. If we assume the light travels along the radial direction and all angular changes are zero, we have $\mathrm{d}r = \mathrm{d}\chi$ in Eq. (1.8) and the comoving distance between the observer and the object is given by (Dodelson & Schmidt, 2020):

$$\chi(t) = \int_t^{t_0} \frac{\mathrm{d}t'}{a(t')} \ . \tag{1.19}$$

One can rewrite the above equation as an integration of the redshift z. But before that we need to introduce an important quantity defined to describe the expansion rate of the Universe called the Hubble parameter (Dodelson & Schmidt, 2020):

$$H \equiv \frac{\dot{a}}{a} \ , \tag{1.20}$$

and if we combine this definition with Eq. (1.3), we can rewrite Eq. (1.19) as:

$$\chi(t) = \int_{a(t)}^1 \frac{\mathrm{d}a'}{\dot{a}(t')a(t')} = \int_{a(t)}^1 \frac{\mathrm{d}a'}{H(a')a^2(t')} = \int_0^z \frac{\mathrm{d}z'}{H(z')} \ , \tag{1.21}$$

where we exploit the relation that the redshift is zero at the current value of the scale factor $a(t_0)$.

In cosmological observations, when we know the physical size of an extended object in the sky is l and it subtends an angle of θ, we can define another distance measurement called the angular diameter distance (Dodelson & Schmidt, 2020):

$$d_A \equiv \frac{l}{\theta} \ . \tag{1.22}$$

The comoving size of the extended object can be represented as l/a and the comoving distance from the observer to that object is exactly what we derive in Eq. (1.21). Therefore we can write the angular size of the object in terms of comoving length scales $\theta = (l/a)/\chi$ and once we substitute the angular size in Eq. (1.22) with this expression, the angular diameter distance becomes (Dodelson & Schmidt, 2020):

$$d_A = a\chi = \frac{\chi}{1+z} \ , \tag{1.23}$$

where it's again important to remember that this is the angular diameter distance expressed in a Euclidean space.

1.3 Standard ΛCDM parametrization

The Big Bang paradigm mentioned at the beginning of this chapter can be parametrized by the ΛCDM model, also called the standard model of cosmology. It is a model with six independent free parameters which can well explain the major results of the current cosmological observations (Dodelson & Schmidt, 2020):

- The origin and anisotropies of Cosmic Microwave Background (CMB).
- The distribution of galaxies at large scales.
- The abundance of light elements such as hydrogen and helium today.

In short, ΛCDM model proposes a Euclidean universe dominated in its energy budget by the non-baryonic cold dark matter and dark energy, with its current large scale structure originated from the initial perturbations generated by the inflation. To better understand this model, we have to know the constituents of cold dark matter (CDM) and dark energy (Λ) which are both beyond the Standard Model of particle physics. But before that, we would like to first set up the framework depicting the energy budget for different constituents of the Universe.

For a flat universe, Eq. (1.14) can be rewritten as:

$$\frac{8\pi G}{3H^2}\sum_i \rho_i = 1 \ , \tag{1.24}$$

where we set $k = 0$ and absorb the cosmological constant term into the energy density. The original energy density ρ is now expressed as the summation over different constituents and the Hubble parameter H comes from the left hand side of Eq. (1.14) according to the definition. The factor $8\pi G/3H^2$ is the reciprocal of the critical density ρ_c when the Hubble parameter takes its current value (Dodelson & Schmidt, 2020):

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \ , \tag{1.25}$$

and by convention cosmologists usually use concepts of density parameter or physical density parameter to describe the energy fraction of a specific constituent in the Universe. The former is defined as:

$$\Omega_i = \frac{\rho_i(t_0)}{\rho_c} \ , \tag{1.26}$$

where $\rho_i(t_0)$ is the current energy density for species i and the latter is:

$$\omega_i = \Omega_i h^2 \ , \tag{1.27}$$

where h is the dimensionless reduced Hubble constant which is equal to the present-day Hubble parameter value divided by 100 km/s/Mpc. For clarity, it is necessary to distinguish between the physical density parameter and the equation of state coefficient.

Cold dark matter: First proposed by Fritz Zwicky in 1933 (Zwicky, 1933) to explain the mismatch between the galaxy velocities within clusters and the estimated mass of the observed luminous objects, cold dark matter particles interact only through gravity and thus prominently affect the structure formation in the Universe. There are many other evidences for its existence such as the location discrepancy between the density concentration of the total matter and the gas in the Bullet Cluster (Clowe et al., 2006) and rotation curves of galaxies. Its name "cold" indicates that the particles do not possess high velocities and are non-relativistic. Since cold dark matter particles cannot be described by nuclear or atomic physics, there are no direct methods to detect it. We can only study it by investigating the properties and distribution of the total matter through probes such as the gravitational lensing. Besides cold dark matter, the total matter also includes **baryonic matter** which is made up of protons and neutrons. Advanced galaxy surveys such as the Dark Energy Survey (DES) points out that the current density parameter for the total matter is $\Omega_{\rm m} = 0.339^{+0.032}_{-0.031}$ (Abbott et al., 2022) through the combined constraints of galaxy clustering and weak lensing. From other independent experiments such as the X-ray emission from galaxy clusters one can approximate the density parameter ratio between the baryonic matter and the total matter which according to Mantz et al. (2014) is $\Omega_{\rm b}/\Omega_{\rm m} = (0.089 \pm 0.012)h^{-3/2}$. Therefore one can have the idea that the matter takes up about 30% of the total energy in our Universe and it consists of approximately 20% baryonic matter.

In conclusion, the cold dark matter and the baryonic matter constitute roughly 25% and 5% of the whole energy budget of the Universe respectively. Both of them can be treated as pressure less matter which means the equation of state coefficients for them are zero: $\omega_{\rm cdm} = \omega_{\rm b} = 0$.

Dark energy: Two pieces of strong evidence support the existence of dark energy. First our Universe is shown to have a Euclidean geometry, therefore the summation of the density parameter of all constituents should be one according to Eq. (1.24). However, the density parameter of the total matter is only about 0.3 and the energy density for radiation is negligible today. This implies that an additional ingredient which does cluster in the same way as does matter is required to fill the energy budget of the Universe. Moreover, experiments of observing distant supernovae events (Riess et al., 1998) show that our Universe is undergoing an accelerated expansion. Matter would only slow down the expansion through gravitational interaction. It supports the claim that there is another dominating constituent in the Universe which does not participate in the gravitational collapse and has a negative equation of state (Dodelson & Schmidt, 2020). This can be seen from Eq. (1.16) where we assume the dominating energy constituent is dark energy, then for the acceleration on the left hand side to be positive we must have:

$$\rho_{\rm DE} + 3\omega_{\rm DE}\rho_{\rm DE} < 0 \ , \tag{1.28}$$

where we use the equation of state for the pressure of dark energy on the right hand side of the equation. This leads to $\omega_{\rm DE} < -1/3$.

Scolnic et al. (2018) show the constraints on the equation of state coefficient for the dark energy $\omega_{\rm DE} = -1.026 \pm 0.041$ from the supernovae observations together with the Planck CMB measurements. This value is consistent with the theoretical modelling of the cosmological constant that has $\omega_\Lambda = -1$. Therefore it is justified to assume that the dark energy is a cosmological constant with $\Omega_{\rm DE} = \Omega_\Lambda \approx 0.7$. However, even if we can successfully establish the existence of the dark energy and relate it to the cosmological constant, the source for its energy density is still an open question today.

Two other important constituents in the Universe are radiation and neutrinos. The majority of the radiation contribute to the total energy budget today through the Cosmic Microwave Background (CMB) photons and its energy density only makes up a tiny fraction (Dodelson & Schmidt, 2020):

$$\omega_\gamma = \Omega_\gamma h^2 = 2.47 \times 10^{-5} \ . \tag{1.29}$$

Therefore radiation is negligible in ΛCDM model today. As for cosmic neutrinos, they do not contribute significantly to the total energy budget as well. However, they are not massless particles as photons and by interacting with cold dark matter and baryons in the late Universe, they influence the structure formation.

1.4 Cosmological experiments

To validate the ΛCDM model and gain deeper insights into the formation and evolution of cosmic structures within its framework, or to test alternative cosmological models, it is essential to conduct diverse cosmological experiments or surveys. These should encompass various observable tracers of the underlying dark matter density field while spanning a wide range of redshifts and length scales.

Below, we reference a figure from Preston et al. (2023) that provides an overview of the approximate scale dependence, in terms of wavenumber k, and redshift dependence z of various cosmological observations. Each observable shown in the figure corresponds to several dedicated cosmological surveys.

The most up-to-date study based on cosmic microwave background (CMB) observations, including CMB lensing, comes from *Planck*[1] (Aghanim et al., 2020). Future CMB experiments such as *CMB-S4*[2](Abazajian et al., 2022) and *LiteBIRD*[3](Hazumi et al., 2020), will achieve unprecedented measurement precision on CMB temperature fluctuations. These missions aim to detect signals such as primordial B-mode polarization and primordial gravitational waves, which are crucial for understanding inflationary physics.

[1]https://www.cosmos.esa.int/web/planck
[2]https://cmb-s4.org/
[3]https://www.isas.jaxa.jp/en/missions/spacecraft/future/litebird.html

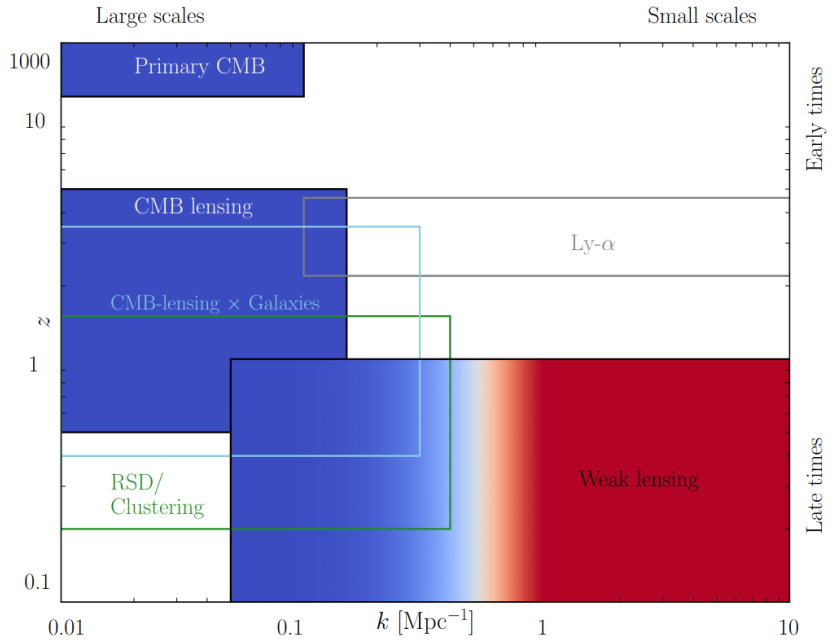

Figure 1.1: A guide of approximate scale and redshift dependence of different cosmological observables. On linear to mildly nonlinear scales, we have Cosmic Microwave Background (CMB), CMB lensing and Redshift Space Distortion (RSD)/galaxy positions from high to low redshifts (blue, filled). On the other hand, weak gravitational lensing probes the nonlinear scales but usually on low redshifts (red, filled). Besides, Lyman-α emission and other line intensity mappings such as Hα and 21cm, can also probe nonlinear scales in our Universe but to a higher redshift range (z ideally up to ~ 20) and thus provide information for the cosmic reionization. Image source: Preston et al. (2023).

In the late Universe, photometric galaxy surveys, including the *Kilo-Degree Survey* (KiDS)[4] (de Jong et al., 2013), the *Dark Energy Survey* (DES)[5](Dark Energy Survey Collaboration et al., 2016) and the *Hyper Suprime-Cam SSP Survey* (HSC)[6](Aihara et al., 2018), focus on weak gravitational lensing galaxy shear measurements. Collectively known as Stage-III galaxy surveys, these studies play a crucial role in constraining cosmological parameters such as the matter density parameter Ω_m, density fluctuations σ_8, and various dark matter and dark energy models, while also accounting for astrophysical effects. These scientific investigations will be further extended by Stage-IV galaxy surveys, including *Euclid*[7](Mellier et al., 2024) and the *Rubin Observatory LSST*[8](Breivik et al., 2022). On the other hand, spectroscopic galaxy surveys such as the *Baryon Oscillation Spectroscopic*

[4]https://kids.strw.leidenuniv.nl/index.php
[5]https://www.darkenergysurvey.org/
[6]https://hsc.mtk.nao.ac.jp/ssp/
[7]https://www.euclid-ec.org/public/data/surveys/
[8]https://www.bnl.gov/lsst/

Survey (BOSS)[9], the *Extended Baryon Oscillation Spectroscopic Survey* (eBOSS)[10] and the *Dark Energy Spectroscopic Instrument* (DESI)[11] provide us with a deep and accurate spatial distribution of the cosmological structures through the observation of different luminous tracers such as bright galaxies, luminous red galaxies (LRGs), emission line galaxies (ELGs) and quasars (Dawson et al., 2012, 2016; Dey et al., 2019). With their measurement of galaxy clustering, we can have a joint constraint on the above cosmological parameters together with weak lensing measurement. Also we gain better understanding of the expansion of our Universe and structure growth through their measurement of Baryon Acoustic Oscillation (BAO) and RSD (see e.g. Hou et al. (2020)).

Multi-line intensity mapping, including Hα, Lyman-α, 21cm emission and so on, provide us with another exciting perspective in studying the history and evolution of our Universe (Bernal & Kovetz, 2022). They have the potential to enhance our observation to a much larger scale both time and space-wise compared to low redshift galaxy surveys. Experiments such as *Square Kilometre Array* (SKA)[12] and *International Low-Frequency Array* (LOFAR)[13] are particularly promising in giving insights into the period of reionization (Mellema et al., 2013; Wiersma et al., 2013).

These different experiments can return results which are either consistent with or contradictory to each other under the framework of ΛCDM model. In other words, we can exploit these independent experimental results to test the standard cosmological model on different fronts. Those occurred discrepancies can be important hints towards alternative or modified theories of dark matter, dark energy and general relativity.

Here we briefly discuss two representative tension scenarios in the inferred Bayesian posteriors of cosmological parameters from different survey experiments:

- H_0 tension: H_0 is the current Hubble parameter which characterizes the expansion rate of our Universe at the moment. Indirect inference of such a parameter from CMB experiments and the assumed ΛCDM model yields a tension compared to the direct inference based on late Universe observables such as supernovae and strong lensing time-delay cosmography (blue points and error bars in the direct measurement section in Fig. 1.2) up to 5σ. Fig. 1.2 shows quite complete results of H_0 measurements from both early and late Universe experiments (They can also be categorized as indirect and direct methods). Various new models and corrections with respect to ΛCDM paradigm have been proposed, including dark radiation, early dark energy and so on, to solve this tension (Readers who are interested can refer to Schöneberg et al. (2022) and references therein). This tension remains an open question until the present.

- S_8 tension: By definition, $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$, characterizes the amplitude of matter clustering in the late Universe where σ_8 is the root mean square of the amplitude of

[9]https://www.sdss4.org/surveys/boss
[10]https://www.sdss4.org/surveys/eboss/
[11]https://www.desi.lbl.gov/
[12]https://www.skao.int/en
[13]https://lofar-surveys.org/index.html

matter perturbation smoothed over 8 h^{-1}Mpc, h is the Hubble constant H_0 in unit of 100 km s^{-1}Mpc^{-1}. The results of S_8 from CMB observations assuming the evolution history of ΛCDM are $2-3\sigma$ larger than those probed by low redshift observables, such as gravitational lensing, galaxy clustering and galaxy cluster number counts. We show in Fig. 1.3 different constraints on S_8 where "WL" stands for "Weak Lensing", "GC" stands for "Galaxy Clustering" and "CC" represents "Cluster Count". There is literature implying that S_8 tension is due to our incomplete modeling of the nonlinear scale physics (Amon & Efstathiou, 2022; Preston et al., 2023). Other analyses show, however, that a mechanism beyond baryonic feedback alone is required to resolve the S_8 tension, i.e. new physics beyond ΛCDM (Terasawa et al., 2025; Salcido & McCarthy, 2024). Similar to H_0 tension, S_8 tension remains debatable until today.

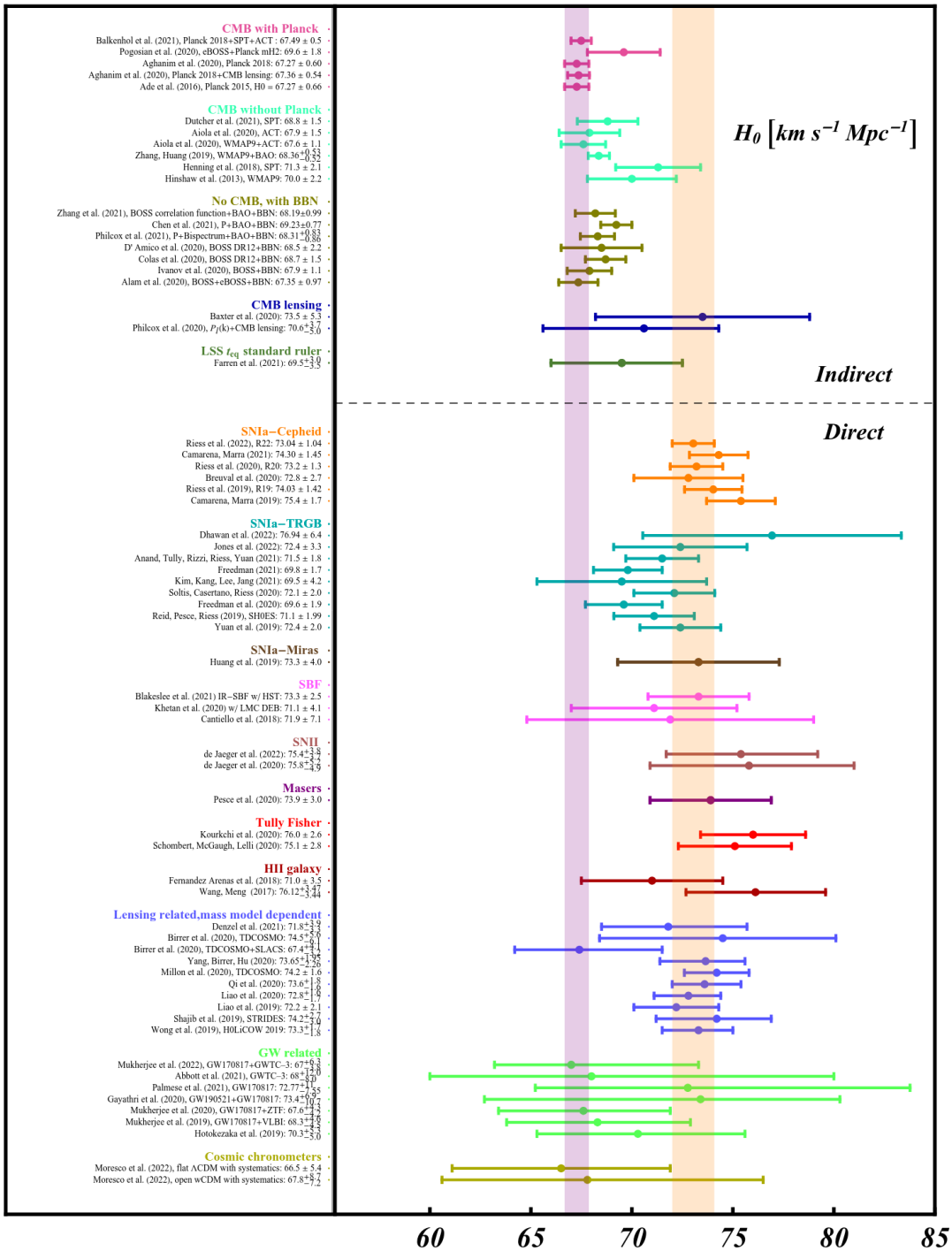

Figure 1.2: 68% confidence level of H_0 values from different cosmological probes. Image source: Abdalla et al. (2022)

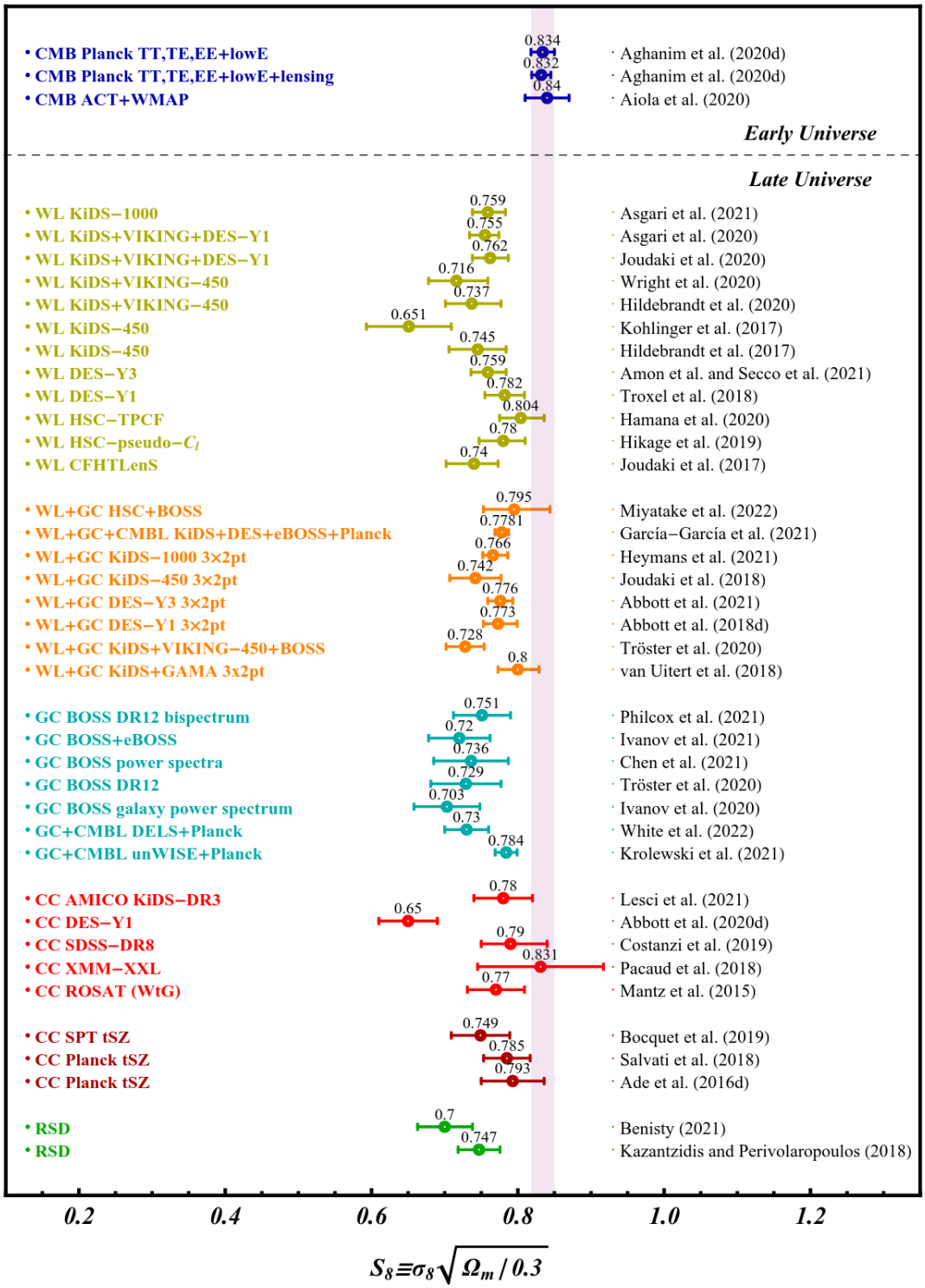

Figure 1.3: 68% confidence level of S_8 values from different cosmological probes. Image source: Abdalla et al. (2022)

Chapter 2

Statistics and perturbation theory of cosmic density fields

In order to probe LSS in the Universe, we need unbiased and accurate statistical tools to extract quantitative information from the observed fields. Due to the nonlinear gravitational evolution of the late Universe, the dark matter density field[1] which forms the basics of the cosmic structure possesses non-Gaussian properties, i.e. it cannot be described completely by the mean and variance of its density fluctuations. In this case, more complicated statistics are needed to help us better understand the evolution and formation of the LSS. One popular and efficient approach to formulate those statistics is the perturbation theory. This methodology is analytical and to a certain degree is complementary to the machine learning-based methods we will discuss in the next chapter.

As dark matter field cannot be directly probed via electromagnetic waves, we can only infer its spatial distribution through luminous objects populated on top of it. This necessarily requires the modeling of the connection between these luminous objects and the underlying dark matter field. Due to the small-scale nonlinear gravitational dynamics, various baryonic effects and the impact from the local environment, the modeling of this connection becomes exceedingly complex. One standard approach to address such a problem is to use bias theory where the luminous objects are regarded as biased tracers of the underlying dark matter field and the connections are parameterized by free parameters, called the bias parameters. In general, one can combine bias theory with perturbation theory to model statistics of tracer fields on large scales. In such a framework, an operator of the underlying matter, with the allowed symmetry of the target tracers (e.g. not only matter density but also tidal fields and their time derivatives), is associated with a bias parameter whose exact values can depend on spatial scales, time evolution and the specific population of the target tracer.

In this chapter, we would like to discuss statistics and physical models that are essential

[1]Our discussions throughout this work are under the context that dark matter exists. We do not consider alternative theories such as the *Scale-Invariant Vacuum* (SIV) Theory (Maeder & Gueorguiev, 2020) where dark matter hypothesis is not necessary.

to the papers presented in this dissertation. In Sec. 2.1, we will introduce the concept of a random field and some of its statistical characteristics. Next in Sec. 2.2, we will first discuss a conventional statistical measure of density fields, i.e. the correlation function, followed by a short overview of standard perturbation theory which provides the mathematical tool to calculate correlation functions. We will then explore their applications in weak gravitational lensing, a phenomenon that directly traces the total matter density distribution, in Sec. 2.3, and examine the modeling of cosmic shear 2PCF and integrated 3-point correlation function (3PCF). We then briefly discuss general biased tracers in Sec. 2.4 within the framework of perturbative bias expansion. Additionally, we demonstrate how galaxies 2PCFs and weak lensing critical points can be derived using this approach. Finally in Sec. 2.5, we briefly touch on 3×2-point analyses and multi-probe cosmology in modern survey experiments which are powerful techniques for advancing our understanding of the Universe.

2.1 Random fields

A spatial random field ρ is a mathematical model that describes a stochastic process that varies across the space. In either a discrete or continuous spatial domain with a given volume V and dimensionality D, ρ assigns a random variable to each location. Suppose for a position $\mathbf{x}_i \in V$, the field value $\rho_i = \rho(\mathbf{x}_1)$ is a random variable. A collection of such random variables would be a spatial random field

$$\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_N) \ , \tag{2.1}$$

where the subscript N above represents N distinct positions within the volume V (Xavier et al., 2016). Each realization of the field ρ represents a possible configuration of values across the space.

2.1.1 Homogeneity, isotropy and ergodicity

The currently accepted explanation for the origin of the LSS observed today is attributed to the quantum fluctuation of the inflation field in the very early Universe (Mukhanov, 2005). This fluctuation is a statistically homogeneous, isotropic and ergodic random process. In other words, the density field produced by this process, although went through various modifications in the subsequent evolution, can be described as a random field and reserves these fundamental properties.

From a statistical perspective, we are not interested in the values of $\rho(\mathbf{x})$ for a particular realization within a survey volume V, but rather its expectation values. If we assume that the same fluctuation and evolution mechanisms generate a multitude of universes and ours is only one of them which is referred to as one *realization*, the corresponding *ensemble average* of the density field is defined as:

$$\big\langle \rho(\mathbf{x}) \big\rangle \equiv \int \mathrm{d}\rho P(\rho) \rho(\mathbf{x}) \ , \tag{2.2}$$

where $P(\rho)$ is the probability density function (PDF) of the density field at one particular position $\mathbf{x}$ and this probability density function is inferred from the ensemble of universes. The integration goes over all realizations at position $\mathbf{x}$. In this sense, the ensemble average value of the density field at one point in the sky can be different from another point.

Statistical homogeneity implies further modification on the above quantity and would cancel its dependency on position. It means that the ensemble average value $\big\langle\rho(\mathbf{x})\big\rangle$ is independent of positions $\mathbf{x}$, and therefore can be simplified as $\big\langle\rho\big\rangle$.

However, in real observations, we can only observe one universe and thus our density field $\rho(\mathbf{x})$ is only one realization. In this case, the natural choice of average value is the *volume average*:

$$\bar{\rho} \equiv \frac{1}{V}\int_V \mathrm{d}\mathbf{x}\rho(\mathbf{x}) \ , \tag{2.3}$$

and the difference between this observational volume average and the ensemble average provided by simulations and theories is called *cosmic variance* and would limit our accuracy in comparing theory against observation (Driver & Robotham, 2010). When the volume V is large enough, the averaged density field of many unconnected sub-volumes can be combined and their average approaches the true ensemble average value. This indicates that the set of those sub-volumes provide a fair sample of the density field. We call this limiting case of equivalence between ensemble and volume average ergodicity (Peacock, 1999). In the following sections, we no longer differentiate between the symbols of these two averages unless specifically mentioned.

2.1.2 PDF, moments and cumulants

The discussion above highlights the significance of the density field PDF, $P(\rho)$, at a given location. Knowing this distribution allows us to determine the probability of the density field adopting a specific amplitude. Furthermore, the time evolution of the density field PDF provides valuable insights into cosmic evolution.

According to the prediction of inflation theory such as the standard single-field slow-roll inflation, the density field distribution immediately after the inflationary period should be Gaussian (Dodelson & Schmidt, 2020). This implies that the density field PDF at that time can be fully described by a Gaussian distribution function, at least on sufficiently large scales, with its two parameters mean $\langle\rho\rangle$ and variance σ_ρ^2, where

$$\sigma_\rho^2 = \langle\rho^2\rangle - \langle\rho\rangle^2 \ , \tag{2.4}$$

and

$$P(\rho) = \frac{1}{\sigma_\rho\sqrt{2\pi}} e^{-\frac{(\rho-\langle\rho\rangle)^2}{2\sigma_\rho^2}} \ . \tag{2.5}$$

However, as we have already pointed out in Chapter. 1, nonlinear gravitational evolution in late time Universe introduces non-Gaussianities into the density field, especially on small

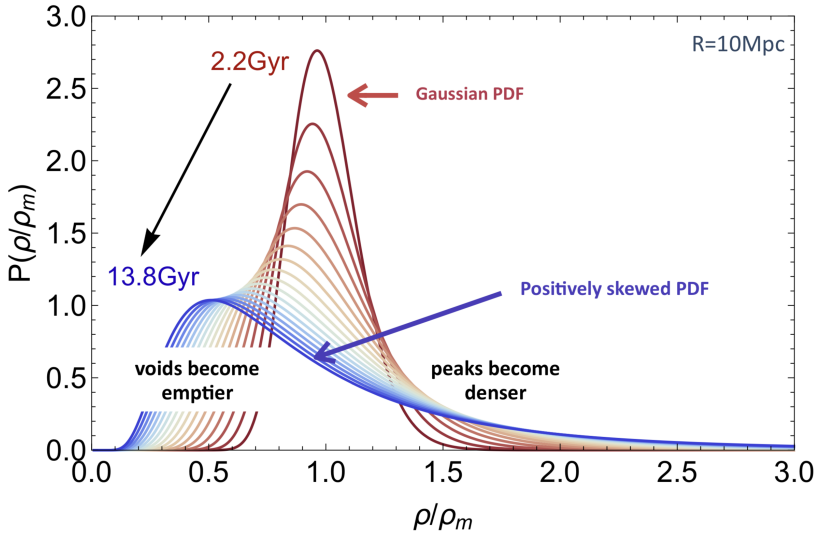

Figure 2.1: A schematic illustration depicts the time evolution of the density field PDF, starting from an initial Gaussian distribution at 2.2 Gyr after the Big Bang (dark red) and gradually transitioning to a positively skewed distribution by $z \approx 0$ (dark blue). This evolution is shown for a density field smoothed on a 10 Mpc scale. Image credit: Sandrine Codis.

scales due to the structure formation. We schematically illustrate this process in Fig. 2.1 where we show the time evolution of the density field PDF with a smoothing scale of 10 Mpc. It is clear from the figure that as redshift z approaches 0 (13.8 Gyr), the PDF becomes highly non-Gaussian with a prominent tail distribution towards high density values. This is consistent with the picture that matter is constantly attracted to the overdense regions during the evolution of our Universe.

In order to describe the density field PDF at late time, Eq. (2.5) is apparently insufficient. We need other statistical features such as moments or cumulants. By definition, an nth order moment of the density field is

$$\mu_n = \int \mathrm{d}\rho \rho^n P(\rho) \ , \tag{2.6}$$

where variance σ_ρ^2 can be related to as

$$\sigma_\rho^2 = \mu_2 - \mu_1^2 \ , \tag{2.7}$$

and is also called the second central moment. And a general definition for the nth order central moment is

$$\mu_{c,n} = \int \mathrm{d}\rho (\rho - \mu_1)^n P(\rho) \ . \tag{2.8}$$

For a Gaussian distribution, all odd-order central moments are zero and all even-order central moments can be expressed in terms of the variance σ_ρ^2. This no longer holds for

a non-Gaussian PDF; thus, higher-order moments become essential for characterizing the density field PDF in the late Universe.

Another way to compute the moments of the PDF is through the so called Moment Generating Function (MGF) which is defined as the Laplace transform of the PDF

$$M_\rho(\lambda) = \mathrm{E}\left(e^{\lambda\rho}\right) = \int \mathrm{d}\rho e^{\lambda\rho} P(\rho) \ , \tag{2.9}$$

and we can see the MGF from the series expansion of the expectation of $e^{\lambda\rho}$

$$M_\rho(\lambda) = \mathrm{E}(1 + \lambda\rho + \frac{(\lambda\rho)^2}{2!} + \ldots + \frac{(\lambda\rho)^n}{n!} + \ldots) = \sum_{n=0}^{+\infty} \frac{\lambda^n \mathrm{E}(\rho^n)}{n!} = \sum_{n=0}^{+\infty} \frac{\lambda^n \mu_n}{n!} \ , \tag{2.10}$$

from which the nth order derivative of the MGF at $\lambda = 0$ is the nth order moment μ_n.

We can define the Cumulant Generating Function (CGF) by taking the logarithm of Eq. (2.10)

$$\phi_\rho(\lambda) = \log\left(M_\rho(\lambda)\right) = \log\left(\mathrm{E}\left(e^{\lambda\rho}\right)\right) = \sum_{n=1}^{+\infty} \frac{\lambda^n K_n}{n!} \ , \tag{2.11}$$

where K_n is the nth order cumulant of the density field PDF. It is important in the context of perturbation theory as it represents the *connected* moments of the density field (e.g. 2nd order cumulant K_2 is the variance of the density field). Cumulants and moments are related to each other and below we list the first few relations

$$\begin{aligned}
\mu_1 &= K_1 \ , \\
\mu_2 &= K_2 + K_1^2 \ , \\
\mu_3 &= K_3 + 3K_1K_2 + K_1^3 \ , \\
\mu_4 &= K_4 + 4K_1K_3 + 3K_2^2 + 6K_1^2K_2 + K_1^4 \ , \\
&\vdots
\end{aligned} \tag{2.12}$$

If the CGF is obtained, the PDF can be reconstructed through an inverse Laplace transform

$$P(\rho) = \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}\lambda}{2\pi i} e^{-\lambda\rho + \phi_\rho(\lambda)} \ , \tag{2.13}$$

one way to construct the CGF for cosmological density fields is through the large deviation theory (LDT). Readers who are interested in details of the method can refer to Chapter. 8 and references therein for applications to both smoothed 3D matter density field and the projected 2D fields of weak lensing convergence and aperture mass fields.

2.2 Density perturbation and its correlation functions

From Chapter. 1 and Sec. 2.1.1, it is clear that the cosmological principle describing our Universe as a homogeneous one can only be held true at its largest length scale. Surveys

like *Sloan Digital Sky Survey* (SDSS) have revealed that our Universe has structures on relatively large scales (Dodelson & Schmidt, 2020). In other words, there are prominent inhomogeneities in the density field developed upon the smooth background as the result of the evolution of the Universe. As a fact, fluctuations were already there after inflation as reflected in the tiny anisotropies in the CMB temperature (Aghanim et al., 2020).

The values of density field ρ themselves cannot effectively display the perturbed feature in the late Universe. Therefore we define another *density perturbation field* which confronts better with our statistical measures (Peacock, 1999)

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} \ , \tag{2.14}$$

where we exploit the homogeneous property of the statistics that leads to $\langle\rho(\mathbf{x})\rangle = \langle\rho\rangle$ and from Eq. (2.14) we have $\langle\delta(\mathbf{x})\rangle = \langle\delta\rangle = (\langle\rho\rangle - \langle\rho\rangle)/\langle\rho\rangle = 0$ follow immediately. Thus it would be more reasonable to describe inhomogeneities with the variance of the density perturbation field, $\langle(\delta - \langle\delta\rangle)^2\rangle = \langle\delta^2\rangle$, rather than its average.

However, the variance $\langle\delta^2\rangle$ just indicates the strength of the inhomogeneity at a single location like what we showed in Eq. (2.4) for the density field itself. It does not take into account the spatial correlations of density perturbations among different locations. Hence we need to introduce another statistical concept named *correlation function*. Here we use the *2-point correlation function* (2PCF) ξ as an example to provide further details (Peacock, 1999)

$$\xi(\mathbf{r}) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \ , \tag{2.15}$$

where $\mathbf{r}$ is the displacement vector between two locations in space. It is positive when the density perturbations at two positions always have the same sign and negative if their signs are opposite. Essentially it investigates how the perturbations of separate locations correlate to each other. We can observe that when $\mathbf{r} = \mathbf{0}$, the correlation function becomes the variance of the density perturbation field: $\xi(\mathbf{0}) = \langle\delta(\mathbf{x})\delta(\mathbf{x})\rangle = \langle\delta^2\rangle$. From homogeneity, statistics like 2PCF would not depend on the positions of the two points but only on the separation between them. Moreover, statistical isotropy implies that its values are independent of the direction of the displacement vector between points but only dependent on the modulus of that vector. Therefore the 2PCF is a spherically symmetric quantity in 3D space and can be simplified as $\xi(\mathbf{r}) = \xi(r)$.

Besides the 3D correlation function $\xi(\mathbf{r})$, though a less direct probe, the *angular correlation function* $w(\theta)$ can be a powerful tool itself owing to large area of sky surveys.

In the case of angular correlation function, only angular positions of density perturbation fields are known but not the distances to them. Therefore when correlating perturbations of two field points in the sky we have

$$w(\hat{r}_1, \hat{r}_2) = \langle\delta(\hat{r}_1)\delta(\hat{r}_2)\rangle \ , \tag{2.16}$$

where $\hat{r}_1$ and $\hat{r}_1$ are unit vectors indicating directions. If the angular separation between two points in the sky is $\boldsymbol{\theta}$, in analogy to the definition of correlation function with statistical

homogeneity and isotropy discussed in the previous section, it can be written as

$$w(\hat{r}_1, \hat{r}_2) = \langle \delta(\hat{r}_1)\delta(\hat{r}_1 + \boldsymbol{\theta})\rangle = w(\boldsymbol{\theta}) = w(\theta) \ , \tag{2.17}$$

which depends only on the angle between two directions $\hat{r}_1$ and $\hat{r}_2$. Usually when the value of angular separation θ is within the limit required by the flat-sky approximation, the angular separation vector $\boldsymbol{\theta}$ can be expressed as a 2D vector in a flat sky plane instead of a spherical sky with respect to a chosen reference coordinate.

Following the discussion of the 2PCF, it is straightforward to extend the definition to higher-order correlation functions. The N-point correlation function (NPCF) and the angular NPCF would be in the form of $\langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}_1)\delta(\mathbf{x}+\mathbf{r}_2)\cdots\delta(\mathbf{x}+\mathbf{r}_{N-1})\rangle$ and $\langle \delta(\hat{r})\delta(\hat{r}+\boldsymbol{\theta}_1)\delta(\hat{r}+\boldsymbol{\theta}_2)\cdots\delta(\hat{r}+\boldsymbol{\theta}_{N-1})\rangle$, respectively.

2.2.1 Power spectrum

It is important to introduce the counterpart of 2PCF in Fourier space which is the power spectrum. To find the formula for power spectrum, we first need to reformulate the correlation function defined in Eq. (2.15).

The two correlated density perturbation field points can be represented by their corresponding Fourier transform and we would have:

$$\begin{aligned}
\xi(\mathbf{r}) &= \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \\
&= \frac{V^2}{(2\pi)^6}\left\langle \int_k \delta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}\mathbf{k}\int_{k'}\delta(\mathbf{k}')e^{i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})}\mathrm{d}\mathbf{k}'\right\rangle \\
&= \frac{V^2}{(2\pi)^6}\int_{k'}\int_k \delta(\mathbf{k})\delta(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}\frac{1}{V}\int_V e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{k}\mathrm{d}\mathbf{k}' \\
&= \frac{V}{(2\pi)^6}\int_{k'}\int_k \delta(\mathbf{k})\delta(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}(2\pi)^3\delta_{\mathrm{D}}(\mathbf{k}+\mathbf{k}')\mathrm{d}\mathbf{k}\mathrm{d}\mathbf{k}' \\
&= \frac{V}{(2\pi)^3}\int_{k'}\delta(-\mathbf{k}')\delta(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}\mathrm{d}\mathbf{k}' \\
&= \frac{V}{(2\pi)^3}\int_k \delta^*(\mathbf{k})\delta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}\mathbf{k} \\
&= \frac{V}{(2\pi)^3}\int_k |\delta(\mathbf{k})|^2 e^{i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}\mathbf{k} \ .
\end{aligned} \tag{2.18}$$

During the above derivation we use the volume average definition for the angular bracket, then we apply the orthogonality relation between different Fourier modes $\mathbf{k}$. The above equations show in the end that 2PCF is the Fourier transform of the power spectrum. If we now take ensemble averages on both sides of Eq. (2.18), which are indeed aims of cosmological studies, we can define the *power spectrum* as (Peacock, 1999):

$$P(\mathbf{k}) \equiv \langle |\delta(\mathbf{k})|^2\rangle \ , \tag{2.19}$$

and because power spectrum cannot contain a preferred direction in an isotropic universe, the definition can be simplified as $P(k) = \langle|\delta(k)|^2\rangle$. Alternatively, another definition of power spectrum is (Dodelson & Schmidt, 2020):

$$\langle\delta^*(\mathbf{k})\delta(\mathbf{k}')\rangle \equiv (2\pi)^3\delta_D(\mathbf{k}'-\mathbf{k})P(\mathbf{k}) \ , \tag{2.20}$$

of which Eq. (2.19) is a special case where $\mathbf{k} = \mathbf{k}'$ and the prefactor $\frac{1}{(2\pi)^3}$ in Eq. (2.18) is included in this definition as well.

The isotropy of power spectrum can lead to further simplification regarding the integration of angular components with spherical symmetry. If we take the angle between $\mathbf{r}$ and $\mathbf{k}$ as the polar angle θ in the spherical coordinate, we would have:

$$\begin{aligned}
\xi(r) &= \frac{V}{(2\pi)^3}\int_0^{2\pi}\int_0^{\pi}\int_k P(k)\cos(kr\cos\theta)k^2\sin\theta\mathrm{d}k\mathrm{d}\theta\mathrm{d}\phi \\
&= \frac{V}{(2\pi)^3}\int_0^{\pi}\int_k P(k)\cos(kr\cos\theta)\sin\theta 2\pi k^2\mathrm{d}k\mathrm{d}\theta \\
&= \frac{V}{(2\pi)^3}\int_k\left(\int_0^{\pi}\cos(kr\cos\theta)\sin\theta\mathrm{d}\theta\right)P(k)2\pi k^2\mathrm{d}k \\
&= \frac{V}{(2\pi)^3}\int_k 2j_0(kr)P(k)2\pi k^2\mathrm{d}k \\
&= \frac{V}{(2\pi)^3}\int_k P(k)\frac{\sin(kr)}{kr}4\pi k^2\mathrm{d}k \ ,
\end{aligned} \tag{2.21}$$

where the exponential factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ only has its real part, $\cos(kr\cos\theta)$, in the integration with θ going from 0 to π since correlation function in this case is a real value quantity. We also use the integral representation of the 0th order spherical Bessel function in the above integration:

$$j_0(x) = \frac{\sin x}{x} = \frac{1}{2}\int_0^{\pi}\cos(x\cos\theta)\sin\theta\mathrm{d}\theta \ . \tag{2.22}$$

The analogous formula to Eq. (2.21) in 2-dimensional space is:

$$\xi^{2D}(r) = \frac{A}{(2\pi)^2}\int_k P^{2D}(k)J_0(kr)2\pi k\mathrm{d}k \ , \tag{2.23}$$

where A is the survey area, J_0 is the 0th order Bessel function of the first kind and r, k are modulus of separation and wave vectors in 2-dimensional space.

From a theoretical perspective, it is more convenient to work in Fourier space and specify models such as power-law spectra. The Fourier or relevant transformations to real space enable prediction of correlation functions based on the theoretical models and compare those predictions with the actual correlation functions calculated from observational data. Higher-order correlation functions also have their corresponding counterpart in Fourier

space, e.g. bispectrum can be transformed to 3PCF and trispectrum can be transformed to 4PCF, following the definition analogous to Eq. (2.20)

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle_{\rm c} = (2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\ , \tag{2.24}$$

$$\langle\delta(\mathbf{k}_1)\cdots\delta(\mathbf{k}_4)\rangle_{\rm c} = (2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_4)T(\mathbf{k}_1,\ldots,\mathbf{k}_3)\ , \tag{2.25}$$

where B and T denote bispectrum and trispectrum respectively and $\langle\cdots\rangle_{\rm c}$ represents the connected part. Note that conjugating $\delta(\mathbf{k})$ on the left-hand side of the above equations changes the plus sign within the Dirac delta function on the right-hand side to a minus sign.

2.2.2 Gravitational dynamics of the density perturbation field

This section only serves as a brief introduction to the gravitational dynamics of the density perturbation field and a better preparation for the discussion in the next section. For a more detailed understanding, readers are referred to Dodelson & Schmidt (2020); Bernardeau et al. (2002) and references therein.

One key equation governing the time evolution of the matter density field in our Universe is the general collisionless Boltzmann equation

$$\frac{{\rm d}f}{{\rm d}t} = \frac{\partial f}{\partial t} + \frac{{\rm d}\mathbf{x}}{{\rm d}t}\cdot\nabla f + \frac{{\rm d}\mathbf{p}}{dt}\cdot\nabla_{\mathbf{p}} f = 0\ , \tag{2.26}$$

where f is the distribution function of the matter density field in phase space $\{\mathbf{x},\mathbf{p}\}$, where $\mathbf{x}$ is the comoving position and $\mathbf{p}$ is the physical momentum. Assuming matter is non-relativistic, from the geodesic equation, one can modify the above equation as

$$\frac{{\rm d}f}{{\rm d}t} = \frac{\partial f}{\partial t} + \frac{\mathbf{p}}{ma}\cdot\nabla f - \left[\frac{m}{a}\nabla\Phi + H\mathbf{p}\right]\cdot\nabla_{\mathbf{p}} f = 0\ , \tag{2.27}$$

where m is the mass of the matter particle, Φ is the gravitational potential, and H is the Hubble parameter. The gravitational potential Φ is related to the density perturbation field δ through the Poisson equation

$$\nabla^2\Phi = \frac{3}{2}\Omega_m(\eta)\mathcal{H}^2\delta\ , \tag{2.28}$$

where $\Omega_m(\eta)$ is the matter density parameter at cosmic conformal time η and $\mathcal{H} = \frac{{\rm d}a}{a{\rm d}\eta}$ is the conformal Hubble parameter where ${\rm d}\eta = {\rm d}t/a$. Eq. (2.27) is called the Vlasov equation and the coupled set of Eqs. (2.27) and (2.28) is known as the Vlasov-Poisson system. It describes the nonlinear evolution of the matter density field. Though it is very difficult to solve, one can still tackle it by using numerical simulations or analytical methods such as perturbation theory.

To simplify the Vlasov-Poisson system, we can take the 0th moment of the Vlasov equation w.r.t the momentum, i.e. taking the integral of $\int {\rm d}^3p/(2\pi)^3$ and then subtract the

background density $\bar{\rho}$ terms to obtain the continuity equation of the density perturbation field

$$\frac{\mathrm{d}\delta}{\mathrm{d}\eta} + \nabla \cdot (1+\delta)\mathbf{u} = 0 \ , \tag{2.29}$$

where $\mathbf{u}$ is the matter peculiar velocity field. Furthermore, by taking the 1st moment of the Vlasov equation w.r.t the momentum and subtract the continuity equation of the full density field multiplied by the velocity field gives the Euler equation

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\eta} + \mathcal{H}\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla\Phi = 0 \ , \tag{2.30}$$

where we adpt the standard assumption that the stress tensor is 0.

For convenient computation, we define the divergence of the velocity field to be $\theta = \nabla \cdot \mathbf{u}$ and take the divergence of Eq. (2.30). Eqs. (2.29) and (2.30) are then

$$\frac{\mathrm{d}\delta}{\mathrm{d}\eta} + \theta = -\delta\theta - (\mathbf{u}\cdot\nabla)\delta \ , \tag{2.31}$$

and

$$\frac{\mathrm{d}\theta}{\mathrm{d}\eta} + \mathcal{H}\theta + \nabla^2\Phi = -\nabla\cdot(\mathbf{u}\cdot\nabla)\mathbf{u} \ , \tag{2.32}$$

respectively. If we set the right-hand side of Eqs. (2.31) and (2.32) to 0, we obtain the linearized continuity and Euler equations. One can then solve the above two equations iteratively to obtain higher-order density perturbative terms.

2.2.3 Standard perturbation theory

As discussed in Sec. 2.2.1, the power spectrum and higher-order spectrum are crucial statistical measures of the density perturbation field. We can apply standard perturbation theory (SPT) which is a powerful analytical tool to compute them in early Universe and down to mildly nonlinear regime in late Universe. SPT is based on the assumption that the density fluctuations of the target fields are small ($\delta(\mathbf{x}) \ll 1$), allowing us to expand the equations governing the dynamics of these fluctuations in a series of perturbative terms.

At a given time during the evolution. The density perturbation field can be expressed as:

$$\delta(\mathbf{x}) = \delta^{(1)}(\mathbf{x}) + \delta^{(2)}(\mathbf{x}) + \delta^{(3)}(\mathbf{x}) + \cdots \ , \tag{2.33}$$

where $\delta^{(1)}$ is the linear term following a Gaussian distribution, $\delta^{(2)}$ is the second-order term, and so on. Each higher-order term approximates part of the nonlinear effects in the density perturbation field and its corresponding amplitude is much smaller than the previous order, $\delta^{(1)}(\mathbf{x}) \gg \delta^{(2)}(\mathbf{x}) \gg \delta^{(3)}(\mathbf{x}) \gg \cdots$.

The linear term $\delta^{(1)}$ evolves according to the linearized continuity and Euler equations as discussed below Eqs. (2.31) and (2.32), leading to the well-known solution:

$$\delta^{(1)}(\mathbf{x},\eta) = D_+(\eta)\delta^{(1)}(\mathbf{x},\eta_0), \tag{2.34}$$

where $\delta^{(1)}(\mathbf{x}, \eta_0)$ is the initial density perturbation field. The scale-independent growth factor $D_+(\eta)$ characterizes the linear growth of density perturbations. It is determined by the energy density parameter in our cosmological background and the equation of state parameter of dark energy (Dodelson & Schmidt, 2020).

Higher-order terms, such as $\delta^{(2)}$ and $\delta^{(3)}$, are proportional to the same order multiplication of the linear term, e.g. $\delta^{(2)} \propto (\delta^{(1)})^2$ and $\delta^{(3)} \propto (\delta^{(1)})^3$. These terms are derived by solving the perturbative expansions of the continuity and Euler equations to higher orders. To be more precise, with the linear solutions to $\delta^{(1)}$ and $\theta^{(1)}$, we can insert them into the right-hand side of Eqs. (2.31) and (2.32) and solve for $\delta^{(2)}$ and $\theta^{(2)}$ which are now on the left-hand side of the equations. This strategy can then be exploited iteratively. The second-order term $\delta^{(2)}(\mathbf{k})$ in Fourier space can be expressed as

$$\delta^{(2)}(\mathbf{k}, \eta) = D_+^2(\eta) \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3} \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) F_2(\mathbf{k}_1, \mathbf{k}_2) \delta^{(1)}(\mathbf{k}_1, \eta_0) \delta^{(1)}(\mathbf{k}_2, \eta_0), \tag{2.35}$$

where $F_2(\mathbf{k}_1, \mathbf{k}_2)$ is the second-order kernel that describes the mode coupling

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{2}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} + \frac{1}{2} \mathbf{k}_1 \cdot \mathbf{k}_2 \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) , \tag{2.36}$$

where we choose the symmetric form of the kernel w.r.t $\mathbf{k}_1$ and $\mathbf{k}_2$ as the integrand in Eq. (2.35) would become symmetric. Note that the above Eq. (2.36) is exact in an Einstein-de Sitter (EdS) Universe, however, it is reasonably accurate for many ΛCDM applications (Dodelson & Schmidt, 2020). Analogous to Eqs. (2.35) and (2.36), the higher-order density $\delta^{(n)}$ can be written as

$$\begin{aligned} \delta^{(n)}(\mathbf{k}, \eta) = D_+^n(\eta) \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3} \cdots \int \frac{\mathrm{d}^3 k_n}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_1 - \cdots - \mathbf{k}_n) \\ \times F_n(\mathbf{k}_1, \cdots, \mathbf{k}_n) \delta^{(1)}(\mathbf{k}_1, \eta_0) \cdots \delta^{(1)}(\mathbf{k}_n, \eta_0) , \end{aligned} \tag{2.37}$$

where the kernel F_n has a fully symmetric form of polynomials w.r.t its arguments and can be computed order by order iteratively (Bernardeau et al., 2002).

As an illustration of the above discussion, we derive the linear power spectrum and tree-level bispectrum using SPT. Following the definition in Eq. (2.20), the linear power spectrum $P_{\mathrm{L}}(k, \eta)$ is given by

$$P_{\mathrm{L}}(k, \eta) = \langle \delta^{(1)}(\mathbf{k}, \eta) \delta^{(1)*}(\mathbf{k}, \eta) \rangle , \tag{2.38}$$

and by solving the Possion equation (Eq. (2.28)) linearly at the reference time $\eta(a)$, we have

$$\delta^{(1)}(\mathbf{k}, a) = \frac{2}{3} \frac{k^2 a}{H_0^2 \Omega_{\mathrm{m}}} \Phi(\mathbf{k}, a) , \tag{2.39}$$

where $\Phi(\mathbf{k}, a)$ is the gravitational potential in Fourier space. It can be related to the initial condition, i.e. the primordial curvature perturbation $\mathcal{R}$ generated during inflation, as

$$\Phi(\mathbf{k}, a) = \frac{3}{5} \frac{D_+(a)}{a} \mathcal{R}(\mathbf{k}) T(k) , \tag{2.40}$$

where $T(k)$ is the transfer function that describes the evolution from the initial pertubation to the late-time matter fluctuation when the relevant constituents of the Universe were only radiation and matter. Combining Eqs. (2.39) and (2.40), we can derive the linear power spectrum as

$$P_{\rm L}(k,\eta) = \frac{4}{25}\left(\frac{k}{H_0}\right)^4 \frac{D_+^2(\eta)}{\Omega_{\rm m}^2} T^2(k) P_{\mathcal{R}}(k) \ , \tag{2.41}$$

where $P_{\mathcal{R}}(k) = (2\pi^2/k^3)\mathcal{A}_s(k/k_p)^{n_s-1}$ (Dodelson & Schmidt, 2020) is the power spectrum of the primordial curvature perturbation. We have parameters $\mathcal{A}_s$, n_s, and k_p as the amplitude, scalar spectral index, and pivot scale of the primordial power spectrum, respectively. We can further express the linear matter power spectrum as $P_{\rm L}(k,\eta) = \frac{D_+^2(\eta)}{\Omega_{\rm m}^2 H_0^4} P_{\rm init}(k)$ where the initial power spectrum $P_{\rm init}(k)$ is

$$P_{\rm init}(k) = \frac{8\pi^2}{25} T(k)^2 \mathcal{A}_s \frac{k^{n_s}}{k_p^{n_s-1}} \ . \tag{2.42}$$

The lowest order bispectrum in SPT is the tree-level bispectrum, which describes the leading-order contribution to the three-point correlation function. Following the definition in Eq. (2.24), we have the expression (here we drop the time dependence for simplicity)

$$(2\pi)^3 \delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) B_{\rm tree}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \langle \delta^{(2)}(\mathbf{k}_1)\delta^{(1)}(\mathbf{k}_2)\delta^{(1)}(\mathbf{k}_3)\rangle_{\rm c} + \text{cyc.} \ , \tag{2.43}$$

where "cyc." denotes the cyclic permutations of $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$. The second-order perturbative term has to be present as $\langle \delta^{(1)}(\mathbf{k}_1)\delta^{(1)}(\mathbf{k}_2)\delta^{(1)}(\mathbf{k}_3)\rangle_{\rm c}$ would vanish according to the Wick's theorem for the Gaussian random fields. We can substitute $\delta^{(2)}$ term in Eq. (2.43) with Eq. (2.35) and compute the tree-level bispectrum using the Wick's theorem as (Scoccimarro et al., 1998)

$$B_{\rm tree}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = 2F_2(\mathbf{k}_1,\mathbf{k}_2) P_{\rm L}(k_1) P_{\rm L}(k_2) + \text{cyc.} \ . \tag{2.44}$$

SPT provides a framework for calculating various statistical measures of the density field, such as the power spectrum and higher-order correlation functions. These calculations are essential for understanding the large-scale structure of the Universe and comparing theoretical predictions with observational data. Here we skip the discussion on the velocity field and its perturbation as they are not the main focus of this thesis. Readers interested in the details of this aspect can refer to Dodelson & Schmidt (2020) and references therein.

As a side remark, the power spectrum and bispectrum on the nonlinear scales are beyond the reach of the perturbation theory. One popular approach to compute them in this case is to build fitting functions of these statistics and calibrated them with N-body simulations. One representative example is the `Halofit` model based on the halo model (Smith et al., 2003; Takahashi et al., 2012) for the power spectrum which is widely used in the cosmological community. For bispectrum, there is also `BiHalofit` (Takahashi et al., 2020).

2.3 Weak lensing cosmology

As predicted by the theory of General Relativity, light coming to us from distant sources would be deflected by intermediate matter. This so called gravitational lensing phenomena provides us a unique tool to map the distribution of matter in our universe and constrain parameters in our cosmological models. In this section, we will introduce the basic concepts of weak gravitational lensing and its applications in cosmology which are relevant to later Chapters. 5 to 8. Readers interested in more details can refer to Bartelmann & Schneider (2001); Schneider et al. (2006); Kilbinger (2015) and references therein.

2.3.1 Weak lensing basics

In all lensing scenarios, emitted light from the source object will be deflected by the intervening matter. The deformation of a lensed image with respect to its unlensed image can be represented by a distortion matrix $\mathcal{A}$:

$$\begin{aligned}\mathcal{A} = \frac{\partial\left(\delta\boldsymbol{\beta}\right)}{\partial\left(\delta\boldsymbol{\theta}\right)} &= \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \\ &= (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos(2\phi) & \sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{pmatrix} ,\end{aligned} \tag{2.45}$$

where $\delta\boldsymbol{\beta}$ and $\delta\boldsymbol{\theta}$ are differential angular position vectors of source and images respectively as denoted in Fig. 2.2 Here we have also introduced several new quantities: κ is the convergence and γ is the modulus of the shear $\gamma = \sqrt{\gamma_1^2+\gamma_2^2}$. Shear itself is a spin-2 field and usually is written in a complex vector format $\gamma = \gamma_1 + i\gamma_2$. Due to the spin-2 transformation property, if we assume ϕ as the rotation angle counterclockwise in the plane of the sky within a Cartesian coordinate, the shear component γ_1 and γ_2 would transform as

$$\begin{aligned}\gamma_1' &= \gamma_1 \cos(2\phi) - \gamma_2 \sin(2\phi) , \\ \gamma_2' &= \gamma_1 \sin(2\phi) + \gamma_2 \cos(2\phi) ,\end{aligned} \tag{2.46}$$

instead of the usual vector rotation transformation.

The effects of distortion represented by Eq. (2.45) can be understood schematically from Fig. 2.3 in which we show how the gravitational lensing distorts the image of a source galaxy whose projected shape in the sky is assumed to be intrinsically circular. From the equation we see that convergence is only attached to an identity matrix, indicating that it is only responsible for magnifying the source image isotropically. Moreover, we have $\kappa \ll 1$ for the weak gravitational lensing. On the other hand, shear is expressed with a rotation-like matrix, implying that it would rotate and stretch the source image, i.e. introducing anisotropy into the shape of a galaxy image. Unlike strong gravitational lensing, weak lensing cannot produce multiple images but only exert the above distortion effects on a single image.

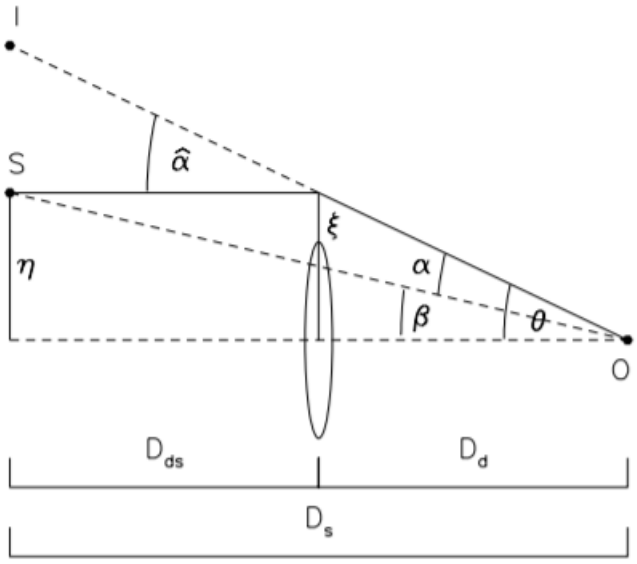

Figure 2.2: Illustration of a lensing system. Point "S", "I" and "O" denote for source, deflected image and observer respectively. At the bottom are distances between source, lens and the observer. Light emitted by a source which has a distance η from the central optical axis and propagates parallel to it is deflected by an angle $\hat{\alpha}$ and then reaches the observer. The resulting image would have an angular separation θ from the optical axis in the sky which is different from the actual separation of the source β. Image source: Narayan & Bartelmann (1996).

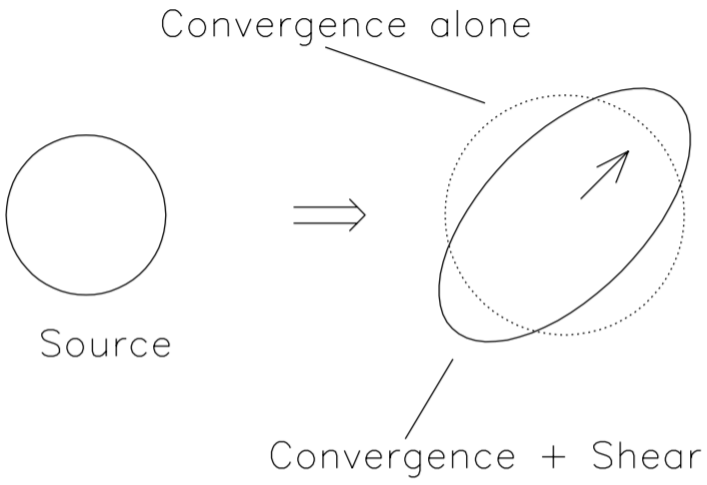

Figure 2.3: Effects of lensing distortion. For an intrinsically circular source galaxy, convergence will magnify its size isotropically and shear will turn it into an ellipse and thus introduce anisotropy in its shape. Image source: Narayan & Bartelmann (1996).

2.3.2 Weak lensing measurement

In weak lensing regime, a source object is usually a distant galaxy whose image is distorted by the foreground matter. What we can observe directly through photometric surveys such as the DES is the shape of each galaxy, characterised by its ellipticity. It is the shear field that induces ellipticities (i.e. results of the image distortion) as mentioned in Sec. 2.3.1 and there are many approaches to infer the shear field from those measured ellipticities assuming that the intrinsic galaxy shapes are on average circular.

The first question we need to address is what is the origin of the shear field. Since it is gravity that deflects the light and distorts the images, the shear field must be induced by the mass concentration between the source and the observer.

Just as the gravitational field can be expressed as the derivative of the gravitational potential, the shear field can also be written as the derivative of a corresponding lensing potential from large scale structures (Kilbinger, 2015):

$$\Psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^{\chi} \mathrm{d}\chi' \frac{f_k(\chi - \chi')}{f_k(\chi) f_k(\chi')} \Phi(f_k(\chi')\boldsymbol{\theta}, \chi') \ , \tag{2.47}$$

where χ is the radial comoving distance and $f_k(\chi)$ is the comoving angular diameter distance in a space with curvature k which would be equal to χ if the Universe is flat. Φ is the Newtonian gravitational potential and here the lensing potential at a point in space is a weighted sum of the Newtonian potential between the observer and that point. c is the speed of light and $\boldsymbol{\theta}$ is the 2D angular position in the sky.

The shear field comes from the second order derivative of the lensing potential with respect to the angular coordinates:

$$\begin{aligned} \gamma_1 &= \frac{1}{2}(\Psi_{,11} - \Psi_{,22}) \ , \\ \gamma_2 &= \Psi_{,12} \ , \end{aligned} \tag{2.48}$$

where $\Psi_{,ij} = \partial\Psi/(\partial\theta_i \partial\theta_j)$. Here the two components of the shear γ_1 and γ_2 are specified w.r.t a chosen Cartesian coordinate at a given location $\boldsymbol{\theta}$ in a 2D flat sky as discussed in Sec. 2.3.1. However, we can also decompose the shear field into two components defined based on the angular position[2] $\boldsymbol{\theta}$ w.r.t a fixed reference point in the sky: the tangential shear γ_t and cross shear $\gamma_\times$. Their relations to the γ_1 and γ_2 are:

$$\begin{aligned} \gamma_\mathrm{t} &= -\gamma_1 \cos 2\vartheta - \gamma_2 \sin 2\vartheta \ , \\ \gamma_\times &= -\gamma_1 \sin 2\vartheta + \gamma_2 \cos 2\vartheta \ . \end{aligned} \tag{2.49}$$

A clear explanation can be seen in Fig. 2.4. If the matter distribution between the source and the observer has the overdensity and spherical symmetry, the weak lensing distortion should perfectly align the lensed galaxy images along the tangential direction with respect

[2]In 2D polar coordinates, the angular posotion $\boldsymbol{\theta}$ can be decomposed into a radial length θ and a polar angle ϑ with respect to a reference point.

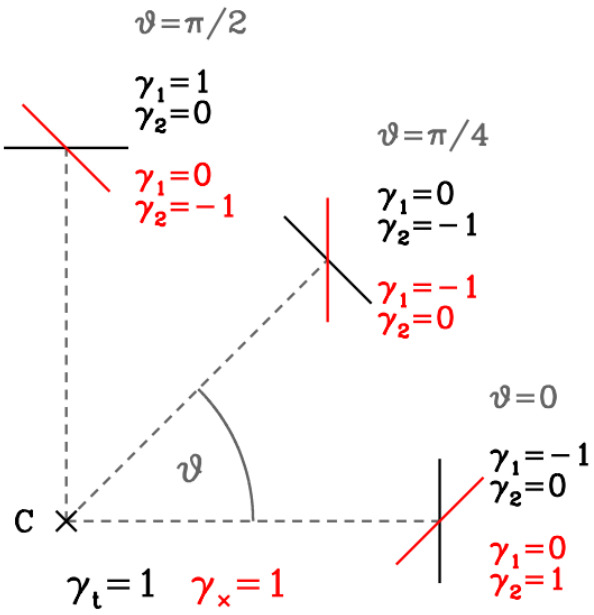

Figure 2.4: C is the central reference point. Tangential shear γ_t is defined perpendicular to or along the line connecting the field and the reference point. Cross shear $\gamma_\times$ is defined with the directions of tangential shear rotating 45°. In this figure, the direction perpendicular to the radial direction is defined as the positive axis for the tangential shear and after rotating it 45° clockwise the positive axis for cross shear is defined. Therefore at three different angular positions, the tangential and cross shear which are represented by the black and red bars are of constants. However, the values of corresponding γ_1 and γ_2 are different following the transformations shown in Eq. (2.46). Image source: Gruen (2015).

to the center of mass concentration and the cross shear will be zero. Analogously, an underdense and spherically symmetric distribution would align the lensed galaxy images along the radial direction with respect to the center of mass concentration and the cross shear will still be zero.Another issue to notice is that the same value of tangential and cross shear correspond to different values of γ_1 and γ_2 for field points at different angular positions in the sky.

Aside from the induced ellipticity by the shear field, the unlensed galaxy image would have an intrinsic ellipticity from its original shape projected in the sky. Assuming there are no additional distortion from observational systematic errors, the relation between the observed ellipticity $e^{\rm obs}$ and the intrinsic ellipticity $e^{\rm int}$ of a galaxy is:

$$e^{\rm obs} = \frac{e^{\rm int} + \gamma}{1 + \gamma^* e^{\rm int}} \ , \tag{2.50}$$

where γ^* represents the complex conjugate of the shear field. The ellipticity itself depends on the light distribution of a galaxy and there can be various definitions (Schneider et al.,

2006). It is infeasible in weak lensing regime to probe the shear field at a point by measuring the ellipticity of just one neighboring galaxy since both observed and intrinsic ellipticities have high-level noises and the results would be highly inaccurate. However, based on one critical assumption that unlensed galaxies are oriented randomly, if we measure sufficiently large number of galaxies around that field point and take the ensemble average of the observed ellipticities, the ensemble average of intrinsic ellipticities would vanish due to their random orientations and we are left with $\langle e^{\rm obs}\rangle \approx \gamma$. This is only an approximation and what scientists actually measure from ellipticities is a quantity called reduced shear g

$$g = \frac{\gamma}{1-\kappa} \ , \tag{2.51}$$

which in the limit of weak lensing regime would become $g \approx \gamma$ since $\kappa \ll 1$.

Like the shear field, the convergence field can be derived from the second order derivative of the lensing potential as well:

$$\kappa = \frac{1}{2}(\Psi_{,11} + \Psi_{,22}) \ . \tag{2.52}$$

We can substitute the lensing potential in Eq. (2.47) with the expression in Eq. (2.52). By introducing the Poisson equation in comoving coordinates, we are able to write the convergence at a given point in space as:

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^\chi \mathrm{d}\chi' \frac{f_k(\chi-\chi')f_k(\chi')}{f_k(\chi)} \frac{\delta_m(f_k(\chi')\boldsymbol{\theta}, \chi')}{a(\chi')} \ , \tag{2.53}$$

where $\delta_m(f_k(\chi)\boldsymbol{\theta}, \chi)$ is the matter density perturbation field in 3D space. One nice interpretation of the equation above is: If a single light source is located at comoving distance χ with angular position $\boldsymbol{\theta}$, then the lensing convergence that a light ray bundle from the source experiences on its way towards the observer is given by a weighted sum of the contributions from all the matter density distribution between the observer and the source.

We can further assume that multiple source objects follow a (normalized) redshift distribution $p_s(z)$ along the line-of-sight with the corresponding distribution in comoving distance being $n_s(\chi)$, i.e. $n_s(\chi)\mathrm{d}\chi = p(z)\mathrm{d}z$. In this case, the effective convergence towards direction $\boldsymbol{\theta}$ becomes:

$$\kappa_{\rm eff}(\boldsymbol{\theta}) = \int_0^{\chi_{max}} \mathrm{d}\chi \ n_s(\chi)\kappa(\boldsymbol{\theta}, \chi) \tag{2.54}$$

$$= \frac{3H_0^2\Omega_m}{2c^2} \int_0^{\chi_{max}} \mathrm{d}\chi \ g(\chi) f_k(\chi) \frac{\delta_m(f_k(\chi)\boldsymbol{\theta}, \chi)}{a(\chi)} \ , \tag{2.55}$$

where χ_{max} is the maximal radial comoving distance (e.g. the horizon of the observable Universe) and $g(\chi)$ reads:

$$g(\chi) = \int_\chi^{\chi_{max}} \mathrm{d}\chi' n_s(\chi') \frac{f_k(\chi'-\chi)}{f_k(\chi')} \ . \tag{2.56}$$

From the above discussion, we know that both shear and convergence are derived from the second order derivative of the lensing potential. Therefore, it is clear that shear and convergence are not independent from each other. It is easier to write down their relation in Fourier space where the 2D angular position $\boldsymbol{\theta}$ transforms to the multipole $\boldsymbol{\ell}$ (Kaiser & Squires, 1993):

$$\gamma(\boldsymbol{\ell}) = \frac{(\ell_x + \mathrm{i}\ell_y)^2}{\ell^2}\kappa(\boldsymbol{\ell}) = e^{2\mathrm{i}\phi_\ell}\kappa(\boldsymbol{\ell}) \ , \qquad \ell \neq 0 \tag{2.57}$$

where $\ell = \sqrt{\ell_x^2 + \ell_y^2}$ and $\phi_\ell = \arctan\left(\frac{\ell_y}{\ell_x}\right)$ is the polar angle of $\boldsymbol{\ell}$.

2.3.3 Weak lensing shear 2PCF and integrated 3PCF

By definition (Schneider & Lombardi, 2003), the 2PCF of shear has two branches: one is the plus shear 2PCF, ξ_+ and the other is the minus shear 2PCF, ξ_-.

Suppose we have two points in the sky with their respective 2D angular positions $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ and these two points are separated by a separation vector $\boldsymbol{\alpha}$. We can write down the tangential and cross shear at each of these two positions as $\gamma_{\mathrm{t},1}$, $\gamma_{\mathrm{t},2}$ and $\gamma_{\times,1}$, $\gamma_{\times,2}$. Then the shear 2PCFs are defined as:

$$\xi_+(\boldsymbol{\alpha}) \equiv \langle\gamma_{\mathrm{t},1}\gamma_{\mathrm{t},2}\rangle + \langle\gamma_{\times,1}\gamma_{\times,2}\rangle = \langle\gamma(\boldsymbol{\theta}_1)\gamma^*(\boldsymbol{\theta}_2)\rangle \tag{2.58}$$

$$\xi_-(\boldsymbol{\alpha}) \equiv \langle\gamma_{\mathrm{t},1}\gamma_{\mathrm{t},2}\rangle - \langle\gamma_{\times,1}\gamma_{\times,2}\rangle = \langle\gamma(\boldsymbol{\theta}_1)\gamma^*(\boldsymbol{\theta}_2)e^{-4i\phi_\alpha}\rangle \ , \tag{2.59}$$

where ϕ_α is the polar coordinate of the separation vector $\boldsymbol{\alpha}$. If we apply the relation between Fourier space 2D shear and convergence in Eq. (2.57) to the definition of power spectrum in Eq. (2.20), we can obtain the relation between 2D power spectrum of shear and convergence

$$P_\gamma^{2D}(\boldsymbol{\ell}) = \langle\gamma^*(\boldsymbol{\ell})\gamma(\boldsymbol{\ell})\rangle = \langle e^{-2i\phi_\ell}\kappa^*(\boldsymbol{\ell})e^{2i\phi_\ell}\kappa(\boldsymbol{\ell})\rangle = P_\kappa^{2D}(\boldsymbol{\ell}) \ . \tag{2.60}$$

Based on Eq. (2.60), we can express the shear 2PCFs in terms of convergence power spectrum. Note here that in ξ_- we have this extra phase factor and its counterpart in Fourier space would be $e^{-4i\phi_\ell}$. In conclusion, the shear 2PCFs can be written down as

$$\xi_+(\alpha) = \mathcal{F}^{-1}\left[P_\kappa^{2D}(\ell)\right] = \int \frac{\mathrm{d}\ell\ell}{2\pi}P_\kappa^{2D}(\ell)J_0(\ell\alpha) \tag{2.61}$$

$$\xi_-(\alpha) = \mathcal{F}^{-1}\left[P_\kappa^{2D}(\ell)e^{-4i\phi_\ell}\right] = \int \frac{\mathrm{d}\ell\ell}{2\pi}P_\kappa^{2D}(\ell)J_4(\ell\alpha) \ , \tag{2.62}$$

where J_0 and J_4 are the zeroth and fourth-order Bessel functions of the first kind respectively.

In practice, one actually estimates the shear 2PCFs in Eqs. (2.58) and (2.59) from the observed ellipticities of galaxies and they are measured as follows (Troxel et al., 2018)

$$\hat{\xi}_\pm(\alpha) = \frac{\sum_{i,j} w_i w_j(\gamma_\mathrm{t}^i\gamma_\mathrm{t}^j \pm \gamma_\times^i\gamma_\times^j)}{\sum_{i,j} w_i w_j} \tag{2.63}$$

where i, j are the indices of a pair of observed galaxies which are separated by an angular separation α. The w_i are the weights associated to each observed galaxy.

From Eq. (2.55), we see that convergence is a weighted line-of-sight projection of matter density perturbation. If we combine it with the projection of 3D power spectrum

$$P^{2D}(\boldsymbol{\ell}) = \int \mathrm{d}\chi \; \frac{q^2(\chi)}{\chi^2} \; P^{3D}\left(|\boldsymbol{k}| = \frac{|\boldsymbol{\ell}|}{\chi}, \chi\right) \; , \tag{2.64}$$

where $q(\chi)$ is the weighting kernel, and we applied the Limber approximation (Limber, 1954; Kaiser, 1992) to the integral, it is straightforward to observe that the 2D convergence power spectrum P_κ^{2D} is a weighted line-of-sight projection of the 3D matter power spectrum P_δ^{3D}. It is equivalent to state that weak lensing convergence and shear directly traces the matter density perturbation in the Universe.

The weighting kernel q_κ in Eq. (2.64) for lensing is (usually referred to as the lensing kernel)

$$q_\kappa(\chi) = \frac{3H_0^2\Omega_m}{2c^2} g(\chi) \frac{f_k(\chi)}{a(\chi)} \; , \tag{2.65}$$

where $g(\chi)$ and $f_k(\chi)$ are the same as in Eqs. (2.55) and (2.56). Then for general case where we have $q_\kappa^i(\chi)$ and $q_\kappa^j(\chi)$ associated with different source redshift distributions, we have the following expression

$$P_\kappa^{2D,ij}(\ell) = \int \mathrm{d}\chi \frac{q_\kappa^i(\chi) q_\kappa^j(\chi)}{\chi^2} P_\delta^{3D}\left(|\boldsymbol{k}| = \frac{|\boldsymbol{\ell}|}{\chi(z)}, z\right) \; , \tag{2.66}$$

where i, j denote tomographic bins of the background source galaxies.

As discussed in Sec. 2.1.2, 2PCFs alone are not sufficient to fully describe the statistical properties of the lensing field in the late-time Universe and on the scales that we are interested in. The 3PCF is the next higher-order statistic that can provide more information about the non-Gaussianity of the lensing field (Schneider & Lombardi, 2003; Takada & Jain, 2004). In a series of papers (see Halder et al. (2021); Halder & Barreira (2022); Halder et al. (2023), including Gong et al. (2023) as Chapter. 6), we have developed a new specific higher-order statistic called the integrated 3PCF (i3PCF) of the projected density fields including weak lensing and galaxy number density. For the purpose of simplicity, here we only discuss its application to weak lensing.

The i3PCF of the weak lensing field can be interpreted as the correlation between (i) a position dependent 1-point aperture mass statistic with (ii) the position-dependent shear 2PCF measured within the same aperture size (for a schematic illustration, readers can refer to Fig. 1 in Halder et al. (2023)). The aperture mass is defined as the convolution of the tangential shear component w.r.t the center of a circular aperture with a specific filter function called the compensated filter function (Eq. (2.3) in Chapter. 6). We skip the derivation of the shear i3PCF here and directly quote the result

$$\zeta_\pm^{ijk}(\boldsymbol{\alpha}) = \langle M_{\rm ap}^i(\boldsymbol{\theta}_C) \xi_\pm^{jk}(\boldsymbol{\alpha}; \boldsymbol{\theta}_C) \rangle = \frac{1}{A_{\rm 2pt}(\alpha)} \int \frac{\mathrm{d}\ell\ell}{2\pi} \mathcal{B}_{\pm,ijk}^{2D}(\boldsymbol{\ell}) J_{0/4}(\ell\alpha) \; , \tag{2.67}$$

where $A_{\rm 2pt}$ is the area normalization factor and $\boldsymbol{\theta}_C$ is the central position of the aperture. Indices i, j, k denote the tomographic bins of the background source galaxies like those in Eq. (2.66). The 2D integrated bispectrum $\mathcal{B}^{2D}_{\pm,ijk}$ is

$$\mathcal{B}^{2D}_{\pm,ijk}(\boldsymbol{\ell}) = \int \mathrm{d}\chi \frac{q^i_\kappa(\chi) q^j_\kappa(\chi) q^k_\kappa(\chi)}{\chi^4} \int_{\boldsymbol{\ell}_1}\int_{\boldsymbol{\ell}_2} B^{3D}_\delta\left(\frac{\boldsymbol{\ell}_1}{\chi}, \frac{\boldsymbol{\ell}_2}{\chi}, \frac{-\boldsymbol{\ell}_{12}}{\chi}; \chi\right) e^{2i(\phi_{\boldsymbol{\ell}_2} \mp \phi_{-\boldsymbol{\ell}_{12}})} \times U(\boldsymbol{\ell}_1) W(\boldsymbol{\ell}_2 + \boldsymbol{\ell}) W(-\boldsymbol{\ell}_{12} - \boldsymbol{\ell}) \,, \tag{2.68}$$

where $\int_{\boldsymbol{\ell}}$ represents the integration form of $\int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2}$ and $-\boldsymbol{\ell}_{12}$ is an abbreviation of $-\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2$. U and W are compensated and top-hat filters respectively.

Compared to the complete shear 3PCF, the i3PCF is more computationally efficient and easier to measure from the data. As it does not require a new 3-point estimator, but only the existing 1-point and 2-point estimators for the aperture mass and position-dependent shear 2PCF separately. From the modeling perspective, shear i3PCF is more sensitive to the squeezed limit bispectrum ($\mathbf{k}_1 \ll \mathbf{k}_2 \approx \mathbf{k}_3$ in $B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$), especially the ζ_- component. This enables us to model the small scale shear i3PCF with the response approach to perturbation theory (Barreira & Schmidt, 2017; Halder & Barreira, 2022) which is a rigorous approach for computing higher-order correlation functions in squeezed limit configurations. This method brings in another advantage that we can incorporate baryonic feedback effects into the model of the shear i3PCF via power spectrum, which is better studied and calibrated compare to the full bispectrum fitting functions on small scales.

2.4 Bias theory in large-scale structure

As we have pointed out multiple times in Sec. 2.3, the weak lensing field is a direct tracer of the matter density perturbation in the Universe. However, another important observable, the galaxy and its distribution are not a perfect tracer of the underlying total matter density field. The galaxy distribution is *biased* w.r.t the matter distribution. Moreover, geometric features such as extrema in the weak lensing field including peaks (maxima), voids (minima) and saddle points are also biased tracers of the matter density field.

The concept of bias is a crucial aspect in the study of large-scale structure and cosmology. It describes the statistical relation between the distribution of biased tracers (e.g. galaxy, weak lensing extrema) and matter. In this section, we will introduce the concept of bias and its applications to prepare for discussions in Chapter. 5. Readers interested in the topic are highly recommended to refer to Desjacques et al. (2018) and references therein.

2.4.1 Galaxy bias

As a complex problem, the essence of the galaxy bias can be summarized as describing the observed statistics of galaxies such as the galaxy 2PCF $\xi_g(r)$, on a certain range of scales,

in terms of a finite number of expressions together with their associated bias parameters. The most general form of the galaxy number density perturbation is[3]

$$\delta_g(\mathbf{x},\eta) = \sum_O b_O(\eta) O(\mathbf{x},\eta) + \epsilon(\mathbf{x},\eta) + \sum_O \epsilon_O(\mathbf{x},\eta) O(\mathbf{x},\eta) \ , \tag{2.69}$$

where O denote different operators composed of the matter density perturbation $\delta(\mathbf{x},\eta)$ and tidal field; b_O, as coefficients in the expansion, are the corresponding bias parameters, and ϵ and ϵ_O are the stochastic terms. From the above equation, we can already interpret the bias parameters to be the response of the tracers to the large-scale perturbations described by the operators O. We need the stochastic terms to account for the indeterministic process in the galaxy formation and the fields ϵ and ϵ_O are uncorrelated with O and among themselves.

From an abstract perspective, the galaxy bias expansion in Eq. (2.69) aims to include all operators that are consistent with the symmetries of the galaxy number density perturbation. Each new operator corresponds to an additional degree of freedom, typically associated with a specific physical process influencing galaxy formation and clustering. As a simple introduction, we only consider local operators of matter density perturbation $\delta(\mathbf{x},\eta)$ and tidal field K_{ij}

$$K_{ij} = \frac{2}{3\Omega_{\rm m}\mathcal{H}^2}\partial_i\partial_j\Phi - \frac{1}{3}\delta_{ij}\delta \ , \tag{2.70}$$

where Φ is the Newtonian gravitational potential and δ_{ij} is the Kronecker delta. We do not show other types of operators such as the convective time derivative of the tidal field $K_{ij}\frac{DK^{ij}}{D\eta}$ which encodes the impact of the tidal forces on galaxies evolution as they move through the large-scale structure ($\frac{D}{D\eta} = \frac{\partial}{\partial\eta} + \mathbf{u}\cdot\nabla$ is the convective time derivative); or the higher-order spatial derivatives of $\delta(\mathbf{x},\eta)$ and K_{ij} such as $\nabla^2\delta$ and $\nabla^2 K_{ij}K^{ij}$ which are related to the non-local effects of the large-scale structure on galaxy formation and clustering as well as all non-gravitational physics influencing the galaxy formation process. Note that we assume the Einstein summation convention in the above expressions.

Due to the symmetry argument discussed above, K_{ij} alone, which is a rank-2 tensor, cannot be an operator in the leading-order of the bias expansion of a scalar quantity galaxy number density perturbation. Therefore we can write down the incomplete galaxy bias expansion in Eulerian space up to the next-to-leading order as

$$\begin{aligned}\delta_g(\mathbf{x},\eta) =& b_1(\eta)\delta(\mathbf{x},\eta) + \frac{b_2(\eta)}{2}\delta^2(\mathbf{x},\eta) + b_K(\eta)K_{ij}(\mathbf{x},\eta)K^{ij}(\mathbf{x},\eta) \\ &+ \epsilon(\mathbf{x},\eta) + \epsilon_1(\mathbf{x},\eta)\delta(\mathbf{x},\eta) + \epsilon_K(\mathbf{x},\eta)K_{ij}(\mathbf{x},\eta)K^{ij}(\mathbf{x},\eta) \ ,\end{aligned} \tag{2.71}$$

where b_1 and b_2 are the linear and quadratic bias parameters respectively, and b_K is the tidal bias parameter. The stochastic terms ϵ, ϵ_1 and ϵ_K are those indicated in Eq. (2.69). The full expansion can be more complicated and readers are referred to Desjacques et al. (2018) for a more detailed discussion.

[3]Here for simplicity we will not discuss the galaxy velocity bias

Furthermore, each operator in the bias expansion can be approximated using perturbation theory as discussed in Sec. 2.2.3. With this fact, there is now a more complete picture of the galaxy bias expansion mathematically: On top layer we have the bias expansion such as the example in Eq. (2.71) where each operator is ordered based on its perturbative scaling or derivative counting relative to leading terms; Below that we have the perturbative expansion of each operator and thus a single bias parameter can be associated with different perturbative orders of the same operator.

To give a concrete example, for Eq. 2.71, only $b_1(\eta)\delta^{(1)}(\mathbf{x},\eta)$ and $\epsilon(\mathbf{x},\eta)$ are in the linear perturbative order. Operator δ^2 can only start from the second order of perturbation theory $(\delta^{(1)})^2$ and the same for the tidal field. Therefore at linear level, we have

$$\delta_g^{(1)}(\mathbf{x},\eta) = b_1(\eta)\delta^{(1)}(\mathbf{x},\eta) + \epsilon(\mathbf{x},\eta) \ . \tag{2.72}$$

With the above equation, in Fourier space we can write the linear galaxy power spectrum as

$$P_g^L(k,\eta) = b_1^2(\eta)P_\delta^L(k,\eta) + P_\epsilon(k,\eta) \ , \tag{2.73}$$

where P_δ^L is the linear matter power spectrum discussed in Sec. 2.2.3 and P_ϵ is the power spectrum of the stochastic term ϵ and should be scale-indepedent. It is often approximated as a Poisson noise though it can be more complicated in reality (e.g. super-Poissonian or sub-Poissonian depending on the tracer under investigation) (Britt et al., 2024; Pezzotta et al., 2024).

2.4.2 Extrema bias in weak lensing field

Extrema in the weak lensing field such as peaks, voids and saddle points are constrained fields compared to the convergence κ alone which traces the matter density distribution directly. This is because to identify extrema, we need to consider not only the amplitude α of the convergence field but also the gradient η_i and curvature ζ_{ij}. The latter two have imposed additional constraints on the convergence field and thus make extrema biased tracers of the convergence as well as matter density field. Note that here η and ζ are not the conformal time and i3PCF we discussed in the previous sections. We adopt the same notations to be consistent with the expression in Chapter. 5. They are defined in a normalized way

$$\alpha = \frac{f}{\sigma_0} \ , \quad \eta_i = \frac{\partial_i f}{\sigma_1} \ , \quad \zeta_{ij} = \frac{\partial_i\partial_j f}{\sigma_2} \ , \tag{2.74}$$

where f is a random 2D density field (e.g. weak lensing convergence κ) and σ_n acts as a normalization constant and is defined as the spectral moment of the field

$$\sigma_n^2 = \int \frac{\mathrm{d}^2k}{(2\pi)^2} k^{2n} P(k) \ , \tag{2.75}$$

in which $P(k)$ is the power spectrum of the 2D random field.

For an extremum to exist in a 2D weak lensing field, the local gradient must vanish and the local Hessian matrix must be negative definite for a peak, positive definite for a void and indefinite for a saddle point. Using peak as an example, its number density is a functional of the convergence field (Kac, 1943; Rice, 1945; Longuet-Higgins, 1957; Bardeen et al., 1986)

$$n_p(\nu) = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \Theta(\alpha - \nu)\delta_{\rm D}(\boldsymbol{\eta})\Theta(\lambda_2)|\det\zeta| \ , \tag{2.76}$$

where we consider all peaks with an amplitude above the threshold ν. The constraint on the gradient is imposed by the Dirac delta function $\delta_{\rm D}(\eta)$ and the constraint on the curvature is imposed by the Heaviside step function Θ whose argument λ_2 is the smallest eigenvalue of the negative Hessian matrix $-\zeta_{ij}$. It is straightforward to change the above expression to voids and saddle points by changing the sign of the curvature constraint.

Like the galaxy bias expansion in Eq. (2.71), the extremum number density such as the one in Eq. (2.76) can also be expanded in terms of the operators (or basis as we call them in Chapter. 5) and the associated bias parameters. However, unlike the galaxy number density, the extremum number density has a clearly defined analytical expression. The reason why we still need the bias expansion is that the probability density functional of the weak lensing field in Eulerian space at late-time Universe is not known and we cannot directly calculate the power spectrum as

$$\frac{\langle\tilde{n}_i(\mathbf{k})\tilde{n}_j(\mathbf{k}')\rangle_{\rm c}}{\langle n_i\rangle\langle n_j\rangle} = (2\pi)^2\delta_{\rm D}(\mathbf{k}+\mathbf{k}')P^{ij}_{\rm ext}(k) \ , \tag{2.77}$$

where i, j denote different types of extrema and $\tilde{n}_i$ is the Fourier transform of the extremum number density n_i. In order to systematically calculate the power spectrum we use the Gram-Charlier A (GCA) series expansion (Blinnikov & Moessner, 1998) to perturbatively expand the probability density functional w.r.t Gaussian distribution (readers can refer to Appendix. A in Chapter. 5 for a detailed derivation). This is analogous to the approach in SPT where we build the clustering of the higher-order perturbative terms of the density perturbation field on linear matter power spectrum which is analytically explicit w.r.t the initial Gaussian distribution. Consistently we expand the extremum number density in terms of the same basis operator, the generalized Wiener-Hermite polynomials, used in the GCA series expansion[4].

To give a conclusion, under this mathematical framework, the bias expansion of the extremum number density in Fourier space is

$$\begin{aligned}\tilde{n}_i(\mathbf{k}) = \sum_{m=0}^{\infty}\frac{1}{m!}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_m}{(2\pi)^2}(2\pi)^2\delta^2_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_m-\mathbf{k}) \\ \times\,\mathcal{G}^i_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)\mathcal{H}^\star_m(\mathbf{k}_1,\ldots,\mathbf{k}_m) \ ,\end{aligned} \tag{2.78}$$

where $\mathcal{H}^\star_m$ is the basis operator of the dual generalized Wiener-Hermite polynomial and $\mathcal{G}_m$ is the scale-dependent bias parameter associated with the operator $\mathcal{H}^\star_m$ which can be

[4]To be precise, we used the dual basis of those in the GCA series expansion

expressed as

$$\mathcal{G}^i_m(\mathbf{k}_1,\ldots,\mathbf{k}_m) = (2\pi)^{2m}\langle\frac{\partial^m n_i(\mathbf{x})}{\partial\tilde{f}(\mathbf{k}_1)\cdots\partial\tilde{f}(\mathbf{k}_m)}\rangle_{\rm G}\ , \tag{2.79}$$

where $\tilde{f}$ is the Fourier transform of the weak lensing field f in Eq. (2.74) and $\langle\ldots\rangle_{\rm G}$ denotes the average over the Gaussian field. Based on Eq. (2.79), the bias parameters $\mathcal{G}_m$ are the response of the extrema to the large-scale perturbations which share the same interpretation as the galaxy bias parameters discussed in Sec. 2.4.1. Readers can refer to Appendix. B in Matsubara (2020) for the calculation of Eq. (2.79) and the results for the first few orders.

Below we skip the derivation which is presented in Chapter. 5 and directly quote the result of the extrema power spectrum up to the next-to-next-to-leading order using the perturbative bias expansion approach discussed above

$$\begin{aligned}
&P^{ij}_{\rm ext}(k) = g^i_1(\mathbf{k})g^j_1(\mathbf{k})P(k)\\
&+\frac{1}{2}\int\frac{{\rm d}^2k_1}{(2\pi)^2}g^i_2(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)g^j_2(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)P(k_1)P(|\mathbf{k}-\mathbf{k}_1|)\\
&+\frac{1}{2}\left[g^i_1(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}g^j_2(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)B(-\mathbf{k},\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)+(i\leftrightarrow j)\right]\\
&+\frac{1}{6}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^i_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)g^j_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)P(k_1)P(k_2)P(|\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|)\\
&+\frac{1}{6}\left[g^i_1(\mathbf{k})P(k)\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^j_4(\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)B(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{1}{2}\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^i_2(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)g^j_3(\mathbf{k}-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)P(|\mathbf{k}-\mathbf{k}_1|)B(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{1}{6}\left[g^i_1(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^j_3(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)T(-\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{1}{4}\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^i_2(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1)g^j_2(\mathbf{k}-\mathbf{k}_2,\mathbf{k}_2)T(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}-\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&-\frac{1}{6}\left[g^i_1(\mathbf{k})g^j_1(\mathbf{k})P(k)\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}g^j_3(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)B(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)+(i\leftrightarrow j)\right]\ ,
\end{aligned} \tag{2.80}$$

where

$$g_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)\equiv\frac{\mathcal{G}_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)}{\mathcal{G}_0}\ , \tag{2.81}$$

and the symbol $(i\leftrightarrow j)$ denotes the addition of a term possessing the same form but exchanging the correspondence of the expansion coefficient function $\mathcal{G}_m$ to the other extremum type. P, B and T on the right-hand side of the Eq. (2.80) are the weak lensing convergence power spectrum, bispectrum and trispectrum respectively, which can be calculated either through perturbation theory or existing fitting functions down to small scales.

2.5 Statistical analysis with multiprobe cosmology

The overlap among different cosmological probes is a powerful tool to break the degeneracies in the cosmological parameter space. The joint analysis of weak lensing, galaxy clustering, CMB lensing, thermal Sunyaev-Zeldovich (tSZ), galaxy clusters and so on can provide a more stringent constraint on the cosmological parameters compared to the analysis of each probe individually.

Here we would like to briefly discuss the approach of 3×2-point analysis of weak lensing and galaxy clustering. It is built on the observation of weak lensing shear and projected galaxy number density, and it has become a standard routine analysis in the current Stage-III galaxy surveys (see (Heymans et al., 2021; Abbott et al., 2022; Sugiyama et al., 2023) to name a few). This approach exploits both the perturbation theory we discussed in Sec. 2.2 and the bias theory in Sec. 2.4. The 3×2-point consists of weak lensing shear 2PCF (Sec. 2.3.3), cross 2PCF between source galaxy shear and lens galaxy position (galaxy-galaxy lensing), and projected galaxy clustering 2PCF.

Like Eqs. (2.61), (2.62) and (2.66), the cross 2PCF between source galaxy shear and lens galaxy position can be expressed as

$$\xi_{\rm t}^{ij}(\alpha) = \langle \delta_g^{2D,i}(\boldsymbol{\theta}) \gamma_{\rm t}^{j}(\boldsymbol{\theta}+\boldsymbol{\alpha}) \rangle = \int \frac{{\rm d}\ell\ell}{2\pi} P_{\rm t}^{ij}(\ell) J_2(\ell\alpha) \ , \tag{2.82}$$

where

$$P_{\rm t}^{ij}(\ell) = \int {\rm d}\chi \frac{q_g^i(\chi) q_\kappa^j(\chi)}{\chi^2} P_{g\delta}^{3D}\left(|\boldsymbol{k}| = \frac{|\boldsymbol{\ell}|}{\chi}, \chi\right) \ . \tag{2.83}$$

In the above two equations, i, j are indices that run over tomographic bins of source or lens galaxies. γ_t represents the tangential shear (check Sec. 2.3.2 for details) w.r.t the lens galaxy located at $\boldsymbol{\theta}$. J_2 is the second order Bessel function of the first kind. $P_{g\delta}^{3D}$ is the galaxy-matter cross power spectrum and q_g is the galaxy kernel which is similar to the lensing kernel q_κ in Eq. (2.65)

$$q_g(\chi) = \frac{n_g(z(\chi))}{\bar{n}_g} \frac{{\rm d}z}{{\rm d}\chi} \ , \tag{2.84}$$

where n_g is the galaxy number density distribution of a given tomographic bin and $\bar{n}_g$ is the corresponding mean galaxy number density.

Following Eqs. (2.82) and (2.83), we can write down the galaxy clustering 2PCF as

$$\xi_g^{ij}(\alpha) = \langle \delta_g^{2D,i}(\boldsymbol{\theta}) \delta_g^{2D,j}(\boldsymbol{\theta}+\boldsymbol{\alpha}) \rangle = \int \frac{{\rm d}\ell\ell}{2\pi} P_g^{ij}(\ell) J_0(\ell\alpha) \ , \tag{2.85}$$

and

$$P_g^{ij}(\ell) = \int {\rm d}\chi \frac{q_g^i(\chi) q_g^j(\chi)}{\chi^2} P_{gg}^{3D}\left(|\boldsymbol{k}| = \frac{|\boldsymbol{\ell}|}{\chi}, \chi\right) \ , \tag{2.86}$$

where P_{gg}^{3D} is the galaxy auto power spectrum. In the above equations, we can use the perturbative bias expansion discussed in Sec. 2.4.1 to model the galaxy power spectrum P_{gg}^{3D} and the galaxy-matter cross power spectrum $P_{g\delta}^{3D}$.

Besides the 3×2-point correlation functions, multiprobe cosmology can include statistics (typically 2PCF) of other observables to jointly constrain important cosmological parameters such as $\Omega_{\rm m}$ and S_8. One such example is CMB lensing and its cross correlation with weak lensing shear and galaxy number density perturbation, to construct the so called 10×2-point or even 12×2-point analysis (Eifler et al., 2021; Fang et al., 2024; Reeves et al., 2024). One can further add galaxy cluster statistics such as the cluster abundance to the multiprobe analysis (Bocquet et al., 2024).

Chapter 3

Basic machine learning concepts and their applications to cosmology

Machine learning (ML) is a branch of artificial intelligence (AI) that focuses on developing algorithms capable of learning patterns from data and making predictions or decisions without being explicitly programmed. It has gained widespread application across various fields due to its ability to process large datasets and uncover complex relationships that traditional statistical methods may struggle to identify.

Recent advancements in deep learning, a subset of ML that leverages neural networks with multiple layers including the multilayer perceptron (MLP) and convolutional neural network (CNN), have significantly improved performance in tasks such as image recognition, model classification, and scientific data analysis. In cosmology and astrophysics, ML techniques are increasingly applied to problems like weak gravitational lensing, large-scale structure (LSS) inference, and Line Intensity Mapping (LIM) studies, providing a powerful framework for extracting insights from vast observational datasets.

As ML continues to evolve, its integration with simulation-based inference (SBI), which exploits simulations to estimate complicated probability distribution functions and perform parameter inference by skipping explicit likelihood functions, is opening new frontiers in data-driven discovery. By efficiently learning from simulations, SBI enables robust parameter estimation and model comparison, even in high-dimensional and nonlinear settings, making it a powerful tool for modern cosmological analyses.

In this chapter, we will introduce the basic concepts of ML and its applications to cosmology, focusing on MLP (Sec. 3.1), CNNs (Sec. 3.2), and SBI (Sec. 3.3). We will discuss the principles behind these techniques and how they are applied to our works in the following chapters. To be more precise, we used the MLP architecture to build emulators for the integrated shear 3PCF (Sec. 2.3.3) which is numerically expensive to compute across the cosmological parameter space in Chapter. 6; We adopted and modified the conventional architecture of CNN to develop a new interpretable ML model named cosmological correlator convolutional neural network (C3NN) in Chapter. 7 and we would like to further explore the power of this model by combining it with SBI framework. After understanding

the capabilities of ML and its potential impact on cosmology, we can better utilize these data-driven methods in improving our understanding of the Universe. For more details in deep learning, we recommend the readers to refer to Chollet (2017) while for SBI, we recommend the review of Cranmer et al. (2020).

3.1 Multilayer perceptron (MLP)

A multilayer perceptron is a fundamental type of deep learning neural network. It is usually composed of the following components

- **Neuron**: A neuron is the basic computational unit that processes and transmits information within the neural network. Each neuron receives inputs from the previous layer, applies a weighted sum to them, adds a bias term, and then passes the result through a nonlinear activation function. The mathematical operation can be written down as
$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{3.1}$$
where x_i are the inputs from the previous layer containing n neurons, w_i are the weights, b is the bias, and f is the activation function.

- **Layer**: A layer is composed of connected neurons that take in the same input. The input layer receives the raw data, the hidden layers perform the computation, and the output layer generates the final prediction.

- **Activation function**: An activation function f in Eq. (3.1) introduces nonlinearity into the MLP, allowing it to learn complex patterns in the data. Common nonlinear activation functions include ReLU (Rectified Linear Unit), sigmoid, and tanh. Below we give the expression of ReLU as an example
$$f(x) = \max(0, x)\ , \tag{3.2}$$
where x is the input to the activation function. The activation function is absolutely necessary for MLPs as otherwise consecutive layers can just collapse to one layer.

With the above components, we show in Fig. 3.1 a schematic illustration of a MLP architecture. Here the figure contains a specific application which is for the MLP to take in the cosmological parameter vectors as input and predict the corresponding matter power spectrum on a sequence of k wavenumbers as output.

3.1.1 Training, validation, optimization and testing

To achieve the desired performance shown in Fig. 3.1, we need more than just the architecture. To be explicit, the MLP needs to be trained on a labeled dataset, validated, optimized, and tested.

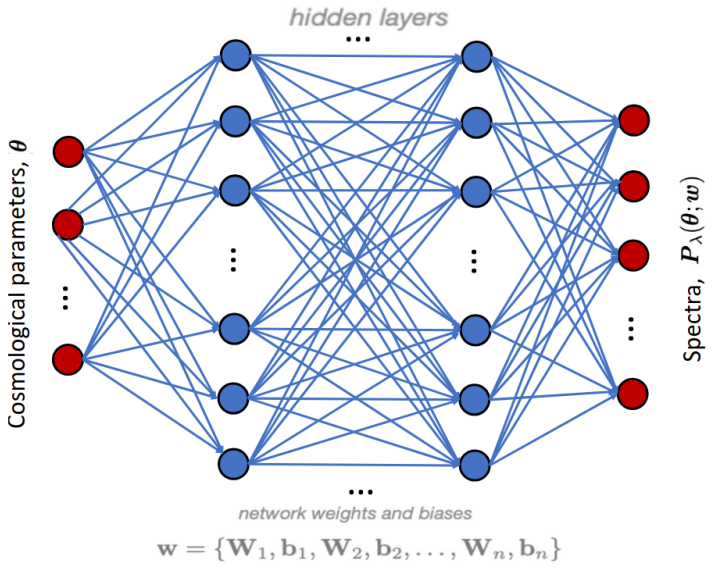

Figure 3.1: Schematic illustration of a MLP architecture. Each neuron is represented by a colored circle. The input layer (red neurons on the left) receives the raw data, here are the cosmological parameter vectors; the hidden layers (blue neurons in the middle) perform the computation showed in Eq. (3.1), and the output layer (red neurons on the right) generates the final prediction which is the matter power spectrum on different k wavenumbers. Image source: Spurio Mancini et al. (2022).

- **Training data**: Training data usually consists of two parts: the feature and the label. Both of them are pre-computed and can be numerical (e.g. from simulations) or categorical (e.g. indicators for different types and groups). The former is a property or characteristic of the data that serves as input to a ML model while the latter is the target variable or ground truth that the ML model is trying to predict[1]. In the case of Fig. 3.1, the cosmological parameter vectors are the features and the matter power spectrum on different k wavenumbers are the labels. The amount of training data needed to obtain the desired performance depends on the complexity of the problem and the size of the model.

- **Validation**: Validation in the ML training process refers to the step where the model's performance is evaluated on a separate dataset, same structure as the training dataset but is not used for optimizing network weights. The validation dataset helps to tune the hyperparameters, prevent overfitting, and ensure that the model generalizes well to unseen data. Here "hyperparameters" refer to the parameters that are initially set to control the learning process, such as the number of layers, the number of neurons in each layer, the learning rate, and the activation function; "Overfitting" occurs when the model performs well on the training data but poorly

[1]In this thesis we only consider supervised learning and therefore labels are necessary. For unsupervised learning, one can refer to Fotopoulou (2024).

on the validation data, indicating that the model has learned the noise in the training data rather than the underlying patterns.

- **Optimization**: The optimization process aims to find the optimal set of weights and biases (Eq. (3.1)) that minimize the loss function. It happens both when the ML model is fed with training and validation dataset. The error between the label and the prediction characterized by the loss function is propagated backward (backpropagation (Chollet, 2017)) through the neural network, updating weights and biases. Common optimization algorithms include stochastic gradient descent (SGD) (Robbins & Monro, 1951), Adam (Kingma & Ba, 2014), and RMSprop (Ruder, 2016). Below we give the expression of the parameter update rule for SGD as an example

$$\theta = \theta - \eta \nabla_\theta L(\theta; x^i, y^i) \ , \tag{3.3}$$

  where θ is the parameter to be updated (like a weight or a bias), η is the learning rate that controls the update step size, L is the loss from a given loss function. The stochasticity of SGD comes from the fact that the gradient is computed on a random subset $\{x^i, y^i\}$ of the training data at each iteration where x is the feature and y the label.

- **Testing**: The testing set is used to evaluate the model's performance on unseen data during training and validation procedures. It provides a final unbiased estimate of the model's accuracy and generalization ability once the whole training process is completed.

With the above discussion on different phases of the MLP or the general ML training process, we can use a simple graph to summarize the workflow as shown in Fig. 3.2.

3.1.2 Emulators for the shear integrated 3PCF

In cases where the computation of numerical simulations or the summary statistics is expensive and cannot be afforded for the Bayesian inference procedures, we can use emulators to approximate these quantities. Emulators are a type of model that leverages ML or other statistical techniques to efficiently predict outputs based on a limited set of precomputed training data. In the context of cosmology, emulators are widely used to approximate the matter power spectrum (as the example shown in Fig. 3.1), the cosmic microwave background (CMB) power spectrum, and the weak lensing statistics, among others, across certain cosmological parameter space.

There are multiple different approaches that one can train an emulator (e.g. Gaussian process (Rasmussen & Williams, 2005), random forest regression (Louppe, 2014), principle component analysis plus interpolation (Gewers et al., 2018) and so on). Here we focus on the use of MLP as discussed in Sec. 3.1.1 and briefly describe how we apply it to the emulation of the shear i3PCF. Readers interested in more details can refer to Sec. 4 in Chapter. 6.

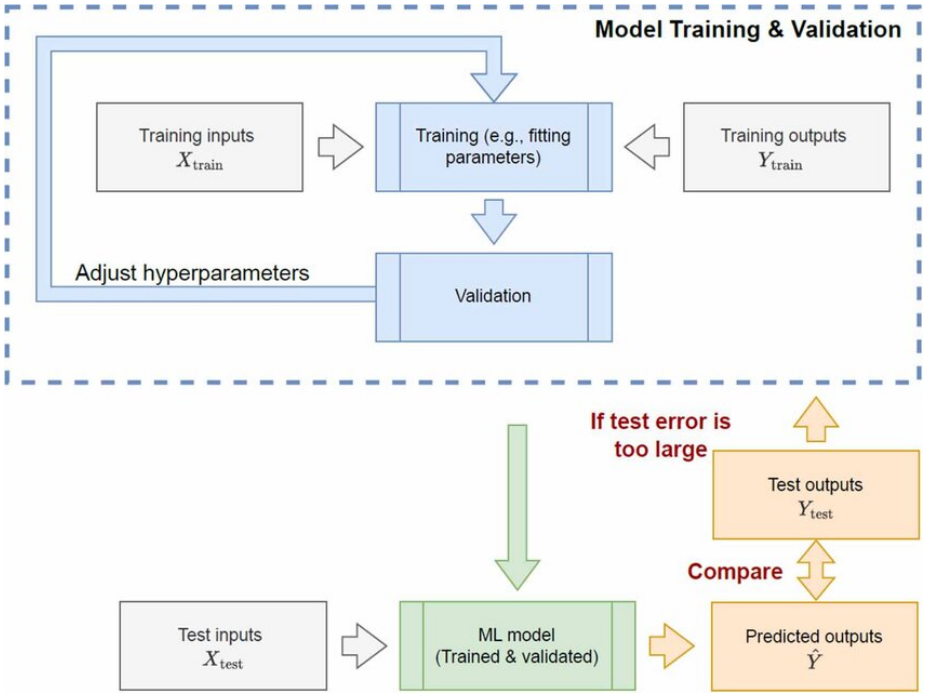

Figure 3.2: A simple graph to summarize the workflow of the ML training process. The training data $\{X_{\text{train}}, Y_{\text{train}}\}$ is used to train the model, the validation data is used to tune the hyperparameters, the optimization process updates the free parameters (e.g. weights and biases), and the testing data $\{X_{\text{test}}, Y_{\text{test}}\}$ is used to evaluate the model's performance. Image source: Bonzanini et al. (2023).

We use the MLP architecture provided by the package `Cosmopower` (Spurio Mancini et al., 2022) to emulate the four-dimensional multipole integration part in Eq. (2.68)

$$\int_{\boldsymbol{\ell}_1}\int_{\boldsymbol{\ell}_2} B_\delta^{3D}\left(\frac{\boldsymbol{\ell}_1}{\chi},\frac{\boldsymbol{\ell}_2}{\chi},\frac{-\boldsymbol{\ell}_{12}}{\chi};\chi\right) e^{2i(\phi_{\boldsymbol{\ell}_2}\mp\phi_{-\boldsymbol{\ell}_{12}})} U(\boldsymbol{\ell}_1) W(\boldsymbol{\ell}_2+\boldsymbol{\ell}) W(-\boldsymbol{\ell}_{12}-\boldsymbol{\ell}) \ , \tag{3.4}$$

across the cosmological and astrophysical parameter space $\{\Omega_{\text{m}}, \ln(10^{10}A_{\text{s}}), w_0, c_{\text{min}}\}$ and redshift z where c_{min} is the baryonic feedback parameter used in a matter power spectrum fitting function `HMcode` (Mead et al., 2021). The ranges of these parameters within which we emulate Eq. (3.4) are given in Tab. 1 of Chapter. 6. The emulator architecture is very similar to what was shown in Fig. 3.1 where we have the above parameters as input and the quantity from Eq. (3.4) on a sequence of ℓ as output. The training data is generated by numerical integration of Eq. (3.4) with its corresponding parameters sampled using Latin hypercube sampling (LHS) which can efficiently populate the parameter space uniformly with a small amount of samples (Devon Lin & Tang, 2022).

This emulation strategy leaves out the part involving the line-of-sight (LOS) integration in Eq. (2.68), but has the advantage of allowing for more flexibility to adjust the source redshift distributions, including bypassing the need to emulate any of the systematic pa-

rameters such as the photometric redshift uncertainty and galaxy intrinsic alignment. For the testing phase, we use our own designed accuracy metric, the fractional difference of log-likelihood surface as in Eq. (4.2) in Chapter. 6, to evaluate the performance of the emulators. The test results are shown in Figs. 5 and 6 of Chapter. 6.

Using emulators to predict various kinds of summary statistics at unseen cosmologies have become a popular approach in cosmology. It can significantly reduce the computational cost of the Bayesian inference procedures and provide a more efficient way to explore the cosmological parameter space. Moreover, there are informative summary statistics that do not have closed analytical forms such as the scattering transform coefficients (Cheng et al., 2020) and weak lensing peak correlation function (Davies et al., 2022). In this case, emulating them with ML techniques across a grid of simulation measurement is a robust approach to make predictions on new cosmological parameters.

3.2 Convolutional neural network (CNN)

Like MLP, CNN is also a specialized type of deep learning model designed for processing structured grid-like data such as images (pixelized maps). It is particularly effective in tasks that require pattern recognition, feature extraction, and spatial hierarchies. Unlike MLP, CNNs use convolutional layers to automatically learn spatial features through small, localized filters, reducing the number of trainable parameters and improving computational efficiency. Meanwhile, CNNs intrinsically possess translational invariance property which significantly help process image-like data.

CNNs' applications to cosmology so far can be approximately categorized into the above mentioned three functions. On the observational side, CNNs are frequently used to identify specific astrophysical systems such as galaxies (Zhu et al., 2019; Cavanagh et al., 2021), supernovae (Brunel et al., 2019; Qu et al., 2021) and strong gravitational lensing systems (Lanusse et al., 2018; Schaefer et al., 2018), just to give an incomplete list of examples. In the cosmological inference tasks, CNNs or CNN-related models are popular ML tools to extract statistical information from cosmological fields such as weak lensing convergence and galaxy number density (Fluri et al., 2022; Lemos et al., 2023b; Lu et al., 2023; Jeffrey et al., 2024). Furthermore, the ability of CNNs to learn spatial hierarchies has been exploited to study the so called interpretability of ML models in cosmology, such as the constructed saliency maps (Simonyan et al., 2013) to interpret information source in weak lensing data (Matilla et al., 2020). Our work presented in Chapter. 7 falls into the joint section of the second and last category where we adapt the conventional CNN model architecture to explicitly extract the spatially hierarchical correlation functions, thus the outputs become interpretable.

3.2.1 Fundamental principles of CNNs

The core building block of a CNN is the convolutional layer, which applies a set of learnable filters (kernels) to input data. Each filter slides across the input, computing dot products

between its weights and the corresponding input data covered by the filter. This operation helps extract features. Mathematically, a convolution operation can be represented as:

$$Y(i,j) = (W * X)(i,j) = \sum_m \sum_n W(m,n) X(i-m, j-n) \ , \tag{3.5}$$

where Y is the output feature map, W is the filter, X is the input, and $*$ denotes the convolution operation. The filter W is applied to each local position of the input X to produce the output Y. m, n are the positional indices within the 2D filter and i, j are the positional indices within the 2D input.

After each convolution operation, usually a nonlinear activation function is applied to introduce nonlinearity into the model. This process is similar to what we described in the beginning of Sec. 3.1. Besides linear convolution and nonlinear activation functions, generally CNNs also include pooling layers which reduce the size of feature maps while preserving the most important information. Two common pooling strategies are (i) max pooling, which takes the maximum value within a pooling region, and (ii) average pooling, which computes the average value within a pooling region. The pooling operation helps to reduce the computational cost and the number of parameters in the network, and also makes the model more robust to spatial translations (Gholamalinezhad & Khosravi, 2020).

Unlike the simplified X input in Eq. (3.5) which is a 2D image, CNNs are designed to be able to process data that has multi-dimensional array structure. For example, in cosmology, the input data can be a 4D tensor with shape (N, C, W, H) of weak lensing convergence maps where N is the number of samples, C is the number of tomographic bins (channels), W is the width, and H is the height of each 2D map in unit of pixels. The corresponding output of the CNN is also a 4D tensor with shape (N, C', W', H') where C' is the number of output channels, W' is the width, and H' is the height of the output after pooling in unit of pixels[2]. The CNN architecture is usually composed of multiple convolutional layers, activation functions and pooling layers, which are stacked together to form a deep neural network. The output of the CNN is then fed into a fully connected layer to make the final prediction.

Like in Fig. 3.1, here we also show in Fig. 3.3 a schematic illustration of a typical CNN model architecture. where we can easily observe that the output feature maps tensor from the convolutional layer conv1 has a dimension of $(64, 224, 224)$ where 224 is the width and height of the feature maps and 64 is the number of output channels. The output then goes through an intermediate pooling layer and both width and height reduce to 112. The second convolutional layer conv2 takes the output of the pooling layer as input and produces a new feature maps tensor with higher values for the dimension of output channel $(128, 112, 112)$. This process then repeats itself until the final fully connected layers which make the final prediction.

[2]Readers can refer to the python API of Pytorch
https://pytorch.org/docs/stable/generated/torch.nn.Conv2d.html
for better numerical understanding in practice.

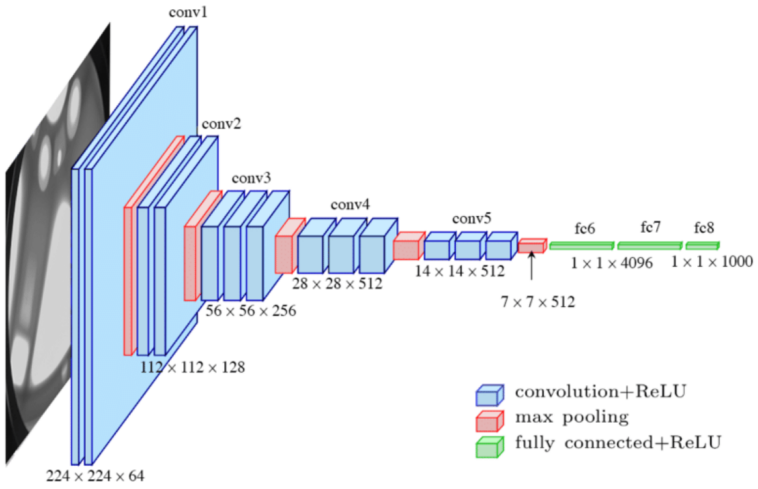

Figure 3.3: Schematic illustration of a typical CNN model architecture. The model is composed of stacked convolutional layers (blue) that contain both linear convolution and nonlinear activation functions. The in-between pooling layers (red) can change the spatial dimensions of the output feature maps from the previous convolutional layer. The output of the CNN is then fed into a fully connected layers (green) to make the final prediction. Image source: Kumar (2023).

3.2.2 Interpretable C3NN

Though MLPs, CNNs and other deep learning models have shown great success in various cosmological applications, they are often criticized as "black boxes" for their lack of interpretability, as the number of free parameters in the models are exceedingly large and we cannot give them clear physical explanation. The black-box nature of these models makes it difficult to understand how they arrive at their predictions, limiting their utility in scientific research. To address this issue, we propose a new interpretable ML model called cosmological correlator convolutional neural network (C3NN) in Chapter. 7. It borrows the convolution operation from a conventional CNN as discussed in Sec. 3.2.1 and fuses it with a designed nonlinear forward calculation such that the output becomes explicitly expressible in terms of the correlation functions at different orders of the input data.

The architecture of C3NN model is shown in Fig. 1 of Chapter. 7 where the outputs $c_\alpha^{(N)}$ are exploited in the context of classification. We skip the derivation here but essentially each $c_\alpha^{(N)}$ can be expressed as

$$c_\alpha^{(i)} = \frac{1}{i!} \sum_{\mathbf{r}_1,\ldots,\mathbf{r}_i,k_1,\ldots,k_i}^{(\mathbf{r}_1,k_1)\neq\cdots\neq(\mathbf{r}_i,k_i)} W_{\alpha,k_1}(\mathbf{r}_1)\cdots W_{\alpha,k_i}(\mathbf{r}_i)\hat{\xi}^{(i)}_{k_1,\ldots,k_i}(\mathbf{r}_1,\ldots,\mathbf{r}_i) \ , \tag{3.6}$$

where $W_{\alpha,k}(\mathbf{r})$ are the weights of the convolutional layer with α as the filter index and k the channel index. $\hat{\xi}^{(i)}_{k_1,\ldots,k_i}(\mathbf{r}_1,\ldots,\mathbf{r}_i)$ is the estimator of ith order correlation function of the input data within the convolutional filter size at positions $\mathbf{r}_1,\ldots,\mathbf{r}_i$.

With the above expression, we developed the interpretability of the C3NN model in the context of classification in three aspects:

- The output $c_\alpha^{(N)}$ of a specific order N can be mathematically expressed in terms of the correlation function at the same order. This is unlike the often hard-to-interpret summary statistics extracted by conventional CNN models.

- Through a regularization path analysis (Efron et al., 2004), which is integrated into the classifier part of our model, we can have a quantitative understanding of the relative importance of the different order $c_\alpha^{(N)}$ in contributing to the model's classification power (see Sec. 2 in Chapter. 7 for more details).

- We can investigate the trained filter weights by connecting individual pixels to form the configuration of any given NPCF. The filter weights for a given NPCF configuration directly allow us to rank different correlation function configurations within a given $c_\alpha^{(N)}$ (Fig. 2 in Chapter. 7).

Besides classification, the C3NN model can be adapted to inference tasks and in the next section we will show some preliminary ideas on how to combine C3NN with simulation-based inference (SBI) framework to perform parameter estimation and information content investigation on the hierarchy of correlation functions.

3.3 Simulation-based inference (SBI)

Traditional methods of parameter inference in cosmology often rely on likelihood-based approaches, which require accurate likelihood functions and well-defined statistical models. However, many modern cosmological observations involve complex, nonlinear, and high-dimensional data that make constructing explicit likelihoods intractable or the Gaussian likelihood assumption insufficient.

Regarding this, SBI, also known as likelihood-free/implicit inference, offers a powerful alternative. Instead of requiring an explicit likelihood function, SBI uses forward simulations of the universe to learn the mapping between cosmological parameters and observational data. This approach is particularly useful in scenarios where the likelihood is unknown or computationally prohibitive to evaluate, such as in Line Intensity Mapping (LIM), weak gravitational lensing, and large-scale structure surveys. Besides the references given in the beginning of Sec. 3.2 about CNN-based inference models, readers can also refer to (Hahn et al., 2024; Tucci & Schmidt, 2024; Lehman et al., 2024; Gatti et al., 2025) for more SBI applications with different statistical models.

By utilizing ML techniques, SBI can construct models such as neural density estimators (Alsing et al., 2019; Kobyzev et al., 2019; Papamakarios, 2019) to approximate the posterior distribution of cosmological parameters given the observed data. This enables efficient inference, even in the presence of non-Gaussian and highly correlated data structures. To

be more precise, different techniques in SBI can be used to approximate different components in the context of Bayesian inference: Neural Posterior Estimation (NPE) (Greenberg et al., 2019; Ward et al., 2022) directly for the posterior; Neural Likelihood Estimation (NLE) (Papamakarios et al., 2018) for the likelihood; and Neural Ratio Estimation (NRE) (Miller et al., 2022) for the likelihood-to-evidence ratio. These different techniques provide great flexibility in solving SBI problems under different scenarios of accuracy requirement, computational efficiency, and applicability to high-dimensions.

Another key element in the SBI framework is to be able to generate synthetic data mimicking the real observations. This component is called the simulator and the process of this mechanism is called the forward modeling. The simulator can be a numerical simulation code, a generative model, or a physical model that can generate synthetic data given a set of cosmological parameters.

During the forward modeling process, one does not only need to generate accurate predictions for the targeted data, but also incorporate every prominent observational and instrumental systematic effects which can to the maximum extent approximate the real observational data. Only then the SBI framework provides them as training data for the ML model as discussed above to learn the mapping between the observational data and cosmological parameter posteriors (likelihoods). In Fig. 3.4 we show flowcharts of different

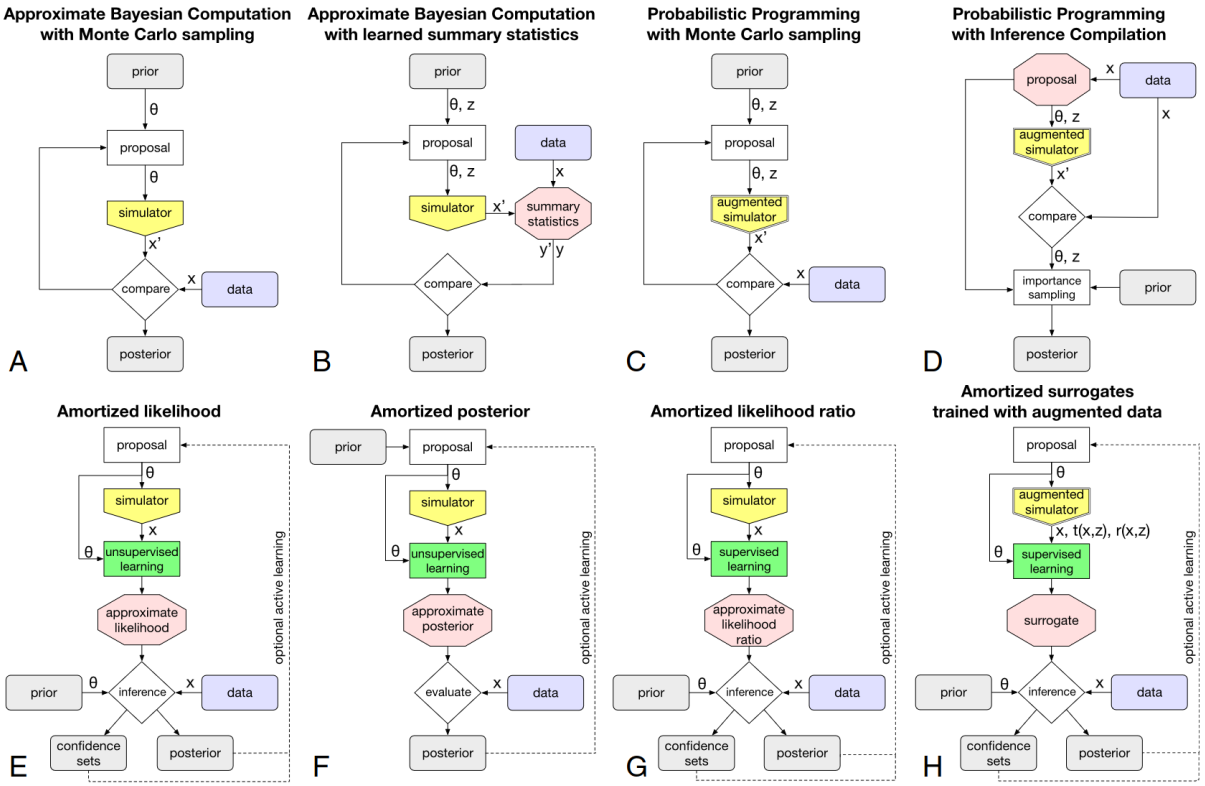

Figure 3.4: An overview of different SBI approaches. Image source: (Cranmer et al., 2020).

SBI approaches, out of which E, F and G apply the NLE, NPE and NRE techniques discussed above respectively. Here "Amortized" means that the inference process in the given SBI approach is precomputed and learned once, so that future inferences can be performed efficiently without running costly forward modeling again.

3.3.1 SBI with C3NN

As already discussed in Sec. 3.2.2, we can replace the classifier part in the C3NN model (Fig. 1 in Chapter. 7) with a neural density estimator to perform SBI. The aim is that we would like to exploit the interpretability of the C3NN model to investigate the cosmological information content within the hierarchy of correlation functions in the context of Bayesian inference. This work so far cannot be done with conventional analysis methods due to the lack of both modeling and estimators for the correlation functions beyond the fourth order.

While the output of the C3NN model shown in Eq, (3.6) is not exactly equivalent to the full-shape correlation functions, it nevertheless provides an efficient estimator up to a user-defined order. It therefore can provide helpful quantitative hints on the constraining power from higher-order correlation functions. In this context, our simulator is the CosmogridV1 simulation suite[3] (Fluri et al., 2022; Kacprzak et al., 2023) which provides us particle shells, we then created full-sky weak lensing convergence maps on 1250 cosmological parameter sets across a 8-dimensional space $\{\Omega_\mathrm{m}, \Omega_\mathrm{b}, \sigma_8, w_0, H_0, n_\mathrm{s}\}$ plus two baryonification parameters by performing line-of-sight projection using `UFalcon` (Sgier et al., 2019, 2021; Reeves et al., 2024). We adopt the amortized posterior approach as shown by F in Fig. 3.4 and the neural posterior estimation algorithm we use is based on masked auto-regressive flows (Papamakarios et al., 2017) as implemented in `sbi`[4] package (Tejero-Cantero et al., 2020). By increasing the order of $c_\alpha^{(N)}$ in the data $\mathbf{x}$ part (F in Fig. 3.4), we can quantify the information contribution from each order based on the credible interval changes in the inferred posterior $P(\boldsymbol{\theta}|\mathbf{x})$.

This is an ongoing work, and just for the purpose of illustration, we show in Fig. 3.5 a preliminary result of the cosmological parameter inference using the C3NN-SBI model. Here we fix the C3NN model to have 1 filter and the training data $c^{(N)}$ comes from noiseless CosmogridV1 weak lensing convergence maps that have 4 channels which correspond to the 4 source tomographic redshift bins in DESY3 analysis. From the first observation, we notice that by adding higher-order output from the C3NN model, the parameter constraints become tighter. This is a proof-of-concept that indeed the C3NN outputs can be used as representations of NPCFs to perform SBI and investigate the information content in each order, or even beyond the ladder of correlation functions.

However, the current work is still in its early stage and there are many aspects that need to be further completed, we list a few below:

[3] www.cosmogrid.ai

[4] https://github.com/sbi-dev/sbi

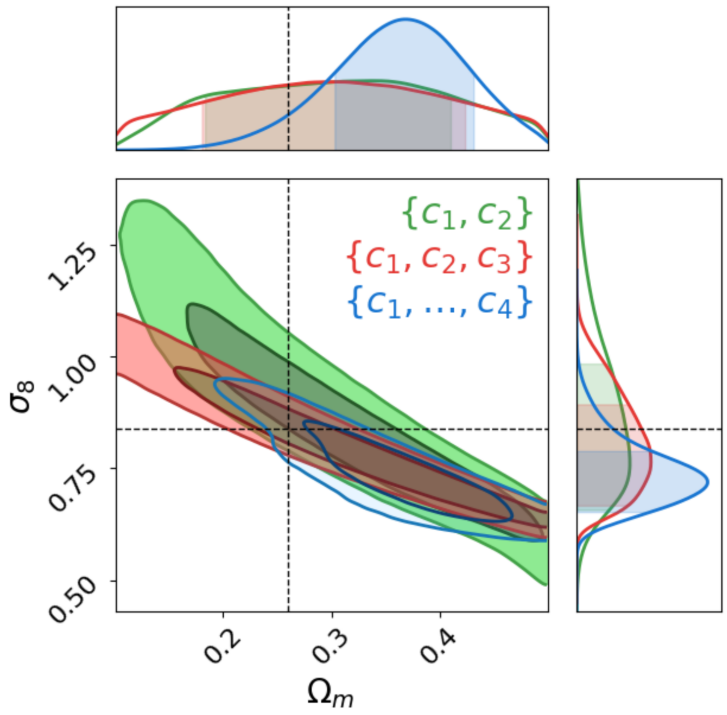

Figure 3.5: A preliminary result of the cosmological parameter inference using the C3NN-SBI model. The corner plot shows the 2D marginalized posterior distributions of the cosmological parameters $\{\Omega_{\rm m}, \sigma_8\}$. The green contour is the inference result by using only $\{c_1, c_2\}$ from the C3NN output. The red contour adds c_3 on top of that. The blue contour includes all the moments from the C3NN output until the fourth order. The true test cosmology is marked by the cross of the dashed lines.

- Each of the three C3NN-SBI models in Fig. 3.5 is not optimized w.r.t its hyperparameters such as learning rate, number of transformations in the NPE and so on. Therefore the quantitative cosmological information from each inference cannot be claimed to be robust. We will use `Optuna`[5] (Akiba et al., 2019) to perform hyperparameter optimization in the future.

- The current work only uses noiseless maps from the CosmogridV1 simulation suite in the forward model. This is not realistic as we have not incorporated the observational and instrumental systematic effects such as shape noise, galaxy intrinsic alignment, point-spread function (PSF) and so on. We will need to include these effects in the simulated data to make the SBI results more reliable.

- No coverage tests have been performed on the C3NN-SBI models. We will need to perform the covarage tests such as the ones in Talts et al. (2018) or Lemos et al. (2023a) to ensure that the credible intervals from the trained amortized posterior

[5]https://optuna.readthedocs.io/en/stable/

estimator are not biased as well as robust, i.e. not under- or overconfident.

Chapter 4

Intermezzo: integrating analytical methods and machine learning in cosmology

The integration of analytical methods with ML is transforming the field of cosmology, enabling more efficient data analyses, enhanced theoretical modeling, and deeper insights into the universe's fundamental properties. Traditional analytical approaches, based on physics (e.g. general relativity and statistical mechanics) and statistics , provide well-defined theoretical frameworks, while ML offers powerful data-driven techniques that can model complex relationships and extract hidden patterns. The synergy between these methodologies is unlocking new frontiers in cosmological research, from large-scale structure formation to precision cosmology.

As we have briefly shown in Chapters. 1 and 2, cosmology has traditionally replied on analytical and numerical approaches to describe the formation and evolution of various kinds of structures in the Universe. These methods, usually based on perturbation theory (Sec. 2.2.3), bias theory (Secs. 2.4.1 and 2.4.2) and statistical inference (Secs. 2.1.2, 2.3.3 and 2.5), are highly interpretable and grounded in solid fundamental physics. However, while applying these methods to cosmological problems, there are often assumptions, approximations and computationally expensive simulations involved such that they are limited in the regime of nonlinearity and high-dimension. We discussed how general perturbative bias expansion would fail on small, nonlinear scales and demonstrate it in Fig. 4 of Chapter. 5 where we utilize this method to compute the clustering of biased critical points in the weak lensing convergence field. In the process of statistical inference, an explicit Gaussian likelihood assumption is usually adopted even for high-dimensional data vectors, to which extent this assumption would impact the final parameter covariance in the context of upcoming large galaxy surveys is still underexplored. On another aspect, certain efficient summary statistics are numerically expensive to calculate (e.g. the i3PCF in Sec. 2.3.3) or can only be extracted via estimators on simulations but cannot be modeled analytically (e.g. scattering transform coefficients (Cheng et al., 2020)). In such a case, the

direct evaluation of these summary statistics is not affordable in a cosmological inference task.

ML, on the other hand, can provide complementary or even alternative solutions to the above problems. As we have discussed in Chapter. 3, we mainly probe the ML applications to cosmology in the following directions:

- Efficient parameter inference: ML is extremely helpful in improving the efficiency of cosmological inference. One major application is the emulation of summary statistics as discussed in Sec. 3.1.2 where the evaluation of summary statistics on observed cosmological fields can be significantly sped up by using ML models such as MLPs through the parameter space, regardless of wether the statistic can be modeled analytically or not. In Chapter. 6 we demonstrate how this approach contributes, together with other components in the pipeline, to the inference of interested cosmological parameters such as A_s and w_0 via summary statistics on cosmic shear beyond 2PCFs. In general, this has become a paradigm in the state-of-the-art statistical analyses of cosmology (Gatti et al., 2022; Harnois-Deraps et al., 2024; Lehman et al., 2025).

- Physics-informed ML: As a developing field, ML models are criticized in many cases as "black boxes" which means it is difficult to understand their internal decision-making mechanism and therefore weaken the robustness of the predictions (Hassija et al., 2024). One way to address this problem is to inject explicit physical laws or properties into the ML model architectures. Through which those models can learn from huge datasets based on physical laws already verified (readers can refer to Karniadakis et al. (2021) and Wetzel et al. (2025) for reviews). Our C3NN model in Chapter. 7 obeys the cosmological principles of homogeneity and isotropy by using the translational invariant convolutional neural network and the rotationally invariant filter in two-dimensional space (Weiler & Cesa, 2019; Cesa et al., 2022) (Sec. 2 in Chapter. 7). Together with the statistically interpretable outputs (Sec. 3.2.2), we have access to verify the model's decision-making process and in turn exploit it to gain new understanding of our cosmological fields' statistical properties. There are other types of ML models that integrate physical laws, constraints, or domain knowledge such as the physics-informed neural networks (PINNs) (Raissi et al., 2019) which respect physics described by general nonlinear partial differential equations while trained to solve supervised learning tasks; and the Hamiltonian Neural Networks (HNNs) (Greydanus et al., 2019) which can learn and respect exact conservative laws in an unsupervised manner.

- Information extraction in high-dimensional data: Traditional statistical analyses of cosmology usually compress the high-dimensional observed field data into certain summary statistics, e.g. correlation functions. Information carried by the original fields would be lost during the compression process. ML models provide an alternative solution to it where models like CNNs can be applied to directly extract features from the data. Those features, though usually not analytical or interpretable, can

act like summary statistics and infer better constraints on cosmological parameters compared to the conventional summary statistics (Lemos et al., 2023b; Lu et al., 2023). The framework that performs such information extraction is SBI as discussed in Sec. 3.3. Down to the bottom, there is the field-level inference formalism which does not even apply ML models like MLPs or CNNs to compress the input data, but rather perform SBI directly on the input pixel maps or catalogues relying on the power of ML techniques dealing with large datasets (Porqueres et al., 2023; Nguyen et al., 2024). Note that this type of inference method still has controversial details and needs to be further investigated before its robustness is generally acknowledged. There are also examples of particular types of ML models that can automatically optimize the amount of information extraction, such as the information maximizing NN (Charnock et al., 2018). The C3NN-SBI framework discussed in Sec. 3.3.1 falls into this category where we integrate C3NN model with the SBI framework to extract cosmological information from weak lensing convergence maps. By doing this, we combine the physics-informed ML models with the information extraction in high-dimensional data.

Besides the above mentioned aspects, there are other useful applications of ML to cosmology which are complementary to the traditional analysis methods. One major example is the pattern recognition or system classification for astrophysical objects. We briefly mentioned and listed relevant references in the beginning of Sec. 3.2. As hundreds of thousands of specific target like supernova, strong lensing system and so on are to be observed in upcoming surveys, ML techniques are necessary to replace human individual selection. Another example is that by combining detailed numerical simulations on small scales with the robust theoretical predictions on large scales, one can construct the so called hybrid SBI model (Modi & Philcox, 2023) through which we can hopefully achieve the reduction of computational cost and improvement on parameter constraining simultaneously. One additional advantage we should mention here is that it is more straightforward to include detailed survey realism to account for observational systematics through forward modeling in SBI frameworks than traditional analytical modeling (Hahn et al., 2022).

So far we have tried to demonstrate how ML techniques can be complementary or even alternative to traditional analysis methods. However, it is essential to notice that the accuracy and robustness of these ML models depend critically on the quality of the training datasets. The reason is the source of the additional information ML models can provide in most cases is on the small scales or field level of numerical simulations, which are commonly used training datasets. However, the behaviour of simulations on those scales is governed by complex physical processes that we do not yet have unanimously agreed numerical recipes to implement, such as baryonic feedback (Schneider & Teyssier, 2015; Giri & Schneider, 2023; Gebhardt et al., 2024; Bigwood et al., 2024), galaxy-halo connection (Berlind & Weinberg, 2002; Nishimichi et al., 2019; Stiskalek et al., 2021) and so on. If the ML model is trained on a suite of simulations that do not properly represent real nature, it is highly likely that the final inference results on cosmology (e.g. from SBI framework) would be biased and inaccurate. This is the effect of model misspecification and there are researches

focusing on testing for it and learning robust statistics under this context (Huang et al., 2023; Anau Montel et al., 2024). For a more abstract discussion on the influence of ML on natural science, one can refer to the interesting paper by Hogg & Villar (2024).

As one suite of simulations may be generated by a different numerical recipe of the above effects from another, it is important to ensure the generalizability of a given ML model. From this perspective, analytical modeling can in turn offer well-defined calibration for the ML model predictions, at least on the scales where analytical modeling is available. This is one of the aims of our work in Chapter. 5.

The fusion of analytical methods with machine learning is redefining how cosmologists analyze data, test theories, and make discoveries. By combining physical principles with data-driven techniques, researchers can potentially tackle previously intractable problems while ensuring robustness and interpretability. As computational methods continue to advance, this integration will play a crucial role in future cosmological experiments, driving deeper insights into the nature of the Universe.

Chapter 5

Clustering of the extreme: A theoretical description of weak lensing critical points power spectra in the mildly nonlinear regime

Bibliographic and copyright information

Gong Z., Barthelemy A., and Codis S., *Clustering of the extreme: A theoretical description of weak lensing critical points power spectra in the mildly nonlinear regime*, 2025, astro-ph.CO, 2502.03457

DOI: 10.48550/arXiv.2502.03457

Outline

This paper has been submitted to Physical Review D and approved by the Dean of the Physics Faculty LMU to be part of the cumulative dissertation.
Our study here presents a theoretical framework for describing the power spectra of weak lensing critical points, focusing on their clustering properties in the mildly nonlinear regime. By leveraging perturbation theory and the theory of bias, we provide new insights into the statistical properties of weak lensing extrema and their connection to large-scale structure. One key success in the paper is that our theoretical framework goes beyond the previous analytical peak clustering works generally assuming Gaussian distributed underlying density field, and includes both voids (minima) and saddle points in the description up to the bispectrum correction accounting for non-Gaussian effects. This can serve as a benchmark test for N-body simulations that are used to measure weak lensing peaks or minima clustering, to prevent those statistics from being biased by simulation systematic effects.

We have also observed features like Baryon Acoustic Oscillations (BAOs) and inflection points on large scales for weak lensing peak 2PCFs. These features could serve as independent standard rulers, independent of galaxy bias, to study the evolution of the large-scale structure and test different cosmological models. As the first author, I was responsible for the following key tasks: (i) deriving the perturbative bias expansion approximation for the power spectrum of critical points up to the next-to-next-to-leading order (NNLO), (ii) developing the code for the numerical computation of the power spectrum of critical points up to the next-to-leading order (NLO) using `Mathematica`, (iii) investigating the BAO features for weak lensing peak 2PCFs and (iv) performing the numerical validation of the theoretical predictions with Monte Carlo (MC) integrations. I also wrote majority of the manuscript. The co-authors contributed to the formulation of critical point bias model, providing valuable scientific discussions, drafting specific sections of the manuscript, and reviewing the final draft.

Clustering of the extreme: A theoretical description of weak lensing critical points power spectra in the mildly nonlinear regime

Zhengyangguang Gong*
Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstraße 1, 81679 München, Germany and Max Planck Institute for Extraterrestrial Physics, Giessenbachstraße 1, 85748 Garching, Germany

Alexandre Barthelemy
Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstraße 1, 81679 München, Germany and Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France

Sandrine Codis
Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France

(Dated: February 7, 2025)

In cosmic web analysis, complementary to traditional cosmological probes, the extrema (e.g. peaks and voids) two-point correlation functions (2PCFs) are of particular interest for the study of both astrophysical phenomena and cosmological structure formation. However most previous studies constructed those statistics via N-body simulations without a robust theoretical derivation from first principles. A strong motivation exists for analytically describing the 2PCFs of these local extrema, taking into account the nonlinear gravitational evolution in the late Universe. In this paper, we derive analytical formulae for the power spectra and 2PCFs of 2D critical points, including peaks (maxima), voids (minima) and saddle points, in mildly non-Gaussian weak gravitational lensing fields. We apply a perturbative bias expansion to model the clustering of 2D critical points. A generalized Gram-Charlier A series expansion is used to describe the probability density functional of the cosmic density field. We successfully derive the power spectrum of weak lensing critical points up to the next-to-next-to-leading order (NNLO) in gravitational perturbation theory, where trispectrum configurations of the weak lensing field have to be included. We numerically evaluate those power spectra up to the next-to-leading order (NLO), which correspond to the inclusion of bispectrum configurations, and transform them to the corresponding 2PCFs. An exact Monte Carlo (MC) integration is performed assuming a Gaussian distributed density field to validate our theoretical predictions. Overall, we find similar properties in 2D compared to the clustering of 3D critical points previously measured from N-body simulations. Contrary to standard lensing power spectra analysis, we find distinct BAO features in the lensing peak 2PCFs due to the gradient and curvature constraints, and we quantify that non-Gaussianity makes for $\sim 10\%$ of the signal at quasi-linear scales which could be important for current stage-IV surveys.

Keywords: cosmology: theory – large-scale structure of the Universe – methods: analytical, numerical – weak gravitational lensing

I. INTRODUCTION

The statistics of critical points, both in 3D and 2D, have attracted significant interests due to their applications to cosmology. In 3D, peaks in the initial Lagrangian density field are key sites for the nonlinear formation of dark matter halos (Ref. [1] and references therein) and their statistics, such as abundance and correlation functions, in Gaussian random fields have been extensively investigated in the past literature [2–10]. Voids, which evolve in the quasi-linear regime, can be effectively modeled using relatively simple linear theory. As such, void statistics can serve as an effective cosmic laboratory for testing modified gravity and exploring dark energy phenomena [11–15]. Additionally, they provide independent and complementary probes for constraining cosmological parameters [16, 17]. Cosmic filaments and walls, while being relatively less studied, provide valuable insights into phenomena such as matter transportation, cosmic web formation and galaxy evolution. Furthermore, their cross-correlations with peaks and voids offer a geometric characterization of the large-scale structure of the Universe [18–23].

In 2D, the statistics of both peaks (maxima) and voids (minima) in weak gravitational lensing have been widely

* lgong@usm.lmu.de

studied and applied to infer cosmological parameters using data from Stage-III surveys, including the Dark Energy Survey (DES), the Kilo-Degree Survey (KiDS), and the Hyper Suprime-Cam SSP Survey (HSC) [24–33]. These studies demonstrate that peaks and voids capture non-Gaussian information in the cosmological field, like other weak lensing higher-order statistics [34–39]. When combined with the weak lensing angular power spectrum, they significantly enhance parameter constraints compared to using the angular power spectrum alone. In particular for the parameter $S_8 = \sigma_8\sqrt{\Omega_{\rm m}/0.3}$, several studies reported improvements in constraints of approximately $35\% \sim 40\%$ [30, 32].

However, all these studies in weak lensing relied on a simulation-based inference approach, where the peak and void statistics were emulated using a grid of N-body simulations spanning various cosmologies, often with machine learning tools such as deep neural networks or Gaussian processes. This approach carries the risk of propagating numerical systematics inherent in the simulations into the emulated statistics, potentially biasing the resulting cosmological inferences [40, 41]. Furthermore, from an analytical modeling perspective, this method lacks a robust theoretical foundation derived from first principles. On the other hand, most previous analytical modeling of peak and void statistics generally assumes a Gaussian cosmological density field. However, the weak lensing fields on our scales of interest are not Gaussian distributed as a consequence of the nonlinear gravitational evolution of the density field in the late Universe.

There is thus a strong motivation to study the statistics of peaks and voids in non-Gaussian weak lensing fields in cosmology (as done for their number densities e.g. in Ref. [42]). Along with saddle points, these features are collectively referred to as critical points. In this work, we focus on the analytical modeling of the power spectrum which leads to the 2-point correlation functions (2PCFs) of critical points in non-Gaussian weak lensing fields, aiming to bridge the gap between current numerical approaches and a theoretical understanding of these statistics. Our analytical method builds on the general formalism proposed in Ref. [43], applied here to 2D weak lensing fields. The approach involves a direct perturbative bias expansion in the Eulerian density field. We derive analytical formulae with perturbative approximations up to the next-to-next-to-leading order (NNLO), which incorporates the trispectrum of the density field. For numerical computation, however, we limit our analysis to the next-to-leading order (NLO), which includes the bispectrum of the density field and represents the lowest-order non-Gaussian correction. All bias coefficients in the derived formulae can be computed analytically with an order-by-order correspondence to operators in the perturbative expansion. These coefficients can be interpreted as response functions of the critical point number densities to variations in the long-wavelength modes of the underlying density field.

Our paper is organized as follows: In Sec. II, we first summarize our results and describe qualitatively the physical behaviour of the extrema 2PCFs before diving into the rigorous derivation of the plots and formulae we schematically present in this section. The following Sec. III provides the formal definitions of general 2D critical points and their number density functions. It also presents the probability density function of critical points under the assumption of a Gaussian distributed density field. In Sec. IV, we derive the 2PCFs of 2D critical points, including both auto- and cross-correlations, up to the NNLO, and discuss the derivation of the bias coefficients. Next in Sec. V, we incorporate weak lensing formalism into our analytical predictions and present numerical results for the 2PCFs of 2D weak lensing critical points up to the NLO. We then validate our perturbative bias expansion predictions by comparing them with results from computationally intensive high-dimensional numerical integrations in Sec. VI. Finally, conclusions are given in Sec. VII.

II. SUMMARY AND INTUITIVE DISCUSSION OF OUR RESULTS

Here we briefly discuss our main results, providing some qualitative arguments for readers to better understand the detailed analytical derivations in subsequent sections.

A. 2PCF of all critical points in mildly non-Gaussian weak-lensing fields

Our major result is a from-first-principles analytical expression for the power spectrum of all pairs of critical points in 2D mildly non-Gaussian fields, which we apply to the projected weak-lensing convergence. The full expression will be rigorously derived in Eq. (35) but we display straight away the resulting 2PCFs among all pairs of critical points in weak lensing convergence field in Fig. 1. The ν variable in the figure, and the remainder of the text, is a parameter that allows us to characterize the "rarity" of the considered extrema by only focusing on extrema of amplitude above the chosen $\nu \equiv \delta/\sigma_0$ where σ_0 can be expressed using Eq. (7). As such, fixing for instance $\nu = 0.3$ can be read as "extrema whose amplitudes are larger than $0.3\sigma_0$ above the mean density of the considered field", here the weak lensing convergence. This choice is arbitrary, and is usually more suited to the study of peaks of the field. Our derivations are generic enough so that another choice, e.g. looking at extrema below a threshold, at a specific amplitude or within an interval, could straightforwardly be derived. We discuss below some symmetries in the formalism that enable to straightforwardly write down some of these cases without any additional derivations. On a side note, $\nu = 0.3$ corresponds to roughly 80% of the total peaks and 6% of voids in a Gaussian random field with standard cosmol-

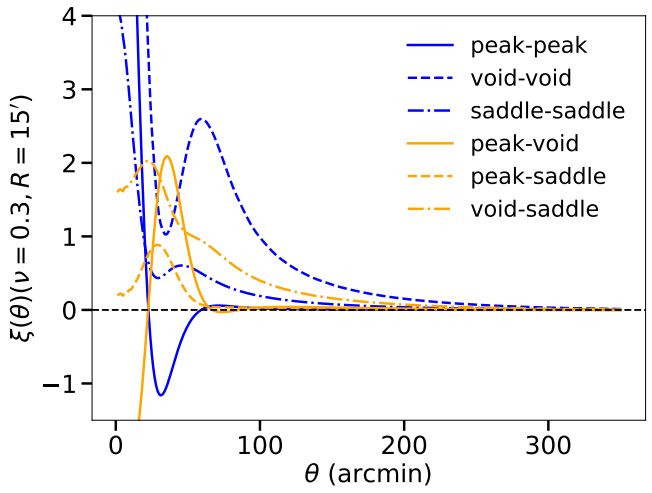

FIG. 1. A summary of 2PCFs of all pairs of critical points in weak lensing convergence fields above a threshold of $\nu = 0.3$. The convergence is smoothed with a Gaussian kernel at scale $R = 15'$. Blue curves represent auto 2PCFs while orange curves display cross 2PCFs (between different types of critical points). Within each color, different curve configurations represent different types of critical points in a 2PCF. Note that the chosen convergence field is for sources located at $z_s = 1.5$.

ogy and lensing parameters. Though subject to change, these order of magnitudes will be preserved for the mildly non-Gaussian fields that we will consider in this paper.

Schematically, our result in Eq. (35) is the combination of two side-by-side perturbative expansions: One is in the development of the gravitational instability where orders are, in the underlying matter density field, expanded in powers of the traditional matter power spectrum (hence amplitude of fluctuations). The second expansion is a bias expansion for critical points that allows us to explicitly bypass the null gradient constraints of extrema by equating it to a series of responses of the extrema one-point distribution to changes in the underlying field at different (coupled) Fourier modes. At leading order (LO), the power spectrum between two critical points i and j is

$$P^{ij}_{\rm extr}(k) = g^i_1(\mathbf{k}) g^j_1(\mathbf{k}) P(k) \tag{1}$$

where the bias function g^i_1 formally depends on the type of critical points, the chosen threshold ν, and characteristics of the field. $P(k)$ is the usual power spectrum of the amplitude of the considered field, here the weak-lensing convergence.

B. Interpretation of LO bias terms

One specificity of our approach is that, without any loss of generality, our bias functions are that of a pure Gaussian field so that the non-Gaussian corrections needed in our formalism appear outside of the bias terms. This allows for a better interpretability of the behaviour of those bias terms. For example, the abundances of peaks and voids in Gaussian fields are symmetric with respect to the mean density, so that the computed biases for peaks above a threshold ν, are exactly those of voids below a threshold $-\nu$. In higher dimensions than 2D, similar statements could be made for different saddle points with symmetric curvature signatures.

Following Eq. (38), the LO bias term is decomposed into

$$g^i_1(\mathbf{k}) = \alpha_i(\nu) + \beta_i(\nu) k^2, \tag{2}$$

where α_i and β_i will b explicitly given in Eqs (47) to (54). Let us note several properties of this LO bias and their consequences for the extrema power spectra:

- Both α_i and β_i are integrals on the specific extrema constraints of the joint distribution of the field amplitude, gradient and second derivatives. Our expression seems to hide this fact through a change of variable that simplifies our calculations, but is effectively a combination of the responses (linear biases) of the amplitude, gradient and second derivatives to a mode fluctuation.

- Following Eq. (1), the small fluctuations in the usual convergence power spectrum are enhanced by k^2 and k^4 terms in the extrema power spectra. Performing the steps described in the above point, it formally allows to determine from which aspects of the field – its amplitude, gradient, second derivatives, or a combination of those – the amplifications come from. In particular, Ref. [6] demonstrated in their Eq. (41) how the response to each successive spatial derivatives of a 3D field impacts the amplification of the power spectrum amplitude in the peak power spectrum calculation. It is apparent that similar expressions hold in our case, though considering all extrema *above* our quoted threshold instead of at its value like in Ref. [6] add extra integrals that prevent us from having similar analytical expressions.

- The schematic behaviour of the bias terms can nevertheless be shown through qualitative arguments that we illustrate in Fig. 2. Indeed, at first order, an extrema of amplitude $\nu_{\rm ext}$ can be thought as tracing the overall matter density fluctuation of amplitude $\nu_{\rm ext}$ plus the curvature of the extrema. As such, the critical point bias will roughly behave as the overall bias of the field amplitude itself at the corresponding value of ν, that is its response to a mode fluctuation of the density. Following Kaiser's formula [44], we thus expect that for extrema of high amplitudes, which we control by a large chosen value of the threshold $\nu_{\rm th}$, the extrema bias will tend towards $\nu_{\rm ext} + \text{curvature}_{\rm ext}$. The curvature of peaks being negative, and that of voids positive, this qualitatively explains why voids tend to

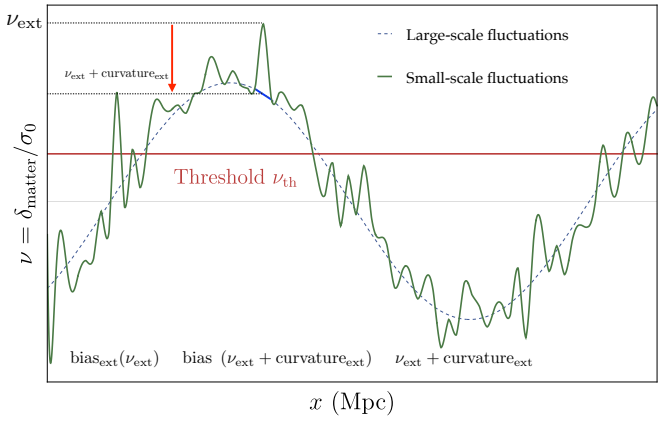

FIG. 2. Illustration of the behaviour of the LO bias functions for extrema. We schematically decompose the field into its large and small-scales fluctuations. The critical point bias can at large ν be approximated by the value of the critical point amplitude plus its curvature.

be more biased than peaks for large values of the threshold. This is indeed what we observe in the rigorous result we will plot in Fig. 8 below, in its top left panel for large values of ν.

C. BAO features in weak-lensing critical points 2PCF

Combining Eqs (1) and (2), the LO critical point auto-spectrum can be written as

$$P^{ii}_{\rm extr}(k) = \alpha_i^2(\nu)\left(1 + 2\frac{\beta_i(\nu)}{\alpha_i(\nu)}k^2 + \left(\frac{\beta_i(\nu)}{\alpha_i(\nu)}\right)^2 k^4\right) P(k). \tag{3}$$

We now decompose the field power spectrum into a dark-matter component, schematically a power law with the index matched to the scales of interest, and a baryonic component, schematically a localized oscillation pattern coming from the Baryonic Acoustic Oscillations (BAO). For the sake of a qualitative argument, we will roughly consider this oscillation to be modeled by a *Sinc* function but the important ingredient in our argument is the damping of the oscillations at large k. We thus have

$$P(k) \approx k^n\left(1 + \frac{\sin(ks)}{ks}\right), \tag{4}$$

where s is the typical scale of the BAO feature in the field. Remember that our application is the weak-lensing convergence field so that s depends on the projection and more precisely on the redshift of the sources and the cosmology. This leads to

$$P^{ii}_{\rm extr}(k) = \alpha_i(\nu)^2 k^n\left[1 + 2\frac{\beta_i(\nu)}{\alpha_i(\nu)}k^2 + \left(\frac{\beta_i(\nu)}{\alpha_i(\nu)}\right)^2 k^4 + 2\left|\frac{\beta_i(\nu)}{\alpha_i(\nu)}\right|\frac{k}{s}\sin(ks)\left(\frac{|\alpha_i(\nu)|}{2|\beta_i(\nu)|k^2} + \text{sign}\left(\frac{\beta_i(\nu)}{\alpha_i(\nu)}\right) + \left|\frac{\beta_i(\nu)}{\alpha_i(\nu)}\right|\frac{k^2}{2}\right)\right], \tag{5}$$

where "sign" in the above equation is the *Sign* function. Eq. (5) shows how the oscillatory behaviour in the field power spectrum can be enhanced in critical point power spectra. This in particular leads to the wiggles in the peak 2PCF that we observe in Fig. 5. However, those wiggles are not observed in the void 2PCF shown in Fig. 4 in our particular setting of a Gaussian smoothing of the field at $15'$ and $\nu = 0.3$. This can also be explained in our simplified model. Indeed, the oscillations would be dampened to the same level as in the field power spectrum if the terms in the last line of Eq. (5) tend to be 0. This would happen if both $\beta_{\rm void}(\nu = 0.3)/\alpha_{\rm void}(\nu = 0.3)$ were negative and $|\beta_{\rm void}(\nu = 0.3)/\alpha_{\rm void}(\nu = 0.3)| \times k^2/2$ at the k of interest were to be of order unity. For the weak-lensing convergence field in the hereby considered case, we indeed have $\beta_{\rm void}(\nu = 0.3)/\alpha_{\rm void}(\nu = 0.3) \sim -3 \times 10^{-5}$ and $k \sim 10^2 - 10^3$ for the wiggles, which qualitatively explains the absence of BAO features in our plots of the void 2PCF, while highlighting the fact that this is specific to the chosen configuration, threshold and smoothing, and does not hold in general in our formalism.

D. Amplitude of the non-Gaussian corrections, impact of our results

One of the other main results of this work is the inclusion of mild (gravitational) non-Gaussianities in the studied fields. In order to illustrate the relative importance of this effect, we plot in Fig. 6 the ratio of our computed non-Gaussian peak 2PCF and its Gaussian counterpart computed at the same order in the critical point bias expansion, as a function of the applied Gaussian smoothing to the convergence field. We perform this comparison at a fixed separation where our formalism typically applies, $\theta = 150'$ given the range of smoothing, the threshold and the redshift of the sources. As expected, the importance of the non-Gaussian terms diminishes as the smoothing increases, but it is worth noting that a smoothing of $\sim 10'$ still shows a difference of 10% at a separation which could be considered important, and thus close to the Gaussian regime, in the cosmological context. A careful analysis including the expected error bars of the critical point 2PCFs in a current stage-IV survey is beyond the scope of this paper and left to future studies.

III. 2D CRITICAL POINTS

To define a critical point in a 2D random field f, one must consider the field amplitude f itself, its derivatives $\partial_i f$ and $\partial_i\partial_j f$ up to the second order, as all local extrema require the gradient to vanish and the curvatures to comply to certain conditions. To be more explicit, the Hessian matrix at a peak location should be negative definite, it is positive definite at a void location and has both positive and negative eigenvalues at a saddle point. Based on the above discussion, we adopt the following three corresponding random variables

$$\alpha = \frac{f}{\sigma_0} \ , \quad \eta_i = \frac{\partial_i f}{\sigma_1} \ , \quad \zeta_{ij} = \frac{\partial_i\partial_j f}{\sigma_2} \ , \tag{6}$$

where σ_n acts as a normalization constant and is defined as the spectral moment of the field

$$\sigma_n^2 = \int \frac{\mathrm{d}^2k}{(2\pi)^2} k^{2n} P(k) \ , \tag{7}$$

in which $P(k)$ is the power spectrum of the 2D random field and is only a function of the magnitude of the wave vector $k = |\mathbf{k}|$ due to the supposedly statistical isotropy of the Universe. These normalization factors are chosen because we have $\langle f^2\rangle = \sigma_0^2$, $\langle(\nabla f)^2\rangle = \sigma_1^2$ and $\langle(\Delta f)^2\rangle = \sigma_2^2$ where Δ represents a Laplacian operator. The power spectrum $P(k)$ is expressed as

$$\langle\tilde{f}(\mathbf{k})\tilde{f}(\mathbf{k}')\rangle = (2\pi)^2\delta_{\rm D}(\mathbf{k}+\mathbf{k}')P(k) \ , \tag{8}$$

where we use $\tilde{f}$ to represent the Fourier counterpart of the random field and the Dirac delta function is a result of the statistical homogeneity.

With the above random variables defined, the number density function of critical points above a given threshold ν (meaning $f \geq \nu\sigma_0$), for example peaks $n_p(\nu)$, can be explicitly expressed as [2, 45–47]:

$$n_p(\nu) = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \Theta(\alpha-\nu)\delta_{\rm D}(\boldsymbol{\eta})\Theta(\lambda_2)|\det\zeta| \ , \tag{9}$$

where Θ is the Heaviside step function and λ_2 is the smallest eigenvalue of the 2×2 matrix $(-\zeta)$ (without loss of generality, we assume $\lambda_1 > \lambda_2$ in the discussion below). For the other two types of critical points, voids and saddle points, their respective number density functions can be derived by modifying the constraint on the eigenvalues of the Hessian matrix ζ as specified in the above equation. For voids, λ_1 must be negative, whereas for saddle points, λ_1 must be positive and λ_2 negative. Eventually, these different constraints will enter the calculation of bias coefficients as will be shown in Sec. IV.

Let us denote the full set of random variables which describes the critical points in 2D as $\mathbf{X} = (\alpha, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}, \zeta_{22})$. The statistics of this multi-variate random vector assuming a Gaussian distributed density field are solely determined by their covariances and reads

$$\begin{aligned}
&\langle\alpha^2\rangle = 1 \ , \quad \langle\alpha\eta_i\rangle = 0 \ , \quad \langle\alpha\zeta_{ij}\rangle = -\frac{\gamma}{2}\delta_{ij} \ , \\
&\langle\eta_i\eta_j\rangle = \frac{\delta_{ij}}{2} \ , \quad \langle\eta_i\zeta_{jk}\rangle = 0 \ , \\
&\langle\zeta_{ij}\zeta_{kl}\rangle = \frac{1}{8}(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \ ,
\end{aligned} \tag{10}$$

where

$$\gamma = \frac{\sigma_1^2}{\sigma_0\sigma_2} \ , \tag{11}$$

and δ_{ij} is the Kronecker delta. The multivariate Gaussian distribution function for the random vector variable $\mathbf{X}$ is therefore:

$$P_{\rm G}(\mathbf{X}) = \frac{1}{(2\pi)^3\sqrt{\det\mathbf{M}}}\exp\left(\frac{1}{2}\mathbf{X}^{\rm T}\mathbf{M}^{-1}\mathbf{X}\right) \ , \tag{12}$$

where the covariance matrix $\mathbf{M}$ of the data vector $\mathbf{X}$ reads

$$\mathbf{M} = \begin{pmatrix}
1 & 0 & 0 & -\frac{\gamma}{2} & 0 & -\frac{\gamma}{2} \\
0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
-\frac{\gamma}{2} & 0 & 0 & \frac{3}{8} & 0 & \frac{1}{8} \\
0 & 0 & 0 & 0 & \frac{1}{8} & 0 \\
-\frac{\gamma}{2} & 0 & 0 & \frac{1}{8} & 0 & \frac{3}{8}
\end{pmatrix} \ . \tag{13}$$

It is very helpful to transform the above probability density function into another representation in terms of rotationally invariant random variables [42, 48]:

$$P_{\rm G}(\mathbf{X}) \propto \mathcal{N}(\alpha, J_1)\exp\left(-\eta^2 - J_2\right) \ , \tag{14}$$

up to a normalization constant, where $\mathcal{N}(\alpha, J_1)$ is a Gaussian joint distribution of α and J_1

$$\mathcal{N}(\alpha, J_1) = \frac{1}{2\pi\sqrt{1-\gamma^2}}\exp\left[-\frac{\alpha^2+J_1^2-2\gamma\alpha J_1}{2(1-\gamma^2)}\right] \ . \tag{15}$$

In the above Eqs. (14) and (15), the new random variables η, J_1, J_2 are defined as:

$$\begin{aligned}
\eta &\equiv \boldsymbol{\eta}\cdot\boldsymbol{\eta} = \eta_1^2+\eta_2^2 \ , \quad J_1 \equiv -\zeta_{ii} = \lambda_1+\lambda_2 \ , \\
J_2 &\equiv 2\tilde{\zeta}_{ij}\tilde{\zeta}_{ji} = \lambda_1^2+\lambda_2^2-2\lambda_1\lambda_2 \ ,
\end{aligned} \tag{16}$$

where the repeated indices follow the Einstein summation convention. The random variable $\tilde{\zeta}_{ij}$ represents the traceless part of the Hessian matrix ζ, $\tilde{\zeta}_{ij} = \zeta_{ij} + \delta_{ij}J_1/2$, and J_1 is the negative trace of the Hessian matrix.

IV. 2PCFS OF 2D CRITICAL POINTS IN MILDLY NON-GAUSSIAN FIELDS

In this paper, our aim is to predict correlation functions of critical points. To do so, let us introduce a

generic functional $\mathcal{F}$ of the density field f. Its power spectrum can be written down as [49]:

$$\frac{\langle\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\rangle_{\rm c}}{\langle\mathcal{F}\rangle^2} = (2\pi)^2\delta_{\rm D}(\mathbf{k}+\mathbf{k}')P_{\mathcal{F}}(k) \ , \tag{17}$$

where $\tilde{\mathcal{F}}$ is the Fourier transform of the real space functional $\mathcal{F}$ and $\langle\ldots\rangle_{\rm c}$ denotes the connected part of the corresponding computed moment. The squared mean of the real space functional $\langle\mathcal{F}\rangle^2$ is a normalization factor to maintain the consistency between the definition of 2PCF and the inverse Fourier transform of the power spectrum. The connected part in Eq. (17) in Fourier space is $\langle\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\rangle_{\rm c} = \langle\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\rangle - \langle\tilde{\mathcal{F}}\rangle^2$ where

$$\langle\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\rangle = \int \mathcal{D}\tilde{f}\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\mathcal{P} \tag{18}$$

is the 2nd-order moment of the functional $\mathcal{F}$ in Fourier space. In the above Eq. (18), $\mathcal{D}\tilde{f}$ represents the volume element of the functional integral over $\tilde{f}$ with appropriate measures and $\mathcal{P}$ is the probability density functional of $\tilde{f}$. We apply the general formalism presented in Ref. [43] to expand both $\mathcal{P}$ and $\tilde{\mathcal{F}}$ in Eq. (18) with a basis composed of the Wiener-Hermite functionals defined as:

$$\mathcal{H}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n) \equiv \frac{(-1)^n}{\mathcal{P}_{\rm G}}\frac{\partial^n\mathcal{P}_{\rm G}}{\partial\tilde{f}(\mathbf{k}_1)\cdots\partial\tilde{f}(\mathbf{k}_n)} \ , \tag{19}$$

where $\mathcal{P}_{\rm G}$ here is the Gaussian probability density functional of $\tilde{f}$, different from those in Eq. (12) and (14). The Wiener-Hermite polynomials $\mathcal{H}_n$ contains the n-point response of the Gaussian probability density functional to the density field and $\mathcal{H}_0 = 1$ when $n = 0$. With this mathematical tool, we can write the expansion of $\mathcal{P}$ and $\tilde{\mathcal{F}}$ as:

$$\begin{aligned}
\mathcal{P}(\tilde{f}) = {} & \mathcal{H}_0\mathcal{P}_{\rm G} + \frac{1}{6}\int {\rm d}^2k_1{\rm d}^2k_2{\rm d}^2k_3\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\mathcal{H}_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\mathcal{P}_{\rm G} \\
& + \frac{1}{24}\int {\rm d}^2k_1\ldots{\rm d}^2k_4\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_4)\rangle_{\rm c}\mathcal{H}_4(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\mathcal{P}_{\rm G} + \frac{1}{120}\int {\rm d}^2k_1\ldots{\rm d}^2k_5\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_5)\rangle_{\rm c}\mathcal{H}_5(\mathbf{k}_1,\ldots,\mathbf{k}_5)\mathcal{P}_{\rm G} \\
& + \frac{1}{720}\int {\rm d}^2k_1\ldots{\rm d}^2k_6\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_6)\rangle_{\rm c}\mathcal{H}_6(\mathbf{k}_1,\ldots,\mathbf{k}_6)\mathcal{P}_{\rm G} + \cdots \\
& + \frac{1}{72}\int {\rm d}^2k_1\ldots{\rm d}^2k_6\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\langle\tilde{f}(\mathbf{k}_4)\tilde{f}(\mathbf{k}_5)\tilde{f}(\mathbf{k}_6)\rangle_{\rm c}\mathcal{H}_6(\mathbf{k}_1,\ldots,\mathbf{k}_6)\mathcal{P}_{\rm G} + \cdots \ ,
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
& \tilde{\mathcal{F}}(\mathbf{k}) \\
& = \sum_{n=0}^{\infty}\frac{1}{n!}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_n}{(2\pi)^2}(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_n-\mathbf{k}) \\
& \times\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\mathcal{H}_n^\star(\mathbf{k}_1,\ldots,\mathbf{k}_n) \ .
\end{aligned} \tag{21}$$

The detailed derivation of Eq. (20) can be found in Appendix. A of this paper or in Appendix. A of Ref. [43]. The expansion in Eq. (20) is a generalization of the Gram-Charlier A (GCA) series [50]. The coefficients $\langle\tilde{f}(\mathbf{k}_1)\ldots\tilde{f}(\mathbf{k}_n)\rangle_{\rm c}$ in the expansion are the corresponding n-th order cumulants of the Fourier density field. They are related to the definition of the higher-order spectrum of the density field, for example:

$$\begin{aligned}
\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c} = {} & (2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) \\
& \times B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \ ,
\end{aligned} \tag{22}$$

where $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ is the bispectrum of the density field and

$$\begin{aligned}
\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_4)\rangle_{\rm c} = {} & (2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_4) \\
& \times T(\mathbf{k}_1,\ldots,\mathbf{k}_3) \ ,
\end{aligned} \tag{23}$$

where $T(\mathbf{k}_1,\ldots,\mathbf{k}_4)$ is the trispectrum. In Eq. (21) we use the dual Wiener-Hermite functionals $\mathcal{H}_n^\star$ whose definition is:

$$\begin{aligned}
\mathcal{H}_n^\star(\mathbf{k}_1,\ldots,\mathbf{k}_n) = {} & (2\pi)^{2n}P(k_1)\cdots P(k_n) \\
& \times\mathcal{H}_n(-\mathbf{k}_1,\ldots,-\mathbf{k}_n) \ ,
\end{aligned} \tag{24}$$

and it has a convenient property of being orthogonal to the Wiener-Hermite functionals with respect to the Gaussian probability density functional $\mathcal{P}_{\rm G}$ [51]

$$\begin{aligned}
& \langle\mathcal{H}_n^\star(\mathbf{k}_1,\ldots,\mathbf{k}_n)\mathcal{H}_m(\mathbf{k}_1',\ldots,\mathbf{k}_m')\rangle_{\rm G} \\
& \qquad = \delta_{nm}\Big[\delta_{\rm D}(\mathbf{k}_1-\mathbf{k}_1')\cdots\delta_{\rm D}(\mathbf{k}_n-\mathbf{k}_m') \\
& \qquad\quad +{\rm perm}(\mathbf{k}_1,\ldots,\mathbf{k}_n)\Big]
\end{aligned} \tag{25}$$

where ${\rm perm}(\mathbf{k}_1,\ldots,\mathbf{k}_n)$ stands for the $(n!-1)$ terms to symmetrize the previous term $\delta_{\rm D}(\mathbf{k}_1-\mathbf{k}_1')\cdots\delta_{\rm D}(\mathbf{k}_n-\mathbf{k}_m')$ with respect to the permutations of its arguments $\mathbf{k}_1,\ldots,\mathbf{k}_n$. The Dirac delta function that appears in Eq. (21) is due to the (statistical) translational invariance of space. The expansion coefficient functions $\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)$ can be derived based on the orthogonality relation in Eq. (25). If we multiply $\mathcal{H}_m(\mathbf{k}_1,\ldots,\mathbf{k}_m)$ on

both sides of Eq. (21) and take their expectation value with respect to the Gaussian probability density functional, we would have

$$\begin{aligned}&(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_n-\mathbf{k})\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\\&=(2\pi)^{2n}\langle\tilde{\mathcal{F}}(\mathbf{k})\mathcal{H}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\rangle_{\rm G}\\&=(2\pi)^{2n}\left\langle\frac{\partial^n\tilde{\mathcal{F}}(\mathbf{k})}{\partial\tilde{f}(\mathbf{k}_1)\cdots\partial\tilde{f}(\mathbf{k}_n)}\right\rangle_{\rm G}\ ,\end{aligned}\tag{26}$$

where the first equation is based on the orthogonal relation in Eq. (25) and the second equation makes use of the definition of the Wiener-Hermite functional in Eq. (19) followed by an integration by parts. To derive the final expression for $\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)$, we apply Fourier transform to the above equation with respect to $\mathbf{k}$:

$$\begin{aligned}&\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\\&=(2\pi)^{2n}e^{i(\mathbf{k}_1+\cdots+\mathbf{k}_n)\cdot\mathbf{x}}\langle\mathcal{F}(\mathbf{x})\mathcal{H}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\rangle_{\rm G}\\&=(2\pi)^{2n}e^{i(\mathbf{k}_1+\cdots+\mathbf{k}_n)\cdot\mathbf{x}}\left\langle\frac{\partial^n\mathcal{F}(\mathbf{x})}{\partial\tilde{f}(\mathbf{k}_1)\cdots\partial\tilde{f}(\mathbf{k}_n)}\right\rangle_{\rm G}\ ,\end{aligned}\tag{27}$$

where we can further set $\mathbf{x}=0$ due to the translational invariance of $\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)$ and conveniently evaluate it to be

$$\mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)=(2\pi)^{2n}\left\langle\frac{\partial^n\mathcal{F}(\mathbf{x})}{\partial\tilde{f}(\mathbf{k}_1)\cdots\partial\tilde{f}(\mathbf{k}_n)}\right\rangle_{\rm G}\ .\tag{28}$$

Following the above equation, the expansion coefficient functions can be interpreted as the Gaussian n-point response of the 2D functional $\mathcal{F}$ to the underlying density field. Conceptually this is analogous to the large-scale galaxy bias and therefore can be thought of in the same way for 2D critical points in this work. Note that similar expansion coefficients have been used in the context of galaxy clustering [52, 53].

With a proper understanding of Eq. (20) and (21), we can substitute them with the corresponding terms in Eq. (18) and expand the whole equation. During the process, one recurrent term is $\langle\mathcal{H}^\star_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\mathcal{H}^\star_m(\mathbf{k}'_1,\ldots,\mathbf{k}'_m)\mathcal{H}_l(\mathbf{k}^{''}_1,\ldots,\mathbf{k}^{''}_l)\rangle_{\rm G}$. One can compute these terms by solving both $\mathcal{H}_n$ and $\mathcal{H}^\star_n$ at each order explicitly using Eq. (19)and (24). Here we give examples of the first few $\mathcal{H}^\star_n$ expressions:

$$\begin{aligned}\mathcal{H}^\star_0=&\ 1\ ,\\\mathcal{H}^\star_1(\mathbf{k})=&\ \tilde{f}(\mathbf{k})\ ,\\\mathcal{H}^\star_2(\mathbf{k}_1,\mathbf{k}_2)=&\ \tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)-(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2)P(k_1)\ ,\\\mathcal{H}^\star_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)=&\ \tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)-\Big[(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2)P(k_1)\tilde{f}(\mathbf{k}_3)+{\rm sym}\Big]\ ,\\\mathcal{H}^\star_4(\mathbf{k}_1,\ldots,\mathbf{k}_4)=&\ \tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\tilde{f}(\mathbf{k}_4)+\big[(2\pi)^4\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2)\delta_{\rm D}(\mathbf{k}_3+\mathbf{k}_4)P(k_1)P(k_3)+{\rm sym}\big]\\&-\Big[(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2)P(k_1)\tilde{f}(\mathbf{k}_3)\tilde{f}(\mathbf{k}_4)+{\rm sym}\Big]\ ,\end{aligned}\tag{29}$$

where "sym" stands for all non-repeating symmetric expressions of the previous term with respect to the $\mathbf{k}_1,\ldots,\mathbf{k}_n$ arguments. Such terms can be inserted into $\langle\mathcal{H}^\star_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\mathcal{H}^\star_m(\mathbf{k}'_1,\ldots,\mathbf{k}'_m)\mathcal{H}_l(\mathbf{k}^{''}_1,\ldots,\mathbf{k}^{''}_l)\rangle_{\rm G}$ and the whole expression can be evaluated by applying Wick's theorem for Gaussian statistics. As a result, $\langle\mathcal{H}^\star_n\mathcal{H}^\star_m\mathcal{H}_l\rangle_{\rm G}$ has nonzero value only when $n+m+l$ is an even number. Though straightforward, this computation becomes tedious very quickly. For example, in the case of NNLO where $n+m+l=8$, Wick's theorem predicts 105 terms from the contraction of eight density field $\tilde{f}$ alone. It is more convenient to evaluate such expressions using the diagrammatic method. We show the formalism in Appendix. B where we also derive all existing $\langle\mathcal{H}^\star_n\mathcal{H}^\star_m\mathcal{H}_l\rangle_{\rm G}$ factors up to NNLO. Readers can also refer to Appendix. A in Ref. [51] where the same formalism is presented but in real space.

With the above discussion, we can now show the resulting equation of $\langle\tilde{\mathcal{F}}(\mathbf{k})\tilde{\mathcal{F}}(\mathbf{k}')\rangle$ in Eq. (18) up to NNLO

$$
\begin{aligned}
&\langle\tilde{\mathcal{F}}^i(\mathbf{k})\tilde{\mathcal{F}}^j(\mathbf{k}')\rangle = (2\pi)^4\mathcal{G}_0^i\mathcal{G}_0^j + (2\pi)^2\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\mathcal{G}_1^i(\mathbf{k})\mathcal{G}_1^j(\mathbf{k})P(k)\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\int\frac{{\rm d}^2k_1}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)\mathcal{G}_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)P(k_1)P(|\mathbf{k}-\mathbf{k}_1|)\\
&+\frac{1}{6}\left[(2\pi)^2\mathcal{G}_0^i\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\int\frac{{\rm d}^2k_3}{(2\pi)^2}(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3-\mathbf{k})\mathcal{G}_3^j(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}\mathcal{G}_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)B(-\mathbf{k},\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_3^i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)\mathcal{G}_3^j(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)P(k_1)P(k_2)P(|\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|)\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})P(k)\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_4^j(\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)B(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)\mathcal{G}_3^j(\mathbf{k}-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)P(|\mathbf{k}-\mathbf{k}_1|)B(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^4}{24}\left[\mathcal{G}_0^i\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\int\frac{{\rm d}^2k_4}{(2\pi)^2}\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4-\mathbf{k})\mathcal{G}_4^j(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\tilde{f}(\mathbf{k}_4)\rangle_{\rm c}+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_3^j(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)T(-\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{4}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1)\mathcal{G}_2^j(\mathbf{k}-\mathbf{k}_2,\mathbf{k}_2)T(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}-\mathbf{k}_2)+(i\leftrightarrow j)\right]\ , \qquad (30)
\end{aligned}
$$

where we exploited the following parity symmetries:

$$
\begin{aligned}
\mathcal{G}_n(-\mathbf{k}_1,\ldots,-\mathbf{k}_n) &= \mathcal{G}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\\
B(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{k}_3) &= B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\
T(-\mathbf{k}_1,\ldots,-\mathbf{k}_4) &= T(\mathbf{k}_1,\ldots,\mathbf{k}_4)\ . \qquad (31)
\end{aligned}
$$

There are two terms in the above equation which contain $\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}$ and $\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\tilde{f}(\mathbf{k}_4)\rangle_{\rm c}$ respectively. We use such notations to distinguish them from the rest as these two terms represent the unconnected parts in the 2nd-order moment of $\tilde{\mathcal{F}}$ and will be subtracted off as we will show below. Additionally, in Eq. (30) we characterize the functional $\tilde{\mathcal{F}}$ with indices i, j which denote different types of critical points (i.e., i,j = peaks, voids, saddle points) through which we can construct both auto- and cross-2PCFs. The symbol $(i\leftrightarrow j)$ denotes the addition of a term possessing the same form but exchanging the correspondence of the expansion coefficient function $\mathcal{G}_n$ to the other functional type. From Eq. (30) we observe that the leading-order (LO) result is composed of the power spectrum $P(k)$ of the underlying density field. The NLO result is proportional to $P(k)^2$ including the bispectrum contribution (at tree-level) which is also the lowest-order non-Gaussian correction. Finally the NNLO result is proportional to $P(k)^3$ which includes contribution from $P\times B$ and T terms where both bispectrum and trispectrum are at tree-level.

To compute the connected part of the 2nd-order moment, we need $\langle\tilde{\mathcal{F}}\rangle$ which can be calculated based on Eqs. (20), (21) and (25)

$$
\begin{aligned}
&\langle\tilde{\mathcal{F}}\rangle = (2\pi)^2\mathcal{G}_0 + \frac{1}{6}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\frac{{\rm d}^2k_2}{(2\pi)^2}\frac{{\rm d}^2k_3}{(2\pi)^2}(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3-\mathbf{k})\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\mathcal{G}_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\
&+\frac{1}{24}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_4}{(2\pi)^2}(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_4-\mathbf{k})\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_4)\rangle_{\rm c}\mathcal{G}_4(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\\
&+\frac{1}{72}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_6}{(2\pi)^2}(2\pi)^2\delta_{\rm D}(\mathbf{k}_1+\cdots+\mathbf{k}_6-\mathbf{k})\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\langle\tilde{f}(\mathbf{k}_4)\tilde{f}(\mathbf{k}_5)\tilde{f}(\mathbf{k}_6)\rangle_{\rm c}\mathcal{G}_6(\mathbf{k}_1,\ldots,\mathbf{k}_6)\\
&+\cdots\ , \qquad (32)
\end{aligned}
$$

where we can easily observe that the first three lowest-order terms in the expansion of $\langle\tilde{\mathcal{F}}^i\rangle\langle\tilde{\mathcal{F}}^j\rangle$ are the constant $(2\pi)^4\mathcal{G}_0^i\mathcal{G}_0^j$ and the two terms we mentioned above in Eq. (30). By subtracting them from Eq. (30), we can summarize the connected 2nd-order moment of the functional $\tilde{\mathcal{F}}$ to be

$$
\begin{aligned}
&\langle\tilde{\mathcal{F}}^i(\mathbf{k})\tilde{\mathcal{F}}^j(\mathbf{k}')\rangle_{\rm c} = (2\pi)^2\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\mathcal{G}_1^i(\mathbf{k})\mathcal{G}_1^j(\mathbf{k})P(k)\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\int\frac{{\rm d}^2k_1}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)\mathcal{G}_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)P(k_1)P(|\mathbf{k}-\mathbf{k}_1|)\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}\mathcal{G}_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)B(-\mathbf{k},\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_3^i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)\mathcal{G}_3^j(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)P(k_1)P(k_2)P(|\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|)\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})P(k)\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_4^j(\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)B(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{2}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)\mathcal{G}_3^j(\mathbf{k}-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_1+\mathbf{k}_2)P(|\mathbf{k}-\mathbf{k}_1|)B(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{6}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\mathcal{G}_1^i(\mathbf{k})\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_3^j(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)T(-\mathbf{k},-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}+\mathbf{k}_1+\mathbf{k}_2)+(i\leftrightarrow j)\right]\\
&+\frac{(2\pi)^2}{4}\delta_{\rm D}(\mathbf{k}+\mathbf{k}')\left[\int\frac{{\rm d}^2k_1}{(2\pi)^2}\int\frac{{\rm d}^2k_2}{(2\pi)^2}\mathcal{G}_2^i(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1)\mathcal{G}_2^j(\mathbf{k}-\mathbf{k}_2,\mathbf{k}_2)T(\mathbf{k}-\mathbf{k}_1,\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}-\mathbf{k}_2)+(i\leftrightarrow j)\right]\ . \qquad (33)
\end{aligned}
$$

The last piece needed to complete the derivation of $P_{\mathcal{F}}(k)$ is $\langle\mathcal{F}\rangle$ which can be calculated based on Eqs. (20) and (27)

$$
\begin{aligned}
\langle\mathcal{F}\rangle =&\mathcal{G}_0+\frac{1}{6}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\frac{{\rm d}^2k_2}{(2\pi)^2}\frac{{\rm d}^2k_3}{(2\pi)^2}\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\mathcal{G}_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\
&+\frac{1}{24}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_4}{(2\pi)^2}\langle\tilde{f}(\mathbf{k}_1)\ldots\tilde{f}(\mathbf{k}_4)\rangle_{\rm c}\mathcal{G}_4(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\\
&+\frac{1}{72}\int\frac{{\rm d}^2k_1}{(2\pi)^2}\cdots\frac{{\rm d}^2k_6}{(2\pi)^2}\langle\tilde{f}(\mathbf{k}_1)\tilde{f}(\mathbf{k}_2)\tilde{f}(\mathbf{k}_3)\rangle_{\rm c}\langle\tilde{f}(\mathbf{k}_4)\tilde{f}(\mathbf{k}_5)\tilde{f}(\mathbf{k}_6)\rangle_{\rm c}\mathcal{G}_6(\mathbf{k}_1,\ldots,\mathbf{k}_6)+\ldots\ . \qquad (34)
\end{aligned}
$$

By combining Eqs. (33) and (34), we can compute $P_{\mathcal{F}}(k)$ in Eq. (17). Note that the non-Gaussian corrections of Eq. (34) in the denominator of Eq. (17) would not contribute to the NLO but they contribute to higher-order results in general. With this in mind, we can derive the expression for the power spectrum up to NNLO as

$$
\begin{aligned}
P^{ij}_{\mathcal{F}}(k) =& g^i_1(\mathbf{k}) g^j_1(\mathbf{k}) P(k) \\
&+\frac{1}{2}\int \frac{\mathrm{d}^2k_1}{(2\pi)^2} g^i_2(\mathbf{k}_1, \mathbf{k}-\mathbf{k}_1) g^j_2(\mathbf{k}_1, \mathbf{k}-\mathbf{k}_1) P(k_1) P(|\mathbf{k}-\mathbf{k}_1|) \\
&+\frac{1}{2}\left[g^i_1(\mathbf{k}) \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} g^j_2(\mathbf{k}_1, \mathbf{k}-\mathbf{k}_1) B(-\mathbf{k}, \mathbf{k}_1, \mathbf{k}-\mathbf{k}_1) + (i \leftrightarrow j) \right] \\
&+\frac{1}{6}\int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^i_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}-\mathbf{k}_1-\mathbf{k}_2) g^j_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}-\mathbf{k}_1-\mathbf{k}_2) P(k_1) P(k_2) P(|\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|) \\
&+\frac{1}{6}\left[g^i_1(\mathbf{k}) P(k) \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^j_4(\mathbf{k}, -\mathbf{k}_1, -\mathbf{k}_2, \mathbf{k}_1+\mathbf{k}_2) B(-\mathbf{k}_1, -\mathbf{k}_2, \mathbf{k}_1+\mathbf{k}_2) + (i \leftrightarrow j) \right] \\
&+\frac{1}{2}\left[\int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^i_2(\mathbf{k}_1, \mathbf{k}-\mathbf{k}_1) g^j_3(\mathbf{k}-\mathbf{k}_1, -\mathbf{k}_2, \mathbf{k}_1+\mathbf{k}_2) P(|\mathbf{k}-\mathbf{k}_1|) B(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1-\mathbf{k}_2) + (i \leftrightarrow j) \right] \\
&+\frac{1}{6}\left[g^i_1(\mathbf{k}) \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^j_3(-\mathbf{k}_1, -\mathbf{k}_2, \mathbf{k}+\mathbf{k}_1+\mathbf{k}_2) T(-\mathbf{k}, -\mathbf{k}_1, -\mathbf{k}_2, \mathbf{k}+\mathbf{k}_1+\mathbf{k}_2) + (i \leftrightarrow j) \right] \\
&+\frac{1}{4}\left[\int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^i_2(\mathbf{k}-\mathbf{k}_1, \mathbf{k}_1) g^j_2(\mathbf{k}-\mathbf{k}_2, \mathbf{k}_2) T(\mathbf{k}-\mathbf{k}_1, \mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}-\mathbf{k}_2) + (i \leftrightarrow j) \right] \\
&-\frac{1}{6}\left[g^i_1(\mathbf{k}) g^j_1(\mathbf{k}) P(k) \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \int \frac{\mathrm{d}^2k_2}{(2\pi)^2} g^j_3(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1-\mathbf{k}_2) B(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1-\mathbf{k}_2) + (i \leftrightarrow j) \right] \ , \qquad (35)
\end{aligned}
$$

where

$$
g_n(\mathbf{k}_1, \ldots, \mathbf{k}_n) \equiv \frac{\mathcal{G}_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)}{\mathcal{G}_0} \ , \qquad (36)
$$

and the last line is the contribution from non-Gaussian corrections in the denominator of Eq. (17). One can then apply a Hankel transformation to Eq. (35) to evaluate the correlation function

$$
\xi^{ij}_{\mathcal{F}}(r) = \int \frac{k\mathrm{d}k}{2\pi} J_0(kr) P^{ij}_{\mathcal{F}}(k) \ , \qquad (37)
$$

where $J_n(x)$ is the Bessel function. Note that in the above Eq. (35), if the underlying density field is Gaussian distributed, all terms containing bispectrum or trispectrum would vanish and we will recover the perturbative bias expansion in the context of Gaussian approximation.

So far, all the above results are general for any 2D functionals $\mathcal{F}$ of a density field f. We still need to answer one question before we can actually compute the 2PCFs for 2D critical points in a mildly non-Gaussian regime using Eqs. (35) and (37), that is what are the expressions for $g_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ for 2D critical points. From the previous Eq. (28), we know that they can be derived directly by substituting $\mathcal{F}$ in the equation with the corresponding number density function of a critical point type, such as the one for peaks in Eq. (9). The required functional derivatives contain very technical calculation and for our purpose of numerically evaluating Eq. (35) to the NLO, we directly present the results below. Readers who are interested in the derivation details can refer to the method presented in Appendix. B of Ref. [43].

$$
g_1(\mathbf{k}) = g_{10000} + g_{01000} k^2 \ , \qquad (38)
$$

$$
\begin{aligned}
g_2(\mathbf{k}_1, \mathbf{k}_2) =& g_{20000} + g_{11000}(k_1^2 + k_2^2) + g_{02000} k_1^2 k_2^2 \\
&- 2 g_{00100} \mathbf{k}_1 \cdot \mathbf{k}_2 \\
&+ 4 g_{00010} \left[(\mathbf{k}_1 \cdot \mathbf{k}_2)^2 - \frac{1}{2} k_1^2 k_2^2 \right] \ , \qquad (39)
\end{aligned}
$$

where the coefficients g_{ijklm} include the constraints imposed on the density field by the critical point functional and can be expressed as

$$
g_{ijklm} = \frac{G_{ijklm}}{\sigma_0^i \sigma_1^{2k} \sigma_2^{j+2l+3m} G_{00000}} \ , \qquad (40)
$$

and the numerator factor G_{ijklm} is

$$
\begin{aligned}
G_{ijklm}(\nu) = \frac{1}{2\pi} \left(\frac{\sigma_2}{\sqrt{2}\sigma_1} \right)^2 X_k \int \mathrm{d}x \, H_{i-1,j}(\nu, x) \\
\times \mathcal{N}(\nu, x) f_{lm}(x) \ , \qquad (41)
\end{aligned}
$$

which is a function of the threshold ν as discussed in Eq. (9). In the integration above, x is exactly the trace of the negative Hessian matrix $(-\zeta)$ and equivalent to J_1 defined in Eq. (16). X_k is a constant and from Eqs. (38) and (39) we only need X_0 and X_1 which are 1 and -1 respectively. The function $H_{ij}(\alpha, J_1)$ is the multivariate Hermite polynomials defined as

$$
H_{ij}(\alpha, J_1) = \frac{1}{\mathcal{N}(\alpha, J_1)} \left(-\frac{\partial}{\partial \alpha} \right)^i \left(-\frac{\partial}{\partial J_1} \right)^j \mathcal{N}(\alpha, J_1) \ , \qquad (42)
$$

where the $\mathcal{N}(\alpha, J_1)$ function is defined previously in Eq. (15). In the case of g_{0jklm}, we need to calculate $H_{-1,j}$ which is

$$
H_{-1,j}(\alpha, J_1) = \frac{1}{\mathcal{N}(\alpha, J_1)} \int_\alpha^\infty \mathrm{d}\beta H_{0j}(\beta, J_1) \mathcal{N}(\beta, J_1) \ . \qquad (43)
$$

Again from Eqs. (38) and (39) we observe that we only need functions $f_{l0}(x)$ which are (for the general definition equation of f_{lm}, please refer to Ref. [43])

$$f_{l0} = 8\int \mathrm{d}y y e^{-4y^2}(x^2-4y^2)(-1)^l L_l(4y^2)\ , \tag{44}$$

where y is defined as $y \equiv (\lambda_1-\lambda_2)/2$ and L_l is the generalized Laguerre polynomial $L_l^{(n)}$ with the index $n=1$

$$L_l(x) = \frac{e^x}{xl!}\frac{\mathrm{d}^l}{\mathrm{d}x^l}(x^{l+1}e^{-x})\ . \tag{45}$$

By replacing G_{ijklm} factors in Eq. (40) with Eq. (41), we can have the expression

$$g_{ijklm}(\nu) = \frac{X_k\int \mathrm{d}x H_{i-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{lm}(x)}{\sigma_0^i\sigma_1^{2k}\sigma_2^{j+2l+3m}\int \mathrm{d}x H_{-1,0}(\nu,x)\mathcal{N}(\nu,x)f_{00}(x)}\ , \tag{46}$$

which can be evaluated by Eq. (15) and Eqs. (42) to (45).

As discussed in Sec. III, different critical points are characterized by their eigenvalues of the Hessian matrix. This characterization is reflected in the integration limits of Eq. (46) through which we can then compute the g_{ijklm} factors for peaks, voids and saddle points separately [42]. As discussed above, the integration variable $x = J_1 = \lambda_1+\lambda_2$, which ranges from 0 to ∞ for peaks, $-\infty$ to 0 for voids and $-\infty$ to ∞ for saddle points. Another integration limit needs considering is in Eq. (44), where y is strictly positive as we have already assumed $\lambda_1 > \lambda_2$. Furthermore we have $y^2 = \left[(\lambda_1+\lambda_2)^2 - 4\lambda_1\lambda_2\right]/4 = (x^2-4\lambda_1\lambda_2)/4$ where both peaks and voids have $\lambda_1\lambda_2 > 0$, therefore we have $y < x/2$ ($x > 0$) for peaks and $y < -x/2$ ($x < 0$) for voids. On the other hand, saddle points always have $\lambda_1\lambda_2 < 0$, thus the integration limit in Eq. (44) would become $y > |x|/2$. To summarize, we have the following equations for g_{ijklm} factors for different types of critical points

$$g_{ijklm}^{peak} = \frac{X_k\int_0^\infty \mathrm{d}x H_{i-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{lm}^{peak}(x)}{\sigma_0^i\sigma_1^{2k}\sigma_2^{j+2l+3m}\int_0^\infty \mathrm{d}x H_{-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{00}^{peak}(x)}\ , \tag{47}$$

$$f_{lm}^{peak} = f_{l0}^{peak} = 8\int_0^{\frac{x}{2}} \mathrm{d}y y e^{-4y^2}(x^2-4y^2)(-1)^l L_l(4y^2)\ , \tag{48}$$

$$g_{ijklm}^{void} = \frac{X_k\int_{-\infty}^0 \mathrm{d}x H_{i-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{lm}^{void}(x)}{\sigma_0^i\sigma_1^{2k}\sigma_2^{j+2l+3m}\int_{-\infty}^0 \mathrm{d}x H_{-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{00}^{void}(x)}\ , \tag{49}$$

$$f_{lm}^{void} = f_{l0}^{void} = 8\int_0^{-\frac{x}{2}} \mathrm{d}y y e^{-4y^2}(x^2-4y^2)(-1)^l L_l(4y^2)\ , \tag{50}$$

and the analytical integration results of the f_{lm} function are the same for both peaks and voids

$$\begin{aligned} f_{00}^{peak/void}(x) &= e^{-x^2}+x^2-1 \\ f_{10}^{peak/void}(x) &= (1+x^2)e^{-x^2}-1\ , \end{aligned} \tag{51}$$

whereas

$$g_{ijklm}^{saddle} = \frac{X_k\int_{-\infty}^\infty \mathrm{d}x H_{i-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{lm}^{saddle}(x)}{\sigma_0^i\sigma_1^{2k}\sigma_2^{j+2l+3m}\int_{-\infty}^\infty \mathrm{d}x H_{-1,j}(\nu,x)\mathcal{N}(\nu,x)f_{00}^{saddle}(x)}\ , \tag{52}$$

$$f_{lm}^{saddle} = f_{l0}^{saddle} = 8\int_{\frac{|x|}{2}}^\infty \mathrm{d}y y e^{-4y^2}(x^2-4y^2)(-1)^l L_l(4y^2)\ , \tag{53}$$

and

$$\begin{aligned} f_{00}^{saddle} &= -e^{-x^2} \\ f_{10}^{saddle} &= -e^{-x^2}(1+x^2)\ . \end{aligned} \tag{54}$$

In Appendix C, we display plots of the seven g_{ijklm} factors in Eqs. (38) and (39) as functions of the threshold ν for different types of critical points.

V. RESULTS FOR THE 2PCFS OF 2D WEAK LENSING CRITICAL POINTS

Building on the formalism for calculating the 2PCFs of 2D critical points in a mildly non-Gaussian regime introduced in the previous section, we apply it to the weak lensing convergence field κ. This particular 2D field can be understood as the weighted line-of-sight projection of the 3D cosmic matter density contrast field [54, 55]

$$\kappa = \int \mathrm{d}\chi q(\chi)\delta(\chi)\ , \tag{55}$$

where $\delta(\chi)$ is the 3D matter density contrast at a comoving radial distance χ and $q(\chi)$ is the weight function of the convergence field along the line-of-sight [56]

$$q(\chi) \equiv \frac{3H_0^2\Omega_{m,0}}{2a(\chi)c^2}\frac{\chi(\chi_s-\chi)}{\chi_s}\ , \tag{56}$$

where H_0, $\Omega_{m,0}$ and c are the Hubble constant, matter density parameter at the present and the speed of light respectively. The $a(\chi)$ function is the scale factor of the Universe and χ_s is the comoving distance to the source galaxies. Here we only consider the case where all source galaxies are located at a Dirac delta like source redshift distribution, but it is straightforward to extend the description of $q(\chi)$ to a general distribution of source galaxies [55].

Adopting the flat-sky and Limber approximations [57], one can derive the power spectrum and bispectrum of the convergence field [58]

$$P_\kappa(k) = \int \mathrm{d}\chi \frac{q^2(\chi)}{\chi^2}P_m\left(\frac{k}{\chi},\chi\right)\ , \tag{57}$$

$$B_\kappa(k_1,k_2,k_3) = \int \mathrm{d}\chi \frac{q^3(\chi)}{\chi^4}B_m\left(\frac{k_1}{\chi},\frac{k_2}{\chi},\frac{k_3}{\chi},\chi\right)\ , \tag{58}$$

where P_m and B_m are the 3D matter power spectrum and bispectrum.

We smooth the convergence field with a smoothing kernel $W(kR)$, where R is the smoothing angle, before we apply Eqs. (57) and (58) to Eq. (35). The resulting smoothed convergence power spectrum and bispectrum then read $P(k) = W(kR)^2 P_\kappa(k)$ and $B(k_1, k_2, k_3) = W(k_1R)W(k_2R)W(k_3R)B_\kappa(k_1, k_2, k_3)$. In practice, we use a Gaussian smoothing kernel, $W(kR) = e^{-k^2R^2/2}$. From there, the power spectrum of weak lensing critical points up to NLO reads

$$\begin{aligned}P^{ij}(k)=& g_1^i(\mathbf{k})g_1^j(\mathbf{k})W(kR)^2P_\kappa(k)\\ &+\frac{1}{2}\int\frac{\mathrm{d}^2k_1}{(2\pi)^2}g_2^i(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)g_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)W(k_1R)^2W(|\mathbf{k}-\mathbf{k}_1|R)^2P_\kappa(k_1)P_\kappa(|\mathbf{k}-\mathbf{k}_1|)\\ &+\frac{1}{2}\left[g_1^i(\mathbf{k})\int\frac{\mathrm{d}^2k_1}{(2\pi)^2}g_2^j(\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)W(kR)W(k_1R)W(|\mathbf{k}-\mathbf{k}_1|R)B_\kappa(-\mathbf{k},\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1)+(i\leftrightarrow j)\right] . \end{aligned} \tag{59}$$

In the above equation, the spectral moments from Eq. (7) exploited in g_n functions has the following form

$$\sigma_n^2 = \int \mathrm{d}\chi \frac{q^2(\chi)}{\chi^2}\int\frac{k\mathrm{d}k}{2\pi}k^{2n}W(kR)^2P_m\left(\frac{k}{\chi},\chi\right) , \tag{60}$$

and we are going to use the tree-level 3D matter bispectrum in Eq. (58) [59]

$$B_m(k_1,k_2,k_3) = 2F_2(\mathbf{k}_1,\mathbf{k}_2)P_m(k_1)P_m(k_2) + \text{perm} , \tag{61}$$

where $F_2(\mathbf{k}_1, \mathbf{k}_2)$ is the symmetric coupling kernel

$$F_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2} + \frac{1}{2}\mathbf{k}_1\cdot\mathbf{k}_2\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right) , \tag{62}$$

and "perm" in Eq. (61) represents the same term but with cyclic permutations on arguments $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ (such that the bispectrum ends up being a sum of three such terms). Note that we use the nonlinear 3D matter power spectrum in Eqs. (57) and (61). For this purpose we apply the fitting function HaloFit [60, 61] adopted in the Boltzmann solver package Class [62].

In the following computation, we adopt the flat ΛCDM model with Planck 2018 cosmological parameters [63]: $\Omega_{\rm cdm}h^2 = 0.1201$, $\Omega_b h^2 = 0.02238$, $h = 0.6732$, $n_{\rm s} = 0.9660$ and $\sigma_8 = 0.8120$. We assume that all source galaxies are located at $z = 1.5$. The smoothing scale is $R = 15'$ for the Gaussian kernel which corresponds to approximately 20 Mpc at the source redshift with the above background cosmology. We applied a fast and accurate numerical evaluation of Eq. (59) using the method of separation of integration variables. After the separation we perform angular integrations first and the resulting expression can be computed by multiple one-dimensional Fourier transform. Readers can refer to Appendix. D or Sec. IIIB in Ref. [43] for more technical details.

We show in Fig. 3 the predicted auto power spectrum of peak-peak, void-void, and saddle-saddle, as well as the cross power spectrum of peak-void, peak-saddle and void-saddle, at a threshold $\nu = 0.3$ where the value is taken with respect to the smoothed σ_0 shown in Eq. (60). Above this given threshold, together with the above cosmology and lensing parameters, we compute the abundances of peaks and voids to be approximately 80% and 6% of the total number using a Monte Carlo (MC) integration method, assuming an underlying Gaussian random field. We first notice that there is a discrepancy on $k \to 0$ scale between the LO and 2nd-order Gaussian approximation for peaks and voids. It has been suggested in previous works that the exclusion zone in 2PCFs for peaks and halos would non-trivially impact the power spectrum on large scales [9, 64–66]. It has also been shown in Ref. [43] for 3D peaks that this nonzero value in the limit of $k \to 0$, corresponding to unphysical component in the perturbative expansion, only exists in the 2nd-order Gaussian approximation term in the NLO, but not in other components. In Fig. 3, we observe this effect not only in 2D weak lensing peak power spectrum $P_{pp}(k)$, but also in voids $P_{vv}(k)$ which is caused by the exclusion zone between two voids. Meanwhile, such effect exists but not significant for the saddle point power spectrum $P_{ss}(k)$ and its cross power spectrum with peaks and voids $P_{ps}(k)$, $P_{vs}(k)$. This suggests that there is no strong exclusion effect between saddle points and other types of critical points (at the same threshold) since the matter flows through filaments (a type of saddle point) that are closely connected to either peaks or voids. This is because when the thresholds are the same, curvature and gradient constraints can be smoothly mapped from one to the other, contrary to peak-void for instance where the gradient constraint and the sign of the curvature impose two configurations that are incompatible in the zero separation limit. When transformed to real space for 2PCFs calculation, the above mentioned zero-lag value would turn into a Dirac delta-like function on small angular separations and thus not impact the convergence among different orders of perturbative bias expansion on large angular scales. This is confirmed in Fig. 4 where we show the corresponding 2PCFs.

For all 2PCFs, results from different orders of perturbative bias expansion converge with respect to each other on large angular separations. There are amplitude incre-

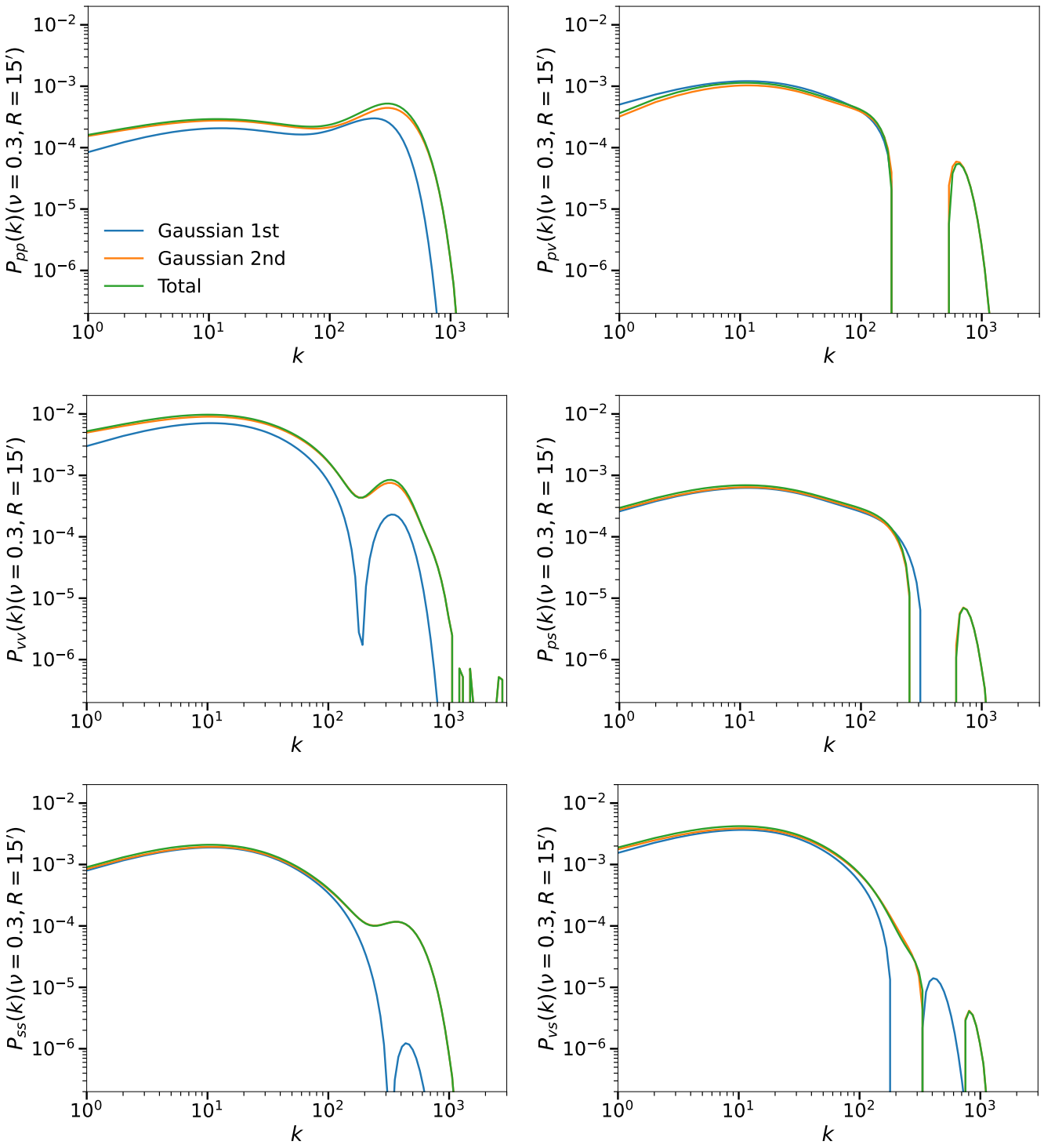

FIG. 3. Auto and cross power spectrum of different critical points in 2D weak lensing fields above a given threshold $\nu = 0.3$ and a smoothing scale $R = 15'$. The subscript "p" represents peaks while "v" and "s" stand for voids and saddle points respectively. Blue curve represents the LO in the power spectrum, corresponding to the first line on the right hand side of Eq. (59). Orange curve is the sum of the LO and the 2nd-order Gaussian approximation ($\propto P(k)^2$ term) in the NLO, which is the second line term in Eq. (59). Green curve is the full NLO prediction including the bispectrum correction expressed by the third line term in Eq. (59). The color curves in all the other sub-panels have the same representation as that denoted in the top left subplot. The fluctuations on large k scales are the residuals of the unphysical components from the perturbative bias expansion after smoothing. They will not impact the 2PCFs on intermediate and large angular separations after the Hankel transform as we will show in Sec. VI with the peak 2PCF as an example.

ments for ξ_{pp}, ξ_{vv}, ξ_{ss} and ξ_{pv} towards small angular scales starting between $30'$ and $70'$, those are caused by

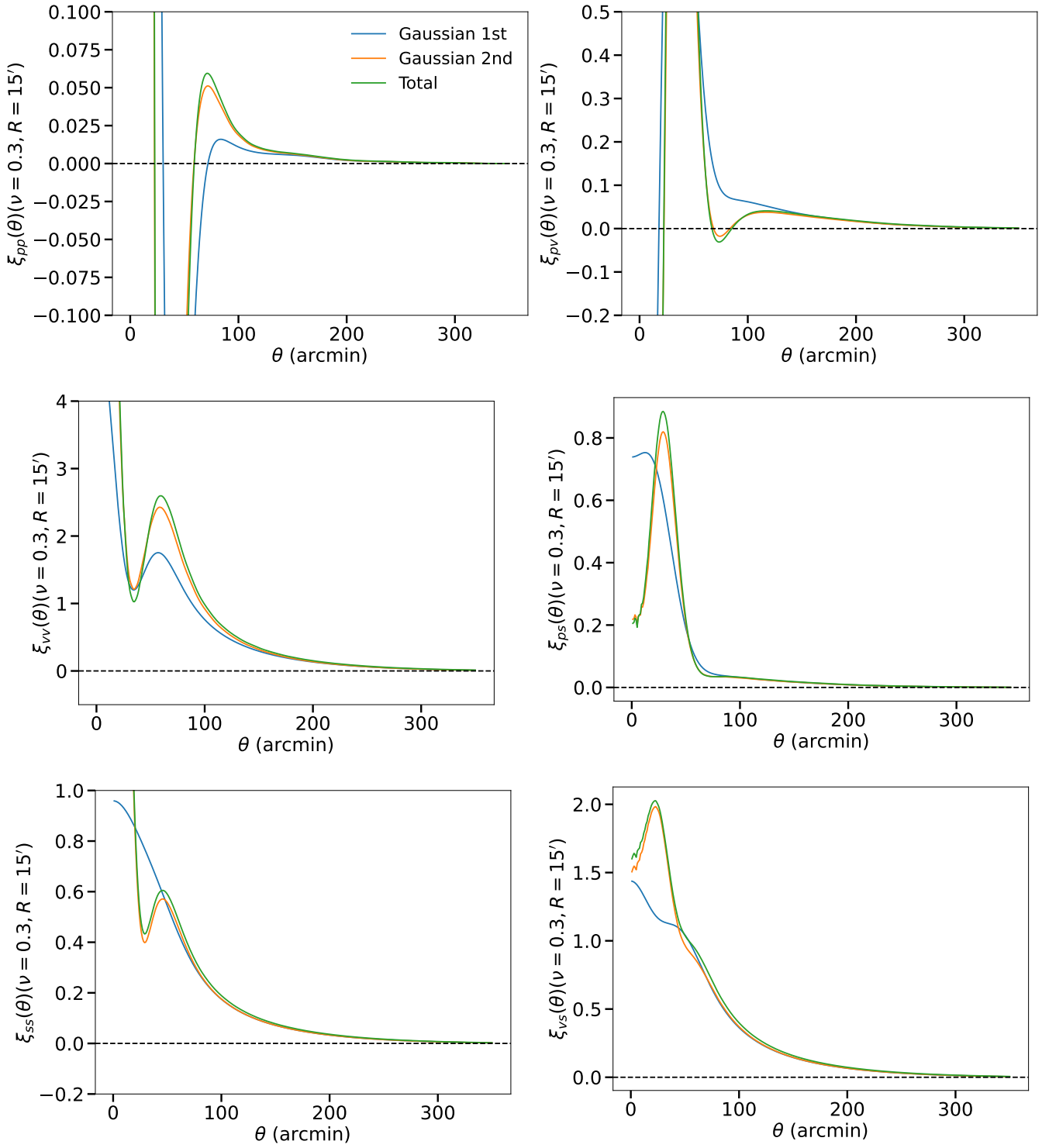

FIG. 4. Auto and cross 2PCF of different critical points in 2D weak lensing fields above a given threshold $\nu = 0.3$ and a smoothing scale $R = 15'$. The subscript "p" represents peaks while "v" and "s" stand for voids and saddle points respectively. Blue curve represents the LO in the 2PCF, which is Hankel transformed from the first line on the right hand side of Eq. (59). The orange curve is the sum of the LO and the 2nd-order Gaussian approximation (whose Fourier counterpart is the $\propto P(k)^2$ term) in the NLO, which is the sum of the Hankel transform of the first two line terms in Eq. (59). The green curve is the full NLO prediction including the bispectrum correction, i.e. the Hankel transform of the complete expression of Eq. (59). The color curves in all the other sub-panels have the same representation as that denoted in the top left subplot.

the unphysical component in the perturbative expansion prediction which cannot correctly capture the non per-

turbative exclusion zones. Compared to the 2nd-order Gaussian approximation, the shape of the non-Gaussian correction is almost proportional to the Gaussian contribution on most of the scales. Thereby, the total shape of the extreme 2PCF does not change much by the effect of non-Gaussianity, but the amplitude does change, especially around their maxima beyond the unphysical angular scales. Among the auto 2PCFs, ξ_{pp} reaches its maximum around $70'$ which is larger than that of ξ_{vv} at $60'$ and in turn larger than $45'$ for ξ_{ss}. This relationship is consistent with what was found in 3D [21] where peak 2PCF has the largest maximum separation, followed by that of voids and then saddle points. This might be due to the different rarity of critical points above the same threshold, thus causes different characteristic separations in their clustering.

If we examine ξ_{ps} and ξ_{vs} instead, we would find that their 2PCF maxima are on much smaller angular separations compared to the other four 2PCFs. This again indicates that saddle points have a weak exclusion effect with respect to peaks and voids above the same threshold, reflecting a shell-like structure in the distribution of saddle points around a given extremum, reminiscent of a cubic crystal lattice as pointed out by Refs. [21, 66]. It implies that cross 2PCFs between saddle points and other extreme may better explore small-scale physics.

Another interesting feature to observe is the existence of oscillations in ξ_{pp} on scales beyond $\theta = 100'$ once we multiply the signal by θ^2. The corresponding 2PCF is plotted in Fig. 5. At least two additional oscillation peaks at $\theta \approx 150'$ and $300'$ can be observed. This wiggly feature is related to the effect of BAOs in the underlying matter power spectrum. If we remove the baryon component from the underlying matter power spectrum and recalculate the peak 2PCF, we would obtain the result shown in the right panel of Fig. 5. The peaks at $150'$ and $300'$ both vanish and overall amplitude of the correlation function is reduced. It is well-known that baryonic features are highly suppressed in weak lensing power spectrum due to the line-of-sight projection shown in Eq. (57). However, these features can be enhanced in 3D peak correlation functions for both Gaussian and mildly non-Gaussian density field as pointed out in Refs. [6, 43, 67]. In this paper we confirm this property in 2D mildly non-Gaussian weak lensing fields. We do not observe such significant BAO related features in other types of correlation functions. We believe this is because peaks have larger curvatures in very overdense regions compared to other critical points, e.g. voids, and therefore can better amplify the strongly suppressed BAO features in the weak lensing convergence power spectrum. Since our $g_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ functions are computed with respect to the Gaussian random field (Eqs. (28) and (36)), based on the symmetry argument, the 2PCF of voids as a critical point should be able to exhibit BAO features if we modify our modeling and probe voids below an underdense threshold. However, we leave this aspect to future investigations.

We explore the influence of the Gaussian smoothing kernel scale, R, on the computed 2PCFs, focusing on the non-Gaussian effects arising from the inclusion of the bispectrum correction compared to the second-order Gaussian approximation. In Fig. 6, we vary R in the computation of the peak power spectrum (Eq. (59)) from $10'$ to $25'$. Using an angular separation of $150'$, which is significantly larger than the range of smoothing scales, we calculate the difference between the full NLO computation of the peak 2PCF and its second-order Gaussian approximation. The results show that the fractional difference decreases as the smoothing scale increases, indicating that stronger smoothing of the underlying cosmic density field reduces the non-Gaussian effects introduced by the bispectrum correction.

With the above discussion, we establish the analytical 2PCFs, including the mildly non-Gaussian correction, among all pairs of critical points in 2D weak lensing convergence field on large angular separations. Previous literature such as those cited in Sec. I showed that peak 2PCF is sensitive to cosmological parameters such as Ω_m and σ_8, and it can add constraining power to the inference of these parameters complementary to peak number count. However, those simulation-based models for the peak 2PCF do not extend to large angular separations where our model is valid and fast in its predictions. Therefore our model can contribute effectively to the cosmological inference. Additionally, the BAO features on large scales for peak 2PCFs can serve as an independent standard ruler [67] without galaxy bias by just looking at the amplitude of the fields in weak lensing maps, to study the evolution of LSS and test different cosmological models, an area that is currently not sufficiently explored.

VI. NUMERICAL PREDICTIONS COMPARED TO MONTE CARLO INTEGRATIONS

In this section, we aim to validate our perturbative bias expansion approach. In order to do this, we compare one of our predictions for 2PCFs of critical points in the previous section, the peak 2PCF $\xi_{pp}(\theta)$, to a full numerical integration of the peak 2PCF obtained by a MC integration method in MATHEMATICA (for comparison of other critical point 2PCFs, please refer to Appendix. E). In the MC integration, we assume a Gaussian probability density distribution for the underlying density field. This assumption guarantees that our full numerical integration result is exact and can be used to validate our theoretical prediction from perturbative bias expansion approach on large angular separations.

In practice, we use the same MC integration method as that presented in Sec. IV of Ref. [9] but in 2D. We draw random numbers of dimension 8 from the joint Gaussian conditional probability of (α, ζ_{ij}) at position $\mathbf{x}_1$ and (α, ζ_{ij}) at position $\mathbf{x}_2$ which satisfy $\eta_i = 0$. We only keep the drawn sample if α is above the amplitude threshold ν and eigenvalues of ζ are negative. With $\det|\zeta_{ij}^{(k)}(\mathbf{x}_1)|$

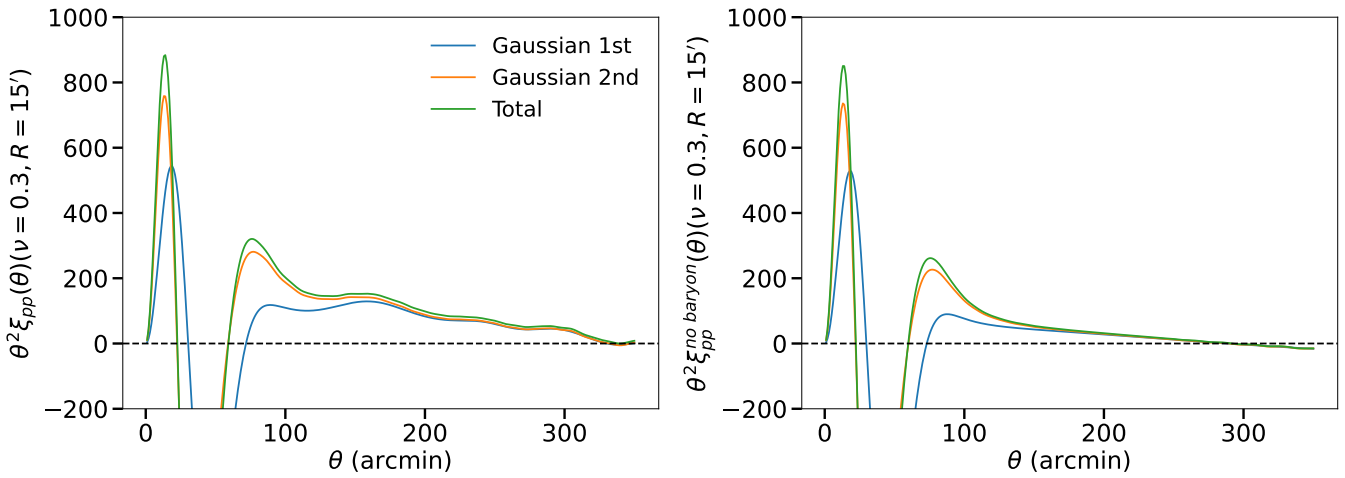

FIG. 5. *Left*: The same peak 2PCF as in Fig. 4 but multiplied by θ^2. *Right*: Same as the left panel, but the underlying matter power spectrum is calculated without baryons while kept at the same total matter density parameter.

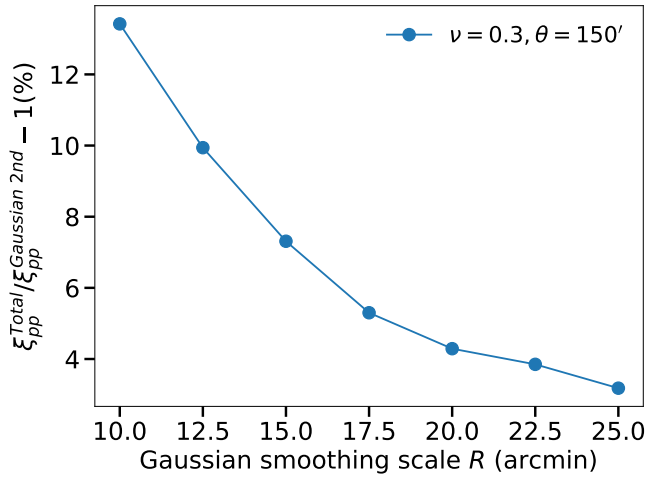

FIG. 6. The fractional difference in percentage between the full NLO computation of peak 2PCF and its corresponding 2nd-order Gaussian approximation at different smoothing scales, where the Gaussian smoothing scale R varies from $10'$ to $25'$. All fractional differences are calculated at a single angular separation $\theta = 150'$ and above the same threshold $\nu = 0.3$. Note that the spectral moments that enter for different smoothing scales follow that in Eq. (60).

and $\det|\zeta_{ij}^{(k)}(\mathbf{x}_2)|$ computed, we have

$$\langle n_p(\nu,\mathbf{x}_1)n_p(\nu,\mathbf{x}_2)\rangle \approx \frac{\mathcal{P}_G\left[\boldsymbol{\eta}(\mathbf{x}_1)=\boldsymbol{\eta}(\mathbf{x}_2)=\mathbf{0}\right]}{N}\sum_{k\in\mathcal{S}}\det|\zeta_{ij}^{(k)}(\mathbf{x}_1)|\det|\zeta_{ij}^{(k)}(\mathbf{x}_2)|\ , \tag{63}$$

where N is the total number of drawn sample, $\mathcal{S}$ is the subset of the drawn indices that correspond to the sample satisfying the conditions on the eigenvalues and amplitude. We can use the same procedure to evaluate the expectation value of the peak number density $\langle n_p(\nu,\mathbf{x})\rangle$. The peak 2PCF $\xi_{pp}(\theta)$ would therefore be

$$\xi_{pp}(\theta) = \frac{\langle n_p(\nu,\mathbf{x}_1)n_p(\nu,\mathbf{x}_2)\rangle}{\langle n_p(\nu,\mathbf{x})\rangle^2} - 1\ , \tag{64}$$

where the angular θ dependence is from the covariance matrix of the joint Gaussian conditional probability in $\langle n_p(\nu,\mathbf{x}_1)n_p(\nu,\mathbf{x}_2)\rangle$. The high dimensionality of the above integration makes the computation expensive, however, we parallelized the algorithm on a local cluster such that the calculation is completed within a reasonable period of time. Another subtlety in our MC integration method is that on small angular separations, $\theta \lesssim 40'$, the covariance matrix between two points at $\mathbf{x}_1$ and $\mathbf{x}_2$ would become non-invertible due to the numerical instability in the integrand of some entries. In order to qualitatively show the exclusion effect of the critical point clustering, which our perturbative bias expansion is incapable of fully capturing, we approximate the weak lensing power spectrum by a power law $P_\kappa(k) \propto k^n$ when performing MC integration on these scales. The power index n is determined by solving $\gamma = \sqrt{(n+2)/(n+4)}$, where γ is from Eq. (11) computed with the weak lensing power spectrum from Boltzmann solver, and the right side comes from expressing γ in terms of $P_\kappa(k) \propto k^n$.

In Fig. 7, for each angular separation, we perform 60 estimations of the 2PCF to obtain the mean value and the associated estimated standard deviation. For each MC integration, we draw 20 million times 8 random numbers for which evaluation is parallelized on 16 cores. On the local cluster, one such estimation for all angular separations took averagely half an hour (with some variability).

We observe from Fig. 7 that on angular scales $\theta \geq 100'$, the theoretical predictions are almost identical to the MC integration result. This proves that the convergence behavior of our theoretical prediction among different orders of approximation is correct. On angular scales

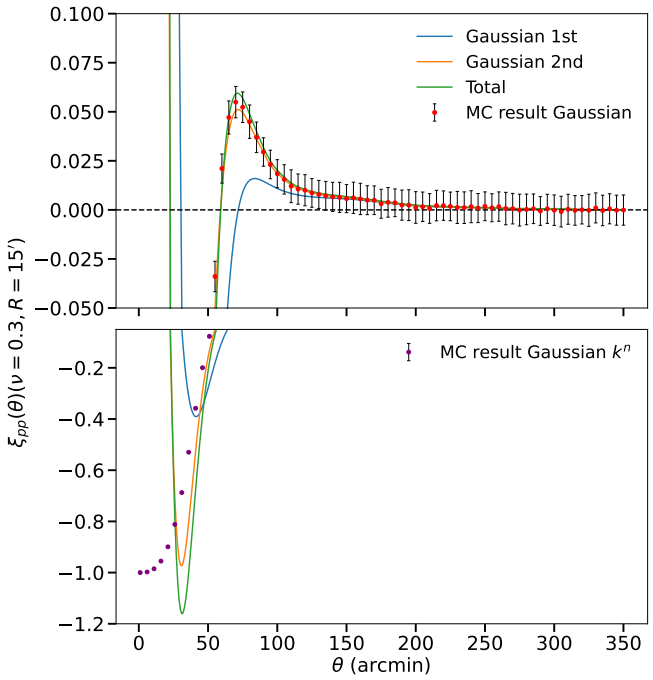

FIG. 7. 2PCF for peaks above the threshold $\nu = 0.3$ with a flat ΛCDM model under Planck 2018 cosmological parameters. The underlying density field is smoothed by a Gaussian kernel with a smoothing scale $R = 15'$. The color curves exploit the same convention to those in Fig. 4. Red dots and the corresponding error bars in the top panel are the mean and standard deviation from 60 estimations of the MC integration. Purple dots and the corresponding error bars in the bottom panel are the mean and standard deviation from 60 estimations of the MC integration but with a power law approximation for the weak lensing convergence power spectrum. The full 2PCF is divided into two panels, each covering a different range of linear scales, allowing for a clearer examination of the small amplitude at larger separations.

$\theta \approx 55'$, the 2nd-order Gaussian approximation already reproduces quite accurately the 2D weak lensing peak exclusion effect [65, 66] as demonstrated by the MC result on the same scale. The non-Gaussian correction from bispectrum does not add significant changes to the predicted exclusion scale if one compares the orange to the green curve on $\theta \approx 55'$. On further smaller scales, there is an increase of the predicted peak 2PCF and the convergence of predictions from different orders of approximation is very poor. This is a well-known feature that the perturbative bias expansion on small scales cannot capture the non-perturbative exclusion zone, as demonstrated by the power law approximation MC result in the bottom panel of Fig. 7.

On angular scales between $50'$ and $100'$, the theoretical prediction from 2nd-order Gaussian approximation is much closer to the MC integration result compared to the LO prediction as expected. Note that in principle one could extend the perturbative expansion to higher orders in the context of Gaussian approximation (e.g, the 3rd-order Gaussian approximation would include the $P(k)^3$ term in NNLO in Eq. (35)) but the convergence of such high-order bias expansion is known to be slow because of the non-perturbative nature of the small-scale exclusion zone (as also shown in Ref. [9] for 3D peaks in Gaussian distributed density field). Adding a bispectrum correction on top of it leads to an excess of 2PCF amplitude which gradually deviates from the exact Gaussian MC result on $\theta \approx 100'$ and reaches its maximum around $70'$, within the context of our chosen smoothing scale and threshold. Around $\theta = 70'$, the amplitude of the 2PCF including the bispectrum (non-Gaussian) correction is about 17% larger than that from the 2nd-order Gaussian approxiamtion. This discrepancy reduces to about 8% on $\theta \approx 100'$. When comparing the theoretical NLO prediction to the exact MC integration result under Gaussian assumption, the discrepancy due to non-Gaussianity at tree-level bispectrum is within the MC sample standard deviation. However, note that with higher-order non-Gaussian corrections or more accurate numerical and survey experiments, we will be able to statistically distinguish them. A further investigation of how well the bispectrum correction characterizes the non-Gaussian part in the 2PCFs of critical points requires a detailed comparison to the N-body simulations. This is beyond the scope of this paper and we leave it to future works.

VII. CONCLUSION

In this paper, we extended the general formalism presented in Ref. [43] for peak statistics in mildly non-Gaussian density field to 2PCFs of 2D critical points, including peaks, voids and saddle points. We applied this formalism to the case of mildly non-Gaussian weak lensing convergence field. Analytically we derived the perturbative bias expansion up to the NNLO, taking the linear terms of trispectrum induced by nonlinear evolution of gravitational instability into account, in Eq. (35). For the numerical calculation, we only consider the lowest-order non-Gaussian correction as in Eq. (59), which is composed of linear terms of the bispectrum. In order to evaluate correlation functions for different types of critical points in 2D weak lensing fields, one needs to compute the g_{ijklm} terms in Eq. (46) and adjust the integration limits within there to a specific critical point type accordingly (Eqs. (47), (49) and (52)), where in Appendix. C we show plots of them as functions of the density field threshold ν.

As a demonstration, we calculated six power spectra and their corresponding 2PCFs from all possible combinations of the three types of critical points in 2D weak lensing field, above a given threshold chosen here to be $\nu = 0.3$ and with a specific Gaussian smoothing scale $R = 15'$. We observed similar properties (shown by Figs. 3 and 4) as for the clustering of 3D critical points measured from N-body simulations in previous works

such as Ref. [21]. The angular separation where the correlation function reaches its maximum beyond the exclusion zone is largest for peaks, followed by that of voids and then saddle points. Saddle points are implied to have a weak exclusion effect with respect to peaks and voids, reflecting a shell-like structure in the distribution of saddle points around a given extremum. Meanwhile we also notice that our theoretical prediction is incapable of fully capturing the nonperturbative exclusion zone on small angular scales, as expected due to its very nonlinear nature. Another interesting feature is that the effect of BAOs is enhanced in 2D peak 2PCF (Fig. 5), compared to the underlying weak lensing convergence power spectrum where the effect is suppressed due to the line-of-sight projection. We believe this enhancement is due to the derivative constraint (curvature) on the weak lensing convergence peak field as discussed in Sec. V and can be used as a BAO probe for weak lensing data.

In order to validate the theoretical prediction, we chose the peak 2PCF as an example and compared it to the exact and yet computationally intensive MC integration result, which assumes a Gaussian distributed underlying density field. The two are almost identical on large angular separations down to approximately $100'$ (Fig. 7). Interestingly, part of the exclusion zone on the outer edge can be described by the perturbative bias expansion. The non-Gaussian correction from the bispectrum contribution modifies the correlation function most significantly around the maximum region.

Eventually, the main purpose of this paper has been to provide the analytical framework for the clustering of critical points in 2D weak lensing field, which is definitely non-Gaussian on scales of interests. There are certain directions to further extend and apply this work. One is to serve as a benchmark test for N-body simulations that are used to measure weak lensing peaks or minima clustering, to prevent those statistics from being biased by simulation systematic effects. One can also combine fast and accurate analytical predictions on large angular scales with detailed measurement from simulations on small scales, which saves significant time and computational resources, and could allow us to achieve a hybrid summary statistic of critical points clustering. This hybrid summary statistic could then be exploited in inference tasks in the ongoing and next-generation weak lensing surveys to extract more information from the large-scale structure of the Universe. Additionally, we have observed features like BAOs and inflection points (observed in Ref. [21] from simulations) on large scales for weak lensing peak 2PCFs. These features could serve as independent standard rulers, independent of galaxy bias, to study the evolution of the large-scale structure and test different cosmological models, an area that is currently under-explored. We hope to address the possibility of the above applications in the near future.

ACKNOWLEDGMENTS

The authors thank Takahiko Matsubara, Stella Seitz, Masahiro Takada and Zhenyuan Wang for fruitful discussions. The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-T-24-05 on "Theory and Data Analysis Challenges for Cosmological Large-Scale Structure Observations" were useful to complete this work. This work has made use of the Infinity cluster hosted by the Institut d'Astrophysique de Paris. We warmly thank S. Rouberol for running it smoothly.

[1] V. Desjacques, D. Jeong, and F. Schmidt, Large-scale galaxy bias, Phys. Rep. **733**, 1 (2018), arXiv:1611.09787 [astro-ph.CO].
[2] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, The Statistics of Peaks of Gaussian Random Fields, Astrophys. J. **304**, 15 (1986).
[3] S. L. Lumsden, A. F. Heavens, and J. A. Peacock, The clustering of peaks in a random Gaussian field, MNRAS **238**, 293 (1989).
[4] J. A. Peacock and A. F. Heavens, Alternatives to the Press-Schechter cosmological mass function, MNRAS **243**, 133 (1990).
[5] E. Regos and A. S. Szalay, Density and velocity correlations of peaks from random Gaussian fluctuations, MNRAS **272**, 447 (1995).
[6] V. Desjacques, Baryon acoustic signature in the clustering of density maxima, Phys. Rev. D **78**, 103503 (2008), arXiv:0806.0007 [astro-ph].
[7] V. Desjacques and R. K. Sheth, Redshift space correlations and scale-dependent stochastic biasing of density peaks, Phys. Rev. D **81**, 023526 (2010), arXiv:0909.4544 [astro-ph.CO].
[8] V. Desjacques, Local bias approach to the clustering of discrete density peaks, Phys. Rev. D **87**, 043505 (2013), arXiv:1211.4128 [astro-ph.CO].
[9] T. Matsubara and S. Codis, Large-separation expansion of peak clustering in Gaussian random fields, Phys. Rev. D **101**, 063504 (2020), arXiv:1910.09561 [astro-ph.CO].
[10] T. Baldauf, S. Codis, V. Desjacques, and C. Pichon, Nonperturbative halo clustering from cosmological density peaks, Phys. Rev. D **103**, 083530 (2021), arXiv:2012.14404 [astro-ph.CO].
[11] E. G. P. Bos, R. van de Weygaert, K. Dolag, and V. Pettorino, The darkness that shaped the void: dark energy and cosmic voids, MNRAS **426**, 440 (2012), arXiv:1205.4238 [astro-ph.CO].
[12] A. Barreira, M. Cautun, B. Li, C. M. Baugh, and S. Pascoli, Weak lensing by voids in modified lensing potentials, JCAP **2015** (8), 028, arXiv:1505.05809 [astro-ph.CO].
[13] E. L. D. Perico, R. Voivodic, M. Lima, and D. F. Mota, Cosmic voids in modified gravity scenarios, arXiv e-prints , arXiv:1905.12450 (2019), arXiv:1905.12450 [astro-ph.CO].
[14] S. Contarini, F. Marulli, L. Moscardini, A. Veropalumbo,

C. Giocoli, and M. Baldi, Cosmic voids in modified gravity models with massive neutrinos, MNRAS **504**, 5021 (2021), arXiv:2009.03309 [astro-ph.CO].

[15] C. M. Correa, Cosmic voids as cosmological laboratories, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina **64**, 159 (2023), arXiv:2210.17459 [astro-ph.CO].

[16] N. Hamaus, A. Pisani, J.-A. Choi, G. Lavaux, B. D. Wandelt, and J. Weller, Precision cosmology with voids in the final BOSS data, JCAP **2020** (12), 023, arXiv:2007.07895 [astro-ph.CO].

[17] S. Contarini, A. Pisani, N. Hamaus, F. Marulli, L. Moscardini, and M. Baldi, Cosmological Constraints from the BOSS DR12 Void Size Function, Astrophys. J. **953**, 46 (2023), arXiv:2212.03873 [astro-ph.CO].

[18] T. Sousbie, The persistent cosmic web and its filamentary structure - I. Theory and implementation, MNRAS **414**, 350 (2011), arXiv:1009.4015 [astro-ph.CO].

[19] M. Cautun, R. van de Weygaert, B. J. T. Jones, and C. S. Frenk, Evolution of the cosmic web, MNRAS **441**, 2923 (2014), arXiv:1401.7866 [astro-ph.CO].

[20] F. Sarron, C. Adami, F. Durret, and C. Laigle, Preprocessing of galaxies in cosmic filaments around AMASCFI clusters in the CFHTLS, A&A **632**, A49 (2019), arXiv:1903.02879 [astro-ph.GA].

[21] J. Shim, S. Codis, C. Pichon, D. Pogosyan, and C. Cadiou, The clustering of critical points in the evolving cosmic web, MNRAS **502**, 3885 (2021), arXiv:2011.04321 [astro-ph.CO].

[22] J. Feldbrugge and R. van de Weygaert, What makes a cosmic filament? The dynamical origin and identity of filaments I. fundamentals in 2D, arXiv e-prints , arXiv:2405.20475 (2024), arXiv:2405.20475 [astro-ph.CO].

[23] D. Galárraga-Espinosa, C. Cadiou, C. Gouin, S. D. M. White, V. Springel, R. Pakmor, B. Hadzhiyska, S. Bose, F. Ferlito, L. Hernquist, R. Kannan, M. Barrera, A. Maria Delgado, and C. Hernández-Aguayo, Evolution of cosmic filaments in the MTNG simulation, A&A **684**, A63 (2024), arXiv:2309.08659 [astro-ph.CO].

[24] L. Marian, R. E. Smith, S. Hilbert, and P. Schneider, The cosmological information of shear peaks: beyond the abundance, MNRAS **432**, 1338 (2013), arXiv:1301.5001 [astro-ph.CO].

[25] H. Y. Shan, J.-P. Kneib, J. Comparat, E. Jullo, A. Charbonnier, T. Erben, M. Makler, B. Moraes, L. Van Waerbeke, F. Courbin, G. Meylan, C. Tao, and J. E. Taylor, Weak lensing mass map and peak statistics in Canada-France-Hawaii Telescope Stripe 82 survey, MNRAS **442**, 2534 (2014), arXiv:1311.1319 [astro-ph.CO].

[26] C. T. Davies, M. Cautun, and B. Li, The self-similarity of weak lensing peaks, MNRAS **488**, 5833 (2019), arXiv:1905.01710 [astro-ph.CO].

[27] C. T. Davies, M. Cautun, B. Giblin, B. Li, J. Harnois-Déraps, and Y.-C. Cai, Constraining cosmology with weak lensing voids, MNRAS **507**, 2267 (2021), arXiv:2010.11954 [astro-ph.CO].

[28] J. Harnois-Déraps, N. Martinet, T. Castro, K. Dolag, B. Giblin, C. Heymans, H. Hildebrandt, and Q. Xia, Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1, MNRAS **506**, 1623 (2021), arXiv:2012.02777 [astro-ph.CO].

[29] C. T. Davies, M. Cautun, B. Giblin, B. Li, J. Harnois-Déraps, and Y.-C. Cai, Cosmological forecasts with the clustering of weak lensing peaks, MNRAS **513**, 4729 (2022), arXiv:2110.10164 [astro-ph.CO].

[30] D. Zürcher, J. Fluri, R. Sgier, T. Kacprzak, M. Gatti, C. Doux, L. Whiteway, A. Réfrégier, C. Chang, N. Jeffrey, B. Jain, P. Lemos, D. Bacon, A. Alarcon, A. Amon, K. Bechtol, M. Becker, G. Bernstein, A. Campos, R. Chen, A. Choi, C. Davis, J. Derose, S. Dodelson, F. Elsner, J. Elvin-Poole, S. Everett, A. Ferte, D. Gruen, I. Harrison, D. Huterer, M. Jarvis, P. F. Leget, N. Maccrann, J. Mccullough, J. Muir, J. Myles, A. Navarro Alsina, S. Pandey, J. Prat, M. Raveri, R. P. Rollins, A. Roodman, C. Sanchez, L. F. Secco, E. Sheldon, T. Shin, M. Troxel, I. Tutusaus, B. Yin, M. Aguena, S. Allam, F. Andrade-Oliveira, J. Annis, E. Bertin, D. Brooks, D. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. Castander, R. Cawthon, C. Conselice, M. Costanzi, L. da Costa, M. E. da Silva Pereira, T. Davis, J. De Vicente, S. Desai, H. T. Diehl, J. Dietrich, P. Doel, K. Eckert, A. Evrard, I. Ferrero, B. Flaugher, P. Fosalba, D. Friedel, J. Frieman, J. Garcia-Bellido, E. Gaztanaga, D. Gerdes, T. Giannantonio, R. Gruendl, J. Gschwend, G. Gutierrez, S. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. James, K. Kuehn, N. Kuropatkin, O. Lahav, C. Lidman, M. Lima, M. Maia, J. Marshall, P. Melchior, F. Menanteau, R. Miquel, R. Morgan, A. Palmese, F. Paz-Chinchon, A. Pieres, A. Plazas Malagón, K. Reil, M. Rodriguez Monroy, K. Romer, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla, M. Smith, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, J. Weller, R. Wilkinson, and DES Collaboration, Dark energy survey year 3 results: Cosmology with peaks using an emulator approach, MNRAS **511**, 2075 (2022), arXiv:2110.10135 [astro-ph.CO].

[31] J. Harnois-Déraps, S. Heydenreich, B. Giblin, N. Martinet, T. Tröster, M. Asgari, P. Burger, T. Castro, K. Dolag, C. Heymans, H. Hildebrandt, B. Joachimi, and A. H. Wright, KiDS-1000 and DES-Y1 combined: cosmology from peak count statistics, MNRAS **534**, 3305 (2024), arXiv:2405.10312 [astro-ph.CO].

[32] G. A. Marques, J. Liu, M. Shirasaki, L. Thiele, D. Grandón, K. M. Huffenberger, S. Cheng, J. Harnois-Déraps, K. Osato, and W. R. Coulton, Cosmology from weak lensing peaks and minima with Subaru Hyper Suprime-Cam Survey first-year data, MNRAS **528**, 4513 (2024), arXiv:2308.10866 [astro-ph.CO].

[33] C. T. Davies, J. Harnois-Déraps, B. Li, B. Giblin, C. Hernández-Aguayo, and E. Paillas, Constraining modified gravity with weak-lensing peaks, MNRAS **533**, 3546 (2024), arXiv:2406.11958 [astro-ph.CO].

[34] A. Halder, O. Friedrich, S. Seitz, and T. N. Varga, The integrated three-point correlation function of cosmic shear, MNRAS **506**, 2780 (2021), arXiv:2102.10177 [astro-ph.CO].

[35] M. Gatti, B. Jain, C. Chang, M. Raveri, D. Zürcher, L. Secco, L. Whiteway, N. Jeffrey, C. Doux, T. Kacprzak, D. Bacon, P. Fosalba, A. Alarcon, A. Amon, K. Bechtol, M. Becker, G. Bernstein, J. Blazek, A. Campos, A. Choi, C. Davis, J. Derose, S. Dodelson, F. Elsner, J. Elvin-Poole, S. Everett, A. Ferte, D. Gruen, I. Harrison, D. Huterer, M. Jarvis, E. Krause, P. F. Leget, P. Lemos, N. Maccrann, J. Mccullough, J. Muir, J. Myles, A. Navarro, S. Pandey, J. Prat, R. P. Rollins, A. Rood-

man, C. Sanchez, E. Sheldon, T. Shin, M. Troxel, I. Tutusaus, B. Yin, M. Aguena, S. Allam, F. Andrade-Oliveira, J. Annis, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, R. Cawthon, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, K. Eckert, A. E. Evrard, I. Ferrero, J. García-Bellido, E. Gaztanaga, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, C. Lidman, M. A. G. Maia, J. L. Marshall, P. Melchior, F. Menanteau, R. Miquel, R. Morgan, A. Palmese, F. Paz-Chinchón, A. Pieres, A. A. Plazas Malagón, K. Reil, M. Rodriguez-Monroyv, A. K. Romer, E. Sanchez, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, and DES Collaboration, Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps, Phys. Rev. D **106**, 083509 (2022), arXiv:2110.10141 [astro-ph.CO].

[36] A. Halder and A. Barreira, Response approach to the integrated shear 3-point correlation function: the impact of baryonic effects on small scales, MNRAS **515**, 4639 (2022), arXiv:2201.05607 [astro-ph.CO].

[37] Z. Gong, A. Halder, A. Barreira, S. Seitz, and O. Friedrich, Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators, JCAP **2023** (7), 040, arXiv:2304.01187 [astro-ph.CO].

[38] A. Halder, Z. Gong, A. Barreira, O. Friedrich, S. Seitz, and D. Gruen, Beyond 3×2-point cosmology: the integrated shear and galaxy 3-point correlation functions, JCAP **2023** (10), 028, arXiv:2305.17132 [astro-ph.CO].

[39] D. Anbajagane, C. Chang, A. Banerjee, T. Abel, M. Gatti, V. Ajani, A. Alarcon, A. Amon, E. J. Baxter, K. Bechtol, M. R. Becker, G. M. Bernstein, A. Campos, A. Carnero Rosell, M. Carrasco Kind, R. Chen, A. Choi, C. Davis, J. DeRose, H. T. Diehl, S. Dodelson, C. Doux, A. Drlica-Wagner, K. Eckert, J. Elvin-Poole, S. Everett, A. Ferté, D. Gruen, R. A. Gruendl, I. Harrison, W. G. Hartley, E. M. Huff, B. Jain, M. Jarvis, N. Jeffrey, T. Kacprzak, N. Kokron, N. Kuropatkin, P. F. Leget, N. MacCrann, J. McCullough, J. Myles, A. Navarro-Alsina, S. Pandey, J. Prat, M. Raveri, R. P. Rollins, A. Roodman, E. S. Rykoff, C. Sánchez, L. F. Secco, I. Sevilla-Noarbe, E. Sheldon, T. Shin, M. A. Troxel, I. Tutusaus, L. Whiteway, B. Yanny, B. Yin, Y. Zhang, T. M. C. Abbott, S. Allam, M. Aguena, O. Alves, F. Andrade-Oliveira, J. Annis, D. Bacon, J. Blazek, D. Brooks, R. Cawthon, L. N. da Costa, M. E. S. Pereira, T. M. Davis, S. Desai, P. Doel, I. Ferrero, J. Frieman, G. Giannini, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, O. Lahav, J. L. Marshall, J. Mena-Fernández, F. Menanteau, R. Miquel, A. Palmese, A. Pieres, A. A. Plazas Malagón, K. Reil, E. Sanchez, M. Smith, M. E. C. Swanson, G. Tarle, P. Wiseman, and DES Collaboration, Beyond the 3rd moment: a practical study of using lensing convergence CDFs for cosmology with DES Y3, MNRAS **526**, 5530 (2023), arXiv:2308.03863 [astro-ph.CO].

[40] P. Cannon, D. Ward, and S. M. Schmon, Investigating the Impact of Model Misspecification in Neural Simulation-based Inference, arXiv e-prints , arXiv:2209.01845 (2022), arXiv:2209.01845 [stat.ML].

[41] B. Horowitz and P. Melchior, Plausible Adversarial Attacks on Direct Parameter Inference Models in Astrophysics, arXiv e-prints , arXiv:2211.14788 (2022), arXiv:2211.14788 [astro-ph.CO].

[42] C. Gay, C. Pichon, and D. Pogosyan, Non-Gaussian statistics of critical sets in 2D and 3D: Peaks, voids, saddles, genus, and skeleton, Phys. Rev. D **85**, 023011 (2012), arXiv:1110.0261 [astro-ph.CO].

[43] T. Matsubara, Statistics of peaks of weakly non-Gaussian random fields: Effects of bispectrum in two- and three-dimensions, Phys. Rev. D **101**, 043532 (2020), arXiv:2001.05702 [astro-ph.CO].

[44] N. Kaiser, On the spatial correlations of Abell clusters., The Astrophysical Journal Letters **284**, L9 (1984).

[45] M. Kac, On the average number of real roots of a random algebraic equation, Bulletin of the American Mathematical Society **49**, 314 (1943).

[46] S. O. Rice, Mathematical Analysis of Random Noise-Conclusion, Bell System Technical Journal **24**, 46 (1945).

[47] M. S. Longuet-Higgins, The Statistical Analysis of a Random, Moving Surface, Philosophical Transactions of the Royal Society of London Series A **249**, 321 (1957).

[48] D. Pogosyan, C. Gay, and C. Pichon, Invariant joint distribution of a stationary random field and its derivatives: Euler characteristic and critical point counts in 2 and 3D, Phys. Rev. D **80**, 081301 (2009), arXiv:0907.1437 [astro-ph.CO].

[49] P. J. E. Peebles, *The large-scale structure of the universe* (Princeton University Press, 1980).

[50] S. Blinnikov and R. Moessner, Expansions for nearly Gaussian distributions, A&A **130**, 193 (1998), arXiv:astro-ph/9711239 [astro-ph].

[51] T. Matsubara, Diagrammatic Methods in Statistics and Biasing in the Large-Scale Structure of the Universe, ApJS **101**, 1 (1995), arXiv:astro-ph/9501056 [astro-ph].

[52] T. Matsubara, Nonlinear perturbation theory integrated with nonlocal bias, redshift-space distortions, and primordial non-Gaussianity, Phys. Rev. D **83**, 083518 (2011), arXiv:1102.4619 [astro-ph.CO].

[53] A. Eggemeier, R. Scoccimarro, and R. E. Smith, Bias loop corrections to the galaxy bispectrum, Phys. Rev. D **99**, 123514 (2019), arXiv:1812.03208 [astro-ph.CO].

[54] M. Bartelmann and P. Schneider, Weak gravitational lensing, Phys. Rep. **340**, 291 (2001), arXiv:astro-ph/9912508 [astro-ph].

[55] P. Schneider, C. S. Kochanek, and J. Wambganss, *Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro* (Springer Science & Business Media, 2006) arXiv:astro-ph/0407232 [astro-ph].

[56] M. Kilbinger, Cosmology with cosmic shear observations: a review, Reports on Progress in Physics **78**, 086901 (2015), arXiv:1411.0115 [astro-ph.CO].

[57] D. N. Limber, The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field., Astrophys. J. **117**, 134 (1953).

[58] N. Kaiser, Weak Lensing and Cosmology, Astrophys. J. **498**, 26 (1998), arXiv:astro-ph/9610120 [astro-ph].

[59] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, Nonlinear Evolution of the Bispectrum of Cosmological Perturbations, Astrophys. J. **496**, 586 (1998), arXiv:astro-ph/9704075 [astro-ph].

[60] 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, MNRAS **341**, 1311 (2003), arXiv:astro-ph/0207664 [astro-ph].
[61] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Revising the Halofit Model for the Nonlinear Matter Power Spectrum, Astrophys. J. **761**, 152 (2012), arXiv:1208.2701 [astro-ph.CO].
[62] D. Blas, J. Lesgourgues, and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes, JCAP **2011** (7), 034, arXiv:1104.2933 [astro-ph.CO].
[63] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca, Planck 2018 results. VI. Cosmological parameters, A&A **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].
[64] T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus, and V. Desjacques, Halo stochasticity from exclusion and nonlinear clustering, Phys. Rev. D **88**, 083507 (2013), arXiv:1305.2917 [astro-ph.CO].
[65] T. Baldauf, S. Codis, V. Desjacques, and C. Pichon, Peak exclusion, stochasticity and convergence of perturbative bias expansions in 1+1 gravity, MNRAS **456**, 3985 (2016), arXiv:1510.09204 [astro-ph.CO].
[66] S. Codis, D. Pogosyan, and C. Pichon, On the connectivity of the cosmic web: theory and implications for cosmology and galaxy formation, MNRAS **479**, 973 (2018), arXiv:1803.11477 [astro-ph.CO].
[67] V. Desjacques, M. Crocce, R. Scoccimarro, and R. K. Sheth, Modeling scale-dependent bias on the baryonic acoustic scale with the statistics of peaks of Gaussian random fields, Phys. Rev. D **82**, 103529 (2010), arXiv:1009.3449 [astro-ph.CO].
[68] S. Ma, *Statistical Mechanics* (World Scientific, 1985).

Appendix A: Generalization of the Gram-Charlier A series expansion with the Wiener-Hermite functionals

In this paper, we adopt the following Fourier transform convention

$$\tilde{f}(\mathbf{k}) = \int \mathrm{d}^2x e^{-i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) \ , f(\mathbf{x}) = \int \frac{\mathrm{d}^2k}{(2\pi)^2} e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{f}(\mathbf{k}) \ . \tag{A1}$$

To derive the probability density functional $\mathcal{P}(\tilde{f})$, we start with the partition function, or the moment generating functional of the density field

$$\mathcal{Z}(\mathbf{J}) = \int \mathcal{D}\tilde{f} \exp\left[i \int \frac{\mathrm{d}^2k}{(2\pi)^2} \mathbf{J}(\mathbf{k}) \tilde{f}(\mathbf{k}) \right] \mathcal{P}(\tilde{f}) \ , \tag{A2}$$

where $\mathcal{D}\tilde{f}$ is the same to that mentioned in Eq. (18). On the other hand, the partition function can also be expressed in terms of a series expansion using the cumulant expansion theorem [68]

$$\begin{aligned} \ln \mathcal{Z}(\mathbf{J}) = & \sum_{n=1}^{\infty} \frac{i^n}{n!} \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \cdots \int \frac{\mathrm{d}^2k_n}{(2\pi)^2} \\ & \times \langle \tilde{f}(\mathbf{k}_1) \cdots \tilde{f}(\mathbf{k}_n) \rangle_{\mathrm{c}} \mathbf{J}(\mathbf{k}_1) \cdots \mathbf{J}(\mathbf{k}_n) \ , \end{aligned} \tag{A3}$$

and if we take the exponential on both sides of the above equation, we would have

$$\begin{aligned} \mathcal{Z}(\mathbf{J}) = & \exp\left[-\frac{1}{2} \int \frac{\mathrm{d}^2k}{(2\pi)^2} P(k) \mathbf{J}(\mathbf{k}) \mathbf{J}(-\mathbf{k}) \right] \\ & \times \exp\left[\sum_{n=3}^{\infty} \frac{i^n}{n!} \int \frac{\mathrm{d}^2k_1}{(2\pi)^2} \cdots \int \frac{\mathrm{d}^2k_n}{(2\pi)^2} \right. \\ & \left. \times \langle \tilde{f}(\mathbf{k}_1) \cdots \tilde{f}(\mathbf{k}_n) \rangle_{\mathrm{c}} \mathbf{J}(\mathbf{k}_1) \cdots \mathbf{J}(\mathbf{k}_n) \right] \ . \end{aligned} \tag{A4}$$

where we used the definition of the density field power spectrum similar to Eqs. (22) and (23)

$$\langle \tilde{f}(\mathbf{k}_1) \tilde{f}(\mathbf{k}_2) \rangle_{\mathrm{c}} = (2\pi)^2 \delta_{\mathrm{D}}(\mathbf{k}_1 + \mathbf{k}_2) P(k) \ . \tag{A5}$$

We invert the transformation in Eq. (A2) and replace $\mathcal{Z}(\mathbf{J})$ with the expression in Eq. (A4)

$$\begin{aligned}\mathcal{P}(\tilde{f})=&\int\hat{\mathcal{D}}\mathbf{J}\mathcal{Z}(\mathbf{J})\exp\left[-i\int\frac{\mathrm{d}^2k}{(2\pi)^2}\mathbf{J}(\mathbf{k})\tilde{f}(\mathbf{k})\right]\\&=\int\hat{\mathcal{D}}\mathbf{J}\exp\left[\sum_{n=3}^{\infty}\frac{i^n}{n!}\int\frac{\mathrm{d}^2k_1}{(2\pi)^2}\cdots\int\frac{\mathrm{d}^2k_n}{(2\pi)^2}\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_n)\rangle_{\mathrm{c}}\mathbf{J}(\mathbf{k}_1)\cdots\mathbf{J}(\mathbf{k}_n)\right]\exp\left[-\frac{1}{2}\int\frac{\mathrm{d}^2k}{(2\pi)^2}P(k)\mathbf{J}(\mathbf{k})\mathbf{J}(-\mathbf{k})\right]\\&\times\exp\left[-i\int\frac{\mathrm{d}^2k}{(2\pi)^2}\mathbf{J}(\mathbf{k})\tilde{f}(\mathbf{k})\right]\\&=\exp\left[\sum_{n=3}^{\infty}\frac{(-1)^n}{n!}\int\mathrm{d}^2k_1\cdots\int\mathrm{d}^2k_n\langle\tilde{f}(\mathbf{k}_1)\cdots\tilde{f}(\mathbf{k}_n)\rangle_{\mathrm{c}}\frac{\delta^n}{\delta\tilde{f}(\mathbf{k}_1)\cdots\delta\tilde{f}(\mathbf{k}_n)}\right]\mathcal{P}_{\mathrm{G}}(\tilde{f})\ ,\end{aligned}\tag{A6}$$

where $\mathcal{P}_{\mathrm{G}}(\tilde{f})$ is the Gaussian probability density functional

$$\begin{aligned}&\mathcal{P}_{\mathrm{G}}(\tilde{f})\\&\quad=\int\hat{\mathcal{D}}\mathbf{J}\exp\left[-\frac{1}{2}\int\frac{\mathrm{d}^2k}{(2\pi)^2}P(k)\mathbf{J}(\mathbf{k})\mathbf{J}(-\mathbf{k})\right.\\&\quad\left.-i\int\frac{\mathrm{d}^2k}{(2\pi)^2}\mathbf{J}(\mathbf{k})\tilde{f}(\mathbf{k})\right]\\&\quad\propto\exp\left[-\frac{1}{2}\int\frac{\mathrm{d}^2k}{(2\pi)^2}\frac{\tilde{f}(\mathbf{k})\tilde{f}(-\mathbf{k})}{P(k)}\right]\ ,\end{aligned}\tag{A7}$$

up to a normalization constant and in the last line of Eq. (A6), we used the following identity

$$\begin{aligned}&\int\hat{\mathcal{D}}\mathbf{J}i^n\mathbf{J}(\mathbf{k}_1)\cdots\mathbf{J}(\mathbf{k}_n)\mathcal{P}_{\mathrm{G}}(\mathbf{J})\\&\quad\times\exp\left[-i\int\frac{\mathrm{d}^2k}{(2\pi)^2}\mathbf{J}(\mathbf{k})\tilde{f}(\mathbf{k})\right]\\&\quad=(-1)^n\times(2\pi)^{2n}\frac{\delta^n}{\delta\tilde{f}(\mathbf{k}_1)\cdots\delta\tilde{f}(\mathbf{k}_n)}\mathcal{P}_{\mathrm{G}}(\tilde{f})\ .\end{aligned}\tag{A8}$$

With Eq. (A6), we can apply the power series expansion to the exponential term. This will give the functional derivatives of the Gaussian probability density functional which can be expressed using Eq. (19) as $\mathcal{H}_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\mathcal{P}_{\mathrm{G}}(\tilde{f})/(-1)^n$. Substituting this into the power series expansion, we can obtain the generalized Gram-Charlier A series expansion in Eq. (20).

Appendix B: Diagrammatic method to evaluate $\langle\mathcal{H}_n^\star\mathcal{H}_m^\star\mathcal{H}_l\rangle_{\mathrm{G}}$ factors

Based on Appendix. A of Ref. [51], here we present the diagrammatic rules in Fourier space for the products of generalized Wiener-Hermite functionals $\langle\mathcal{H}_n^\star\mathcal{H}_m^\star\mathcal{H}_l\rangle_{\mathrm{G}}$.

1. Corresponding to each $\mathcal{H}_{n_i}^\star$ or $\mathcal{H}_{n_i}$, draw n_i points labeled by $\mathbf{k}_i^{(1)},\ldots,\mathbf{k}_i^{(n_i)}$, representing each mode.
2. Create $\sum_i n_i/2$ pairs out of all the points such that two points in the same $\mathcal{H}_{n_i}^\star$ or $\mathcal{H}_{n_i}$ are not paired. If $\sum_i n_i/2$ is an odd number, then $\langle\mathcal{H}_n^\star\mathcal{H}_m^\star\mathcal{H}_l\rangle_{\mathrm{G}}=0$.
3. Associate factors $(2\pi)^2\delta_{\mathrm{D}}(\mathbf{k}_i^{(p)}+\mathbf{k}_j^{(q)})P(k_i^{(p)})$ ($p\in\{1,\ldots,n_i\},q\in\{1,\ldots,n_j\}$) for each pair if the two points are from $\mathcal{H}_{n_i}^\star$ and $\mathcal{H}_{n_j}^\star$ respectively. If the two points are from $\mathcal{H}_{n_i}^\star$ and $\mathcal{H}_{n_j}$ separately, the associated factors change to $\delta_{\mathrm{D}}(\mathbf{k}_i^{(p)}-\mathbf{k}_j^{(q)})$ ($p\in\{1,\ldots,n_i\},q\in\{1,\ldots,n_j\}$) for each pair. We then make products of these factors.
4. Sum up these products from all possible pair configurations.

With the above diagrammatic rules, it is very convenient to compute any $\langle\mathcal{H}_n^\star\mathcal{H}_m^\star\mathcal{H}_l\rangle_{\mathrm{G}}$ factors with $n+m+l$ equal to an even number. Below we show all non-zero results needed in deriving Eq. (30).

$$\langle \mathcal{H}_0^\star \mathcal{H}_0^\star \mathcal{H}_0 \rangle_{\rm G} = 1 \ , \tag{B1}$$

$$\langle \mathcal{H}_1^\star(\mathbf{k}_1) \mathcal{H}_1^\star(\mathbf{k}_1') \mathcal{H}_0 \rangle_{\rm G} = (2\pi)^2 P(k_1) \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_1') \ , \tag{B2}$$

$$\langle \mathcal{H}_2^\star(\mathbf{k}_1, \mathbf{k}_2) \mathcal{H}_2^\star(\mathbf{k}_1', \mathbf{k}_2') \mathcal{H}_0 \rangle_{\rm G} = (2\pi)^4 P(k_1) P(k_2) \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2 + \mathbf{k}_2') + {\rm sym} \ , \tag{B3}$$

$$\langle \mathcal{H}_0^\star \mathcal{H}_3^\star(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \mathcal{H}_3(\mathbf{k}_1', \mathbf{k}_2', \mathbf{k}_3') \rangle_{\rm G} = \delta_{\rm D}(\mathbf{k}_1 - \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2 - \mathbf{k}_2') \delta_{\rm D}(\mathbf{k}_3 - \mathbf{k}_3') + {\rm sym} \ , \tag{B4}$$

$$\langle \mathcal{H}_1^\star(\mathbf{k}_1) \mathcal{H}_2^\star(\mathbf{k}_1', \mathbf{k}_2') \mathcal{H}_3(\mathbf{k}_1'', \mathbf{k}_2'', \mathbf{k}_3'') \rangle_{\rm G} = \delta_{\rm D}(\mathbf{k}_1 - \mathbf{k}_1'') \delta_{\rm D}(\mathbf{k}_2' - \mathbf{k}_2'') \delta_{\rm D}(\mathbf{k}_3' - \mathbf{k}_3'') + {\rm sym} \ , \tag{B5}$$

$$\langle \mathcal{H}_3^\star(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \mathcal{H}_3^\star(\mathbf{k}_1', \mathbf{k}_2', \mathbf{k}_3') \mathcal{H}_0 \rangle_{\rm G} = (2\pi)^6 P(k_1) P(k_2) P(k_3) \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2 + \mathbf{k}_2') \delta_{\rm D}(\mathbf{k}_3 + \mathbf{k}_3') + {\rm sym} \ , \tag{B6}$$

$$\begin{aligned} &\langle \mathcal{H}_1^\star(\mathbf{k}_1) \mathcal{H}_4^\star(\mathbf{k}_1', \ldots, \mathbf{k}_4') \mathcal{H}_3(\mathbf{k}_1^{''}, \mathbf{k}_2^{''}, \mathbf{k}_3^{''}) \rangle_{\rm G} \\ &= (2\pi)^2 \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2' - \mathbf{k}_1^{''}) \delta_{\rm D}(\mathbf{k}_3' - \mathbf{k}_2^{''}) \delta_{\rm D}(\mathbf{k}_4' - \mathbf{k}_3^{''}) P(k_1) + {\rm sym} \ , \end{aligned} \tag{B7}$$

$$\begin{aligned} &\langle \mathcal{H}_2^\star(\mathbf{k}_1, \mathbf{k}_2) \mathcal{H}_3^\star(\mathbf{k}_1', \mathbf{k}_2', \mathbf{k}_3') \mathcal{H}_3(\mathbf{k}_1^{''}, \mathbf{k}_2^{''}, \mathbf{k}_3^{''}) \rangle_{\rm G} \\ &= (2\pi)^2 \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2 - \mathbf{k}_1^{''}) \delta_{\rm D}(\mathbf{k}_2' - \mathbf{k}_2^{''}) \delta_{\rm D}(\mathbf{k}_3' - \mathbf{k}_3^{''}) P(k_1) + {\rm sym} \ , \end{aligned} \tag{B8}$$

$$\langle \mathcal{H}_0^\star \mathcal{H}_4^\star(\mathbf{k}_1, \ldots, \mathbf{k}_4) \mathcal{H}_4(\mathbf{k}_1', \ldots, \mathbf{k}_4') \rangle_{\rm G} = \delta_{\rm D}(\mathbf{k}_1 - \mathbf{k}_1') \delta_{\rm D}(\mathbf{k}_2 - \mathbf{k}_2') \delta_{\rm D}(\mathbf{k}_3 - \mathbf{k}_3') \delta_{\rm D}(\mathbf{k}_4 - \mathbf{k}_3') + {\rm sym} \ , \tag{B9}$$

$$\langle \mathcal{H}_1^\star(\mathbf{k}_1) \mathcal{H}_3^\star(\mathbf{k}_1', \mathbf{k}_2', \mathbf{k}_3') \mathcal{H}_4(\mathbf{k}_1^{''}, \ldots, \mathbf{k}_4^{''}) \rangle_{\rm G} = \delta_{\rm D}(\mathbf{k}_1 - \mathbf{k}_1^{''}) \delta_{\rm D}(\mathbf{k}_1' - \mathbf{k}_2^{''}) \delta_{\rm D}(\mathbf{k}_2' - \mathbf{k}_3^{''}) \delta_{\rm D}(\mathbf{k}_3' - \mathbf{k}_4^{''}) + {\rm sym} \ , \tag{B10}$$

$$\langle \mathcal{H}_2^\star(\mathbf{k}_1, \mathbf{k}_2) \mathcal{H}_2^\star(\mathbf{k}_1', \mathbf{k}_2') \mathcal{H}_4(\mathbf{k}_1^{''}, \ldots, \mathbf{k}_4^{''}) \rangle_{\rm G} = \delta_{\rm D}(\mathbf{k}_1 - \mathbf{k}_1^{''}) \delta_{\rm D}(\mathbf{k}_2 - \mathbf{k}_2^{''}) \delta_{\rm D}(\mathbf{k}_1' - \mathbf{k}_3^{''}) \delta_{\rm D}(\mathbf{k}_2' - \mathbf{k}_4^{''}) + {\rm sym} \ , \tag{B11}$$

where all "sym" expressions in the above equations denote all following additional terms that have the same form as the previous one but are composed of other pair configurations in the diagrammatic scheme.

Appendix C: Plots of g_{ijklm} factors and g_1 function

Here we show the seven $g_{ijklm}(\nu)$ functions in Eqs. (38) and (39) in Fig. 8 where we adopt the same cosmological parameters as those in Sec. V and a smoothing scale of $15'$.

By combing Fig. 8 and Eqs. (35), (38) and (39), we can observe that g_{10000} and g_{20000} are the two dominant factors that determine the amplitude of the power spectrum for different critical points. In both subplots, the corresponding factor for voids is greater than that for saddle points, which in turn is greater than that for peaks within the range $0 < \nu < 6$. This numerically explains the amplitude relation among the power spectrum of peaks, voids and saddle points we computed in Sec. V.

As a schematic illustration, we show in the left panel of Fig. 9 the g_1 functions of all three types of critical points above the threshold $\nu = 0.3$. And we do observe that on small to intermediate k scales, voids have a larger function value than that of saddle points which in turn larger than that of peaks. Furthermore, in the right panel of Fig. 9, we show how g_1 function of peaks, which has a quadratic form, varies along with different threshold ν. For higher threshold, larger the value of leading order Gaussian response function g_1 would become, which leads to a higher clustering amplitude in the 2PCF. This holds true for voids and saddle points as well.

Appendix D: Angular integration of the peak power spectrum

We demonstrate here how to simplify the type of integrals that appear in Eq. (59) to obtain the extrema power spectra of 2D fields. This appendix borrows from the derivation presented in Ref. [43] and is only shown here for completeness.

We are interested in constrained integrals of the form

$$A = \int_{\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}} \left(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right)^l X(k_1) Y(k_2) \ , \tag{D1}$$

where $\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2$ is the cosine of the angle θ between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. In 2D, this constraint can be explicitly written as

$$\begin{aligned} A = &\int {\rm d}^2 r e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \int \frac{{\rm d}^2\boldsymbol{k}_1}{(2\pi)^2} \frac{{\rm d}^2\boldsymbol{k}_2}{(2\pi)^2} e^{i(\boldsymbol{k}_1 + \boldsymbol{k}_2)\cdot r} \\ &\times \left(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right)^l X(k_1) Y(k_2) \ . \end{aligned} \tag{D2}$$

Rotational invariance of the system makes the result of the $\boldsymbol{k}_i$ integrals only dependent on the amplitude r of $\boldsymbol{r}$ so that we can directly perform the angular integration replacing the exponentials by their angular averages given by Bessel functions of the first kind:

$$\begin{aligned} A = &2\pi \int {\rm d}r J_0(kr) \int \frac{{\rm d}^2\boldsymbol{k}_1}{(2\pi)^2} \frac{{\rm d}^2\boldsymbol{k}_2}{(2\pi)^2} J_0(|\boldsymbol{k}_1 + \boldsymbol{k}_2| r) \\ &\times \left(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right)^l X(k_1) Y(k_2) \ . \end{aligned} \tag{D3}$$

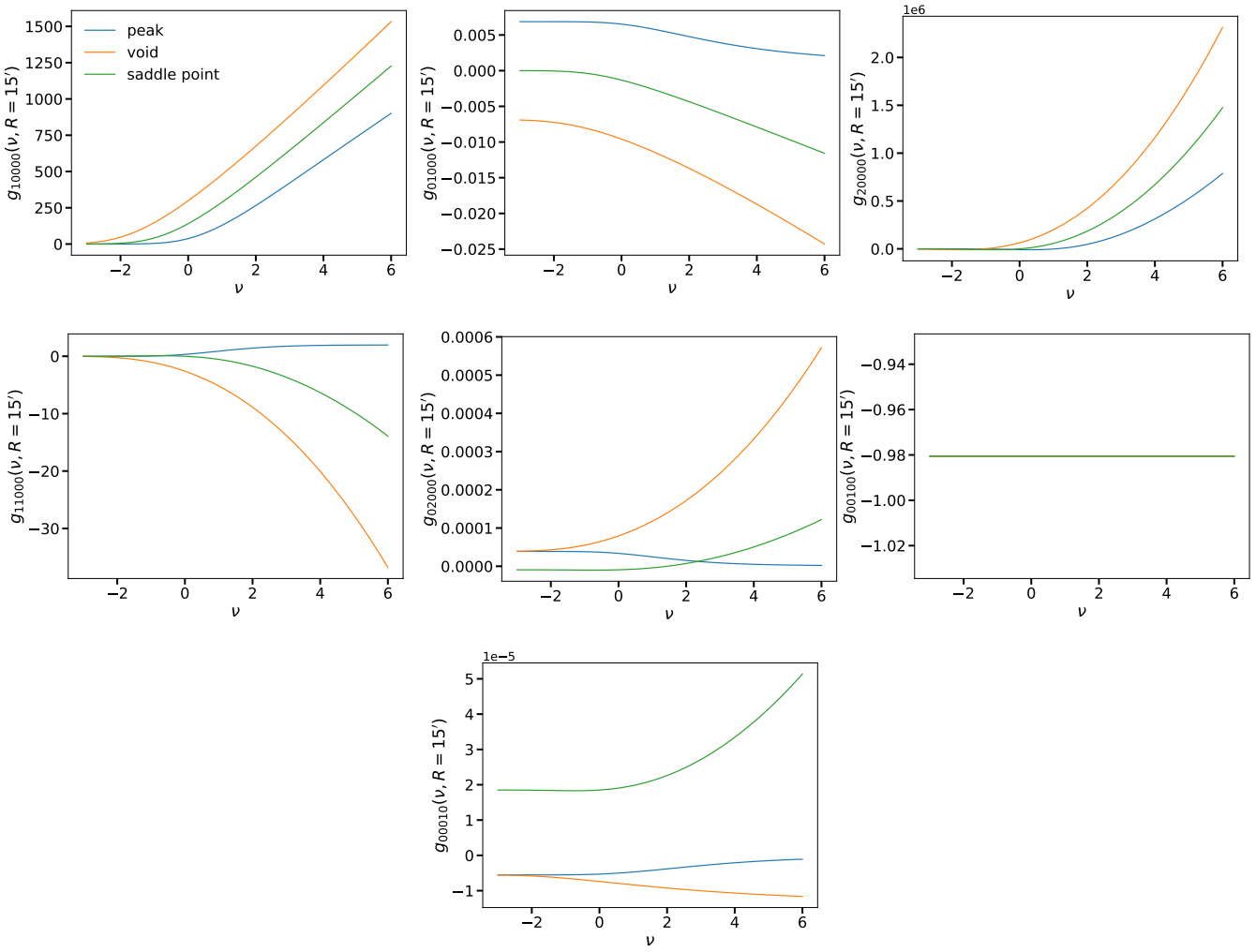

FIG. 8. Plots of g_{ijklm} factors as functions of the threshold ν. The range of ν here is from -3 to 6. The color curves in all the other sub-panels have the same representation as that denoted in the top left subplot. For $g_{00100}(\nu)$ specifically, peaks, voids and saddle points have the same constant function.

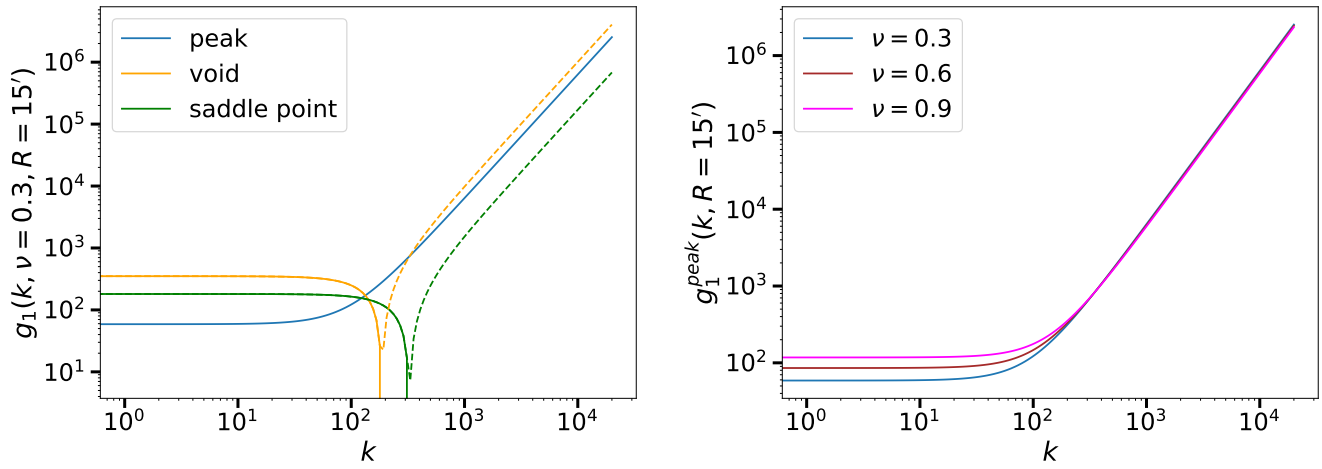

FIG. 9. Plots of $g_1(\mathbf{k})$ in Eq. (38) under a Gaussian smoothing kernel with the smoothing scale $R = 15'$. The left panel shows the g_1 function of different critical point above the same threshold $\nu = 0.3$. The dashed lines are the absolute value of the negative part of the function. The right panel shows the g_1 function of peak above different thresholds.

A theorem of Bessel function enables us to write

$$J_0\left(|\boldsymbol{k}_1 + \boldsymbol{k}_2|\, r\right) = \sum_{n=-\infty}^{\infty} (-1)^n J_n\left(k_1 r\right) J_n\left(k_2 r\right) e^{in\theta} \,, \tag{D4}$$

and the final trick consists in expressing $\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 = \cos(\theta)$

as a sum of exponentials thanks to Euler's formula, and expand its $l^{\rm th}$ power thanks to the Binomial theorem. We get

$$(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2)^l = \frac{1}{2^l} e^{-il\theta} \sum_{m=0}^{l} \binom{l}{m} e^{2im\theta} \ . \tag{D5}$$

Finally, combining Eqs. (D4) and (D5) into Eq. (D3), and noticing that every integral in which $n \neq l - 2m$ is 0 since it gives the integral of $e^{i\theta}$ over the full circle, we get

$$\begin{aligned} A =& 2\pi \int \mathrm{d}r J_0(kr) \frac{1}{2^l} \sum_{m=0}^{l} \binom{l}{m} \\ & \times (-1)^{l-2m} \int \frac{k\mathrm{d}k}{2\pi} J_{l-2m}(kr) X(k) \\ & \times \int \frac{k\mathrm{d}k}{2\pi} J_{l-2m}(kr) Y(k) \ . \end{aligned} \tag{D6}$$

We have thus reduced the computation of A to a product of two 1D integrals which are Fourier transforms (or more specifically Hankel) of X and Y, and can thus be very easily implemented using traditional methods such as FFTs, and a final radial integration. This enables the efficient numerical evaluation of the extrema power spectra presented in this paper.

Appendix E: Comparison of other critical point 2PCFs to MC integration

In the main text, we showed a comparison between our analytical predictions and the exact MC integration results with Gaussian assumption only for peak 2PCF. In this appendix, we show the same comparison but for all pairs of critical points. These include the auto 2PCFs of both voids and saddle points, as well as the cross 2PCFs among all three types of critical points. In Fig. 10, we show the two auto 2PCFs. In both panels, the MC integration with the power law approximation of the weak lensing convergence power spectrum (purple dots) qualitatively exhibits the exclusion zone on small angular scales where our analytic theory is limited in its prediction as discussed in Sect VI. On large angular separations ($\theta > 100'$), our analytic predictions from different orders of perturbative bias expansion converge and agree with the exact MC integration results under the Gaussian assumption. It is worth noticing that on the high amplitude part of both 2PCFs ($\theta \approx 60'$ for voids and $45'$ for saddle points), the 2nd-order Gaussian approximation has quite a discrepancy with respect to the MC integration result, this indicates that the 3rd-order Gaussian approximation term in NNLO would have a more significant role for clustering of voids and saddle points compared to what was shown for peaks. Similar to the peak clustering case, the bispectrum correction here changes the amplitude of the 2PCF but not the overall shape.

In Fig. 11 we show the comparison between all three cross 2PCFs among different types of critical points and their respective MC integration results. In the top panel, we observe that there is not only an exclusion zone on small angular scales between peaks and voids, but also a turnover feature with negative amplitude on $\theta \approx 70'$, which is captured fairly accurately by our analytic predictions. This implies that there are two angular scales on which the clustering between a peak and a void above the same threshold is negatively correlated, different from what we have shown above. On the other hand, MC results in the bottom two plots confirm that there are no exclusion zones between saddle points and the other two types of critical points, although the amplitude of the cross 2PCFs on small angular scales is not well described by the analytic predictions.

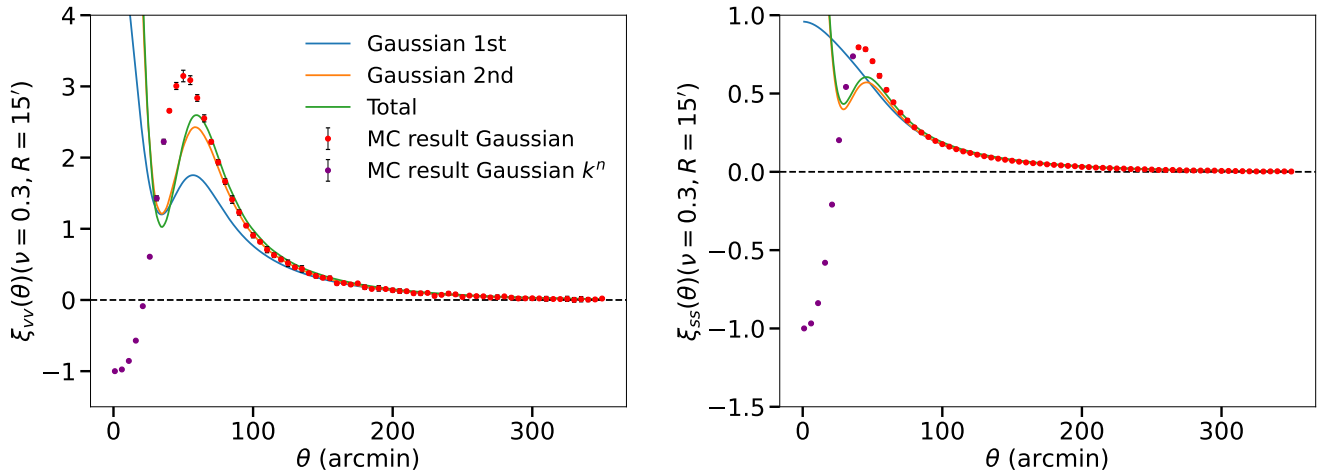

FIG. 10. *Left*: The auto 2PCFs of weak lensing voids (minima) above a threshold $\nu = 0.3$ with a Gaussian smoothing scale of $15'$. All colored curves and scattered dots with error bars have the same representation to those in Fig. 7. *Right*: The same auto 2PCFs as those in the left panel but for weak lensing saddle points.

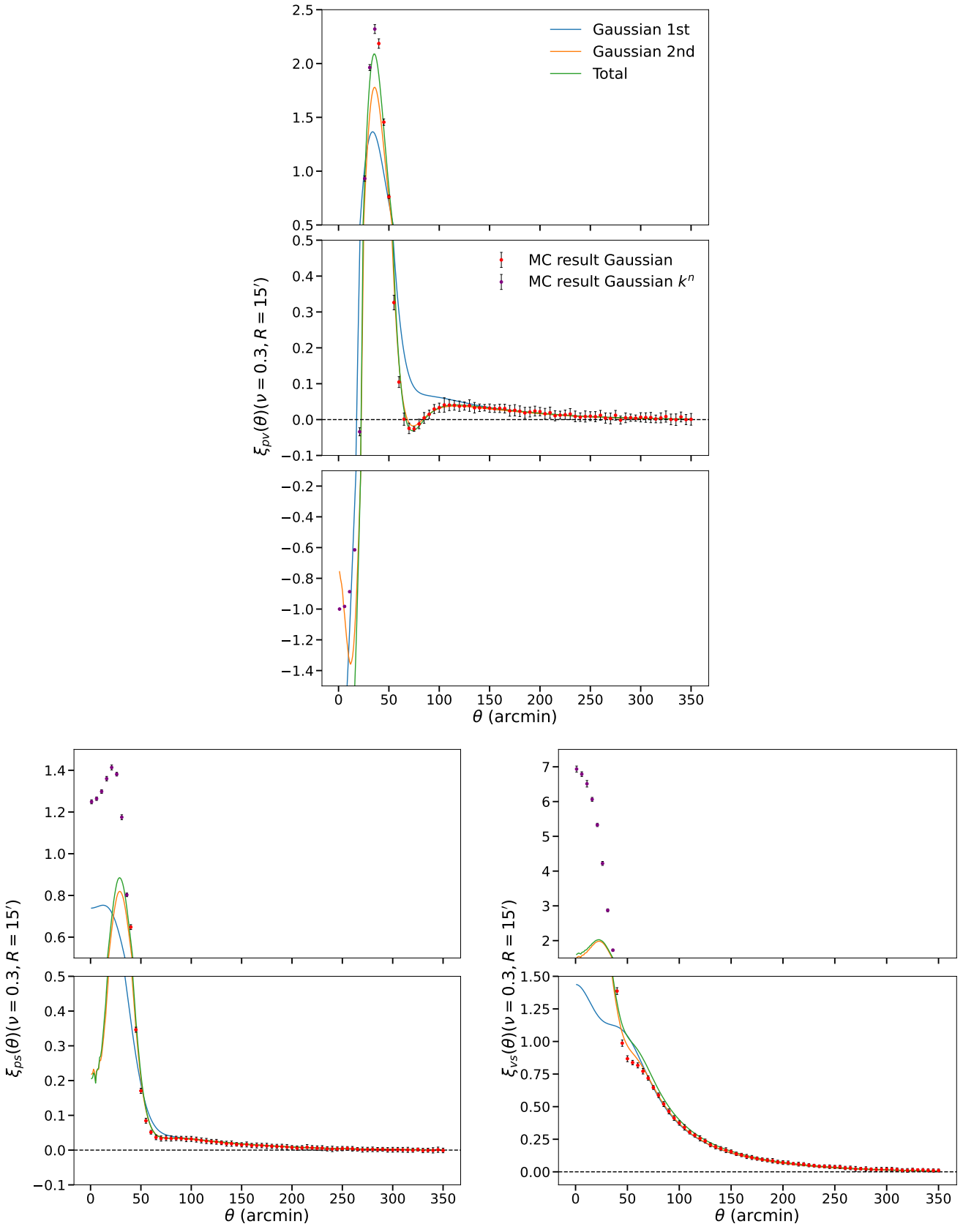

FIG. 11. *Top*: The cross 2PCF between peaks and voids above a threshold $\nu = 0.3$ with a Gaussian smoothing scale of $15'$. All colored curves and scattered dots with error bars have the same representation to those in Fig. 7. The bottom left and right plots show the cross 2PCFs of peak-saddle point, and void-saddle point respectively. We divide each cross 2PCF plot into two or three panels, covering a range of linear scales with different intervals, allowing for a clearer examination of the small amplitude at larger separations.

Chapter 6

Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators

Bibliographic and copyright information

Gong Z., Halder A., Barreira A., Seitz S., and Friedrich O., *Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators*, JCAP, 07 (2023) 040

DOI: 10.1088/1475-7516/2023/07/040

Outline

This paper is part of a broader series, including Halder et al. (2021), Halder & Barreira (2022) and Halder and **Gong** et al. (in prep). The overarching scientific aim of this series is to apply the integrated shear 3PCF $\zeta_\pm$ to the statistical analysis of the Dark Energy Year 3 (DESY3) (Gatti et al., 2021) weak lensing shear data, extracting non-Gaussian information beyond the standard cosmic shear 2PCF $\xi_\pm$. A key focus is on improving parameter constraints, particularly for $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$ and the dark energy equation of state parameter w_0, through the joint inference of $\zeta_\pm$ and $\xi_\pm$. This paper serves as a crucial step in preparing for the final data analysis by constructing a complete statistical inference pipeline and validating it on simulated, realistic datasets. As the first author, I was responsible for the following key tasks: (i) developing a machine-learning emulator for $\zeta_\pm$ to accelerate its calculation using the `cosmopower` framework (Spurio Mancini et al., 2022), (ii) implementing the code to estimate the data covariance of $\xi_\pm$ and $\zeta_\pm$ from

the simulated T17 weak lensing shear maps (Takahashi et al., 2017), (iii) programming a GPU-based Markov Chain Monte Carlo (MCMC) algorithm for efficient Bayesian posterior sampling of cosmological parameters[1] and (iv) conducting validation and optimization tests discussed in the paper. I also wrote majority of the manuscript. The co-authors contributed to modeling the survey and astrophysical systematic effects, refining the machine-learning model, preparing simulation data, providing valuable scientific discussions, drafting specific sections of the manuscript, and reviewing the final draft.

[1]The sampling program is based on the package https://github.com/justinalsing/affine

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators

Zhengyangguang Gong,[a,b] **Anik Halder,**[a,b] **Alexandre Barreira,**[c,d] **Stella Seitz**[a,b] **and Oliver Friedrich**[a]

[a]Universitäts-Sternwarte, Fakultät für Physik,
Ludwig-Maximilians Universität München,
Scheinerstraße 1, 81679 München, Germany
[b]Max Planck Institute for Extraterrestrial Physics,
Giessenbachstraße 1, 85748 Garching, Germany
[c]Excellence Cluster ORIGINS,
Boltzmannstraße 2, 85748 Garching, Germany
[d]Ludwig-Maximilians-Universität,
Schellingstraße 4, 80799 München, Germany

E-mail: lgong@usm.lmu.de, ahalder@usm.lmu.de, alex.barreira@origins-cluster.de, stella@usm.lmu.de, oliver.friedrich@physik.uni-muenchen.de

Received April 18, 2023
Revised June 27, 2023
Accepted June 28, 2023
Published July 13, 2023

Abstract. The integrated shear 3-point correlation function $\zeta_\pm$ measures the correlation between the local shear 2-point function $\xi_\pm$ and the 1-point shear aperture mass in patches of the sky. Unlike other higher-order statistics, $\zeta_\pm$ can be efficiently measured from cosmic shear data, and it admits accurate theory predictions on a wide range of scales as a function of cosmological and baryonic feedback parameters. Here, we develop and test a likelihood analysis pipeline for cosmological constraints using $\zeta_\pm$. We incorporate treatment of systematic effects from photometric redshift uncertainties, shear calibration bias and galaxy intrinsic alignments. We also develop an accurate neural-network emulator for fast theory predictions in MCMC parameter inference analyses. We test our pipeline using realistic cosmic shear maps based on N-body simulations with a DES Y3-like footprint, mask and source tomographic bins, finding unbiased parameter constraints. Relative to $\xi_\pm$-only, adding $\zeta_\pm$ can lead to $\approx 10-25\%$ improvements on the constraints of parameters like A_s (or σ_8) and w_0. We find no evidence in $\xi_\pm + \zeta_\pm$ constraints of a significant mitigation of the impact of systematics.

https://doi.org/10.1088/1475-7516/2023/07/040

JCAP07(2023)040

We also investigate the impact of the size of the apertures where $\zeta_\pm$ is measured, and of the strategy to estimate the covariance matrix (N-body vs. lognormal). Our analysis solidifies the strong potential of the $\zeta_\pm$ statistic and puts forward a pipeline that can be readily used to improve cosmological constraints using real cosmic shear data.

Keywords: weak gravitational lensing, cosmological parameters from LSS, Machine learning

ArXiv ePrint: 2304.01187

JCAP07(2023)040

Contents

1 Introduction **1**

2 Theoretical formalism **3**
- 2.1 Integrated shear 3-point correlation function 3
- 2.2 The three-dimensional matter bispectrum model 5
- 2.3 Systematic error effects 6

3 Data vector and covariance from simulations **7**
- 3.1 Shear maps from N-body simulations 7
- 3.2 Shear maps from lognormal realizations 8
- 3.3 Data vector and covariance measurements 8

4 Emulators for $\boldsymbol{\xi_\pm}$ and $\boldsymbol{\zeta_\pm}$ **11**

5 Results: simulated likelihood analyses with MCMC **14**
- 5.1 Validation on the T17 cosmic shear maps 14
- 5.2 The impact of the aperture size 14
- 5.3 The impact of systematics and their modelling 17
- 5.4 The impact of different covariance estimates 19

6 Summary & conclusion **20**

A The modelling of intrinsic alignments **22**

JCAP07(2023)040

1 Introduction

The weak gravitational lensing effect is the bending of the light of background source galaxies by foreground gravitational potentials [1, 2]. This induces a coherent distortion pattern in the observed shape of the background galaxies that is called the *cosmic shear field*. The statistics of this field depend on the three-dimensional large-scale structure, hence cosmic shear studies offer a powerful way to address key questions in cosmology such as the structure formation history, the nature of dark energy and dark matter, and the laws of gravity on large scales. Indeed, cosmic shear is one of the most active research areas in large-scale structure today: the DES [3], KiDS [4] and HSC-SSP [5] surveys have recently presented cosmological constraints from their cosmic shear data, and more accurate and bigger data sets will be available soon with missions like Euclid [6], Vera Rubin's LSST [7] and Nancy Roman [8].

The majority of cosmic shear analyses are based on the shear 2-point correlation function (2PCF), or its Fourier counterpart the lensing power spectrum. These statistics completely characterize the information content of Gaussian random fields, which our Universe was close to at the earliest stages of its evolution, as well as today on sufficiently large-scales. At late times, however, the evolution of matter density fluctuations becomes nonlinear on small scales, inducing non-Gaussian features in the cosmic shear field that cannot be described by 2PCF alone. Higher-order statistics are thus needed to access the non-Gaussian information.

The shear 3-point correlation function (3PCF; or its Fourier counterpart the lensing bispectrum) is the natural first step beyond the 2PCF [9–14]. However, being a more complicated statistic, it is more challenging to measure observationally, as well as to predict

theoretically. Concretely, compared to the 2PCF which depends only on the distance between two points in the survey footprint, the 3PCF is a function of the size and shape of triangles connecting three points, which requires more demanding estimators. Additionally, theoretical predictions require accurate prescriptions for the nonlinear matter bispectrum, which despite recent progress [15, 16], are still not as developed as the matter power spectrum that enters the shear 2PCF. Further complications arise by the need to account for baryonic feedback effects, as well as systematics effects such as photometric redshift uncertainties, shear multiplicative bias and galaxy intrinsic alignments (IA). This helps explain why existing real-data constraints using higher-order shear information are based not on the full 3-point correlation function, but on other statistics including aperture moments [17–21], lensing peaks [22–24], density-split statistics [25–29] and persistent homology of cosmic shear [30, 31]. The shear 3PCF was recently measured using DES Year 3 (Y3) data [20], although only in patches over the survey and not over the whole footprint as that would be too computationally demanding.

In this paper, we focus on a particular kind of shear 3PCF called the *integrated shear 3-point correlation function* [32]. This statistic corresponds to the correlation between the shear 2PCF measured in patches of the sky with the 1-point shear aperture mass in those patches.[1] Physically, this statistic describes the modulation of the local shear 2PCF by long-wavelength features in the cosmic shear field. The integrated shear 3PCF enjoys two key advantages relative to other higher-order shear statistics. The first is that it is straightforward to measure from the data as it requires only conventional and well-tested shear 2PCF estimators. The second is that, as shown in ref. [38], this statistic is sensitive to the squeezed matter bispectrum that can be evaluated accurately in the nonlinear regime using the response approach to perturbation theory [39]. Importantly, the response approach allows to account for the impact of baryonic feedback on small scales, which is crucial to design scale cuts and/or marginalize over these uncertainties in real data analyses.

Our goal here is to develop and test a likelihood analysis pipeline to reliably extract cosmology from real cosmic shear data using the integrated shear 3PCF. Concretely, we incorporate the impact of baryonic feedback (as in ref. [38]), as well as of photometric redshift uncertainties, shear multiplicative bias and galaxy IA. We also develop a neural-network (NN) emulator for the theory model to enable fast theory predictions in Monte-Carlo Markov Chain (MCMC) parameter inference analyses. We test our analysis pipeline on simulated cosmic shear maps with DES Y3-like survey footprints and source galaxy redshift distributions. We study in particular (i) the ability of the theory model to return unbiased parameter constraints,[2] (ii) the impact of the size of the aperture where the integrated shear 3PCF is measured, (iii) the ability of combined 2PCF and 3PCF analyses to mitigate the impact of systematic uncertainties, and (iv) the impact of different data vector covariance estimates.

In terms of constraining power, we find that the integrated shear 3PCF leads to improvements of $\approx 10-25\%$ on the constraints of parameters like the amplitude of primordial density fluctuations A_s (or equivalently σ_8) or the dark energy equation of state parameter w_0. This is consistent with the previous findings of refs. [32, 38] based on idealized Fisher matrix forecasts, but now in the context of realistically simulated MCMC likelihood analyses. Our results thus strongly motivate as next steps exploring the power of this statistic to improve cosmological constraints using real cosmic shear data.

[1] See also refs. [33, 34] for earlier applications of the same idea in the context of the three-dimensional galaxy distribution, and refs. [35–37] for studies of the Fourier counterpart of the integrated shear 3PCF.

[2] Throughout the paper we loosely use the term "unbiased constraints" to mean that the 68% posterior credible intervals encompass the true model parameter values.

JCAP07(2023)040

This paper is structured as follows: In section 2 we review the theoretical formalism behind the integrated shear 3PCF and describe how we incorporate lensing systematic effects. In section 3 we describe the construction of our DES Y3-like cosmic shear maps, as well as the measurements of the shear 2PCF, integrated 3PCF and their (cross) covariance matrices. We describe and discuss the performance of our NN emulator of the theory predictions for fast MCMC likelihood analyses in section 4. Our main numerical results are shown in section 5. We summarize and conclude in section 6. Appendix A describes our modelling of the galaxy IA.

2 Theoretical formalism

In this section we describe the theory behind the integrated shear 3PCF. We begin with a recap of the model of refs. [32, 38], and then discuss how we incorporate lensing systematics.

2.1 Integrated shear 3-point correlation function

The integrated shear 3PCF, $\zeta_{\pm,ijk}(\boldsymbol{\alpha})$, is defined as

$$\zeta_{\pm,ijk}(\boldsymbol{\alpha}) \equiv \langle M_{\mathrm{ap},i}(\boldsymbol{\theta}_C)\hat{\xi}_{\pm,jk}(\boldsymbol{\alpha};\boldsymbol{\theta}_C)\rangle\,, \tag{2.1}$$

where $M_{\mathrm{ap},i}(\boldsymbol{\theta}_C)$ is the 1-point aperture mass statistic measured on a patch of the survey centered at angular position $\boldsymbol{\theta}_C$, and $\hat{\xi}_{\pm,jk}(\boldsymbol{\alpha};\boldsymbol{\theta}_C)$ is the shear 2PCF measured on the same patch of the sky; $\boldsymbol{\alpha}$ describes angular separations. The angle brackets denote ensemble average (or in practice, averaging over all positions $\boldsymbol{\theta}_C$) and the subscripts i, j, k denote tomographic source bins, i.e. $\hat{\xi}_{\pm,jk}$ is the 2PCF of the shear fields from galaxy shape measurements at the redshift bins j and k. This equation makes apparent the interpretation of the shear 3PCF as describing the spatial modulation of the local 2PCF by the local shear mass aperture, which describes larger-scale features in the shear field.

The aperture mass $M_{\mathrm{ap}}(\boldsymbol{\theta}_C)$ is defined as [2, 40]

$$M_{\mathrm{ap}}(\boldsymbol{\theta}_C) = \int \mathrm{d}^2\boldsymbol{\theta}\ \kappa(\boldsymbol{\theta})U(\boldsymbol{\theta}_C - \boldsymbol{\theta})\,, \tag{2.2}$$

where $\kappa(\boldsymbol{\theta})$ is the lensing convergence field, and U is an azimuthally symmetric filter function with angular size θ_{ap}. The convergence field is not directly observable, but if U is a compensated filter satisfying $\int \mathrm{d}^2\boldsymbol{\theta}\ U(\boldsymbol{\theta}_C - \boldsymbol{\theta}) = 0$, then $M_{\mathrm{ap}}(\boldsymbol{\theta}_C)$ can be expressed as

$$M_{\mathrm{ap}}(\boldsymbol{\theta}_C) = \int \mathrm{d}^2\boldsymbol{\theta}\ \gamma_{\mathrm{t}}(\boldsymbol{\theta}, \phi_{\boldsymbol{\theta}_C-\boldsymbol{\theta}})Q(\boldsymbol{\theta}_C - \boldsymbol{\theta})\,, \tag{2.3}$$

where γ_{t} is the tangential component of the shear field (which is directly observable), $\phi_{\boldsymbol{\theta}_C-\boldsymbol{\theta}}$ is the polar angle of the angular separation between $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}$, and Q is a filter function related to U. As in previous works, we adopt the following form for U and Q [41]

$$U(\theta) = \frac{1}{2\pi\theta_{\mathrm{ap}}^2}\left(1 - \frac{\theta^2}{2\theta_{\mathrm{ap}}^2}\right)\exp\left(-\frac{\theta^2}{2\theta_{\mathrm{ap}}^2}\right)\,, \tag{2.4}$$

$$Q(\theta) = \frac{\theta^2}{4\pi\theta_{\mathrm{ap}}^2}\exp\left(-\frac{\theta^2}{2\theta_{\mathrm{ap}}^2}\right)\,; \tag{2.5}$$

JCAP07(2023)040

note the filters depend only on the magnitude of the arguments because of the azimuthal symmetry. The Fourier transform of U, which appears in equations below, is given by

$$U(\ell) = \int \mathrm{d}^2\boldsymbol{\theta}\; U(\theta) e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} = \frac{\ell^2\theta_{\rm ap}^2}{2}\exp\left(-\frac{\ell^2\theta_{\rm ap}^2}{2}\right), \tag{2.6}$$

where $\boldsymbol{\ell}$ is a two-dimensional wavevector on the sky (we assume the flat-sky approximation).

The other term in eq. (2.1), $\hat{\xi}_\pm(\boldsymbol{\alpha};\boldsymbol{\theta}_C)$, is the 2PCF of the *windowed* shear field $\gamma(\boldsymbol{\theta};\boldsymbol{\theta}_C) \equiv \gamma(\boldsymbol{\theta})W(\boldsymbol{\theta}_C - \boldsymbol{\theta})$, where the window function W is a top-hat of size $\theta_{\rm T}$ at position $\boldsymbol{\theta}_C$. The two 2PCFs are defined as

$$\begin{aligned}
\hat{\xi}_+(\boldsymbol{\alpha};\boldsymbol{\theta}_C) &\equiv \frac{1}{A_{\rm 2pt}(\boldsymbol{\alpha})}\int \mathrm{d}^2\boldsymbol{\theta}\;\gamma(\boldsymbol{\theta};\boldsymbol{\theta}_C)\gamma^*(\boldsymbol{\theta}+\boldsymbol{\alpha};\boldsymbol{\theta}_C)\\
\hat{\xi}_-(\boldsymbol{\alpha};\boldsymbol{\theta}_C) &\equiv \frac{1}{A_{\rm 2pt}(\boldsymbol{\alpha})}\int \mathrm{d}^2\boldsymbol{\theta}\;\gamma(\boldsymbol{\theta};\boldsymbol{\theta}_C)\gamma(\boldsymbol{\theta}+\boldsymbol{\alpha};\boldsymbol{\theta}_C)e^{-4i\phi_{\boldsymbol{\alpha}}},
\end{aligned} \tag{2.7}$$

where * denotes complex conjugation, $\phi_{\boldsymbol{\alpha}}$ is the polar angle of $\boldsymbol{\alpha}$, and $A_{\rm 2pt}(\boldsymbol{\alpha}) \equiv \int \mathrm{d}^2\boldsymbol{\theta}\; W(\boldsymbol{\theta}_C - \boldsymbol{\theta})W(\boldsymbol{\theta}_C - \boldsymbol{\theta} - \boldsymbol{\alpha})$. The Fourier transform of W appears in equations below, and is given by

$$W(\boldsymbol{l}) = W(l) = 2\pi\theta_{\rm T}^2\,\frac{J_1(l\theta_{\rm T})}{l\theta_{\rm T}}, \tag{2.8}$$

where J_n is the nth-order Bessel function of the first kind.

Skipping the details of the derivation [32], the two 3PCF in eq. (2.1) can be written as

$$\zeta_{+,ijk}(\boldsymbol{\alpha}) = \frac{1}{A_{\rm 2pt}(\alpha)}\int\frac{\mathrm{d}\ell\ell}{2\pi}\mathcal{B}^{\rm 2D}_{+,ijk}(\boldsymbol{\ell})J_0(\ell\alpha)\,, \tag{2.9}$$

$$\zeta_{-,ijk}(\boldsymbol{\alpha}) = \frac{1}{A_{\rm 2pt}(\alpha)}\int\frac{\mathrm{d}\ell\ell}{2\pi}\mathcal{B}^{\rm 2D}_{-,ijk}(\boldsymbol{\ell})J_4(\ell\alpha)\,, \tag{2.10}$$

where $\mathcal{B}^{\rm 2D}_\pm$ is called the integrated lensing bispectrum, and it is given by (in the Limber approximation)

$$\begin{aligned}
\mathcal{B}^{\rm 2D}_{\pm,ijk}(\boldsymbol{\ell}) = \int \mathrm{d}\chi\frac{q^i(\chi)q^j(\chi)q^k(\chi)}{\chi^4}\int\frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\boldsymbol{\ell}_2}{(2\pi)^2}B^{\rm 3D}_\delta\left(\frac{\boldsymbol{\ell}_1}{\chi},\frac{\boldsymbol{\ell}_2}{\chi},\frac{-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2}{\chi},\chi\right)\\
\times\, e^{2i(\phi_2\mp\phi_{-1-2})}U(\boldsymbol{\ell}_1)W(\boldsymbol{\ell}_2+\boldsymbol{\ell})W(-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2-\boldsymbol{\ell})\,.
\end{aligned} \tag{2.11}$$

In this equation, B^{3D}_δ is the 3-dimensional matter bispectrum (discussed below), ϕ_2 is the polar angle of $\boldsymbol{\ell}_2$, ϕ_{-1-2} is the polar angle of $-\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2$, and $q(\chi)$ is the lensing kernel

$$q^i(\chi) = \frac{3H_0^2\Omega_m}{2c^2}\frac{\chi}{a(\chi)}\int_\chi \mathrm{d}\chi' n_s^i(\chi')\frac{\chi'-\chi}{\chi'}\,, \tag{2.12}$$

where $n_s^i(\chi)$ is the galaxy source number density distribution for the redshift tomographic bin i, χ denotes comoving distances, H_0 is the Hubble parameter, Ω_m is the cosmic matter density parameter today, c is the speed of light and $a(\chi)$ is the scale factor; note that throughout the paper we always assume spatially flat cosmologies.

In our results, we will consider also the *global* shear 2PCF, which can be evaluated as

$$\xi_{+,ij}(\alpha) = \int\frac{\mathrm{d}\ell\ell}{2\pi}P_{\kappa,ij}(\ell)J_0(\ell\alpha)\,, \tag{2.13}$$

$$\xi_{-,ij}(\alpha) = \int\frac{\mathrm{d}\ell\ell}{2\pi}P_{\kappa,ij}(\ell)J_4(\ell\alpha)\,, \tag{2.14}$$

JCAP07(2023)040

where $P_{\kappa,ij}$ is the convergence power spectrum given by (in the Limber approximation)

$$P_{\kappa,ij}(\ell) = \int \mathrm{d}\chi \frac{q^i(\chi) q^j(\chi)}{\chi^2} P_\delta^{3D}\left(k = \frac{\ell}{\chi}, \chi\right) , \tag{2.15}$$

with P_δ^{3D} the three-dimensional matter power spectrum.

2.2 The three-dimensional matter bispectrum model

A key ingredient to evaluate $\zeta_\pm$ is the three-dimensional matter bispectrum $B_\delta^{\rm 3D}$ in eq. (2.11), which we evaluate following ref. [38] as

$$B_\delta^{\rm 3D}(\boldsymbol{k_1}, \boldsymbol{k_2}, \boldsymbol{k_3}, \chi) = \begin{cases} B_{\delta,\rm RF}^{\rm 3D}\,, & f_{\rm sq} \geq f_{\rm sq}^{\rm thr} \Longrightarrow \text{squeezed} \\ B_{\delta,\rm GM}^{\rm 3D}\,, & \text{otherwise} \end{cases} , \tag{2.16}$$

where $B_{\delta,\rm RF}^{\rm 3D}$ is the bispectrum expression of the response function approach valid for squeezed configurations, and $B_{\delta,\rm GM}^{\rm 3D}$ is the bispectrum fitting formula of ref. [42]. The parameter $f_{\rm sq}$ is defined as $f_{\rm sq} = k_m/k_s$, with k_s (k_m) the smallest (intermediate) of the amplitudes of the three modes $\boldsymbol{k_i}$. As explained in ref. [38], this equation guarantees that the response function branch correctly evaluates the squeezed matter bispectrum configurations in the nonlinear regime, which determine the value of $\zeta_\pm$ on small angular scales. The value of $f_{\rm sq}^{\rm thr}$ is the threshold that defines whether a given bispectrum configuration is dubbed as squeezed or not. Ref. [38] found that a range of values around $f_{\rm sq}^{\rm thr} \approx 7$ yield good fits to simulation measurements; in this paper we adopt $f_{\rm sq}^{\rm thr} = 7$.

The response function branch in eq. (2.16) is given by

$$B_{\delta,\rm RF}^{3D}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, z) = \left[R_1(k_h, z) + \left(\mu^2_{\boldsymbol{k}_h,\boldsymbol{k}_s} - \frac{1}{3}\right) R_K(k_h, z)\right] P_\delta^{3D}(k_h, z) P_{\delta,L}^{3D}(k_s, z)\,, \tag{2.17}$$

where $\boldsymbol{k}_h$ denotes the mode $\boldsymbol{k}_i$ with the highest magnitude, $\mu_{\boldsymbol{k}_i,\boldsymbol{k}_j}$ is the cosine of the angle between $\boldsymbol{k}_i$ and $\boldsymbol{k}_j$, $P_{\delta,L}^{3D}$ is the three-dimensional linear matter power spectrum and $R_1(k,z)$ and $R_K(k,z)$ are the first-order response functions of the matter power spectrum to large-scale density and tidal fields:

$$R_1(k,z) = 1 - \frac{1}{3}\frac{\mathrm{d}\ln P_\delta^{\rm 3D}(k,z)}{\mathrm{d}\ln k} + G_1(k,z), \tag{2.18}$$

$$R_K(k,z) = G_K(k,z) - \frac{\mathrm{d}\ln P_\delta^{\rm 3D}(k,z)}{\mathrm{d}\ln k}. \tag{2.19}$$

In these expressions, G_1 and G_K are the so-called *growth-only* response functions, which can be measured in the nonlinear regime of structure formation using separate universe simulations. Just as in ref. [38], we use the results of ref. [43] for G_1 and ref. [44] for G_K.

The GM branch is in turn given by

$$B_{\delta,\rm GM}^{3D}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, z) = 2F_2^{\rm eff}(\boldsymbol{k}_1, \boldsymbol{k}_2, z) P_\delta^{3D}(k_1, z) P_\delta^{3D}(k_2, z) + \text{cyclic permutations}\,, \tag{2.20}$$

where $F_2^{\rm eff}(\boldsymbol{k}_1, \boldsymbol{k}_2, z)$ is a modified version of the 2-point mode coupling kernel with free functions calibrated against N-body simulations [42].

In this paper, we evaluate the nonlinear matter power spectrum using the `HMcode` [45] implementation inside the publicly available Boltzmann code `CLASS` [46]; to model the impact

JCAP07(2023)040

of baryonic feedback effects, we adopt the single parameter $c_{\rm min}$ parametrization, where $c_{\rm min}$ roughly describes the strength of feedback by active galactic nuclei (AGN). As discussed in refs. [47, 48], mode-coupling terms like $F_2^{\rm eff}$, G_1 and G_K are expected to be very weakly dependent on baryonic physics. This way, the impact of baryonic effects on the bispectrum is trivially propagated by that on the power spectrum; note that in practice the baryonic effects impact only the response function branch in eq. (2.16), since the GM branch contributes only on large scales [38] where baryonic effects have a negligible role.

2.3 Systematic error effects

Reference [38] has shown how to include the impact of baryonic feedback effects on $\zeta_\pm$, which are one of the main non-cosmological contaminants in cosmic shear analyses. In this subsection we describe how we take into account a series of other important systematic effects, namely photometric redshift uncertainties, multiplicative shear bias and galaxy IA.

Photometric redshift (photo-z) uncertainties have a direct impact on the galaxy source redshift distribution. Here, we follow a strategy commonly adopted in real-data analyses and parametrize their effect through a single shift parameter Δz defined as

$$n_s^i(z) = \hat{n}_s^i(z + \Delta z^i)\,, \tag{2.21}$$

where $\hat{n}_s^i$ is the default estimate for the galaxy source redshift bin i. This simple way to account for photo-z uncertainties was found sufficient at the statistical power of DES-Y3 analyses (see figure 10 in ref. [49]).

Again, as common in the literature, we model biases from the shear measurement pipeline with multiplicative factors $1 + m_i$ for each tomographic bin i. In practice, this implies the following transformations of $\xi_\pm$ and $\zeta_\pm$,

$$\xi_{\pm,ij}(\alpha) \longrightarrow (1+m_i)(1+m_j)\xi_{\pm,ij}(\alpha)\,, \tag{2.22}$$
$$\zeta_{\pm,ijk}(\alpha) \longrightarrow (1+m_i)(1+m_j)(1+m_k)\zeta_{\pm,ijk}(\alpha)\,. \tag{2.23}$$

We assume that any additive bias component is well calibrated by lensing image simulations and removed from the measurement pipeline [50].

Finally, we consider the effect of galaxy IA that describe intrinsic correlations between the shapes of source galaxies and their local tidal fields, i.e. correlations that are not induced by the gravitational lensing effect. We adopt the *nonlinear linear alignment* (NLA) model [51, 52] for both $\xi_\pm$ and $\zeta_\pm$. In practice, the incorporation of IA in our theory predictions is equivalent to transforming the lensing kernels as (see appendix A for more details)

$$q^i(\chi) \longrightarrow q^i(\chi) + f_{\rm IA}(z(\chi))\frac{n_s^i(\chi)}{\bar{n}_s^i}\frac{{\rm d}z}{{\rm d}\chi}\,, \tag{2.24}$$

with $\bar{n}_s^i$ the mean source galaxy density in tomographic bin i and [53, 54]

$$f_{\rm IA}(z) = -A_{\rm IA,0}\left(\frac{1+z}{1+z_0}\right)^{\alpha_{\rm IA}}\frac{c_1\rho_{\rm crit}\Omega_{\rm m,0}}{D(z)}\,, \tag{2.25}$$

where $A_{\rm IA,0}$ is the IA amplitude, $\alpha_{\rm IA}$ is a power index and $D(z)$ is the linear growth factor. We adopt $z_0 = 0.62$, $c_1\rho_{\rm crit} = 0.0134$ [52, 55]. In our results below we keep the power index fixed to $\alpha_{\rm IA} = 0$ for simplicity; note that simultaneously varying $A_{\rm IA,0}$ and $\alpha_{\rm IA}$ in MCMC

JCAP07(2023)040

constraints can lead to posterior projection effects that could artificially bias the marginalized constraints of $A_{\rm IA,0}$ towards zero.

In our modelling of IA, $\xi_\pm$ acquires terms $\propto f_{\rm IA}, f_{\rm IA}^2$, and $\zeta_\pm$ terms $\propto f_{\rm IA}, f_{\rm IA}^2, f_{\rm IA}^3$. These are different from the terms displayed in ref. [56]; further notice that our eq. (2.25) differs from the corresponding eq. (27) in ref. [56] by a multiplicative factor $1/(1+z)$. We shall return to the impact of different IA treatments when we discuss our numerical results.

3 Data vector and covariance from simulations

In this section we describe the DES Y3-like simulated cosmic shear maps that we use to measure the $\xi_\pm$ and $\zeta_\pm$ data vectors and to estimate their covariance matrices.

3.1 Shear maps from *N*-body simulations

Our main cosmic shear maps are obtained using the publicly available N-body simulation data developed by Takahashi et al. [57] (hereafter referred to as T17). In particular, we make use of the 108 independent full-sky cosmic shear maps for several Dirac-delta source distributions at redshifts between $z = 0.05$ and $z = 5.3$. The cosmology of the simulations is flat ΛCDM with parameters: $\Omega_m = 0.279$, $\Omega_b = 0.046$, $h = 0.7$, $\sigma_8 = 0.82$, $n_s = 0.97$.

We consider DES Y3-like galaxy source redshift distributions to construct our cosmic shear maps. For simplicity, rather than considering the four source bins utilized in the DES Y3 analysis, we merge them into two as follows. Let N_1 and N_2 be the total number of galaxies in the first two DES source distributions $n_{s,\rm DES1}(z)$ and $n_{s,\rm DES2}(z)$, respectively (see figure 6 and 11 in ref. [58]). Then, our first source redshift bin is obtained as $\hat{n}_s^1 = (N_1 n_{s,\rm DES1} + N_2 n_{s,\rm DES2})/(N_1 + N_2)$; and similarly for our second source redshift, using the third and fourth DES Y3 source distributions. The source redshift distributions that we consider in this paper are shown on the left of figure 1. For each of the 108 T17 realizations, we build two full-sky shear maps by summing the T17 shear maps weighted by each of the two source redshift distributions. The vertical lines on the left of figure 1 mark the source redshift of the T17 maps we use.

We then apply the DES Y3 footprint to each of the full-sky shear maps. In order to maximize the utility of each full-sky map, we place 5 footprints in each with minimal overlap, as illustrated on the right of figure 1. For each of our two source bins, this provides us with $108 \times 5 = 540$ DES Y3-like shear maps on which we can measure $\xi_\pm$, $\zeta_\pm$ and their covariance.

Finally, we add DES Y3 levels of shape noise to our maps as follows. Using the angular positions of the source galaxies in the DES Y3 shape catalogue [59], we assign to each of our pixels the galaxy ellipticities and measurement weights that are also present in those catalogues. We then randomly rotate the ellipticities of the galaxies assigned to each pixel. The shape noise $\gamma_{\rm noise}$ is the average of these randomly rotated ellipticities weighted by the corresponding measurement weights. This is added to the shear values of the T17 maps $\gamma_{\rm sim}$ to generate the shear measurement in each pixel $\gamma_{\rm pix}$. Concretely,

$$\gamma_{\rm pix} = \gamma_{\rm noise} + \gamma_{\rm sim} = \frac{\sum_{j=1}^N \omega_j \gamma_{j,\rm DES} \exp(i\phi_j)}{\sum_{j=1}^N \omega_j} + \gamma_{\rm sim}\,, \tag{3.1}$$

where N is the number of galaxies in a given pixel, $\gamma_{j,\rm DES}$ and ω_j are the measured ellipticity and weight of the jth galaxy and each angle ϕ_j is drawn uniformly from $[0, 2\pi]$; note that the average value of $\gamma_{\rm noise}$ across all pixels is zero, but each pixel has in general nonzero values.

JCAP07(2023)040

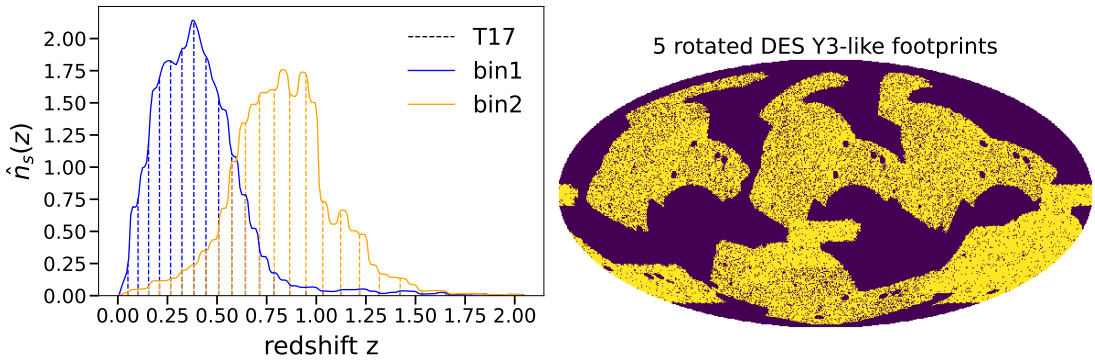

Figure 1. *Left panel*: The two galaxy source redshift distributions that we consider in this paper. Each is a combination of two of the four DES Y3 source distributions. The vertical dashed lines mark the source redshifts of the T17 shear maps, which are weighted by the galaxy source distributions to produce our shear maps. *Right panel*: The Mollweide projection map with the placement of 5 DES Y3 survey footprints after the selection with Q filters of 90 arcmin in a full-sky map; dark blue pixels indicate masked/unobserved regions. This allows us to measure 5 DES Y3-like realizations of $\xi_\pm$ and $\zeta_\pm$ from each full-sky map.

3.2 Shear maps from lognormal realizations

In addition to the T17-based shear maps, we also consider DES Y3-like maps from lognormal lensing realizations generated with the Full-sky Lognormal Astro-fields Simulation Kit [60] (hereafter referred to as FLASK). FLASK takes as input the lensing convergence power spectrum, which we compute theoretically for the T17 cosmology and our two galaxy source redshift distributions. FLASK requires also the value of a logshift parameter, which we obtain by fitting a lognormal probability distribution function (PDF) to the PDF of the T17 maps (see section 4.2 of [32] for more details about the generation of our FLASK shear maps). For each of our two source bins, we generate a total of 300 independent FLASK full-sky cosmic shear maps, on which we place 5 DES Y3-like footprints analogously to the T17 full-sky maps (cf. right panel of figure 1). We add shape noise following the strategy described above for the T17 maps. For each of the two source bins then, we have a total of $5 \times 300 = 1500$ lognormal realizations of a DES Y3-like footprint on which we can measure $\xi_\pm$ and $\zeta_\pm$.

3.3 Data vector and covariance measurements

We use the `Treecorr` code [61] to measure $\xi_{\pm,ij}(\alpha)$ on 15 log-spaced angular bins between 5 and 250 arcmin; these are scales comparable to those adopted in the DES Y3 analysis [62]. We measure the auto- and cross-correlation of the two source redshift bins, yielding a total of 6 shear 2PCFs. The measurements from the T17 maps are shown by the black dots in figure 2.

In order to measure $\zeta_{\pm,ijk}(\alpha)$, we use the `Treecorr` code to compute the position-dependent shear 2PCF and 1-point aperture masses within patches of the footprint; we assume the same size $\theta_{\rm ap}$ and $\theta_{\rm T}$ for the aperture mass and position-dependent 2PCF. The 2PCF in each patch is measured in 15 log-spaced angular bins between 5 and $2\theta_{\rm T} - 10$ arcmin, and the 1-point aperture mass is evaluated using eq. (2.3) with the integral up to $5\theta_{\rm ap}$. The $\zeta_\pm$ is obtained by averaging the product of the shear 2PCF and 1-point aperture mass across

JCAP07(2023)040

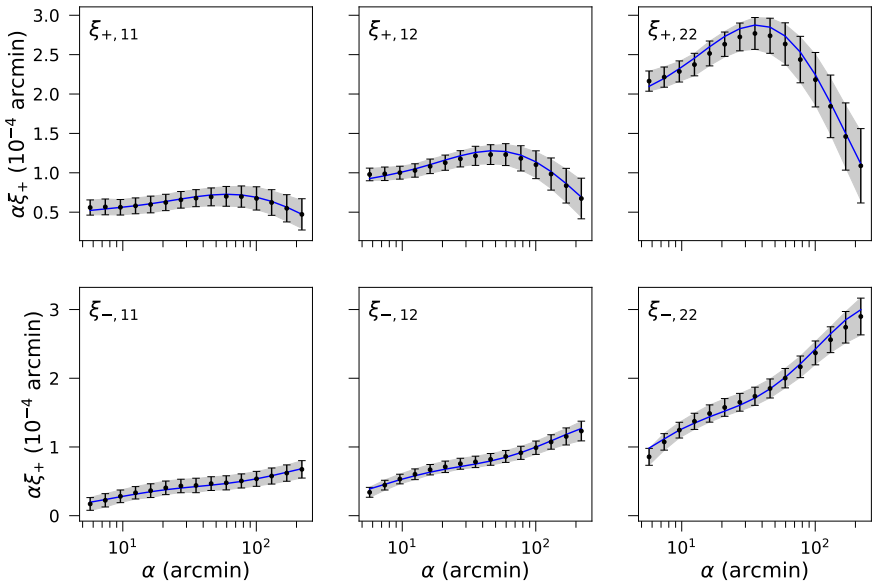

Figure 2. The shear 2PCF $\xi_\pm(\alpha)$ measured from our DES Y3-like footprints for two galaxy source redshift bins. The black dots with the error bars show the mean and the standard deviation of the measurements from the 540 T17 shear maps. For comparison, the grey shaded bands show the standard deviation computed using the 1500 FLASK shear maps. The blue curves show the theoretical result obtained using eqs. (2.13) and (2.14).

all patches selected in the footprint. For our two source redshift bins, we have 8 integrated auto- and cross-3PCF $\zeta_{\pm,ijk}(\alpha)$. The measurements from the T17 maps are shown by the black dots in figure 3 for an aperture size of $\theta_{\rm ap} = \theta_{\rm T} = 90$ arcmin.[3]

We estimate the covariance matrix of our data vectors as

$$\hat{C} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} (\hat{d}_i - \hat{d})(\hat{d}_i - \hat{d})^{\rm T} \,, \tag{3.2}$$

where N_s is the number of footprint realizations (540 for T17 and 1500 for FLASK), $\hat{d}_i$ is the data vector of the i-th realization and $\hat{d}$ is the mean data vector across all realizations. When evaluating the inverse covariance matrix, we correct it as

$$\hat{C}^{-1} = \left[\frac{N_{\rm s} - N_{\rm d} - 2}{N_{\rm s} - 1}\right][1 + A + B(N_{\rm p} + 1)]C^{-1} \,, \tag{3.3}$$

where

$$A = \frac{2}{(N_{\rm s} - N_{\rm d} - 1)(N_{\rm s} - N_{\rm d} - 4)} \,, \tag{3.4}$$

$$B = \frac{N_{\rm s} - N_{\rm d} - 2}{(N_{\rm s} - N_{\rm d} - 1)(N_{\rm s} - N_{\rm d} - 4)} \,, \tag{3.5}$$

[3] As a technical point, in our measurements of $\zeta_\pm$ we consider only survey patches where the fraction of unmasked pixels is larger than 80% for the top-hat filter W and larger than 70% for the Q filter up to $5\theta_{\rm ap}$ of aperture radius. Holes and masked pixels inside the footprint contribute to the counting of these fractions, in addition to pixels outside the survey footprint. This ensures our measurements are not affected by too many unmasked pixels in the patches, as confirmed by their excellent agreement with the theory predictions for both $\xi_\pm$ and $\zeta_\pm$ in figures 2 and 3, respectively.

JCAP07(2023)040

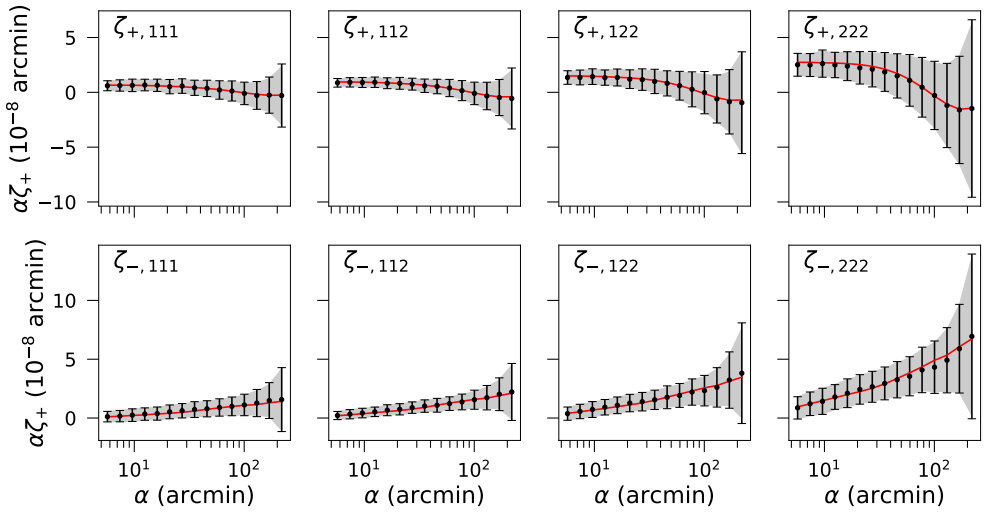

Figure 3. The integrated shear 3PCF $\zeta_\pm(\alpha)$ measured from our DES Y3-like footprints for two galaxy source redshift bins. The black dots with the error bars show the mean and the standard deviation of the measurements from the 540 T17 shear maps for a filter size of 90 arcmin. The grey shaded bands show the standard deviation computed using the 1500 FLASK shear maps. The red curves show the theoretical result from eqs. (2.9) and (2.10).

and N_d is the size of the data vector ($N_\mathrm{d} = 90$ for $\xi_\pm$, $N_\mathrm{d} = 120$ for $\zeta_\pm$, and $N_\mathrm{d} = 210$ for their combination), N_p is the number of inference parameters and C^{-1} is the directly inverted covariance. The first term in brackets is the bias correction on the inverse covariance from ref. [63], while the second term is a correction factor from ref. [64].

The FLASK covariance matrix has the advantage of having less numerical noise because of the larger N_s, but the disadvantage of corresponding to lognormal realizations of cosmic shear maps, which are not as realistic as the T17 ones from N-body simulations. The left panel of figure 4 compares the correlation matrix $r_{mn} = \hat{C}_{mn}/\sqrt{\hat{C}_{mm}\hat{C}_{nn}}$ from the FLASK (upper triangle) and T17 (lower triangle) maps; the indices m, n run over the data vector entries. Reassuringly, the two covariance matrices display broadly the same correlations. There are however some differences that are better seen in the right panel of figure 4 which shows the relative difference between the two covariances. We investigate the impact of these differences in the parameter constraints when we discuss our results below.

We note that both our covariance matrices do not appropriately account for super-sample covariance (SSC) [65, 66], i.e. the variance induced by the gravitational coupling between observed modes inside the survey and unobserved modes with wavelengths larger than the survey size. The SSC is the dominant off-diagonal contribution in 2-point function analyses [67], and it is expected to be a smaller contribution to the squeezed bispectrum configurations that dominate the small-scale $\zeta_\pm$ [68]. Our quoted error bars for $\xi_\pm$-only analyses are thus expected to be underestimated, and consequently, our quoted improvements from $\zeta_\pm$ are conservative; i.e. the relative improvement from $\zeta_\pm$ is expected to be larger in analyses that appropriately account for SSC. We defer the inclusion of SSC to future work.

JCAP07(2023)040

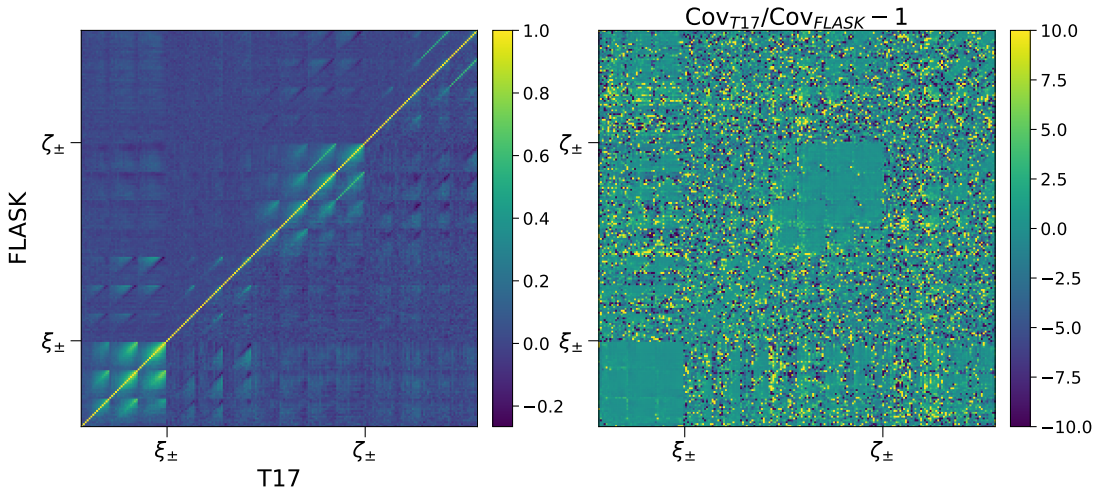

Figure 4. Comparison between the T17 and FLASK covariance matrices. The left panel shows the correlation coefficient from FLASK in the upper triangle and T17 in the lower triangle part of the matrix. The right panel shows the relative difference between the two covariance estimates (the color coding is limited to ±10 to exclude a few extreme values for visibility). The ordering of the matrix entries is according to: $\{\xi_{+,11}, \xi_{+,12}, \xi_{+,22}, \xi_{-,11}, \xi_{-,12}, \xi_{-,22}, \zeta_{+,111}, \zeta_{+,112}, \zeta_{+,122}, \zeta_{+,222}, \zeta_{-,111}, \zeta_{-,112}, \zeta_{-,122}, \zeta_{-,222}\}$. The result shown for $\zeta_\pm$ is estimated using 90 arcmin apertures.

4 Emulators for $\xi_\pm$ and $\zeta_\pm$

The evaluation of the integrated lensing bispectrum $\mathcal{B}^{\rm 2D}_{\pm,ijk}(\boldsymbol{\ell})$ is the key computational bottleneck when evaluating $\zeta_\pm$ using eqs. (2.9) and (2.10), and thus the quantity that we wish to emulate. However, rather than emulating $\mathcal{B}^{\rm 2D}_{\pm,ijk}(\boldsymbol{\ell})$ directly, we emulate only the part of the integrand in eq. (2.11) given by

$$\int \frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{\mathrm{d}^2\boldsymbol{\ell}_2}{(2\pi)^2} B_\delta^{\rm 3D}\left(\frac{\boldsymbol{\ell}_1}{\chi}, \frac{\boldsymbol{\ell}_2}{\chi}, \frac{-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2}{\chi}, \chi\right) \times e^{2i(\phi_2 \mp \phi_{-1-2})} U(\boldsymbol{\ell}_1) W(\boldsymbol{\ell}_2+\boldsymbol{\ell}) W(-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2-\boldsymbol{\ell}). \tag{4.1}$$

This leaves out the part involving the line-of-sight integration in eq. (2.11), but has the advantage of allowing for more flexibility to adjust the source redshift distributions, including bypassing the need to emulate any of the systematic parameters mentioned in section 2.3. The training of the emulator still needs to be redone for different sizes of the U and W filters. The direct evaluation of $\xi_\pm$ in an MCMC exploration of the parameter space would not impose a serious computational burden, but we emulate its calculation anyway for extra speed. In this case we emulate simply the three-dimensional matter power spectrum P_δ^{3D} in eq. (2.15).

We build our emulator by training a neural network (NN) on a Latin hypercube with 10^5 training nodes. The emulated parameters comprise the cosmological parameters $\{\Omega_{\rm m}, A_s, w_0\}$, the baryonic feedback parameter $c_{\rm min}$, as well as the redshift z which we need to emulate to perform the line-of-sight integrations in eqs. (2.11) and (2.15). The ranges of the cosmological and baryonic parameters are listed in table 1 (note we rescale A_s to $\ln(10^{10}A_s)$), and for redshift we consider $z \in [0, 2.1]$. The NN architecture is that of the `Cosmopower` code [69],[4]

[4] https://alessiospuriomancini.github.io/cosmopower/.

JCAP07(2023)040

	Prior range
Cosmological parameters (emulated)	
$\Omega_{\rm m}$	$U\,[0.16, 0.45]$
$\ln(10^{10}A_s)$	$U\,[1.61, 4.20]$
w_0	$U\,[-3.33, -0.33]$
Baryonic feedback parameter (emulated)	
$c_{\rm min}$	$U\,[1.0, 5.5]$
Systematic parameters (not emulated)	
Δz_1	$\mathcal{N}(0.0, 0.023)$
Δz_2	$\mathcal{N}(0.0, 0.020)$
m_1	$\mathcal{N}(0.0261, 0.012)$
m_2	$\mathcal{N}(-0.061, 0.011)$
$A_{\rm IA,0}$	$U\,[-5.0, 5.0]$
$\alpha_{\rm IA}$	0 (fixed)

Table 1. Model parameters considered in this paper. The parameters that enter our NN emulator are the cosmological parameters $\Omega_{\rm m}$, $\ln(10^{10}A_s)$, w_0, and the baryonic feedback parameter $c_{\rm min}$. The photo-z, shear calibration and IA systematic parameters do not need to be emulated because the predictions for different values are fast to obtain. In our MCMC analyses we vary these parameters within the listed uniform prior ranges (U) or assuming Gaussian priors $\mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ. The listed priors for the systematic parameters are inspired by those assumed in the DES Y3 analyses [54, 62].

which was originally developed to emulate 2-point statistics, but which can be straightforwardly applied to emulate eq. (4.1). The input layers of the NN are the cosmological, baryonic and redshift parameters. For $\zeta_\pm$, the output of the NN is the quantity in eq. (4.1) in 100 log-spaced ℓ bins between $\ell = 2$ and $\ell = 15000$. In the training set, the supervised learning labels are the same quantity obtained by directly evaluating eq. (4.1) using Monte-Carlo integration. For $\xi_\pm$ the output is the three-dimensional matter power spectrum in 100 log-spaced ℓ bins as in the right-hand side of eq. (2.15) between $\ell = 2$ and $\ell = 15000$.

We test the emulators using another Latin hypercube with 10^3 test nodes with the same prior ranges of the training set. We quantify the performance of the emulator with the expression

$$\epsilon \equiv \left| \frac{\chi^2_{{\rm emu},i}}{\chi^2_{{\rm test},i}} - 1 \right| , \tag{4.2}$$

where $\chi^2_{{\rm emu},i}$ is the χ^2 value associated with the ith test node, defined w.r.t. the data vector $\hat{d}_{\rm T17}$ generated by the theory model at T17 cosmological parameters. Concretely, $\chi^2_{{\rm emu},i} = \left(\hat{d}_{{\rm emu},i} - \hat{d}_{\rm T17}\right)^t \hat{C}^{-1} \left(\hat{d}_{{\rm emu},i} - \hat{d}_{\rm T17}\right)$, with $\hat{d}_{{\rm emu},i}$ the emulator prediction and $\hat{C}^{-1}$ the T17 inverse covariance matrix. The quantity $\chi^2_{{\rm test},i}$ is defined analogously, but replacing the emulator result at each test node with the test label prediction. The ϵ metric describes how similar the emulator would behave to the theory model in likelihood analyses. The smaller the value of ϵ, the better the accuracy of the emulator.

Figures 5 and 6 show the outcome of this test for $\xi_\pm$ and $\zeta_\pm$, respectively. We show ϵ projected only on the $\Omega_{\rm m}$ — w_0 plane, but the takeaways are common to other projections. For $\xi_\pm$, effectively all of the test nodes have χ^2 relative differences $\epsilon < 0.05$. The performance

JCAP07(2023)040

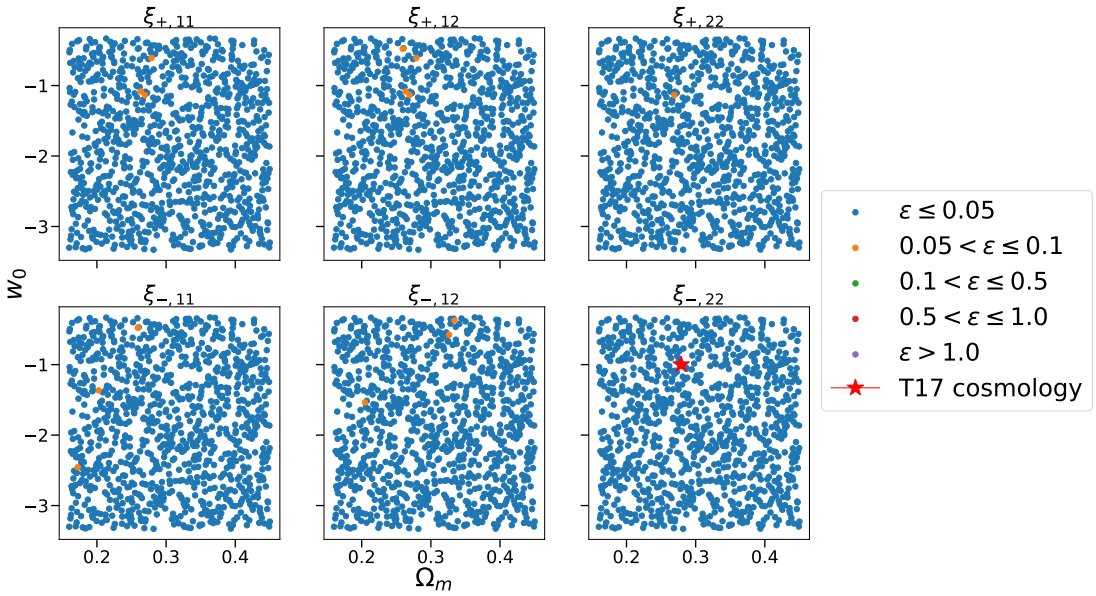

Figure 5. Performance of the $\xi_\pm$ emulator on 10^3 test nodes projected on the Ω_m — w_0 plane. The colors show the absolute value of the χ^2 relative difference ϵ defined in eq. (4.2); if $\epsilon < 0.05$, this means the emulator describes the χ^2 w.r.t. the T17 cosmology (marked by the red star) to better than 5%.

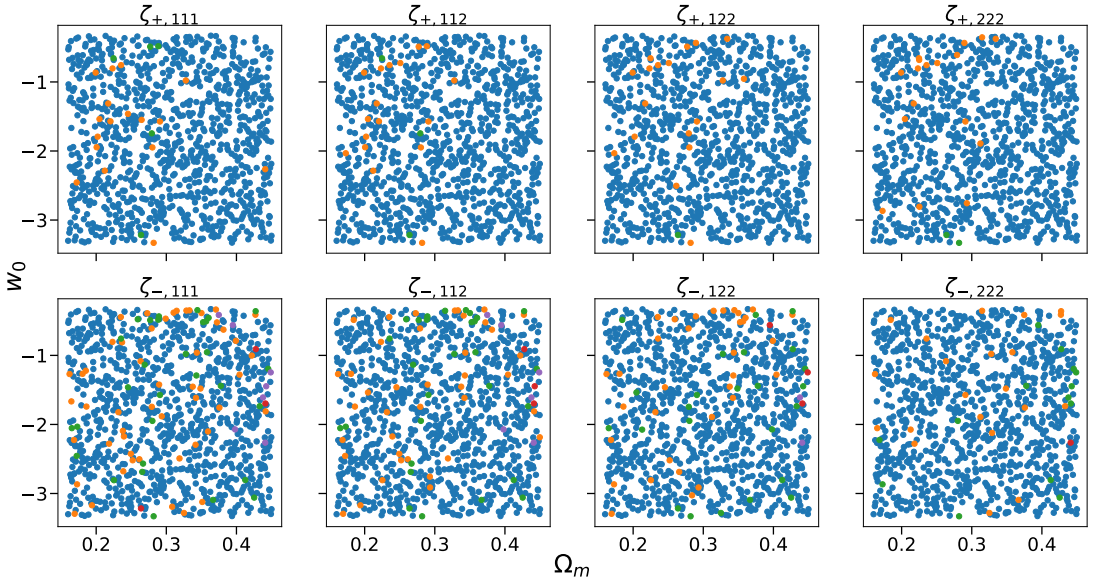

Figure 6. Same as figure 5, but for the performance of the $\zeta_\pm$ emulator, instead of $\xi_\pm$. The result is for 90 arcmin apertures. The color coding is the same as in figure 5.

gets reduced slightly for $\zeta_\pm$ with 92% (95%) of the test nodes having $\epsilon < 0.05$ ($\epsilon < 0.1$); the result in figure 6 is for apertures with 90 arcmin, but we have checked the performance is equivalent for other apertures as well. If the true χ^2 value of some point in parameter space is $\chi^2_\mathrm{test} = 1$, then $\epsilon < 0.1$ implies $\chi^2_\mathrm{emu} \in [0.9, 1.1]$. Effectively all of the test nodes for both $\xi_\pm$ and $\zeta_\pm$ satisfy this satisfactory criterion.

JCAP07(2023)040

5 Results: simulated likelihood analyses with MCMC

In this section we present our main numerical results from simulated likelihood analyses with MCMC. Unless otherwise specified, we consider the parameter priors listed in table 1, and sample the parameter space assuming a Gaussian likelihood function,

$$\mathcal{L}(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}\left(\mu(\boldsymbol{\theta}) - \hat{d}\right)^t C^{-1}\left(\mu(\boldsymbol{\theta}) - \hat{d}\right)\right], \tag{5.1}$$

where $\hat{d}$ is the assumed data vector, C the covariance matrix and $\mu(\boldsymbol{\theta})$ the theory prediction for model parameters $\boldsymbol{\theta}$. We utilize the sampler code `affine`[5] based on `tensorflow`. With the available NVIDIA A100 GPU (Graphics Processing Unit) hardware, emulator and sampler, we are able to sample an order of 10^6 points in an hour's timescale.

Next, we validate our model using the T17 $\xi_\pm$ and $\zeta_\pm$ data vectors in section 5.1, investigate the impact of the aperture size in $\zeta_\pm$ constraints in section 5.2, discuss the impact of the systematic parameters in section 5.3, and check the impact from using the T17 or FLASK covariance matrices in section 5.4. All of the marginalized two-dimensional constraints shown throughout display contours with the 1σ and 2σ confidence regions.

5.1 Validation on the T17 cosmic shear maps

Figure 7 shows the constraints on the cosmological and baryonic feedback parameters for the data vector from the T17 shear maps (cf. black points in figures 2 and 3) and the FLASK covariance matrix. The result is for $\zeta_\pm$ measured using 90 arcmin apertures. We keep the systematic parameters fixed to zero in these constraints, which is the case for our T17 maps. In addition to the correction factors in eq. (3.3), in this section we consider also the factor $[1 + B(N_\mathrm{d} - N_\mathrm{p})]^{-1}$ from ref. [70] due to statistical noise in our covariance matrix estimate.

The key takeaway from figure 7 is that our theory model and emulator recover unbiased constraints: the T17 parameters (dashed black lines) are contained well within the 1σ confidence levels for both the $\xi_\pm$-only (green) and $\xi_\pm + \zeta_\pm$ constraints (red). The ability of our theory model to recover unbiased cosmological constraints could have already been anticipated from the good agreement between theory and simulations in figures 2 and 3.

As a test, we have repeated the analysis in figure 7 but adopting the t-distribution likelihood function from ref. [71], instead of a Gaussian likelihood. The result (not shown) is practically indistinguishable from that in figure 7 for both $\xi_\pm$ and $\xi_\pm + \zeta_{\pm,\{90'\}}$, suggesting the exact choice of the likelihood function does not critically affect our results.

5.2 The impact of the aperture size

When measuring $\zeta_\pm$ one of the decisions concerns the choice of the apertures on which to measure the 1-point shear aperture mass and local $\xi_\pm$. To investigate the impact of this, we perform likelihood analyses with a noiseless data vector generated with the theory model using the T17 parameters. In these tests, we use the FLASK covariance, and vary also the systematic parameters with the priors listed in table 1. The main result is shown in table 2, which lists the relative improvement of the combined $\xi_\pm + \zeta_\pm$ constraints relative to $\xi_\pm$-only, for different aperture sizes and combinations. Figure 8 shows the actual parameter constraints for two aperture choices: a single aperture with 90 arcmin (blue) and the combination of five apertures with sizes $\{50, 70, 90, 110, 130\}$ arcmin (red).

[5] https://github.com/justinalsing/affine.

JCAP07(2023)040

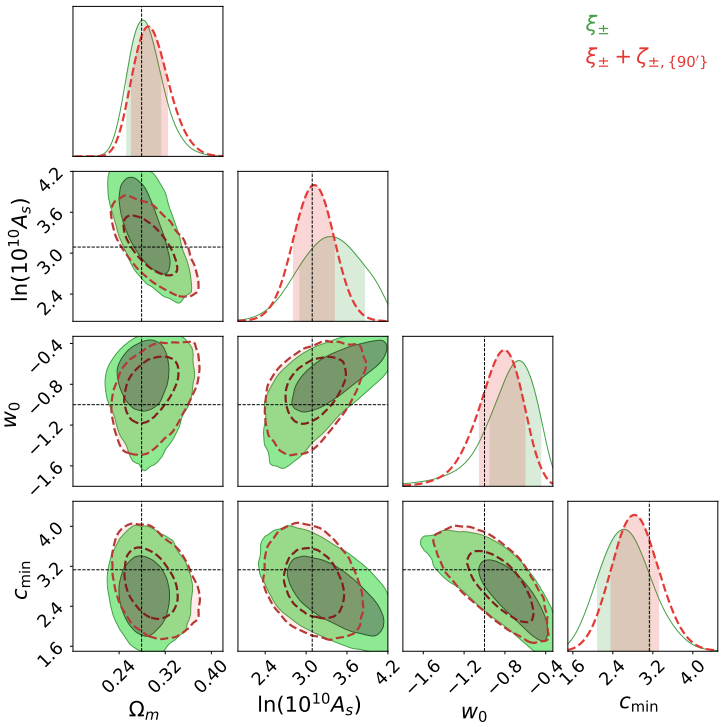

Figure 7. Parameter constraints obtained with the T17 data vector and FLASK covariance matrix. The constraints in green and red are for the $\xi_\pm$-only and $\xi_\pm + \zeta_\pm$ data vectors, respectively. The black dashed lines mark the T17 parameters.

Aperture sizes (arcmin)	$\Omega_{\rm m}$	$\ln(10^{10}A_s)$	w_0	$c_{\rm min}$
50	1.2%	9.0%	18.1%	4.8%
70	1.2%	16.9%	31.9%	11.6%
90	**3.7%**	**20.2%**	**38.4%**	**15.1%**
110	1.2%	19.1%	34.1%	11.0%
130	1.2%	16.9%	32.6%	12.3%
{50, 70, 90}	2.5%	24.7%	39.1%	15.8%
{50, 90, 130}	3.7%	23.6%	41.3%	16.4%
{70, 90, 110}	6.2%	25.8%	39.1%	15.1%
{90, 110, 130}	8.6%	25.9%	42.8%	15.8%
{50, 70, 90, 110, 130}	**12.4%**	**28.1%**	**44.9%**	**19.9%**

Table 2. Relative improvement of combined $\xi_\pm + \zeta_\pm$ constraints relative to $\xi_\pm$-only for different values and combinations of the aperture sizes. The best single- and combined-filter cases are highlighted in bold. The result is for a noiseless data vector from the theory model, the FLASK covariance, and marginalizing over the systematic parameters.

JCAP07(2023)040

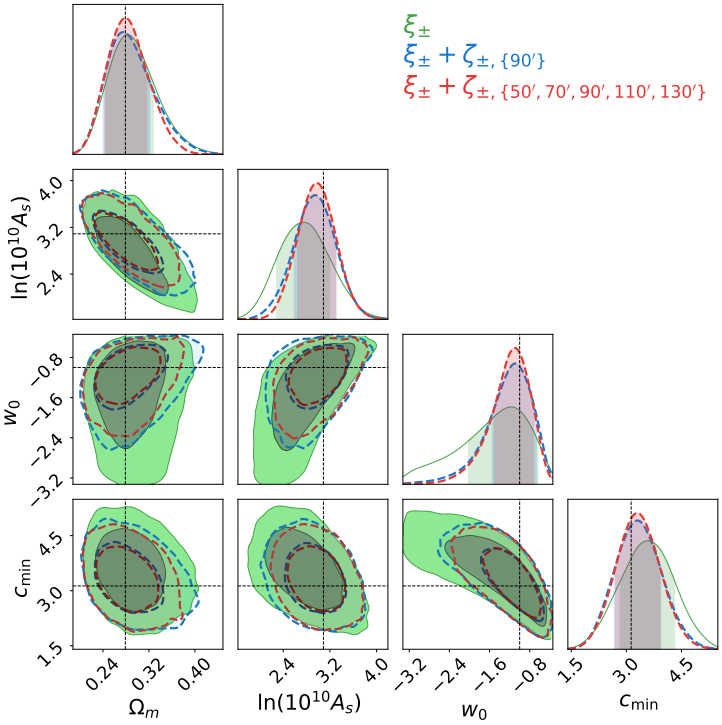

Figure 8. Impact of the aperture size in $\xi_\pm + \zeta_\pm$ constraints. The contours in green are for $\xi_\pm$-only constraints. The contours in blue are for $\xi_\pm + \zeta_\pm$ constraints using a single aperture with size 90 arcmin, and in red for the combination of five filter sizes $\{50, 70, 90, 110, 130\}$ arcmin. The result is for a noiseless data vector from the theory model with the T17 parameters (dashed lines), the FLASK covariance, and marginalizing over the systematic parameters.

Regarding the single aperture cases, table 2 shows that the constraints improve first from 50 to 90 arcmin, but then degrade from 90 to 130 arcmin. This follows from the combination of the following effects. Smaller apertures have the advantage of providing $\zeta_\pm$ with higher signal-to-noise ratio since there are more apertures over which the average of eq. (2.1) can be taken. They have, however, the disadvantage that the local $\xi_\pm$ is measured over a more reduced range of angular scales inside each patch. Conversely, bigger apertures allow to probe the local $\xi_\pm$ on larger scales, but at the price of less signal-to-noise as one averages over a smaller number of patches on the sky.[6] In general, different aperture sizes are sensitive to different configurations of the small-scale squeezed-limit bispectrum [38], which can contain varying cosmological information and impact the final parameter constraints.

For the aperture sizes shown, the balance between these effects is optimal for apertures with 90 arcmin, which gives the best constraints. Concretely, the addition of $\zeta_\pm$ to the constraints leads to improvements of 4% for $\Omega_{\rm m}$, 20% for $\ln(10^{10}A_s)$, 38% for w_0 and 15% for $c_{\rm min}$. These figures are in line with the previous findings of refs. [32, 38] based on idealized Fisher-matrix forecasts, but extended here to more realistic MCMC-based analyses.

The $\zeta_\pm$ measured over slightly different aperture sizes are expected to be substantially correlated due to the large overlap of the regions where the local $\xi_\pm$ is measured. However, the

[6] In particular, in the limit of very large apertures, the $\xi_\pm$ measured in the patches become almost perfectly correlated with the $\xi_\pm$ of the whole survey, effectively contributing with no independent information.

JCAP07(2023)040

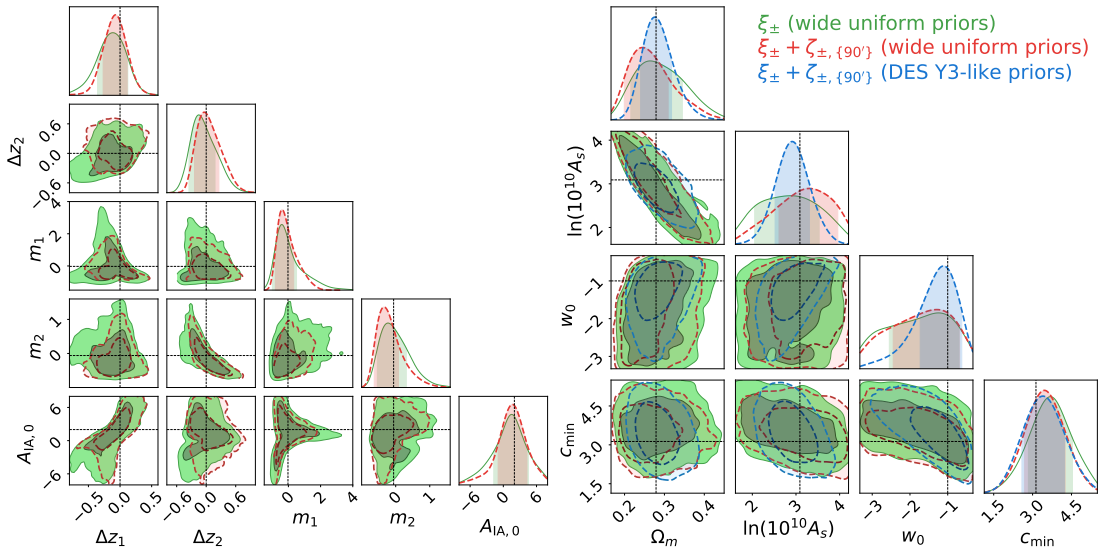

Figure 9. Parameter constraints for different priors on the systematic parameters. The green and red contours are for $\xi_\pm$ and $\xi_\pm + \zeta_\pm$ assuming wide uniform priors on the systematic parameters. The blue contours are for $\xi_\pm + \zeta_\pm$ with DES Y3-like Gaussian priors on the systematic parameters. The left panel shows the systematic parameter constraints, and the right panel shows the constraints on the cosmological and baryonic feedback parameters. The left panels do not show the contours with DES Y3-like priors as they are too small to be clearly seen. The result is for a noiseless data vector drawn from the theory model with the T17 parameters (except we set $A_{\rm IA,0} = 2$; cf. dashed lines), and using the FLASK covariance.

lower part of table 2 shows that there is still enough independent information to improve the constraints further by combining different apertures. For the cases shown, the best constraints are obtained when combining all apertures $\{50, 70, 90, 110, 130\}$ arcmin: the improvements become 12% for $\Omega_{\rm m}$, 28% for $\ln(10^{10}A_s)$, 45% for w_0 and 20% for $c_{\rm min}$. These improvements need however to be contrasted with the complications that they add to the analyses. For example, this comes with the price of a much larger data vector, which puts pressure on the numerical requirements for reliable covariance estimates from simulations. In this paper, this pressure was still manageable for a DES Y3-like survey with two tomographic bins, but future survey analysis settings will have larger areas and more source redshift bins as well. The decision of how many filters to combine should thus be made case by case.

5.3 The impact of systematics and their modelling

We turn our attention now to the impact of systematics (photo-z, shear calibration and IA) in $\zeta_\pm$ constraints. This is interesting as $\xi_\pm$ and $\zeta_\pm$ depend differently on systematics, and so combined analyses can potentially mitigate the degradation caused by these additional free parameters, leading to better cosmological constraints [73–75]. Indeed, this has been studied recently in ref. [56], where it was shown that combining lensing 2- and 3-point correlation function information in a survey like Euclid could lead even to the self-calibration of the systematic parameters to levels that reduce the need for external calibration data sets.

The green and red contours in figure 9 show the constraints for $\xi_\pm$ and $\xi_\pm + \zeta_\pm$, but instead of the tight DES Y3-like priors that we have assumed so far for the systematic parameters

JCAP07(2023)040

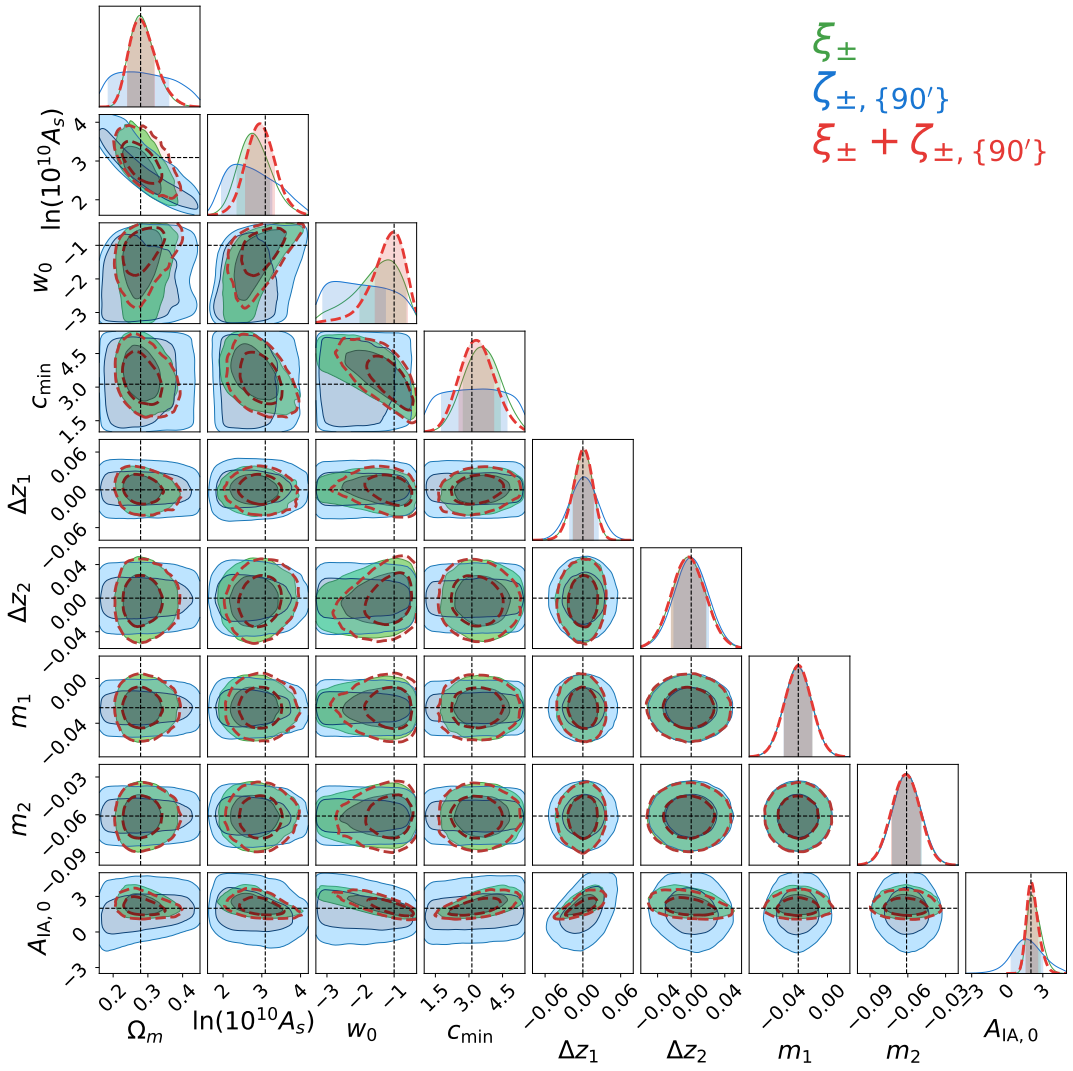

Figure 10. Parameter constraints obtained with our default NLA IA modelling, but on a data vector generated with the NLA IA model used in refs. [54, 56, 72]. The result is for the T17 parameters (except we set $A_{\mathrm{IA},0} = 2$; cf. dashed lines) and the FLASK covariance. The green, blue and red contours are for $\xi_\pm$, $\zeta_\pm$ and $\xi_\pm + \zeta_\pm$ constraints. We use DES Y3-like Gaussian priors for the systematic parameters.

(cf. table 1), we assume now wide priors for them. The result is for a noiseless realization of the data vector for the T17 parameters, with the exception that we set $A_{\mathrm{IA},0} = 2$ in this subsection. The improvements on the cosmological and baryonic parameters from adding $\zeta_\pm$ are 15.4% for Ω_m, 8.8% for $\ln(10^{10}A_s)$, 4.9% for w_0 and 8.8% for c_min. Compared to the case where we marginalize over tight DES Y3-like Gaussian priors, varying the systematic parameters over wide priors degrades the improvement by factors of 2.3, 7.8 and 1.7 for $\ln(10^{10}A_s)$, w_0 and c_min respectively. Furthermore, contrary to the case in ref. [56], the improvements that still exist do not appear to be associated with a significant self-calibration of the systematic parameters. This can be seen also on the left of figure 9, where the constraints on the systematic parameters in the combined $\xi_\pm + \zeta_\pm$ case (red) show improvements of 21% for

JCAP07(2023)040

Δz_1, 8% for Δz_2, 24% for m_1, 17% for m_2 and 18% for $A_{\rm IA,0}$. There is indeed a visible level of systematics self-calibration from combining $\xi_\pm$ with $\zeta_\pm$, but which still yields constraints that are substantially larger than using the externally calibrated DES Y3-like priors (blue).

The quantitative differences to the analysis of ref. [56] could be at least partly due to some of the following reasons. First, ref. [56] considers 3-point correlation function information by taking the equilateral lensing bispectrum as the data, whereas we consider $\zeta_\pm$ that probes predominantly the squeezed lensing bispectrum [32, 38]. Second, ref. [56] considers a treatment of the NLA IA model that is not the same as ours (cf. appendix A). Further, the results of ref. [56] are based on Fisher matrix analyses, whereas ours are for simulated likelihood analyses with MCMC sampling. This can be especially important given how strongly non-Gaussian the marginalized posteriors of the systematic parameters are on the left of figure 9. Finally, our analysis is for a DES Y3-like survey assuming two tomographic bins, whereas ref. [56] considers a larger Euclid-like survey with five tomographic bins, and thus a higher-dimensional subspace of systematic parameters. A deep investigation of the origin of the differences between the results of the two works would be interesting to pursue, but that is beyond the scope of the present paper.

We investigate also potential biases in the constraints of the $A_{\rm IA,0}$ parameter from assuming different IA models in shear 3-point correlation function analyses. In particular, we wish to contrast the NLA model used in this paper (cf. section 2.3 and appendix A) with that in ref. [54] which comes from refs. [56, 72]. To do so we generate a noiseless data vector with the T17 parameters and $A_{\rm IA,0} = 2$ *assuming the IA parametrization of ref. [54]*, which we subsequently analyse by running MCMC constraints *assuming our IA modelling strategy*. At the $\xi_\pm$ level, the two IA treatments are equivalent, but there are differences at the level of the 3-point correlation functions (cf. appendix A).[7] Figure 10 shows the corresponding constraints for $\xi_\pm$ (green), $\zeta_\pm$ (blue) and $\xi_\pm + \zeta_\pm$ (red), with all yielding unbiased constraints, including $A_{\rm IA,0}$. That is, at the level of the constraining power of our DES Y3-like setup, the differences between the two NLA IA models do not have any significant impact. We note, however, that whether the same conclusion holds for other survey setups should be checked on a case-by-case basis.

5.4 The impact of different covariance estimates

We compare in figure 11 the parameter constraints obtained with the T17 covariance matrix (left) with those obtained using FLASK (right). In order to make a fair comparison, in this subsection we constructed a new FLASK covariance with the same number of footprint realizations as T17 ($N_s = 540$), and with a source redshift distribution matching the discretized one of the T17 simulations described in section 3.1. The result in figure 11 is for a noiseless realization of the data vector from the theory model at the T17 parameters, and with the systematic parameters marginalized with the DES Y3-like Gaussian priors. Table 3 lists the corresponding improvements from adding $\zeta_\pm$ information to the constraints.

The two covariance matrices yield effectively the same parameter posteriors for $\xi_\pm$-only constraints; cf. similarity between the green contours on the left and right of figure 11. There are however some differences in the combined $\xi_\pm + \zeta_\pm$ constraints shown in red, with the FLASK covariance yielding smaller parameter error bars for most parameters. In particular, the improvements from $\zeta_\pm$ can be factors of $\approx 1.2 - 1.9$ larger with the FLASK covariance compared to T17.

[7] Among other, the model of ref. [54] includes terms $\propto A^4_{\rm IA,0}$, whereas ours stops at third order $\propto A^3_{\rm IA,0}$, as expected for a three-point correlation function.

JCAP07(2023)040

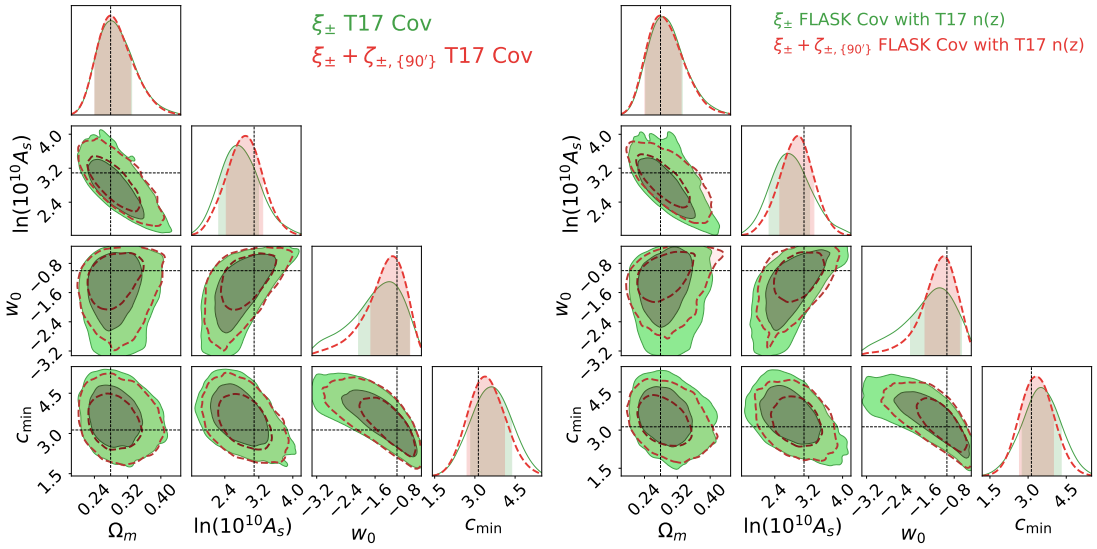

Figure 11. Impact of the covariance matrix estimate on the parameter constraints. The left and right panels show the result for the covariance estimated from the T17 and FLASK shear maps, respectively. In both panels, the result is for a noiseless realization of the data vector from the theory model at the T17 parameters (dashed lines); the green and red contours are for $\xi_\pm$-only and $\xi_\pm + \zeta_\pm$, respectively. The result is for apertures with sizes 90 arcmin and systematics marginalized with the DES Y3-like Gaussian priors.

Covariance type	$\Omega_{\rm m}$	$\ln(10^{10}A_s)$	w_0	$c_{\rm min}$
FLASK (lognormal)	1.1%	16.7%	32.1%	12.4%
T17 (N-body simulations)	3.5%	8.8%	26.1%	8.7%

Table 3. Impact of the covariance matrix estimate on the improvement of $\xi_\pm + \zeta_\pm$ constraints, relative to $\xi_\pm$-only. The result is for the aperture with size 90 arcmin.

The T17 and FLASK covariances in this subsection are estimated from ensembles of 540 shear maps and so have the same noise level. These differences may indicate they are intrinsic to the different ability of N-body simulations and lognormal realizations to capture the covariance of $\zeta_\pm$,[8] or due to residual statistical fluctuations for $N_s = 540$. We leave a more detailed investigation of the impact of the covariance matrix, including covariances calculated analytically [13, 68], to future work.

6 Summary & conclusion

The integrated shear 3PCF $\zeta_\pm$ [32, 38] is a higher-order cosmic shear statistic that measures the correlation between the shear 2PCF measured in patches of the sky and the shear aperture mass in the same patches (cf. eq. (2.1)). On small scales, $\zeta_\pm$ probes primarily the cosmological

[8] The covariance of a 3-point function contains terms up to the 6-point function, which are not as faithfully captured in lognormal realizations, compared to N-body simulations.

JCAP07(2023)040

information encoded in the squeezed-limit lensing bispectrum. Two of the key advantages of $\zeta_\pm$ compared to other higher-order cosmic shear statistics are that (i) it can be straightforwardly evaluated from the data using efficient and well-tested 2-point correlation function estimators (i.e. it does not explicitly require dedicated and more expensive 3-point estimators) and (ii) it admits a theoretical model based on the response approach to perturbation theory [38] that is accurate in the nonlinear regime of structure formation, allowing to reliably account for the impact of baryonic physics.

In this paper, we developed an analysis pipeline that can be directly applied to real cosmic shear data to obtain cosmological constraints from $\zeta_\pm$ and its combination with $\xi_\pm$. Compared to previous works on $\zeta_\pm$, the main significant advances in this paper are (i) the incorporation of lensing systematics associated with photo-z uncertainties, shear calibration biases and galaxy IA (cf. section 2.3), and (ii) the development of a NN-based emulator for fast theory predictions to enable MCMC parameter inference. We tested our pipeline on a set of realistic cosmic shear maps based on N-body simulations, with DES Y3-like survey footprint, mask and source redshift distributions (cf. section 3).

In our tests of the analysis pipeline we have investigated in particular (i) the accuracy of the theory model (cf. section 5.1), (ii) the impact of the size of the apertures used to measure $\zeta_\pm$ (cf. section 5.2), (iii) the impact of lensing systematics (cf. section 5.3) and (iv) the impact of N-body simulation vs. lognormal estimates of the data vector covariance matrix (cf. section 5.4). Our main findings can be summarized as follows:

- Our analysis pipeline is accurate (cf. figures 2 and 3) and able to yield unbiased parameter constraints from our N-body simulation DES Y3-like data vectors (cf. figure 7).

- For the range of aperture sizes $\{50, 70, 90, 110, 130\}$ arcmin, 90 arcmin is what results in the largest information gain from $\zeta_\pm$. The combination of several filter sizes can improve the constraints further (cf. table 2), but at the cost of dealing with a larger data vector and covariance matrix.

- Although $\xi_\pm$ and $\zeta_\pm$ depend differently on the systematic parameters, we do not find significant improvements in their constraints in combined $\xi_\pm + \zeta_\pm$ analyses; i.e. the mitigation of systematic effects still requires prior calibration from external data (cf. figure 9). This is in contrast with the findings in ref. [56], although this may be due to differences in the 3-point correlation function studied, survey setup and other analysis details. At the level of the DES Y3 constraining power, different modelling strategies for IA lead also to no significant biases in parameter constraints (cf. figure 10).

- Relative to $\xi_\pm$-only constraints with the N-body covariance matrix, adding $\zeta_\pm$ leads to improvements of 4% for Ω_m, 9% for $\ln\left(10^{10}A_s\right)$, 26% for w_0 and 9% for c_min. Except for Ω_m, these are factors of $\approx 1.2 - 1.9$ smaller compared to the FLASK covariance. This may be due to residual statistical fluctuations at the level of our number of simulation realizations ($N_s = 540$), or simply that lognormal realizations do not provide reliable estimates of the $\zeta_\pm$ covariance matrix.

Overall, our results corroborate with a realistic MCMC-based simulated likelihood analysis the encouraging findings from previous idealized Fisher matrix forecasts [32, 38]. The analysis pipeline developed and tested here can be readily applied to real survey data, enabling the exploration of the potential of the integrated shear 3PCF $\zeta_\pm$ to improve cosmological parameter constraints using cosmic shear observations.

JCAP07(2023)040

Acknowledgments

We would like to thank Pierre Burger, Juan M. Cruz-Martinez, Chris Davies, Mariia Gladkova, Eiichiro Komatsu, Elisabeth Krause and Alessio Spurio-Mancini for very helpful comments and discussions at various stages of this project. We acknowledge support from the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2094-390783311. Most of the numerical calculations have been carried out on the ORIGINS computing facilities of the Computational Center for Particle and Astrophysics (C2PAP). We would like to particularly thank Anthony Hartin for the support in accessing these computing facilities. The results in this paper have been derived using the following publicly available libraries and software packages: `healpy` [76], `Treecorr` [61], `CLASS` [46], `FLASK` [60], `Vegas` [77], `Cosmopower` [69], `GPflow` [78] and `Numpy` [79]. We also acknowledge the use of `matplotlib` [80] and `ChainConsumer` [81] python packages in producing the figures shown in this paper.

Data availability. The numerical data underlying the analysis of this paper may be shared upon reasonable request to the authors.

A The modelling of intrinsic alignments

In this appendix we describe our modelling of galaxy intrinsic alignments in $\xi_\pm$ and $\zeta_\pm$.

General considerations

The observed galaxy ellipticity in cosmic shear observations $\epsilon_{\rm obs}$ is a combination of the gravitational (G) lensing shear component γ and the intrinsic (I) ellipticity of the galaxies $\epsilon_{\rm I}$ induced by correlations with local gravitational tidal fields at the source (in this appendix, we ignore the random stochastic component that would contribute as shape noise):

$$\epsilon^i_{\rm obs}(\boldsymbol{\theta}) = \gamma^i(\boldsymbol{\theta}) + \epsilon^i_{\rm I}(\boldsymbol{\theta}), \tag{A.1}$$

where i denotes a specific source galaxy redshift bin. The lensing shear is related to the lensing convergence κ as [2]

$$\gamma(\boldsymbol{\ell}) = e^{2i\phi_\ell}\kappa(\boldsymbol{\ell})\,; \quad \kappa(\boldsymbol{\ell}) = \int {\rm d}^2\boldsymbol{\theta}\ \kappa(\boldsymbol{\theta})e^{-i\boldsymbol{\ell}\boldsymbol{\theta}}\,; \quad \kappa(\boldsymbol{\theta}) = \int {\rm d}\chi\ q(\chi)\delta_m(\boldsymbol{\theta}\chi,\chi), \tag{A.2}$$

where δ_m is the three-dimensional matter density contrast.[9] In analogy, we can write for the intrinsic component $\epsilon_{\rm I}$

$$\epsilon_{\rm I}(\boldsymbol{\ell}) = e^{2i\phi_\ell}\kappa_{\rm I}(\boldsymbol{\ell})\,; \quad \kappa_{\rm I}(\boldsymbol{\ell}) = \int {\rm d}^2\boldsymbol{\theta}\ \kappa_{\rm I}(\boldsymbol{\theta})e^{-i\boldsymbol{\ell}\boldsymbol{\theta}}\,; \quad \kappa_{\rm I}(\boldsymbol{\theta}) = \int {\rm d}\chi\ n(\chi)\delta_{\rm I}(\boldsymbol{\theta}\chi,\chi), \tag{A.3}$$

where $\delta_{\rm I}$ is a three-dimensional field that determines *effectively* the intrinsic alignment (IA) of the galaxies with their local gravitational tidal fields; note also that the line-of-sight kernel is now just the source galaxy distribution $n(\chi)$, and not the lensing kernel $q(\chi)$.

In the popular *nonlinear linear alignment* (NLA) model [51, 52], one writes

$$\delta_{\rm I}(\boldsymbol{x}, z) = f_{\rm IA}(z)\delta_m(\boldsymbol{x}, z), \tag{A.4}$$

[9] To ease the notation, we distinguish between real- and harmonic-space variables by their arguments. For example, $\kappa(\boldsymbol{\theta})$ and $\kappa(\boldsymbol{\ell})$ are the lensing convergence in real and harmonic space, respectively.

JCAP07(2023)040

treating δ_m as the nonlinear matter density contrast. The amplitude $f_{\rm IA}(z)$ is

$$f_{\rm IA}(z) = -A_{\rm IA,0}\left(\frac{1+z}{1+z_0}\right)^{\alpha_{\rm IA}} \frac{c_1\rho_{\rm crit}\Omega_{\rm m,0}}{D(z)}\,, \tag{A.5}$$

where $A_{\rm IA,0}$, $\alpha_{\rm IA}$ are free redshift-independent parameters, $c_1 = 5\times10^{-14}\ (h^2 M_\odot/{\rm Mpc}^3)^{-1}$ [52], $\rho_{\rm crit}$ is the critical cosmic energy density, $D(z)$ is the growth factor normalized to unity today, and z_0 is some reasonable pivot redshift value.

Note that this is only an *effective* parametrization of the impact of IA in cosmic shear observations. A more rigorous approach would involve a description of the relation of galaxy shapes and tidal fields in 3D, subsequently projected to the sky plane. This is the approach described in refs. [82, 83] based on bias expansions in effective field theory, which is however valid only in the quasi-linear, large-scale regime of structure formation. Extensions of the NLA model to include nonlinear corrections to eq. (A.4) also exist [84].

Contributions to $\boldsymbol{\xi_\pm}$

The two shear 2PCF $\xi_\pm$ are given by

$$\xi^{ij}_{+,\rm obs}(\alpha) = \langle \epsilon^i_{\rm obs}(\boldsymbol{\theta})\epsilon^{j*}_{\rm obs}(\boldsymbol{\theta}+\boldsymbol{\alpha})\rangle \tag{A.6}$$

$$\xi^{ij}_{-,\rm obs}(\alpha) = \langle \epsilon^i_{\rm obs}(\boldsymbol{\theta})\epsilon^{j}_{\rm obs}(\boldsymbol{\theta}+\boldsymbol{\alpha})e^{-4i\phi_\alpha}\rangle, \tag{A.7}$$

and each can be decomposed into GG, GI, IG and II terms as

$$\xi^{ij}_{\pm,\rm obs} = \xi^{ij}_{\pm,\rm GG} + \xi^{ij}_{\pm,\rm GI} + \xi^{ij}_{\pm,\rm IG} + \xi^{ij}_{\pm,\rm II}. \tag{A.8}$$

The GI case of $\xi^{ij}_{+,\rm obs}$, for example, is given by (the derivations are analogous for all terms):

$$\begin{aligned}\xi^{ij}_{+,\rm GI}(\alpha) &= \langle \gamma^i(\boldsymbol{\theta})\epsilon^{j*}_{\rm I}(\boldsymbol{\theta}+\boldsymbol{\alpha})\rangle \\ &= \int \frac{{\rm d}\ell\ell}{2\pi} P^{ij}_{\kappa\kappa_{\rm I}}(\ell) J_0(\ell\alpha),\end{aligned} \tag{A.9}$$

where $P^{ij}_{\kappa\kappa_{\rm I}}(\ell)$ is defined as $(2\pi)^2 P^{ij}_{\kappa\kappa_{\rm I}}(\ell)\delta_D(\boldsymbol{\ell}+\boldsymbol{\ell}') = \langle \kappa^i(\boldsymbol{\ell})\kappa^j_{\rm I}(\boldsymbol{\ell}')\rangle$ and given by

$$P^{ij}_{\kappa\kappa_{\rm I}}(\ell) = \int {\rm d}\chi \frac{q^i(\chi)n^j(\chi)}{\chi^2} P^{\rm 3D}_{\delta_m\delta_{\rm I}}(\ell/\chi,\chi). \tag{A.10}$$

The $P^{\rm 3D}_{\delta_m\delta_{\rm I}}$ is defined as $(2\pi)^3 P^{\rm 3D}_{\delta_m\delta_{\rm I}}(\boldsymbol{k}_1+\boldsymbol{k}_2) = \langle \delta_m(\boldsymbol{k}_1)\delta_{\rm I}(\boldsymbol{k}_2)\rangle$, and in the NLA model it is

$$P^{\rm 3D}_{\delta_m\delta_{\rm I}}(k,z) = f_{\rm IA}(z) P^{\rm 3D}_{\delta_m\delta_m}(k,z). \tag{A.11}$$

That is, the GI contribution to $\xi^{ij}_{+,\rm obs}$ can be obtained by replacing the jth lensing kernel $q^j(\chi)$ in the expression of the GG term with $n^j(\chi)f_{\rm IA}$. It follows as a result that all contributions from GG, GI, IG and II can be obtained by replacing all lensing kernels $q(\chi)$ with $q(\chi)+n(\chi)f_{\rm IA}$, as in eq. (2.24). This yields terms $\propto f^0_{\rm IA}$ (GG), $\propto f_{\rm IA}$ (GI, IG) and $\propto f^2_{\rm IA}$ (II).

JCAP07(2023)040

Contributions to $\zeta_\pm$

The observed integrated shear 3PCF $\zeta_\pm$ is defined as

$$\zeta^{ijk}_{\pm,\rm obs}(\alpha) = \Big\langle \hat{M}^i_{\rm ap,obs}(\boldsymbol{\theta}_C)\hat{\xi}^{jk}_{\pm,\rm obs}(\alpha;\boldsymbol{\theta}_C)\Big\rangle. \tag{A.12}$$

The position-dependent shear 2PCF $\hat{\xi}^{jk}_{\pm,\rm obs}(\alpha;\boldsymbol{\theta}_C)$ also contains GG, GI, IG and II terms. Further, the IA terms also contribute to the 1-point aperture mass $\hat{M}^i_{\rm ap,obs}(\boldsymbol{\theta}_C)$, which contains G and I terms as

$$\hat{M}^i_{\rm ap,obs}(\boldsymbol{\theta}_C) = \int \mathrm{d}^2\boldsymbol{\theta}\Big[\kappa^i(\boldsymbol{\theta}) + \kappa^i_{\rm I}(\boldsymbol{\theta})\Big]U(\boldsymbol{\theta}_C - \boldsymbol{\theta}). \tag{A.13}$$

This thus generates the following 8 contributions to $\zeta^{ijk}_{\pm,\rm obs}(\alpha)$:

$$\zeta^{ijk}_{\pm,\rm obs} = \zeta^{ijk}_{\pm,\rm GGG} + \zeta^{ijk}_{\pm,\rm GGI} + \zeta^{ijk}_{\pm,\rm GIG} + \zeta^{ijk}_{\pm,\rm GII} + \zeta^{ijk}_{\pm,\rm IGG} + \zeta^{ijk}_{\pm,\rm IGI} + \zeta^{ijk}_{\pm,\rm IIG} + \zeta^{ijk}_{\pm,\rm III}. \tag{A.14}$$

Again, just as a single example, the IIG case for $\zeta_{+,\rm obs}$ can be written as

$$\zeta^{ijk}_{+,\rm IIG}(\alpha) = \frac{1}{A(\alpha)}\int \frac{\mathrm{d}\ell\ell}{2\pi}\mathcal{B}^{ijk}_{+,\rm IIG}(\ell)J_0(\ell\alpha), \tag{A.15}$$

where

$$\mathcal{B}^{ijk}_{+,\rm IIG}(\ell) = \frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\int \frac{\mathrm{d}^2\boldsymbol{\ell}_2}{(2\pi)^2} B^{ijk}_{\kappa_{\rm I}\kappa_{\rm I}\kappa}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,-\boldsymbol{\ell}_{12})\, e^{2i(\phi_2-\phi_{-1-2})}U(\boldsymbol{\ell}_1)W(\boldsymbol{\ell}+\boldsymbol{\ell}_2)W(-\boldsymbol{\ell}-\boldsymbol{\ell}_{12}), \tag{A.16}$$

with $\boldsymbol{\ell}_{12} = \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2$ and

$$B^{ijk}_{\kappa_{\rm I}\kappa_{\rm I}\kappa_{\rm G}}(\boldsymbol{\ell}_a,\boldsymbol{\ell}_b,\boldsymbol{\ell}_c) = \int \mathrm{d}\chi \frac{n^i(\chi)n^j(\chi)q^k(\chi)}{\chi^4} B^{\rm 3D}_{\delta_{\rm I}\delta_{\rm I}\delta_m}\left(\frac{\boldsymbol{\ell}_a}{\chi},\frac{\boldsymbol{\ell}_b}{\chi},\frac{\boldsymbol{\ell}_c}{\chi};\chi\right). \tag{A.17}$$

The derivation of these expressions is the same as the usual gravitational lensing GGG expression, except one replaces the first two instances of κ by $\kappa_{\rm I}$. In the NLA model, $B^{\rm 3D}_{\delta_{\rm I}\delta_{\rm I}\delta_m} = f^2_{\rm IA}B^{\rm 3D}_{\delta_m\delta_m\delta_m}$. That is, the IIG contribution to $\zeta^{ijk}_{+,\rm obs}(\alpha)$ term can be obtained from GGG by simply replacing the ith and jth lensing kernels $q^i(\chi)$, $q^j(\chi)$ with $n^i(\chi)f_{\rm IA}$ and $n^j(\chi)f_{\rm IA}$. It follows as a result that all of the 8 contributions to $\zeta^{ijk}_{\pm,\rm obs}(\alpha)$ can be obtained by replacing all lensing kernels $q(\chi)$ with $q(\chi) + n(\chi)f_{\rm IA}$, as in eq. (2.24). This yields terms $\propto f^0_{\rm IA}$ (GGG), $\propto f_{\rm IA}$ (GGI, GIG, IGG) and $\propto f^2_{\rm IA}$ (GII, IGI, IIG) and $\propto f^3_{\rm IA}$ (III).

These 3-point contributions from galaxy IA are different than those derived in ref. [56] using also the NLA model. Among other differences, their III term is $\propto f^4_{\rm IA}$ and their GII + IGI + IIG terms are $\propto f^3_{\rm IA}$ (cf. their eqs. (30–32)). Reference [56] does not provide a detailed derivation of their expressions, which keeps us from inspecting this issue further. We emphasise, however, that the NLA model is in itself only an approximation of the effect of galaxy IA on small-scales, and so even our expressions should be interpreted in light of this.

References

[1] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, *Phys. Rept.* **340** (2001) 291 [astro-ph/9912508] [INSPIRE].

[2] C.S. Kochanek, *The Saas Fee Lectures on strong gravitational lensing*, in the proceedings of the *33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro*, Les Diablerets, Switzerland, April 7–12, 2003, [astro-ph/0407232] [INSPIRE].

JCAP07(2023)040

[3] DES collaboration, *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [arXiv:2105.13549] [INSPIRE].

[4] KiDS collaboration, *KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics*, *Astron. Astrophys.* **645** (2021) A104 [arXiv:2007.15633] [INSPIRE].

[5] T. Hamana et al., *Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data*, *Publ. Astron. Soc. Jap.* **72** (2020) PublicationsoftheAstronomicalSocietyofJaan,Volume72,Issue1,February2020,16,https://doiorg/101093/asj/sz138 [arXiv:1906.06041] [INSPIRE].

[6] Euclid collaboration, *Euclid preparation — VI. Verifying the performance of cosmic shear experiments*, *Astron. Astrophys.* **635** (2020) A139 [*Erratum ibid.* **638** (2020) C2] [arXiv:1910.10521] [INSPIRE].

[7] LSST Dark Energy Science collaboration, *Large Synoptic Survey Telescope: Dark Energy Science Collaboration*, arXiv:1211.0310 [INSPIRE].

[8] M. Yamamoto et al., *Weak gravitational lensing shear estimation with metacalibration for the Roman High-Latitude Imaging Survey*, *Mon. Not. Roy. Astron. Soc.* **519** (2023) 4241 [arXiv:2203.08845] [INSPIRE].

[9] P. Schneider and M. Lombardi, *The three-point correlation function of cosmic shear: I. The natural components*, *Astron. Astrophys.* **397** (2003) 809 [astro-ph/0207454] [INSPIRE].

[10] M. Takada and B. Jain, *Cosmological parameters from lensing power spectrum and bispectrum tomography*, *Mon. Not. Roy. Astron. Soc.* **348** (2004) 897 [astro-ph/0310125] [INSPIRE].

[11] P. Schneider, M. Kilbinger and M. Lombardi, *The three-point correlation function of cosmic shear. 2. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures*, *Astron. Astrophys.* **431** (2005) 9 [astro-ph/0308328] [INSPIRE].

[12] S. Dodelson and P. Zhang, *The weak lensing bispectrum*, *Phys. Rev. D* **72** (2005) 083001 [astro-ph/0501063] [INSPIRE].

[13] I. Kayo, M. Takada and B. Jain, *Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 344 [arXiv:1207.6322] [INSPIRE].

[14] M. Sato and T. Nishimichi, *Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation*, *Phys. Rev. D* **87** (2013) 123538 [arXiv:1301.3588] [INSPIRE].

[15] N. McCullagh, D. Jeong and A.S. Szalay, *Toward accurate modelling of the non-linear matter bispectrum: standard perturbation theory and transients from initial conditions*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 2945 [arXiv:1507.07824] [INSPIRE].

[16] R. Takahashi et al., *Fitting the nonlinear matter bispectrum by the Halofit approach*, *Astrophys. J.* **895** (2020) 113 [arXiv:1911.07886] [INSPIRE].

[17] E. Semboloni et al., *Weak lensing from space: first cosmological constraints from three-point shear statistics*, *Mon. Not. Roy. Astron. Soc.* **410** (2011) 143 [arXiv:1005.4941] [INSPIRE].

[18] L. Fu et al., *CFHTLenS: Cosmological constraints from a combination of cosmic shear two-point and three-point correlations*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 2725 [arXiv:1404.5469] [INSPIRE].

[19] A. Barthelemy, S. Codis and F. Bernardeau, *Probability distribution function of the aperture mass field with large deviation theory*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 5204 [arXiv:2012.03831] [INSPIRE].

[20] DES collaboration, *Dark Energy Survey Year 3 Results: Three-point shear correlations and mass aperture moments*, *Phys. Rev. D* **105** (2022) 103537 [arXiv:2201.05227] [INSPIRE].

JCAP07(2023)040

[21] S. Heydenreich, L. Linke, P. Burger and P. Schneider, *A roadmap to cosmological parameter analysis with third-order shear statistics — I. Modelling and validation*, *Astron. Astrophys.* **672** (2023) A44 [arXiv:2208.11686] [INSPIRE].

[22] DES collaboration, *Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 3653 [arXiv:1603.05040] [INSPIRE].

[23] J. Harnois-Déraps et al., *Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 1623 [arXiv:2012.02777] [INSPIRE].

[24] DES collaboration, *Dark energy survey year 3 results: Cosmology with peaks using an emulator approach*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2075 [arXiv:2110.10135] [INSPIRE].

[25] DES collaboration, *Density Split Statistics: Joint Model of Counts and Lensing in Cells*, *Phys. Rev. D* **98** (2018) 023508 [arXiv:1710.05162] [INSPIRE].

[26] DES collaboration, *Density Split Statistics: Cosmological Constraints from Counts and Lensing in Cells in DES Y1 and SDSS Data*, *Phys. Rev. D* **98** (2018) 023507 [arXiv:1710.05045] [INSPIRE].

[27] P. Burger et al., *An adapted filter function for density split statistics in weak lensing*, *Astron. Astrophys.* **642** (2020) A161 [arXiv:2006.10778] [INSPIRE].

[28] P. Burger, O. Friedrich, J. Harnois-Déraps and P. Schneider, *A revised density split statistic model for general filters*, *Astron. Astrophys.* **661** (2022) A137 [arXiv:2106.13214] [INSPIRE].

[29] P.A. Burger et al., *KiDS-1000 cosmology: Constraints from density split statistics*, *Astron. Astrophys.* **669** (2023) A69 [arXiv:2208.02171] [INSPIRE].

[30] S. Heydenreich, B. Brück and J. Harnois-Déraps, *Persistent homology in cosmic shear: constraining parameters with topological data analysis*, *Astron. Astrophys.* **648** (2021) A74 [arXiv:2007.13724] [INSPIRE].

[31] S. Heydenreich et al., *Persistent homology in cosmic shear — II. A tomographic analysis of DES-Y1*, *Astron. Astrophys.* **667** (2022) A125 [arXiv:2204.11831] [INSPIRE].

[32] A. Halder, O. Friedrich, S. Seitz and T.N. Varga, *The integrated three-point correlation function of cosmic shear*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 2780 [arXiv:2102.10177] [INSPIRE].

[33] C.-T. Chiang, C. Wagner, F. Schmidt and E. Komatsu, *Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum*, *JCAP* **05** (2014) 048 [arXiv:1403.3411] [INSPIRE].

[34] C.-T. Chiang et al., *Position-dependent correlation function from the SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 10 CMASS Sample*, *JCAP* **09** (2015) 028 [arXiv:1504.03322] [INSPIRE].

[35] D. Munshi and P. Coles, *The Integrated Bispectrum and Beyond*, *JCAP* **02** (2017) 010 [arXiv:1608.04345] [INSPIRE].

[36] G. Jung et al., *The integrated angular bispectrum of weak lensing*, *JCAP* **06** (2021) 055 [arXiv:2102.05521] [INSPIRE].

[37] D. Munshi et al., *Position-dependent correlation function of weak-lensing convergence*, *Phys. Rev. D* **107** (2023) 043516 [arXiv:2104.01185] [INSPIRE].

[38] A. Halder and A. Barreira, *Response approach to the integrated shear 3-point correlation function: the impact of baryonic effects on small scales*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 4639 [arXiv:2201.05607] [INSPIRE].

[39] A. Barreira and F. Schmidt, *Responses in Large-Scale Structure*, *JCAP* **06** (2017) 053 [arXiv:1703.09212] [INSPIRE].

JCAP07(2023)040

[40] P. Schneider, *Detection of (dark) matter concentrations via weak gravitational lensing*, *Mon. Not. Roy. Astron. Soc.* **283** (1996) 837 [astro-ph/9601039] [INSPIRE].

[41] R.G. Crittenden, P. Natarajan, U.-L. Pen and T. Theuns, *Discriminating weak lensing from intrinsic spin correlations using the curl-gradient decomposition*, *Astrophys. J.* **568** (2002) 20 [astro-ph/0012336] [INSPIRE].

[42] H. Gil-Marín et al., *An improved fitting formula for the dark matter bispectrum*, *JCAP* **02** (2012) 047 [arXiv:1111.4477] [INSPIRE].

[43] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, *Separate Universe Simulations*, *Mon. Not. Roy. Astron. Soc.* **448** (2015) L11 [arXiv:1409.6294] [INSPIRE].

[44] A.S. Schmidt, S.D.M. White, F. Schmidt and J. Stücker, *Cosmological N-Body Simulations with a Large-Scale Tidal Field*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 162 [arXiv:1803.03274] [INSPIRE].

[45] A. Mead et al., *An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 1958 [arXiv:1505.07833] [INSPIRE].

[46] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *JCAP* **07** (2011) 034 [arXiv:1104.2933] [INSPIRE].

[47] A. Barreira et al., *Separate Universe Simulations with IllustrisTNG: baryonic effects on power spectrum responses and higher-order statistics*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 2079 [arXiv:1904.02070] [INSPIRE].

[48] S. Foreman, W. Coulton, F. Villaescusa-Navarro and A. Barreira, *Baryonic effects on the matter bispectrum*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 2887 [arXiv:1910.03597] [INSPIRE].

[49] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration*, *Phys. Rev. D* **105** (2022) 023514 [arXiv:2105.13543] [INSPIRE].

[50] DES collaboration, *Dark Energy Survey Y3 results: blending shear and redshift biases in image simulations*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 3371 [arXiv:2012.08567] [INSPIRE].

[51] C.M. Hirata et al., *Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: Luminosity and redshift scalings and implications for weak lensing surveys*, *Mon. Not. Roy. Astron. Soc.* **381** (2007) 1197 [astro-ph/0701671] [INSPIRE].

[52] S. Bridle and L. King, *Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements*, *New J. Phys.* **9** (2007) 444 [arXiv:0705.0166] [INSPIRE].

[53] DES collaboration, *Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses*, arXiv:1706.09359 [INSPIRE].

[54] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps*, *Phys. Rev. D* **106** (2022) 083509 [arXiv:2110.10141] [INSPIRE].

[55] A.J.S. Hamilton, *Formulae for growth factors in expanding universes containing matter and a cosmological constant*, *Mon. Not. Roy. Astron. Soc.* **322** (2001) 419 [astro-ph/0006089] [INSPIRE].

[56] S. Pyne and B. Joachimi, *Self-calibration of weak lensing systematic effects using combined two- and three-point statistics*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 2300 [arXiv:2010.00614] [INSPIRE].

[57] R. Takahashi et al., *Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments*, *Astrophys. J.* **850** (2017) 24 [arXiv:1706.01472] [INSPIRE].

JCAP07(2023)040

[58] DES collaboration, *Dark Energy Survey Year 3 results: redshift calibration of the weak lensing source galaxies*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 4249 [arXiv:2012.08566] [INSPIRE].

[59] DES collaboration, *Dark energy survey year 3 results: weak lensing shape catalogue*, *Mon. Not. Roy. Astron. Soc.* **504** (2021) 4312 [arXiv:2011.03408] [INSPIRE].

[60] H.S. Xavier, F.B. Abdalla and B. Joachimi, *Improving lognormal models for cosmological fields*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 3693 [arXiv:1602.08503] [INSPIRE].

[61] M. Jarvis, G. Bernstein and B. Jain, *The skewness of the aperture mass statistic*, *Mon. Not. Roy. Astron. Soc.* **352** (2004) 338 [astro-ph/0307393] [INSPIRE].

[62] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty*, *Phys. Rev. D* **105** (2022) 023515 [arXiv:2105.13544] [INSPIRE].

[63] J. Hartlap, P. Simon and P. Schneider, *Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix*, *Astron. Astrophys.* **464** (2007) 399 [astro-ph/0608064] [INSPIRE].

[64] W.J. Percival et al., *The Clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Including covariance matrix errors*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2531 [arXiv:1312.4841] [INSPIRE].

[65] M. Takada and W. Hu, *Power Spectrum Super-Sample Covariance*, *Phys. Rev. D* **87** (2013) 123504 [arXiv:1302.6994] [INSPIRE].

[66] A. Barreira, E. Krause and F. Schmidt, *Complete super-sample lensing covariance in the response approach*, *JCAP* **06** (2018) 015 [arXiv:1711.07467] [INSPIRE].

[67] A. Barreira, E. Krause and F. Schmidt, *Accurate cosmic shear errors: do we need ensembles of simulations?*, *JCAP* **10** (2018) 053 [arXiv:1807.04266] [INSPIRE].

[68] A. Barreira, *The squeezed matter bispectrum covariance with responses*, *JCAP* **03** (2019) 008 [arXiv:1901.01243] [INSPIRE].

[69] A. Spurio Mancini et al., *CosmoPower: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 1771 [arXiv:2106.03846] [INSPIRE].

[70] S. Dodelson and M.D. Schneider, *The Effect of Covariance Estimator Error on Cosmological Parameter Constraints*, *Phys. Rev. D* **88** (2013) 063537 [arXiv:1304.2593] [INSPIRE].

[71] W.J. Percival, O. Friedrich, E. Sellentin and A. Heavens, *Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3207 [arXiv:2108.10402] [INSPIRE].

[72] S. Pyne, A. Tenneti and B. Joachimi, *Three-point intrinsic alignments of dark matter haloes in the IllustrisTNG simulation*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 1829 [arXiv:2204.10342] [INSPIRE].

[73] D. Huterer, M. Takada, G. Bernstein and B. Jain, *Systematic errors in future weak lensing surveys: Requirements and prospects for self-calibration*, *Mon. Not. Roy. Astron. Soc.* **366** (2006) 101 [astro-ph/0506030] [INSPIRE].

[74] E. Semboloni, C. Heymans, L. van Waerbeke and P. Schneider, *Sources of contamination to weak lensing three-point statistics: constraints from N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **388** (2008) 991 [arXiv:0802.3978] [INSPIRE].

[75] M.A. Troxel and M. Ishak, *Self-Calibration Technique for 3-point Intrinsic Alignment Correlations in Weak Lensing Surveys*, *Mon. Not. Roy. Astron. Soc.* **419** (2012) 1804 [arXiv:1109.4896] [INSPIRE].

JCAP07(2023)040

[76] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [INSPIRE].

[77] G.P. Lepage, *Adaptive multidimensional integration: VEGAS enhanced*, *J. Comput. Phys.* **439** (2021) 110386 [arXiv:2009.05112] [INSPIRE].

[78] A.G.G. Matthews et al., *GPflow: A Gaussian process library using TensorFlow*, arXiv:1610.08733 [DOI:10.48550/ARXIV.1610.08733].

[79] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [arXiv:2006.10256] [INSPIRE].

[80] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [INSPIRE].

[81] S. Hinton, *ChainConsumer*, *J. Open Source Softw.* **1** (2016) 45.

[82] Z. Vlah, N.E. Chisari and F. Schmidt, *An EFT description of galaxy intrinsic alignments*, *JCAP* **01** (2020) 025 [arXiv:1910.08085] [INSPIRE].

[83] Z. Vlah, N.E. Chisari and F. Schmidt, *Galaxy shape statistics in the effective field theory*, *JCAP* **05** (2021) 061 [arXiv:2012.04114] [INSPIRE].

[84] J. Blazek, N. MacCrann, M.A. Troxel and X. Fang, *Beyond linear galaxy alignments*, *Phys. Rev. D* **100** (2019) 103506 [arXiv:1708.09247] [INSPIRE].

JCAP07(2023)040

Chapter 7

C3NN: Cosmological Correlator Convolutional Neural Network an Interpretable Machine-learning Framework for Cosmological Analyses

Bibliographic and copyright information

Gong Z., Halder A., Bohrdt, A., Seitz, S., and Gebauer, D., *C3NN: Cosmological Correlator Convolutional Neural Network an Interpretable Machine-learning Framework for Cosmological Analyses*, 2024, ApJ 971 156

DOI: 10.3847/1538-4357/ad582e

Outline

This work can potentially be the starting point of a series of papers, trying to address the questions of (i) how the output of a Convolutional Neural Network (CNN) based model can be interpreted in terms of the N-point correlation functions (NPCFs), (ii) whether this property can help to distinguish and specify different physical models under the context of dark energy, dark matter and primordial non-Gaussianity etc., (iii) can this model be exploited to investigate quantitatively the information content of cosmological fields (e.g., galaxy density and weak lensing) within the hierarchical ladder of the correlation functions and (iv) the improvement in the model performance by changing its architecture (e.g., from CNNs to transformers) while maintaining its interpretability. As the first author, I was responsible for the following key tasks: (i) studying the initial model proposed in

Miles et al. (2021) and adapting it from the research field of quantum many-body systems to cosmology, (ii) developing mathematical formulae to explain the relation between the output of C3NN at a given Nth order and the corresponding N-point correlation function, and (iii) performing all the proof-of-concept tests with Gaussian and log-normal random fields, as well as the more advanced test using simulated weak lensing convergence maps of different values of the dark energy equation of state w_0. I also wrote majority of the manuscript. The co-authors contributed to preparing simulation data, suggesting valuable tests and analysis methods, drafting specific sections of the manuscript, and reviewing the final draft.

The Astrophysical Journal, 971:156 (16pp), 2024 August 20 https://doi.org/10.3847/1538-4357/ad582e

C3NN: Cosmological Correlator Convolutional Neural Network an Interpretable Machine-learning Framework for Cosmological Analyses

Zhengyangguang Gong[1,2], Anik Halder[1,2], Annabelle Bohrdt[3,4], Stella Seitz[1,2], and David Gebauer[1,2]
[1] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstraße 1, 81679 München, Germany; lgong@usm.lmu.de
[2] Max Planck Institute for Extraterrestrial Physics, Giessenbachstraße 1, 85748 Garching, Germany; ahalder@usm.lmu.de
[3] University of Regensburg, Universitätsstraße 31, Regensburg D-93053, Germany; annabelle.bohrdt@physik.uni-regensburg.de, stella@usm.lmu.de, dgebauer@usm.lmu.de
[4] Munich Center for Quantum Science and Technology, Schellingstraße 4, Munich D-80799, Germany
Received 2024 March 18; revised 2024 June 12; accepted 2024 June 12; published 2024 August 14

Abstract

Modern cosmological research in large-scale structure has witnessed an increasing number of machine-learning applications. Among them, convolutional neural networks (CNNs) have received substantial attention due to their outstanding performance in image classification, cosmological parameter inference, and various other tasks. However, many models based on CNNs are criticized as "black boxes" due to the difficulties in relating their outputs intuitively and quantitatively to the cosmological fields under investigation. To overcome this challenge, we present the Cosmological Correlator Convolutional Neural Network (C3NN)—a fusion of CNN architecture and cosmological N-point correlation functions (NPCFs). We demonstrate that its output can be expressed explicitly in terms of the analytically tractable NPCFs. Together with other auxiliary algorithms, we can open the "black box" by quantitatively ranking different orders of the interpretable outputs based on their contribution to classification tasks. As a proof of concept, we demonstrate this by applying our framework to a series of binary classification tasks using Gaussian and log-normal random fields and relating its outputs to the NPCFs describing the two fields. Furthermore, we exhibit the model's ability to distinguish different dark energy scenarios ($w_0 = -0.95$ and -1.05) using N-body simulated weak-lensing convergence maps and discuss the physical implications coming from their interpretability. With these tests, we show that C3NN combines advanced aspects of machine learning architectures with the framework of cosmological NPCFs, thereby making it an exciting tool to extract physical insights in a robust and explainable way from observational data.

Unified Astronomy Thesaurus concepts: Astrostatistics techniques (1886); Classification (1907); Convolutional neural networks (1938); Weak gravitational lensing (1797); Cosmological parameters (339)

1. Introduction

In recent years, numerous machine-learning methods have found applications in cosmology and astrophysics ranging from classification and regression tasks to acceleration of computational methods (see Dvorkin et al. 2022 for a recent review). Among the various machine-learning techniques, convolutional neural networks (CNNs; LeCun et al. 2015) have been used extensively. Briefly, CNNs can compress a large data set (e.g., images) into several feature representations through a series of alternative linear and nonlinear transformations. The large number of parameters in a CNN model is typically determined by training the model to numerous simulated data that aim at reproducing actual scientific phenomena. In the context of astronomy and cosmology, these compressed features can be used for classification, such as searching for strong gravitational lensing systems (Rojas et al. 2022) and classifying different galaxy morphologies (Domínguez Sánchez et al. 2022), or for inference analyses such as constraining parameters in various cosmological models (Fluri et al. 2019, 2022; Lu et al. 2023), to name a few.

However, this impressive development of CNNs has raised an important question: How to interpret the output feature representations of a CNN? Ideally, we would like to establish a complete and controllable process in the application of CNN models in cosmology. However, most of the output features of a conventional CNN are notoriously difficult to interpret and hence they have often been termed "black boxes." To open such a black box, one good approach would be to connect the output feature representations from a CNN to either the analyzable statistical properties underlying the training data or visually understandable images that resemble the input training maps. If this can be achieved, we can understand what physical information in the training data the CNN pays the most attention in order to complete an assigned task. Moreover, the mechanism that the model adopts to extract this physical information can potentially provide us with certain knowledge that cannot be acquired by conventional CNN methods.

To address this problem of interpretability in CNN models, different methods and architectures have been developed such as saliency maps (Simonyan et al. 2013; see, e.g., Matilla et al. 2020 and Villanueva-Domingo & Villaescusa-Navarro 2021 for a few applications), the three-dimensional CNN framework in Lucie-Smith et al. (2024), variational autoencoders applied to find a compressed representation of dynamical dark energy (DE-VAE) models (Piras & Lombriser 2024), and so on. However, these techniques are either more visual-based checks, which can only lead to qualitative conclusions, or involve complex nonlinear transformations such that one cannot establish a straightforward mathematical relation between the output features and the input data. This can to a certain degree

limit our interpretation of the physical information captured by CNNs.

To overcome this challenge, we introduce the Cosmological Correlator Convolutional Neural Network (C3NN) for cosmological analyses, adapted from the model initially proposed by Miles et al. (2021) in studies of correlated quantum matter. The output of this model can be explicitly related via a one-to-one correspondence to different orders of cosmological N-point correlation functions (NPCFs). The whole framework of NPCFs and their applications in cosmology (Bernardeau et al. 2002; Peebles 2020) have been subject to extensive theoretical as well as practical developments over the last several decades. The two-point correlation function (2PCF), or its Fourier space counterpart the power spectrum, is currently the most widely used statistical method in many fields of cosmology such as weak gravitational lensing (Asgari et al. 2021; Secco et al. 2022; Dalal et al. 2023), galaxy clustering (Ivanov et al. 2020; Chen et al. 2022) and cosmic microwave background (Planck Collaboration et al. 2020; Rosenberg et al. 2022) analyses. Within the correlation function framework, active research has been done to push the theoretical modeling beyond 2PCF to access non-Gaussian cosmological information, e.g., in the studies of weak lensing and projected galaxy density fields such as those of Takada & Jain (2003a, 2003b), Semboloni et al. (2010), Friedrich et al. (2018), Halder et al. (2021, 2023), Halder & Barreira (2022), Gatti et al. (2022), Gong et al. (2023), Burger et al. (2024), Heydenreich et al. (2023), Anbajagane et al. (2023), Barthelemy et al. (2024) to name a few. However, due to the difficulties in modeling, measurement, and the treatment of systematic effects, extensive analyses of higher-order correlation functions in actual data remain challenging. Therefore, it is difficult to quantify the amount of additional cosmological information contained in three-point, four-point, or even higher-order correlation functions. From this perspective, C3NN can provide some novel insights as we design it such that:

1. It can efficiently extract a series of statistical features, which we call moments. These moments are directly related to compressed NPCFs of the input data up to a desired order N (user specified) by construction. C3NN relies on simulations for its training wherein one can include all the desired physics and systematic effects at the field level. Therefore, in extracting these output moments, C3NN sidesteps the need for involved theoretical modeling of higher-order correlation functions.
2. At the same time, if one has a theoretical model for a given NPCF it can be directly used to interpret the corresponding order C3NN moment through straightforward mathematical relations as explained later.
3. Using a certain postprocessing numerical method called regularization path analysis, we can easily understand the relative importance of the output moments in tasks such as binary classification, e.g., to distinguish between different cosmological model scenarios beyond ΛCDM. This in turn can guide corresponding NPCF analyses in real observations.
4. Finally, the filter weights that are learned by C3NN can be used to pick out the specific configurations within a given NPCF. This is extremely desirable as it also allows us to understand the relative importance of a given order moment as well as the configurations that carry the most distinct information for performing an assigned task.

As this is the first time we have introduced C3NN in cosmology, we only consider the task of binary classification with two-dimensional simulated maps in this work. However, the same architecture can be exploited in cosmological parameter inference or classification with three-dimensional simulated boxes, etc. We leave such topics to future works. This paper is organized as follows: In Section 2, we describe the architecture of C3NN. We then give clear mathematical expressions for the output moments and explicitly show their relations to the conventional correlation functions in cosmology. We also discuss the differences among C3NN, CNN, and scattering transforms. In the following Sections 3.1 and 3.2, we present as a proof of concept, two test cases of C3NN classifying between (i) two different Gaussian random fields, and (ii) between Gaussian and log-normal random fields. Then in Section 4 we further apply C3NN to distinguish between different dark energy scenarios using N-body simulated weak-lensing convergence maps and discuss the physical implications coming from its interpretability. Finally, we summarize and conclude in Section 5.

2. C3NN Model Architecture and Interpretability

In this section, we present the architecture of C3NN whose overall structure is shown in Figure 1. We split it into two parts and name them as CNN-based N-point moment map generator and moment-map-based classifier, respectively.

CNN-based N-point moment map generator: The generator part of the architecture performs convolution operations recursively in order to produce moment maps $C_\alpha^{(N)}$ (of order N), which we define as follows:

$$
\begin{aligned}
C_\alpha^{(1)}(\boldsymbol{x}) &= \sum_{\boldsymbol{a},k} w_{\alpha,k}(\boldsymbol{a}) S_k(\boldsymbol{x}+\boldsymbol{a}) \\
C_\alpha^{(2)}(\boldsymbol{x}) &= \frac{1}{2!}\left[\sum_{\boldsymbol{a},\boldsymbol{a}_1,k,k'}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}_1,k')\}} w_{\alpha,k}(\boldsymbol{a}) w_{\alpha,k'}(\boldsymbol{a}_1) S_k(\boldsymbol{x}+\boldsymbol{a}) S_{k'}(\boldsymbol{x}+\boldsymbol{a}_1)\right] \\
&\ldots \\
C_\alpha^{(N)}(\boldsymbol{x}) &= \frac{1}{N!}\left[\sum_{\boldsymbol{a},\ldots,\boldsymbol{a}_N,k,\ldots,k_N}^{\{(\boldsymbol{a},k)\neq\ldots\neq(\boldsymbol{a}_N,k_N)\}} \prod_{j=1}^{N} w_{\alpha,k_j}(\boldsymbol{a}_j) S_{k_j}(\boldsymbol{x}+\boldsymbol{a}_j)\right].
\end{aligned}
\tag{1}
$$

Here, w is the filter weight exploited in standard CNN filters and S denotes the input map (e.g., a pixelized image of a 2D field). During training, we enforce w to take its absolute value to aid interpretation. The curly brackets above the summation include constraints for the summed variables. Position vectors $\boldsymbol{a}$, $\boldsymbol{a}_1$ ... $\boldsymbol{a}_N$ of the filter weight run over every pixel within the finite support of the filter. The factorial factor removes repeated counting of the same combination of filter weights and input map pixels. Index α represents different filters while index k goes through different channels. Each filter can have several channels that correspond to those in the input data, e.g., different tomographic bins of an observable. From the above equations, we see clearly that for $N \geqslant 2$, $C_\alpha^{(N)}(\mathbf{x})$ is the weighted summation of all possible pixel configurations within the filter at position $\mathbf{x}$. For $N=1$, $C_\alpha^{(1)}$ represents the usual convolution of the input data with the standard CNN filter. In the summation, we strictly exclude zero-lag contributions for any order. The summation over k enables C3NN to measure cross correlations among different channels of the input data. It is exactly relying on this property that we can straightforwardly

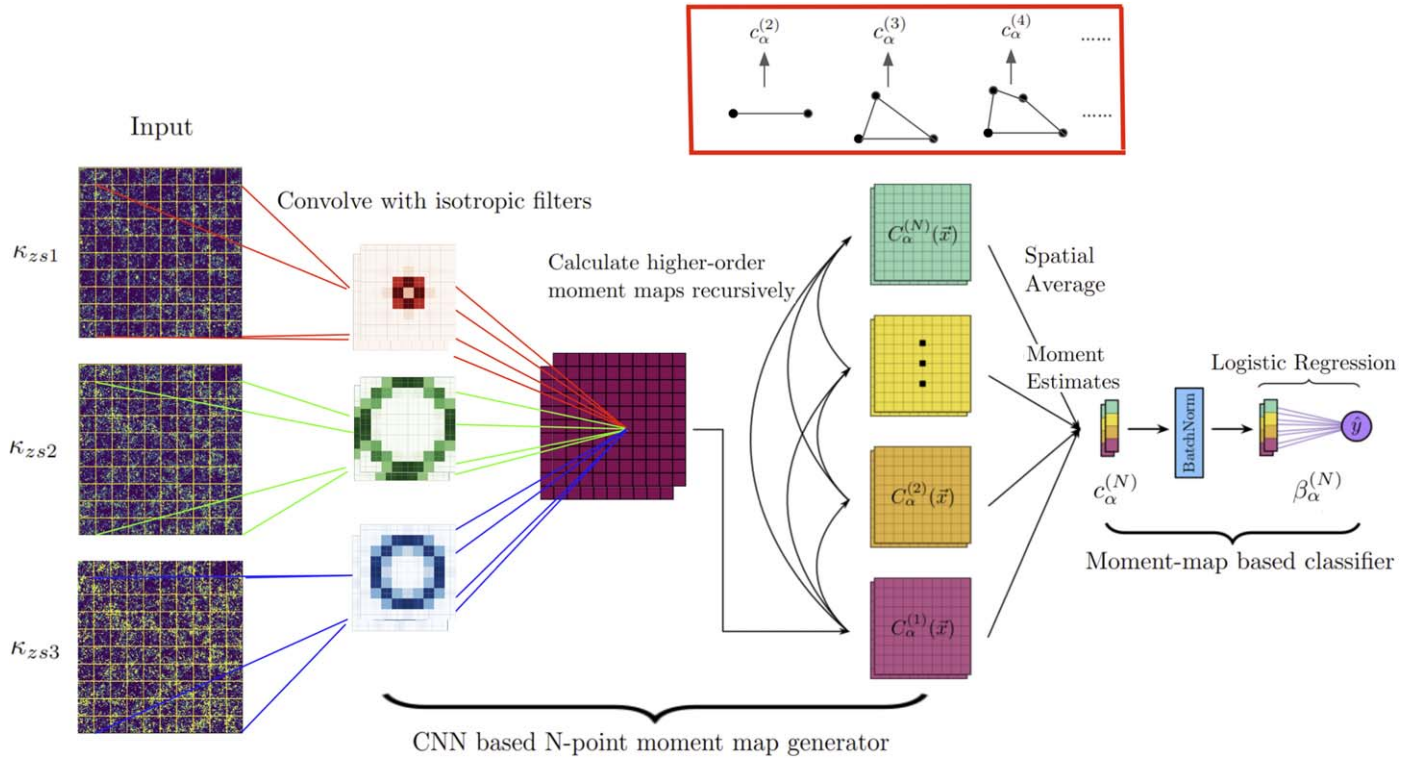

Figure 1. A schematic illustration of the C3NN architecture. The input data can be multichannel 2D maps (e.g., weak-lensing convergence maps with multiple distinct source redshift bins). The dimensionality of the input data tensor is (D, C, W, and H) where D is the number of simulation realizations, C the number of channels, and W and H describe the width and height (in pixels) of each map. This input data tensor is then convolved with isotropic filters built with `ESCNN` (see details in Section 2) in order to produce the first-moment map $C_\alpha^{(1)}$. With this single convolution operation, one can calculate higher-order moment maps of the input field (up to order N, which is user defined) recursively (see Section 2 for details). The maps are then spatially averaged to obtain compressed scalar quantities that we call moments $c_\alpha^{(N)}$. As shown in the top red rectangular box, these $c_\alpha^{(N)}$ are associated with the corresponding orders of NPCFs. The calculated moments can then be passed through the classifier part of the model to perform a binary classification task. Part of this figure has been adapted from Miles et al. (2021).

implement C3NN within the context of a tomographic analysis as we will discuss later in Section 4. Moreover, in the case when one provides different observables as input to the different channels, C3NN would capture cross correlations among these quantities, e.g., the 3×2-point probes of cosmic shear and galaxy clustering that are routinely analyzed in galaxy imaging surveys (Heymans et al. 2021; Abbott et al. 2022). We defer such investigations to future works.

One difficulty in computing moment maps in Equation (1) is the increasing computational cost along with the order N. The cost can be approximated as $\mathcal{O}((KP)^N)$ per input map pixel where K is the number of input data channels and P is the number of pixels in the convolutional filter. To solve this problem, Miles et al. (2021; see their Section S II) have proved that $C_\alpha^{(N)}$ can be calculated using a recursive formula:

$$C_\alpha^{(N)}(\boldsymbol{x}) = \frac{1}{N}\sum_{\ell=1}^{N}(-1)^{\ell-1}\left(\sum_{\boldsymbol{a},k} w_{\alpha,k}^{\ell}(\boldsymbol{a}) S_k^{\ell}(\boldsymbol{x}+\boldsymbol{a})\right) C_\alpha^{(N-\ell)}(\boldsymbol{x}), \tag{2}$$

where the power operation is taken on every pixel and we set $C_\alpha^{(0)}(\boldsymbol{x})$ to be 0. Through this relation, the computational cost decreases to $\mathcal{O}(N^2KP)$. Instead of direct computation at each order of the moment map, we can now convolve the input only once to produce $C_\alpha^{(1)}$. Then $C_\alpha^{(2)}$ can be calculated from $C_\alpha^{(1)}$. Similarly, $C_\alpha^{(3)}$ can be calculated from $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ and so on via Equation (2). This recursive relation holds until we truncate at an order N.

Moment-map-based classifier: This classifier part of the architecture starts with compressing the $C_\alpha^{(N)}$ maps obtained from the generator into scalars $c_\alpha^{(N)}$ by taking the spatial average over the maps:

$$c_\alpha^{(N)} = \frac{1}{N_{\rm pix}}\sum_{\mathbf{x}} C_\alpha^{(N)}(\mathbf{x}), \tag{3}$$

where $N_{\rm pix}$ is the total number of pixels in the input map S. Based on Equations (1) and (3), we can explicitly show the relation between $c_\alpha^{(N)}$ and the corresponding NPCF both at the discrete filter pixel level and in the continuous limit. As an example, when $N = 2$, we have

$$\begin{aligned} c_\alpha^{(2)} &= \frac{1}{N_{\rm pix}}\sum_{\boldsymbol{x}}\frac{1}{2!}\left[\sum_{\boldsymbol{a},\boldsymbol{a}_1,k,k'}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}_1,k')\}} w_{\alpha,k}(\boldsymbol{a}) w_{\alpha,k'}(\boldsymbol{a}_1) S_k(\boldsymbol{x}+\boldsymbol{a}) S_{k'}(\boldsymbol{x}+\boldsymbol{a}_1)\right] \\ &= \frac{1}{2!}\sum_{\boldsymbol{a},\boldsymbol{a}_1,k,k'}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}_1,k')\}} w_{\alpha,k}(\boldsymbol{a}) w_{\alpha,k'}(\boldsymbol{a}_1) \\ &\quad \times\left[\frac{1}{N_{\rm pix}}\sum_{\boldsymbol{x}} S_k(\boldsymbol{x}+\boldsymbol{a}) S_{k'}(\boldsymbol{x}+\boldsymbol{a}_1)\right] \\ &= \frac{1}{2!}\sum_{\boldsymbol{a},\boldsymbol{a}_1,k,k'}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}_1,k')\}} w_{\alpha,k}(\boldsymbol{a}) w_{\alpha,k'}(\boldsymbol{a}_1)\hat{\xi}_{kk'}(\boldsymbol{a}_1-\boldsymbol{a}) \\ &= \frac{1}{2!}\sum_{\boldsymbol{a},\boldsymbol{r},k,k'}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}+\boldsymbol{r},k')\}} w_{\alpha,k}(\boldsymbol{a}) w_{\alpha,k'}(\boldsymbol{a}+\boldsymbol{r})\hat{\xi}_{kk'}(\boldsymbol{r}), \end{aligned} \tag{4}$$

where $\hat{\xi}_{kk'}$ is the volume average estimator for auto/cross 2PCFs at a separation $\mathbf{r}$ between channels k and k'. Note that both $\boldsymbol{a}$ and $\boldsymbol{a}+\boldsymbol{r}$ in the last line should be strictly contained within the filter. Therefore, the scale of correlation functions that C3NN can probe is restricted by the filter size. Similarly, $c_\alpha^{(3)}$ can be written as

$$c_\alpha^{(3)} = \frac{1}{3!}\Bigg[\sum_{\boldsymbol{a},\boldsymbol{r},\boldsymbol{r}',k,k_1,k_2}^{\{(\boldsymbol{a},k)\neq(\boldsymbol{a}+\boldsymbol{r},k_1)\neq(\boldsymbol{a}+\boldsymbol{r}',k_2)\}} \times w_{\alpha,k}(\boldsymbol{a})w_{\alpha,k_1}(\boldsymbol{a}+\boldsymbol{r})w_{\alpha,k_2}(\boldsymbol{a}+\boldsymbol{r}')\hat{\zeta}_{kk_1k_2}(\boldsymbol{r},\boldsymbol{r}',-\boldsymbol{r}-\boldsymbol{r}')\Bigg], \quad (5)$$

where $\hat{\zeta}_{kk_1k_2}$ is the volume average estimator for auto/cross three-point correlation functions (3PCFs) where closed triangles can be formed by 3 pixels within the filter. To complete the demonstration, we consider the calculation of $c_\alpha^{(2)}$ in the limit of a continuous filter function:

$$c_\alpha^{(2)} = \frac{1}{2!}\sum_{k,k'}\int \boldsymbol{d}^2\boldsymbol{a}\; W_{\alpha,k}(\boldsymbol{a})\int_{\boldsymbol{r}\neq\boldsymbol{0}} \boldsymbol{d}^2\boldsymbol{r}\; W_{\alpha,k'}(\boldsymbol{a}+\boldsymbol{r})\hat{\xi}_{kk'}(\boldsymbol{r}), \quad (6)$$

where $W_{\alpha,k}$ acts as a continuous two-dimensional kernel function that has a sharp boundary outside which its values are always 0. This takes the finite filter size into consideration. A similar calculation for $c_\alpha^{(3)}$ follows analogously. These explicitly show the relation between $c_\alpha^{(N)}$ and NPCFs thereby robustly identifying them as interpretable summary statistics.

We note that $c_\alpha^{(N)}$ is an integration over all the configurations of a given order NPCF and hence we call it an *N-point moment*. It is interesting to compare $c_\alpha^{(n)}$ to the mass aperture moments used for weak lensing (Schneider et al. 2005; Kilbinger & Schneider 2005; Heydenreich et al. 2023):

$$\langle\mathcal{M}_{\rm ap}^2\rangle(\theta) = \int\frac{d\ell\ell}{2\pi}P_\kappa(\ell)\hat{u}(\theta\ell), \quad (7)$$

$$\langle\mathcal{M}_{\rm ap}^3\rangle(\theta_1,\theta_2,\theta_3) = \frac{2}{(2\pi)^3}\int d\ell_1\ell_1\int d\ell_2\ell_2\int_0^\pi d\phi \times \hat{u}(\theta_1\ell_1)\hat{u}(\theta_2\ell_2)\hat{u}(\theta_3\sqrt{\ell_1^2+\ell_2^2+2\ell_1\ell_2\cos\phi}) \times B_\kappa(\ell_1,\ell_2,\phi), \quad (8)$$

where we show the analytic expressions for the second and third-order mass aperture moments. In the above expressions, P_κ and B_κ are the power spectrum and bispectrum of the underlying cosmological field, here the weak-lensing convergence κ (see Section 3.2), whereas $\hat{u}$ is the Fourier counterpart of the filter function used for computing the aperture mass moments. Though written as integrals in Fourier space, their expressions are essentially similar to those of $\boldsymbol{c}_{\boldsymbol{\alpha}}^{(\boldsymbol{N})}$. The difference is that mass aperture moments exploit *predetermined* analytical kernel functions (the $\hat{u}$ in the above equations) such as compensated filters (Crittenden et al. 2002) that can have different scale radii while the C3NN filters are *automatically learned and optimized* by the training process and the same filter is applied to all input data pixels. One point to note is that measuring aperture mass moments beyond the third order has been difficult to date (see Porth et al. 2020 for some recent progress). On the other hand, C3NN can measure a whole sequence of $c_\alpha^{(n)}$ up to any arbitrary order very efficiently leveraging the recursion relation mentioned in Equation (2). Another point to note regarding $c_\alpha^{(n)}$ is that since it explicitly excludes the measurement of self-correlation, it is also different from the conventional definition of *N*th-order moment, which is composed of zero-lag correlation functions. For simplicity, in the following texts, we will mention $c_\alpha^{(n)}$ as moments unless otherwise specified.

After building all the $c_\alpha^{(N)}$ up to a given order N, we group them together to form a feature vector $\mathbf{c} = \{c_\alpha^{(1)}, c_\alpha^{(2)}, c_\alpha^{(3)},\ldots, c_\alpha^{(N)}\}$. Then the vector goes through a batch normalization, which applies standard normalization to each batch of the training data to scale all entries of $\mathbf{c}$ to the same order of magnitude for numerical stability during the latter training process. Finally, the model passes it through a logistic classifier for a final probabilistic prediction $\hat{y}$ to differentiate two classes:

$$\hat{y} = \frac{1}{1+\mathrm{e}^{-\boldsymbol{\beta}\cdot\boldsymbol{c}+\epsilon}}, \quad (9)$$

where $\boldsymbol{\beta}$ is a trainable vector composed of coefficients $\beta_\alpha^{(N)}$ which have a one-to-one correspondence to $c_\alpha^{(N)}$. The other trainable parameter is the bias scalar ϵ. When $-\boldsymbol{\beta}\cdot\boldsymbol{c}+\epsilon=0$, the probabilistic prediction would always be 0.5, which implies a random classification between two classes. From that naturally the hyperplane $-\boldsymbol{\beta}\cdot\boldsymbol{c}+\epsilon=0$ acts as the decision boundary in the high N-dimensional space. C3NN would classify the feature vector $\mathbf{c}$ that falls to either side of the decision boundary into a specific class. In other words, if $\hat{y}\in[0, 0.5)$, the model would predict the input data to belong to one class. Otherwise, the input data would be considered as the other class.

To further develop the interpretability, we adopt a two-round training strategy: In the first round, we train the whole C3NN exhibited in Figure 1. In this case, we use the binary cross entropy as the loss function with an extra L1 regularization term on the filter weights

$$L_{\rm 1st}(y,\hat{y}) = -y\log\hat{y} - (1-y)\log(1-\hat{y}) + \gamma\sum_{\alpha,k,\boldsymbol{a}} w_{\alpha,k}(\boldsymbol{a}), \quad (10)$$

where γ is the regularization strength. This regularization term helps to turn off unnecessary pixels (Bluecher et al. 2020), which aids our later interpretation of the filter weights. The resulting C3NN model can be directly used to perform binary classification tasks. In the second round, we freeze the filter weights from the previous round and use the moment map generator to produce moment maps. We then use those maps to retrain $\beta_\alpha^{(N)}$ coefficients and $\boldsymbol{\epsilon}$ in Equation (9). It is important to note that the value of $\beta_\alpha^{(N)}$ indicates the relative importance of $c_\alpha^{(N)}$ in its contribution to the classification, e.g., from Equation (9) when $\beta_\alpha^{(N)}=0$, its corresponding $c_\alpha^{(N)}$ would be irrelevant in C3NN's decision making. Based on this property, we use a different L1 regularization term in the binary cross-entropy loss during the second round of training:

$$L_{\rm 2nd}(y,\hat{y}) = -y\log\hat{y} - (1-y)\log(1-\hat{y}) + \lambda\sum_{\alpha,n}|\beta_\alpha^{(n)}|, \quad (11)$$

where λ is the regularization strength. We train the $\beta_\alpha^{(N)}$ coefficients iteratively with gradually decreasing λ values. This

is a process of feature selection using the so-called regularization path analysis (Efron et al. 2004; Tang et al. 2014). Initially, when λ is large, all $\beta_\alpha^{(N)}$ values are close to 0 in order to minimize $L_{\rm 2nd}$ and therefore the model does not possess any classification power. After λ decreases to a certain value, the most important feature represented by a given $\beta_\alpha^{(N)}$ would be activated since at this stage the increase in the loss from $\lambda\sum_{\alpha,n}|\beta_\alpha^{(n)}|$ would be overcompensated by the decrease from $-y\log\hat{y}-(1-y)\log(1-\hat{y})$ and thus minimize the total loss. As $\beta_\alpha^{(N)}$ is coupled to $c_\alpha^{(N)}$, the sequence of the activation of $\beta_\alpha^{(N)}$ can provide us insights into the relative importance of different moments.

Once we find the sequence of the $\beta_\alpha^{(N)}$ activation, we can check the filter weights that are trained during the first round. For each corresponding moment $c_\alpha^{(N)}$ measured by a specific filter, we can quantify the relative importance of a particular NPCF configuration. The method is demonstrated in Figure 2. As an aside, we note here that the smallest configuration of the correlation function at a given order is determined by the pixel resolution of the input map. This method implies that we can not only infer a rank of moments based on their relative importance in classification but also a rank of correlation function configurations based on the relative weights within each moment.

The filters used in the convolution are constructed in such a way that their weights after training are rotationally symmetric about the central filter pixel, i.e., weights that have the same distance to the central pixel are identical. We impose this symmetry because the spatially averaged moments $c_\alpha^{(N)}$ should be invariant to rotations of the training maps. From the perspective of the cosmological principle, this ensures that our model obeys the isotropy of the Universe. In practice, we implement this in our filters using the equivariant steerable convolutional neural network[5] (`ESCNN`; Weiler & Cesa 2019; Cesa et al. 2022) architecture. The requirement of rotational invariance imposed on the filter naturally decreases the number of free weights to be trained and therefore makes the training procedure more numerically efficient as it needs fewer training data.

Besides the actual trainable parameters such as filter weights and $\beta_\alpha^{(N)}$, we also consider the hyperparameters within the model. Parameters like the learning rate or regularization strength cannot update themselves during the training but their values can affect the model performance. Therefore, in practice, we use `Optuna`[6] (Akiba et al. 2019) as the framework for optimizing these hyperparameters.

It is necessary to point out, based on what we have already discussed, the differences between C3NN and some other AI methods at first glance can look similar to our model. For example, in conventional CNNs used in cosmological research (Fluri et al. 2019, 2022; Lu et al. 2023), there are several convolutional layers, each of which is followed by a nonlinear activation function. This type of information extraction compresses the input data into some compact feature vectors, which are then used by the model to undertake tasks such as classification or regression. Although very powerful, one disadvantage of this conventional CNN framework is that it is generally hard to interpret the reasons why the CNN compresses the input data in the way it does to train its model parameters. On the other hand, our C3NN architecture has only one convolution layer. We discard the general nonlinear transformations and instead replace them with the expansion of moment maps via a recursive relation. In other words, our C3NN architecture is simpler than the conventional multilayered CNNs. Meanwhile, the final compressed feature vectors from C3NN can be mathematically related to the concept of NPCFs as discussed above, which are summary statistics with great theoretical tractability and are the basis of statistical analyses in cosmology over the past few decades.

Scattering transforms (Cheng et al. 2020; Cheng & Ménard 2021) are also different from C3NN in the sense that they use predetermined wavelet filters. C3NN filters on the other hand are optimized during the training process. Moreover, unlike C3NN, the final compressed coefficients of the scattering transform (no matter up to which order they are computed) do not represent the corresponding NPCF that is contained in the C3NN output $c_\alpha^{(N)}$.

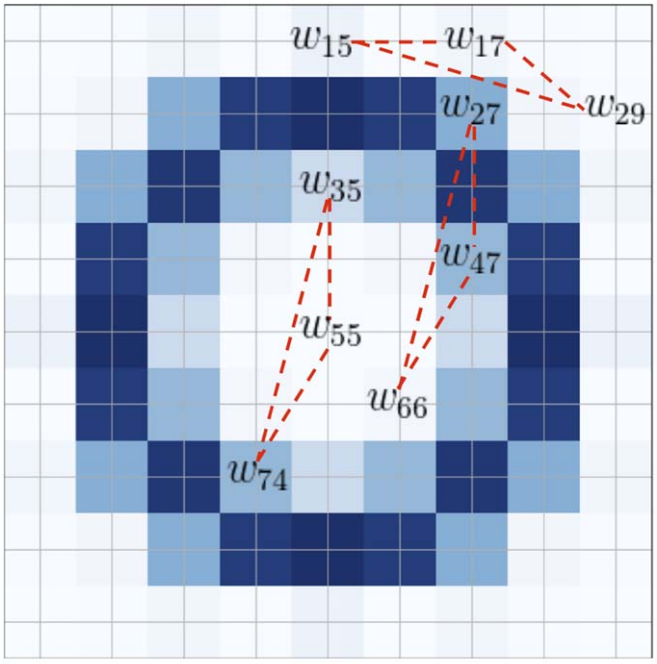

Figure 2. A demonstration of how C3NN filter weights can be analyzed: suppose we have $\beta_1^{(3)}$ first activated in the regularization path analysis, we can then check the filter and quantify the relative importance of different configurations of 3PCF that $c_1^{(3)}$ is composed of. Here, we assume that the trained rotationally symmetric filter is of size 9×9 pixels and it only has one channel. Each intersection between a vertical and a horizontal line represents the center of a pixel. The trained weight is denoted by $w_{\rm ij}$ and any distinct triplet would form a closed triangle that is included in the measurement of $c_1^{(3)}$. As an example, the three triangles drawn in red dashed lines are weighted as $w_{35}w_{55}w_{74}$, $w_{27}w_{47}w_{66}$, and $w_{15}w_{17}w_{29}$, respectively, according to Equation (1). Since they are equivalent triangles, their weights can be summed, along with all the other equivalent triangles that can be found within the filter. The summation is the total weight W of this particular triangle configuration. We want to emphasize that all configurations of the 3PCF mentioned here are in real space, which is different from triangles in Fourier space often discussed in the context of bispectrum studies in literature. The same approach can be applied to two-point separations, quadrilaterals, etc.

[5] Currently hosted at https://github.com/QUVA-Lab/escnn.
[6] Currently hosted at https://optuna.org/.

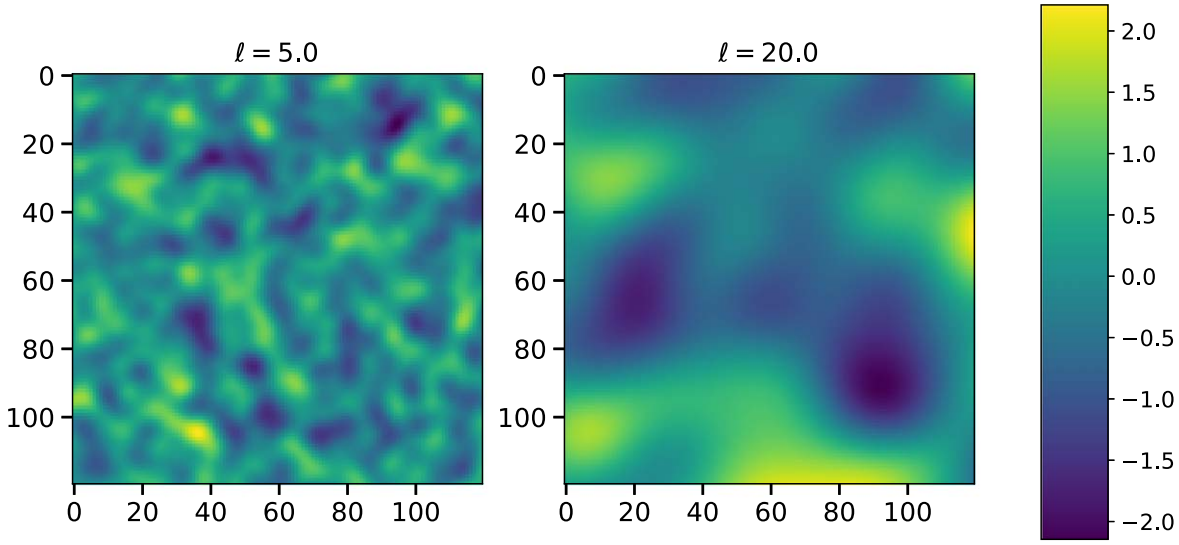

Figure 3. Two classes of Gaussian random fields generated by `GSTools` where we show one example for each. Both classes have the same variance amplitude of 1.0 and a map size of 120 × 120 pixels. They differ from each other on the correlation length, with one having $\ell = 5.0$ (left panel) and the other having $\ell = 20.0$ (right panel), with the correlation lengths specified in units of pixels. The color bar to the right shows the scale of the field amplitude.

Table 1
Four Optimized Hyperparameters in C3NN Trained on Two Classes of Gaussian Random Fields with Different Correlation Lengths[a]

Parameter	γ	Learning Rate (lr)	Learning Rate Decaying Ratio (ϕ)	Optimizer
Value	2.33	0.47	0.02	"Adam"

Notes. γ is the regularization strength in Equation (10). We use a learning rate scheduler that decays the initial learning rate (lr) of each parameter group by a factor ϕ after every epoch. We simultaneously also search for the type of optimizer that performs the best and we obtain Adam, which is derived from adaptive moment estimation (Kingma & Ba 2014).
[a] See Figure 3.

3. Proof of Concept Tests

In this section, we test C3NN on two binary classification tasks: (i) Gaussian random fields with different correlation lengths (Section 3.1), and (ii) Gaussian and log-normal random fields with the same power spectrum but different higher-order spectra (Section 3.2). Using these two tests we also demonstrate how the regularization path analysis and filter weights analysis relate to our model's interpretability as mentioned in the previous section.

3.1. Test Case of Gaussian Random Fields

As the first proof of concept, we apply C3NN to distinguish two classes of Gaussian random fields, each with a different correlation length. We use the `GSTools`[7] (Müller et al. 2022) framework to generate these Gaussian random fields as square 2D maps. While keeping the variance amplitude of the two fields the same, we vary their field correlation lengths ℓ as shown in Figure 3 where the realization in the left panel (smaller correlation length) clearly shows more fluctuations on small scales than the right one (larger correlation length).

We prepare 5000 maps for each class, and every map only has one channel. We split this sample into 4500 for training and 500 for validation. Some model parameters including filter number, filter size, and the highest correlation order to measure are predetermined to be 1, 31 × 31 pixels and 3, respectively. The filter size is chosen to be large enough to capture the correlation scales of both training classes. We then use `Optuna` for a grid search to optimize the hyperparameters in the model such that their combination can maximize the validation accuracy. The results are shown in Table 1. The model with the above hyperparameter values can return a validation accuracy of 99.2% after 100 training epochs.

We implement the regularization path analysis as discussed in the previous section and results are shown in the left panel of Figure 4.

It is clear that the second-order moment $c_1^{(2)}$ is the most important feature in this classification as $\beta_1^{(2)}$ is activated first. Once it is activated, both the training and validation accuracy increase from 50%, i.e., no classification power, to more than 90%. Afterward, both the training and validation accuracy experience a slow increase while the value of $\beta_1^{(2)}$ goes through a fine-tuning along the regularization path. At the activation of $\beta_1^{(1)}$ and $\beta_1^{(3)}$, the validation accuracy has almost reached 100% already. Therefore, we interpret these two activations as overfitting. This result is consistent with our expectation that the two classes of training data are completely distinguished based on the second-order moment $c_1^{(2)}$ since for a Gaussian random field, one can characterize it completely by its expectation value and covariance. Since the two classes of training data have the same expectation value, the dominant power to distinguish them should come from the quantity that relates to the covariance, i.e., the second-order moment $c_1^{(2)}$. The right panel in Figure 4 shows that C3NN successfully captures this statistical property by using its trained filter to map the two classes of Gaussian random fields into two distinct distributions in the $\{c_1^{(n)}\}$ space, particularly along the marginalized dimension of $c_1^{(2)}$. It is based on these different distributions that the classifier learns the decision boundary between the two classes.

[7] https://geostat-framework.readthedocs.io/projects/gstools/en/stable/

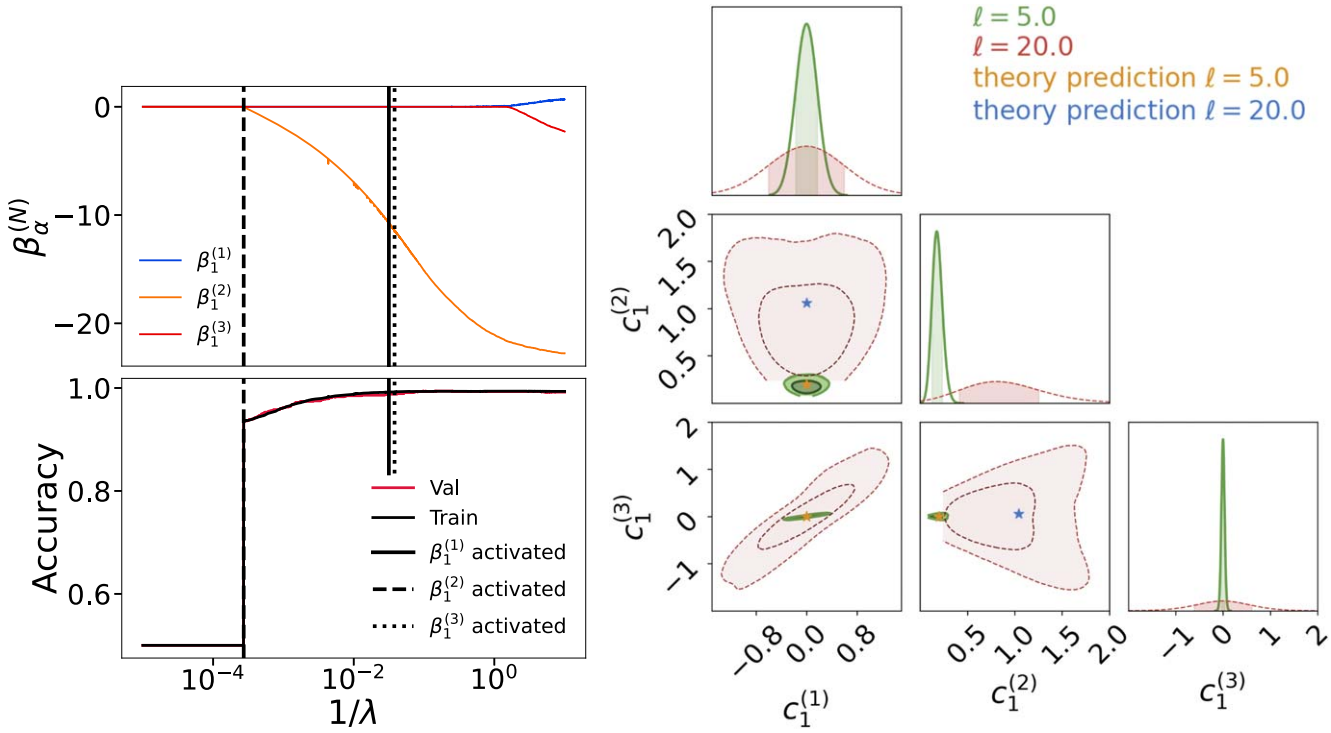

Figure 4. Left: the regularization path analysis result of C3NN trained on two classes of Gaussian random fields with correlation lengths $\ell = 5.0$ and 20.0 (see Figure 3). The upper panel shows the evolution of the $\beta_\alpha^{(N)}$ coefficients along with the decrease in the regularization strength λ in Equation (11) (an increase in $1/\lambda$). The bottom panel shows the corresponding changes in training and validation accuracy. Different types of vertical lines indicate the first λ value at which each $\beta_\alpha^{(N)}$ becomes nonzero (i.e., gets activated). Right: contours of moment $c_\alpha^{(N)}$ distributions mapped by passing the full training data set into the moment map generator and spatial averaging after the first round of training. The diagonal subplots show the marginalized distributions of each order of moments from the two training classes. The green contours represent the moment distribution of the Gaussian random field with $\ell = 5.0$ and the red contours represent that with $\ell = 20.0$. The orange and blue stars are the theoretical predictions of different $c_\alpha^{(N)}$ for the two classes based on the trained filter weights and analytical expressions of correlation functions for a Gaussian random field.

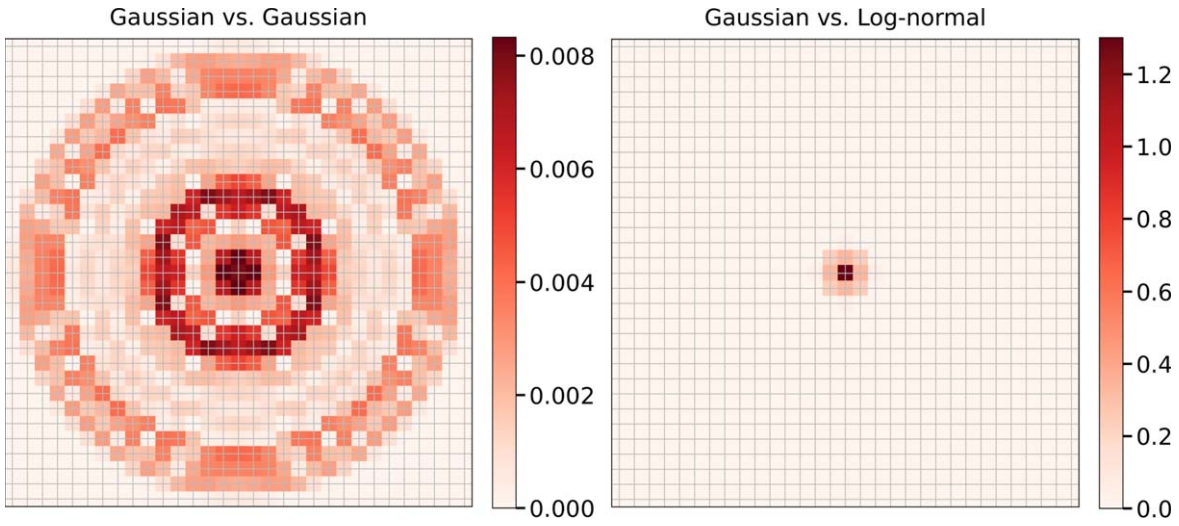

Figure 5. Left: the trained filter of C3NN classifying two Gaussian random fields with different correlation lengths (see Figure 3). Right: the trained filter of C3NN classifying Gaussian and log-normal random fields with the same power spectrum but different higher-order NPCFs (see Figure 5). In both cases, the filter is of size 31×31 pixels and all weights are non-negative. Both color bars show the scales of the filter weights.

Looking at the filter in the left panel of Figure 5 we see that the learned filter weights show a rotationally symmetric pattern as required by our architecture. There are two prominent annuli on different scales. This indicates that there is a correspondence between the annulus scales in the filter and the characteristic scales in the Gaussian random fields. The outer annulus has a much larger radius than the correlation length 5.0; hence, it can measure correlation function signals that only appear in the training data class with $\ell = 20.0$. The inner annulus, which has a radius of approximately 5 pixels, is more efficient in extracting correlation function signals from the training data class with $\ell = 5.0$. As $\beta_1^{(2)}$ is first activated, we can quantify the relative importance of 2PCF with different separations following the approach described in Section 2 and Figure 2.

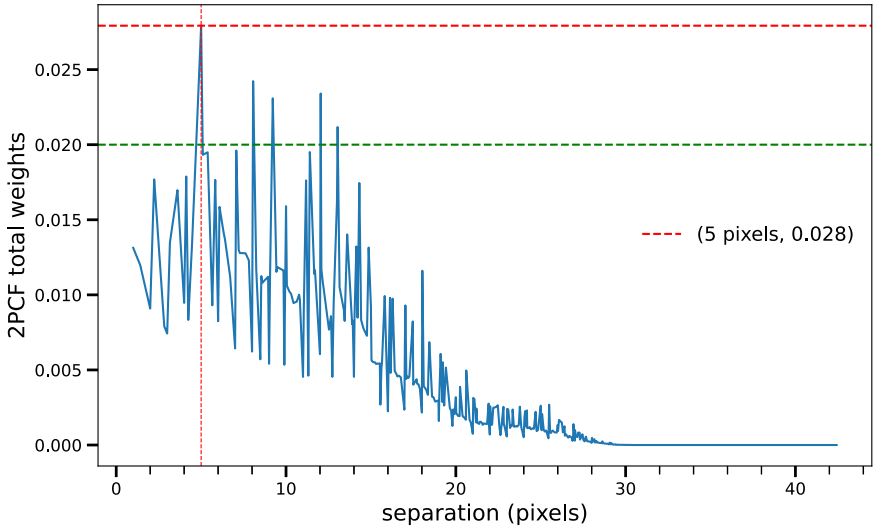

Figure 6. All possible 2PCF configurations from the filter in the left panel of Figure 5 with their corresponding total weights. The green horizontal dashed line indicates a total weight of 0.02 and one can observe that there are five separations with weight contributions exceeding this threshold. These separations are 5.0, 8.06, 9.22, 12.04, and 13.04 pixels, respectively. The intersection of red dashed lines marks the separation with the largest weight contribution (i.e., 5.0 pixels with a total weight of 0.028).

We show in Figure 6 all possible 2PCF separations together with their total weights.

We find that the most important 2PCF that contributes to the classification is exactly the one with a separation of 5.0 pixels with a corresponding total weight is 0.028.

The above filter can even be used to compute $\{c_1^{(1)}, c_1^{(2)}, c_1^{(3)}\}$ analytically. We know theoretically for the training Gaussian random fields, their expectation value and all connected 3PCFs should be 0.0. Therefore, $c_1^{(1)}$ and $c_1^{(3)}$ for both training data classes would be 0.0. For $c_1^{(2)}$, first, we select all possible 2PCF configurations within the filter and calculate the total weight for each of them, then by following Equation (4) we combine these total weights with the corresponding 2PCF amplitudes that can be computed analytically for Gaussian random fields. The equation we use in this case is the one used by `GSTools` to create the Gaussian maps (Webster & Oliver 2007):

$$\gamma(r) = \sigma^2\left(1 - \exp\left[-\left(\frac{sr}{\ell}\right)^2\right]\right), \tag{12}$$

where γ represents the 2PCF amplitude at a given separation r. The standard rescale factor is $s = \sqrt{\pi}/2$. The variance amplitude is denoted by σ^2 and $\ell$is the correlation length of the Gaussian random field. The results are $c_1^{(2)} = 0.199$ for $\ell = 5.0$ and 1.049 for $\ell = 20.0$. They are marked by orange and blue stars in the right panel of Figure 4 and are well contained within the 1σ confidence interval of the distributions mapped by C3NN from the input maps of the two classes. This confirms that the moments $c_\alpha^{(N)}$ measured by C3NN can be interpreted in terms of correlation functions.

3.2. Test Case of Gaussian and Log-normal Distributed Weak-lensing Convergence Fields

In cosmology, one of the central quantities of interest is the 3D matter density contrast field δ^{3D} whose evolution is governed by the interplay of large-scale gravitational and small-scale baryonic feedback processes. One way to probe the δ^{3D} field is through gravitational lensing. Gravitational lensing is the bending of light rays coming from background light sources, e.g., galaxies, by the gravitational potential of foreground lens objects, e.g., galaxy clusters, resulting in our observation of shifted, magnified, and distorted images. In most cases where a light ray passes far away from the centers of galaxy clusters where the gravitational potential is weak, it only experiences slight deflections by many foreground lenses distributed along its trajectory to us. Hence, we only observe a minuscule distortion in the image of the source. This is usually the case when light from sources passes through the foreground matter distribution of the large-scale structure (LSS). The distortion is so small that one can only see this effect on a statistical basis through correlations of the alignment of the weakly but coherently distorted images of many background source galaxies. This is known as weak gravitational lensing and serves as a probe for investigating the distribution of matter in the LSS. This field can be interpreted as the *shear* caused by a weighted line-of-sight projection of the 3D matter density contrast field—known as the weak-lensing convergence field. The weak-lensing convergence field $\kappa(\boldsymbol{\theta})$ can be written as a line-of-sight projection of the 3D matter density contrast field (Bartelmann & Schneider 2001)

$$\kappa(\boldsymbol{\theta}) = \int \mathrm{d}\chi\, q(\chi)\delta^{3D}(\chi\boldsymbol{\theta}, \tau_0 - \chi), \tag{13}$$

where χ is the radial comoving distance, $\boldsymbol{\theta}$ is a 2D planar vector denoting positions on the sky and τ_0 is the conformal time today. The term $q(\chi)$ is the projection kernel (also known as lensing efficiency) and for the case when all source galaxies are located in a Dirac-δ function-like tomographic bin at χ_s it becomes

$$q(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2}\frac{\chi}{a(\chi)}\frac{\chi_s - \chi}{\chi_s}; \qquad \text{with } \chi \leqslant \chi_s. \tag{14}$$

However, it is straightforward to write $q(\chi)$ for an ensemble of sources (instead of a single source) in terms of a distribution in a tomographic redshift bin following a normalized probability

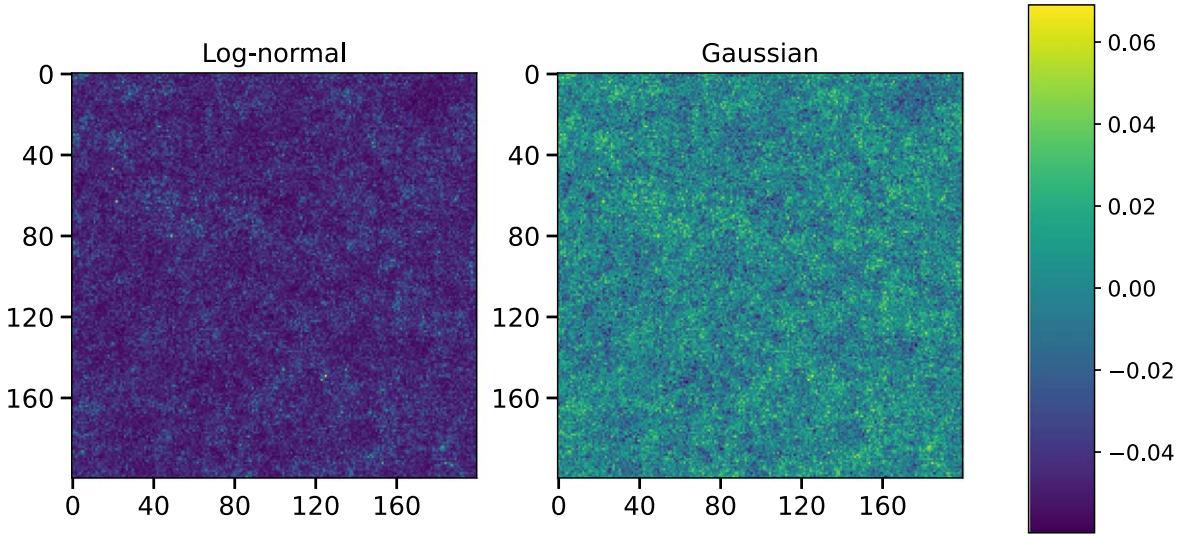

Figure 7. Simulated Gaussian and log-normal square maps with the same underlying cosmology and power spectrum but different higher-order correlation functions. Both classes of maps have a size of 200 × 200 pixels spanning an area of 20 × 20 deg^2 on the spherical sky maps from which they are partitioned (see Section 3.2 for details). The color bar indicates the field amplitude.

density function (PDF) $p(\chi')$ (Kilbinger 2015):

$$q_\kappa(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2}\frac{\chi}{a(\chi)}\int_\chi^{\chi_\mathrm{lim}} \mathrm{d}\chi' p(\chi')\frac{\chi'-\chi}{\chi'} \times \text{ with } \chi \leqslant \chi_\mathrm{lim}. \quad (15)$$

In the equations above Ω_m is the total matter density parameter of the Universe today, H_0 the Hubble parameter today, a the scale factor, and c the speed of light.

In this section, we are interested in applying C3NN to a cosmological scenario where we study how well our architecture can distinguish simulated convergence fields that follow two different distributions: one a Gaussian random field and another a log-normal random field. Log-normal random fields have been extensively studied and shown to closely approximate the one-point PDF of the convergence field (Hilbert et al. 2011; Xavier et al. 2016). In order to simulate these sets of maps, we use the publicly available `FLASK` tool (Xavier et al. 2016) to create realizations of Gaussian/log-normal fields on the celestial sphere with a Dirac-δ source redshift distribution at $z = 1.0334$ in `HEALPix` format with `NSIDE` = 2048 (Górski et al. 2005; Zonca et al. 2019). The detailed process of creating the maps is described in Section 4.2 of Halder et al. (2021). The generated Gaussian and log-normal maps share the same power spectrum (or the 2PCF) but differ in the higher-order correlations of the field, which in the case of the log-normal field is induced by the so-called log-normal shift parameter (see Xavier et al. 2016). This toy scenario is hence well set for us to apply C3NN in a cosmological context and study its power in tapping into higher-order information in the lensing convergence field.

With the full-sky simulation maps from `FLASK`, we partition them into nonoverlapping square patches on which we can implement C3NN. We adopt the approach in Ferlito et al. (2023; see their Section 2.4 and Figure 1) for partitioning the spherical sky. Each square map is 20 × 20 deg^2 with an angular pixel resolution of 6′. In Figure 7 we show one example from Gaussian and log-normal random field respectively. Both random fields have the same underlying cosmological parameters and power spectrum. As in Section 3.1, we prepare 5000 maps for each class where 4500 is for training and 500 for validation. Each map only has one channel, which in this case is a single source redshift bin lensing convergence field. We fix the filter number, filter size, and the highest correlation order to be 1, 31 × 31 pixels and 4, respectively. The optimized hyperparameters from `Optuna` are shown in Table 2. With these parameters, the model is able to reach a validation accuracy of 100% after 100 training epochs.

Table 2
Four Optimized Hyperparameters in C3NN Trained on Gaussian and Log-normal Random Fields[a]

Parameter	γ	Learning Rate (lr)	Learning Rate Decaying Ratio (ϕ)	Optimizer
Value	0.0026	0.15	0.66	"RMSprop"

Notes. γ is the regularization strength in Equation (10). We use a learning rate scheduler that decays the initial learning rate (lr) of each parameter group by a factor ϕ after every epoch. We also simultaneously search for the best optimizer and obtain RMSprop (Graves 2013).
[a] See Figure 7.

The regularization path analysis in the upper left panel of Figure 8 shows that the third-order moment $c_1^{(3)}$ is activated first (dotted line) and contributes most to the classification accuracy. This is consistent with our expectation as the Gaussian and log-normal random fields that we have constructed should only start to differ at the 3PCF level. The second and fourth-order moments only become activated when both training and validation accuracy have reached 100% so that their contributions can be counted as overfitting. One notices that $c_1^{(1)}$, which according to Equation (3) is the average of the field, is critical in complementing $c_1^{(3)}$ in the classification between the two classes. The reason behind this is explained in the bottom panel of Figure 8 where we show the marginalized distributions of feature vectors from training data in the $c_1^{(1)} - c_1^{(3)}$ plane. As described in Section 2, we can visualize the evolution of the

decision boundary in the projected $c_1^{(1)} - c_1^{(3)}$ plane, along with the changes in the regularization strength λ. At the very beginning of $\beta_1^{(3)}$ activation, the decision boundary can only be drawn based on the marginalized distribution of $c_1^{(3)}$ since all the other dimensions are still suppressed by the large λ value. Once $\beta_1^{(1)}$ is activated, the decision boundary can be probed within a two-dimensional plane. It turns out that the mapped distributions of $\{c_1^{(1)}, c_1^{(3)}\}$ from the two random fields have the same degeneracy so that C3NN can rotate the decision boundary to achieve higher classification accuracy. This process is exactly depicted in the rotation from the green solid line ($\beta_1^{(1)}$ activation) to the blue dashed one ($\beta_1^{(2)}$ activation) in the figure. The blue decision boundary completely separates the two classes. Activation of $\beta_1^{(2)}$ together with others until the end of the regularization path makes almost no difference to the classification.

As in the previous discussion, we can also combine the filter weights (see the right panel of Figure. 5) and the theoretical expressions for NPCFs to predict the corresponding moments $c_1^{(N)}$ for both the Gaussian and log-normal fields. Any 2PCF of a given angular separation within the filter size is the same for both random fields and can be computed from the power spectrum. For the Gaussian random field, all 4PCFs directly depend on the 2PCFs within the quadrilateral configurations while all its odd correlation functions are zero. For the log-normal random field, any of its 3PCFs and 4PCFs can be written in terms of the 2PCF within a given configuration and the log-normal shift parameter mentioned above. Readers who are interested in the exact equations are referred to Appendix B of Hilbert et al. (2011). The calculations give us $\{c_1^{(1)}, c_1^{(2)}, c_1^{(3)}, c_1^{(4)}\} = \{0.0, 2.414 \times 10^{-4}, 0.0, 1.315 \times 10^{-8}\}$ for the Gaussian random field and $\{c_1^{(1)}, c_1^{(2)}, c_1^{(3)}, c_1^{(4)}\} = \{0.0, 2.414 \times 10^{-4}, 8.078 \times 10^{-7}, 1.613 \times 10^{-8}\}$ for the log-normal random field. These are marked by the orange and blue stars in the upper right panel of Figure 8, respectively. Once again, the calculations fall well within the bulk of the moments distribution output by C3NN, which we obtain by passing all the training data through our trained model. The right panel of Figure 5 shows that the dominant weights that hold the classification power mainly concentrate in the central part of the filter. Qualitatively, we can understand this as statistically speaking, log-normal random fields asymptotically resemble Gaussian random fields when smoothed on large scales. The differences between the two fields are mainly captured through the higher-order correlation functions which become increasingly more significant on small scales. An efficient way to distinguish the two classes therefore would be to select correlation functions beyond 2PCFs on small scales. Then it is no wonder that out of all triangular configurations that can be selected from the filter constituting $c_1^{(3)}$, C3NN considers the smallest one the most important in aiding the classification, which is the isosceles triangle with two side lengths equal to $6'$ (neighboring pixel separation) and the third equal to $6\sqrt{2}$ arcmin.

4. Results from *N*-body Cosmological Simulations

In Sections 3.1 and 3.2, we have shown that C3NN can classify different random fields by correctly capturing the underlying statistical properties. In this section, we are interested in applying the framework to weak-lensing fields from realistic *N*-body simulations to investigate whether a successful classification is still possible and what statistical features C3NN extracts from the training data. We shall also investigate the impact of shape noise and smoothing scale of the simulated maps on the model's performance.

4.1. Weak-lensing Convergence Maps from N-body Simulations

In the rest of this work, we use weak-lensing convergence maps from the publicly available `CosmoGridV1`[8] simulations (Fluri et al. 2022; Kacprzak et al. 2023). We focus on the simulation products around the fiducial cosmology adopted by `CosmoGridV1`, which has the following parameter values: $\Omega_m = 0.26$, $\Omega_b = 0.0493$, $\sigma_8 = 0.84$, $w_0 = -1.0$, $n_s = 0.9649$, and $H_0 = 67.3\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. The two classes of maps we use as training data share the same parameters as the fiducial one except for w_0. One class of simulation has $w_0 = -1.05$, which is in the realm of "phantom dark energy" (Ludwick 2017) and the other has $w_0 = -0.95$, which is characterized as "quintessence" (Tsujikawa 2013). For each simulation set, there are 200 independent full-sky simulations. Each simulation contains a series of light cone shells along the redshift until $z = 3.5$. We project these shells following the four source galaxy redshift distributions of the Dark Energy Survey Year 3 (DES Y3) data release (see Figure 11 in Myles et al. 2021) to produce the corresponding full-sky weak-lensing convergence maps. Readers can refer to Sgier et al. (2019, 2021) and Reeves et al. (2024) for details of the methodology we adopt for the shell projection and convergence map generation. These maps from the four different source redshift distributions constitute the four channels in the input data and enable us to perform a tomographic analysis with C3NN.

On top of these noiseless maps, we also add shape noise to make more realistic tests. We directly add noise to each pixel in the full-sky convergence map by sampling independently from a Gaussian distribution with zero mean and variance $\sigma^2 = \epsilon^2/(A n_g)$ where ϵ is the dispersion of galaxy ellipticity, n_g is the galaxy surface number density and A is the pixel area. We adopt DES Y3 values for ϵ and n_g, which are 0.255, and $\{1.476, 1.479, 1.484, 1.461\}$ arcmin^{-2} for each source redshift bin, respectively (Jeffrey et al. 2021). The pixel area of the full-sky convergence map is 2.95 arcmin^2. Besides shape noise, we also take into account the smoothing of the input maps. This is motivated by the concern that in real observations we may need to discard small-scale measurements of correlation functions since they are impacted by the baryonic feedback effects that we cannot realistically model (Abbott et al. 2022) in *N*-body gravity-only simulations. For map smoothing, we apply Gaussian kernels with full widths at half maximum (FWHM) of $\{10', 20', 30', 40'\}$ to the tomographic full-sky maps.

In this analysis, we use the same partition method mentioned in the previous section to obtain 5800 nonoverlapping square maps for each cosmology and source redshift bin. We split 90% of the data for training and 10% for validation. Regardless of the smoothing scales, the square map always has an area of 20×20 deg^2 and a size of 200×200 pixels. The difference of w_0 between the two cosmologies while other parameters are fixed to the same values is about $1/6$ of the 1σuncertainty as constrained by the state-of-the-art DES Y3 analysis (Abbott et al. 2022) marginalized over other cosmological parameters as well as nuisance parameters. This constraint comes from a combination of cosmic shear and galaxy clustering, the so-called 3×2pt analysis, within the w_0CDM model. Only when

[8] http://www.cosmogrid.ai/

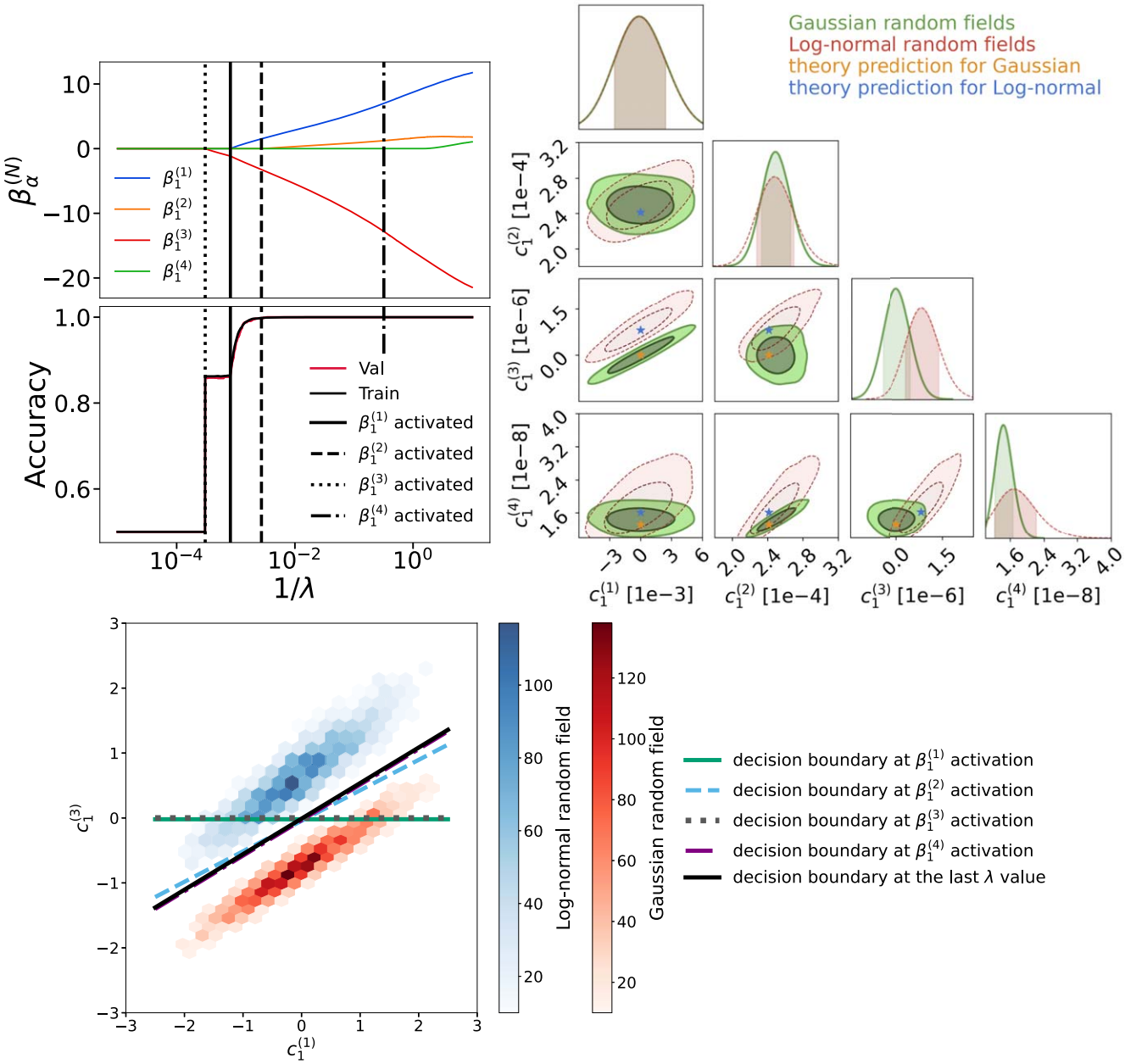

Figure 8. Upper left: the regularization path analysis applied to C3NN trained on two classes of Gaussian and log-normal random fields (see Figure 7). The upper part shows the evolution of the $\beta_\alpha^{(N)}$ coefficients along with the decrease of the regularization strength λ in Equation (11) (an increase in $1/\lambda$). The lower part shows the corresponding changes in training and validation accuracy. Different types of vertical dashed lines indicate the first λ value at which each $\beta_\alpha^{(N)}$ becomes nonzero (i.e., gets activated). Upper right: contours of moment $c_\alpha^{(N)}$ distributions mapped by passing the full training data set into the moment map generator and spatial averaging after the first round of training. The diagonal subplots show the marginalized distributions of each order of moments from the two training classes. The green contours represent the moment distribution of the Gaussian random field and the red contours represent that of the log-normal random field. The orange and blue stars are the theoretical predictions of different $c_\alpha^{(N)}$ for the two classes based on the trained filter weights and analytical expressions of correlation functions for the two fields. Bottom: the distribution of $\{c_1^{(1)}, c_1^{(3)}\}$ with decision boundaries from the classifier. The values of both moments are batch normalized compared to the corresponding upper right panel. The blue distribution is from the log-normal random field and the red one is from the Gaussian random field. Lines with different colors and styles corresponding to the vertical lines in the upper left panel represent the decision boundaries of C3NN at activations of $\beta_\alpha^{(N)}$ coefficients up to $N = 4$, as well as the boundary at the last regularization strength value.

data from 3×2pt and the external low-redshift probes (including baryon acoustic oscillations, supernovae Type Ia, and redshift space distortions) are combined can the marginalized 1σ uncertainty of w_0 be constrained approximately to 0.1. Based on these contexts, it would require multiple powerful cosmological probes to accurately distinguish between the two dark energy scenarios we are testing. Hence, in the context of the above training setups, we would like to address the following questions: (i) with weak-lensing convergence alone (i.e., no external probes), how much generic improvement can higher-order correlation functions as captured by the moments $c_\alpha^{(N)}$ contribute to differentiating the two dark energy equation of state parameters; (ii) what is the relative importance of each order of moment in such a classification task; (iii) how do shape noise and different smoothing scales affect the model performance.

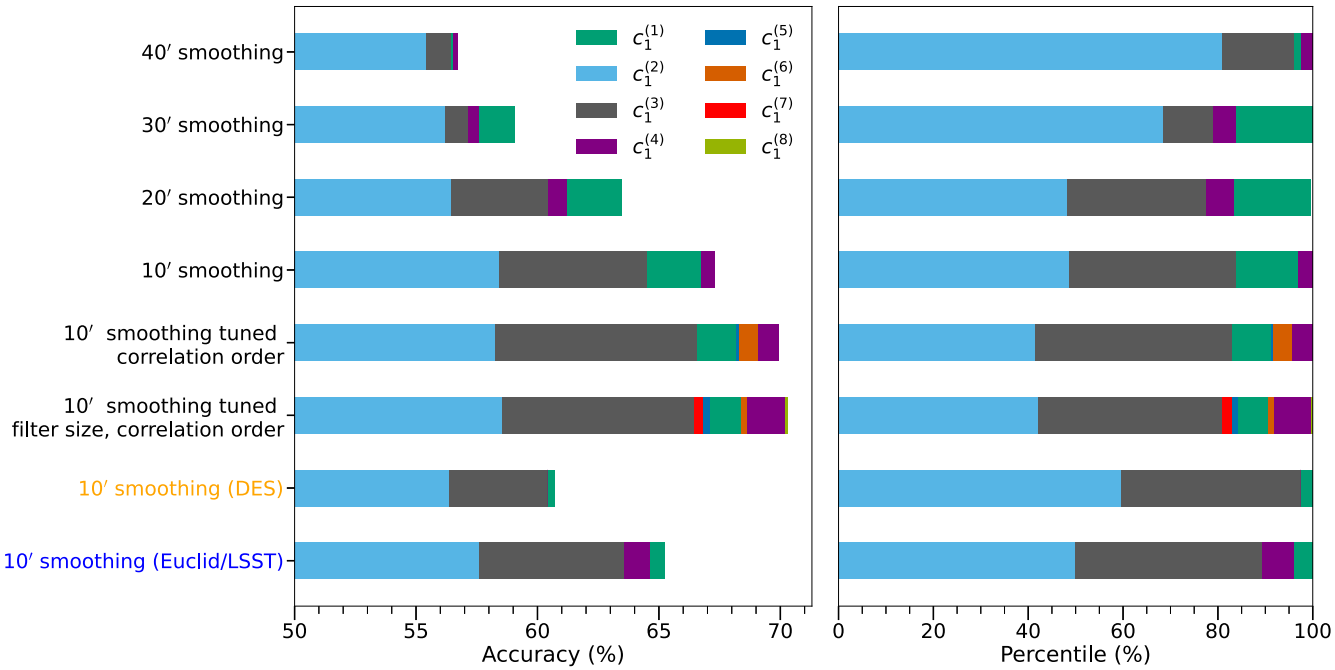

Figure 9. Left: the validation accuracy from regularization path analysis (see Section 2) of C3NN trained on two classes of tomographic weak-lensing convergence maps with different w_0 values (-1.05 and -0.95). The four source redshift distributions follow that of DES Y3 analyses. Test names in black indicate that the analyses are with noiseless data. The first four tests deviate from each other in the FWHM of the Gaussian kernels used for smoothing. In the fifth row of the bar chart, we give the model one extra free parameter by allowing `Optuna` to optimize the highest correlation order while still fixing the filter number and size. In the sixth row, we further add filter size to the set of tuning parameters. The sixth and seventh rows of the bar charts are results from different noisy training data with the same model setups as the corresponding noiseless test with 10' smoothing. The name in orange represents that the test is with noisy data produced with DES Y3 ϵ and n_g values while the one in blue uses training data with n_g rescaled to 7.5 arcmin^{-2} for each tomographic bin (as expected for a Stage IV survey like Euclid/LSST). In each single row of the bar chart, from left to right, the different color components form the sequence of the activation of moments in the regularization path analysis. The length of every color component represents the improvement in validation accuracy added by the current activation combined with all the previous ones. The total validation accuracy starts from 50% (random classification) and ends with the summation of all color components. Right: the percentage of improved validation accuracy from the activation of different moments with respect to the total validation accuracy in each test. It shares the same color labels and test names as the left panel. Color components in each bar are also arranged according to the activation sequence.

4.2. Noiseless Tomographic Training Results with Varying Smoothing Scales

With the above-mentioned noiseless training data, we apply C3NN to them at each smoothing scale and summarize the results in the left panel of Figure 9. To keep the model simple, at the beginning we freeze the filter number, filter size, and the highest correlation order (1, 31×31 pixels and 4, respectively) while changing the smoothing scales of the training data. The tuned hyperparameters are shown in Table 3.

From the figure we can observe that the second and third-order moments are the major contributors to the classification. The activation of $\beta_1^{(3)}$ (onset of the gray bars) always follows that of $\beta_1^{(2)}$ (end of the light blue bars), indicating that when one only searches for a single moment to distinguish the above two w_0CDM cosmologies, the second-order moment $c_1^{(2)}$ is most effective. However, once the third-order moment $c_1^{(3)}$ is activated and combined with $c_1^{(2)}$, it can significantly improve the validation accuracy. This is analogous to what Gong et al. (2023) found from the perspective of parameter inference that w_0 constraint can be particularly tightened by combining 2PCF and a compressed version of the 3PCF (called the integrated 3PCF). As for the other higher-order moments, here $c_1^{(4)}$ specifically, can only bring minor validation accuracy improvement compared to $c_1^{(2)}$ and $c_1^{(3)}$. This can be explained by either the cosmic variance caused by the lack of sampling of 4PCF configurations within the filter due to the limited 400 deg^2 simulated map area resulting in a lower signal-to-noise compared to lower order moments, or the map resolution, which prohibits the higher-order moments in probing smaller and smaller scales where their importance become increasingly more significant. Both effects can lead to the weakening of the classification capability of higher-order moments. Another explanation for the weak classification power contribution from $c_1^{(4)}$ under the current setups is that there is not much information contained in the fourth-order moment of weak-lensing convergence that can be added to differentiate between these two different cosmologies. One supportive evidence is that the increased validation accuracy after activating $\beta_1^{(4)}$ is not as significant as the increment brought by $c_1^{(3)}$.

Table 3
Same as Table 1, the Optimized Hyperparameters of C3NN for Varying Smoothing Scales Applied to Noiseless Tomographic Weak-lensing Convergence Maps with Different w_0 Values (-1.05 and -0.95)[a]

	γ	Learning Rate (lr)	Learning Rate Decaying Ratio (ϕ)	Optimizer
40′ smoothing	0.06	0.018	0.58	"Adam"
30′ smoothing	1.10	0.041	0.24	"Adam"
20′ smoothing	0.0023	0.043	0.34	"Adam"
10′ smoothing	0.0022	0.064	0.14	"Adam"

Notes. Here, we fix the filter number, filter size, and the highest correlation order.
[a] See Section 4.2.

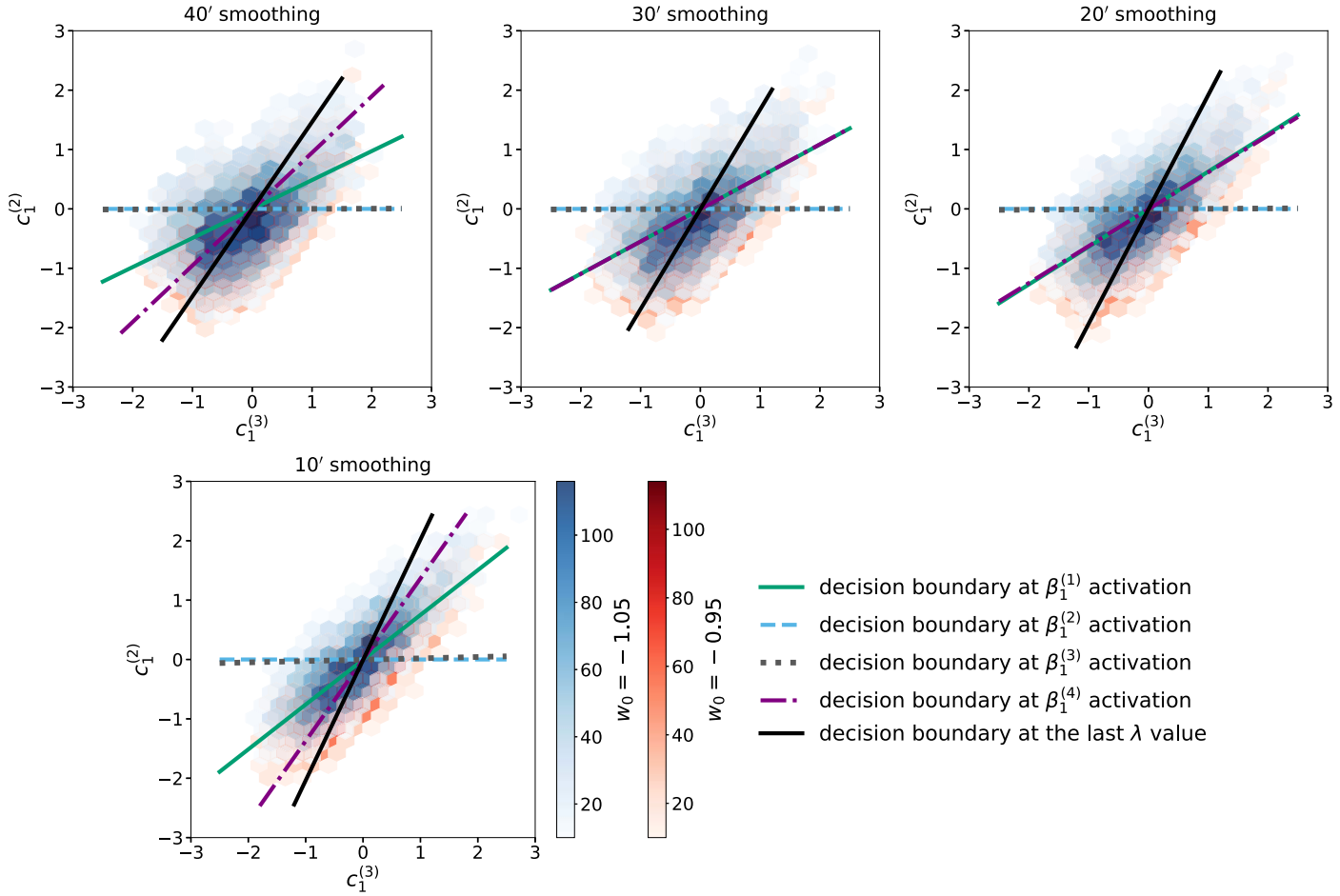

Figure 10. The batch-normalized $\{c_1^{(3)}, c_1^{(2)}\}$ joint distributions mapped by C3NN trained on simulated noiseless weak-lensing convergence data for different smoothing scales. The blue distribution represents the mapping from data with $w_0 = -1.05$ and the red distribution represents the mapping from data with $w_0 = -0.95$. Lines with different colors and styles are the decision boundaries learned by the classifier during the regularization path analysis at the activation of different orders of moments.

One should also notice that although the amount of validation accuracy associated with $c_1^{(2)}$ (i.e., the size of the light blue bars in the first four rows in the left panel of Figure 9) increases along with the decrease of the smoothing scale, the results are still similar to each other i.e., they make up 6%–8% of the total validation accuracy. In comparison, the additional validation accuracy brought by $c_1^{(3)}$ (size of the gray bars) depends more drastically on the smoothing scale. The result from 40′ smoothing is approximately six times smaller than the result from 10′ smoothing. This different dependence indicates that the joint distribution of $\{c_1^{(2)}, c_1^{(3)}\}$ mapped by the trained model from two classes has a larger overlapping region as we increase the smoothing scale. Similar to the bottom panel in Figure 8, we show in Figure 10 the batch-normalized $\{c_1^{(2)}, c_1^{(3)}\}$ joint distributions for the different smoothing scales. It is clear that the initial activation of $\beta_1^{(2)}$ leads to the horizontal decision boundary in each panel. Those decision boundaries then gradually rotate to the direction along the degeneracy of $c_1^{(2)}$ and $c_1^{(3)}$ after other activations. However, as the smoothing scale increases, the distributions from two classes in $c_1^{(2)} - c_1^{(3)}$ plane become more extended along the orthogonal direction of the degeneracy. This naturally enlarges the overlapping region of the two distributions and reduces the validation accuracy C3NN can obtain with the activation of $\beta_1^{(3)}$.

To ensure that our validation accuracy is not limited by the fixed model parameters, i.e., filter number, filter size, and the highest correlation order, we alternatively treat them as hyperparameters and optimize them using `Optuna`. The 5th and 6th rows of the bar charts in the left panel of Figure 9 show the results of tuning the highest correlation order and simultaneously tuning it with the filter size, respectively (see Tables 4 and 5 for the results).

In both cases where C3NN can measure moments beyond fourth order, the validation accuracy only shows little improvement. The dominant contribution still comes from the joint distribution of $\{c_1^{(1)}, c_1^{(2)}, c_1^{(3)}\}$. We even further allow the tuning process to vary the number of filters in the model but it does not help to substantially increase the validation accuracy either. We do not show the corresponding result in the plot as the optimized filter number is 5 and the highest correlation order is 8 according to `Optuna`, which is difficult to fit into the plotting space. This is another sign suggesting that moments of convergence beyond the third order in total may not contain sufficient information to classify different dark energy equation of state parameters at the desired precision of $\Delta w_0 = 0.1$ within the context of our test setup. To address this issue, one promising approach would be to include more observables such as projected galaxy number density into the training data and measure the cross correlations among these

Table 4
Optimized Hyperparameters Including the Highest Correlation Order for the C3NN Model Trained on Simulated Noiseless Tomographic Convergence Maps with 10′ Smoothing[a]

Parameter	Correlation Order	γ	Learning Rate (lr)	Learning Rate Decaying Ratio (ϕ)	Optimizer
Value	6	0.0039	0.18	0.75	"Adam"

Note.
[a] See Section 4.2.

Table 5
Same as Table 4 but also Allowing the Optimization of the Filter Size along with the Highest Correlation Order in the Training

Parameter	Filter Size	Correlation Order	γ	Learning Rate (lr)	Learning Rate Decaying Ratio (ϕ)	Optimizer
Value	11×11	8	0.002	0.019	0.93	"Adam"

quantities rather than keep adding higher-order moments beyond the third order in the data vector. We defer this study to future works.

From the above discussion, we show that C3NN can be a useful tool to investigate the sensitivity of moments with respect to the w_0 parameter within the w_0CDM model. Broadly speaking, it is possible to extend this functionality to any given binary classification tasks of two cosmological models that can impact differently the matter distribution in our Universe, e.g., between the standard cold dark matter and other exotic models such as wave dark matter (Hui 2021) or self-interacting dark matter (Tulin & Yu 2018). Moreover, it can provide a quantitative understanding of the relative importance of different orders of moments, which are directly related to correlation functions, in classifying different cosmological parameters or models. All exact numeric values in Figure 9 may vary slightly from one evaluation to another, but the comparison among the total validation accuracy of different tests is certain. Moreover, we can always confirm the dominance of second and third-order moments as the first and second-activated features that carry, relatively, the most significant information in the classification.

4.3. Noisy Tomographic Training Results

For simplicity, we only show the training result from data at 10′ smoothing scale with shape noise with DES Y3 dispersion of galaxy ellipticity and surface number density. This is the seventh row of the bar chart in the left panel of Figure 9. The validation accuracy of the other smoothing scales with the same noise addition gradually decreases as in the noiseless cases. Compared to the corresponding noiseless result (fourth row), here the total validation accuracy reduces by approximately 6.6%, and $c_1^{(3)}$ contributes less significantly to the classification power. Moreover, the contributions from both $\beta_1^{(1)}$ and $\beta_1^{(4)}$ almost vanish in the activation sequence. Overall, the effects of shape noise are similar to that of smoothing: it mainly decreases the total validation accuracy through moments beyond second order.

Besides adding the shape noise associated with DES Y3 analyses, we also rescale the galaxy surface number density mentioned in Section 4.1 to approximately the value that would be provided by upcoming Stage IV lensing surveys such as Euclid (Euclid Collaboration et al. 2022) and Vera Rubin Observatory's LSST (LSST Dark Energy Science Collaboration 2012). We adopt the new total galaxy surface number density to be 28 galaxies arcmin^{-2}, which is between those of the two surveys (Euclid Collaboration et al. 2020; Chang et al. 2013). We retain the four tomographic bins in the training data and equally divide this new number density among all of them. The original shape noise thus should be rescaled by a factor of $\{0.444, 0.444, 0.445, 0.441\}$ for each tomographic bin, respectively. Through this, we obtain a crude estimation of C3NN performance on next-generation survey data. One should bear in mind that in this test the map size of the training data (20×20 deg^2) is much smaller than the actual footprints of these Stage IV surveys and that we also use fewer tomographic bins. The result is shown in the eighth row of the bar chart in the left panel of Figure 9. Clearly, with a higher signal-to-noise ratio, the total validation accuracy exceeds that of noisy simulated data using DES Y3 shape noise parameters. Out of all moments, $c_1^{(3)}$ gains the largest relative growth in its classification power. Although relatively small compared to $c_1^{(2)}$ and $c_1^{(3)}$, we also find that $c_1^{(1)}$ and $c_1^{(4)}$ can bring non-negligible improvements to the total validation accuracy. Another point to observe is that though it is not as powerful as the model trained on the corresponding noiseless data as expected, there is not a great contrast between the results that these two models manifest. This implies that with the forthcoming Stage IV surveys, C3NN can potentially be close to reaching its best performance.

Another useful perspective to understand all the above results is shown in the right panel of Figure 9. The total validation accuracy of each test is used as a normalization constant, based on which we calculate the percentage of increased validation accuracy associated with each order moment. We notice that with larger smoothing scale or shape noise, the percentage of the second-order moment increases while that of higher orders, here particularly third and fourth-order, decreases. This is consistent with our previous findings that the power of second-order moment is robust against varying smoothing scales or shape noise but higher-order moments can be significantly impacted. Another point the right panel helps to stress is that $c_1^{(3)}$ is powerfully complementary to $c_1^{(2)}$, e.g., in the test case of noisy data with Euclid/LSST galaxy surface number density, the combination of $c_1^{(3)}$ and $c_1^{(2)}$ almost double the classification ability of the model compared

to $c_1^{(2)}$ alone. Unfortunately, we do not observe this feature in moments beyond third order within our test setups.

5. Summary and Conclusion

C3NN, originally proposed by Miles et al. (2021) in the context of correlated quantum matter, is an interpretable machine-learning architecture that we have introduced for cosmological analyses. It is composed of two parts (see Figure 1). The first part comprises a CNN-based N-point moment map generator that outputs a series of statistics $c_\alpha^{(N)}$, which we call N-point moments and have an explicit mathematical connection to traditional NPCFs that we are familiar within cosmology. In this part of the model we first perform a single convolution on the input data and then without applying a nonlinear transformation as the usual practice, we construct moment maps directly from the initial convolved map through a recursive procedure (see Equation (2)). We adapt the filters of the convolution to be rotationally invariant (Cesa et al. 2022) such that the output remains stationary for arbitrary rotations of the input. This benefits the training data efficiency of our analyses and the interpretation of the trained filter weights. The second part of the model, which is a moment-map-based classifier, exploits these output moment maps to perform classification tasks by passing them through consecutive layers of spatial averaging, batch normalization, and logistic regression. We optimize all hyperparameters of C3NN using the package `Optuna`.

The interpretability of C3NN mainly exists in the following three aspects (Section 2): (i) the output moment $c_\alpha^{(N)}$ of a specific order N can be mathematically expressed in terms of the correlation function at the same order. This is unlike the often hard-to-interpret summary statistics extracted by conventional CNN models, (ii) through a regularization path analysis (Efron et al. 2004; Tang et al. 2014), which is integrated into the classifier part of our model, we can have a quantitative understanding of the relative importance of the different order moments $c_\alpha^{(N)}$ in contributing to the model's classification power. Moreover, (iii) we can investigate the trained filter weights by connecting individual pixels to form the configuration of any given NPCF. The filter weights for a given NPCF configuration directly allow us to rank different correlation function configurations within a given moment.

Since this is the first time we introduce this architecture into the field of cosmology, we focus on the relatively simple task of binary classification and implement C3NN on three tests including both proof of concept as well as application to simulated cosmological data. The results are summarized as follows:

1. *Gaussian random fields with different correlation lengths* (Section 3.1). The performance of C3NN to distinguish between two Gaussian random fields with different correlation lengths is in line with theoretical expectations. The model successfully achieves 100% classification based on the second moment $c_1^{(2)}$ alone. Its trained filter weights reflect the respective correlation lengths of the two classes of training data. We find good agreement between the distributions of each order of moment measured by C3NN and the corresponding theoretical prediction where we combine the trained filter weights and the analytical calculation of all possible correlation functions at each order.
2. *Gaussian and log-normal random fields with the same 2PCF but different higher-order moments* (Section 3.2). In this test, we apply C3NN to distinguish a Gaussian from a log-normal random field, which starts to differ from each other at the third moment. C3NN reaches a perfect classification by first activating and tapping into the information in the third moment of the field $c_\alpha^{(3)}$ and then via the field's first moment $c_1^{(1)}$ (arising due to finite field map sizes). The $c_1^{(3)}$ activation is consistent with the theoretical expectation since the two random fields share the same 2PCF but not the 3PCF. We further demonstrate that the activation of $c_1^{(1)}$ can separate the mapped distribution from the two classes in $c_1^{(1)} - c_1^{(3)}$ plane completely (see lower panel in Figure 8). Similar to the previous test, we also find good agreement between the distributions of each order of moment output by C3NN and the corresponding theoretical predictions.
3. *N-body simulated weak-lensing convergence fields with* $w_0 = -1.05$ and -0.95 (Sections 4.2 and 4.3). Finally, we investigate the classification power of C3NN on weak-lensing convergence maps between two simulated dark energy scenarios of $w_0 = -1.05$ and -0.95. C3NN can maximally reach a classification accuracy of around 70% with our training setups for a 400 deg^2 survey map. We find that the classification power of the second moment $c_1^{(2)}$ is robust against varying smoothing scales or shape noise but higher-order moments can be significantly impacted. In all our tests, $c_1^{(2)}$ and $c_1^{(3)}$ dominate the classification accuracy and the activation of $c_1^{(3)}$ is the major complementary component to $c_1^{(2)}$ (Figure 9). On the other hand, higher-order moments of convergence beyond the third order in total do not contain sufficient information to classify between the two different dark energy equation of state parameters, at least within the context of our investigation setup. This suggests that including more observables besides convergence in the input data and using C3NN (or the conventional 2PCFs and 3PCFs) to measure the cross correlations among these different observables may be a more efficient approach.

Overall, we show that the architecture of C3NN contains novel features that can be robustly and quantitatively interpreted and are rarely seen in the application of machine-learning tools to cosmology nowadays. Through multiple tests, we prove its validity and reveal its potential to provide us with physical insights. C3NN can have many interesting extensions such as simulation-based inference with the CNN-based N-point moment generator or 3D C3NN, which can measure moments in the context of galaxy clustering. The former can be built by replacing the moment-map-based classifier with a neural network-based likelihood-free-inference pipeline using methods such as normalizing flows or mixture density networks (Alsing et al. 2019). The latter can be constructed by extending the current 2D architecture in Equation (2) to 3D. All these can offer us promising opportunities for future exciting applications of C3NN in cosmology.

Acknowledgments

We would like to thank Alexandre Barthelemy, Gabriele Cesa, Sandrine Codis, Fulvio Ferlito, Oliver Friedrich, Laila Linke, and Tilman Tröster for very helpful comments and discussions at various stages of this project. We would also like to give credits to `CosmoGridV1`, which was created by Janis Fluri, Tomasz Kacprzak, Aurel Schneider, Alexandre Refregier, and Joachim Stadel at the ETH Zurich and the University of Zurich. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC2111 - 390814868. Some of the numerical calculations have been carried out on the ORIGINS computing facilities of the Computational Center for Particle and Astrophysics (C2PAP) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2094 - 390783311.

Software: `ChainConsumer` (Hinton 2016), `ESCNN` (Cesa et al. 2022), `FLASK` (Xavier et al. 2016), `healpy` (Górski et al. 2005; Zonca et al. 2019), `Optuna` (Akiba et al. 2019), `UFalcon` (Sgier et al. 2019, 2021; Reeves et al. 2024).

ORCID iDs

Zhengyangguang Gong https://orcid.org/0009-0002-7361-4073
Anik Halder https://orcid.org/0000-0002-0352-9351
David Gebauer https://orcid.org/0009-0002-2889-3704

References

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, PhRvD, 105, 023520
Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. 2019, arXiv:1907.10902
Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440
Anbajagane, D., Chang, C., Banerjee, A., et al. 2023, MNRAS, 526, 5530
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Bartelmann, M., & Schneider, P. 2001, PhR, 340, 291
Barthelemy, A., Halder, A., Gong, Z., & Uhlemann, C. 2024, JCAP, 2024, 060
Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, PhR, 367, 1
Bluecher, S., Kades, L., Pawlowski, J. M., Strodthoff, N., & Urban, J. M. 2020, PhRvD, 101, 094507
Burger, P. A., Porth, L., Heydenreich, S., et al. 2024, A&A, 683, A103
Cesa, G., Lang, L., & Weiler, M. 2022, in Int. Conf. on Learning Representations, https://openreview.net/forum?id=WE4qe9xlnQw
Chang, C., Jarvis, M., Jain, B., et al. 2013, MNRAS, 434, 2121
Chen, S.-F., Vlah, Z., & White, M. 2022, JCAP, 2022, 008
Cheng, S., & Ménard, B. 2021, MNRAS, 507, 1012
Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, MNRAS, 499, 5902
Crittenden, R. G., Natarajan, P., Pen, U.-L., & Theuns, T. 2002, ApJ, 568, 20
Dalal, R., Li, X., Nicola, A., et al. 2023, PhRvD, 108, 123519
Domínguez Sánchez, H., Margalef, B., Bernardi, M., & Huertas-Company, M. 2022, MNRAS, 509, 4024
Dvorkin, C., Mishra-Sharma, S., Nord, B., et al. 2022, arXiv:2203.08056
Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. 2004, Ann. Statist., 32, 407
Euclid Collaboration, Paykari, P., Kitching, T., et al. 2020, A&A, 635, A139
Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022, A&A, 662, A112
Ferlito, F., Springel, V., Davies, C. T., et al. 2023, MNRAS, 524, 5591
Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, PhRvD, 100, 063514
Fluri, J., Kacprzak, T., Lucchi, A., et al. 2022, PhRvD, 105, 083518
Friedrich, O., Gruen, D., DeRose, J., et al. 2018, PhRvD, 98, 023508
Gatti, M., Jain, B., Chang, C., et al. 2022, PhRvD, 106, 083509
Gong, Z., Halder, A., Barreira, A., Seitz, S., & Friedrich, O. 2023, JCAP, 2023, 040
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Graves, A. 2013, arXiv:1308.0850
Halder, A., & Barreira, A. 2022, MNRAS, 515, 4639
Halder, A., Friedrich, O., Seitz, S., & Varga, T. N. 2021, MNRAS, 506, 2780
Halder, A., Gong, Z., Barreira, A., et al. 2023, JCAP, 2023, 028
Heydenreich, S., Linke, L., Burger, P., & Schneider, P. 2023, A&A, 672, A44
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140
Hilbert, S., Hartlap, J., & Schneider, P. 2011, A&A, 536, A85
Hinton, S. R. 2016, JOSS, 1, 00045
Hui, L. 2021, ARA&A, 59, 247
Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, JCAP, 2020, 042
Jeffrey, N., Gatti, M., Chang, C., et al. 2021, MNRAS, 505, 4626
Kacprzak, T., Fluri, J., Schneider, A., Refregier, A., & Stadel, J. 2023, JCAP, 2023, 050
Kilbinger, M. 2015, RPPh, 78, 086901
Kilbinger, M., & Schneider, P. 2005, A&A, 442, 69
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436
LSST Dark Energy Science Collaboration 2012, arXiv:1211.0310
Lu, T., Haiman, Z., & Li, X. 2023, MNRAS, 521, 2050
Lucie-Smith, L., Peiris, H. V., Pontzen, A., Nord, B., & Thiyagalingam, J. 2024, PhRvD, 109, 063524
Ludwick, K. J. 2017, MPLA, 32, 1730025
Matilla, J. M. Z., Sharma, M., Hsu, D., & Haiman, Z. 2020, PhRvD, 102, 123506
Miles, C., Bohrdt, A., Wu, R., et al. 2021, NatCo, 12, 3905
Müller, S., Schüler, L., Zech, A., & Heße, F. 2022, GMD, 15, 3161
Myles, J., Alarcon, A., Amon, A., et al. 2021, MNRAS, 505, 4249
Peebles, P. J. E. 2020, The Large-scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Piras, D., & Lombriser, L. 2024, PhRvD, 110, 023514
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Porth, L., Smith, R. E., Simon, P., Marian, L., & Hilbert, S. 2020, MNRAS, 499, 2474
Reeves, A., Nicola, A., Refregier, A., Kacprzak, T., & Machado Poletti Valle, L. F. 2024, JCAP, 2024, 042
Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73
Rosenberg, E., Gratton, S., & Efstathiou, G. 2022, MNRAS, 517, 4620
Schneider, P., Kilbinger, M., & Lombardi, M. 2005, A&A, 431, 9
Secco, L. F., Samuroff, S., Krause, E., et al. 2022, PhRvD, 105, 023515
Semboloni, E., Schrabback, T., van Waerbeke, L., et al. 2010, MNRAS, 410, 143
Sgier, R., Fluri, J., Herbel, J., et al. 2021, JCAP, 2021, 047
Sgier, R. J., Réfrégier, A., Amara, A., & Nicola, A. 2019, JCAP, 2019, 044
Simonyan, K., Vedaldi, A., & Zisserman, A. 2013, arXiv:1312.6034
Takada, M., & Jain, B. 2003a, MNRAS, 340, 580
Takada, M., & Jain, B. 2003b, MNRAS, 344, 857
Tang, J., Alelyani, S., & Liu, H. 2014, Data Classification: Algorithms and Applications (Boca Raton, FL: CRC Press), 37
Tsujikawa, S. 2013, CQGra, 30, 214003
Tulin, S., & Yu, H.-B. 2018, PhR, 730, 1
Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2021, ApJ, 907, 44
Webster, R., & Oliver, M. 2007, Geostatistics for Environmental Scientists (New York: Wiley)
Weiler, M., & Cesa, G. 2019, arXiv:1911.08251
Xavier, H. S., Abdalla, F. B., & Joachimi, B. 2016, MNRAS, 459, 3693
Zonca, A., Singer, L., Lenz, D., et al. 2019, JOSS, 4, 1298

Chapter 8

Making the leap. Part I. Modelling the reconstructed lensing convergence PDF from cosmic shear with survey masks and systematics

Bibliographic and copyright information

Barthelemy A., Halder A., **Gong Z.**, and Uhlemann, C., *Making the leap. Part I. Modelling the reconstructed lensing convergence PDF from cosmic shear with survey masks and systematics*, JCAP 03 (2024) 060

DOI: 10.1088/1475-7516/2024/03/060

Outline

This work focuses on the improvement of weak lensing convergence PDF modeling by incorporating realistic survey masks and systematics effects. The modeling of the PDF itself is based on the large deviation theory (LDT) framework and its applications to projected density fields such as weak lensing convergence have also been demonstrated in previous works such as Barthelemy et al. (2021). In this work, the survey mask considered is that of the third data release of the Dark Energy Survey (DES) and its impact on the theoretical prediction is introduced into the model via the pseudo-C_ℓ formalism. Other systematic effects considered include the intrinsic alignment, photometric redshift uncertainties, shear additive and multiplicative biases and baryonic feedback. As a co-author, my contributions to this work are (i) together with the first author Barthelemy A. and the second author Halder A., we developed the systematic effects modeling as listed above (except for the baryonic feedback effect which was mainly investigated by Uhlemann C. and Barthelemy

A.), (ii) writing parts of the manuscript, and (iii) preparing the neural network-based emulators of the PDF statistic and the GPU-based cosmological parameter inference pipeline for the real data analysis on DES Y3 data (Barthelemy et al. in prep).

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *July 21, 2023*
Revised: *December 14, 2023*
Accepted: *January 12, 2024*
Published: *March 28, 2024*

Making the leap. Part I. Modelling the reconstructed lensing convergence PDF from cosmic shear with survey masks and systematics

Alexandre Barthelemy[a], **Anik Halder**[a,b], **Zhengyangguang Gong**[a,b] **and Cora Uhlemann**[c]

[a]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstraße 1, München 81679, Germany*
[b]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstraße 1, Garching 85748, Germany*
[c]*School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne NE1 7RU, U.K.*

E-mail: A.Barthelemy@lmu.de, ahalder@usm.lmu.de, lgong@usm.lmu.de, cora.uhlemann@newcastle.ac.uk

Abstract: The last few years have seen the development of a promising theoretical framework for statistics of the cosmic large-scale structure — the theory of large deviations (LDT) for modelling weak-lensing one-point statistics in the mildly nonlinear regime. The goal of this series of papers is to make the leap and lay out the steps to perform an actual data analysis with this theoretical tool. Building upon the LDT framework, in this work (paper I) we demonstrate how to accurately model the Probability Distribution Function (PDF) of a reconstructed Kaiser-Squires convergence field under a realistic mask, that of the third data release of the Dark Energy Survey (DES). We also present how weak lensing systematics and higher-order lensing corrections due to intrinsic alignments, shear biases, photo-z errors and baryonic feedback can be incorporated in the modelling of the reconstructed convergence PDF. In an upcoming work (paper II) we will then demonstrate the robustness of our modelling through simulated likelihood analyses, the final step required before applying our method to actual data.

Keywords: weak gravitational lensing, galaxy surveys, gravitational lensing, cosmological parameters from LSS

ArXiv ePrint: 2307.09468

https://doi.org/10.1088/1475-7516/2024/03/060

JCAP03(2024)060

Contents

1 Introduction 1

2 Modelling the (observationally inaccessible) true convergence PDF 4
- 2.1 Statistical definitions 4
- 2.2 Large deviation theory of the matter density field 5
- 2.3 From matter density to convergence 8

3 Measuring the KS reconstructed convergence PDF 9
- 3.1 From galaxy ellipticities to shear field 9
- 3.2 From shear to convergence 10
- 3.3 Reconstructing the convergence PDF from simulated shear maps 11

4 Modelling the KS reconstructed convergence PDF 12
- 4.1 Schematic view 12
- 4.2 Reconstructed cumulant generating function 14
- 4.3 Reconstructed variance along the line-of-sight 15

5 Testing the model 16
- 5.1 Mock data preparation and simulation-specific corrections 16
- 5.2 DES Y3 mask and BNT transform 18
- 5.3 DES Y3 mask for sources in the fourth DES Y3 redshift bin 21

6 Incorporating observational systematics in the modelling 21
- 6.1 Intrinsic alignments 22
- 6.2 Baryonic feedback 24
- 6.3 Additive and multiplicative shear biases 25
- 6.4 Higher-order lensing corrections 26
- 6.5 Photometric redshift uncertainties 27

7 Discussion and conclusion 28

A Mode-coupling matrices in the pseudo-C_ℓ formalism 30

B Masking effects on the Kaiser-Squires reconstructed convergence moments 31

C Square mask for sources in the DES Y3 4^{th} bin 32

1 Introduction

In weak gravitational lensing (WL), light rays from background source galaxies propagate through the foreground inhomogeneous distribution of baryonic and dark matter. This induces (de)magnification of the brightness of galaxies and a coherent distortion pattern in their observed shapes [1, 2]. The statistics of the projected WL fields depend on the three-dimensional large-scale structure (LSS) and hence provide a powerful way to probe and address critical questions in cosmology such as structure formation history, the nature

JCAP03(2024)060

of dark energy and dark matter, and the laws of gravity. WL cosmology is therefore an active area of research in currently ongoing wide-area galaxy imaging surveys such as DES [3], KiDS [4] and HSC-SSP [5]. The promising results from these surveys have also motivated the development of new generation surveys such as Euclid [6] and Vera Rubin's LSST [7] which will soon provide data with unprecedented quality.

One standard approach to extract cosmological information from WL fields is to analyse their power spectra (in Fourier space) or the real space 2-point correlation function (2PCF). These statistics can completely characterise the information contained in Gaussian random fields which describe our Universe accurately at its earliest stage of evolution [8] and on linear (large) scales. However, at late times, the evolution of matter density fluctuations becomes nonlinear due to gravitational instability and develops non-Gaussian features on nonlinear (small) scales. As a consequence, to probe the rich information contained in this non-Gaussian late-time cosmic density field one needs to continue standard 2PCF analyses as well as go beyond and investigate non-Gaussian statistical tools. We follow this path in our work.

There are various kinds of non-Gaussian statistics that have already been or that will be applied to weak-lensing survey data. We mention as examples the cosmic shear 3-point correlation function (3PCF) [9, 10], the integrated 3PCFs [11–14], aperture-mass moments [15–17], lensing peaks [18, 19] and density-split statistics [20–22] (to name a few). All of these have distinct advantages and disadvantages based on both their measurement and modelling strategies. A typical WL survey analysis strategy would rely on a baseline 2PCF analysis and further complement that with non-Gaussian statistics [13, 23] to obtain stringent constraints on cosmological parameters. Extracting non-Gaussian information is a promising avenue as field-level approaches have been shown to enhance the constraining power for σ_8 and Ω_m by a factor of 3 and 5, respectively and to break the weak lensing degeneracy [24, 25]. The Probability Distribution Function (PDF) of the weighted projection of the matter density fluctuation field along the line-of-sight, the weak lensing convergence field (κ), was shown to contain a significant amount of cosmological information [23] and to hold the potential to break cosmological parameter degeneracies that are exhibited in standard weak lensing 2PCF analyses [26–28]. By measuring the smoothed κ field value inside apertures (or cells), the one-point κ-PDF statistic has the advantage of being straightforward to measure compared to other non-Gaussian WL convergence field probes such as bispectrum (counting triangular configurations) or Minkowski functionals (topological measurement). An emulation approach to constrain cosmological parameters using the κ/σ-PDF (σ is the κ standard deviation) jointly with convergence power spectrum was applied to HSC Year 1 (Y1) data [29]. A similar analysis was carried out for KiDS-1000 and LSST mock data as well [30]. Both these works relied on small patches obtained from N-body simulations as models for the κ-PDF to perform their cosmological analyses. In this series of work we instead use a from-first-principles theoretical modelling framework based on the large deviation theory (LDT) for the κ-PDF in the mildly nonlinear regime, giving access to the cosmological signal [28] as well as the covariance [31]. This is complementary to the halo-model approach aimed to describe the more nonlinear regime [32].

However, one major difficulty in applying κ-PDF to real data analyses is that κ itself is not a direct observable. Reference [33] showed that the κ and the more directly observable weak lensing cosmic shear γ are related to each other through a convolution and pointed out

JCAP03(2024)060

that the κ field can be reconstructed from the shear field through an inversion called the Kaiser-Squires (KS) inversion. An important point to note is that this KS reconstruction procedure is exact only if one has access to the shear field at every location on the celestial sphere (full-sky), which unfortunately is not feasible. As a consequence, since the KS inversion of the shear field is non-local, the reconstructed κ field is severely affected by the unobserved regions on the sky due to survey geometry, the presence of masks and holes in the data as well as the field borders. The KS reconstructed κ field can hence be significantly different from the true inaccessible κ field. Thus, it is absolutely crucial that one includes the accurate modelling/control of the masking effect when analysing any κ-statistic from a reconstructed κ-field. To side step this problem, recent works have adopted informed priors in the reconstruction of κ fields to infer the underlying true κ fields, e.g. the usage of a log-normal prior [34] or a sparse wavelet prior with nulled B-modes inside the survey mask [35]. Some of these informed priors were shown to work better than the KS inversion procedure at the map-level in reconstructing the κ field [36], and recent machine-learning-based reconstruction techniques seem to be extremely accurate [37, 38]. However, these other strategies come with the disadvantage that the impact of the reconstruction under the presence of mask is not at all straightforward to include in theoretical models of κ-field summary statistics. On the other hand, this can instead be achieved within the context of the KS reconstruction procedure and for example, has been done in the modelling of the KS reconstructed 1-point κ-map moments [39] and as we shall present for the first time in this paper — is also possible in the case of the KS reconstructed κ-PDF.

Previous works [15, 40, 41] have developed an accurate theoretical model for the cumulant generating function and also the probability distribution function of the lensing-aperture mass or the κ field using LDT.[1] In this paper, we present for the first time a few-percent accurate (within cosmic variance for the DES Y3 survey area) analytical prediction of the KS reconstructed κ-PDF based on LDT while incorporating a realistic survey mask into our modelling through the pseudo-C_l formalism [44], mitigation of convergence E to B mode leakage and modification of the scaling relations between the cumulants of the density contrast field through purely geometric corrections. We show that the reconstructed κ-PDF from N-body simulations and the corresponding theoretical prediction is consistent within statistical uncertainty of cosmic variance. We also present the framework to include the modelling of different astrophysical and survey systematic effects in the reconstructed κ-PDF, including shape noise, galaxy intrinsic alignments, additive and multiplicative shear biases, photometric redshift uncertainties and higher-order lensing corrections. Using hydrodynamic simulations, we also quantify the impact of baryonic feedback on the κ-PDF and find that it mainly affects the variance of the PDF. All these components can be treated theoretically and thus significantly strengthens the case for a κ-PDF analysis in real data using a theoretical framework which is from-first-principles up to a re-scaling by the amplitude of fluctuations.

This paper is structured as follows: we first recap in section 2 the base model for the true but observationally inaccessible (i.e. non-reconstructed from shear field) convergence PDF from LDT. In section 3 we detail the KS reconstruction procedure of the κ field from a simulated full-sky shear field with a realistic mask, that of the DES Year 3 (Y3) data release.

[1] Readers interested in the application of LDT in LSS cosmology are referred to refs. [42, 43].

JCAP03(2024)060

We show the modification of our original LDT theoretical model of the κ-PDF to account for this reconstruction with the realistic mask in section 4 and test it against measurements from N-body simulations in section 5. Furthermore, in section 6 we present the strategy for modelling higher-order lensing, astrophysical and survey systematic effects in the κ-PDF. We summarise and conclude in section 7. Appendix A recaps details about the treatment of the masking effect on the cosmic shear power spectrum in the pseudo-C_l formalism. In appendix B we quantify the effect of the KS reconstruction on the moments/cumulants of the reconstructed κ-field. Finally, in appendix C we present a simplistic study of the impact of a square mask (without holes) on the KS reconstructed κ-PDF.

2 Modelling the (observationally inaccessible) true convergence PDF

In this section we review how one can model the "true" but observationally inaccessible 1-point convergence κ-PDF based on large deviation theory (LDT). Readers already familiar with this model originally described in e.g. [40, 45] can skip to section 3.

2.1 Statistical definitions

We use different statistical quantities that we briefly introduce here for clarity. From the PDF $\mathcal{P}_X$ of some continuous random variable X, one can define the Moment Generating Function (MGF) as the Laplace transform of the PDF

$$M_X(\lambda) = \mathbb{E}\left(e^{\lambda X}\right) = \int_{-\infty}^{+\infty} \mathrm{d}x \; e^{\lambda x} \mathcal{P}_X(x), \tag{2.1}$$

or equivalently as the expectation value[2] of the random variable $e^{\lambda X}$. Note however that the existence of an MGF is not guaranteed for all possible random fields. For example, the MGF of a strictly lognormal field is undefined for real positive λ. For the MGF to exist (in cases where the PDF also exists), the PDF needs to decay faster than the exponential $e^{\lambda x}$ for the integral in equation (2.1) to exist. In the case of the one-point statistics of the cosmic density field, as computed within the large deviation framework (see section 2.2), there actually exists a critical positive real value $\lambda_{\rm c}$ — hereafter dubbed *critical point* — beyond which the MGF is not defined. In practice and for a field sampled within a finite volume, the MGF along the real axis will always exist and will simply tend towards $e^{\lambda X_{\rm max}}$ for $\lambda \geq \lambda_{\rm c}$, where $X_{\rm max}$ is the maximum value of X in the finite field.

The moment generating function, as its name implies, can be used to compute the moments of the distribution as can be seen from the series expansion of the expectation of $e^{\lambda X}$,

$$M_X(\lambda) = \mathbb{E}\left(e^{\lambda X}\right) = \sum_{n=0}^{+\infty} \frac{\lambda^n \mathbb{E}\left(X^n\right)}{n!}, \tag{2.2}$$

[2] Note that we make use throughout this work of the ergodicity hypothesis, in which one assumes that ensemble averages are equivalent to spatial averages ($\mathbb{E}(\cdot) \to \langle\cdot\rangle$) over one realisation of a random field at one fixed time. This requires that spatial correlations decay sufficiently rapidly with separation so that one has access to many statistically independent volumes within one realisation.

JCAP03(2024)060

so that the nth derivative of the MGF at $\lambda = 0$ is equal to the nth-order moment, $\mathbb{E}(X^n)$. The logarithm of the MGF is the Cumulant Generating Function (CGF)

$$\phi(\lambda) = \log(M(\lambda)) = \log(\mathbb{E}(e^{\lambda X})) = \sum_{n=1}^{+\infty} k_n \frac{\lambda^n}{n!}\,, \tag{2.3}$$

where k_n are the cumulants (i.e. the connected moments) of the distribution and where we have dropped the subscript X for clarity.

The reduced cumulants are defined as

$$S_n = \frac{k_n}{k_2^{n-1}}, \qquad n \geq 1 \tag{2.4}$$

where k_2 is the variance. These are important in the context of cosmological structure formation because the S_n of the cosmic matter density field have been shown to be independent of redshift down to mildly nonlinear scales [46, 47] and thus introduce relevant scaling relations between the cumulants. We thus also define the scaled cumulant generating function (SCGF hereafter) as

$$\varphi(\lambda) = \lim_{k_2 \to 0} \sum_{n=1}^{+\infty} S_n \frac{\lambda^n}{n!} = \lim_{k_2 \to 0} k_2\, \phi\left(\frac{\lambda}{k_2}\right), \tag{2.5}$$

which we also extrapolate to non-zero values of the variance (i.e. evaluate at finite k_2 on our chosen smoothing scale). One can reconstruct the PDF from the CGF using an inverse Laplace transform (inverting equation (2.1))

$$\mathcal{P}(x) = \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}\lambda}{2\pi i} \exp\left(-\lambda x + \phi(\lambda)\right)\,. \tag{2.6}$$

2.2 Large deviation theory of the matter density field

The large deviation theory (LDT) framework in large-scale structure has mainly been used to model the one-point statistics of the smoothed 3D matter density field [see, for example, 42, 43, 48] and also of projected 2D quantities such as weak lensing convergence and aperture-mass fields [15, 28, 40]. It has also been extended to the joint distribution between densities measured at some distance [49] and projected quantities between different source redshift bins [50] (i.e. n-point PDF). The results are most simply applied for highly symmetrical window functions such as two- or three-dimensional top-hats, but can be generalised to other smoothing schemes [15, 42, 51]. We begin here by recalling some of the results of LDT for the one-point statistics of the matter density contrast smoothed in two-dimensional disks (which replicates the dynamics within long cylinders), which in turn will allow us to compute the one-point statistics of projected 2D quantities like the convergence field.

A set of random variables $\{\rho^\epsilon\}_\epsilon$ with PDF $\mathcal{P}_\epsilon(\rho^\epsilon)$ is said to satisfy a large deviation principle if the limit

$$\Psi_{\rho^\epsilon}(\rho^\epsilon) = -\lim_{\epsilon \to 0} \epsilon \log\left[\mathcal{P}_\epsilon(\rho^\epsilon)\right] \tag{2.7}$$

exists, where ϵ is the *driving parameter*. Ψ is known as the rate function of ρ^ϵ and describes the exponential decay of its PDF. The driving parameter ϵ indexes the random variables

JCAP03(2024)060

with respect to some evolution, for example an evolution in time. In the case of the matter density field smoothed on a single scale R, this driving parameter is the variance, which acts as a clock from initial to late times ($\epsilon \equiv \sigma_R^2$). We now omit the ϵ sub/superscripts in our notation for simplicity.

The existence of a large deviation principle for the random variable ρ implies that its SCGF φ_ρ is given through Varadhan's theorem as the Legendre-Fenchel transform of the rate function Ψ_ρ [52–54]

$$\varphi_\rho(\lambda) = \sup_\rho \left[\lambda\rho - \Psi_\rho(\rho)\right], \tag{2.8}$$

where the Legendre-Fenchel transform reduces to a simple Legendre transform when Ψ_ρ is convex. In that case,

$$\varphi_\rho(\lambda) = \lambda\rho - \Psi_\rho(\rho), \tag{2.9}$$

where ρ is a function of λ through the stationary condition[3]

$$\lambda = \frac{\partial \Psi_\rho(\rho)}{\partial \rho}. \tag{2.10}$$

Another consequence of the large-deviation principle is the so-called contraction principle. This principle states that for a random variable τ satisfying a large deviation principle and related to ρ through the continuous mapping f, the rate function of ρ can be computed as

$$\Psi_\rho(\rho) = \inf_{\tau : f(\tau)=\rho} \Psi_\tau(\tau). \tag{2.11}$$

This is called the contraction principle because f can be many-to-one, in which case we are *contracting* information about the rate function of one random variable down to the other. In physical terms, this states that an improbable fluctuation of ρ is brought about by the most probable of all improbable fluctuations of τ.

For the case of the normalised 3D matter density field $\rho \equiv \rho/\bar{\rho}$, the rate function of the late-time normalised density field at different scales can be computed from initial conditions if the most likely mapping between the two is known — that is, if one is able to identify the leading field configuration that will contribute to the infimum of equation (2.11). In cylindrically symmetric configurations, as for a disk of radius R at redshift z (in 2D space) or alternatively a very long 3D cylinder centered on this disk, the most likely mapping between final and initial conditions should preserve the symmetry [55, 56].[4] This in turn leads to initial conditions also being cylindrically symmetric and the dynamics between the two being that of cylindrical collapse.

[3] The ρ_c value at which Ψ_ρ ceases to be convex leads to a λ_c value which corresponds to the critical point mentioned when discussing equation (2.1).

[4] This is only true for a certain range of density contrasts around zero, very much sufficient for our purposes. However, one could note that there are counter-examples in which the spherical or cylindrical symmetry does not lead to spherical/cylindrical collapse being the most likely dynamics, for example in 1D for very high values of the density.

JCAP03(2024)060

Thus, starting from Gaussian initial conditions,[5] the rate function $\Psi_{\bar{\tau}}$ of the most probable initial (linear) density field $\bar{\tau}$ is simply a quadratic term. From the dynamics of cylindrical collapse that maps the most probable initial and late-time fields, the rate function $\Psi_{\rm cyl}$ of the normalised late-time density field ρ in a disk of radius R is then given by

$$\Psi_{\rm cyl}(\rho) = \sigma_R^2 \frac{\bar{\tau}^2}{2\sigma_r^2}, \tag{2.12}$$

where σ_R^2 — our driving parameter — is the variance of the nonlinear density field in the disk, σ_r^2 is the variance of the linear density field inside the initial disk (before collapse) of radius $r = R\,\rho^{1/2}$ (from mass conservation), and $\bar{\tau}$ is the linear density contrast obtained through the most probable mapping between the linear and late-time density fields. This mapping is given by 2D spherical (cylindrical) collapse, for which an accurate parametrisation is given by [59]

$$\zeta(\bar{\tau}) = \rho = \left(1 - \frac{\bar{\tau}}{\nu}\right)^{-\nu}. \tag{2.13}$$

This parametrisation is in the spirit of previous works involving the density filtered in spherical cells, while the value of ν in this parametrisation of ζ is chosen to be $\nu = 1.4$, so as to reproduce the value of the leading-order (tree-level) skewness in cylinders as computed from Eulerian perturbation theory [48].

Then, as a straightforward consequence of the contraction principle, the rate function given by equation (2.12) is also the rate function of any monotonic transformation of ρ, so that for the density contrast $\delta = \rho - 1$, we have $\Psi_\delta(\delta) = \Psi_\rho(\rho(\delta))$. Thus, plugging equation (2.12) into equation (2.9) gives us the SCGF $\varphi_{\rm cyl}$ of the matter density contrast in a disk at redshift z.

Finally, one of the key aspects of the large deviation formalism in the cosmological context is that we apply the result for the SCGF beyond the $\sigma_R^2 \to 0$ limit. This extrapolation of the exact result allows us to obtain a realistic CGF $\phi_{\rm cyl}$ of the real density field for non-vanishing σ_R^2 by rescaling the SCGF by the driving parameter (the nonlinear variance) at the scale and redshift being considered. This yields:

$$\phi_{\rm cyl}(\lambda) = \frac{1}{\sigma_R^2}\varphi_{\rm cyl}(\lambda\sigma_R^2). \tag{2.14}$$

This is physically meaningful because we thus construct a CGF naturally matching its quasi-linear limit and since the reduced cumulants S_n from the cylindrical/spherical collapse dynamics have been shown to be very robust over a large range of scales and redshifts down to mildly nonlinear scales ($\gtrsim 5\,{\rm Mpc}/h$ at $z \gtrsim 0$, see for example figure 2 in [43] and figure A1 in [48]) so that rescaling by the nonlinear variance σ_R^2 allows access to the full one-point statistics of the nonlinear density field. In LDT terms, the SCGF given by the large deviation principle is the well-defined asymptotic form taken by the reduced cumulants of the field in the regime where the variance goes to zero, and we simply keep this form for our predictions of the nonlinear CGF.

Finally, note that though equations (2.9) and (2.12) have been known and used for three decades in the context of count-in-cells statistics, that is the density field filtered in

[5] Primordial non-Gaussianity can also straightforwardly be accounted for in this formalism as shown by [57, 58].

JCAP03(2024)060

3D top-hat windows [see for example 60, 61], their re-derivation through large deviation statistics is more general and allows to set up a framework for the computation of different probabilities in the cosmological context.

2.3 From matter density to convergence

Let us recall that for a flat cosmology, the convergence κ can be interpreted as a line-of-sight projection of the matter density contrast between the observer and the source and can be written as [62]

$$\kappa(\boldsymbol{\vartheta}) = \int_0^{+\infty} \mathrm{d}z \frac{\mathrm{d}\chi}{\mathrm{d}z}\, \omega_{n(z)}(\chi)\, \delta(\chi, \chi\boldsymbol{\vartheta}), \tag{2.15}$$

where χ is the comoving radial distance and the generalised lensing kernel $\omega_{n(z)}$ for a wide distribution of sources following the normalised distribution $n(z)$ is

$$\omega_{n(z)}(z) = \frac{3\Omega_\mathrm{m} H_0^2}{2c^2} \int \mathrm{d}z_\mathrm{s}\, n(z_\mathrm{s}) \frac{[\chi(z_\mathrm{s}) - \chi(z)]\,\chi(z)}{\chi(z_\mathrm{s})} \mathcal{H}(z_\mathrm{s} - z)(1+z), \tag{2.16}$$

where the Heaviside $\mathcal{H}$ ensures that the integrand vanishes for $z \geq z_\mathrm{s}$. Hereafter we only note the lensing kernel ω for simplicity. Under the small-angle/Limber approximation, it can be shown that correlators of the (smoothed) convergence field can be seen as a juxtaposition (by which we mean an integral along the line-of-sight) of the 2D correlators of the underlying density field, as if each 2D slice along the line-of-sight is statistically independent of the others [40, 45]. In terms of the one-point statistics of the smoothed κ field within a top-hat window function of angular radius θ, this translates to saying that the CGF of κ is a sum along the line-of-sight of the CGF of independent 2D slices of the matter density contrast:[6]

$$\phi_{\kappa,\theta}(\lambda) = \int_0^{+\infty} \mathrm{d}z \frac{\mathrm{d}\chi}{\mathrm{d}z}\, \phi_\mathrm{cyl}^{<\chi(z)\theta}(\omega_{n(z)}(z)\lambda, \chi(z)), \tag{2.17}$$

where $\phi_\mathrm{cyl}^{<\chi(z)\theta}$ is the CGF of the density contrast filtered within a disk of radius $\chi(z)\theta$ so as to reproduce the geometry of the light cone.

Equation (2.17) thus reduces the complexity of the problem down to computing the one-point statistics of the 2D matter density in each two-dimensional slice (or equivalently within long 3D cylinders at the same redshift up to some factor depending only on the length of the cylinder) along the line-of-sight, which we have already done in section 2.2. Using these results, we can then build the nonlinear CGF $\phi_{\kappa,\theta}$ of the convergence field. Note that equation (2.17) highlights the nice property of the projected CGF being expressible simply as a sum of independent redshift slices. This is an important property when considering more complicated joint distributions (e.g. joint convergence CGF of multiple source tomographic redshift bins), in which the only modification in this integral would be the replacement of

[6] Rigorously, this result applies for a juxtaposition of very long cylinders centered on the slices and of length $L \to \infty$. Since the symmetry and thus the most likely dynamics of these long cylinders are the same as for a 2D slice in a 2D space, and since the results are independent of L, we refer to '2D slices' for clarity. This emphasises that correlations along the line-of-sight are negligible compared to those in the transverse directions.

JCAP03(2024)060

the $\omega_{n(z)}(z)\lambda$ term by a term depending on more than one λ variable. This form is much simpler than that of the corresponding multi-variate PDF [50].

Finally, it is important when working with the CGF (albeit to compute the PDF) to know the approximate location of the (theoretical) critical point of the convergence field, λ_c. First, the critical points δ_c in each redshift slice are calculated by finding where the second derivative of the rate function becomes zero. The corresponding critical λ values can be obtained by applying the stationary condition (equation (2.10)). The minimum λ_c along the line-of-sight is then taken as the critical point of the convergence CGF.

The convergence PDF $\mathcal{P}(\kappa)$ is then computed from its CGF $\phi_{\kappa,\theta}$ using the inverse Laplace transform of equation (2.6) as

$$\mathcal{P}(\kappa) = \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\lambda}{2\pi i} \exp\left(-\lambda\kappa + \phi_{\kappa,\theta}(\lambda)\right), \tag{2.18}$$

and thus explicitly depends on the lensing kernel $\omega_{n(z)}$ from equation (2.16). This computation involves a continuation of the CGF in the complex plane that can be performed using different techniques. We refer to [40] and [15] for technical discussions on the two possible methods that have been used in the literature to perform this complex continuation in the LDT context. Here, we use the one of [15] that relies on an informed fitting function of the CGF.

3 Measuring the KS reconstructed convergence PDF

In this section we describe how one performs a Kaiser-Squires inversion on an observed masked shear field in order to create a reconstructed convergence κ map as well as how one can measure a meaningful κ-PDF from this map.

3.1 From galaxy ellipticities to shear field

In real-life weak-lensing experiments, the convergence field is not a direct observable (except maybe through magnification, see for example [63]). What we actually observe are the shapes of source galaxies, their ellipticities, which are a noisy estimate of the reduced shear g:

$$\epsilon = \frac{g + \epsilon_{\mathrm{IA}} + \epsilon_{\mathrm{n}}}{1 + g(\epsilon_{\mathrm{IA}} + \epsilon_{\mathrm{n}})} \tag{3.1}$$

with $g = \gamma/(1-\kappa)$, ϵ_{IA} the intrinsic shape of the galaxy and ϵ_{n} the shape measurement noise. In the weak-lensing regime, both the shear γ and the convergence κ are $\ll 1$ so that the ellipticities can serve as a noisy estimate of the shear field through

$$\epsilon \approx \gamma + \epsilon_{\mathrm{IA}} + \epsilon_{\mathrm{n}}. \tag{3.2}$$

In the following we will ignore intrinsic alignments whose introduction in the modelling will be described later in section 6.1. On the other hand, the contribution of noise is estimated in the literature by randomly rotating the shape of the galaxies to erase the cosmological contribution which would lead to pure-noise ellipticity, shear and then convergence fields (whose reconstruction we detail below) of zero average but non-negligible variance. By virtue of the central limit theorem, this noise is expected to become Gaussian for large numbers of

JCAP03(2024)060

galaxies and is also expected to be closely independent from the reconstructed field. Since the simulations we use in this work contain no intrinsic shape noise, we will assume that those two hypotheses are valid and shape noise can thus be trivially taken into account in the theoretical modelling of the convergence PDF by simple convolution of the clean PDF with a Gaussian of the correct shape noise amplitude (we will test this in paper II of the series). For a survey with varying source galaxy density across the sky, this would lead to a mean shape noise variance which would still convolve the clean PDF. Note however that the Gaussian hypothesis can be easily lifted if needed — simply by replacing the Gaussian convolution by another one — as well as the hypothesis of zero correlation between the noise and reconstructed field which can be tested by computing the joint PDF between the noise map (obtained through random rotations) and the observed convergence. A denoised estimator of the observed convergence would then be obtained by marginalisation over the noise in this joint PDF.

3.2 From shear to convergence

Crucially, for a footprint of approximately $5000\,\mathrm{deg}^2$ such as the one from the DES Y3 — and up to three times that for a Euclid-like experiment — we need a full-sky, spherical harmonics approach to estimate the convergence field from the shear [36, 64]. In this formalism, the (Born-)projected Poisson equation reads [65]

$$\kappa = \frac{1}{4}(\eth\bar{\eth} + \bar{\eth}\eth)\psi, \tag{3.3}$$

with ψ the usual projected (under Born approximation) gravitational potential and $\eth$, $\bar{\eth}$ are respectively the raising and lowering operators acting on the spin-weighted spherical harmonics [see appendix A of 65]. Similarly, the usual complex shear field equations translate into

$$\gamma = \gamma_1 + i\gamma_2 = \frac{1}{2}\eth\eth\psi. \tag{3.4}$$

Finally, expanding the projected gravitational potential and convergence (scalar, spin-0 fields) as well as the spin-2 complex shear on the spin-weighted spherical harmonics basis ${}_0Y_{lm}$ and ${}_2Y_{lm}$ respectively, we get

$$\psi(\theta,\phi) = \sum_{l,m} \psi_{lm}\, {}_0Y_{lm}(\theta,\phi), \tag{3.5}$$

$$\kappa = \kappa_E + i\kappa_B = \sum_{l,m} (\kappa_{E,lm} + i\kappa_{B,lm})\, {}_0Y_{lm}, \tag{3.6}$$

$$\gamma = \gamma_1 + i\gamma_2 = 2\sum_{l,m} (\gamma_{E,lm} + i\gamma_{B,lm})\, {}_2Y_{lm}. \tag{3.7}$$

Both the convergence and the shear have been decomposed into curl-free E-modes and divergence-free B-modes, and we can relate the shear to the convergence as

$$\kappa_{E,lm} + i\kappa_{B,lm} = \frac{1}{2} l(l+1)\psi_{lm}, \tag{3.8}$$

$$\gamma_{E,lm} + i\gamma_{B,lm} = \frac{1}{2}\left[l(l+1)(l-1)(l+2)\right]^{1/2}\psi_{lm}, \tag{3.9}$$

$$\implies \kappa_{E,lm} + i\kappa_{B,lm} = \sqrt{\frac{l(l+1)}{(l+2)(l-1)}}(\gamma_{E,lm} + i\gamma_{B,lm}). \tag{3.10}$$

JCAP03(2024)060

Equation (3.10) is the harmonic space spherical-sky generalisation of the Kaiser-Squires (KS) inversion formula (originally proposed by [33] for flat-sky convergence field reconstruction from the observed shear field). An inverse spherical-harmonic transform on the full-sky allows one to then obtain the KS reconstructed κ_E and κ_B fields on the celestial sphere. Strictly speaking, the equations above are correct assuming no couplings between the lenses that would appear when relaxing the Born approximation and taking into account corrections of the same order in the Sachs' equation of gravitational lensing [66, 67]. In this paradigm, the only allowed B-modes are those produced by the effect of masking the full-sky or observational systematics. On the other hand, allowing for couplings between lenses along the line-of-sight would slightly change the definitions of the shear and convergence and as a consequence their relationship. This is not a fundamental issue since those higher-order corrections can be modelled respectively for the shear and the convergence in cases where one or the other can be accessed [see for example 67–69], and since these corrections are small, one could still reconstruct a field from equation (3.10) and call it *convergence*. Alternatively, the knowledge of what are both the true shear and convergence has also permitted to build Bayesian reconstruction schemes of the convergence based on an observed shear field with the help of informed priors [see for example 34–36] as well as machine learning reconstruction algorithms [37] and more recently clever mix of the two [38]. In this paper, we reconstruct a convergence field by applying the KS inversion equation (3.10) on a masked full-sky shear field which contains higher-order lensing corrections and thus admits B-modes. We then only work with the reconstructed κ_E field and theoretically model the E-modes one-point PDF assuming that the main difference between the reconstructed and "true" PDFs are sourced only by masking the full-sky shear field.[7]

3.3 Reconstructing the convergence PDF from simulated shear maps

In actual observations, one would start from a discrete shear catalogue that has to be interpolated (pixelised) to create a shear map. We here skip that step and directly use a set of publicly available full-sky weak lensing shear maps in `Healpix` format [70][8] generated from the ray-traced N-body simulation suite from Takahashi et al. [71].[9] We have 108 realisations of several fixed source redshifts up to $z_s \sim 5$ and combine them to mimic a realistic source distribution from an input $n(z)$ (see section 5 for the description of how we do that in practice to mimic the 4$^{\text{th}}$ source redshift bin of the DES Y3). For each set of full-sky shear maps, we multiply those by a binary mask, shown in figure 1, mocking observed maps from DES Y3 and reconstruct the E-modes full-sky convergence field using KS reconstruction equation (3.10). This reconstruction makes use of the functions MAP2ALM and ALM2MAP of the `Healpy` package [72]. We then smooth the maps with top-hat filters of radius θ in

[7] The theoretical model for the non-reconstructed κ-PDF does not natively include higher-order lensing corrections (though they can also be included in the model if desired as discussed in section 6.4) hence the quotation marks around "true". However, the simulated convergence fields we use are either obtained through ray-tracing hence beyond Born-approximation or reconstructed from ray-traced shear fields thus admitting physical B-modes. We have thus checked explicitly that those are fortunately negligible at the scales and redshifts we consider.

[8] https://healpix.sourceforge.io/.

[9] http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/.

JCAP03(2024)060

harmonic space and the only remaining step is to select the pixels on the full-sky map to then construct a meaningful/physical convergence PDF. Indeed, since large parts of the shear field were set to zero due to the binary survey mask (and though the reconstruction of the convergence is in principle non-local) the reconstruction of κ_E in the previously masked areas yields unphysically small values compared to the fluctuations of the true convergence $\kappa_{\rm true}$ values (obtained directly from the simulation suite) in those regions. We illustrate this point in figure 2 where we display a full-sky (FS) reconstructed field $\kappa_{E,\rm FS}$ smoothed with a top-hat of radius 20 arcmin. We thus only take into account pixels whose smoothed convergence values $\kappa_{E,\rm unmasked}$ results from unmasked regions on the sky.[10] That way, we can expect a loss of power of κ E-modes to B-modes in the reconstruction due to the presence of the mask and the non-locality of the transform, but sufficiently mitigated such that physical scaling relations for example between higher-order cumulants of the matter density field still hold. We illustrate this fact more quantitatively in the next section in equation (4.1). In practice, this can be done by smoothing the binary mask with the same top-hat kernel as the field, and only keeping pixels where the smoothed masked has values sufficiently close to 1. For the DES Y3 mask, we find in practice that keeping all pixels where the smoothed mask is higher than 0.98 is a good equilibrium between keeping the maximum possible portion of the survey volume and mitigating the influence of masked pixels. This approach might be improved by upweighting the pixels that are partially masked as done in the density-split statistics context [20, 21] which is similar to ours. It might allow to consider more pixels but we do not implement here such a strategy.

4 Modelling the KS reconstructed convergence PDF

Having described the LDT based framework for modelling the observationally inaccessible κ-PDF in section 2 and then the actual procedure to reconstruct and measure the κ-PDF under the presence of a realistic survey mask using KS inversion in section 3, we now describe how to modify our original framework to incorporate the effect of the survey mask and develop a model for the KS reconstructed κ-PDF.

4.1 Schematic view

Let us first have a look at how the reconstruction scheme presented in the previous section affects the amplitude of convergence fluctuations. Schematically, masking the field reduces the overall shear power spectrum and thus the reconstructed convergence power spectrum by a factor $f_{\rm mix}$,[11] due to mode-mixing which is at first order comparable to the observable fraction of the full-sky $f_{\rm sky}$. On the other hand, the reconstructed convergence in the masked regions (e.g. far away from the survey footprint) is mostly 0 so that we could model the full-sky reconstructed PDF as a sum of the distribution of κ_E values in the unmasked regions and a

[10] Measuring the full-sky convergence PDF (i.e. using all the pixels over the entire sky) reconstructed from a partial sky shear field would yield a distribution with very high kurtosis (highly peaked around zero) illustrating the fact that most values are close to zero around which we observe fluctuations that correspond to the unphysically small (close to zero) reconstructed convergence values in the masked regions.

[11] Note that this is an effective and schematic treatment. In reality the loss of power is modelled by the convolution of the underlying true power spectrum by a mode-mixing matrix as described in appendix A.

JCAP03(2024)060

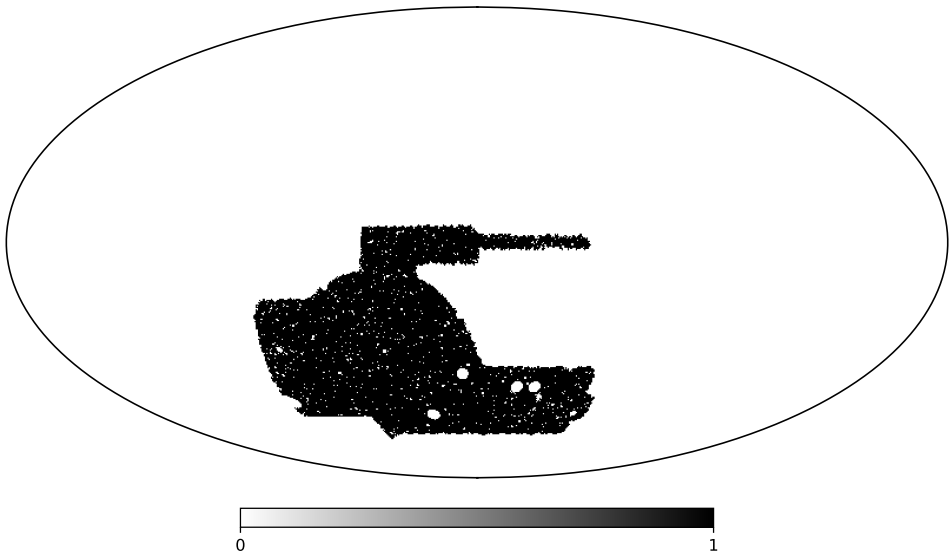

Figure 1. Illustrative Mollweide projection of the binary mask applied to the full-sky shear field to mimic DES Y3 observations. The mask is as close as possible to the real one, notably keeping the multiple holes of different sizes across the field of view. The observed fraction of the full-sky is $f_{\rm sky} = 0.1149$.

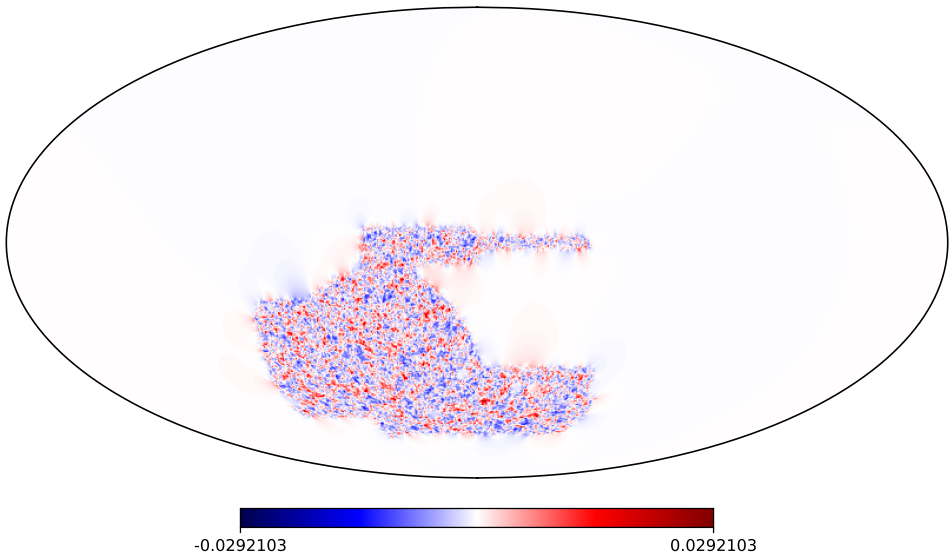

Figure 2. Illustrative Mollweide projection of the full-sky reconstructed $\kappa_{E,\rm FS}$ field for sources mimicking the DES Y3 fourth redshift bin and from a simulated shear field under the DES Y3 mask shown in figure 1. The reconstructed κ_E field has been further smoothed by a top-hat window of radius 20 arcmin. Observed carefully, one can spot the non-locality of the KS reconstruction through the blurring of the mask boundaries and holes. It should also be noted that the majority of the pixel values far outside the survey footprint fluctuate very closely around zero (but are not exactly zero), as expected.

JCAP03(2024)060

Dirac-delta distribution, both weighted by $f_{\rm sky}$ and $(1-f_{\rm sky})$ respectively. As a consequence, the variance of the reconstructed κ_E in the unmasked region of the sky $\sigma^2_{\kappa_E,\rm unmasked}$ is roughly the variance $\sigma^2_{\kappa_E,\rm FS}$ of the full-sky $\kappa_{E,\rm FS}$ reconstructed from the masked shear field divided by $f_{\rm sky}$ and we have

$$\sigma^2_{\kappa_E,\rm unmasked} \approx \frac{1}{f_{\rm sky}}\sigma^2_{\kappa_E,\rm FS} \approx \underbrace{\frac{f_{\rm mix}}{f_{\rm sky}}}_{O(1)}\sigma^2_{\kappa,\rm true}\,. \tag{4.1}$$

This is further illustrated with actual values of the κ_E variance under the DES Y3 mask in appendix B.

Additionally, the core idea behind the large deviation approach to the statistics of the cosmic matter density field is to identify asymptotic scaling relations between the cumulants of the field such that specifying the value of the variance serves both as a dial controlling the proximity to the asymptotic limit and as a closure relation allowing to compute all cumulants of the field from the variance. For now neglecting the projection along the line-of-sight (and since the convergence reconstruction scheme only affects the amplitude of fluctuation mildly), it could seem natural to assume that at first order the reconstruction affects the successive cumulants of the convergence field in a manner consistent with their scaling relations with the variance based on constant reduced cumulants S_n (see equation (2.4)). This would typically mean that the loss of power from E to B modes in the region of the sky where it is mitigated the most (which we call the "unmasked" regions, see section 3.3), preserves core physical properties of the field. As such, neglecting the projection along the line-of-sight, the $n^{\rm th}$ cumulants of the reconstructed κ_E field in the unmasked regions can schematically be related to the cumulants of the true but observationally inaccessible κ field as

$$\langle\kappa_E^n\rangle_{c,\rm unmasked} \approx \left(\frac{f_{\rm mix}}{f_{\rm sky}}\right)^{n-1}\langle\kappa^n\rangle_{c,\rm true}. \tag{4.2}$$

In the above equation (4.2), the subscript c denotes the connected parts of the moments, i.e. the cumulants.

4.2 Reconstructed cumulant generating function

More precisely, the large deviation principle applied to the cosmic matter density field implies that the scaling relations (constant S_n in equation (2.4)) between cumulants are correct for the matter density field, but not for the projected convergence field itself. However, under the Born and Limber approximations, the total lensing effect that we observe can be treated as the independent combination of successive lensing events along the line-of-sight. As a consequence, taking into account the convergence reconstructing scheme while preserving the scaling relations between cumulants must be understood at the level of each lensing event, that is for the cumulants of the matter density contrast along the line-of-sight. In terms of the reconstructed CGF ϕ_{κ_E}, this can be written by modifying equation (2.17) as

$$\phi_{\kappa_E,\theta}(\lambda) = \int_0^{+\infty}{\rm d}z\frac{{\rm d}\chi}{{\rm d}z}\frac{\langle\delta^2_{\rm true}\rangle}{\langle\delta^2_E\rangle}\phi^{<\chi\theta}_{\rm cyl}\left(\frac{\langle\delta^2_E\rangle}{\langle\delta^2_{\rm true}\rangle}\omega(\chi)\lambda,\chi(z)\right) \tag{4.3}$$

JCAP03(2024)060

where $\langle\delta^2_{\rm true}\rangle$ is the true variance of slices of the density field at redshift z and of radius $\chi(z)\theta$, while $\langle\delta^2_E\rangle$ is to be understood as the variance that results from the reconstruction scheme (as schematically done in equation (4.1) for the whole projection) and whose modelling we describe in the next subsection.

4.3 Reconstructed variance along the line-of-sight

We now take a more detailed look at the $\langle\delta^2_E\rangle$ term that appears in equation (4.3). In the Limber approximation, it corresponds to a variance measure[12] of a 2D matter density contrast slice of radius $\chi(z)\theta$ at redshift z which acts locally as the (reconstructed) convergence from a single lens plane. In the absence of a mask, it can be written as

$$\left\langle\delta^2_{\rm true}\right\rangle = \sum_l \frac{2l+1}{4\pi\chi(z)^2} W_l\left(\chi(z)\theta\right)^2 P_{\rm NL}(l/\chi(z), z), \tag{4.4}$$

where $P_{\rm NL}$ is the nonlinear matter power spectrum and W_l is the angular top-hat window function in harmonic space given by

$$W_l\left(\theta_0\right) = \frac{P_{l-1}\left(\cos\left(\theta_0\right)\right) - P_{l+1}\left(\cos\left(\theta_0\right)\right)}{\left(2l+1\right)\left(1-\cos\left(\theta_0\right)\right)} \tag{4.5}$$

where P_l are Legendre polynomial of degree l.

In the presence of masks, equation (4.4) is modified as follows [39]

$$\left\langle\delta^2_E\right\rangle = \frac{1}{f_{\rm sky}} \sum_l \frac{2l+1}{4\pi\chi(z)^2} W_l\left(\chi(z)\theta\right)^2 f_l^{-1} \sum_{l'} M^{EE,EE}_{ll'} P_{\rm NL}\left(l'/\chi(z), z\right) f_{l'}, \tag{4.6}$$

where $M^{EE,EE}_{ll'}$ is the element of the mode-mixing matrix that accounts for the transfer of power in the power spectrum from the unmasked shear E-modes to the masked shear E-modes. This term is explained in more detail in appendix A. Here, $f_l = [(l+2)(l-1)]/[l(l+1)]$ and accounts for the passage from convergence to the shear power spectrum in order to apply the mode-mixing formalism to the shear field directly. Finally, the $f_{\rm sky}$ factor is not necessarily to be understood as the true sky fraction observed by the survey. Instead, it actually comes from the parametrisation of the full-sky reconstructed convergence PDF as a sum of the fluctuations within the mask weighted by $f_{\rm sky}$ and a Dirac-delta (for the unobserved sky regions) weighted by $(1-f_{\rm sky})$. However, this parametrisation, though very accurate in practice, is not exact at the boundaries and holes of the mask, and even more so after smoothing the reconstructed convergence field. Having said that, this $f_{\rm sky}$ term is not free either as it can directly — without any theoretical input — be estimated from data as the ratio of the reconstructed smoothed full-sky variance to the one within the considered pixels. Note moreover that the reconstructed κ_E unmasked variance values measured after masking from the Takahashi simulation maps tend to be equal to the κ_E full-sky variance divided by the true fraction of the full-sky observed in the mask up to less than a percent for the DES Y3 mask and for the range of smoothing scales that we tested (up to $\sim 30\,$arcmin), hinting to the fact that our parametrisation is very accurate.

[12] It formally has a unit of length (Mpc or Mpc/h) because the derivation of the projection formula used in equations (2.17) and (4.3) makes use of cylinders of infinite length L (and not 2D slices) which cancels out in the end result but for which the variance is formally $\langle\delta^2_E\rangle/L$.

JCAP03(2024)060

5 Testing the model

In this section we implement the theoretical formalism for the Kaiser-Squires reconstructed convergence PDF described in the previous section and compare it to measurements made in the Takahashi simulations. Given that the simulated shear maps also contain higher-order lensing corrections, the comparison presented here allows us to test that the mask modelling is accurate even when neglecting other sources of B-modes than the ones created by the mask. We study two test cases and a third in appendix C. In subsection 5.2 we test the reconstruction scheme in a regime where the large deviation theory is known to be very accurate, that of the Bernardeau-Nishimichi-Taryuya (BNT) transform [73] which allows us to construct lensing observables only sensitive to the dynamics of the quasi-linear regime of the matter density field. It allows us to understand how well we are probing the Kaiser-Squires reconstruction without mixing additional inaccuracies that the theory could present. In subsection 5.3, we test our theoretical κ_E PDF for a source distribution mimicking that of the 4$^{\text{th}}$ bin from DES Y3 analysis and where we applied the real DES Y3 mask to the full-sky shear field. We present in appendix B supplementary information to subsection 5.3 but at the level of the cumulants and in appendix C the effect of a reconstruction similar to subsection 5.3 but where we replace the DES Y3 mask by a square patch with no holes and of the same area ($\sim 70 \times 70\,\text{deg}^2$).

5.1 Mock data preparation and simulation-specific corrections

As mentioned previously, the simulated true convergence and cosmic shear maps used in this work are obtained from the publicly available N-body simulations by Takahashi et al. [71]. The simulation suite provides weak lensing maps for $N = 38$ fixed source redshift planes between $z_s = 0.05$ and $z_s = 5.3$ in `Healpix` format. We combine these maps according to a source distribution $n(z)$ inspired by the 4$^{\text{th}}$ tomographic bin of the DES Y3 analysis to simulate a DES Y3-like lensing map in its 4$^{\text{th}}$ redshift bin. The procedure to do that is as follows:[13]

$$\kappa_{n(z)} = \sum_{i=1}^{N} s_i \kappa_{z_s^i} \,, \tag{5.1}$$

where s_i is a specific weight for a given full-sky lensing map $\kappa_{z_s^i}$ at source plane z_s^i. The weights used for combining the discrete source planes from the simulations are shown in figure 3. This final simulated convergence map can then be expressed as a line-of-sight projection of the matter density contrast through equation (2.15) with a lensing kernel $w_{n(z)}$:

$$w_{n(z)}(z) = \sum_{i=1}^{N} s_i w_{z_s^i}(z) \,. \tag{5.2}$$

The $w_{z_s^i}$ is in turn the lensing kernel for a given discrete source plane at redshift z_s^i and reads

$$w_{z_s^i}(z) = \frac{3\Omega_{\text{m}} H_0^2}{2c^2} \frac{[\chi_{z_s^i} - \chi(z)]\chi(z)}{\chi_{z_s^i}} \mathcal{H}(z_s - z)(1+z) \,. \tag{5.3}$$

[13] Although we show here the procedure on the κ field, the same source plane combination scheme is applied to create the shear maps.

JCAP03(2024)060

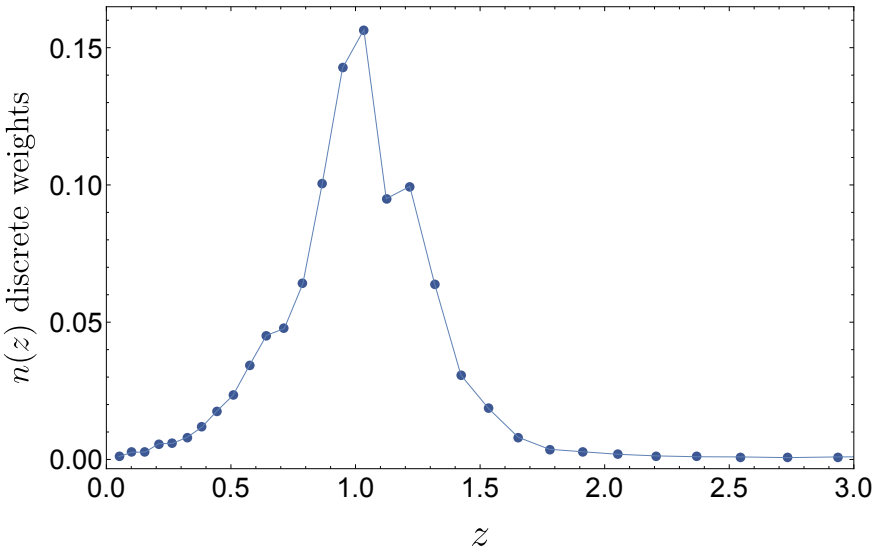

Figure 3. Weights applied to the discrete source planes of the Takahashi simulation to mimic the DES Y3 source distribution in its fourth redshift bin. The weights are normalised so that their sum is equal to one. The thin line is only there to guide the eye and does not have any meaning.

To fairly compare the theory and our measurement from the simulations, we further include a few corrections to the theoretical modelling described in section 4. These corrections are specific to the Takahashi simulation and take into account inaccuracies in the simulation itself rather than additional effects present in a real survey. Firstly, we take into account the fact that the simulation has discrete spherical lens shells of thickness 150 Mpc/h. This is done both by replacing the continuous integrations along the line-of-sight by discrete sums

$$\int \mathrm{d}\chi f(\chi) \to \sum_i 150 \times f[150(i-0.5)], \tag{5.4}$$

and at the level of the nonlinear power spectrum by correcting it following (equation (28) in appendix B of ref. [71]):

$$P_{\rm NL}(k,z) \to \frac{(1+c_1 k^{-\alpha_1})^{\alpha_1}}{(1+c_2 k^{-\alpha_2})^{\alpha_3}} P_{\rm NL}(k,z), \tag{5.5}$$

with $c_1 = 9.5171 \times 10^{-4}$, $c_2 = 5.1543 \times 10^{-3}$, $\alpha_1 = 1.3063$, $\alpha_2 = 1.1475$ and $\alpha_3 = 0.62793$. Note that the wave-modes k are in units of h/Mpc and the power spectrum in units of $(\mathrm{Mpc/h})^3$. Finally, in a manner analogous to taking into account the pixel window function at the map level, the resolution effects of the Takahashi maps can be taken into account by multiplying the nonlinear power-spectrum by a damping factor that depends on the NSIDE of the `Healpix` map (equation (5) in ref. [71]):

$$P_{\rm NL}(l/\chi(z),z) \to \frac{P_{\rm NL}(l/\chi(z),z)}{1+(l/1.6/\mathrm{NSIDE})^2}\,. \tag{5.6}$$

Note that taking into account corrections (5.5) and (5.6) at the level of the nonlinear power spectrum has some effects on all other higher-order (density contrast) cumulants through their scaling relations with the variance while keeps the S_n ratios constant in equation (2.4).

JCAP03(2024)060

5.2 DES Y3 mask and BNT transform

One issue faced by theoretical approaches that aim at describing quantities projected along the line-of-sight is the mixing of both very nonlinear scales not accurately probed by standard from-first-principles approaches such as ours, and reasonably larger (quasi-linear) scales more accessible to the theory. As such, usual weak lensing statistical probes are (i) modelled accurately only on sufficiently large angular scales with scale-cuts on small angular scales, so as to mitigate the influence of small nonlinear physical scales at the tip of the light-cone, (ii) modelled by more phenomenological approaches such as halo models that can also take into account nonlinear and baryonic physics which becomes important on small scales [74], (iii) or by making use of numerical recipes and simulations for e.g. specifically incorporating baryonic feedback effects [75] which have unfortunately not been tested in great detail, especially for higher-order non-Gaussian statistics.

Alternatively, a theoretical strategy to disentangle quasi- and non-linear scales in lensing quantities known as the Bernardeau-Nishimichi-Taruya (BNT) transform or nulling strategy was proposed by [73]. It allows for very accurate theoretical predictions in the context of power spectrum analysis and has recently been extended to the convergence and aperture-mass PDFs [15, 40]. This nulling strategy was used recently in [76] to remove the sensitivity to the poorly modelled small scales for the cosmic shear 2PCF, and therefore improve cosmological constraints using the DES Y1 shear data. This strategy could become even more relevant for future lensing experiments with better knowledge of redshifts and the division of sources in more redshift bins.

This BNT transform can only be used in the context of a tomographic analysis of at least 3 source redshifts (or redshift bins, although not treated here) and is a linear transformation M applied to the set of lensing kernels $\omega_i \equiv \omega(\chi, \chi_{s,i})$, giving rise to a new set of re-weighted kernels

$$\tilde{\omega}^j = M^{ij}\omega_i. \tag{5.7}$$

For a set of 3 source planes labeled from $j = i-2$ to $j = i$ arranged by ascending order, it was shown in [73] that M must satisfy the system

$$\begin{cases} \displaystyle\sum_{j=i-2}^{i} M^{ji} = 0, \\ \displaystyle\sum_{j=i-2}^{i} \frac{M^{ji}}{\chi_{s,j}} = 0\,. \end{cases} \tag{5.8}$$

This system of equations is under-constrained and hence we also impose by convention $M^{ii} = 1$. The elements of M can thus be computed considering sequential triplets of tomographic bins, going from the lowest to the highest redshift, such that

$$M^{i-2,i} = \frac{\chi_{i-2}(\chi_{i-1} - \chi_i)}{\chi_i(\chi_{i-2} - \chi_{i-1})}, \tag{5.9}$$

$$M^{i-1,i} = \frac{\chi_{i-1}(\chi_i - \chi_{i-2})}{\chi_i(\chi_{i-2} - \chi_{i-1})}. \tag{5.10}$$

JCAP03(2024)060

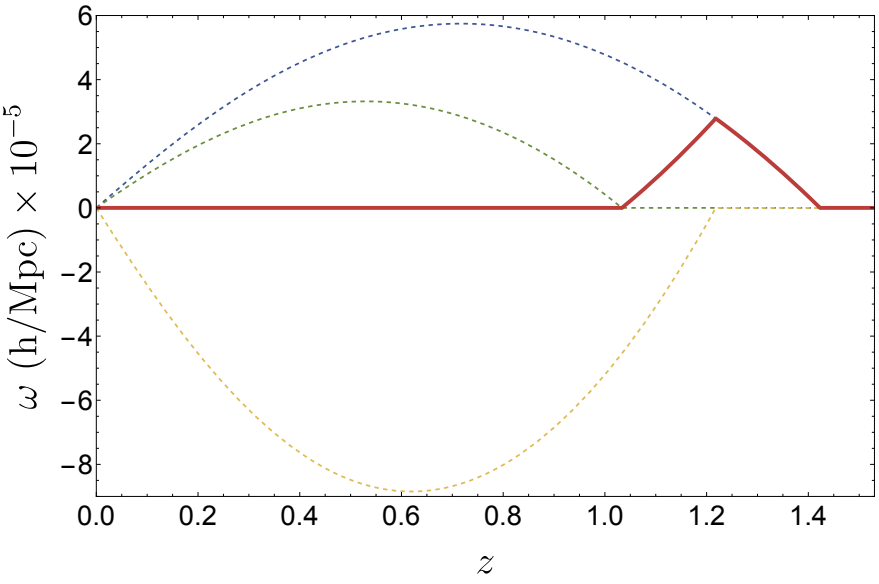

Figure 4. Illustration of the effect of the BNT transform on lensing kernels. The green, yellow and blue dashed lines are the kernels up to $z_s = 1.0, 1.2, 1.4$ respectively re-weighted by their appropriate BNT coefficients, $M^{ij} = [0.80, -1.80, 1]$ where j is fixed and equal to 3 if the blue kernel is the third of a tomographic analysis. The thick red line is the sum of the 3 re-weighted kernel. The effect of nulling is to set to zero the contribution of all lenses below the closest source plane.

We display in figure 4 an example for a set of 3 source planes located at $z_s = 1.0, 1.2, 1.4$. This is the kernel we use in this subsection. The green, yellow and blue dashed lines are the kernels up to $z_s = 1.0, 1.2, 1.4$ respectively re-weighted by their appropriate BNT coefficients while the thick red line is the sum of the 3 re-weighted kernels. Note that the blue dashed line is also the original kernel since its BNT coefficient is set to 1. One can thus clearly see that the effect of nulling is to set to zero the contribution of all lenses below the closest source plane and therefore remove the contribution of small scales (at the tip of our light-cone) which are very nonlinear and where the effect of baryonic physics becomes non-negligible.

For our purpose, the BNT transform — which boils down to a simple linear combination of the maps — is applied at the level of the masked shear field[14] and is also straightforward to implement in our theoretical approach to the convergence PDF since we only need to replace the original kernel with its nulled counterpart.

We then show in figure 5 — in the regime where the traditional BNT transformed PDF is perfectly described by the LDT formalism — how the theoretical integration of the convergence reconstruction scheme performs. The BNT convergence field is smoothed with a top-hat window of radius $\theta = 10\,\text{arcmin}$. By looking at the top panels, one can appreciate that

[14]Since the BNT transform is a linear transformation, it can without distinction be applied at the level of the (masked) shear or the convergence. The consideration of nonlinear high-order lensing or reduced shear corrections would formally break the exactness of the nulling but would still mitigate the influence of small scales so that it could still be used on real data. The same comment applies to the inexact knowledge of the underlying cosmology: an inexact nulling still mitigates the influence of small scales.

JCAP03(2024)060

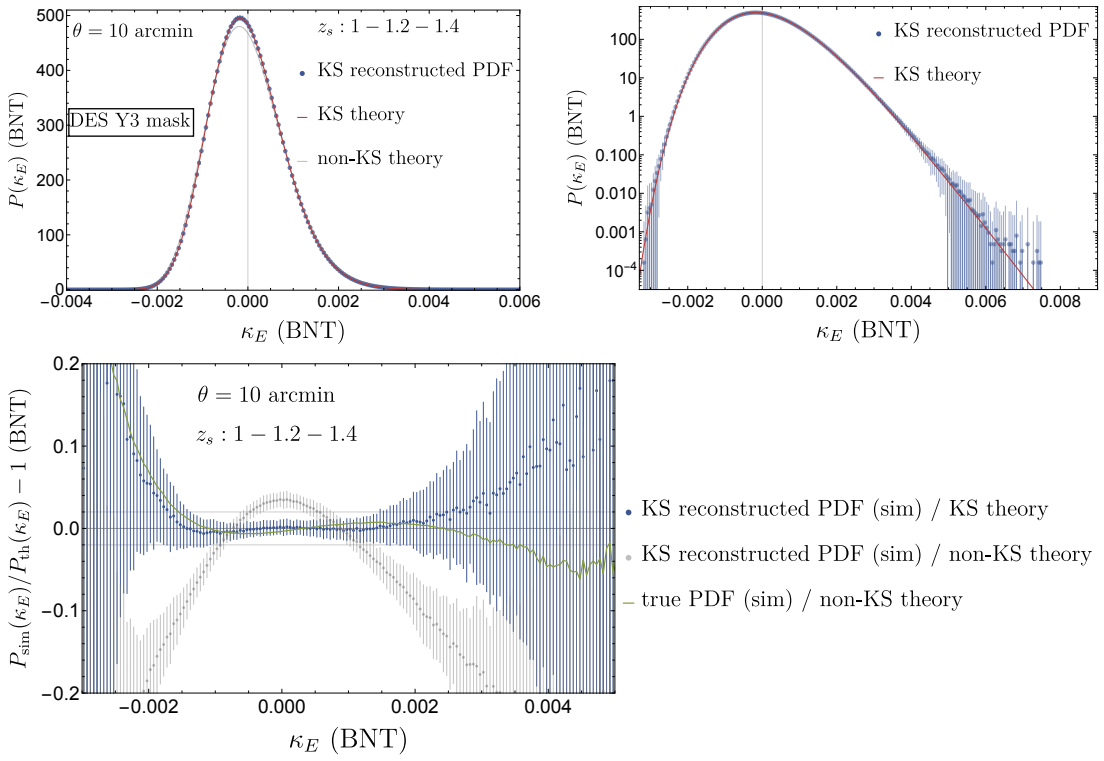

Figure 5. *Top panels*: PDF of the Kaiser-Squires reconstructed nulled convergence map using source planes located at $z_s = 1.0$, 1.2 and 1.4 under the DES Y3 mask. The red line is the theory described by equation (4.3) and the blue points with error bars are the mean and 1σ fluctuations of the measured PDF in 108 realisations. The grey solid line is the theory described by equation (2.17) that is without taking into account the KS reconstruction masking effect. *Bottom panel*: residuals between the theory and the simulation. The green line is the residual of the original non-KS theory to the true, non-reconstructed simulated convergence PDF. It is highly compatible with the residuals between the KS-theory and the simulated reconstructed KS PDF (blue points). Neglecting the KS reconstruction effects (grey points) significantly worsens the quality of the theoretical prediction while taking them into account does not improve nor worsen the predictive power of our LDT formalism.

the exponential cut-off in the tails of the PDF, a prediction of our formalism, is well-observed once one reduces the lensing kernel down to scales accessible to from-first-principles theoretical modelling (such as ours). More interestingly, this behaviour is still observed through the Kaiser-Squires reconstruction scheme presented in section 3, and our theoretical modelling of this reconstruction does not reduce our ability to capture the shape of the PDF in this regime. This hints at the fact that our model is accurate enough to reproduce the reconstruction effects, at least in the regime where our initial large-deviation formalism is accurate. This is further illustrated in the bottom panel of figure 5 where blue points denote the residuals of our KS theory with respect to the measurement from the KS reconstructed simulations. The green line describes the residual between the original LDT theory without reconstruction and the measurement of the PDF from the true κ map (i.e. not reconstructed from the shear field using KS method). Here we observe that the two residuals are highly compatible. Moreover, we also find that taking the reconstruction into account does not reduce in the

JCAP03(2024)060

slightest our ability to describe the PDF in this regime, while neglecting it (grey points) would seriously damage our predictive power.

5.3 DES Y3 mask for sources in the fourth DES Y3 redshift bin

For the more traditional case of a source distribution mimicking the 4^{th} source bin of the DES Y3 analysis (that we replicate using the source redshift distribution shown in figure 3), and for a shear field under the DES Y3 mask, we show in figure 6 the results of our theoretical prediction for the reconstructed κ_E PDF. The field is smoothed with a top-hat window of radius $\theta = 20\,\text{arcmin}$. By looking at the top-left panel, one can appreciate that the theoretical formalism seems to capture well the bulk of the PDF when considering the 1σ fluctuations across the 108 Takahashi realisations estimating the diagonal of the PDF covariance matrix. This is to be contrasted with the theoretical formalism that neglects the KS reconstruction effects and thus performs poorly which justifies the need for its modelling. Looking at the PDF on a log-scale (top-right panel), one can however observe an increasingly larger departure of the theory from the simulation as one enters the high and low-density tails. This is expected since cumulants from the large-deviations formalism are only valid in the quasi-linear regime and the mixing of scales (that the BNT transform would prevent as shown in subsection 5.2) worsens the quality of the theoretical prediction in the tails.

This is clearly seen in the residuals between the theory and the simulation (bottom panel) where a typical mismatched skewness modulation[15] is visible, showing that higher order correction to the skewness would be needed if we had smaller error bars, especially since only the cosmic variance is here used to obtain those error bars. The alternatives include looking at larger angular scales, considering sources at higher redshifts, incorporating modelling errors in the error budget or making use of the BNT transform shown in subsection 5.2 which we consider the most principled approach.

Focusing on the accuracy of the modelling of the KS reconstruction, we now look at the green line in the bottom panel which describes the residual of the original LDT theory without KS reconstruction to the measurement of the PDF from the true κ map (i.e. not reconstructed from the shear field using KS method). As in the previous subsection, we once again observe that the two residuals are highly compatible, highlighting the fact that the reconstruction does not weaken nor improve our ability to describe the PDF in this regime. This again hints at the fact that our implementation of the KS effects on the PDF is accurate enough.

6 Incorporating observational systematics in the modelling

In this section we discuss the modelling of additional effects in the reconstructed κ-PDF, including galaxy intrinsic alignments, baryonic feedback effects, additive and multiplicative shear biases, higher-order lensing corrections and photometric redshift uncertainties.

[15] Let us remind here that the skewness enters the Edgeworth expansion of the PDF at the first non-Gaussian correction order and multiplies a third order Hermite polynomial of the convergence field as follows

$$\mathcal{P}(\kappa) = \mathcal{G}(\kappa)\left[1 + \sigma\frac{S_{3,\kappa}}{3!}H_3\left(\frac{\kappa}{\sigma}\right) + \mathcal{O}(\sigma^2)\right],$$

where $H_3(x) = x^3 - x$. We emphasise that the predictions shown in this paper do not use or assume an Edgeworth expansion. It is however useful to interpret the shape of the residuals between the theory and the simulation.

JCAP03(2024)060

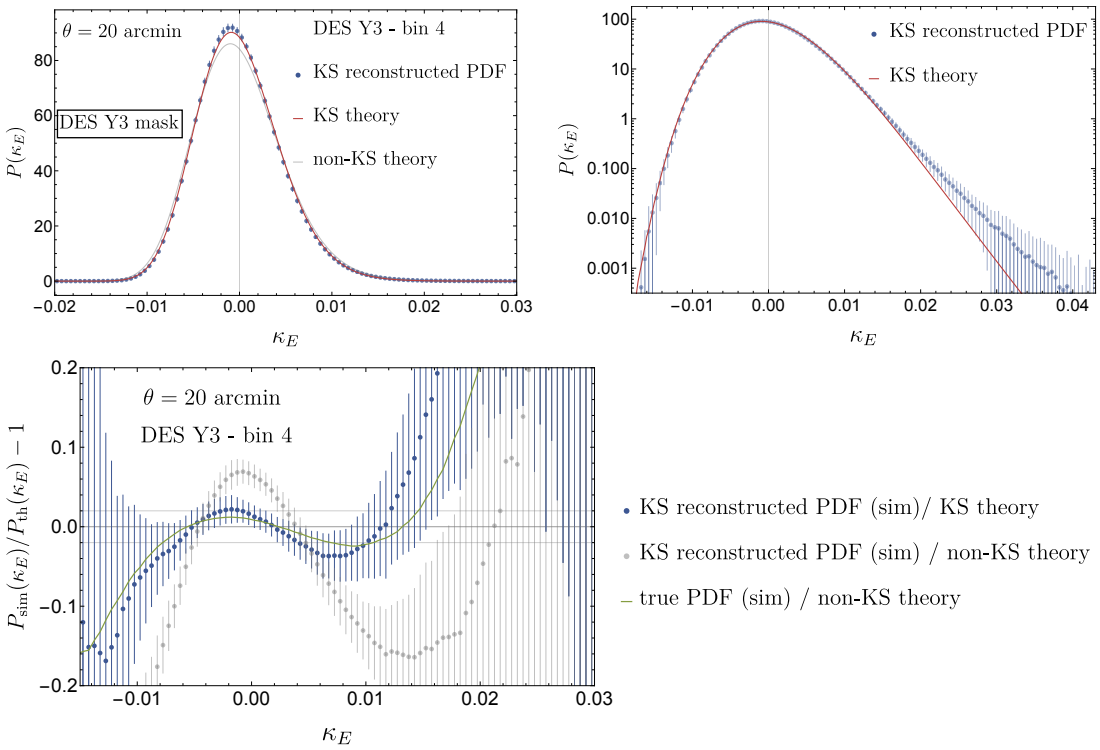

Figure 6. *Top panels*: PDF of the Kaiser-Squires reconstructed convergence map for the simulated DES Y3-like 4$^{\text{th}}$ bin in the Takahashi simulation and under the DES Y3 mask. The red line is the theory described by equation (4.3) and the blue points with error bars are the mean and 1σ fluctuations of the measured PDF in 108 masked Takahashi maps. The grey solid line is the theory described by equation (2.17) that is without taking into account the KS reconstruction. *Bottom panel*: residuals between the theory and the simulation. The green line is the residual of the original non-KS theory to the true, non-reconstructed simulated convergence PDF. It is highly compatible with the residuals between the KS-theory and the simulated reconstructed KS PDF (blue points). Neglecting the KS reconstruction effects (grey points) significantly worsens the quality of the theoretical prediction while taking them into account preserves the predictive power of our LDT formalism. We note that the slight increase in the size of residuals is due to the complicated shape of the mask and presence of holes as demonstrated in figure 9 for the case of the square mask of similar area.

6.1 Intrinsic alignments

As noted in equation (3.2), the observed source galaxy ellipticity is a combination of the gravitational lensing shear component γ, the intrinsic ellipticity of the galaxies ϵ_{IA} induced by correlations with local gravitational tidal fields at the source, and the random stochastic component that contributes as shape noise. In this section we include the ϵ_{IA} (also known as intrinsic alignment) component in our framework using the popular *non-linear tidal alignment* (NLA) model [77, 78]. Indeed, recent work based on a perturbative field-level forward modelling of weak lensing fields [25] indicates that on scales of about 15 arcmin, the NLA model is adequate and leads to conservative and unbiased cosmology constraints even when analysing data generated through a more complex tidal alignment and tidal torquing (TATT) intrinsic alignment model [79].

JCAP03(2024)060

As shown in section 3.2, through second derivatives of the projected gravitational potential ψ, the lensing shear γ is related to the lensing convergence κ, the latter being expressed as a line-of-sight integration of the matter density contrast δ (see equation (2.15)). In analogy, for the intrinsic ellipticity component $\epsilon_{\rm IA}$ we can define a quantity $\kappa_{\rm IA}$

$$\kappa_{\rm IA}(\boldsymbol{\vartheta}) \equiv \int_0^{+\infty} \mathrm{d}z \frac{\mathrm{d}\chi}{\mathrm{d}z}\, n(z)\, \delta_{\rm IA}(\chi(z), \chi(z)\boldsymbol{\vartheta}), \tag{6.1}$$

where $\delta_{\rm IA}$ is a three-dimensional field that *effectively* determines the intrinsic alignment (IA) of the source galaxies with their local gravitational tidal fields. Note that the line-of-sight projection kernel in equation (6.1) is the source galaxy redshift distribution $n(z)$, and not the lensing kernel $\omega_{n(z)}(z)$ as in equation (2.15).

In the NLA model, $\delta_{\rm IA}$ can be expressed as a first order bias expansion with the nonlinear matter density contrast field[16] [77, 78]:

$$\delta_{\rm IA}(\chi, \chi\boldsymbol{\vartheta}) = f_{\rm IA}(\chi(z))\ \delta(\chi, \chi\boldsymbol{\vartheta})\,. \tag{6.2}$$

The $f_{\rm IA}(z)$ term is the amplitude of the intrinsic alignment of the specific source galaxies and is given by

$$f_{\rm IA}(z) = -A_{\rm IA} \left(\frac{1+z}{1+z_0}\right)^{\alpha_{\rm IA}} \frac{c_1 \rho_{\rm crit} \Omega_{\rm m}}{D(z)}\,, \tag{6.3}$$

where $A_{\rm IA}$, $\alpha_{\rm IA}$ are redshift-independent parameters, $c_1 = 5 \times 10^{-14}\,(h^2\, M_\odot/{\rm Mpc}^3)^{-1}$ [80], $\rho_{\rm crit}$ is the critical energy density, $D(z)$ is the growth factor normalised to unity today, and z_0 is a pivot redshift. The fact that c_1 was calibrated in h^2 units allows us to fix $c_1 \rho_{\rm crit} = 0.0134$. Therefore, the intrinsic alignment of source galaxies can thus be readily included in the overall convergence signal as an additive component on top of the lensing κ

$$\kappa \to \kappa + \kappa_{\rm IA}\,. \tag{6.4}$$

In practice, the inclusion of the intrinsic alignment effect in the modelling of the PDF $P(\kappa)$ within the NLA framework is straightforward and can be achieved by replacing the lensing kernel in equation (2.15) with

$$\omega_{n(z)}(z) \to \omega_{n(z)}(z) + n(z) f_{\rm IA}(z)\,. \tag{6.5}$$

One can treat these intrinsic alignment terms as nuisance parameters which can be marginalised over in order to constrain cosmological parameters of interest when analysing $\mathcal{P}(\kappa)$. Note that even though marginalized constraints over the IA parameters in the DES Y3 shear 2PCF analysis are compatible with no IA (see for example [81] or [82]), neglecting the IA effect would bias the constraints on other parameters of interest. Based on those analyses, the typical order of magnitude of the IA effect would be around $A_{\rm IA} \sim 0.4$ and $\alpha_{\rm IA} \sim 1.7$ but again one should not fix those values in an analysis, also because the uncertainty on

[16] Strictly speaking, in the *linear tidal alignment* model one performs the bias expansion around the linear matter density contrast field $\delta_{\rm lin}$ (as is correct in perturbation theory), whereas in the NLA model one simply replaces $\delta_{\rm lin}$ with the nonlinear matter density field δ (as adopted in our work).

JCAP03(2024)060

σ_κ $[10^{-3}]$	DM only	fid AGN	high AGN	low AGN
full $n(z)$	3.955	3.919 $[-0.9\%]$	3.874 $[-2.0\%]$	3.944 $[-0.27\%]$
z-bin 2 $[0.43, 0.63]$	3.483	3.442 $[-1.2\%]$	3.398 $[-2.4\%]$	3.468 $[-0.45\%]$
z-bin 4 $[0.9, 1.3]$	6.149	6.129 $[-0.33\%]$	6.080 $[-1.2\%]$	6.154 $[+0.08\%]$

Table 1. Impact of bayronic effects on the standard deviation of the weak lensing convergence at scale $\theta = 10'$ for a DES Y1 $n(z)$ as measured from the BAHAMAS simulations.

the IA nuisance parameters is one of the limiting aspect of cosmic shear analysis that one might want to mitigate through the use of high-order statistics. For the PDF, the interplay with cosmological parameters will be studied in more detail in our upcoming work (paper II in the series). One potentially important follow-up line of work would consist in extending the current bias-like IA expansions to the PDF so that higher-order terms in models such like TATT to alleviate for potential biases in the IA treatment. Current DES Y3 analysis of convergence moments [39] or third-order aperture mass in KiDS-1000 [83] which both contain NLA modelling are nevertheless respectively compatible with the main 3×2 points analysis from DES Y3 which is made with TATT [82] and the KiDS-1000 2-point analysis [83].

6.2 Baryonic feedback

We determine the effect of baryonic feedback on the mildly nonlinear convergence PDF by relying on simulated DES Y1-like lensing maps from the BAHAMAS hydrodynamical simulation suite [84], where the strength of AGN feedback has been varied in various simulation runs. The convergence PDF for the full DES Y1 $n(z)$ and the separate redshift bins is illustrated in the upper panel of figure 7. Note that for our purpose, the distribution of sources between DES Y1 and Y3 is similar enough, specifically, the four bins roughly peak respectively at redshift $z_s = 0.3$, 0.5, 0.8 and 1. In the lower panel we show the residual between the PDFs including baryonic feedback and the dark-matter only runs, which shows a clear signature of a changed standard deviation, whose values we summarise in table 1. Conjecturing that the main impact is on the variance, we obtain the PDFs of the weak lensing convergence divided by its variance $\nu_\kappa = \kappa/\sigma_\kappa$ and show exemplary results for the whole $n(z)$ and two redshift bins in figure 8. The excellent agreement of the PDFs for the standard deviation-normalised convergence field demonstrates that modelling the impact of baryons at the level of the variance is likely sufficient, as assumed in the recent analysis of HSC Y1 data [29]. Using conservation of probability, this implies that the κ-PDF in the presence of baryonic feedback is given by

$$\mathcal{P}_b(\kappa_b) = \mathcal{P}_{\rm DM}\left(\kappa = \kappa_b \frac{\sigma_{\rm DM}}{\sigma_b}\right) \frac{\sigma_{\rm DM}}{\sigma_b}\,. \tag{6.6}$$

For the purpose of obtaining the variance correction factor, the baryonic feedback model within HMcode can be used as illustrated in figure 5 of [74] showing a simple single-parameter model reproducing the 3D matter power spectrum in BAHAMAS.

We have further checked (not shown) that our the results are fully compatible with similar measurements from kappaTNG maps for single source redshifts [85] obtained from

JCAP03(2024)060

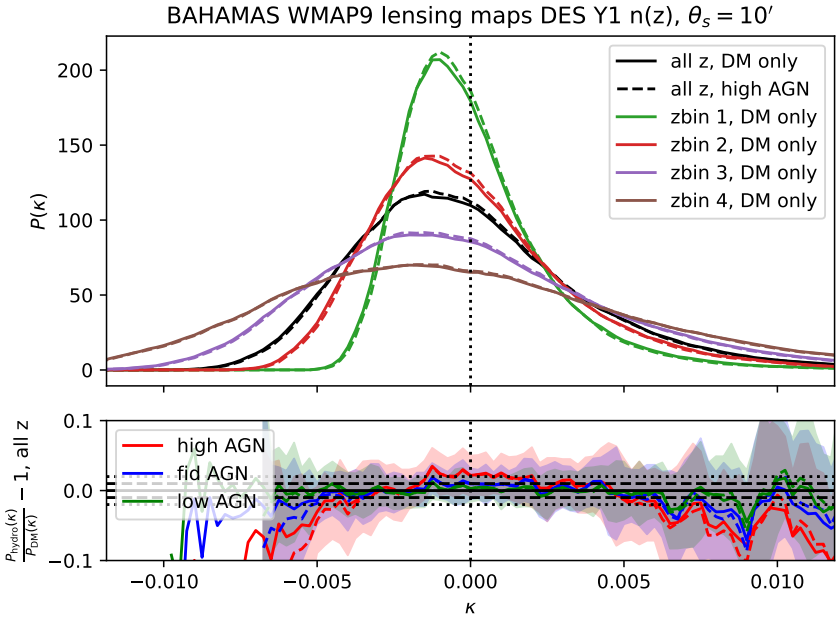

Figure 7. *Top panel*: smoothed convergence PDF from the BAHAMAS runs for the full DES Y1 $n(z)$ (black) and the 4 tomographic bins (colours) for the DM-only run (solid) and the high AGN run (dashed) averaged over 25 lensing cones. The dotted vertical line at $\kappa = 0$ highlights the non-Gaussian shape of the κ-PDF. *Bottom Panel*: residual of convergence PDF comparing the hydro BAHAMAS runs with fiducial (blue), high (red) and low (green) AGN feedback to the DM-only result for the whole $n(z)$. The solid lines indicate the residual between the means (averaged over 25 runs) while the shaded bands indicate the mean and standard deviation of the residual ratio in individual runs (obtained from 25 runs). The horizontal black dashed and dotted lines indicate a residual of 2% and 1%, respectively.

the IllustrisTNG hydrodynamical simulations [86], and for tomographic bins following the Euclid $n(z)$ as adopted by the Magneticum simulations [18, 87].

6.3 Additive and multiplicative shear biases

We adopt the modelling of any biases coming from the shear measurement pipeline, such as noise bias (e.g. [88]), model-fitting bias (e.g. [89]), selection bias (e.g. [90]) and bias from the imperfect correction of the image point spread function (PSF; e.g. [91]), with a multiplicative factor $1 + m$ to any instance of the estimated shear as is common in literature. As the weak lensing shear and convergence fields are connected to each other via linear transformations, we can propagate the shear measurement biases to convergence with the following transformation using a multiplicative $1 + m$ and an additive bias c term

$$\kappa \longrightarrow (1 + m)\kappa + c\,. \tag{6.7}$$

This linear relation holds in the weak lensing regime where κ is small. It is common in the literature [39, 82, 92] to consider that these biases are redshift and scale independent within a given source redshift distribution $n(z)$, and thus fixed at the map level. In that

JCAP03(2024)060

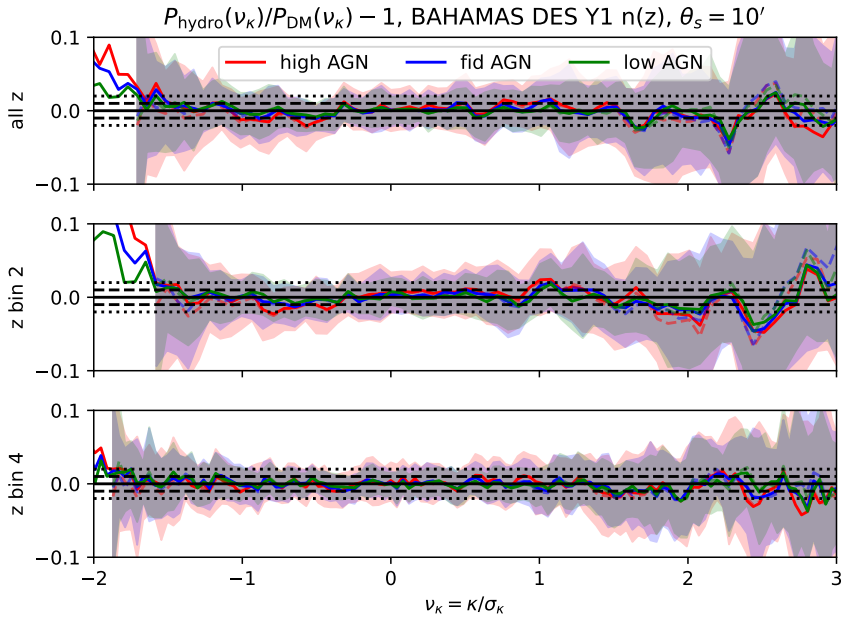

Figure 8. Residual of $\nu = \kappa/\sigma$ lensing PDF comparing the hydro BAHAMAS runs with fiducial (blue), high (red) or low (green) AGN feedback to the DM-only result averaged over 25 lensing cones smoothed for the whole $n(z)$ and in two different tomographic bins corresponding to DES Y1 $n(z)$. Bin 2 corresponds to the range $z \in [0.43, 0.63]$ while bin 4 corresponds to $z \in [0.9, 1.3]$. The solid lines indicate the residual between the means (averaged over 25 runs) while the shaded bands indicate the mean and standard deviation of the residual ratio in individual runs.

case, since the KS reconstruction of the convergence forces us to fix the $l = 0$ wave mode to zero as a straightforward consequence of the mass-sheet degeneracy, we are forcing the mean $\kappa_{(E/B)}$ to be zero which renders inconsistent the consideration of an additive bias c. This is consistent with the assumption that any additive bias component can be perfectly removed from the measurement pipeline for compactness through calibration with image simulations on average [93]. The PDF of the multiplicative biased shear $\tilde{\kappa} = (1 + m)\kappa$ is straightforwardly obtained by conservation of probability as

$$\mathcal{P}(\tilde{\kappa}) = \mathcal{P}\left(\kappa = \tilde{\kappa}/(1 + m)\right)/(1 + m)\,. \tag{6.8}$$

Previous works have indicated that the presence of shear biases enhances the complementarity of the shear 2PCF and the convergence PDF [26] which may help to further lift parameter degeneracies appearing at the 2-point level.

6.4 Higher-order lensing corrections

At the scales and redshift where the large-deviations formalism can be considered accurate enough to be applied to real data analysis, the corresponding weak-lensing PDFs of either the convergence or the aperture mass can be generated from a finite set of cumulants in the sense that a correct variance, skewness, and a consistent manner to generate higher-order cumulants

JCAP03(2024)060

yield good results [15, 40, 67]. In that context, studying the corrections to the PDF induced by higher-order lensing corrections amounts to computing the leading-order corrections on the variance and skewness which can then be incorporated straightforwardly to the formalism for the non-linear variance, and by slight modification of the spherical collapse ν parameter in equation (2.13) to match the new values of the skewness along the line-of-sight [see for example 67, that does it for post-Born corrections in the LDT context]. We detail in the following the corrections that could be considered in our theoretical formalism and explain why we could discard them for our DES Y3 analysis of the reconstructed κ_E PDF. This is consistent with what was done for the weak-lensing moments analysis in DES [39, 94].

Among all possible corrections, most of the traditional ones have been estimated in past works and in the LDT context. More precisely, the relaxation of the Born approximation and accounting for the coupling between lenses was studied in detail in [67] and in appendix F1 of [15]. These terms will tend to Gaussianise the convergence field since they characterise the introduction of random deflections along the light path which will in turn tend to diminish the impact of the non-linear clustering of matter. The heuristic picture one could form is that of clustered chunks of matter blurred by these lensing terms. It was shown that the effect matters for higher source redshifts (e.g. CMB lensing) but is totally negligible for cosmic shear experiments.

Reduced shear corrections which at first order account to replacing the shear field by $\gamma \to \gamma + \gamma\kappa$ was studied in the LDT context in [51] and in appendix F2 of [15] and was shown to induce only a percent-level change in the skewness at our scales and source-redshifts of interest. This tiny effect can be ignored for the analysis we propose here since the gravity-induced skewness directly implemented in the LDT formalism is itself not accurate to the percent-level because of the mixing of scales discussed in section 5.

Finally, individual galaxies can be (de)magnified through lensing effects and thus their fluxes are de-/increased. At the flux limit of a survey, this can cause fainter sources to be included in the observed sample while they would, in the absence of lensing, be excluded. At the same time, the density of galaxies in the small region around this source appears (increased) reduced since the area of the region is also (de)magnified. As such, the net effect depends on the slope of the intrinsic, unlensed, galaxy luminosity function at the survey's flux limit. This is known as the magnification bias and it also induces a correction on the overall measured skewness of the convergence field. However, as for the reduced shear correction, it was shown in appendix F3 of [15] to matter very little and we hence could discard its implementation in the theoretical formalism used to analyse DES Y3-like survey data in our upcoming works.

6.5 Photometric redshift uncertainties

In cosmic shear surveys the redshifts of source galaxies are determined using photometric methods. Any systematic error in the photometric estimates of the galaxies can lead to biases in the redshift distribution of the source galaxies which can in turn lead to biased cosmological parameter inference. In order to include the effect of such a systematic uncertainty on the source redshift distribution, one can propagate this through the theory by updating the weak lensing kernel equation (2.16). We could either compute this for a set of redshift distributions

JCAP03(2024)060

(obtained through different methods, for example the HYPERRANK software developed in the context of DES [95]) or parametrise it through a single shift parameter Δz via [39, 82, 95]

$$n(z) = n'(z + \Delta z)\,, \tag{6.9}$$

where $n(z)$ denotes the shifted redshift distribution of the source galaxies and $n'(z)$ the default redshift distribution estimate. This simple parametrization of a single mean redshift uncertainty (one for each tomographic redshift bin) is reported to be sufficient within the statistical power of surveys such as DES Y3 2-point main analysis (see figure 10 of [81] and figure 12 of [95]), and will be considered sufficient for this series of papers, in coherence with the previous high-order statistics analysis of current data sets [39, 83].

7 Discussion and conclusion

The weak lensing convergence κ denotes a weighted projection of the three-dimensional matter density fluctuations. Intuitively, it quantifies a projected 'mass' of all the late-time, non-Gaussian distributed foreground structures which contribute to the deformation of a light beam emanating from a background source galaxy on its way to us. Therefore, studying the full 1-point Probability Distribution Function (PDF) of the smoothed κ field inside apertures (see [32, 41]) is a promising way to access the non-Gaussian cosmological information of the foreground matter density field beyond the variance (or the widely used 2PCF) and holds the potential to tighten constraints on cosmological parameters.

Unfortunately, the κ field itself is not a direct observable as what is actually seen in lensing surveys is the *cosmic shear* field — the weak coherent distortions imprinted in the shapes of the source galaxies. Nevertheless, one can apply the widely known Kaiser-Squires (KS) inversion technique on the shear field to reconstruct the "true" inaccessible convergence field. However, the KS inversion is exact and recovers the "true" convergence only when one has information of the shear field over the entire celestial sphere.[17] This is of course not the case in practice as one has access only to the shear information inside a given footprint on the sky due to survey masks (consisting of holes, complicated survey geometry and boundaries). Applying the full-sky KS inversion on the masked shear field thus results in a reconstructed κ field which differs from the "true" inaccessible κ field. Hence, for any κ-statistic which one desires to measure and analyse using a KS reconstructed κ map from observed masked shear data, it is of paramount importance to correctly quantify and account for the effect of the survey mask in the theoretical modelling of the corresponding κ-statistic. In this paper we have presented for the first time how to do that in a from-first-principles theoretical modelling of the KS reconstructed κ-PDF. Our main achievements on this front can be summarised as follows:

- Our reconstructed full-sky convergence PDF is obtained from the "true" one [15, 40] inside the survey footprint and purely geometric factors which take into account the effect of the survey window and the survey area on the variance (4.6) and hence the series of cumulants (4.3). Explicitly, this is achieved by both an accurate parametrisation

[17] The KS inversion to reconstruct κ at a given location involves the convolution of the shear field over the full-sky with a specific kernel.

JCAP03(2024)060

(without free parameters) of the reconstructed full-sky convergence PDF as a function of the one inside the survey footprint and modifying the scaling relations of the matter density contrast field cumulants (which are needed to compute the line-of-sight projected κ cumulants that are in turn required in the modelling of the KS reconstructed κ-PDF) by purely geometric factors which take into account the effect of the survey window and the survey area.

- We have applied the recipe for the full-sky KS κ-map reconstruction under the presence of a realistic survey mask (in our case, DES Year 3 survey mask) to the simulated cosmic shear data from the Takahashi suite of weak lensing N-body simulations and measured the reconstructed κ-PDF. We have tested our theoretical modelling of the same against the measurements and found excellent agreement between them within cosmic variance (figure 6). We further find that using the baseline theoretical model for the "true" κ-PDF without accounting for the masking effect significantly deviates from simulation measurements of the reconstructed PDF. These conclusions are valid for scales and source redshifts relevant for the baseline theoretical model which is accurate on quasi-linear scales as we have also demonstrated in figure 5 where we have applied a *nulling strategy* to construct lensing observables less-sensitive to very small non-linear scales normally probed through scale-mixing in the projection along the line-of-sight.

- In preparation for an upcoming real data analysis of the KS reconstructed κ-PDF we have also discussed and laid down the strategy to include several effects such as astrophysical and measurement systematics as well as higher-order lensing corrections to the theoretical model for the reconstructed κ-PDF. We included a modelling for intrinsic alignments based on an adaptation of the weak lensing kernel (6.5) and tested in simulations that baryonic feedback can be included through a rescaling of the variance (6.6). We describe how the lensing PDF is affected by a multiplicative shear biases (6.8) and how to propagate photometric redshift uncertainties through our theoretical model.

Thus, our work not only presents a proper theory-based modelling framework for a real analysis of the KS reconstructed κ-PDF under the presence of realistic survey masks but it also underlines the susceptible errors when analysing any statistic from a KS reconstructed κ-map with a theoretical model that does not correctly include the E/B mode mixing due to the presence of survey masks.

Overall, our results indicate that the κ-PDF measured from the straightforward to implement spherical-sky KS reconstructed κ-map on the observed shear field (in the presence of masks) can be treated accurately within a theoretical framework without the need for any forward-modelling simulation-based approach. In particular, though some of the systematics modelling presented in this paper might need to be improved for Stage-IV surveys, the modelling of masks in combination with LDT in the context of the BNT transform will remain valid. This paves the way for us to explore the power of the κ-PDF in probing higher-order information in current lensing surveys such as DES, and in the future using Euclid and Vera Rubin's LSST data. We will perform cosmological inference analyses on simulated and real data in the following papers of this 'MAKING THE LEAP' series.

JCAP03(2024)060

Acknowledgments

AB's work is supported by the ORIGINS excellence cluster. CU is supported by the STFC Astronomy Theory Consolidated Grant ST/W001020/1 from U.K. Research & Innovation. The authors are grateful to Ryuichi Takahashi for making the weak lensing simulation suite used in this paper publicly available. The authors thank Ian McCarthy for making the BAHAMAS simulation maps publicly available, as well as Tiago Castro and Klaus Dolag for providing access to convergence maps from the Magneticum simulations. The authors are indebted to Nick Kaiser for his numerous and wide-ranging contributions to the field of large-scale structures, one of which is at the centre of the present work. May he rest in peace in the knowledge that his legacy is immortal.

Data availability. We will release all codes written for our theoretical predictions and map-making at the release of paper II of this 'MAKING THE LEAP' series. This will come with an emulator for our theoretical κ_E PDFs which renders easier MCMC analysis of either simulated or real data.

A Mode-coupling matrices in the pseudo-C_ℓ formalism

For completeness and better readability of this paper, we reproduce here key-results and broad derivation steps of the pseudo-C_ℓ formalism [44] that aims at taking masks into account in the computation of the shear power spectra. For the shear field which formally admits both E and B modes, we can define its full power spectra as a vector $\mathbf{C}_l \equiv (C_l^{EE}, C_l^{EB}, C_l^{BB})$ with

$$C_l^{ij} = \frac{1}{2l+1} \sum_m \gamma_{i,lm} \gamma^*_{j,lm}, \qquad \text{with} \quad i,j \in \{E,B\}. \tag{A.1}$$

We will then see that its masked power spectra $\hat{\mathbf{C}}_l$ can be seen as a convolution with a mode-mixing matrix $\mathbf{M}_{ll'}$ through

$$\hat{\mathbf{C}}_l = \sum_{l'} \mathbf{M}_{ll'} \mathbf{C}_{l'}. \tag{A.2}$$

The formalism thus takes into account transfers of power due to masking from pure E-modes spectra to pure B-modes and all other possible combinations. In this work, since we neglect any source of shear B-modes prior to masking, we are mostly interested in the element $M_{ll'}^{EE,EE}$ that allows to compute the contribution from the unmasked shear EE spectrum to the masked $\hat{C}_l^{EE}$. Let us now sketch the derivation of these terms.

In the presence of a footprint $K(\theta,\phi)$ (in our case the DES Y3 mask) the observed shear field becomes

$$\hat{\gamma}_1(\theta,\phi) + i\hat{\gamma}_2(\theta,\phi) = K(\theta,\phi)\left(\gamma_1(\theta,\phi) + i\gamma_2(\theta,\phi)\right). \tag{A.3}$$

Written in terms of spherical harmonics coefficients, this leads to the generation of both pseudo-E and pseudo-B modes

$$\hat{\gamma}_{E,lm} \pm i\hat{\gamma}_{B,lm} = \int \mathrm{d}\Omega \left[K(\theta,\phi)\left(\gamma_1(\theta,\phi) \pm i\gamma_2(\theta,\phi)\right)\right] {}_{\pm2}Y^*_{lm}(\theta,\phi) \tag{A.4}$$

JCAP03(2024)060

which can be expressed in terms of the original E and B modes via

$$\hat{\gamma}_{E,lm} \pm i\hat{\gamma}_{B,lm} = \sum_{l'm'} \left(\gamma_{E,lm} \pm i\gamma_{B,lm}\right) {}_{\pm2}W_{ll'mm'} \tag{A.5}$$

where

$$\begin{aligned} {}_{\pm2}W_{ll'mm'} &= \int \mathrm{d}\Omega {}_{\pm2}Y_{l'm'}(\theta,\phi) K(\theta,\phi) {}_{\pm2}Y^*_{lm}(\theta,\phi) \\ &= \sum_{l''m''} K_{l''m''} (-1)^m \sqrt{\frac{(2l+1)\,(2l'+1)\,(2l''+1)}{4\pi}} \begin{pmatrix} l & l' & l'' \\ \pm2 & \mp2 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ m & m' & m'' \end{pmatrix}. \end{aligned} \tag{A.6}$$

In the above, K_{lm} are the conjugate spherical harmonics coefficients of the window function $\int \mathrm{d}\Omega K(\theta,\phi) Y^*_{lm}(\theta,\phi)$, and $\left(\begin{smallmatrix} l & l' & l'' \\ m & m' & m'' \end{smallmatrix}\right)$ are the usual Wigner $3j$ symbols. Finally, combining the last result in equation (A.5) with the definition of the power spectra (A.1) for the pseudo-E and pseudo-B modes, one can then access the mode-mixing matrix coefficients and obtain

$$M^{EE,EE}_{ll'} = M^{BB,BB}_{ll'} = \frac{2l'+1}{8\pi} \sum_{l''} (2l''+1)\, K_{l''} \left[1 + (-1)^{l+l'+l''}\right] \begin{pmatrix} l & l' & l'' \\ 2 & -2 & 0 \end{pmatrix}^2, \tag{A.7}$$

$$M^{EE,BB}_{ll'} = M^{BB,EE}_{ll'} = \frac{2l'+1}{8\pi} \sum_{l''} (2l''+1)\, K_{l''} \left[1 - (-1)^{l+l'+l''}\right] \begin{pmatrix} l & l' & l'' \\ 2 & -2 & 0 \end{pmatrix}^2, \tag{A.8}$$

$$M^{EB,EB}_{ll'} = \frac{2l'+1}{4\pi} \sum_{l''} (2l''+1)\, K_{l''} \begin{pmatrix} l & l' & l'' \\ 2 & -2 & 0 \end{pmatrix}^2, \tag{A.9}$$

with $K_l \equiv 1/(2l+1) \sum_m K_{lm} K^*_{lm}$, the power spectrum of the mask. All other elements of the mode-mixing matrix are equal to zero.

B Masking effects on the Kaiser-Squires reconstructed convergence moments

Here we quantify the effect of the KS reconstruction scheme on the moments/cumulants of the κ_E field. The goal of this appendix is (i) to illustrate that the KS reconstruction has an important quantifiable effect on the variance and skewness, thus justifying the need for its theoretical modelling and (ii) to demonstrate that the theoretical model captures well this reconstruction effect on the variance and the skewness.

We provide in table 2 the values of the variance and skewness of the true convergence fields as measured in the Takahashi simulation (i.e. without any reconstruction) and the analytically computed (theory) values with a `halofit` power spectrum [96]. We also provide those same values for the reconstructed κ_E field. As expected from the mode-mixing formalism used in this paper, the modelling of the KS reconstruction perfectly recovers the effect on the variance as measured. The roughly $\sim 15\%$ loss in the value of the skewness as measured in the simulations (cf. the fractional difference of the measured $\langle\kappa^3\rangle_{c,\mathrm{true}}$ and $\langle\kappa^3_E\rangle_{c,\mathrm{unmasked}}$ values) is also well recovered by the theory modelling (cf. the fractional difference of the theory

JCAP03(2024)060

	$\sigma^2_{\kappa,\text{true}}$ (10^{-5})	$\sigma^2_{\kappa_E,\text{FS}}$ (10^{-6})	$\sigma^2_{\kappa_E,\text{unmasked}}$ (10^{-5})	$\langle\kappa^3\rangle_{c,\text{true}}$ (10^{-8})	$\langle\kappa^3_E\rangle_{c,\text{unmasked}}$ (10^{-8})
Measured	2.29 ± 0.02	2.43 ± 0.05	2.07 ± 0.05	5.56 ± 0.18	4.86 ± 0.47
Theory	2.30	2.44	2.08	4.97	4.05

Table 2. Smoothed second and third cumulants ($\theta = 20\,\text{arcmin}$) of the reconstructed κ_E under the DES Y3 mask on the shear field and for sources in the 4$^{\text{th}}$ DES Y3 redshift bin. All measurements are made in the Takahashi simulation and this table is supplementary to figure 6. We give values for the reconstructed full-sky κ_E variance and the variance and skewness values across the pixels under the mask that we take into account as described at the end of section 3.3.

$\langle\kappa^3\rangle_{c,\text{true}}$ and $\langle\kappa^3_E\rangle_{c,\text{unmasked}}$ values). As expected, the reconstruction does not significantly improve nor reduce the accuracy of the large-deviations formalism describing the underlying dynamics of the matter density field. The values given in table 2 are supplementary to figure 6 and are thus taken from a reconstructed field under the DES Y3 mask and for sources in the Takahashi simulation mimicking the DES Y3 4$^{\text{th}}$ bin displayed in figure 3.

C Square mask for sources in the DES Y3 4$^{\text{th}}$ bin

We here replicate the results of section 5.3, that is a source distribution (shown in figure 3) mimicking the 4$^{\text{th}}$ source bin of the DES Y3 analysis but this time using a square mask without holes of roughly the same area as the DES Y3 mask. Since by construction the KS effects on the variance are perfectly taken into account, any discrepancy in the modelling of the KS effects comes at the level of the skewness of the reconstructed κ_E field. As before, we quantify the quality of the KS modelling by comparing the residuals of the theoretical KS PDF to the simulated KS reconstruction PDF and the residuals of those same PDFs without any reconstruction. That way we visually do not include the intrinsic quality of the LDT prediction in our assessment.

Looking at the bottom panel of figure 9 we observe that the green line matches better the mean residuals of the theoretical KS PDF to the simulated KS reconstruction PDF (blue points) than what we could observe in the case of the true DES Y3 mask in figure 6. At the same time the impact of neglecting the KS effects in the modelling are less pronounced for this idealised mask (grey points). This is understandable to a certain extent. Indeed, the presence of multiple holes of various sizes within the DES Y3 mask makes the mixing of wave-modes in the E to E/B modes leakage more complicated to a degree where our mitigation (performed by restricting ourselves to the less affected pixels, cf. section 3.3) is less effective than in this simpler square-mask case. Note that similarly, finding it harder to mitigate the E/B mode leakage, would happen if the area of the square patch is small since the boundary effects would then affect more seriously the entire available area under the mask. The next step to improve the modelling would be to compute explicitly the convolution kernel on the shear bispectrum induced by the mask and include the new correction on the skewness in a manner analogous to how we would include higher-order weak lensing correction, i.e. by modification of the spherical collapse parameter ν in equation (2.13) [67]. This however does not seem to be necessary at this stage for a DES Y3 analysis.

JCAP03(2024)060

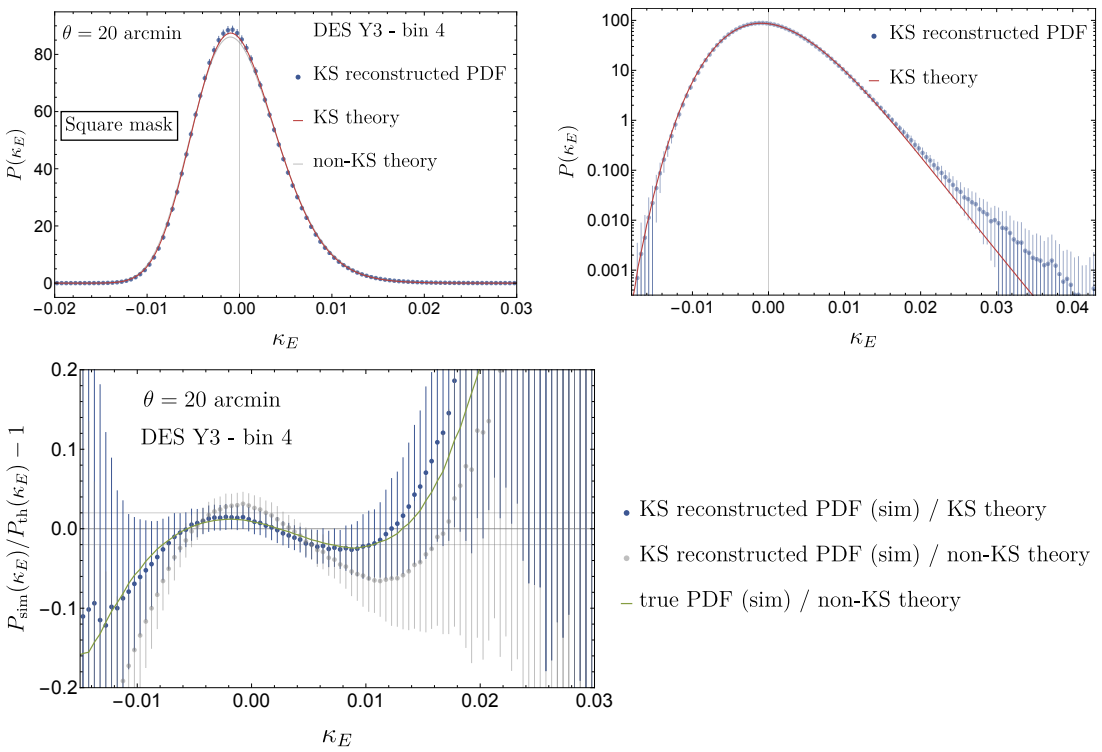

Figure 9. *Top panels*: PDF of the Kaiser-Squires weak-lensing convergence map for the DES Y3-like 4^{th} bin in the Takahashi simulation and under an idealised square-patch mask of 70×70 deg^2. The red line is the theory described by equation (4.3) and the blue points with error bars are the mean and 1σ fluctuations of the measured PDF in 108 Takahashi maps. The grey solid line is the theory described by equation (2.17) that is without taking into account the KS reconstruction. *Bottom panel*: residuals between the theory and the simulation. The green line is the residual of the original non-KS theory to the true, non-reconstructed simulated convergence PDF. It is highly compatible with the residuals of the KS-theory to the simulated reconstructed KS PDF (blue points). Though their inclusion allows to recover the full predictive power of our LDT formalism, neglecting the KS reconstruction effects is less imperative for this simple idealised square-mask, as opposed to the more realistic DES Y3 mask. This is understandable as discussed in appendix C.

References

[1] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, *Phys. Rep.* **340** (2001) 291 [astro-ph/9912508] [INSPIRE].

[2] C.S. Kochanek, *The Saas Fee Lectures on strong gravitational lensing*, in the proceedings of the *33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro*, Les Diablerets, Switzerland, 7–12 April 2003, astro-ph/0407232 [INSPIRE].

[3] DES collaboration, *The Dark Energy Survey: more than dark energy — an overview*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1270 [arXiv:1601.00329] [INSPIRE].

[4] J.T.A. de Jong, G.A. Verdoes Kleijn, K.H. Kuijken and E.A. Valentijn, *The Kilo-Degree Survey*, *Exp. Astron.* **35** (2013) 25 [arXiv:1206.1254] [INSPIRE].

JCAP03(2024)060

[5] H. Aihara et al., *The Hyper Suprime-Cam SSP Survey: Overview and Survey Design*, *Publ. Astron. Soc. Jpn.* **70** (2018) S4 [arXiv:1704.05858] [INSPIRE].

[6] R. Laureijs et al., *Euclid Definition Study Report*, arXiv:1110.3193 [INSPIRE].

[7] Ž. Ivezić et al., *LSST: from Science Drivers to Reference Design and Anticipated Data Products*, *Astrophys. J.* **873** (2019) 111 [arXiv:0805.2366] [INSPIRE].

[8] PLANCK collaboration, *Planck 2018 results. Part I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1 [arXiv:1807.06205] [INSPIRE].

[9] P. Schneider and M. Lombardi, *The Three-point correlation function of cosmic shear. Part I. The Natural components*, *Astron. Astrophys.* **397** (2003) 809 [astro-ph/0207454] [INSPIRE].

[10] P. Schneider, M. Kilbinger and M. Lombardi, *The Three-point correlation function of cosmic shear. Part 2. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures*, *Astron. Astrophys.* **431** (2005) 9 [astro-ph/0308328] [INSPIRE].

[11] A. Halder, O. Friedrich, S. Seitz and T.N. Varga, *The integrated three-point correlation function of cosmic shear*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 2780 [arXiv:2102.10177] [INSPIRE].

[12] A. Halder and A. Barreira, *Response approach to the integrated shear 3-point correlation function: the impact of baryonic effects on small scales*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 4639 [arXiv:2201.05607] [INSPIRE].

[13] Z. Gong, A. Halder, A. Barreira, S. Seitz and O. Friedrich, *Cosmology from the integrated shear 3-point correlation function: simulated likelihood analyses with machine-learning emulators*, *JCAP* **07** (2023) 040 [arXiv:2304.01187] [INSPIRE].

[14] A. Halder, Z. Gong, A. Barreira, O. Friedrich, S. Seitz and D. Gruen, *Beyond* 3×2*-point cosmology: the integrated shear and galaxy 3-point correlation functions*, *JCAP* **10** (2023) 028 [arXiv:2305.17132] [INSPIRE].

[15] A. Barthelemy, S. Codis and F. Bernardeau, *Probability distribution function of the aperture mass field with large deviation theory*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 5204 [arXiv:2012.03831] [INSPIRE].

[16] DES collaboration, *Dark Energy Survey Year 3 Results: Three-point shear correlations and mass aperture moments*, *Phys. Rev. D* **105** (2022) 103537 [arXiv:2201.05227] [INSPIRE].

[17] S. Heydenreich, L. Linke, P. Burger and P. Schneider, *A roadmap to cosmological parameter analysis with third-order shear statistics. Part I. Modelling and validation*, *Astron. Astrophys.* **672** (2023) A44 [arXiv:2208.11686] [INSPIRE].

[18] J. Harnois-Déraps et al., *Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 1623 [arXiv:2012.02777] [INSPIRE].

[19] DES collaboration, *Dark energy survey year 3 results: Cosmology with peaks using an emulator approach*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2075 [arXiv:2110.10135] [INSPIRE].

[20] DES collaboration, *Density Split Statistics: Cosmological Constraints from Counts and Lensing in Cells in DES Y1 and SDSS Data*, *Phys. Rev. D* **98** (2018) 023507 [arXiv:1710.05045] [INSPIRE].

[21] DES collaboration, *Density Split Statistics: Joint Model of Counts and Lensing in Cells*, *Phys. Rev. D* **98** (2018) 023508 [arXiv:1710.05162] [INSPIRE].

[22] P.A. Burger et al., *KiDS-1000 cosmology: Constraints from density split statistics*, *Astron. Astrophys.* **669** (2023) A69 [arXiv:2208.02171] [INSPIRE].

JCAP03(2024)060

[23] EUCLID collaboration, *Euclid Preparation. Part XXVIII. Forecasts for ten different higher-order weak lensing statistics*, *Astron. Astrophys.* **675** (2023) A120 [arXiv:2301.12890] [INSPIRE].

[24] N. Porqueres, A. Heavens, D. Mortlock and G. Lavaux, *Lifting weak lensing degeneracies with a field-based likelihood*, *Mon. Not. Roy. Astron. Soc.* **509** (2022) 3194 [arXiv:2108.04825] [INSPIRE].

[25] N. Porqueres, A. Heavens, D. Mortlock, G. Lavaux and T.L. Makinen, *Field-level inference of cosmic shear with intrinsic alignments and baryons*, arXiv:2304.04785 [INSPIRE].

[26] K. Patton et al., *Cosmological constraints from the convergence* 1*-point probability distribution*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 439 [arXiv:1611.01486] [INSPIRE].

[27] J. Liu and M.S. Madhavacheril, *Constraining neutrino mass with the tomographic weak lensing one-point probability distribution function and power spectrum*, *Phys. Rev. D* **99** (2019) 083508 [arXiv:1809.10747] [INSPIRE].

[28] A. Boyle et al., *Nuw CDM cosmology from the weak-lensing convergence PDF*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 2886 [arXiv:2012.07771] [INSPIRE].

[29] L. Thiele, G.A. Marques, J. Liu and M. Shirasaki, *Cosmological constraints from the Subaru Hyper Suprime-Cam year 1 shear catalogue lensing convergence probability distribution function*, *Phys. Rev. D* **108** (2023) 123526 [arXiv:2304.05928] [INSPIRE].

[30] B. Giblin, Y.-C. Cai and J. Harnois-Déraps, *Enhancing cosmic shear with the multiscale lensing probability density function*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 1721 [arXiv:2211.05708] [INSPIRE].

[31] C. Uhlemann et al., *It takes two to know one: Computing accurate one-point PDF covariances from effective two-point PDF models*, *Open J. Astrophys.* **6** (2023) 1 [arXiv:2210.07819] [INSPIRE].

[32] J.M. Zorrilla Matilla, M. Sharma, D. Hsu and Z. Haiman, *Interpreting deep learning models for weak lensing*, *Phys. Rev. D* **102** (2020) 123506 [arXiv:2007.06529] [INSPIRE].

[33] N. Kaiser and G. Squires, *Mapping the dark matter with weak gravitational lensing*, *Astrophys. J.* **404** (1993) 441 [INSPIRE].

[34] P. Fiedorowicz, E. Rozo, S.S. Boruah, C. Chang and M. Gatti, *KaRMMa — kappa reconstruction for mass mapping*, *Mon. Not. Roy. Astron. Soc.* **512** (2022) 73 [arXiv:2105.14699] [INSPIRE].

[35] S. Pires et al., *Euclid: Reconstruction of Weak Lensing mass maps for non-Gaussianity studies*, *Astron. Astrophys.* **638** (2020) A141 [arXiv:1910.03106] [INSPIRE].

[36] DES collaboration, *Dark Energy Survey Year 3 results: curved-sky weak lensing mass map reconstruction*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 4626 [arXiv:2105.13539] [INSPIRE].

[37] N. Jeffrey, F. Lanusse, O. Lahav and J.-L. Starck, *Deep learning dark matter map reconstructions from DES SV weak lensing data*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 5023 [arXiv:1908.00543] [INSPIRE].

[38] B. Remy et al., *Probabilistic Mass Mapping with Neural Score Estimation*, *Astron. Astrophys.* **672** (2023) A51 [arXiv:2201.05561] [INSPIRE].

[39] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps*, *Phys. Rev. D* **106** (2022) 083509 [arXiv:2110.10141] [INSPIRE].

[40] A. Barthelemy, S. Codis, C. Uhlemann, F. Bernardeau and R. Gavazzi, *A nulling strategy for modelling lensing convergence in cones with large deviation theory*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 3420 [arXiv:1909.02615] [INSPIRE].

JCAP03(2024)060

[41] A. Boyle, A. Barthelemy, S. Codis, C. Uhlemann and O. Friedrich, *The cumulant generating function as a novel observable to cumulate weak lensing information*, *Open J. Astrophys.* **6** (2023) 10351 [arXiv:2212.10351] [INSPIRE].

[42] F. Bernardeau and P. Reimberg, *Large deviation principle at play in large scale structure cosmology*, *Phys. Rev. D* **94** (2016) 063520 [arXiv:1511.08641] [INSPIRE].

[43] C. Uhlemann, S. Codis, C. Pichon, F. Bernardeau and P. Reimberg, *Back in the saddle: Large-deviation statistics of the cosmic log-density field*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1529 [arXiv:1512.05793] [INSPIRE].

[44] C. Hikage, M. Takada, T. Hamana and D. Spergel, *Shear Power Spectrum Reconstruction using Pseudo-Spectrum Method*, *Mon. Not. Roy. Astron. Soc.* **412** (2011) 65 [arXiv:1004.3542] [INSPIRE].

[45] F. Bernardeau and P. Valageas, *Construction of the one-point PDF of the local aperture mass in weak lensing maps*, *Astron. Astrophys.* **364** (2000) 1 [astro-ph/0006270] [INSPIRE].

[46] P.J.E. Peebles, *The large-scale structure of the universe*, Princeton University Press (1980).

[47] C.M. Baugh, E. Gaztanaga and G. Efstathiou, *A Comparison of the evolution of density fields in perturbation theory and numerical simulations. Part 2. Counts-in-cells analysis*, *Mon. Not. Roy. Astron. Soc.* **274** (1995) 1049 [astro-ph/9408057] [INSPIRE].

[48] C. Uhlemann et al., *Cylinders out of a top hat: counts-in-cells for projected densities*, *Mon. Not. Roy. Astron. Soc.* **477** (2018) 2772 [arXiv:1711.04767] [INSPIRE].

[49] S. Codis, F. Bernardeau and C. Pichon, *The large-scale correlations of multicell densities and profiles: implications for cosmic variance estimates*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1598 [arXiv:1602.03562] [INSPIRE].

[50] A. Barthelemy, F. Bernardeau, S. Codis and C. Uhlemann, *Numerical complexity of the joint nulled weak-lensing probability distribution function*, *Phys. Rev. D* **105** (2022) 043537 [arXiv:2106.11632] [INSPIRE].

[51] P. Reimberg and F. Bernardeau, *Large deviation principle at work: Computation of the statistical properties of the exact one-point aperture mass*, *Phys. Rev. D* **97** (2018) 023524 [arXiv:1708.00252] [INSPIRE].

[52] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, in *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag (1985).

[53] H. Touchette, *The large deviation approach to statistical mechanics*, *Phys. Rep.* **478** (2009) 1 [arXiv:0804.0327].

[54] H. Touchette, *A basic introduction to large deviations: Theory, applications, simulations*, arXiv:1106.4146.

[55] P. Valageas, *Dynamics of gravitational clustering. Part 2. Steepest-descent method for the quasi-linear regime*, *Astron. Astrophys.* **382** (2002) 412 [astro-ph/0107126] [INSPIRE].

[56] M.M. Ivanov, A.A. Kaurov and S. Sibiryakov, *Non-perturbative probability distribution function for cosmological counts in cells*, *JCAP* **03** (2019) 009 [arXiv:1811.07913] [INSPIRE].

[57] C. Uhlemann, E. Pajer, C. Pichon, T. Nishimichi, S. Codis and F. Bernardeau, *Hunting high and low: Disentangling primordial and late-time non-Gaussianity with cosmic densities in spheres*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 2853 [arXiv:1708.02206] [INSPIRE].

JCAP03(2024)060

[58] O. Friedrich, C. Uhlemann, F. Villaescusa-Navarro, T. Baldauf, M. Manera and T. Nishimichi, *Primordial non-Gaussianity without tails — how to measure $f_{\rm NL}$ with the bulk of the density PDF*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 464 [arXiv:1912.06621] [INSPIRE].

[59] F. Bernardeau, *The angular correlation hierarchy in the quasilinear regime*, *Astron. Astrophys.* **301** (1995) 309 [astro-ph/9502089] [INSPIRE].

[60] F. Bernardeau, *The Nonlinear evolution of rare events*, *Astrophys. J.* **427** (1994) 51 [astro-ph/9311066] [INSPIRE].

[61] P. Valageas, *Structure formation: A Spherical model for the evolution of the density distribution*, *Astron. Astrophys.* **337** (1998) 655 [astro-ph/9807033] [INSPIRE].

[62] Y. Mellier, *Probing the universe with weak lensing*, *Ann. Rev. Astron. Astrophys.* **37** (1999) 127 [astro-ph/9812172] [INSPIRE].

[63] J. Alsing, D. Kirk, A. Heavens and A. Jaffe, *Weak Lensing with Sizes, Magnitudes and Shapes*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 1202 [arXiv:1410.7839] [INSPIRE].

[64] C.G.R. Wallis, M.A. Price, J.D. McEwen, T.D. Kitching, B. Leistedt and A. Plouviez, *Mapping dark matter on the celestial sphere with weak gravitational lensing*, *Mon. Not. Roy. Astron. Soc.* **509** (2022) 4480 [arXiv:1703.09233] [INSPIRE].

[65] P.G. Castro, A.F. Heavens and T.D. Kitching, *Weak lensing analysis in three dimensions*, *Phys. Rev. D* **72** (2005) 023516 [astro-ph/0503479] [INSPIRE].

[66] R.K. Sachs, *Gravitational waves in general relativity. Part 6. The outgoing radiation condition*, *Proc. Roy. Soc. Lond. A* **264** (1961) 309 [INSPIRE].

[67] A. Barthelemy, S. Codis and F. Bernardeau, *Post-Born corrections to the one-point statistics of (CMB) lensing convergence obtained via large deviation theory*, *Mon. Not. Roy. Astron. Soc.* **494** (2020) 3368 [arXiv:2002.03625] [INSPIRE].

[68] G. Fabbian, M. Calabrese and C. Carbone, *CMB weak-lensing beyond the Born approximation: a numerical approach*, *JCAP* **02** (2018) 050 [arXiv:1702.03317] [INSPIRE].

[69] G. Fabbian, A. Lewis and D. Beck, *CMB lensing reconstruction biases in cross-correlation with large-scale structure probes*, *JCAP* **10** (2019) 057 [arXiv:1906.08760] [INSPIRE].

[70] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[71] R. Takahashi et al., *Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments*, *Astrophys. J.* **850** (2017) 24 [arXiv:1706.01472] [INSPIRE].

[72] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [INSPIRE].

[73] F. Bernardeau, T. Nishimichi and A. Taruya, *Cosmic shear full nulling: sorting out dynamics, geometry and systematics*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 1526 [arXiv:1312.0430] [INSPIRE].

[74] A. Mead, S. Brieden, T. Tröster and C. Heymans, *HMCODE-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 1401 [arXiv:2009.01858] [INSPIRE].

[75] A. Schneider et al., *Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation*, *JCAP* **03** (2019) 020 [arXiv:1810.08629] [INSPIRE].

JCAP03(2024)060

[76] P.L. Taylor, F. Bernardeau and E. Huff, *x-cut Cosmic Shear: Optimally Removing Sensitivity to Baryonic and Nonlinear Physics with an Application to the Dark Energy Survey Year 1 Shear Data*, *Phys. Rev. D* **103** (2021) 043531 [arXiv:2007.00675] [INSPIRE].

[77] C.M. Hirata et al., *Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: Luminosity and redshift scalings and implications for weak lensing surveys*, *Mon. Not. Roy. Astron. Soc.* **381** (2007) 1197 [astro-ph/0701671] [INSPIRE].

[78] S. Bridle and L. King, *Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements*, *New J. Phys.* **9** (2007) 444 [arXiv:0705.0166] [INSPIRE].

[79] J. Blazek, N. MacCrann, M.A. Troxel and X. Fang, *Beyond linear galaxy alignments*, *Phys. Rev. D* **100** (2019) 103506 [arXiv:1708.09247] [INSPIRE].

[80] M.L. Brown, A.N. Taylor, N.C. Hambly and S. Dye, *Measurement of intrinsic alignments in galaxy ellipticities*, *Mon. Not. Roy. Astron. Soc.* **333** (2002) 501 [astro-ph/0009499] [INSPIRE].

[81] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration*, *Phys. Rev. D* **105** (2022) 023514 [arXiv:2105.13543] [INSPIRE].

[82] DES collaboration, *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty*, *Phys. Rev. D* **105** (2022) 023515 [arXiv:2105.13544] [INSPIRE].

[83] P.A. Burger et al., *KiDS-1000 cosmology: Combined second- and third-order shear statistics*, arXiv:2309.08602 [INSPIRE].

[84] I.G. McCarthy, J. Schaye, S. Bird and A.M.C. Le Brun, *The BAHAMAS project: Calibrated hydrodynamical simulations for large-scale structure cosmology*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2936 [arXiv:1603.02702] [INSPIRE].

[85] K. Osato, J. Liu and Z. Haiman, *κTNG: effect of baryonic processes on weak lensing with IllustrisTNG simulations*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 5593 [arXiv:2010.09731] [INSPIRE].

[86] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [arXiv:1707.03397] [INSPIRE].

[87] T. Castro, M. Quartin, C. Giocoli, S. Borgani and K. Dolag, *The effect of baryons in the cosmological lensing PDFs*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 1305 [arXiv:1711.10017] [INSPIRE].

[88] T. Kacprzak et al., *Measurement and Calibration of Noise Bias in Weak Lensing Galaxy Shape Estimation*, *Mon. Not. Roy. Astron. Soc.* **427** (2012) 2711 [arXiv:1203.5049] [INSPIRE].

[89] L.M. Voigt and S.L. Bridle, *Limitations of model fitting methods for lensing shear estimation*, *Mon. Not. Roy. Astron. Soc.* **404** (2010) 458 [arXiv:0905.4801] [INSPIRE].

[90] C.M. Hirata and U. Seljak, *Shear calibration biases in weak lensing surveys*, *Mon. Not. Roy. Astron. Soc.* **343** (2003) 459 [astro-ph/0301054] [INSPIRE].

[91] S. Paulin-Henriksson, A. Amara, L. Voigt, A. Refregier and S.L. Bridle, *Requirements on PSF Calibration for Dark Energy from Cosmic Shear*, *Astron. Astrophys.* **484** (2008) 67 [arXiv:0711.4886] [INSPIRE].

[92] DES collaboration, *Dark Energy Survey Y3 results: blending shear and redshift biases in image simulations*, *Mon. Not. Roy. Astron. Soc.* **509** (2022) 3371 [arXiv:2012.08567] [INSPIRE].

JCAP03(2024)060

[93] B.R. Gillis and A.N. Taylor, *The Effects of Calibration on the Bias of Shear Measurements*, *Mon. Not. Roy. Astron. Soc.* **482** (2019) 402 [arXiv:1809.09540] [INSPIRE].

[94] DES collaboration, *Dark Energy Survey Year 3 results: cosmology with moments of weak lensing mass maps — validation on simulations*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 4060 [arXiv:1911.05568] [INSPIRE].

[95] DES collaboration, *Dark Energy Survey Year 3 results: marginalization over redshift distribution uncertainties using ranking of discrete realizations*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2170 [arXiv:2109.09636] [INSPIRE].

[96] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the Halofit Model for the Nonlinear Matter Power Spectrum*, *Astrophys. J.* **761** (2012) 152 [arXiv:1208.2701] [INSPIRE].

JCAP03(2024)060

Chapter 9

Summary and conclusions

The current and upcoming galaxy photometric and spectroscopic surveys provide us with a rich information to study the large-scale structure (LSS) in the Universe and gain insights into the mysterious energy components, the dark matter and the dark energy. However, the amount of data that we need to perform statistical analyses with high precision is enormous, and together with the task of optimally extracting cosmological information from them, they impose challenges that require spontaneous responses. Researchers have been dedicated to the statistical analyses on galaxy positions (2D or 3D) and shapes induced by weak gravitational lensing on two fronts: the analytical modeling and machine learning (ML) techniques. In this thesis, I contribute to the progress in both directions and explored the potential of combining the two approaches in different ways, with weak gravitational lensing fields (both shear and convergence) as the main target.

One representative statistic in analyzing the LSS is the 2-point correlation function (2PCF) which can be used to study the clustering of galaxies and weak lensing signals. In Chapter. 2, I briefly reviewed the analytical modeling of the 2PCF in the context of weak lensing, and introduced the concept of multiprobe cosmology where the 2PCFs of different observables are combined to extract cosmological information. However, besides the weak lensing fields themselves, it was shown that the 2PCFs of critical points, which are biased tracers of the underlying weak lensing fields such as peaks (maxima), voids (minima) and saddle points, can provide complementary constraining power on cosmological parameters. In Chapter. 5, I present the analytical modeling of the auto- and cross-2PCFs among critical points, considering a mild non-Gaussianity in the underlying weak lensing fields. This is the first time that the clustering of critical points in weak lensing has taken the non-Gaussian probability density distribution (PDF) of the underlying fields into account. Compared to the previous works that assumed Gaussian distribution, this new approach of modeling is more realistic in the context of LSS in late Universe. As current mainstream method for analyzing weak lensing maxima and minima clustering is based on N-body simulations and ML techniques, this analytical modeling can provide a well-defined calibration and complementary calculation for these statistics on large scales.

Back to the ML side, in Chapter. 3 I introduce the basic concepts of ML and some of its pop-

ular model architectures such as multi-layer perceptrons (MLPs) (Sec. 3.1), convolutional neural networks (CNNs) (Sec. 3.2). Some more advanced topics are also discussed such as the simulation-based inference (SBI) (Sec. 3.3). I demonstrate how these ML techniques can be utilized to enhance the statistical analyses of weak lensing fields. In particular, in Chapter. 6, I present the application of MLPs to the emulation of a specific weak lensing shear statistic, the integrated 3-point correlation function (i3PCF), which is promised to extract higher-order information of the cosmological parameters beyond Gaussianity accessed by the 2PCFs. This emulation, together with other GPU-based algorithms, greatly reduce the computational cost of the parameter inference task using i3PCFs. With this efficient pipeline, we are able to show on simulated likelihood analysis that the i3PCF can improve the constraints on A_s, the amplitude of the primordial power spectrum, and w_0, the dark energy equation of state parameter, by $10-25\%$ on top of the 2PCFs. We are currently working on the real Dark Energy Survey (DES) Year 3 data analysis with this higher-order statistic and the ML-supported pipeline, and the results will be published soon.

One step further, in Chapter. 7, I present a special ML model which fuses the CNN architecture and the formalism of correlation functions, the cosmological correlator convolutional neural network (C3NN). This model is designed to directly learn and output summary statistics that can be mathematically expressed in terms of N-point correlation functions (NPCFs). It possesses the property of being interpretable both in the outputs and the learned weights of the CNN layer. I demonstrate how helpful it can be in classifying different cosmological models with their evolved weak lensing convergence maps as inputs. We are currently working on embedding the C3NN model into the SBI framework, with aim of performing the cosmological parameter inference and the quantitative investigation of the information content in each order of the NPCFs, as well as the information existing beyond this correlation function ladder.

For an overarching view of the interactions between analytical methods and ML techniques, I give another discussion in Chapter. 4 as a summary for the previous introduction chapters (Chapter. 2 to 3) and a starter for the following Chapters. 5 to 7. In Chapter. 8, I show another co-author work on the modeling of weak lensing convergence PDF with large deviation theory (LDT), including major realistic survey and astrophysical systematic effects.

In the end, the thesis shows our works on major aspects of the weak lensing statistical analyses. And I would like to convey the message that the fusion of analytical methods with ML is redefining how cosmologists analyze data, test theories, and make discoveries. By combining physical principles with data-driven techniques, researchers can potentially tackle previously intractable problems while ensuring robustness and interpretability. As computational methods continue to advance, this integration will play a crucial role in future cosmological experiments, driving deeper insights into the nature of the Universe.

Bibliography

Abazajian, K., et al. 2022, arXiv:2203.08024, doi: 10.48550/arXiv.2203.08024

Abbott, T. M., Aguena, M., Alarcon, A., et al. 2022, Physical Review D, 105, 023520

Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. 2022, Journal of High Energy Astrophysics, 34, 49

Aghanim, N., et al. 2020, Astron. Astrophys., 641, A1, doi: 10.1051/0004-6361/201833880

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, Publications of the Astronomical Society of Japan, 70, S4

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. 2019, arXiv e-prints, arXiv:1907.10902, doi: 10.48550/arXiv.1907.10902

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, , 488, 4440, doi: 10.1093/mnras/stz1960

Amon, A., & Efstathiou, G. 2022, Monthly Notices of the Royal Astronomical Society, 516, 5355–5366, doi: 10.1093/mnras/stac2429

Anau Montel, N., Alvey, J., & Weniger, C. 2024, arXiv e-prints, arXiv:2412.15100, doi: 10.48550/arXiv.2412.15100

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, , 304, 15, doi: 10.1086/164143

Barreira, A., & Schmidt, F. 2017, , 2017, 053, doi: 10.1088/1475-7516/2017/06/053

Bartelmann, M., & Schneider, P. 2001, , 340, 291, doi: 10.1016/S0370-1573(00)00082-X

Barthelemy, A., Codis, S., & Bernardeau, F. 2021, , 503, 5204, doi: 10.1093/mnras/stab818

Berlind, A. A., & Weinberg, D. H. 2002, , 575, 587, doi: 10.1086/341469

Bernal, J. L., & Kovetz, E. D. 2022, The Astronomy and Astrophysics Review, 30, doi: 10.1007/s00159-022-00143-0

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, PHYSREP, 367, 1, doi: 10.1016/S0370-1573(02)00135-7

Bigwood, L., et al. 2024, Mon. Not. Roy. Astron. Soc., 534, 655, doi: 10.1093/mnras/stae2100

Blinnikov, S., & Moessner, R. 1998, , 130, 193, doi: 10.1051/aas:1998221

Bocquet, S., Grandis, S., Krause, E., et al. 2024, arXiv preprint arXiv:2412.07765

Bonzanini, A. D., Shao, K., Graves, D. B., Hamaguchi, S., & Mesbah, A. 2023, Plasma Sources Science Technology, 32, 024003, doi: 10.1088/1361-6595/acb28c

Breivik, K., Connolly, A. J., Ford, K., et al. 2022, arXiv preprint arXiv:2208.02781

Britt, D., Gruen, D., Friedrich, O., Yuan, S., & Ried Guachalla, B. 2024, , 689, A253, doi: 10.1051/0004-6361/202450266

Brunel, A., Pasquet, J., Pasquet, J., et al. 2019, arXiv e-prints, arXiv:1901.00461, doi: 10.48550/arXiv.1901.00461

Cavanagh, M. K., Bekki, K., & Groves, B. A. 2021, , 506, 659, doi: 10.1093/mnras/stab1552

Cesa, G., Lang, L., & Weiler, M. 2022, in International Conference on Learning Representations. https://openreview.net/forum?id=WE4qe9xlnQw

Charnock, T., Lavaux, G., & Wandelt, B. D. 2018, , 97, 083004, doi: 10.1103/PhysRevD.97.083004

Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, , 499, 5902, doi: 10.1093/mnras/staa3165

Chollet, F. 2017, Deep Learning with Python (Manning)

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, APJL, 648, L109, doi: 10.1086/508162

Cranmer, K., Brehmer, J., & Louppe, G. 2020, Proceedings of the National Academy of Science, 117, 30055, doi: 10.1073/pnas.1912789117

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270, doi: 10.1093/mnras/stw641

Davies, C. T., Cautun, M., Giblin, B., et al. 2022, , 513, 4729, doi: 10.1093/mnras/stac1204

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2012, The Astronomical Journal, 145, 10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, The Astronomical Journal, 151, 44

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25, doi: 10.1007/s10686-012-9306-1

Desjacques, V., Jeong, D., & Schmidt, F. 2018, , 733, 1, doi: 10.1016/j.physrep.2017.12.002

Devon Lin, C., & Tang, B. 2022, arXiv e-prints, arXiv:2203.06334, doi: 10.48550/arXiv.2203.06334

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, The Astronomical Journal, 157, 168

Dodelson, S., & Schmidt, F. 2020, Modern Cosmology (Elsevier Science). https://books.google.de/books?id=GGjfywEACAAJ

Driver, S. P., & Robotham, A. S. G. 2010, MNRAS, 407, 2131, doi: 10.1111/j.1365-2966.2010.17028.x

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. 2004, arXiv Mathematics e-prints, math/0406456, doi: 10.48550/arXiv.math/0406456

Efstathiou, G., & Gratton, S. 2020, MNRAS, 496, L91, doi: 10.1093/mnrasl/slaa093

Eifler, T., Miyatake, H., Krause, E., et al. 2021, , 507, 1746, doi: 10.1093/mnras/stab1762

Fang, X., Krause, E., Eifler, T., et al. 2024, , 527, 9581, doi: 10.1093/mnras/stad3808

Fluri, J., Kacprzak, T., Lucchi, A., et al. 2022, , 105, 083518, doi: 10.1103/PhysRevD.105.083518

Fotopoulou, S. 2024, Astronomy and Computing, 48, 100851, doi: 10.1016/j.ascom.2024.100851

Gatti, M., Sheldon, E., Amon, A., et al. 2021, Monthly Notices of the Royal Astronomical Society, 504, 4312

Gatti, M., et al. 2022, Phys. Rev. D, 106, 083509, doi: 10.1103/PhysRevD.106.083509

Gatti, M., Campailla, G., Jeffrey, N., et al. 2025, , 111, 063504, doi: 10.1103/PhysRevD.111.063504

Gebhardt, M., Anglés-Alcázar, D., Borrow, J., et al. 2024, , 529, 4896, doi: 10.1093/mnras/stae817

Gewers, F. L., Ferreira, G. R., de Arruda, H. F., et al. 2018, arXiv e-prints, arXiv:1804.02502, doi: 10.48550/arXiv.1804.02502

Gholamalinezhad, H., & Khosravi, H. 2020, arXiv e-prints, arXiv:2009.07485, doi: 10.48550/arXiv.2009.07485

Giri, S. K., & Schneider, A. 2023, BCemu: Model baryonic effects in cosmological simulations, Astrophysics Source Code Library, record ascl:2308.010

Gong, Z., Halder, A., Barreira, A., Seitz, S., & Friedrich, O. 2023, JCAP, 2023, 040, doi: 10.1088/1475-7516/2023/07/040

Greenberg, D. S., Nonnenmacher, M., & Macke, J. H. 2019, arXiv e-prints, arXiv:1905.07488, doi: 10.48550/arXiv.1905.07488

Greydanus, S., Dzamba, M., & Yosinski, J. 2019, arXiv e-prints, arXiv:1906.01563, doi: 10.48550/arXiv.1906.01563

Gruen, D. 2015, Weak lensing by galaxy clusters, Ludwig-Maximilians-Universität München. http://nbn-resolving.de/urn:nbn:de:bvb:19-183024

Hahn, C., Eickenberg, M., Ho, S., et al. 2022, arXiv e-prints, arXiv:2211.00723, doi: 10.48550/arXiv.2211.00723

—. 2024, , 109, 083534, doi: 10.1103/PhysRevD.109.083534

Halder, A., & Barreira, A. 2022, Monthly Notices of the Royal Astronomical Society, 515, 4639–4654, doi: 10.1093/mnras/stac2046

Halder, A., Friedrich, O., Seitz, S., & Varga, T. N. 2021, Monthly Notices of the Royal Astronomical Society, 506, 2780–2803, doi: 10.1093/mnras/stab1801

Halder, A., Gong, Z., Barreira, A., et al. 2023, JCAP, 2023, 028, doi: 10.1088/1475-7516/2023/10/028

Harnois-Deraps, J., et al. 2024, Mon. Not. Roy. Astron. Soc., 534, 3305, doi: 10.1093/mnras/stae2249

Hassija, V., Chamola, V., Mahapatra, A., et al. 2024, Cognitive Computation, 16, 45, doi: 10.1007/s12559-023-10179-8

Hazumi, M., et al. 2020, Proc. SPIE Int. Soc. Opt. Eng., 11443, 114432F, doi: 10.1117/12.2563050

Heymans, C., et al. 2021, Astron. Astrophys., 646, A140, doi: 10.1051/0004-6361/202039063

Hogg, D. W., & Villar, S. 2024, arXiv e-prints, arXiv:2405.18095, doi: 10.48550/arXiv.2405.18095

Hou, J., et al. 2020, Mon. Not. Roy. Astron. Soc., 500, 1201, doi: 10.1093/mnras/staa3234

Huang, D., Bharti, A., Souza, A., Acerbi, L., & Kaski, S. 2023, arXiv e-prints, arXiv:2305.15871, doi: 10.48550/arXiv.2305.15871

Jeffrey, N., et al. 2024, Mon. Not. Roy. Astron. Soc., 536, 1303, doi: 10.1093/mnras/stae2629

Kac, M. 1943, Bulletin of the American Mathematical Society, 49, 314

Kacprzak, T., Fluri, J., Schneider, A., Refregier, A., & Stadel, J. 2023, , 2023, 050, doi: 10.1088/1475-7516/2023/02/050

Kaiser, N. 1992, APJ, 388, 272, doi: 10.1086/171151

Kaiser, N., & Squires, G. 1993, APJ, 404, 441, doi: 10.1086/172297

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., et al. 2021, Nature Reviews Physics, 3, 422, doi: 10.1038/s42254-021-00314-5

Kilbinger, M. 2015, Reports on Progress in Physics, 78, 086901, doi: 10.1088/0034-4885/78/8/086901

Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980, doi: 10.48550/arXiv.1412.6980

Kobyzev, I., Prince, S. J. D., & Brubaker, M. A. 2019, arXiv e-prints, arXiv:1908.09257, doi: 10.48550/arXiv.1908.09257

Kumar, A. 2023, Different Types of CNN Architectures Explained: Examples. https://vitalflux.com/different-types-of-cnn-architectures-explained-examples/

Lanusse, F., Ma, Q., Li, N., et al. 2018, , 473, 3895, doi: 10.1093/mnras/stx1665

Lehman, K., Krippendorf, S., Weller, J., & Dolag, K. 2024, arXiv e-prints, arXiv:2411.08957, doi: 10.48550/arXiv.2411.08957

Lehman, K., Schuster, N., Lucie-Smith, L., et al. 2025. https://arxiv.org/abs/2502.05262

Lemos, P., Coogan, A., Hezaveh, Y., & Perreault-Levasseur, L. 2023a, 40th International Conference on Machine Learning, 202, 19256, doi: 10.48550/arXiv.2302.03026

Lemos, P., Parker, L. H., Hahn, C., et al. 2023b, in Machine Learning for Astrophysics, 18, doi: 10.48550/arXiv.2310.15256

Lesgourgues, J., Mangano, G., Miele, G., & Pastor, S. 2013, Neutrino Cosmology

Limber, D. N. 1954, APJ, 119, 655, doi: 10.1086/145870

Longuet-Higgins, M. S. 1957, Philosophical Transactions of the Royal Society of London Series A, 249, 321, doi: 10.1098/rsta.1957.0002

Louppe, G. 2014, arXiv e-prints, arXiv:1407.7502, doi: 10.48550/arXiv.1407.7502

Lu, T., Haiman, Z., & Li, X. 2023, , 521, 2050, doi: 10.1093/mnras/stad686

Maeder, A., & Gueorguiev, V. G. 2020, Universe, 6, 46, doi: 10.3390/universe6030046

Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2014, MNRAS, 440, 2077, doi: 10.1093/mnras/stu368

Matilla, J. M. Z., Sharma, M., Hsu, D., & Haiman, Z. 2020, , 102, 123506, doi: 10.1103/PhysRevD.102.123506

Matsubara, T. 2020, , 101, 043532, doi: 10.1103/PhysRevD.101.043532

Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, MNRAS, 502, 1401, doi: 10.1093/mnras/stab082

Mellema, G., Koopmans, L. V., Abdalla, F. A., et al. 2013, Experimental Astronomy, 36, 235

Mellier, Y., Abdurroúf, A., Barroso, J. A., et al. 2024, Astronomy & Astrophysics

Miles, C., Bohrdt, A., Wu, R., et al. 2021, Nature Communications, 12, doi: 10.1038/s41467-021-23952-w

Miller, B. K., Weniger, C., & Forré, P. 2022, arXiv e-prints, arXiv:2210.06170, doi: 10.48550/arXiv.2210.06170

Modi, C., & Philcox, O. H. E. 2023, arXiv e-prints, arXiv:2309.10270, doi: 10.48550/arXiv.2309.10270

Mukhanov, V. 2005, Physical Foundations of Cosmology, doi: 10.2277/0521563984

Narayan, R., & Bartelmann, M. 1996, arXiv e-prints, astro. https://arxiv.org/abs/astro-ph/9606001

Nguyen, N.-M., Schmidt, F., Tucci, B., Reinecke, M., & Kostić, A. 2024, , 133, 221006, doi: 10.1103/PhysRevLett.133.221006

Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, , 884, 29, doi: 10.3847/1538-4357/ab3719

Papamakarios, G. 2019, arXiv e-prints, arXiv:1910.13233, doi: 10.48550/arXiv.1910.13233

Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, arXiv:1705.07057, doi: 10.48550/arXiv.1705.07057

Papamakarios, G., Sterratt, D. C., & Murray, I. 2018, arXiv e-prints, arXiv:1805.07226, doi: 10.48550/arXiv.1805.07226

Peacock, J. A. 1999, Cosmological Physics

Pezzotta, A., Moretti, C., Zennaro, M., et al. 2024, Astronomy & Astrophysics, 687, A216

Porqueres, N., Heavens, A., Mortlock, D., Lavaux, G., & Makinen, T. L. 2023, arXiv e-prints, arXiv:2304.04785, doi: 10.48550/arXiv.2304.04785

Preston, C., Amon, A., & Efstathiou, G. 2023, , 525, 5554, doi: 10.1093/mnras/stad2573

Qu, H., Sako, M., Möller, A., & Doux, C. 2021, , 162, 67, doi: 10.3847/1538-3881/ac0824

Raissi, M., Perdikaris, P., & Karniadakis, G. E. 2019, Journal of Computational Physics, 378, 686, doi: 10.1016/j.jcp.2018.10.045

Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (The MIT Press), doi: 10.7551/mitpress/3206.001.0001

Reeves, A., Nicola, A., Refregier, A., Kacprzak, T., & Machado Poletti Valle, L. F. 2024, , 2024, 042, doi: 10.1088/1475-7516/2024/01/042

Rice, S. O. 1945, Bell System Technical Journal, 24, 46, doi: 10.1002/j.1538-7305.1945.tb00453.x

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, , 116, 1009, doi: 10.1086/300499

Robbins, H., & Monro, S. 1951, The Annals of Mathematical Statistics, 22, 400. http://www.jstor.org/stable/2236626

Ruder, S. 2016, arXiv e-prints, arXiv:1609.04747, doi: 10.48550/arXiv.1609.04747

Salcido, J., & McCarthy, I. G. 2024. https://arxiv.org/abs/2409.05716

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. P. 2018, , 611, A2, doi: 10.1051/0004-6361/201731201

Schneider, A., & Teyssier, R. 2015, , 2015, 049, doi: 10.1088/1475-7516/2015/12/049

Schneider, P., Kochanek, C. S., & Wambsganss, J. 2006, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek, & J. Wambsganss, 269–451

Schneider, P., & Lombardi, M. 2003, , 397, 809, doi: 10.1051/0004-6361:20021541

Schöneberg, N., Abellán, G. F., Sánchez, A. P., et al. 2022, Physics Reports, 984, 1

Scoccimarro, R., Colombi, S., Fry, J. N., et al. 1998, , 496, 586, doi: 10.1086/305399

Scolnic, D. M., et al. 2018, Astrophys. J., 859, 101, doi: 10.3847/1538-4357/aab9bb

Sgier, R., Fluri, J., Herbel, J., et al. 2021, , 2021, 047, doi: 10.1088/1475-7516/2021/02/047

Sgier, R. J., Réfrégier, A., Amara, A., & Nicola, A. 2019, , 2019, 044, doi: 10.1088/1475-7516/2019/01/044

Simonyan, K., Vedaldi, A., & Zisserman, A. 2013, arXiv e-prints, arXiv:1312.6034, doi: 10.48550/arXiv.1312.6034

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311, doi: 10.1046/j.1365-8711.2003.06503.x

Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, Monthly Notices of the Royal Astronomical Society, 511, 1771–1788, doi: 10.1093/mnras/stac064

Stiskalek, R., Desmond, H., Holvey, T., & Jones, M. G. 2021, , 506, 3205, doi: 10.1093/mnras/stab1845

Sugiyama, S., et al. 2023, Phys. Rev. D, 108, 123521, doi: 10.1103/PhysRevD.108.123521

Takada, M., & Jain, B. 2004, , 348, 897, doi: 10.1111/j.1365-2966.2004.07410.x

Takahashi, R., Hamana, T., Shirasaki, M., et al. 2017, The Astrophysical Journal, 850, 24, doi: 10.3847/1538-4357/aa943d

Takahashi, R., Nishimichi, T., Namikawa, T., et al. 2020, , 895, 113, doi: 10.3847/1538-4357/ab908d

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, APJ, 761, 152, doi: 10.1088/0004-637X/761/2/152

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. 2018, arXiv e-prints, arXiv:1804.06788, doi: 10.48550/arXiv.1804.06788

Tejero-Cantero, A., Boelts, J., Deistler, M., et al. 2020, Journal of Open Source Software, 5, 2505, doi: 10.21105/joss.02505

Terasawa, R., et al. 2025, Phys. Rev. D, 111, 063509, doi: 10.1103/PhysRevD.111.063509

Troxel, M. A., et al. 2018, Phys. Rev. D, 98, 043528, doi: 10.1103/PhysRevD.98.043528

Tucci, B., & Schmidt, F. 2024, , 2024, 063, doi: 10.1088/1475-7516/2024/05/063

Ward, D., Cannon, P., Beaumont, M., Fasiolo, M., & Schmon, S. M. 2022, arXiv e-prints, arXiv:2210.06564, doi: 10.48550/arXiv.2210.06564

Weiler, M., & Cesa, G. 2019, in Conference on Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/1911.08251

Wetzel, S. J., Ha, S., Iten, R., Klopotek, M., & Liu, Z. 2025, arXiv e-prints, arXiv:2503.23616, doi: 10.48550/arXiv.2503.23616

Wiersma, R. P. C., Ciardi, B., Thomas, R. M., et al. 2013, , 432, 2615, doi: 10.1093/mnras/stt624

Xavier, H. S., Abdalla, F. B., & Joachimi, B. 2016, , 459, 3693, doi: 10.1093/mnras/stw874

Zhu, X.-P., Dai, J.-M., Bian, C.-J., et al. 2019, , 364, 55, doi: 10.1007/s10509-019-3540-1

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Acknowledgments

The past three and a half years have been a journey full of excitement and challenges. I can still clearly remember the day Stella asked me if I wanted to stay in the group to pursue a PhD. Looking back, I feel incredibly fortunate to have given an unwavering yes without hesitation. My experience is too limited to make a definitive statement, but based on what I have encountered, Stella Seitz has been an absolutely outstanding supervisor. Her passion for science, her constant pursuit of knowledge, and her great patience with students, encouraging them to explore their own paths while providing thoughtful and steady guidance, have all made me deeply grateful to her. She is undoubtedly one of the most important people in my journey to becoming an independent researcher. It is a regret that, due to unavoidable circumstances, we were unable to spend more time together on science. But I guess this imperfection exactly highlights the value and respectability of Stella as a supervisor. Another person I must thank is Ralf Bender. Without his affirmation, I would not have been able to embark on this period of life journey. It is also because of his and Stella's generosity and support that I have had the opportunity to communicate and collaborate with physicists from around the world, visit many renowned academic institutions, and attend numerous conferences that have greatly benefited me. There is an ancient Chinese saying, "A teacher for a day, a father for life," and Stella and Ralf certainly deserve this quote in my academic career.

How could I not mention Anik Halder? I can never forget how he helped me during my master's program and my first PhD project in life. We spent so much time together discussing, programming, testing, writing, traveling, laughing, feeling frustrated, teasing each other, and even arguing. He was almost always there for me through the ups and downs of my early experiences in research. He has been both a mentor and a friend to me. I believe he can always do well whether in academy or not. He is a man whom I respect.

Alexandre Barthelemy is someone who may seem difficult to get along with at first, and in reality, he does have a strong personality. But I would say that I really enjoy interacting with him in both work and life. I can feel his sincerity and dedication. He is a mentor who opened up a new path for me in cosmology research. It was a truly pleasant experience which we spent together on the project and he has taught me a lot. He has also introduced me to the science, work, social life, and of course, the cuisine in Paris. Most of all, I still want to hear his jokes, whether inappropriate or not, from time to time.

The collaboration with Annabelle Bohrdt means a lot to me. This was the first time I proactively, yet nervously, reached out to other physicists for a collaboration. What

I experienced was kindness and warmth. It allowed me to feel the joy and confidence that come from connecting with researchers who, although initially strangers and from completely different cultural and linguistic backgrounds, come together through science.
I would also like to express my gratitude to Sandrine Codis, who hosted me with friendliness and warmth at CEA-Saclay. Together with Alexandre, she introduced me to the world of cosmological study with perturbation and bias theories, though so far I am only able to understand a small part of it. Both Annabelle and Sandrine made me feel the support and help that senior researchers provide to juniors in this field. This is the type of behaviours that I aspire to follow in the future.
I am grateful to Oliver Friedrich, Daniel Gruen, Ariel Sanchez and Jochen Weller for their lectures and discussions with me on cosmology which form the basic framework of my understanding about this research field. Alex Barreira helped me a lot during my first project with his insights, guidance and patience. The discussion with him on topics of covariance matrix and response approach to perturbation theory has left me with a deep impression.
During my PhD study, I was surrounded by supportive and helpful colleagues. With some of them I had collaborations and others daily interactions. I would like to thank David, Kai (special thanks to both who helped proof-read ML chapters in my thesis), Giacomo, Leon, Luis, Raphael, Jan-Niklas, Michael, Parth, Jed, Yun-Hsin, Naman, Sarah, together we had many great moments during lunch time, Friday afternoon tea time, pre-colloquiums and after-work beers. I also want to thank Andrea, Matteo, Carlos, Sofia, Jiamin from OPINAS MPE and Arno, Steffen, Chris, Henrique, Sanka and Zekang from USM. Sijin, I am happy for you that you have found your interests in cosmological research. You have done a good job in your master thesis project in our group and I am very pleased to be part of it. Just do not give up and keep going, the road will open for you ahead. Xu, Qiang and you are my Chinese best friends in Munich, I really appreciate your company and sincerely wish all of you can achieve your goals in the future.
Mom, Dad, it is quite amazing that I have been able to come this far. I know you always have confidence in me, it is such a dear love and source from which I can draw strength. You may not understand my work and focus, but we always share our feelings and emotions. You raised me up with decent education, independence and dignity, and wrap them up with your love. You are truly the best parents in the world and I am so proud to be your son. Waigong and Nainai, so much time you have spent on me while I was a kid, I learned how to be a good person from you. You have passed your expectations and best wishes in this world to me, and I will always try my best not to fail them. It is such a pity that Waipo and Yeye cannot see this moment. I miss them so much. Uncles and aunts, during my childhood you cared about me as if I were your own child. I am forever grateful for that.
After all I think I am a lucky person with all the love and support received from surroundings. It is a fortune and yet a responsibility. Let me end this thesis with a quote from *Zhou Yi - Qian Hexagram* as a form as self-expectation: 天行健，君子以自强不息; 潜龙勿用，阳在下也; 见龙在田，德施普也; 终日乾乾，反复道也; 或跃在渊，进无咎也; 飞龙在天，大人造也; 亢龙有悔，盈不可久也。