Aix*Marseille université
Socialement engagée

NNT : 2023AIXM0001

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 6 février 2024 par

Renan Isquierdo Boschetti

Advancements towards precision cosmology with Void-Lensing: Observational aspects and modelling

Discipline
Physique et Sciences de la Matière

Spécialité
Astrophysique et Cosmologie

École doctorale
ED 352 Physique et Sciences de la Matière

Laboratoire/Partenaires de recherche
CPPM (Centre de Physique de Particules de Marseille)
LAM (Laboratoire d'Astrophysique de Marseille)
IPhU (Institut de Physique de l'Univers)

Composition du jury

Martin KILBINGER CEA IRFU/DAp-AIM, IAP	Rapporteur
Nico HAMAUS Ludwig-Maximilians-Universität	Rapporteur
Alice PISANI Aix Marseille Université, CPPM	Examinatrice
Sandrine CODIS Université Paris-Saclay, AIM, LCEG	Examinatrice
Ravi K. SHETH University of Pennsylvania	Examinateur
Cristinel DIACONU Aix Marseille Université, CPPM	Président du jury
Stephanie ESCOFFIER Aix Marseille Université, CPPM	Directrice de thèse
Eric JULLO Aix Marseille Université, LAM	Co-Directeur de thèse

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM

Institut Physique de l'Univers Aix*Marseille Université

LAM LABORATOIRE D'ASTROPHYSIQUE DE MARSEILLE

Affidavit

Je soussigné, Renan Isquierdo Boschetti, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Stephanie Escoffier, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 23/11/23



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse:

1. "Towards cosmology with Void Lensing: how to find voids sensitive to weak-lensing and numerically interpret them", Renan Boschetti *et al.* 2023. e-Print: 2311.14586. Submitted to JCAP.

Participation aux conférences et écoles au cours de la période de thèse:

1. "L'école de GIF 2021", Marseille, France, September 2021.
2. "Tonale Winter School on Cosmology 2021", Passo Del Tonale, Italy, December 2021.
3. "Euclid consortium meeting 2022", Oslo, Norway, April 2022.
4. "Atelier Action Dark Energy", Marseille, France, May 2022. Présentation: "Void-Lensing as a test of gravity", orateur.
   https://indico.in2p3.fr/event/26251/contributions/109675/
5. "ICTP Summer School on Cosmology 2022", Trieste, Italy, July 2022.
6. "DESI Collaboration Meeting Winter 2022", Cancun, Mexico, December 2022.
7. "Future Cosmology 2023 (IESC)", Cargèse, Corse, April 2023.

Résumé

Dans cette thèse, nous présentons une étude du signal de lentille gravitationnelle des vides ou de l'excès de densité de masse en surface (ESMD) autour des vides cosmiques. Tout d'abord, nous proposons un nouvel algorithme de recherche de vides conçu pour capturer l'ESMD autour des vides. Nous comparons notre algorithme appliqué à des tranches projetées avec celui de ZOBOV et trouvons des profils de faible lentille significativement plus profonds pour les vides définis par notre algorithme dans le cadre d'une simulation réaliste de galaxies. Ensuite, nous testons la cohérence entre les mesures de l'ESMD telles que mesurées par la déformation des galaxies en arrière-plan et calculées directement à travers les profils de densité de matière noire des mêmes vides. Nous avons trouvé des incohérences pour les vides avec un diamètre $\leq 100h^{-1}$Mpc le long de la ligne de visée, mais la cohérence est maintenue pour les vides plus petits, ce qui signifie que nous sondons effectivement le champ de matière noire sous-jacent en mesurant la déformation autour de ces vides. De plus, nous montrons que les vides trouvés dans les tranches projetées, qui sont très sensibles au phénomène de lentille gravitationnelle, sont corrélés aux vides en 3D présentant des alignements intrinsèques entre eux.

Nous présentons également un travail moins avancé sur la modélisation de l'ESMD. Nous nous appuyons sur l'idée que les vides projetés, qui sont plus sensibles au signal ESMD, sont des combinaisons des vides définis dans le champ de densité en 3D. Nous proposons un *Ansatz* pour relier les deux grandeurs. Cette connexion dépend de l'abondance des vides en 3D, ainsi que de la corrélation croisée entre les positions des vides en 3D et 2D. Nous avons effectué des mesures de ces grandeurs sur des simulations N-corps pour tester le modèle. Le modèle est capable de reproduire le profil de vide projeté dans le régime des deux vides, tandis qu'il présente des désaccords dans le terme d'un seul vide.

Mots clés : Cisaillement gravitationnel, Vides cosmiques, Structure à grandes échelles.

Abstract

In this thesis, we present a study of the void lensing signal or the excess surface mass density (ESMD) around cosmic voids. First, we propose a new void-finder algorithm that is designed to capture the ESMD around voids. We compare our algorithm applied to projected slices with the ZOBOV void finder and find significantly deeper weak-lensing profiles for voids defined by our algorithm in the context of a realistic galaxy mock. Then we test the consistency between the measurements of the ESMD as measured through the shear of background galaxies and directly calculated through the dark matter density profiles of the same voids. We found inconsistencies for voids with diameter $\leq 100h^{-1}$Mpc along the line- of-sight, but the consistency holds for smaller voids, meaning that we are indeed probing the underlying dark matter field by measuring the shear around these voids. Moreover, we show that voids found in the projected slices, which are highly sensitive to lensing, are correlated to 3D voids exhibiting intrinsic alignments between them.

We also present a less mature work on the ESMD modelling. We rely on the idea that the projected voids, which are more sensitive to the ESMD signal, are combinations of the voids defined in the 3D density field. We propose an *Ansatz* to connect both quantities. This connection depends on the 3D void abundance, as well as on the cross-correlation between the 3D and 2D void positions. We performed measurements of these quantities on N-body simulations to test the model. The model is capable of reproducing the projected void profile in the two-void regime, whereas present discrepancies in the one-void term.

Keywords: Weak-Lensing, Cosmic voids, Large-scale structure.

Résumé étendu

Introduction

Ce travail s'inscrit dans un domaine de connaissance appelé « cosmologie ». La cosmologie est l'étude de l'Univers dans son ensemble. Autrement dit, elle vise à comprendre l'origine de l'univers, son destin et tout ce qui se passe entre les deux.

On dit généralement que la cosmologie moderne est née avec l'invention de la Relativité Générale en 1915. Elle permet, pour la première fois, de s'interroger sur l'évolution de l'univers dans son ensemble, sa géométrie et son contenu énergétique. Depuis, nous avons fait un progrès étonnant, au point où nous sommes capables d'expliquer quantitativement toutes les observations cosmologiques que nous avons faites jusqu'à présent.

Malgré ce succès, nous pouvons affirmer avec certitude que nous ne savons pas tout ce qu'il faut savoir pour connaître l'Univers. Malgré l'ajustement de toutes les données disponibles, nous disposons encore de beaucoup d'espace pour tester la nouvelle physique. C'est l'objectif des projets à grandes échelles actuels et à venir.

Des études de structure, qui cartographieront essentiellement tout l'Univers observable disponible. Ces données nous permettront de tester le modèle standard de cosmologie dans les régimes que nous n'avons pas testés jusqu'à présent.

Il y a des indications claires qu'il y a quelque chose de fondamental dans la physique qu'on ne comprend pas. Par exemple, les deux principaux ingrédients du modèle standard de la cosmologie sont partiellement ou totalement inconnus, à savoir la matière noire et l'énergie noire. La matière noire est responsable d'environ 25% de l'univers et n'est pas incluse dans le modèle standard de la physique des particules. Nous avons une bonne idée du comportement collectif de ces particules aux échelles cosmologiques, à savoir qu'elles sont froides, se déplacent à des vitesses non relativistes et qu'elles sont sans collision, c'est-à-dire qu'elles ne se couplent pas avec le champ électromagnétique. Nous savons aussi comment elles se sont comportées tout au long de l'histoire cosmique et ont formé les modèles que nous

mesurons grâce à des traceurs lumineux. Cependant, nous ne savons pas quelle est la nature fondamentale de la matière noire, qu'elle soit composée de particules que nous devons inclure dans le modèle standard, ou si elle est composée de trous noirs primordiaux, de WIMP ou d'axions. Ce sont quelques-unes des possibilités qui s'offrent à nous au moment où cette thèse est rédigée. Le pire des cas est qu'elle s'agit d'une nouvelle particule qui n'interagit avec le champ électromagnétique, et nous ne pourrons jamais le détecter directement, mais seulement spéculer à ce sujet.

À ce stade, notre stratégie consiste à nous tourner vers la nature et à rechercher à partir d'une « pointe » de la façon de procéder. C'est, selon la vision du présent auteur, la raison pour laquelle nous nous engageons dans cet énorme effort de cartographie des structures à grande échelle. Ces données pourraient nous montrer avec une grande précision là où exactement le modèle standard de la cosmologie est incomplet. Par conséquent, nous devons extraire toutes les informations possibles des données. Cette tâche n'est pas anodine et constitue l'un des principaux thèmes sous-jacents à de nombreux travaux en cosmologie observationnelle des dernières décennies. Ce travail fait partie de ce programme.

Le fait que les vides sont des traceurs de structures à grande échelle avec des biais linéaires négatifs les fait transporter des informations complémentaires aux traceurs positivement biaisés.

Les vides sont également intéressants en eux-mêmes. Puisqu'ils sont sous-denses en matière, ils peuvent être considérés comme les meilleurs « laboratoires » pour détecter les signatures de processus physiques qui sont moins efficaces dans les environnements à haute densité. Cela montre que l'abondance du vide est une sonde sensible des modèles d'énergie sombre, des neutrinos massifs et des modifications de la gravité.

Nous pouvons « voir » les vides dans le domaine de la matière totale en mesurant leur effet sur les formes des galaxies d'arrière-plan. En d'autres termes, pour mesurer la lentille faible profil de cisaillement autour des vides, qui est essentiellement la projection de la densité des vides profil le long de la ligne de visée. Cette ligne d'investigation est connue sous le nom de Void-Lensing.

L'état de l'art de la science de la lentille du Vide est : nous savons qu'elle est hautement sensible aux modifications de la gravité notamment, et nous en avons détecté quelques fois avec un rapport signal/bruit relativement important, mais jamais réalisé d'analyse avec cet observable. Ce travail vise à apporter quelques

contributions au développement de ce domaine.

Le paradigme ΛCDM

Les données suggèrent fortement que l'Univers traverse une phase d'expansion. Cela implique que l'Univers soit dominé par un constant cosmologique à bas redshift. Nous disposons également de preuves solides, grâce à la mesure des anisotropies dans la distribution de température du Fond diffus cosmologique (CMB), que l'Univers au redshift $z \simeq 1090$ était presque parfaitement homogène et isotrope, avec de petites perturbations de l'ordre de $\simeq 10^{-5}$. La forme précise du spectre de puissance CMB est en accord étonnant avec un modèle plat à six paramètres (ΛCDM), où les paramètres sont les fractions densités de baryons, matière noire, profondeur optique due à la réionisation, Hubble paramètre, l'indice spectral scalaire et l'amplitude du spectre de puissance scalaire. La figure 1 montre l'accord entre le modèle ΛCDM à six paramètres libres contre les données du télescope Planck. On peut soutenir que cette mesure a déclenché ce qu'on appelle « l'ère de la cosmologie de précision », dans laquelle le modèle standard est capable d'ajuster les données avec des erreurs de sous-pourcentage.

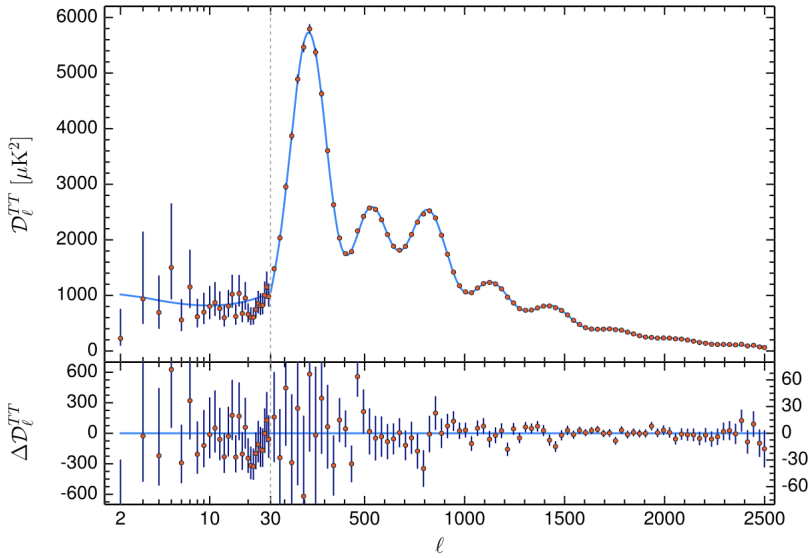

FIGURE 1 : Anisotropies dans le CMB mesurées par le télescope Planck. La ligne bleue représente l'ajustement du modèle ΛCDM à six paramètres libres. L'axe x correspond aux moments multipolaires et l'axe $y-$ est la variance à chaque l. Les l faibles correspondent aux grandes échelles, tandis que les l élevés correspondent aux petites échelles. Extrait de Nabila AGHANIM et al. 2020

La phénoménologie des vides cosmiques

La science des vides remonte aux années 70, où certains auteurs détectaient de vastes régions de rayon $\simeq 20h^{-1}$Mpc presque vides de galaxies dans la distribution des galaxies proches. Mais après la découverte d'un vide de rayon $\simeq 60h^{-1}$Mpc dans la constellation du Bouvier, les vides ont commencé à recevoir davantage d'attention dans la littérature, remettant en question s'ils sont une caractéristique commune ou une exception dans les grandes structures. Depuis, les enquêtes sur les galaxies ont montré que les vides sont si fréquents qu'ils occupent une grande fraction du volume de l'Univers, constituant ainsi une caractéristique essentielle des structures à grande échelle. L'existence de vides découle directement de la configuration des conditions initiales et de l'effondrement gravitationnel, principal moteur de la formation des structures. Alors que la matière s'accumule dans des régions sur-denses, déterminées par les conditions initiales, les régions sous-denses s'étendent et occupent des volumes de plus en plus grands.

L'observable le plus important des vides est la fonction de rayon des vides, ou le

nombre de vides par intervalle de rayon, pour un volume donné.

La théorie des ensembles d'excursions est basée sur l'effondrement sphérique (expansion pour les vides), dans lequel une surdensité (sous-densité) isolée évolue dans un arrière-plan Einstein de-Sitter et finit par s'effondrer et se virialiser pour former des halos, ou subit des croisements de coques entre les coques internes qui se dilatent plus rapidement que les coques externes pour les vides. Ensuite, la valeur extrapolée linéairement (à partir des conditions initiales) du contraste de densité pour lequel la virialisation (traversée de coquille) se produit est utilisée comme seuil pour définir un halo (vide). Ces valeurs sont $\delta_c = 1,686$ et $\delta_v = -2,7$, pour les halos et les vides respectivement.

Dans l'ensemble d'excursions, le champ de densité lagrangien est lissé à une certaine échelle R comme

$$S(R) \equiv \sigma^2(R) = \left\langle \delta^2(x,R) \right\rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_L(k) W_R^2(k), \tag{0.1}$$

où $P_L(k)$ est le spectre de puissance linéaire et $W_R(k)$ est une fonction de lissage. Le champ lissé $\delta(S)$ est ensuite utilisé pour effectuer une marche aléatoire, à partir de $S = 0$ ($R \to \infty$). L'ensemble des excursions prédit la fraction des promenades qui traverseront le seuil δ_c pour la première fois dans l'intervalle de masse $[M, M + dM]$, $f(M)$. Cette fraction est ensuite convertie en densité numérique d'objets par intervalle de masse comme

$$S(R) \equiv \sigma^2(R) = \left\langle \delta^2(x,R) \right\rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_L(k) W_R^2(k), \tag{0.2}$$

où $P_L(k)$ est le spectre de puissance linéaire et $W_R(k)$ est une fonction de lissage. Le champ lissé $\delta(S)$ est ensuite utilisé pour effectuer une marche aléatoire, à partir de $S = 0$ ($R \to \infty$). L'ensemble des excursions prédit la fraction des promenades qui traverseront le seuil δ_c pour la première fois dans le bin de masse $[M, M + dM]$, $f(M)$. Cette fraction est ensuite convertie en densité numérique d'objets par bin de masse comme

$$\frac{dn}{d\ln M} \equiv \frac{d^2N}{dV\, d\ln M} = \bar{\rho}_m f(M). \tag{0.3}$$

En cas de $\delta(S)$ performer une marche Markovian, la fonction $f(M)$ est donne par

$$\frac{dn}{d\ln M} = \frac{\bar{\rho}_m}{M} f_h(\nu) \frac{d\ln \sigma^{-1}}{d\ln M}, \tag{0.4}$$

où $\nu = \delta_c/\sigma$ est la fonction de multiplicité pour les halos est défini par

$$f_h(\nu) = \sqrt{\frac{2}{\pi}} \nu e^{-\nu^2/2}. \tag{0.5}$$

La version pour les vides est obtenue en utilisant deux barrières de densité, une pour la formation de halos et une additionnelle pour la formation des vides, δ_c et δ_v. La fonction de multiplicité pour les vides qu'on utilise dans ce travail est donnée par

$$\begin{aligned} f_v^{\mathrm{2LDB}}(\sigma) = 2\left(1+D_v\right) \exp\left[-\frac{\beta_v^2\sigma^2}{2\left(1+D_v\right)} - \frac{\beta_v \delta_c}{1+D_v}\right] \\ \times \sum_n \frac{n\pi}{\delta_T^2}\sigma^2 \sin\left(\frac{n\pi\delta_c}{\delta_T}\right) \exp\left[-\frac{n^2\pi^2\left(1+D_v\right)}{2\delta_T^2}\sigma^2\right]. \end{aligned} \tag{0.6}$$

Optimum Centering Void Finder (OCVF)

Dans ce travail, nous proposons un nouvel algorithme de recherche de vides. Le fonctionnement de cet algorithme peut être résumé par les étapes suivantes :

- Effectuez la triangulation de Delaunay pour obtenir l'ensemble des points qui sont candidats comme positions des vides.
- Développez des cercles (ou des sphères) autour d'eux jusqu'à ce que la densité moyenne de ces cercles atteigne un certain seuil de densité, spécifié par $\bar{\rho}_v(< r_v)$.
- Le plus grand cercle sera le premier vide du catalogue et tous les autres vides qui l'interceptent seront écartés. Le même processus sera répété pour le deuxième plus grand vide restant. Ce processus sera répété jusqu'à ce qu'un vide dont le rayon est inférieur au rayon de coupe R_c soit inclus dans le catalogue.

Nous avons utilisé une simulation de matière noire pour valider l'algorithme. La figure 2 montre que les vides trouvés par notre algorithme sont bien capables de reproduire la prédiction théorique.

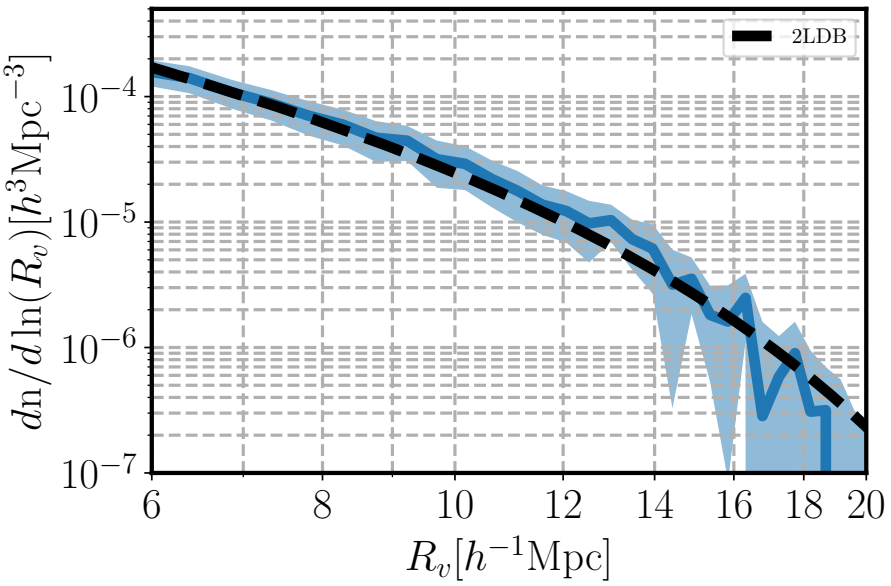

FIGURE 2 : L'abondance mesurée dans une simulation à N corps DM uniquement de taille $L = 500h^{-1}$Mpc par rapport à la prédiction du modèle 2LDB.

Comparaison des algorithmes

Nous avons comparé le signal de lentilles par les vides mesurés par notre algorithme avec un algorithme déjà existant dans la littérature. La Figure 3 montre la comparaison entre le signal de lensing ($\Delta\Sigma$) autour des vides `OCVF` et `ZOBOV`. L'algorithme `OCVF` présente des profils plus profonds par rapport aux vides `ZOBOV`. Ce fait se réalise à cause du fonctionnement de notre algorithme et aussi dû aux vides `OCVF` qui sont 2D au lieu de 3D comme les vides `ZOBOV`.

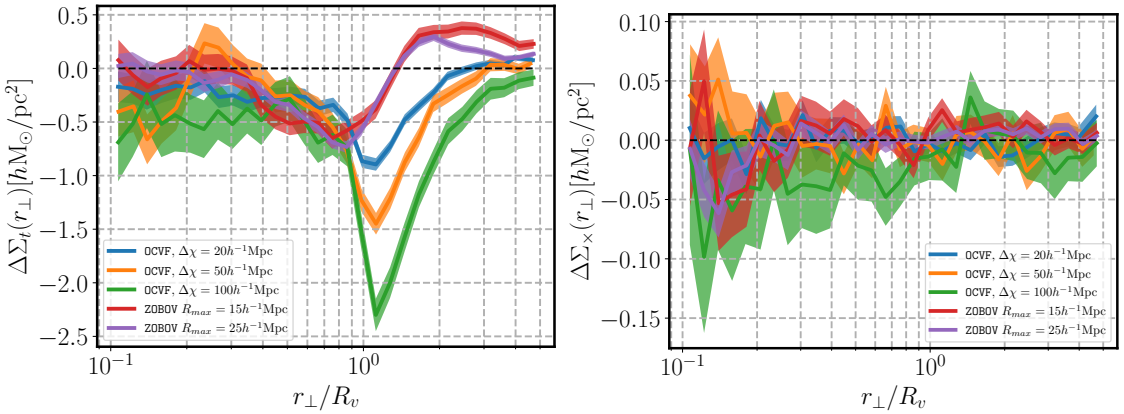

FIGURE 3 : Left : Comparison between the $\Delta\Sigma_t$ measurements performed using the `OCVF` in slices of width $\Delta\chi = 20,\ 50,\ 100\,h^{-1}\,\mathrm{Mpc}$ (blue, orange, and green) and `ZOBOV` in the bins of radius $[10, 15]h^{-1}$Mpc (red) and $[10, 25]h^{-1}$Mpc (purple). Right : The same comparison for the cross component $\Delta\Sigma_\times$.

Ensuite, nous avons vérifié la cohérence entre le profil $\Delta\Sigma$ mesuré en utilisant le shear tangentiel, γ_t, par rapport au même profil mesuré directement en utilisant les particules de matière noire autour des mêmes vides. La figure 5 montre la cohérence entre ces deux pour trois choix d'épaisseur du bin, en utilisant les vides `OCVF`. Notamment, les profils pour le choix de bin $\Delta\chi = 100h^{-1}$Mpc ne sont pas complètement consistants. Cette incohérence se déroule à cause de l'approximation de lentille faible, laquelle n'est pas plus valide pour les vides plus grands. Nous avons trouvé un résultat similaire pour les vides `ZOBOV`.

En plus, nous avons montée que les vides 2D sont bien corrélés avec les vides 3D, lesquels présentent un profil de densité anisotropique.

La modélisation du signal de lentille par les vides

Une fois que les vides 2D possèdent des profils plus profonds, il est désirable d'avoir un modèle pour ces profils de densité. Nous proposons le modèle suivant

$$\delta_{2D}(r_\perp|R_{2D},\Delta_{2D},\Delta_{3D}) = \frac{1}{N}\int dR_{3D}\,\frac{dn_v}{dR_{3D}}(R_{3D}|\Delta_{3D}) \times \int dx_\perp d\phi\, P(x_\perp|R_{3D},R_{2D},\Delta_{3D},\Delta_{2D}) \times \int dr_\parallel \delta_{3D}(|\boldsymbol{r}_\perp - \boldsymbol{x}_\perp|,|r_\parallel - x_\parallel||R_{3D},\Delta_{3D}) \tag{0.7}$$

où $\boldsymbol{r}_\perp$ et $\boldsymbol{x}_\perp$ sont respectivement la distance au centre du vide en 2D et la position du centre en 3D dans le plan perpendiculaire à la ligne de visée. Comme le profil empilé est anisotrope, nous prenons $\boldsymbol{r}_\perp$ aligné avec l'axe x et le système de coordonnées centré sur le centre du vide en 2D.

Nous avons argumenté que la fonction P doit être donnée par la corrélation entre les vides trois et deux D.

Ensuite, nous avons mesuré chaque ingrédient du modèle 0.7, à savoir, la corrélation 2D - 3D, la fonction de rayon, $\frac{dn_v}{dR_{3D}}$ et les profils 3D anisotropiques, δ_{3D}.

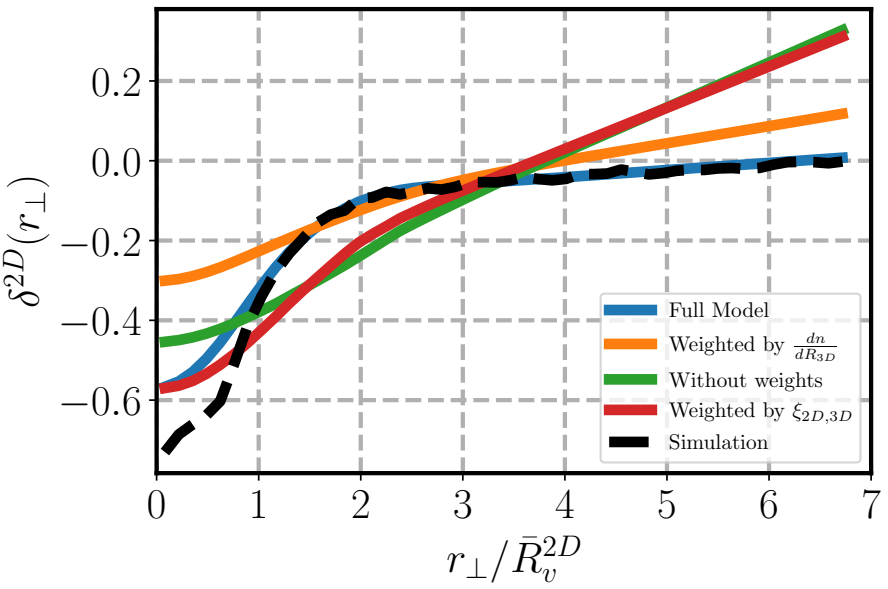

FIGURE 4 : Le modèle complet (en bleu, équation 0.7), le modèle uniquement pondéré par l'abondance (en orange), le modèle uniquement pondéré par les cross-corrélations 2D-3D (en rouge) et le modèle sans pondération (en vert).

La figure 4 montre la performance du modèle par rapport à la mesure directe du profil dans la simulation. Les différentes lignes montrent le modèle sans chaque

ingrédient pour montrer l'impact de chacun sur la prédiction.

Le modèle complet est bien capable de reproduire le profil 2D dans le régime de 2-vides (au-delà du rayon), mais pas à l'intérieur du rayon du vide. La cause de cette incohérence est sujette d'une enquête en cours.

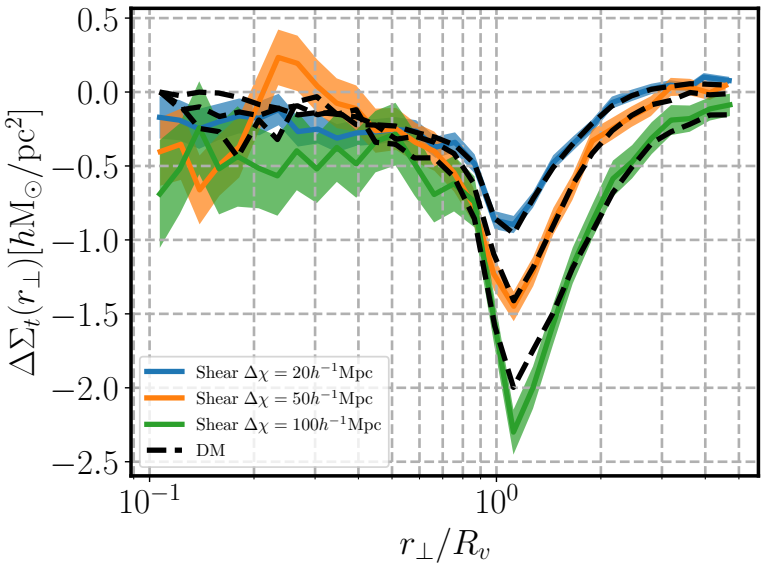

FIGURE 5 : Comparison between $\Delta\Sigma_t$ as measured through the shear (blue, orange, and green) and directly using the DM particles (dashed black).

Mots-clés : géométrie algorithmique, complexe planaire et rectangulaire, géodésique, courbure globale non-positive

Acknowledgments

This work wouldn't have been possible without the support of several key individuals.

Firstly, I want to acknowledge my family for their unwavering support throughout my career. Specifically, I'd like to thank my father, Marco Aurélio Boschetti, and my mother, Eliana Isquierdo Boschetti. Without them, none of this would have been achievable.

I'd also like to express gratitude to my friends, who have been and continue to be pivotal in my growth as a scientist. In particular, I'd like to thank Beatriz Tucci, Caroline Guandalin, and Francisco Maion. I extend special thanks to Rodrigo Voivodic for his contribution to the Void-Lensing modeling presented in the final chapter of this work.

Appreciation goes out to all the individuals at CPPM, with a special mention of Mme Brigitte Pantat for her remarkable work there. Without her, accomplishing this work would have been significantly more challenging.

I'm grateful to my supervisors, Stephanie Escoffier and Eric Jullo, for accepting me into this PhD program.

Lastly and most important, I want to give a special thanks to my wife, Fernanda Leite. Without her, none of this work would have been possible. Apart from her unwavering support over the years, she also contributed to creating illustrative figures in this thesis, as well as in reviewing the text.

Table des matières

Affidavit 2

Liste de publications et participation aux conférences 3

Résumé 4

Abstract 5

Résumé étendu 6

Acknowledgments 16

Table des matières 17

1 Introduction 19

1.1 The broader context of this work . . . 19
1.2 How this work is contextualized in modern cosmology . . . 23
1.3 The ΛCDM paradigm . . . 28
1.3.1 A brief history of background expansion . . . 29
1.3.2 The inhomogeneous Universe in a nutshell . . . 32
1.3.3 Gauge-invariant variables . . . 33
1.3.4 Classification of perturbations . . . 34
1.3.5 Gauge transformations . . . 35
1.3.6 Qualitative analysis of linear perturbations . . . 36
1.3.7 Quantitative analysis of linear perturbations . . . 40
1.4 The cosmological constant problem . . . 49

2 Void Phenomenology 54

2.1 Introduction . . . 54
2.2 The excursion set for Halo formation . . . 56
2.2.1 The spherical collapse . . . 57

2.2.2 The Press-Schechter argument 59
2.3 The Excursion set - The void abundance 69

3 Weak-Lensing **77**
3.1 Overview 77
3.2 Light deflection 79
3.2.1 The Lens equation 82
3.2.2 Ellipticities and shear 92
3.2.3 Theoretical challenges 95

4 Void-Lensing **98**
4.1 The Void-Lensing measurement and numerical interpretation 98
4.1.1 Introduction 98
4.1.2 Optimum Centering Void Finder (OCVF) 103
4.1.3 Void lensing on galaxies 113
4.1.4 Void catalogues 115
4.1.5 The $\Delta\Sigma$ Estimator and the TL approximation 116
4.1.6 Void finder comparison 117
4.1.7 Numerical interpretation 121
4.1.8 Conclusions 125
A How the thin lens approximation affects the VL profiles 126
B The void intrinsic alignment around projected voids 127
4.2 The Void-Lensing model 128
4.2.1 Introduction 128
4.2.2 The relation between 2D and 3D voids 129
4.2.3 The void abundance 132
4.2.4 The cross-correlation between 2D and 3D voids. 133
4.2.5 Anisotropic void profiles 135
4.2.6 Testing the model 136

Conclusion **139**

Bibliographie **142**

1 Introduction

1.1 The broader context of this work

Unlike the other animals in this planet, we have an unusual need : to understand the Universe in all its details. For millennia, we have accumulated knowledge and created a method for extracting some truth from nature[1].

This endeavour has led us surprisingly far. Regarding physics, we have built models which are capable of predicting everything we can measure, in all scales of the Universe[2].

This work falls into one field of knowledge called "cosmology". Cosmology is the study of the Universe as a whole. That is, it aims to understand the origin of the Universe, its fate and everything that happens in between. Of course, it is hard to trace a well defined line between what is cosmology and astrophysics, for instance, since at some point there are topics which are in between the both fields. But we can stick with the definition that cosmology is more concerned about macro aspects of the Universe, rather than specific micro objects, such as the life-cycle of a star.

The next natural question is why to study cosmology and how this knowledge can impact human societies. Some development of some type of cosmology is not unique of our modern society. If we understand cosmology as any attempt of understanding the broader context of existence, we have evidence that indigenous cultures, for instance, also have their stories for the Universe. In essence, other types of cosmology do something similar to what we understand as cosmology in

[1] Here, I take a realism philosophical position, under which the physical theories are telling something about reality, rather than simply being a mathematical model which fits the data, as positivism claims. The true is probable in between realism and positivism. Moreover, it is highly dependent on the context.

[2] Here I am treating the recent tensions in cosmology as possibly arising from systematic errors. The standard model of cosmology and, in particular, of particles physics is well established and capable of describing all data, despite some inconsistencies. Of course this does not mean that our knowledge of fundamental physics is complete, but rather that it is hard to push beyond the standard models.

our scientific context[3] : they identify patterns in nature and came up with a story which is capable of explaining these patterns[4]. Therefore, we can say that any type of cosmology comes from the same human need : the need for rationalizing the world around us. Of course, the collective reason why we engage in this program of identifying patterns in nature has not a unique answer. One can argue in a couple of different lines. For instance, from a purely Darwinian point of view, we can understand it as an evolutionary resource. Our ancestors survived «also» due to their ability of identifying patterns and, according to this argument, nature drives us to this activity, ultimately, the scientific activity. Or, one can argue that there is something that we cannot explain in this impulse. Simply we urge to understand who we are and how it is possible that we are alive in this Universe. This is connected with an amazement that we have when we face something much bigger than us. According to this, the reason why we do science is close to the reason why people engage in religious rituals - it is about the transcendence, the contact with something bigger.

Whatever is the objective reason we search for explanations, the narratives we develop about the world around us have the power to deeply impact our culture beyond the scope of science. Scientific discoveries are capable of inspiring art of all sorts, ideologies, political discourse, etc. It is hard to know to what extent scientific discoveries impact human societies, even in cases in which there is no trivial technological applications.

Modern cosmology is usually said to be born with the invention of General Relativity in 1915. It allowed, for the first time, inquiries about the evolution of the Universe as a whole, its geometry and its energy content. Since then, we have made an astonishing progress, up to the point in which we are able to quantitatively explain every cosmological observation we have done so far[5].

Despite this success, we can firmly say that we don't know everything there is to know about the Universe. Despite fitting all the available data, we still have a lot of

[3] Here "scientific context" is emphasising that there is a key difference between stories of creation, or explanations for why nature does what it does and modern science. It is a topic which is not covered in this work, but we can highlight that the cumulative, the auto-regulated and quantitative character play a key role in the successes of modern science. Therefore, we can say that, at least at some level, our theories capture something about reality.

[4] I am using "explanation" in a broader sense, not necessarily meaning "explanation" in the scientific context.

[5] Here we use the word "explanation" as synonymous of having a model which is capable of fitting the data.

room for testing new physics. That is the goal of current and upcoming large-scale structure surveys, which will map basically all the available observable Universe. This data will allow us to test the standard model of cosmology in regimes we have not tested so far.

Puzzles in physics

There are clear indications that there is something fundamental about physics that we don't understand. For instance, the two main ingredients of the standard model of cosmology are partially or completely unknown, namely, dark matter and dark energy.

Dark matter is responsible for approximately 25% of the energy budget of the Universe and it is not included in the standard model of particle physics. We have a good idea of the collective behaviour of these particle on cosmological scales, namely, that they are cold, or move in non-relativistic speeds and that they are collisionless, i.e. that it does not couple with the electromagnetic field. We also know how it behaved throughout cosmic history and formed the patterns we measure through luminous tracers. However, we do not know what is the fundamental nature of dark matter, whether it is composed of particles we have to include into the standard model, or whether it is composed of primordial black holes, WIMPs, or axions. These are some of the possibilities on the table at the time this thesis is being written. The worst case scenario is that it is a new particle and does not interact with the electromagnetic field at all, then we will never be able to detect it directly, but only speculate about it is nature.

We call by dark energy a couple of different ideas which accounts for the same observable fact : the Universe is expanding and the recession velocity of galaxies is increasing with time. Given that the Universe is described by General Relativity on large-scales, we have to add a *cosmological constant* in one of the sides of Einsteins equations to reproduce this accelerated expansion at the background level. Despite of being freedom of the theory, i.e., the Einstein-Hilbert action is defined up to a constant term, we don't know how to physically justify the existence of this constant. We can interpret it either as a curvature of space-time, by adding it to the left-hand side of Einstein's equations, or as a constant term in the energy-momentum tensor of a homogeneous and isotropic fluid. In the latter case, the cosmological constant is the contribution of the zero-point energies of the fields in the standard model

of particle physics. We can predict a value for the cosmological constant by using the most well tested physical theory we have, namely quantum field theory, and the result will be many orders of magnitude different than the inferred value from cosmology. This is commonly known as "the worst prediction in the history of physics".

If the cosmological constant is really responsible for the late cosmic acceleration, we have to find some mechanism that cancels out the vacuum contribution and make it have the tiny value we observe. In string-theory, the most popular candidate to quantum gravity, the vacuum is not unique, but rather it is an ensemble of more than 10^{500} vacua (or cosmological constants), each one corresponding to a different way of compactifying extra dimensions. Some argue that this "landscape" could be solution for the cosmological problem, since we are living in an Universe drawn from this landscape with a vacuum that allows the formation of structures and, eventually, life. This argument is known as "the anthropic principle". Despite of being an attractive idea, is is highly controversial amongst physicists.

Since the Universe extends beyond the horizon, it is natural to ask whether it is the same in arbitrarily far regions, or whether we observe a particular Universe, emerged in a landscape of different vacua, or even more radically different Universes. As discussed in Tegmark 2007, there might be four types of Multiverse. The first one is trivial : the simplest inflationary model predicts that there is an infinite number of Hubble volumes, each one being a particular realisation of the initial conditions. This is nothing but a prediction of inflation. The second level allows for different physical constants amongst the Hubble volumes. The string theory landscape is include at this level. The third and fourth types are, respectively, the many worlds of quantum mechanics and a mathematical multiverse, where the laws of physics are different. The two latter are highly speculative. However, the first two are natural consequences of our current paradigm in cosmology. In the case that the Universe is Euclidean, which is favored by data Aghanim et al. 2018, either it continues indefinitely, or at some point things start to look different w.r.t. physical constants, for instance. We can call it a "cosmological multiverse", or simply a Universe where things are very different from place to place on ultra-large scales Carroll 2019.

1.2 How this work is contextualized in modern cosmology

At this point, our strategy is to look to nature and search from some "tip" of how to proceed. This is, in the vision of the present author, why we are engaging in this huge effort of mapping out the large-scale structures. This data might show us with high precision where exactly the standard model of cosmology is incomplete. Therefore, we need to extract every possible piece of information from the data. This task is not trivial and is one of the main underlying theme in many works in observational cosmology in the last decades. This work is part of this program.

In a galaxy survey, we always measure the distribution of galaxies. This distribution depends on aspects of galaxy formation which are not fully comprehended, baryonic process and non-linear perturbation theory which are hard to model. On the other hand, given that dark matter is collisionless, we can straightforwardly simulate its evolution from initial conditions and by using semi-analytical models to assign galaxies to dark matter halos, we can simulate the observed galaxy distribution. Figure 1.1 shows the comparison between the distribution of galaxies in real data and the one we obtain by assigning galaxies to dark matter halos of a N-body simulation. It is incredible how similar they are.

Despite of being able to reproduce the distribution of tracers very well, to extract cosmological information from it is challenging. N-body simulations are time consuming and the semi-analytical models to assign galaxies to halos are merely empirical. Therefore it is important to develop analytical models to compare to the data.

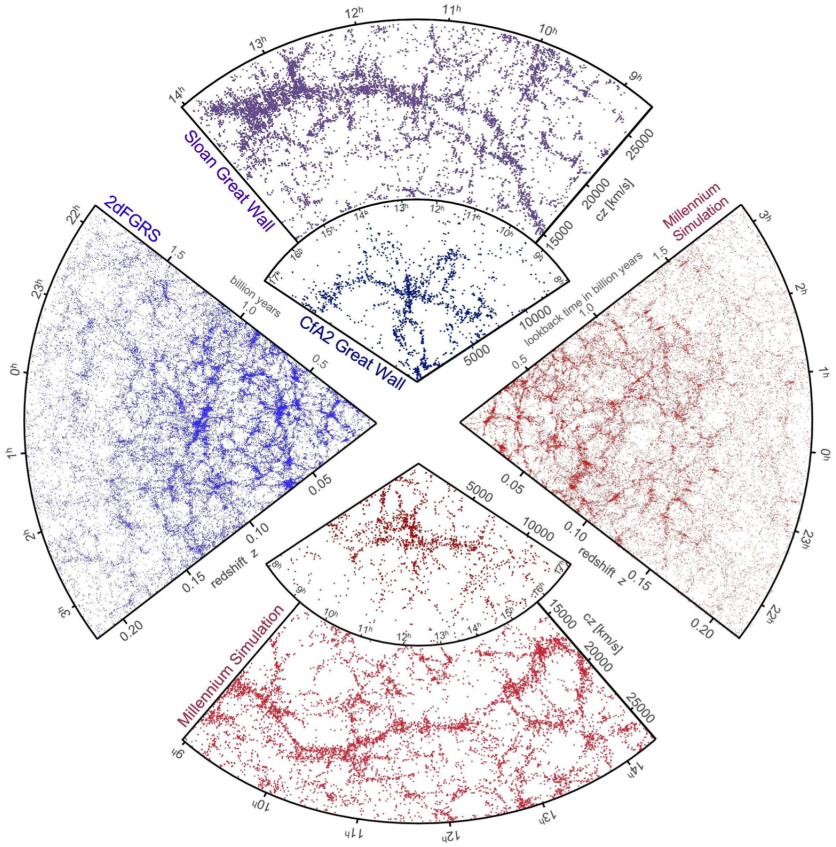

FIGURE 1.1 : A pie diagram of the distribution of galaxies measured by CfA2, 2dF and SDSS surveys. The lower right pies show galaxies assigned to dark matter halos through an HOD prescription. Figure extracted from SPRINGEL et al. 2006

Since dark matter is easier to treat analytically, we can apply perturbation theory to the Einstein's equations to find the evolution of linear perturbations in the homogeneous and isotropic dark matter fluid. To connect this prediction to observations, we encapsulate the theoretical unknowns into a single *bias* parameter, $b_g^{(1)}$, and express the linear perturbations of any tracer at linear order as :

$$\delta_g(\boldsymbol{x}) = \frac{n_g(\boldsymbol{x}) - \bar{n}_g}{\bar{n}_g} = b_g^{(1)} \left(\frac{\rho_m(\boldsymbol{x}) - \bar{\rho}_m}{\bar{\rho}_m} \right) \equiv b_g^{(1)} \delta_g(\boldsymbol{x}), \tag{1.1}$$

where n_g is the galaxy number density, ρ_m is the dark matter density field and bars indicate comoving averaged quantities. We use the subscript g to indicate that a quantity refers to any possible tracer of large-scale structure, such as quasars, Ly-α forests, cluster of galaxies, voids, etc.

The linear bias relation describes the two-point correlation function of galaxy number density on sufficient large scales $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$. The relation between the linear bias and the number density of clusters can be obtained through the "peak-background split" Mo et al. 1996 ; Sheth et Tormen 1999. In this picture, the linear bias parameter is the "response" of the abundance of clusters w.r.t. changes in long wave-length perturbations. The abundance of halos, as we further discuss in section 2.2, is determined by regions where the smoothed density contrast exceeds a certain threshold.

Different tracers of large-scale structure will then respond differently to long wave-length perturbations. Halos tend to be formed where the long-wave length perturbations have larger values, because small scale perturbations in these regions have a higher probability of exceeding the threshold for halo formation (see Fig. 1.2), whereas voids will have a higher probability of exceeding the threshold for void formation where the long-wave length perturbations are underdense.

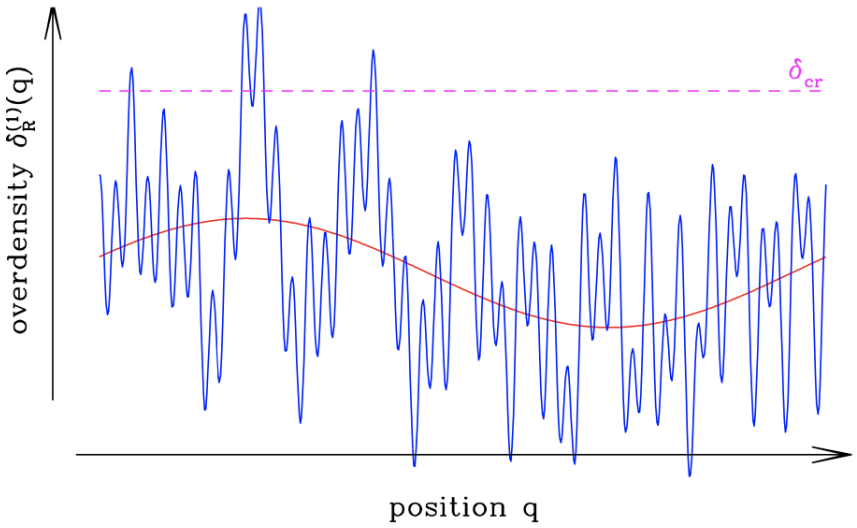

Figure 1.2 : Representation of the overdensity field in one dimension. Overdensities which belongs to long-wave lentgh perturbations have a higher probability of crossing the threshold for halo formation. Extracted from Desjacques et al. 2018

Therefore, the derivative (or "response") of halos and voids might be, respectively, positive and negative, and so their linear biases.

The fact that voids are tracers of large-scale structure with negative linear biases[6]

[6]Whether the linear bias of a void sample is negative or positive depend on the details of the

makes them carry complementary information to positively biased tracers. It has been shown that the multi-tracer Fisher matrix is not bounded as in the single tracer case (for instance, Abramo et al. 2013), as a consequence the precision with which cosmological parameters can be measured is not limited by cosmic variance, but only limited by shot-noise[7]. Given that voids trace the large-scale structures in a particular way compared to any other tracer found mostly in overdensities, then it is expected that in a multi-tracer analysis, the complementary information from voids is valuable Zhao et al. 2022; Contarini, Pisani et al. 2023.

Voids are also interesting on their own. Since they are underdense in matter, they can be thought as the best "laboratories" to detect signatures of physical process which are less effective on high-density environments. It has been shown that the void abundance is a sensitive probe of dark energy models Pisani et al. 2015, massive neutrinos Massara, Villaescusa-Navarro et al. 2015 and modifications to gravity Contarini, Marulli et al. 2021; Perico et al. 2019.

Despite being promising, the void cosmology is in its infancy, having some challenges ahead. The most obvious limitation is the shot-noise, i.e., the fact that voids are the most sparse tracer of large-scale structure. Furthermore, since voids are underdense in matter, there are less galaxies in voids. Since voids can only be observed through galaxies, the detection of voids will be affected by the poor sampling of galaxies inside voids.

On the theoretical side, we don't have so far a model to describe the density and velocity profiles around voids. All that we can do is to model how redshift-space distortions affect the density profile of voids in configuration space, or how voids are affected by Alcock-Paczynski effect.

Regarding the fact that we only observe voids through galaxies, a natural way around this is to "see" voids in the total-matter field by measuring their effect on the shapes of background galaxies. In other words, to measure the weak-lensing shear profile around voids, which is basically the projection of the void density profile along the line-of-sight. This line of investigation is known as Void-Lensing.

void finder algorithm, as well as the void size. But it safe to say that the largest void in large-scale structure, or sub-voids inside them, will tend to have negative linear bias. The positive linear bias is more related to voids-in-clouds.

[7] In fact, the shot-noise limitation in a multi-tracer analysis will be relevant, since that by splitting the galaxy sample into sub-samples, we inevitably increase the shot-noise of each sub-sample. The multi-tracer analysis is expected to be more efficient when a large number density of galaxies is observed.

The possibility that we could be able to measure weak-lensing around voids was first appreciated by Amendola et al. 1999. Since then, a few measurements were made Carles Sánchez et al. 2016; Fang et al. 2019; Melchior et al. 2014, as well as analytical and numerical investigations have shown the sensitivity of this observable to modifications of gravity Barreira et al. 2015; Baker et al. 2018; Davies, Cautun et al. 2019.

Arguably, Void-Lensing science is much less developed than void science in general. For instance, we cannot derive cosmological constraints from it so far, since the theoretical prediction for density profiles around voids is not known. Everything previous measurements did was to detect weak-lensing around voids, but no work has interpreted it. Furthermore, since that the void definition is not unique and arguably the "variance"[8] in void definition is higher amongst different void finder than the halo definition is amongst different halo finders, it is not clear which strategy works best in a fairly realist set-up.

In summary, the state-of-the-art of Void-Lensing science is : we know it is highly sensitive to modifications to gravity[9] in particular, and we have detected a few times with a relative significant signal-to-noise, but never performed cosmological analysis with this observable.

This work aims to give some contributions to the development of this field. We explore the freedom of choice in the void definition and show that for a suitable choice, the void-lensing signal can be drastically increased. In particular, voids found in the projected field provide a significantly deeper S/N. We show that these voids, for which there is no theoretical prediction, are correlated with voids defined in the 3D field, for which we know how to predict the abundance. We also propose a model to predict the profiles of voids in the projected field with profiles of voids found in the 3D field and their abundance. We believe that these contributions are of the crucial importance for future analysis involving Void-Lensing.

[8] Here, I use the term "variance" in an informal way to express the fact that the void definition amongst void finders varies more than the halo definition in halo-finder. This happens because voids are not virialized objects such as halos.

[9] The sensitivity to modifications to gravity strongly varies with the choice of the void-finder algorithm, as shown in Cautun et al. 2018.

1.3 The ΛCDM paradigm

Data strongly suggests that the Universe is passing through a phase of accelerated expansion Riess et al. 1998. This implies the Universe to be dominated by a cosmological constant at low redshift. We also have strong evidence, through the measurement of the anisotropies in the temperature distribution of the Cosmic Microwave Background (CMB), that the Universe at redshift $z \simeq 1090$ was almost perfectly homogeneous and isotropic, with small perturbations of order of $\simeq 10^{-5}$ Nabila Aghanim et al. 2020.

The precise shape of the CMB power spectrum is in astonishing accordance with a flat-six-parameters model (ΛCDM), where the parameters are the fractional densities of baryons, dark-matter, optical depth due to reionization, the Hubble parameter, the scalar spectral index and the scalar power-spectrum amplitude. Figure 1.3 shows the agreement between the six free-parameter ΛCDM model against the data from Planck telescope. Arguably, this measurement started the so-called "precision cosmology era", in which the standard model is capable of fitting the data with sub-percent errors.

The observation that the CMB is almost homogeneous and isotropic, implies that the metric of the Universe is the one of an homogeneous and isotropic Universe on sufficient large scales, unless there is an unknown mechanism which acts on scales of the order of the horizon which breaks the isotropy with time. In fact, we have evidence that the Universe is indeed homogeneous and isotropic at low redshift on scales $\geq 70h^{-1}$Mpc Ntelis 2016; Sarkar et al. 2009.

1.3.1 A brief history of background expansion

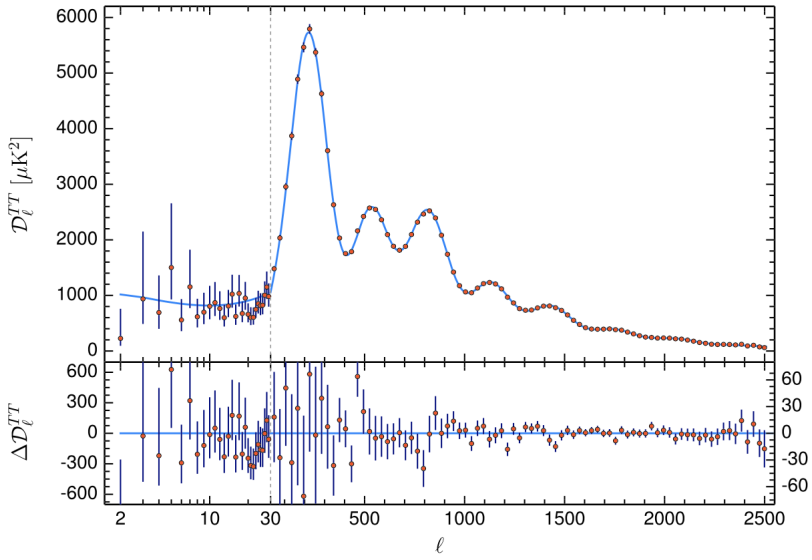

FIGURE 1.3 : Anisotropies in the CMB measured by the Planck telescope. The blue line is the fit of the six free-parameters ΛCDM model. The x-axis is the multipole moments and the $y-$axis is the variance at each l. The low ls correspond to large-scales, whereas high ls correspond to small scales. Extracted from Nabila AGHANIM et al. 2020

Given that there is no such mechanism, the evolution of the background Universe is determined by the *scale factor* $a(t)$, which is the stretching factor. The relation between the redshift and the scale factor is :

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{a_{\text{obs}}}{a_{\text{emit}}} = \frac{1}{a_{\text{emit}}}, \tag{1.2}$$

where by convenience, it is defined that the scale factor at present time is unity. The metric on sufficient large scales can then be written as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a(t)^2 d\Sigma^2, \tag{1.3}$$

where $d\Sigma^2$ is a 3D hypersurface of constant t :

$$d\Sigma^2 = \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2(\theta) d\phi^2). \tag{1.4}$$

In the above expressions, $K = 0, 1, -1$ corresponds to flat, closed and open Universe, $c = 1$[10] and we use spherical coordinates for the spatial section, $(x^1, x^2, x^3) = (r, \theta, \phi)$. This metric is known as Friedmann-Lamaître-Robert-Walkson (FLRW) metric. Notice that the time evolution of the metric is completely determined by the scale factor.

In a homogeneous and isotropic Universe the energy momentum tensor reduces to the one of a perfect fluid :

$$T_v^\mu = (\rho + P)u^\mu u_v + P\delta_v^\mu \ , \tag{1.5}$$

where $u = (-1, 0, 0, 0)$ is the four-velocity of the fluid in comoving coordinates, ρ and P are, respectively, the energy density and pressure of the fluid and δ_μ^ν is the Kronecker delta. The pressure can be related to the density (on large-scales) as $P = w\rho$.

The evolution of $a(t)$ and $\rho(t)$ can be found by plugging these ingredients into the Einstein's equations[11] :

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}, \tag{1.6}$$

where $R_{\mu\nu}$ is the Riemann tensor, defined as

$$R_{\mu v} = \Gamma^\alpha_{\mu\nu,\alpha} - \Gamma^\alpha_{\mu\alpha,\nu} + \Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\alpha\beta} - \Gamma^\alpha_{\mu\beta}\Gamma^\beta_{\alpha\nu} \tag{1.7}$$

where

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2}g^{\alpha\beta}\left(g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}\right) \tag{1.8}$$

are the Christoffel symbols (or connection) and $R \equiv g_{\mu\nu}R^{\mu\nu}$ is the Ricci scalar.

By plugging these ingredients into the Einstein's equations, we obtain the Friedmann equations

$$\begin{aligned} H(t)^2 &= \frac{8\pi G}{3}\rho(t) - \frac{K}{a(t)^2} \ , \\ 3H(t)^2 + 2\dot{H(t)} &= -8\pi G P - \frac{K}{a(t)^2}, \end{aligned} \tag{1.9}$$

[10] This convention will not always hold through out this work. We will explicitly inform whenever $c = 1$ is used.

[11] Notice that we are not including the cosmological constant on the left-hand side. That's because here we understand the cosmological constant as a form of energy, which is included in ρ.

where

$$H(t) \equiv \frac{1}{a}\frac{da}{dt} \tag{1.10}$$

is the Hubble parameter.

The continuity equation follows from the Bianchi identities

$$\nabla_\mu G^\nu_\mu \equiv \frac{\partial G^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu} G^\alpha_\nu - \Gamma^\alpha_{\nu\mu} G^\mu_\alpha = \nabla_\mu T^\nu_\mu = 0, \tag{1.11}$$

where ∇_μ denotes covariant derivatives. The last equality gives the continuity equation

$$\dot{\rho}(t) + 3H(t)(\rho(t) + P(t)) = 0. \tag{1.12}$$

We can rewrite equation 1.9 as

$$\Omega_M + \Omega_K = 1, \tag{1.13}$$

where

$$\Omega_M \equiv \frac{8\pi G \rho_M}{3H(t)^2}, \quad \Omega_K \equiv -\frac{K}{(a(t)H(t))^2}, \tag{1.14}$$

with M denoting the sum of matter, radiation and dark-energy components.

Let's consider the case in which the Universe is flat and dominated by a single component. Then we can solve the Friedmann equations analytically and get to the following solutions :

$$\rho(t) \propto a(t)^{-3(1+w)}, \quad a(t) \propto t^{2/(3(1+w))}. \tag{1.15}$$

For radiation, $w = 1/3$ and for matter, $w = 0$. In order to reproduce the late cosmic acceleration, the equation of state for dark energy must be $w < -1/3$. Since the data is consistent with a constant dark-energy, the equation of state must be $w = -1$. The negative pressure is a requirement for having a constant energy while the volume expands. Therefore, in the cases of a radiation, matter and dark-energy dominates universes, we have, respectively $\rho_r(t) \propto a(t)^{-4}$ ($a(t) \propto t^{1/2}$), $\rho_m(t) \propto a(t)^{-3}$ ($a \propto t^{2/3}$) and $\rho_\Lambda = cte$ ($a \propto \exp(Ht)$).

Therefore, at early times, the radiation dominates, at $a_{eq} = 4.15 \times 10^{-5} \Omega_m^{-1} h^{-2}$ switches to $a \propto t^{2/3}$ in the matter-dominated time and then to exponential growth at late times, when dark-energy takes over. The fact that dark-energy only takes

over very recently is referred as "the coincidence problem", because we happen to live right at this transition between matter and dark energy domination.

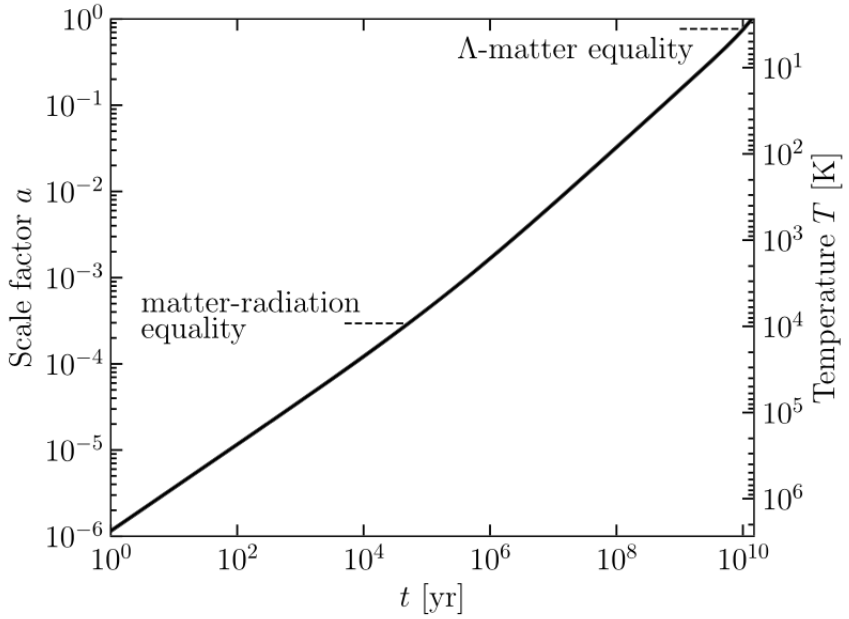

FIGURE 1.4 : The evolution of the scale factor as a function of time. At early times, the universe was radiation-dominated and the expansion of the scale factor was proportional to $t^{1/2}$, eventually it turns to the matter-domination era, with $a(t) \propto t^{3/2}$ and very recently dark energy takes over . Extracted from DODELSON et SCHMIDT 2020

1.3.2 The inhomogeneous Universe in a nutshell

Our Universe is not as simple as a homogeneous and isotropic spacetime with an equally smooth matter distribution. Instead, the Universe is a complicated place where radiation, dark matter and baryons interact in many ways, in particular gravitationally, leading to the formation of all kinds of structures, from atomic nuclei, atoms and molecules (driven by nuclear and electromagnetic interactions) to star, galaxies and clusters (driven by gravity).

Empirically, we know that matter is distributed in a peculiar way throughout space. The way by which matter organized itself into the Large-Scale Structure (LSS) is a booming field of research. The LSS encapsulates a lot of information about fundamental features of nature, and its detailed study is a promising way to shed light upon deep facts concerning the underlying laws of physics.

If we look upon spherical regions large enough ($\sim 200\ h^{-1}$ Mpc), the contrast between the matter density inside this sphere and the background density is much less than 1, and we can use linear theory to obtain many useful results. This is the subject of this section.

1.3.3 Gauge-invariant variables

In GR we are allowed to choose any coordinate system. This freedom of choice, or *gauge freedom*, is a fundamental feature of covariant theories of gravity, but it can lead to difficulties interpreting the physical meaning of perturbations. In particular, it may happen that with one choice of coordinate system we compute density perturbations which, upon closer inspection, are not manifested in the physical observables.

Following the argument by Mukhanov MUKHANOV 2005, consider a homogeneous and isotropic universe where the energy density is distributed evenly throughout the space, i.e. $\rho(\boldsymbol{x}, t) = \rho(t)$. Since any coordinate system is allowed, we can make the particular choice such that the time coordinate relates to the old one as $\tilde{t} = t + \delta t(\boldsymbol{x}, t)$, where $\delta t(\boldsymbol{x}, t) \ll t$. Thus, the energy density in the new coordinate system $\rho(\tilde{t}, \boldsymbol{x}) = \rho(t(\boldsymbol{x}, \tilde{t}))$ will depend on $\boldsymbol{x}$ in general. Furthermore, with this coordinate choice we generate perturbations which are not present in the old coordinate system. Indeed, if we expand $\rho(t)$,

$$\rho(t) = \rho(\tilde{t} - \delta t(\boldsymbol{x}, t)) \simeq \rho(\tilde{t}) - \frac{\partial \rho}{\partial t}\delta t\ , \tag{1.16}$$

we see that the energy density splits into a background term, $\rho(\tilde{t})$, plus a non-physical perturbation, which is entirely due to our choice of coordinates. Conversely, it is also possible to remove a real perturbation by choosing a coordinate system such that the hypersurfaces of constant energy density coincide with the hypersurfaces of constant time.

We could think that if *gauge-invariant* perturbations exist, we could check if the perturbations due to a particular choice of coordinate system is fictitious or not. If these gauge-invariant perturbations vanish in one coordinate system, they must vanish in any coordinate system. Therefore, if there are perturbations in any coordinate system and the gauge-invariant perturbations vanish, then these perturbations are fictitious and can be removed by a change of coordinates.

Here we will briefly discuss gauge transformations. For a exhausting discussion on this topic, see Ma et al. 1995.

1.3.4 Classification of perturbations

The perturbed FLRW spacetime can be expressed in terms of the metric

$$ds^2 = [g^{(0)}_{\mu\nu} + \delta g_{\mu\nu}(x^\alpha)]dx^\mu dx^\nu \ . \tag{1.17}$$

It is useful to define the *conformal time*

$$\eta \equiv \int \frac{dt}{a(t)} \ . \tag{1.18}$$

The background metric is written as :

$$g^{(0)}_{\mu\nu}dx^\mu dx^\nu = a^2(\eta)[-d\eta^2 + \delta_{ij}dx^i dx^j] \ . \tag{1.19}$$

The perturbations on the metric $\delta g_{\mu\nu}$ can be split into scalar, vector and tensor perturbations which, in the most general form, can be written, respectively, as

$$\delta g^{scalar}_{ij} = \begin{pmatrix} 2a^2\phi & B,i \\ B,i & 2a^2(\psi\delta_{ij} + E,ij) \end{pmatrix} \tag{1.20}$$

for scalar perturbations,

$$\delta g^{vector}_{ij} = \begin{pmatrix} 0 & S_i \\ S_i & a^2(F_{i,j} + F_{j,i}) \end{pmatrix} \tag{1.21}$$

for vector perturbations, and

$$\delta g^{tensor}_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & h_{ij} \end{pmatrix} \tag{1.22}$$

for tensor perturbations. In the expressions above, S_i and F_i are divergenceless ($S^i{}_{,i} = F^i{}_{,i} = 0$), so each has two independent components, and h_{ij} is a traceless and transverse tensor, i.e, $h^i_i = h^i_{j,i} = 0$. Since h^i_j is a symmetric tensor (6 independent components), the traceless condition eliminates one component and the transverse conditions eliminate 3 components, leaving two independent components. Therefore

the scalar, vector and tensor independent functions give ten independent functions.

The only perturbations we are interested in when treating cosmological inhomogenities are scalar perturbations, typically because they are the only ones that can be sourced by energy density perturbations.

1.3.5 Gauge transformations

Consider the infinitesimal transformation :

$$x^\rho \longrightarrow \tilde{x}^\rho = x^\rho + \xi^\rho \ , \tag{1.23}$$

where $\xi^\rho = (\xi^0, \xi^i)$ and $\xi^i = \xi^i_\perp + \zeta^{,i}$ can be split into a 3-vector with zero divergence plus the spatial derivative of a scalar function ζ. The metric calculated in the new coordinate system will transform through the usual tensor transformation law. The perturbed part transforms in a non-trivial way Mukhanov 2005 :

$$\delta g_{\alpha\beta} \to \delta\tilde{g}_{a\beta} = \delta g_{\alpha\beta} - g^{(0)}_{a\beta,\gamma}\xi^\gamma - g^{(0)}_{\gamma\beta}\xi^\gamma_{,\alpha} - g^{(0)}_{\alpha\delta}\xi^\delta{}_{,\beta} \ . \tag{1.24}$$

Using the transformation law (1.24) and the scalar part of the perturbed metric (1.20), we easily find how the scalar perturbation functions transform under (1.23) :

$$\begin{aligned} &\phi \to \tilde{\phi} = \phi - \frac{1}{a}\left(a\xi^0\right)', \quad B \to \tilde{B} = B + \zeta' - \xi^0 \\ &\psi \to \tilde{\psi} = \psi + \frac{a'}{a}\xi^0, \quad E \to \tilde{E} = E + \zeta \ . \end{aligned} \tag{1.25}$$

That is, the way that scalar perturbations transform when we pass from the background metric to any other coordinate system is totally defined by the functions ξ^0 and ζ. We can choose an infinitesimal transformation whose effect is to vanish any of the scalar functions (1.25). However, if we can make them all vanish, then in the new coordinates we would not see any perturbation, and the homogeneous and isotropic background would be exact. In other words, in the presence of perturbations there is a minimum set of physical degrees of freedom that cannot be made to vanish by any choice of coordinates.

It is easy to check that the combinations

$$\Phi \equiv \phi - \frac{1}{a}\left[a\left(B - E'\right)\right]', \quad \Psi \equiv \psi + \frac{a'}{a}\left(B - E'\right) \ , \tag{1.26}$$

are gauge invariant – i.e. they do not depend on the transformation (1.23). Thus, if the functions (1.26) vanish in one coordinate system, then they vanish in any coordinate system and there are no real perturbations.

The particular choice of ξ^0 and ζ corresponds to a gauge choice. A widely used gauge in the literature is to treat scalar inhomogeneities as fixed on the background, corresponding to $\xi^0 = \zeta = 0$. In this gauge – the Newtonian gauge – the gauge invariant functions are simply $\Phi = \phi$ and $\Psi = \psi$.

In the Newtonian gauge, the invariant distance interval takes the form

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\phi)\mathrm{d}\eta^2 + (1+2\psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] . \tag{1.27}$$

1.3.6 Qualitative analysis of linear perturbations

The primordial Universe was a very smooth and hot plasma, which was also very opaque : the mean free path of a photon one second after the Big Bang was only about the size of an atom, while today a photon can travel almost freely over cosmological distances. The initial conditions of the Universe are usually described by the mechanism known as cosmic inflation. Although we are not sure if inflation really is the theory which describes the primordial universe, it is the best explanation for the so-called *horizon* and *flatness* problems. The horizon problem expresses the fact that the CMB is extremely isotropic, despite the CMB photons arriving to us from different directions which, at the time of decoupling, apparently never had the chance to have causal contact. This condition is best expressed in terms of the conformal time η : this is also the maximum comoving distance traveled by a photon since the Big Bang. When $\eta k \ll 1$, the wavelength of a perturbation is much larger than the maximum distance traveled by a photon since the Big Bang, and no causal physics could have affected the evolution of such perturbations.

Inflationary theory proposes a mechanism by which the causal contact between apparently disconnected places was possible in the very early Universe. This mechanism consists of exponential expansion at the very beginning, and it can be caused by a scalar field – the inflaton Baumann 2009. It was also found, early in the development of inflation, that this mechanism can also provide the initial conditions for cosmic perturbations. With the next generation of LSS surveys we are going to impose stringent constraints on inflationary models. In particular, the bispectrum measurement can impose interesting constraints on the f_{NL} parameter,

which parameterizes the deviation from Gaussianity in the primordial fluctuations Tellarini et al. 2016. Although the treatment of inflation is out of the scope of this work, we will implicitly assume initial conditions for density fluctuations which are consistent with inflation.

Before discussing perturbation theory quantitatively, let us briefly expose what one might expect from the math. The dynamics of the density contrast responds mainly to two forces, namely, pressure and gravity as represented schematically by :

$$\ddot{\delta} + [\text{Pressure} - \text{Gravity}]\delta = 0 \, . \tag{1.28}$$

If gravity dominates, one expects the density contrast to grow exponentially ; on the other hand, if pressure is not negligible, the density contrast oscillates in time.

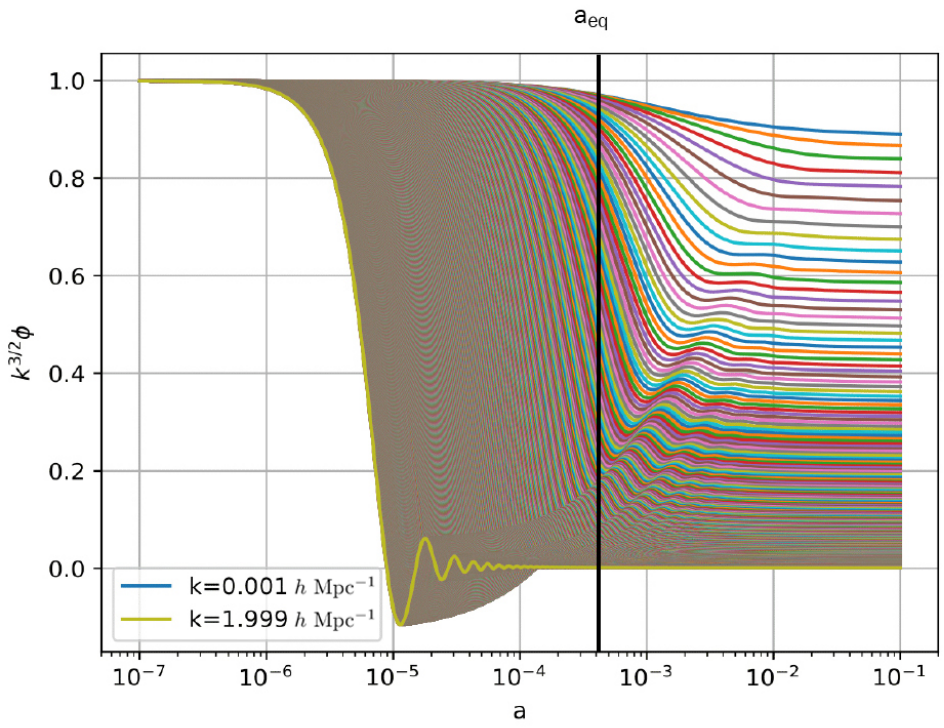

Figure 1.5 : The linear evolution of the gravitational potential ϕ.

This qualitative picture leads naturally to three stages of evolution of cosmological perturbations. At $a_{eq} \simeq 4 \times 10^{-4}$ the energy densities of radiation and matter were equal, so before that time radiation dominates, and after that time it is matter

which dominates. As shown in Fig. (1.5), at early times ($\eta k \ll 1$) all modes are outside the horizon and therefore none of them evolve (they are frozen). At $a \simeq 10^{-6}$ the small scale mode $k = 1.9\,h$ Mpc^{-1} enters the horizon and begins to decay (this happens because radiation pressure tends to dilute perturbations). The modes which entered the horizon after the matter-radiation show an evolution very different from that of the modes which entered before. Finally, at late times, when the universe is matter dominated, all modes evolve in the same way.

The period of transition between radiation and matter domination is described by the transfer function, which is defined as the ratio between the potential for mode k well after matter starts to dominate (a_{late}) and the potential for an extremely large-scale mode at the same time :

$$T(k) = \frac{\Phi(k, a_{late})}{\Phi_{Large-Scale}(k, a_{late})} \, . \tag{1.29}$$

We will show that the large-scale solution is the primordial potential decreased by a factor of 9/10. Thus,

$$\Phi(k, a_{late}) = \frac{9}{10}\Phi_P T(k) \, . \tag{1.30}$$

In the above result, Φ_P is the primordial potential (predicted by some inflationary model). Roughly speaking, the transfer function encodes the information of how modes change in the matter-radiation equality period.

After the matter-radiation equality, all modes evolve equally and the evolution of perturbations does not depend on k (although the initial conditions in the matter era do, as evidenced by the scale-dependence of the transfer function), in such a way that the evolution is determined by a function of the scale factor, or equivalently a function of redshift. Once the potentials are set out, one expects matter to be attracted by regions where there is more matter (overdense regions). The growth of such regions is described by the Growth function, which is defined by the ratio between the potential at some time and its value well before matter starts to dominate :

$$\frac{D(a)}{a} = \frac{\Phi(a)}{\Phi(a_{late})}, \qquad (a > a_{late}) \, . \tag{1.31}$$

Therefore, the evolution of the potential can be written as :

$$\Phi(k,a) = \frac{9}{10}\Phi_P(k)T(k)\frac{D(a)}{a}, \qquad (a > a_{late}) \ . \tag{1.32}$$

If one could measure some quantity which is related to the potential, then the model for inhomogeneities in the Universe could be compared to the data. We can relate the potential to the density contrast through the Poisson equation, and the density contrast is measured directly in LSS surveys – to be more precise, what we actually measure is the matter power spectrum $P(\boldsymbol{k})$ or the correlation function $\xi(\boldsymbol{r})$.

The Fourier version of Poisson's equation reads :

$$\Phi = \frac{4\pi G\rho_m a^2\delta}{k^2} \ . \tag{1.33}$$

Using (1.32), the background density of matter $\rho_m = \Omega_m\rho_{cr}/a^3$ and the critical density $\rho_{cr} = (3/2)H_0^2/(4\pi G)$, the density contrast is :

$$\delta(k,a) = \frac{3}{5}\frac{k^2}{\Omega_m H_0^2}\Phi_P(k)T(k)D(a), \qquad (a > a_{late}) \ . \tag{1.34}$$

Finally, the density contrast is related to the power spectrum as :

$$P(\boldsymbol{k},a) = \langle|\delta(\boldsymbol{k},a)|^2\rangle \ , \tag{1.35}$$

which is the most basic observable in cosmology[12].

After the inflationary period, the expansion rate decelerates and perturbations start to fall back inside the Hubble horizon H^{-1}. The small scales enter the horizon before suffering the effect of radiation pressure, which tends to dilute perturbations. There is a value of k ($k_{eq} \simeq 0.02\ h\ \mathrm{Mpc}^{-1}$) which happens to come into the horizon exactly at the time of equality. This means that k_{eq} denotes the smallest scale that does not suffer the effects of radiation pressure, and therefore is not diluted. Now, taking into account the term k^2 in Eq. (2.37), which comes from the Poisson equation, we conclude that k_{eq} corresponds to the peak of the power spectrum : scales smaller than k_{eq} entered the horizon before the equality and were diluted by pressure gradients ; and scales larger than that are suppressed by the k^2 term.

[12]Of course, in real observations we never have access to the matter power-spectrum apart from observational effects, such as the galaxy bias and the redshift-space distortions.

1.3.7 Quantitative analysis of linear perturbations

The set of equations we need in order to characterize the photons and matter perturbations are the Boltzmann equations DODELSON 2003 :

$$\begin{aligned} \dot{\Theta}_{r.0} + k\Theta_{r.1} &= -\dot{\Phi}, \\ \dot{\Theta}_{r.1} - \frac{k}{3}\Theta_{r.0} &= \frac{-k}{3}\Phi, \\ \dot{\delta} + ikv &= -3\dot{\Phi}, \\ \dot{v} + \frac{\dot{a}}{a}v &= ik\Phi, \end{aligned} \tag{1.36}$$

where $\Theta_{r.0}$ and $\Theta_{r.1}$ are the monopole and dipole of radiation perturbations (photons + neutrinos), and v is the velocity field of dark matter. These equations are complemented by the relativistic Poisson equation for the potential, which is given by the time-time component of Einstein's equations :

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0}\right] \ , \tag{1.37}$$

where ρ_{dm} is the dark matter density. We can also write an algebraic equation for the potential, obtained through the combination of space-time Einstein's equations with the Poisson equation :

$$k^2\Phi = 4\pi G a^2 \left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}\left(i\rho_{\rm dm}v + 4\rho_r\Theta_{r,1}\right)\right] \ . \tag{1.38}$$

This set of equations needs some remarks, since they are not the full set of Boltzmann equations in all its glory ibid., instead, they are result of two simplifications. First, the baryons were neglected, since they compose only a small fraction of matter. Second, the higher moments of photon perturbations $(\Theta_2, \Theta_3, ...)$ were neglected. This latter simplification is justified because before the recombination photons are strongly coupled to matter. Since perturbations in non-relativistic matter are only described by the two first momenta, which correspond to δ and v, then the photon perturbations are well described by the two first momenta Θ_0 and Θ_1.

Analytical solutions for the full set of equations are impossible to obtain. Hence, we will perform approximations and obtain analytical solutions valid at some times and on some scales. These approximations are illustrated by Figure (1.6). On

super-horizon scales we can neglect terms multiplying k, since they will be $k\eta$ times smaller than the other terms, and on super-horizon scales $k\eta \ll 1$. On scales which enter the horizon before the equality, we know that the solution for the potential is a constant (after the transfer function regime). At the radiation-dominated epoch we can neglect matter perturbations and at the matter dominated epoch we can neglect radiation perturbations.

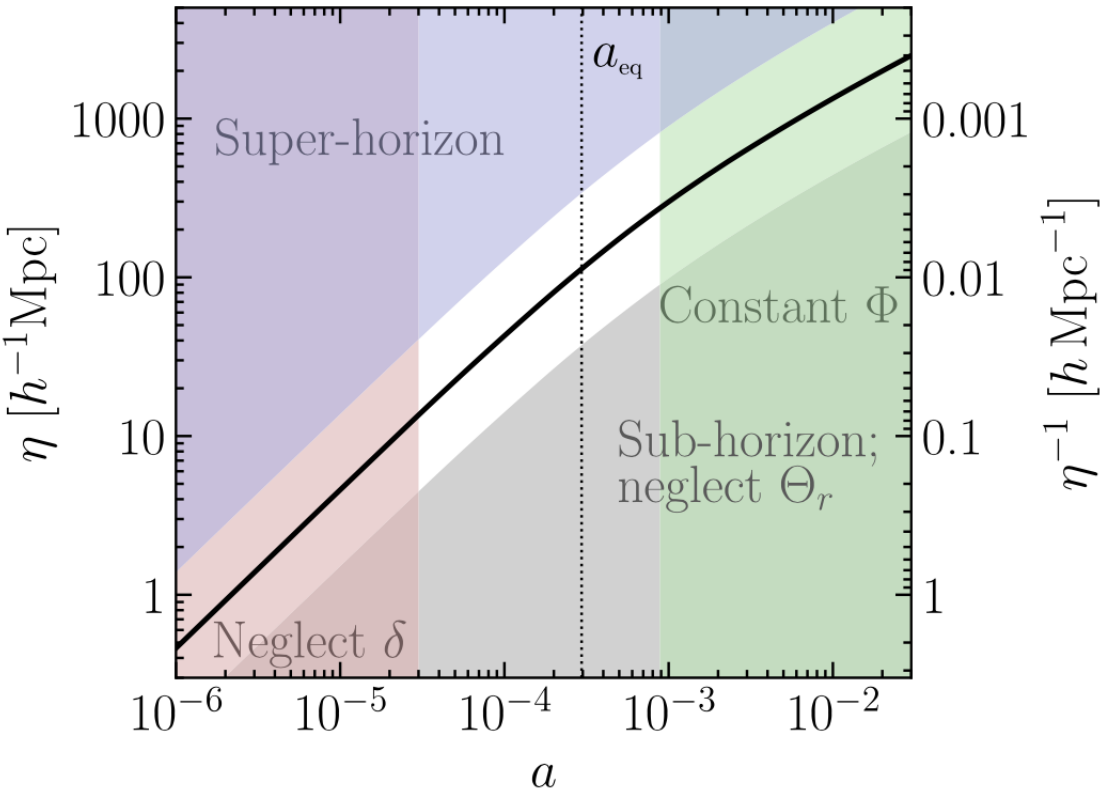

FIGURE 1.6 : Regimes where it is possible to perform approximations and derive analytical solutions. Figure extracted from DODELSON 2003

Large scales

On large and super-horizon scales, we can neglect terms involving k, and therefore the set of Eqs. (1.36)-(1.38) reduces to the three equations :

$$\dot{\Theta}_{r,0} = -\dot{\Phi}, \tag{1.39}$$

$$\dot{\delta} = -3\dot{\Phi}, \tag{1.40}$$

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0}\right] . \tag{1.41}$$

Therefore, at first approximation, the dynamics of perturbations does not depend on the velocity field v nor on the dipole Θ_1. This feature can be intuitively understood as the independence of the large scale perturbations on local fluxes of matter and radiation. On large scales the small-scale fluxes cancel, and only the monopole term determines the dynamics.

The first two equations lead to $\delta - 3\Theta_{r,0} = const.$ The adiabatic initial conditions set this constant to zero. Using it in the Einstein's equation (equation (1.41)) yields :

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \rho_{\rm dm}\delta\left[1 + \frac{4}{3y}\right] , \tag{1.42}$$

where $y \equiv \frac{a}{a_{eq}} \equiv \frac{\rho_{dm}}{\rho_r}$ will be the evolution variable. Transforming the derivative $\frac{d}{d\eta} = H\, y\, \frac{d}{dy}$, the Einstein equation becomes :

$$y\,\Phi' + \Phi = \frac{y}{2(y+1)}\delta\left[1 + \frac{4}{3y}\right] = \frac{3y+4}{6(y+1)}\delta , \tag{1.43}$$

where the prime denotes derivatives with respect to y. Using $\delta' = -3\Phi'$ and differentiating with respect to y leads to :

$$\Phi'' + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\Phi' + \frac{\Phi}{y(y+1)(3y+4)} = 0 . \tag{1.44}$$

In terms of the variable $u \equiv \frac{y^3}{\sqrt{1+y}}\Phi$, the above equations become :

$$u'' + u'\left[\frac{-2}{y} + \frac{3/2}{1+y} - \frac{3}{3y+4}\right] = 0 . \tag{1.45}$$

Integrating and exponentiating yield :

$$u' = A\frac{y^2(3y+4)}{(1+y)^{3/2}}. \tag{1.46}$$

With the definition of u we have :

$$\frac{y^3}{\sqrt{1+y}}\Phi = A\int_0^y dy' \frac{y'^2\,(3y'+4)}{(1+y')^{3/2}}\ , \tag{1.47}$$

where A is a constant to be determined. The analytic solution for this equation is

$$\Phi = \frac{\Phi(0)}{10}\frac{1}{y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right]\ . \tag{1.48}$$

This is the analytic form for the potential on super-horizon scales, neglecting baryons. The important result that this solution gives us is that for large y, i.e, in the matter-dominated era, the y^3 term dominates and therefore $\Phi \to \frac{9}{10}\Phi_0$. That is, the largest super-horizon scales are slightly suppressed as the Universe passes from radiation-dominated to matter-dominated era. One might think that super-horizon scales should not be affected by what is happening inside the horizon since these scales are out of causal contact. Actually, inside the super-horizon scales there are smaller scales which are in causal contact, then, in some level perturbations on the largest scales will be affected. Eventually these scales will enter the horizon deep in the matter-dominated era and thus the potential will remain constant.

Small scales

The treatment of small scales can also be divided into two regimes : (i) super-horizon modes crossing the horizon well within the radiation era, and (ii) sub-horizon modes crossing the equality between matter and radiation eras. In the first regime, we can neglect the matter perturbations, since although matter perturbations are influenced by the potential, they do not influence the potential. Then, having a solution for the potential neglecting matter, we can use it as source to the evolution of matter perturbations.

In order to find a solution for the potential in regime (i), we use the algebraic equation (1.38). Neglecting matter, (1.38) reads :

$$\Phi = \frac{6a^2H^2}{k^2}\left[\Theta_{r,0} + \frac{3aH}{k}\Theta_{r,1}\right]. \tag{1.49}$$

In the above equation, we used that $H^2 = 8\pi G\rho_r/3$ in the radiation era, where we can also write $H \simeq H_0\sqrt{\Omega_r a^{-4}} = H_0\sqrt{\Omega_r}a^{-2}$. Thus, a^2H is approximately constant in the radiation era, and since $\eta = \int_0^a \frac{da}{a^2H}$, then $\eta = 1/aH$ well within the radiation era. Therefore, equation (1.49) along with the two equations for radiation perturbation,

$$\dot{\Theta}_{r,0} + k\Theta_{r,1} = -\dot{\Phi}$$

and

$$\dot{\Theta}_{r,1} - \frac{k}{3}\Theta_{r,0} = \frac{-k}{3}\Phi\ ,$$

yield :

$$-\frac{3}{k\eta}\dot{\Theta}_{r,1} + k\Theta_{r,1}\left[1 + \frac{3}{k^2\eta^2}\right] = -\dot{\Phi}\left[1 + \tfrac{k^2\eta^2}{6}\right] - \Phi\tfrac{k^2\eta}{3} \tag{1.50}$$

and

$$\dot{\Theta}_{r,1} + \frac{1}{\eta}\Theta_{r,1} = \tfrac{-k}{3}\Phi\left[1 - \tfrac{k^2\eta^2}{6}\right]\ . \tag{1.51}$$

Combining these two equations and eliminating $\dot{\Theta}_{r,1}$ and $\Theta_{r,1}$, we end up with a second order equation for the potential :

$$\ddot{\Phi} + \frac{4}{\eta}\dot{\Phi} + \frac{k^2}{3}\Phi = 0\ . \tag{1.52}$$

This is the equation we want to solve with the initial condition that the potential is constant (before crossing the horizon). Rewritten in terms of the variable $u \equiv \Phi\eta$, equation (1.52) reads :

$$\ddot{u} + \frac{2}{\eta}\dot{u} + \left(\frac{k^2}{3} - \frac{2}{\eta^2}\right)u = 0\ . \tag{1.53}$$

We recognize the above equation as the spherical Bessel equation of order 1. The general solution of this equation is a combination of the spherical Bessel function,

$j_1(k\eta/\sqrt{3})$ and the spherical Neumann function, $\eta_1(k\eta/\sqrt{3})$. The latter goes to infinity at small argument, then it is not part of our solution. The spherical Bessel function of order 1 can be written as :

$$\Phi = 3\Phi_p \left(\frac{\sin\left(k\eta/\sqrt{3}\right) - (k\eta/\sqrt{3})\cos\left(k\eta/\sqrt{3}\right)}{(k\eta/\sqrt{3})^3} \right) . \tag{1.54}$$

As we expect from our qualitative analysis, when the mode enters the horizon ($k\eta \lesssim 1$), it decreases (due to radiation pressure) and oscillates (due to the gravitational instability of baryons). Figure (1.8) shows two modes, $k = 10\ h\ \mathrm{Mpc}^{-1}$ and $k = 1\ h\ \mathrm{Mpc}^{-1}$, which enter the horizon in the radiation era, the dashed-lines are the numerical solutions (including matter perturbations) and the solid-lines are the analytical approximations. We see that our analytical approximation starts to break for modes entering the horizon at $a \simeq 10^{-5}$, i.e., our analytical approximation is only valid deep in the radiation era, since the equality happens at $a_{eq} \simeq 4 \times 10^{-4}$.

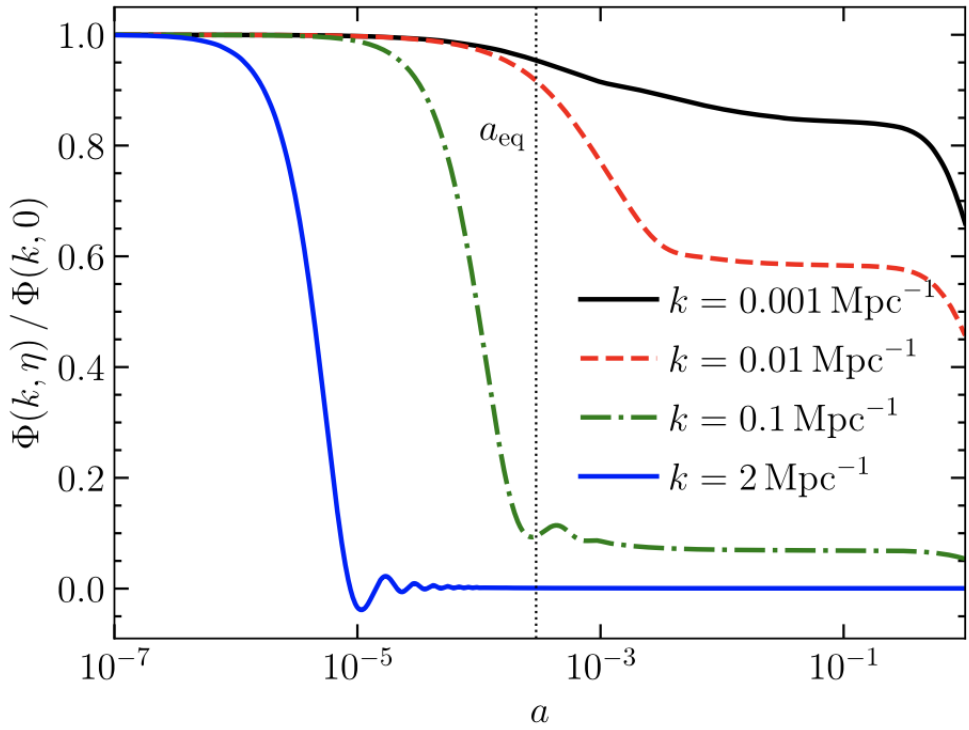

FIGURE 1.7 : The analytical (solid-line) and numerical (dashed-line) solutions for the potential. As soon as the mode enters the horizon in the radiation dominated era, it decreases and oscillates. Figure extracted from DODELSON 2003.

We can use the potential as a source for the evolution of matter perturbations. Combining the third and fourth equations in (1.36), we can write :

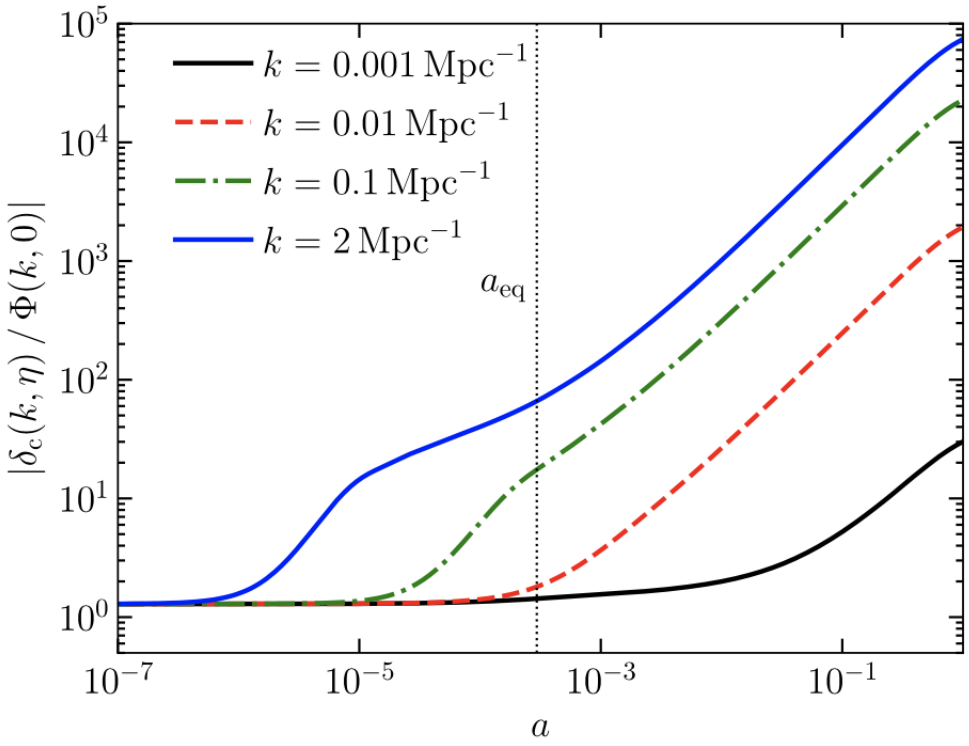

FIGURE 1.8 : The analytical (solid-line) and numerical (dashed-line) solutions for the matter density contrast. Figure extracted from DODELSON 2003.

$$\ddot{\delta} + \frac{1}{\eta}\dot{\delta} = S(k,\eta)\ , \tag{1.55}$$

where,

$$S(k,\eta) = -3\ddot{\Phi} + k^2\Phi - \frac{3}{\eta}\dot{\Phi}\ . \tag{1.56}$$

The solution for the above equation is DODELSON 2003 :

$$\delta(k,\eta) = A\,\Phi_p \ln(Bk\eta)\,. \tag{1.57}$$

Ref. HU et al. 1996 found the values $A = 9.6$ and $B = 0.44$ for the two remaining constants.

This solution can be interpreted as follows. The potential of a certain mode

is frozen until it enters the horizon and also the matter density contrast, this is expressed by the constant term in (1.57), $A\,\Phi_p \ln(B)$. After entering the horizon, the matter density contrast grows, despite the radiation pressure, but the growth is logarithmic (expressed by $A\,\Phi_p \ln(k\eta)$) and slower than in the matter dominated era, where it grows as $\delta \propto a$. Figure (1.8) shows our analytical solution (dashed-lines) and the numerical solution (solid-lines). We see that when getting closer to the end of the radiation era, the density contrast starts to grow faster.

Finally, we will obtain analytic solutions for the regime (ii), where small scales modes are well within the horizon and cross the epoch of equality between matter and radiation.

We can also perform an approximation that will simplify our calculations : neglecting the radiation perturbations. Arguably, when getting closer to the equality, eventually, it will happen that the matter perturbations dominate over the radiation perturbations, since radiation continuously dilutes on scales inside the horizon, despite the fact that ρ_r is still larger than ρ_{dm}. In other words, the potential evolution is dominated by matter perturbations even before the equality.

Therefore, using our prescription of turning three equations into one second order equation, we can use the the third and fourth equations in (1.36) and the algebraic (1.38) to obtain :

$$\delta'' - \frac{ik(2+3y)v}{2aHy^2(1+y)} = -3\Phi'' + \frac{k^2\Phi}{a^2H^2y^2}\,. \tag{1.58}$$

To obtain the above equation, one needs to consider that perturbations are well within the horizon, which leads to $aH/k \ll 1$. We are again using the ratio between the scale factor with its value at the equality, y, and primes denotes derivatives with respect to this variable.

Realizing that the potential is much smaller than δ on sub-horizon scales, we can use the equation

$$\delta' + \frac{ikv}{aHy} = -3\Phi'\,, \tag{1.59}$$

to replace $ikv/(aHy)$ by $-\delta'$ and finally obtaining the *Meszaros equation*,

$$\delta'' + \frac{2+3y}{2y(y+1)}\delta' - \frac{3}{2y(y+1)}\delta = 0\,, \tag{1.60}$$

which has as solution :

$$\delta(k, y) = C_1 D_1(y) + C_2 D_2 \ , \tag{1.61}$$

where $D_1(y) = y + 2/3$ and $D_2(y) = D_1(y) \ln\left[\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right] - 2\sqrt{1+y}$. This solution is valid on small scales and well after the mode entered the horizon, because at this regime the growth is dominated by the matter perturbations. Let us express this as $y \gg y_H$, where y_H is the ratio between the scale factor at the moment the mode k enters the horizon to the scale factor at the equality. Thus, y_H is a function of k.

In order to determine the two unknown coefficients C_1 and C_2, we need to match this solution to our previous solution for modes crossing the horizon well within the radiation era. For these modes, $y_H \ll y \ll 1$. In order to match the two solutions, we need to guarantee that the solutions are equal as well as their first derivatives :

$$A\Phi_p \ln(By_m/y_H) = C_1 D_1(y_m) + C_2 D_2(y_m)$$
$$\text{and}$$
$$\frac{A\Phi_p}{y_m} = C_1 D_1'(y_m) + C_2 D_2'(y_m) \ ,$$

where y_m satisfies the condition $y_H \ll y_m \ll 1$. It is important to note that the Meszaros equations do not depend on k, i.e., all modes evolve identically on the linear regime. It is also important to note that the solution for the Meszaros equation can not be extrapolated for late times, since at late times the energy budget of the Universe is dominated by dark energy. In order to find a solution valid at late times we need to generalize the Meszaros equation taking into account dark energy.

The $y \gg 1$ limit of the Meszaros equation, along with the redefinition of the coefficient multiplying δ, $4\pi G\rho_{dm} = (3/2)H_0^2\Omega_m a^{-3}$, and the continuity equation, lead to :

$$\frac{d^2\delta}{da^2} + \left(\frac{d\ln(H)}{da} + \frac{3}{a}\right)\frac{d\delta}{da} - \frac{3\Omega_m H_0^2}{2a^5 H^2}\delta = 0. \tag{1.62}$$

This equation is now written in terms of a instead of y.

Eq. 1.61 shows that density fluctuations grow as the Universe expands – i.e., initially overdense regions become denser, and underdense regions tend to become even more empty of matter. This *growth of structures* is the result of gravity

attracting matter towards the initial density peaks, and causing the Universe to become increasingly inhomogeneous. Moreover, it also generates peculiar velocities as matter clumps start to fall into the gravitational potential wells of the overdense regions.

1.4 The cosmological constant problem

The classical action for gravity, the Einstein-Hilbert action, is defined up to an additive constant term as

$$S = \frac{1}{2\kappa}\int \mathrm{d}^4x\sqrt{-g}\,(R - 2\Lambda_B) + S_{\text{matter}}\ [g_{\mu\nu}, \Psi]\,, \tag{1.63}$$

where $\kappa \equiv 8\pi G/c^4$. The first term (inside parenthesis) is the geometrical part of the action, where we add the *bare* cosmological constant Λ_B. The bare cosmological constant is simply a free parameter of the action, which is to be constrained by data. The second term is the matter action and Ψ is a generic field.

Variation of the action 1.63 w.r.t. the metric yields the Einstein's equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda_{\mathrm{B}}g_{\mu\nu} = \kappa T_{\mu\nu}, \tag{1.64}$$

where the stress-energy tensor is defined by

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}. \tag{1.65}$$

Before the observational fact that the Universe is passing through a phase of accelerated expansion Riess et al. 1998, it was thought that the cosmological constant would vanish, i.e., that some mechanism would cancel out all possible contributions to it. After ibid., within the framework of general relativity, the accelerated expansion requires that the Einstein's equations have a constant term, which is interpreted as an energy component of negative pressure.

But what would be the interpretation of this cosmological constant, and is there any reason to expect that it is there? In the way we expressed the Einstein's equations 1.64, the bare cosmological constant is simply a property of space-time, or an intrinsic curvature of it, i.e., in the absence of a stress-energy tensor, the space-time is not Minkowski, but rather de Sitter space-time. This interpretation is rather misterious since it is not clear the physical origins of this intrinsic property of space-

time. It seems clear that we need a theory of quantum gravity to understand what the cosmological constant at the left hand-side would mean, since the framework of general relativity is merely classical and therefore it is probably only an effective theory of gravity which emerges from fundamental quantum gravity theory [13].

The leading candidate [14] for a quantum theory of gravity is string-theory, which provides a possible mechanism to explain the cosmological constant : different ways of compactifying extra-dimensions would be associated to different false vacua Susskind 2003. This number would be of order of $\simeq 10^{100}$. Therefore, different Universes have different false vacua, or cosmological constants and we ended up living in the one we observe for anthropic reasons. As already mentioned, this explanation is highly speculative and controversial amongst cosmologists.

If we understand the cosmological as a constant term in the stress-energy tensor, or putting it in the right hand side of Einstein's equations 1.64, then we are assuming that there is a matter/energy field which has constant energy through out cosmic expansion. This is what we man by "dark energy". There is indeed a reason to believe that it must be a constant contribution to the stress-energy tensor : the vacuum state of the stress-energy tensor does not vanish and therefore, it must appear as a constant in the right-hand side of Einstein's equations.

Consider, for instance, a scalar field Φ, the corresponding action is

$$S_\Phi = -\int d^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi + V(\Phi)\right]. \tag{1.66}$$

The corresponding stress-energy tensor reads

$$T_{\mu\nu} = \partial_\mu\Phi\partial_\nu\Phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\Phi\partial_\beta\Phi + V(\Phi)\right]. \tag{1.67}$$

The above stress-energy tensor is minimized when the kinetic energy vanishes and the potential is taken to sit at its minimum, i.e.

$$\langle T_{\mu\nu}\rangle = -V(\Phi_{min})g_{\mu\nu} \equiv \rho_{vac}g_{\mu\nu}. \tag{1.68}$$

[13] Some argue that gravitational is intrinsically classical. However, there are clearly regimes in which gravity is relevant at the quantum level, namely, in the singularities of black holes and at the big-bang. Therefore, a quantum description of gravity seems to be mandatory in order to understand these events.

[14] By "leading candidate" I mean the most popular candidate in literature. I completely acknowledge the existence of other candidates for a quantum gravity theory, such as loop-quantum gravity. However, up to my knowledge, string-theory is the only candidate which offers a possible solution for the cosmological constant problem.

We have used the brackets to indicate the zero-point expectation value.
We can rewrite the Einstein's equations as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda_B g_{\mu\nu} = \kappa T_{\mu\nu} + \kappa\langle T_{\mu\nu}\rangle. \tag{1.69}$$

Therefore, we can redefine the cosmological constant as

$$\Lambda_{eff} \equiv \Lambda_B + \kappa\rho_{vac}, \tag{1.70}$$

which is the quantity we have access in observational cosmology.

In the case in which ρ_{vac} is the only source of cosmological constant (as it should be according to the standard model), the prediction for ρ_{vac} has to be contributions from zero-point energy of all possible fields in the Universe. The ideal situation would be to have this prediction which is consistent with observations.

In our example, we are dealing with a classical field. The so-called "classical cosmological constant problem" refers to the disagreement between the prediction for the induced cosmological constant from the eletroweak (EW) and quantum chromo dynamics (QCD) phase transitions (under the assumption of a mass for the Higgs field) and the inferred value from cosmological observations. The inferred value from cosmology is the order of the critical density today, $\rho_\Lambda = \Omega_\Lambda\rho_{crit}$. The predicted value is (see Martin 2012 for details)

$$\rho_{vac}^{EW} \simeq -10^{55}\rho_{crit} \tag{1.71}$$

and

$$\rho_{vac}^{QCD} \simeq 10^{45}\rho_{crit}. \tag{1.72}$$

Even if this discrepancy is solved at the classical level, it rises also when we estimate the contribution from the zero-point quantum fluctuations. The latter is called "The quantum-mechanical cosmological constant problem".

The quantum contribution from a generic quantized field is given by integral of the zero-point energy of a quantum harmonic oscilator :

$$\Lambda_{eff} = \Lambda_B + \frac{\kappa}{(2\pi)^3}\int d\boldsymbol{k}\frac{1}{2}\omega^2(k). \tag{1.73}$$

The obvious problem with this contribution is that the integral blows-up in the UV. In order to extract the finite part from it, we can use dimensional regularization,

which leads to

$$\rho^{Q}_{vac} = \frac{m^4}{64\pi^2} \ln\left(\frac{m^2}{\mu^2}\right), \tag{1.74}$$

where m is the mass of the field and μ is an energy scale introduced in the regularization to fix the dimensionality.

in order to obtain the right prediction, we need to take into account the contributions from all quantum fields in the Universe. The total contribution from zero-point quantum fluctuations reads

$$\rho^{Q}_{vac} = \sum_i n_i \frac{m_i^4}{64\pi} \ln\left(\frac{m_i^2}{\mu^2}\right). \tag{1.75}$$

Taking $n_H = 1$, $m_H \simeq 125\text{GeV}$ for the Higgs, $n_{quarks} = -4$, $m_t \simeq 171.2\text{GeV}$, $m_b \simeq 4.2\text{GeV}$, $m_c \simeq 1.27\text{GeV}$, $m_s \simeq 0.104\text{GeV}$, $m_\mu \simeq 0.24\text{GeV}$ and $m_d \simeq 0.48\text{GeV}$ for the quarks, $n_{leptons} = 4$, $m_e \simeq 0.511\text{MeV}$, $m_\mu \simeq 105\text{MeV}$, $m_\tau \simeq 1.77\text{MeV}$, $n_z = 3$, $m_z \simeq 91\text{GeV}$, $n_{W^\pm} = 3$ and $m_{W^\pm} = 80\text{GeV}$ for the gauge bosons, neglecting the neutrinos and taking $\mu \simeq 3 \times 10^{-25}\text{GeV}$[15] leads to

$$\rho^{Q}_{vac} \simeq -2 \times 10^{8}\text{GeV}^4 \simeq -10^{55}\rho_c. \tag{1.76}$$

Therefore, the quantum contribution presents a discreapancy of ~ 55 orders of magnitude w.r.t. to the inferred vacum energy from cosmology. Clearly, none of the choices made in this prediction will solve this problem, unless we consider new physics. We are ignoring the interactions between the fields, which would worsen the problem.

The cosmological constant problem is a multifaceted problem, which emerges in both classical and quantum level. Hence, it is hard to imagine any avenue to solve it, since even if we come up with a new theory of gravity which is capable of correctly describe how quantum fluctuations interact with gravity suppressing the quantum contribution, then we still have to explain the classical contribution. The other way around is also true.

The observations of cosmic expansion seem to be in highly solid ground at the moment this theses is being written and it seems unlikely that the solution would

[15] Here we follow Koksma et al. 2011, which argues that we should take $\mu \sim \sqrt{E_{grav}E_\gamma}$, where $E_{grav} \simeq H_0 \simeq 3.7\times10^{-41}\text{GeV}$ and E_γ is the energy of photons corresponding to the wavelength $\lambda \simeq 500\text{nm}$. The reasoning behind this is that we measure the cosmological constant through the photons from supernovae, which couples to the expansion rate.

come from new insights given by cosmological data.

However, it is safe to say that the standard model of particle physics is much more well tested than the standard model of cosmology. Therefore, the current and up-coming cosmological data carry a hope of finding new breakthroughs in fundamental physics.

I believe that the cosmological constant problem is the most exciting problem in our current knowledge of fundamental physics. Any advancement in the comprehension of this problem can deeply impact cosmology and particle physics.

By measuring how photons are deviated by underdense structures, we are measuring how photons interact with gravitational potential in environments where the vacuum energy is more relevant than dark matter compared to an average environment in LSS. Therefore, the work developed in this thesis has as ultimate motivation the contribution for the development of an observable which might be sensitive to aspects of the cosmological constant problem.

2 Void Phenomenology

2.1 Introduction

The void science goes back to the 70's, where some authors detected large regions of radius $\simeq 20h^{-1}$Mpc almost empty of galaxies in the nearby galaxy distribution (e.g. Gregory et al. 1978). But after the discovery of a void of radius $\simeq 60h^{-1}$Mpc in the Bootes constellation Kirshner et al. 1981, voids started to receive more attention in literature and raised the question of whether this kind of structure is a common feature of the large-scale structures or exceptions. Since then galaxy surveys have shown that voids are so common that they are responsible for a large fraction of the volume of the Universe and, therefore, is an essential feature of large-scale structures. The existence of voids is a direct consequence of the initial conditions and the gravitational collapse, which is the main driver of structure formation. As matter accretes into overdense fluctuation given by initial conditions, underdense fluctuations expand and occupy then larger and larger volumes.

Since voids represent a large fraction of the volume of the Universe, posses their own substructures (e.g Jaber et al. 2023, Weygaert 2014), are emptier of baryonic feedback and are less non-linear (e.g Paillas et al. 2017) than their overdense counterparts, they emerge as a prolific laboratory for cosmological studies.

In the last two decades voids have been extensively used to constraint cosmological parameters (e.g Sutter et al. 2012, Nico Hamaus, Cousinou et al. 2017), as well as they have shown their potential as discriminators of modified gravity models (Voivodic, Lima, Llinares et David F Mota 2017a, Perico et al. 2019, Contarini, Marulli et al. 2021, Baker et al. 2018), the sum of neutrino masses Massara, Villaescusa-Navarro et al. 2015 and dynamical dark energy models Verza et al. 2019.

Despite the vast literature concerning voids, there are still challenges to overcome in order to achieve the whole potential they provide. The main and most obvious problem is the cosmic variance : since voids are large structures, spanning from

few h^{-1}Mpc up to $\simeq 100h^{-1}$Mpc, we always have a modest number of them on galaxy surveys (usually a few thousands) and, as a consequence, the precision of the statistics we estimate from these samples is always largely limited by cosmic variance. Unfortunately, the only solution for this problem is the mapping of the largest possible volume of the universe by galaxy surveys, being then a fundamental limitation. A possible way of easing this problem is the identification of the smallest possible voids[1] in the Dark Matter distribution, but it is also a huge problem since voids are, by definition, emptier of galaxies, which means that the sparsity of the galaxy field strongly affects the resulting void catalogue, which means that a number of voids produced by a certain void-finder algorithm are simply useless since they are the result of Poisson noise, specially the smaller ones. One possible avenue to circumvent the latter problem is to use the weak-lensing signature of voids, since this signature is sensitive to the total matter field and therefore allows us to "see" this field without the limitation imposed by galaxy shot-noise. As we aim to show in this work, we believe we have advancements in this program.

Another challenge concerns the most obvious statistics we can measure from voids : the two-point cross-correlation function between voids and tracers (galaxies or Dark Matter). See Massara et Sheth 2018 for some effort towards this goal. Unlike the matter-matter auto-correlation, the void-matter cross-correlation is not trivially predicted by standard perturbation theory (and its extensions such as effective field theory of large scale structure). This happens because voids are not a common tracer of LSS, but rather it has an internal structure with radius of order of h^{-1}Mpc, which gives a large stochastic contribution to the void profile[2]. Moreover, the void-finder recipe, which is not unique, will highly affect the shape of this cross-correlation and it is far from clear how to include this dependence in an analytical prescription. One possible avenue to analytically model this cross-correlation is to understand voids as any other tracer of large-scale structure at the 2-Void term and simply write the void-matter cross-correlation as an expansion of density field operators and corresponding bias parameters (see Desjacques et al.

[1] Here I'm assuming a specific definition of what is a void which will be later introduced. Briefly, a void is an expanding structure which experienced shell crossing, since the internal shells expand faster than outer shells. Having this definition in mind, we can talk about the smallest structures in the DM field which satisfy this criteria.

[2] Since the void radius is defined at the moment of shell crossing, the internal void structure cannot be predicted using perturbative methods.

2018 for a review in the context of other tracers) :

$$\delta_{\rm v} = \mathcal{F}\left[\Phi, \Phi_v\right], \tag{2.1}$$

where the functional $\mathcal{F}$ is an expansion of the potential Φ, the velocity potential Φ_v and their derivatives. This idea was studied in Chan, Nico Hamaus et al. 2014 and it was shown that the void profile in its 2V term has contributions from higher order operators, beyond the linear prescription.

Despite the theoretical and observational challenges, we believe this is an exciting field to work on, where several plausible advancements are to be done. This theses is mainly about small steps towards these advancements. In this chapter we aim to summarize the phenomenology concerning voids which has been developed in the literature in the last decades, as well as to present our contribution to void science by presenting a new void finder algorithm and its application in the context of void-lensing.

2.2 The excursion set for Halo formation

Voids can be seen as one of the building blocks of LSSs along with halos, by thinking of walls and filaments as being composed by halos. The usefulness of this way of understanding LSS is that we can actually model the LSSs from these building blocks and come up with predictions for the statistics of it Cooray et al. 2002. Moreover, we can make predictions for the counts of such objects per bin of mass, the so-called halo-mass/void-size function (or void abundance) Sheth et Van De Weygaert 2004. These observables are of central importance in modern cosmology, since they carry valuable cosmological information, in particular regarding the primordial non-gaussianities.

We can roughly understand halos as being virialized objects which can be treated as an effective particle (forming the LSSs) of a given mass and then count the number of these objects per bin of mass - the halo mass function, as well as calculate their cross-correlations. It is worth mention that the precise definition of what is a halo depends on the halo finder algorithm and the parameters in the predictions may vary depending on the algorithm's peculiarities. The halo mass function plays an important role in a variety of astrophysical and cosmological applications, such as, for instance, galaxy formation Risa H Wechsler et Tinker 2018, Halo

Occupation Distribution models (HODs) Berlind et al. 2002 and the prediction of linear halo bias Tinker et al. 2010.

In 1974 Press and Schechter Press et al. 1974 (PS) developed the first (up to the knowledge of the present author) theory for the counts of what they called "clumps" particles in a "gas" which are bounded after having virialized through the spherical collapse. In the modern jargon these "clumps" are known as halos «of DM». After PS, several authors have further developed the theory. We aim to cover from the PS theory until the so-called excursion set theory, first introduced in the classical work of Bond et al. 1991.

In this section we review the basic theory for the halo mass function (or halo abundance).

2.2.1 The spherical collapse

Before proceeding with the first approach to count halos, we shall briefly review the basic assumption of the collapse of a spherical region with uniform density embedded in an Einstein-de Sitter background.

Let's consider then, that a spherical region has initial density contrast Δ_i and radius R_i. Therefore, the enclosed mass is $M = (4\pi/3)R_i^3\bar{\rho}_{m,i}(1+\Delta_i)$, where $\bar{\rho}_{m,i} = 3H_i^2/8\pi G$ is the background density.

Due to the Gauss law, the interior of the spherical region evolves independent of the background and the equation of motion for its radius is

$$\ddot{R} = -\frac{GM}{R^2}. \tag{2.2}$$

By integrating the above equation once we get :

$$\dot{R}^2 = \frac{2GM}{R} - E, \tag{2.3}$$

where E is the integration constant, which is interpreted as the energy of the system.

For $\Delta_i << 1$, we can use the linear growth of peculiar velocities $\dot{R}_i^{pec} = -R_i H_i \Delta_i/3$, where the growth rate is $f = 1$, since $\Omega_m = 1$. The total velocity is the sum of the peculiar velocity and the Hubble flow :

$$\dot{R}_i = \dot{R}_i^{pec} + H_i R_i = H_i R_i (1 - \frac{\Delta_i}{3}). \tag{2.4}$$

That is, the sphere initially expands a little bit slower than the background. We expect that this expansion will turn around at some point, as Δ grows to non-linear values.

By substituting M and $\dot{R}$, the total energy is :

$$E = -\frac{5}{3}\frac{(H_i R_i)^2}{2}\Delta_i \tag{2.5}$$

and by setting $\dot{R} = 0$ we get the turn around radius :

$$R_{\rm ta} = \frac{3}{5}\left(\frac{1+\Delta_i}{\Delta_i}\right) R_i. \tag{2.6}$$

We can then write the cycloid solution of equation 2.2 as

$$R = \frac{R_{\rm ta}}{2}(1-\cos\theta),\ t = \frac{t_{\rm ta}}{\pi}(\theta - \sin\theta), \tag{2.7}$$

where $\theta \in [0, 2\pi]$, with the turn around occurring at $\theta = \pi$.

This parametric solution implies that the radius will eventually collapse, but actually this solution is only valid until the shell crossing. After this point the halo will virialize.

We can write the evolution of the spherical overdensity for all t using the parametric solutions 2.7 as :

$$1+\Delta_i = \frac{\rho(<R)}{\bar{\rho}} = \frac{M/(4\pi/3R^3)}{\bar{\rho}} = \frac{\bar{\rho}_{m,i}(1+\Delta_i)R_i^3}{\bar{\rho}_m R^3} = \frac{9}{2}\frac{(\theta-\sin(\theta))^2}{(1-\cos\theta)^3} \tag{2.8}$$

By assuming that at $\theta = 2\pi$ the virialization is complete, we can estimate the linear density associated to the virialization as :

$$\delta_{\rm cr} \equiv \Delta_L\left(t_{\rm vir}\right) = \Delta_i\left(\frac{t_{\rm vir}}{t_i}\right)^{2/3} = \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}\left(\theta_{\rm vir}-\sin\theta_{\rm vir}\right)^{2/3} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686. \tag{2.9}$$

That is, $\delta_{cr} \simeq 1.686$ is the initial overdensity linearly extrapolated at the time of complete virialization. In the following, we are going to use it as a "clock" to define a virialized object, i.e., it will be used as the linearly extrapolated overdensity of a halo and thus the fraction of objects which are virialized will be the ones having their initial overdensities above δ_{cr} when linearly extrapolated.

This is a rough approximation, specially for smaller halos, which are subjected

to strong tidal forces and therefore might need a higher overdensity to virialize. However, it should be a reasonable approximation for the most massive halos, which are better approximated as being isolated spherical overdensities.

2.2.2 The Press-Schechter argument

The Press-Schechter formalism, as well as the excursion set theory, have few basic assumptions, namely :

- The smoothed density contrast $\delta(R)$ in a scale R follows a Gaussian distribution.
- The collapse of DM halos is approximately spherical
- Protohalo patches are characterized by a linearly extrapolated overdensity equal to a certain threshold δ_c (given by the spherical collapse).

The second conditions means that the PS theory will search for proto-halos in the Initial Conditions (ICs), i.e in the initial Gaussian field. Regions in this ICs which surpass the threshold $\delta_c \simeq 1.686$ (for Einstein-de Sitter) will be destined to virialize into halos. Conversely, the DM inside halos at the present time can be traced back into the ICs and therein this DM content will be inside a region above the threshold. In the light of modern N-body simulations we know that these assumptions are only valid for larger halo masses. The less massive halos are formed by physical processes, such as tidal fields, for which the spherical collapse is not a reasonable approximation Desjacques et al. 2018.

The PS can be summarized as the following. For sufficient small δ, which is a reasonable assumption in the initial conditions, we can write its distribution as being Gaussian :

$$\Pi_{\rm PS}(\delta, S) = \frac{1}{\sqrt{2\pi S}} e^{-\delta^2/(2S)}, \tag{2.10}$$

where

$$S(R) \equiv \sigma^2(R) = \left\langle \delta^2(x, R) \right\rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_L(k) W_R^2(k) \tag{2.11}$$

is the field variance and $\delta(x, R)$ the density field smoothed at scale R by the window function $W_R(k)$. We are going to adopt a top-hat window function for the sake of

simplicity. With this choice, the relation between the smoothing scale and the mass is trivially $M = \bar{\rho}_m 4/3\pi R^3$.

Therefore, for a certain smoothing scale, the volume fraction which will be part of virialized objects is

$$F_{\rm PS}(>M) = \int_{\delta_c}^{\infty} d\delta\, \Pi_{\rm PS}(\delta, S(R)) = \frac{1}{2}{\rm erfc}\left(\frac{\nu(R)}{\sqrt{2}}\right), \tag{2.12}$$

where $\nu(R) = \delta_c/\sigma(R)$ and erfc(x) is the complementary error function. Notice that we use the notation $F_{PS}(>M)$ to make explicit that this volume fraction contain objects with mass *equal or greater than M(R)*. It is easy to see why : if a certain smoothed region has density contrast equal to the density threshold δ_c, then it will have a mass $M_\star$, associated to a scale $R_\star$, typically the size of the corresponding smoothing scale. But it turns out that all the regions encompassing mass $> M_\star$ ($> R_\star$) will also be accounted.

By defining $\bar{n}_h$, such as

$$F_{PS}(>M) = \frac{1}{\bar{\rho}_m}\int_M^{\infty} d{\rm ln}M' M' \bar{n}_h(M') \tag{2.13}$$

then we can convert 2.13 into the number density of halos per logarithmic mass interval as

$$\bar{n}_h(M) \equiv \frac{d^2\bar{N}_h}{dV\, d\ln M} = \bar{\rho}_m f(M) = -\bar{\rho}_m \frac{dF(>M)}{dM}, \tag{2.14}$$

where we have defined

$$f(M) \equiv -\frac{dF(>M)}{dM}. \tag{2.15}$$

However, this result is not properly normalized, since

$$\int_0^{\infty} d\ln M M \bar{n}_h(M) = -\bar{\rho}_m \int_0^{\infty} dM \frac{dF}{dM} = -\bar{\rho}_m \left[p_G(R=\infty) - p_G(R=0)\right] = \frac{1}{2}\bar{\rho}_m \tag{2.16}$$

should be the total density found in virialized objects of all masses, or the mean density of the Universe, given the assumptions of the PS framework. This missing mass is known as the *cloud-in-cloud problem* (as it will soon be clear why) and it was realised in the original PS paper Press et al. 1974. However, they simply include

an *ad hoc* factor of 2 in their result, heuristically justifying with the argument that this "fudge" factor of two is initially into underdensities and will eventually collapse into halos. By introducing this factor of two, we finally obtain the PS halo mass function

$$\bar{n}_{h,PS}(M) = -2\bar{\rho}_m \frac{dF}{dM} = \frac{\bar{\rho}}{M}\sqrt{\frac{2}{\pi}}\nu \exp\left[-\nu^2/2\right]\frac{d\ln\sigma^{-1}}{d\ln M}. \tag{2.17}$$

where $\bar{\rho}_m$ is the average density of the Universe and we have converted radius to mass. This relation is simply $M = 4/3\pi R^3\bar{\rho}_m$ for a top-hat smoothing function. In another words, the number density of objects of a certain mass M can be predicted through the fraction of mass belonging to future virialized objects, which is simply given by the integral 2.12, with the correct normalization $\bar{\rho}_m/M$. The result 2.17 can be extrapolated to any value of redshift by linearly extrapolating the density threshold $\delta_c(z) = \delta_c(0)/D(z)$, where $D(z)$ is the linear growth function.

Notice that $F_{PS}(0) = 1/2$, which means that in the limit in which the variance $\sigma(R) \to 0$, the virialized objects account for only half of the Universe! But clearly in this limit all the mass of the Universe must be inside virialized objects. This problem was realized by PS in their original paper and they basically correct this mismatch by multiplying F_{PS} by a factor of two. They give an heuristic justification in the appendix, arguing that the missing mass is in the surrounding underdensities which will eventually collapse onto the virialized objects. This problem is known as the "cloud-in-cloud" problem. In the following we clarify the nature of this problem.

The solution to the cloud-in-cloud problem

The origin of the "fudge"factor of two in the PS formalism is the fact that in the above derivation, we integrate the probability distribution at a fixed smoothing scale R, but there is also the possibility that for some smoothing scale $R' > R$ a region which is below the threshold at R is now above the threshold. In other words, there is a missing contribution from regions which are embedded in a larger region which is above the threshold. The fact that the probability of having these regions is exact the same probability of the accounted regions (those above the threshold at smoothing scale R) becomes clear with the aid of the figure 2.1, reproduced from Bond et al. 1991. In this figure we are using the language of excursion sets, which will be further developed soon. In this language, the field points $\delta(R)$ perform a random walk as a function of the pseudo-time variable S, starting from $S = 0$,

where $\delta(R) = 0$. At the "instant" $S_\star$, the PS prediction comes from the integral of the Gaussian above the threshold. However, there are trajectories which have already crossed the threshold at an earlier "instant", but happen to be below the threshold at $S_\star$. Since the distribution is Gaussian at all S, the probability that a trajectory is above the threshold is the same as the probability that a trajectory has crossed the threshold earlier and is below the threshold at $S_\star$. Therefore, the missing factor of two is linked to trajectories which cross the threshold multiple times.

One of the first attempts to predict the halo mass function including the fudge factor was made by Peacock et al. 1990. In their work they explicitly include the probability of having an underdense region embedded into a large overdense region above the threshold :

$$F_{PH}(> M) = p_G\left(\delta(R) > \delta_c\right) + \int_{-\infty}^{\delta_c} d\delta\, \frac{dp_G}{d\delta} p_{up}\left(\delta\left(r > R\right) > \delta_c; \delta\right), \tag{2.18}$$

where $p_G \equiv F_{PS}$ is the PS contribution and p_{up} is the probability that a smoothed overdensity in a scale $r > R(\sigma(r) < \sigma(R))$ has already crossed the threshold, weighted by the Gaussian distribution of $\delta(R)$. A bit of reflection makes clear the parallel between eq. 2.18 and figure 2.1. The second term in eq 2.18 is the probability that $\delta(R)$ has already crossed the threshold for larger smoothing scale conditioned by the probability that $\delta(R)$ is Gaussian distributed and is below the threshold at R(or $S(R) = \sigma(R)$). This is exactly what the reflected tail in figure 2.1 represents, since the probability that a trajectory has passed the threshold at $r > R$ and is found above or below the threshold at R is the same, the second term in eq. 2.18 must be equal to the first. Notice that the probability p_{up} is properly normalized, since for $R \to 0\,(\sigma(R) \to \infty)$ the probability that any trajectory has crossed the threshold goes to unity $p_{up} \to 1$.

2.2.2.1 The excursion set - intuition

The basic assumption of the excursion set is the following

> *a generic point* $\boldsymbol{x}$ *belongs to a halo of size* R *only if* R *is the largest possible scale at which the smoothed density contrast* $\delta(R, \boldsymbol{x})$ *centered at* $\boldsymbol{x}$ *crosses the threshold* δ_c

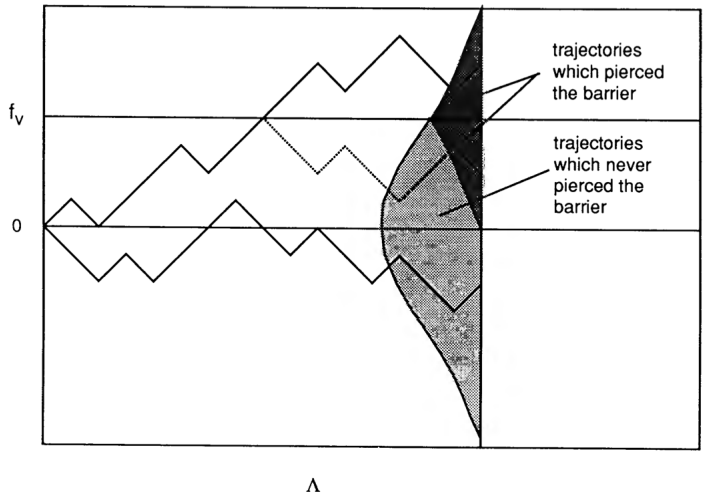

FIGURE 2.1 : Illustrative representation of walks which achieve Λ, which represents here the "time" variable, or the variance of the smoothed density field. The dark shaded region shows that the fraction of trajectories that achieve Λ below and above the threshold of smoothed density contrast f_v is the same, since for every trajectory which pierce the threshold before Λ and reaches Λ above the threshold, there is an equal probability that a reflected trajectory reaches Λ below the threshold. Figure extracted from BOND et al. 1991.

Therefore, we are interested in obtaining the probability that a generic point belongs to a virialized object of size R, which is the largest possible size above the threshold. Once we are able to calculate this probability, we can say that the number density of halos of size $R(M)$ is proportional to this probability, so we can write :

$$\bar{n}_h(M) \propto \frac{dP_{up}(\boldsymbol{x} \in \delta(R(M)) > \delta_c)}{dM}, \tag{2.19}$$

where P_{up} denotes the probability that a generic point $\boldsymbol{x}$ belongs to a region which is the largest one to reach δ_c. It is important to notice that since $S = \sigma^2(R)$ is a monotonically decreasing function of $R(M)$ and therefore we can use interchangeably the variables M, R and S. The conversion between R and M depends on the adopted smoothing kernel W_R, which reflects the halo density profile. See BOND et al. 1991 for a discussion about different kernel choices.

Given this assumption, it is natural to think in a practical algorithm in which we treat $\delta(R)$ as performing a random walk as a function of smoothing scale R, which

starts from $R \to \infty$ and continuously decreases, or equivalently, starts from $S = 0$ and increases. We are going to make an idealised experiment using this algorithm to illustrate the intuition behind the excursion set and then we will put this idea in an analytical form.

Figure 2.2 (extracted from Desjacques et al. 2018) shows an example of realisations of $\delta(R, \boldsymbol{x})$, i.e the random walk of the smoothed density contrast inside a sphere of radius R centered at two different points of a simulation box. The authors in ibid. chose two points presenting different trends, but statistically a generic point has no preferential direction (that is in the core of the cloud-in-cloud problem), but rather the value of the density contrast in a scale R is simply a Gaussian with increasing variance around zero. The algorithm will then work as the following : take a point $\boldsymbol{x}$ and calculate the density contrast around it at the largest possible R, then reduce R as continuously as possible and keep the scale $R_\star$ at which $\delta(R_\star)$ first touches the threshold δ_c. Repeat the process for all points. Then we will have a histogram of number of points that first crossed δ_c at each scale, which is exactly the distribution on the right-hand side of eq. 2.19. Then this distribution can be converted into the number density of halos with corresponding mass $M(R)$ (which depends on the filter kernel).

2.2.2.2 The excursion set - analytical solution for a Markovian walk

Since the excursion set models $\delta(R)$ as a random walking variable, the problem reduces to a diffusion problem, for which we can write a Langevin equation :

$$\frac{d\delta(R)}{dR} = \mathcal{Q}(R), \tag{2.20}$$

where the noise term has covariance

$$\langle \mathcal{Q}(R_1)\mathcal{Q}(R_2)\rangle = \int \frac{d^3k}{(2\pi)^3} P_L(k) \frac{dW(R_1)}{dR} \frac{dW(R_2)}{dR} \tag{2.21}$$

and we have used eq. 2.11.

Notice, however that for a general choice of filter kernel W, this diffusion problem has a noise term which is correlated throughout the random walk, i.e the value the field will assume for a certain pseudo-time S depends on the history $s < S$, where s is a dummy variable.

Therefore, to avoid this complication we can make a suitable choice of the filter

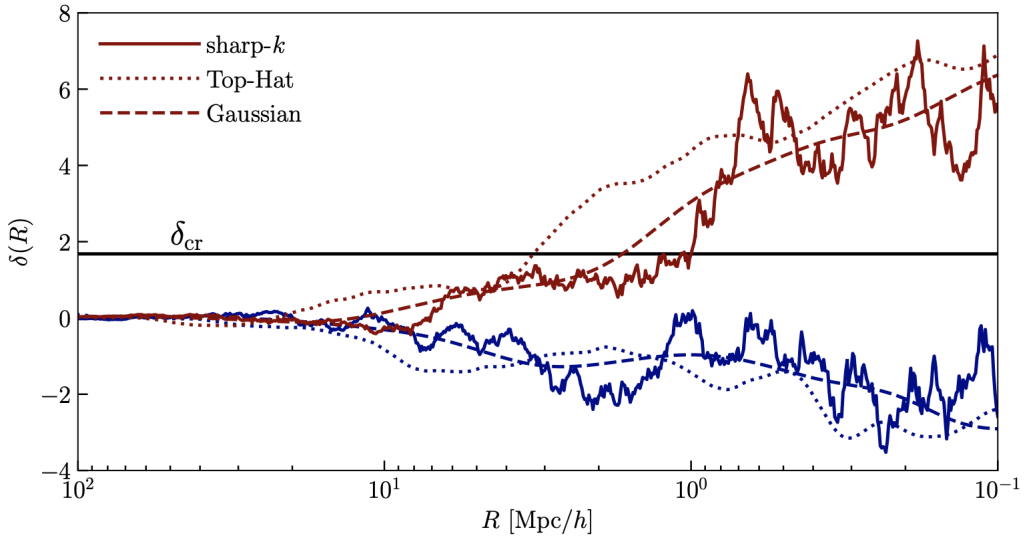

FIGURE 2.2 : The random walk performed by the smoothed density contrast $\delta(R)$ using three different smoothing kernels. The red and blue lines represent randoms walks associated, respectively, to a sphere centered in a overdensity, which crosses the threshold of halo formation, and to a sphere centered in an underdensity, which never crossed the threshold of halo formation. Extracted from DESJACQUES et al. 2018.

kernel for which the covariance 2.21 will be a Dirac delta noise. This choice is clearly a sharp-k filter

$$W(R) = \Theta_H(1 - kR), \tag{2.22}$$

with Θ_H being the Heaviside function. With this choice, the covariance 2.21 becomes

$$\langle \mathcal{Q}(R_1)\, \mathcal{Q}(R_2) \rangle = - \left. \frac{k^2 P_{\rm L}(k)}{2\pi^2} \right|_{k=\frac{1}{R_1}} \frac{1}{R_1^2} \delta_D(R_1 - R_2) \tag{2.23}$$

and $\delta(R)$ can be understood as performing a Markovian walk, since in 2.23 the different steps S(R) are uncorrelated . That is, when we increment the pseudo-time variable from s to $s+ds$, the difference $d\delta = \delta(s+ds) - \delta(s)$ is Gaussian distributed and, since the sum of independent Gaussian distributed variable is also a Gaussian distributed variable, we are allowed to write the probability distribution function of the walking field δ' arrive at S with the value $\delta'(S) = \delta$ as :

$$p_G(\delta)d\delta = \frac{1}{\sqrt{2\pi S}} e^{-\delta^2/2S} d\delta \tag{2.24}$$

Then, following our discussion in the last section, we are interested in the probability that a trajectory $\delta(S)$ first crosses the threshold δ_c. A natural way of doing that is to find the probability that a field $\delta(S)$ has never reached the threshold for any $s < S(R_0)$, and acquires the value $\delta = \delta(R_0)$, $\Pi(\delta; S(R_0))$. Then, the probability that a halo of mass greater than $M(R_0)$ will form due to a first crossing trajectory is

$$F(> M) = 1 - \int_{-\infty}^{\delta_c} d\delta \, \Pi(\delta; S), \tag{2.25}$$

which is exactly the probability we were looking for when we wrote 2.19. The mass fraction inside halos of mass inside the bin $[M, M + dM]$ is then

$$f(M)dM = -\frac{dF(> M)}{dM} dM, \tag{2.26}$$

and we can finally convert into the number density of halos of mass inside the bin $[M, M + dM]$ as

$$\bar{n}_h(M) = \bar{\rho}_m f(M). \tag{2.27}$$

Our goal is now to find the probability distribution $\Pi(\delta, S)$ given the assumption that the filter kernel is k-sharp, allowing us to write the probability that the field acquires the value δ at S as a Gaussian distribution. We could use the knowledge that when a random variable $\delta(R)$ is described by a Langevin equation with a Dirac delta noise, then its probability distribution is described by a Fokker-Planck equation with appropriate boundary condition Bond et al. 1991 :

$$\frac{\partial \Pi}{\partial S} = \frac{1}{2} \frac{\partial^2 \Pi}{\partial \delta^2}, \tag{2.28}$$

where the appropriate boundary condition is $\Pi(\delta_c, S) = 0$. The solution of the FP equation is the distribution we are looking for.

However, here we will adopt a more intuitive approach, which will gives this distribution through simple observations as the following.

The distribution 2.24 is the probability that the field assumes any value of δ at the "instant" S. To exclude trajectories that crossed the threshold at any $s < S$,

we can simply state that $\Pi(\delta, S) = 0$ for $\delta > \delta_c$. We also need to consider the probability of δ to be below the threshold δ_c after crossing the threshold at any $s < S$. Since the probability distribution is Gaussian for all s and independent amongst them, once the trajectory crosses the threshold, the probability that the trajectory will reach a value of $\delta_c + (\delta_c - \delta) = 2\delta_c - \delta$ above the threshold, is the same that it will perform exactly the same trajectory, but reflected around the horizontal line $\delta = \delta_c$ and reach $\delta_c - (\delta_c - \delta)$. Hence, we can simply subtract the probability $p_G(2\delta_c - \delta)$, yielding

$$\Pi(\delta; S) = \begin{cases} \frac{1}{\sqrt{2\pi S}} \left[e^{-\delta^2/2S} - e^{-(2\delta_c - \delta)^2/2S} \right] & (\delta < \delta_c) \\ 0 & (\delta \geq \delta_c) \end{cases} . \tag{2.29}$$

Finally, we can use eqs. 2.25 - 2.27 to estimate the halo mass function :

$$F_{ES}(> M) = 1 - \int_{-\infty}^{\delta_c} d\delta \, \Pi(\delta; S) = \mathrm{erfc} \left[\frac{\delta_c}{\sqrt{2S}} \right] \tag{2.30}$$

and

$$\bar{n}_{h(\mathrm{ES})}(M) = \frac{\bar{\rho}_m}{M} f_{\mathrm{ES}}(\nu) \frac{d\ln\sigma^{-1}}{d\ln \mathrm{M}}, \tag{2.31}$$

where we have defined the *multiplicity function* f_{ES} :

$$f_{\mathrm{ES}}(\nu) = \sqrt{\frac{2}{\pi}} \nu \exp\left\{ [-\nu^2/2] \right\}. \tag{2.32}$$

Notice that this result is exactly equal the PS prediction 2.17, but naturally corrected by the "fudge" factor of two. It is straightforward to check that eq. 2.28 with the given boundary condition leads to the same probability distribution 2.29. Therefore, by applying the boundary condition of an absorbing barrier we are excluding all the possible paths which cross the threshold at any $s < S(R_0)$. This intuition will be useful in order to extend the formalism to voids.

2.2.2.3 Excursion-set with a linear diffusing barrier

It is worth noticing that this result is useful in order to give an intuition for the excursion set, but it has several simplifications, namely, that the collapse is spherical, that the void density profile is a sharp-k profile (which is unphysical), that the random walk $\delta(S)$ is Markovian and that the threshold is a constant.

Despite all the simplifications, this result can qualitatively reproduce halo-mass functions in N-body simulations, but there are discrepancies at the quantitative level (see fig. 1 in the paper I of Maggiore et al. 2010). In order to improve the agreement with N-body simulations, ibid. developed an extension of the excursion set (as done in Bond et al. 1991) for non-Markovian deviations (Maggiore et al. 2010, paper I). This result still has discrepancies w.r.t N-body simulation measurements. The reasons for these discrepancies are, beyond the already cited ones, the fact that halos are not spherical in general, but rather triaxial and most importantly, that there are arbitrary choices in the definition of what is a void in the void finder algorithm. The algorithm's details will introduce variations in the resulting mass function which is not captured by these predictions. Furthermore, the density threshold is likely to not be constant for any environment. Consider, for instance, an environment in which there are strong tidal forces (e.g a filament nearby a massive cluster). A halo forming in this environment will be "stripped" by the tidal field, which will work in the sense of preventing the halo formation. Therefore, it is expected that in this environment, the density threshold must be greater than the one in most "favorable" environments, such as in the center of a long wave-length overdensity.

Motivated by these observations, the second paper in the series of ibid. presented the extended excursion set, which introduces an extension of the first crossing problem in the presence of a moving (diffusing) threshold (barrier). In the following we summarize their result applied to the case of a Markovian walk.

Consider a barrier which fluctuates around the static one $B \equiv \delta_c$ and has variance which is linearly related to the variance of the field δ :

$$\Sigma_B \equiv \langle (B - \langle B \rangle)^2 \rangle = \sqrt{D_B}\sigma(R). \tag{2.33}$$

This *Ansatz* is reasonable since the halo formation is better described by a spherical collapse in regions where $\sigma(R)$ is small, so we expect that the larger deviations (in average) will occur in messy environments, scaling with $\sigma(R)$.

Therefore, the probability distribution which satisfies the Fokker-Planck equation will be a joint probability distribution for δ and B, $\Pi(\delta, B, R)$, with $B = \delta_c$ for $\delta = 0$. Assuming that they perform independent walks, the Fokker-Planck equation for the joint probability distribution reads

$$\frac{\partial \Pi}{\partial S} = \frac{1}{2}\frac{\partial^2 \Pi}{\partial \delta^2} + \frac{D_B}{2}\frac{\partial^2 \Pi}{\partial B^2} \tag{2.34}$$

with the initial conditions $B(S=0) = \delta_c$ and $\delta(S=0) = 0$ and the boundary condition $\Pi = 0$ when $\delta = B$, generalizing the boundary condition in the case of static barrier. The solution produces the multiplicity function (Maggiore et al. 2010, paper II)

$$f_{1LDB} = \sqrt{\frac{2}{\pi}}\sqrt{a}\nu \exp\Big\{-a\nu^2/2\Big\}, \tag{2.35}$$

where $a = 1/(1+D_B)$ and the subscript $1LDB$ stands for one linear diffusing barrier.

2.3 The Excursion set - The void abundance

In this section we extend the reasoning described in the last section for the case of voids. We will call by "void abundance" the equivalent of halo mass function for voids, i.e the mean number of voids in a bin of mass. It is also common in literature the term "void radius function" or "void size function" to refer to the same observable.

Despite not being virialized objects, the formation of voids can be modelled in an analogous manner w.r.t the halo formation, but instead of being product of a spherical collapse, voids are «ideally» the result of spherical expansion suffered by underdense regions in the initial conditions. Since the spherical collapse/expansion models the evolution of idealised isolated overdensities/underdensities, it is expected that any excursion set based on this assumption will be an approximation to the real complex patterns of LSS.

On one hand, voids are more suitable for being described by this idealisation than halos. Arguably, voids are formed where the gravitational forces are better described by a divergent-like field, with less presence of tidal fields inducing curl forces which source deviations from spherical collapse/expansion Icke 1984, specially near the void center. Moreover, voids are way bigger than halos, with a typical radius of order of $\sim 10h^{-1}$Mpc and hence they are less susceptible to the complications of smaller scales. Despite this limitation, the spherical expansion is a reasonable starting point for extending the excursion set to voids and gives us a general intuition and a good

qualitative description of the void abundance.

Figure 2.3 (extracted from Sheth et Van De Weygaert 2004) shows an example of a spherical expansion suffered by an initial underdense top-hat profile. The time steps are $a = 0.05, 0.1, 0.2$ and 0.3. This exercise captures essential features of void evolution : underdensities tend to expand (faster than the background), the density in their interiors decreases with the expansion and a ridge is created at the boundaries. The latter is the result of the faster expansion of inner shells w.r.t outer shells, resulting in an eventual "shell crossing". The inner shells are driven by a larger «in absolute value» averaged density contrast, since their interior are emptier, whereas outer shells are sourced by a smaller averaged density contrast, since it is an average of underdense/overdense inner/outer shells. This exercise captures essential features which are present in void evolution.

Nevertheless, all the complications inherent to overdensities will also have their contribution to void formation, since they are not isolated underdensities and their boundaries will be shaped by the complicated overdense environments Weygaert et Kampen 1993. Moreover, voids are far from evolving isolated from each other, but they collide, merge into each other and form new voids from these processes, as well as filaments and walls Dubinski et al. 1993. This merging of voids into each other is known as the *void-in-void* problem. Furthermore, voids might also inhibit longer wave length overdensities, which tends to collapse and erase these voids. This process is known as the *void-in-cloud* problem. There is also the possibility of halo formation inside voids, or *cloud-in-void* halos, but these halos will form despite inhibiting voids and the larger void will not be affected by eventual halos forming in their inside.

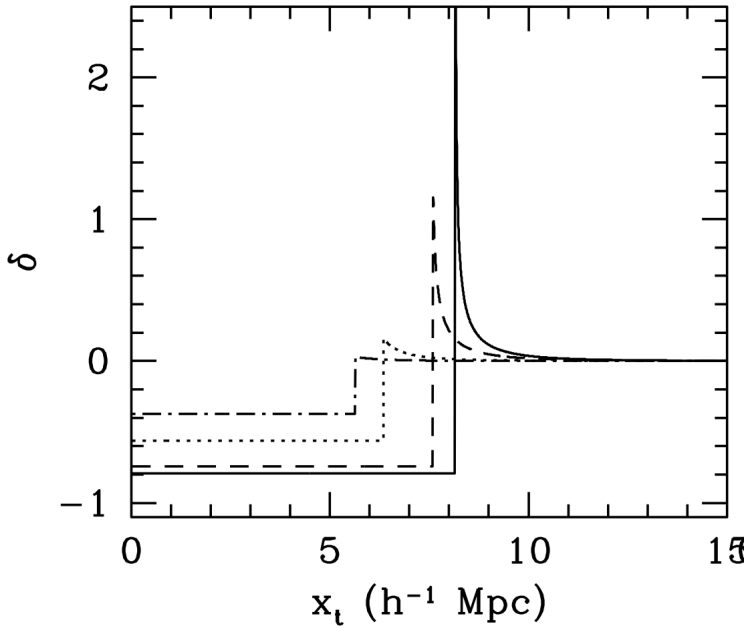

FIGURE 2.3 : The spherical expansion of an initial underdense isolated top-hat profile in four time-steps, $a = 0.05, 0.1, 0., 2$ and 0.3 (dotted-dashed, dotted, dashed and solid lines). Figure extracted from SHETH et VAN DE WEYGAERT 2004.

Therefore, in order to apply the excursion set reasoning for voids we need to worry about the possibility of voids-in-clouds and voids-in-voids. Figure 2.4 (extracted from SHETH et VAN DE WEYGAERT 2004) shows an example of each situation in the language of the excursion set. The left figures are the random walks performed by the linearly extrapolated density contrast in an early time in an N-body simulation, the middle and right plots show two successive time steps in the same simulation at latter times w.r.t the middle ones. The two horizontal dashed lines labeled by δ_c and δ_v are, respectively, the threshold (linearly extrapolated) for halo and void formation, where the value $\delta_v = -2.81$ comes from the spherical expansion of an isolated void which has just passed through the shell-crossing, or has just acquired the average internal density $\rho_v = 0.2\bar{\rho}_m$ (SUTO et al. 1984). In the cloud-in-cloud process the middle and right plots show how smaller overdensities merge to become only one halo. In the cloud-in-void row, it is clear that halos form in despite being inside underdensities and voids are not affected by clouds inside them. The last two rows clearly illustrate that the void-in-void and void-in-cloud process are both important to void formation. In the void-in-void row, existing voids delimited by

some filaments in the mid plot are erased in the right one. In the void-in-cloud row, an existing underdensity in the mid plot is erased by the collapse of the larger overdense region.

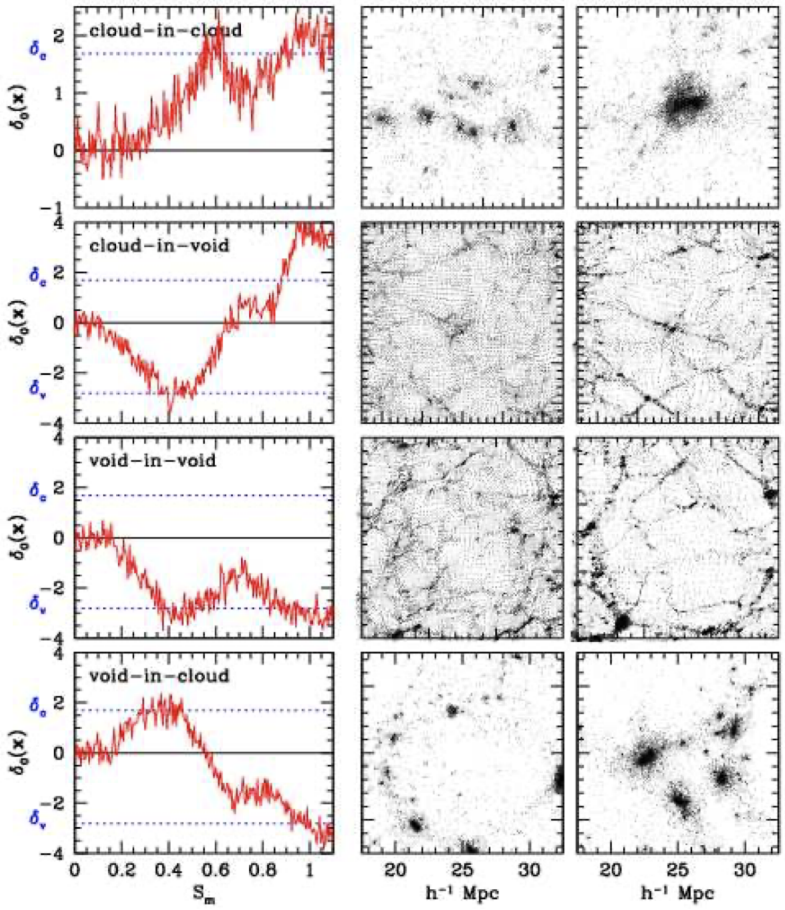

FIGURE 2.4 : Extracted from SHETH et VAN DE WEYGAERT 2004.

This exercise shows that in order to apply the excursion set reasoning to voids, it is necessary to consider the two barriers, δ_c and δ_v. The first will be important to not count voids inside overdensities and the latter will be important to not overcount voids inside voids. That is, the only voids we are going to count are those which first cross δ_v, without ever having crossed δ_c "before".

Given that the excursion set prediction for voids is a two-barrier problem, the *Ansatz* for the probability distribution of first crossing is not as trivial as the one-barrier case. In order to derive the excursion set prediction we are going to make usage of the Fokker-Planck equation and closely follow reference VOIVODIC,

Lima, Llinares et David F Mota 2017a, where is derived the void abundance for two diffuse barriers, with Markovian walks (sharp-k filter).

We begin by considering two diffusing barriers :

$$\begin{aligned} \langle B_c(S)\rangle &= \delta_c + \beta_c S, \\ \langle B_c(S)B_c\left(S'\right)\rangle &= D_c \min\left(S, S'\right), \\ \langle B_v(S)\rangle &= \delta_v + \beta_v S, \\ \langle B_v(S)B_v\left(S'\right)\rangle &= D_v \min\left(S, S'\right), \end{aligned} \tag{2.36}$$

where $B_{c,v}$ is the barrier associated to halos/voids, $\beta_{c,v}$ describes the increase/decrease of the mean barrier along the walk and D_c is the diffusion coefficient and express the variance of the barrier in each S, as well as the correlation between different values of S. The $\min(S, S')$ comes from the choice of a k-sharp smoothing function, or Markovian walk. This means that the field satisfies

$$\langle \delta(S)\delta(S')\rangle = \min(S, S'), \tag{2.37}$$

with $\langle \delta(S)\rangle = 0$. These two equations mean that the random walk is Gaussian and Markovian. Since the field itself is Markovian, we expect that the barrier will also be.

Notice also that in the case in which $\beta_{c(v)}$ is positive, the barrier, which starts at δ_c for $S = 0$ ($R \to \infty$), will have an expected value which scales with the variance. This feature is in accordance with our intuition that smaller scales will be subject to more tidal fields which will act against the collapse and therefore a larger barrier is required.

Since $\delta(S)$ is Markovian, the probability distribution $\Pi(\delta; S)$ satisfies a Fokker-Planck equation

$$\frac{\partial \Pi}{\partial S} = \frac{\partial^2 \Pi}{\partial \delta^2}, \tag{2.38}$$

with the boundary conditions $\Pi(\delta = B_c, S) = \Pi(\delta = B_v, S) = 0$, which is the "absorbing barrier" condition, equivalent to the *Ansatz* we gave in eq. 2.29. The initial condition is given by

$$\Pi(\delta, S = 0) = \delta_D(\delta), \tag{2.39}$$

which means that all the mass of the Universe is contained either in halos or

voids, since

$$\int_{-\infty}^{\infty} d\delta\, \Pi(\delta, S=0) = 1. \tag{2.40}$$

By introducing the variable $Y(S) = B_v(S) - \delta(S)$ and making the assumption that $\beta \equiv \beta_c = \beta_v$ [3], equation 2.37 can be rewritten as

$$\frac{\partial \Pi}{\partial S} = -\beta \frac{\partial \Pi}{\partial Y} + \frac{1+D}{2}\frac{\partial^2 \Pi}{\partial Y^2}, \tag{2.41}$$

where $D \equiv D_c + D_v$. The solution is (see Voivodic, Lima, Llinares et David F Mota 2017a for details) :

$$\begin{aligned} \Pi(Y,S) &= \exp\left[\frac{\beta}{1+D}\left(Y - \frac{\beta S}{2} - \delta_v\right)\right] \\ &\times \sum_{n=1}^{\infty} \frac{2}{\delta_T} \sin\left(\frac{n\pi\delta_v}{\delta_T}\right) \sin\left(\frac{n\pi}{\delta_T} Y\right) \exp\left[-\frac{n^2\pi^2(1+D)}{2\delta_T^2} S\right], \end{aligned} \tag{2.42}$$

where $\delta_T = |\delta_v| + \delta_c$.

The distribution $\Pi(Y,S)$ is the fraction of trajectories which does not cross the "barrier" Y until the "time" S. Then we can proceed as before and calculate the fraction of trajectories which cross the barrier in the interval $[S, S+dS]$ ($[M, M+dM]$) :

$$F_v(>M) = 1 - \int_{-\infty}^{0} dY\, \Pi(Y,S) \tag{2.43}$$

,

$$f_v(M) = -\frac{\partial F_v(>M)}{\partial S}\frac{dS}{dM}. \tag{2.44}$$

and finally

$$\bar{n}_v(M) = \bar{\rho}_m f_v(M). \tag{2.45}$$

Putting everything together yields the number density per logarithm bin of mass :

[3] It is not clear whether this simplification is reasonable or not. If $\beta > 0$, for instance, it means that in smaller scales it is necessary larger overdensities to virialize and smaller underdensities to form voids. This unreasonable assumption can be partially absorbed by the diffusion coefficient D_v though.

$$\bar{n}_v = \frac{\bar{\rho}_m}{M} f_v \frac{d\ln\sigma^{-1}}{d\ln M}, \tag{2.46}$$

where the multiplicity function for voids is given by

$$\begin{aligned} f_v = 2(1+D)\exp\left[-\frac{\beta^2 S}{2(1+D)} + \frac{\beta\delta_v}{(1+D)}\right] \\ \times \sum_{n=1}^{\infty} \frac{n\pi}{\delta_T^2} S \sin\left(\frac{n\pi\delta_v}{\delta_T}\right) \exp\left[-\frac{n^2\pi^2(1+D)}{2\delta_T^2} S\right]. \end{aligned} \tag{2.47}$$

We are going to refer to this solution as 2LDB (Two Linear Diffusing Barriers). By taking the limit of two static barriers, i.e $D = \beta = 0$ we recover the simple case of two static barriers (2SB), first derived by Sheth et Van De Weygaert 2004 :

$$f^{2SB} = 2\sum_{n=1}^{\infty} \frac{n\pi}{\delta_T^2} S \sin\left(\frac{n\pi\delta_v}{\delta_T}\right) \exp\left(-\frac{n^2\pi^2}{2\delta_T^2} S\right). \tag{2.48}$$

It is usual to express the void abundance in terms of the radius instead of the mass. By using $\rho_m/M = 1/V(r)$, then we can rewrite equation 4.20 as

$$\frac{\mathrm{d}\ \bar{n}_{v,L}}{\mathrm{d}\ \ln r_L} = \frac{f_v}{V(r_L)} \frac{\mathrm{d}\ \ln\sigma^{-1}}{\mathrm{d}\ \ln r_L}, \tag{2.49}$$

where we have redefined $\bar{n}_v \to \mathrm{d}\bar{n}_v/\mathrm{dln}r$ and the subscript L stands for linear radius, i.e the radius . This is stressing that r_L is the radius of regions defined as void in the initial conditions, before the expansion and the eventual shell crossing (in the spherical collapse model). At the shell crossing, the spherical collapse says that the initial underdense region expanded by a factor of $\simeq 1.7$ (for instance, Blumenthal et al. 1992).

By noticing that the factor $f_v V(r) d\ln\sigma^{-1}/d\ln r$ is the fraction of voids in the interval $[\ln r, \ln r + d\ln r]$, $dF(r)/d\ln r$, then the cumulative fraction

$$F(>R) = \int_R^{\infty} d\ \ln r V(r) \frac{dn}{d\ \ln\ r} \tag{2.50}$$

is, in general, larger than unity.

The cure for this problem was given by Jennings et al. 2013, where they relax the assumption of conserving void number density, i.e $dn_L = dn$, after the spherical expansion. This is equivalent to relaxing the assumption of *isolated* spherical expansion and therefore assuming that voids might merge into each other in the process of expansion. Thus, instead of number density, the volume fraction, $F(>R)$,

is conserved, that is

$$V(r)dn = V(r_L)dn_L. \tag{2.51}$$

By substituting the above equation into 2.50 it is easy to see that it will be conserved. Therefore, the predicted abundance corrected by the volume fraction conservation is given by

$$\frac{dn}{d\ln\, r} = \frac{f_v}{V(r)} \frac{d\sigma^{-1}}{d\ln\, r_L}, \tag{2.52}$$

where we have used $r = 1.7 r_L$.

3 Weak-Lensing

3.1 Overview

As light passes through mass in the Universe, its trajectory is deflected by gravitational potentials. This effect, the gravitational lensing, is a trivial consequence of GR, appreciated by Einstein immediately when he was developing it. Later, this effect was one of the *smoking-gun* evidences of the validity of GR : the observation of the deflection of the position of a star due to the sun during a solar eclipse by Eddington (one of those observations was made in Sobral, Ceará, in the north east of Brazil). We have largely confirmed this effect since then.

Gravitational lensing was also important as evidence of the existence of collisionless DM. Figure 3.2 shows, for instance, the bullet-cluster. In this image it is possible to see the two merger galaxy clusters (purple) in the foreground distorting the shapes of background galaxies in the background, whereas most of the baryonic matter (detected in X-rays) is concentrated in between the two clusters : the dragging in the baryonic matter was caused by the collision, whereas DM passes through freely. This is a striking evidence for the existence of DM.

These examples are cases where gradient of matter around clusters, or stars, cause a drastic difference between bundles of light which passes through slight different impact parameters with respect to the object sourcing the gravitational field. This strong gradient completely distorts the shapes of background galaxies and makes it detectable at the level of individual galaxies, which generically characterizes the strong lensing regime.

Unlike strong lensing, the *weak lensing* (WL, henceforth) regime is mainly due to light being deflected by large scale structures (order of $\gtrsim h^{-1}$Mpc). These structures are usually classified as halos, filament, walls, and voids, which make up the so-called cosmic web. The first detection of gravitational WL was made in 1990 by Tyson et al. 1990 as statistical tangent alignments around clusters. Since then, it has become one of the main observable to test GR and the ΛCDM model.



FIGURE 3.1 : Credits : NASA, ESA, the Hubble Heritage Team (STScI/AURA), J. Blakeslee (NRC Herzberg Astrophysics Program, Dominion Astrophysical Observatory), and H. Ford (JHU)



FIGURE 3.2 : Credits : X-ray : NASA/CXC/M.Markevitch et al. Optical : NASA/STScI ; Magellan/U.Arizona/D.Clowe et al. Lensing Map : NASA/STScI ; ESO WFI ; Magellan/U.Arizona/D.Clowe et al.

One reason why WL lensing is so interesting for probing LSS is because it is sensitive to the total matter and, as a consequence, we circumvent one of the main problem when performing cosmological analysis with luminous tracers : the galaxy

bias Desjacques et al. 2018. Moreover, WL can also be used as complementary information to usual clustering analysis with luminous tracers and therefore shed some light into the connection between the latter and dark matter.

As a limiting factor for WL, the fact that we always infer the projection of foreground matter hides more detail information about the 3D density distribution. As an attempt to partially overcome this limitation, we can take a tomographic bin and assume that all the WL effect is due to only it, this is the so-called thin-lens approximation.

In this chapter, we aim to briefly review the WL theory and describe how it will be applied to the case in which we are interested in : the WL signature of cosmic voids.

3.2 Light deflection

In this section we derive the most basic result in weak-lensing : how light rays are deflected by gravitational potentials in the regime $\Phi/c^2 \ll 1$.

In order to derive this result we will start from the relativistic form of Fermat's principle, which states that the path traveled by a light bundle between points A and B will extremize the action

$$\tau = \int_A^B dl \frac{n[\boldsymbol{x}(l)]}{c}, \tag{3.1}$$

with n the analogous to the refraction index being the number which converts the speed of light in the presence of a gravitational field from the speed of light in the Minkowski space-time, $n = c/c'$. Therefore, this equation is simply saying that amongst all the possible light paths $\vec{x}[l]$ in the presence of a gravitational field (encoded in n), the one choosen by nature will deliver the least amount of time. Notice that we use "time" here in an imprecise way, since it has no meaning in general relativity. We are simply computing an invariant quantity and saying that the null geodesics will be an extremal of it in the curved space time with "refraction index" n. The following calculation will also assume the limit in which A and B $\to \infty$. In this limit the result does not depend on the source and lens positions.

In order to compute n, we will use the weak-field approximation mentioned earlier[1] by writting the line element as

[1] In this derivation we also assume that the lens is small compared to the system observer,

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \left(1 + \frac{2\Phi}{c^2}\right) c^2\, \mathrm{d}t^2 - \left(1 - \frac{2\Phi}{c^2}\right) (\mathrm{d}\boldsymbol{x})^2. \tag{3.2}$$

At the null geodesic we find

$$c' = \frac{|\mathrm{d}\boldsymbol{x}|}{\mathrm{d}t} = c\sqrt{\frac{1 + \frac{2\Phi}{c^2}}{1 - \frac{2\Phi}{c^2}}} \approx c\left(1 + \frac{2\Phi}{c^2}\right), \tag{3.3}$$

where we used $\Phi/c^2 \ll 1$. Therefore the refraction index is

$$n = c/c' = \frac{1}{1 + \frac{2\Phi}{c^2}} \approx 1 - \frac{2\Phi}{c^2}. \tag{3.4}$$

Since by convention $\Phi < 0$,then the light speed in the presence of a faint gravitational field Φ is less than the speed of light in the vacuum. Back to the action 3.1, we can use the parameterisation

$$dl = \left|\frac{d\vec{x}}{d\lambda}\right| d\lambda. \tag{3.5}$$

Therefore, we will find the trajectory $\boldsymbol{x}(\lambda)$ such that

$$\delta \int_{\lambda_A}^{\lambda_B} \mathrm{d}\lambda n[\boldsymbol{x}(\lambda)] \left|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}\right| = 0, \tag{3.6}$$

where the Lagrangian is

$$n[\boldsymbol{x}(\lambda)] \left|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}\right| \equiv L(\dot{\boldsymbol{x}}, \boldsymbol{x}, \lambda) \tag{3.7}$$

and $\dot{}$ denotes derivatives w.r.t the affine parameter λ.

Then, plugging this Lagrangian into the Euler-Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{\boldsymbol{x}}} - \frac{\partial L}{\partial \boldsymbol{x}} = 0 \tag{3.8}$$

we have

$$\frac{\partial L}{\partial \boldsymbol{x}} = |\dot{\boldsymbol{x}}| \boldsymbol{\nabla} n, \tag{3.9}$$

where we dropped second order derivatives and

lens and source, in such a way that we can consider the lens as a point source along the line-of-sigth.

$$\frac{\partial L}{\partial \dot{\boldsymbol{x}}} = n\frac{\partial}{\partial \dot{\boldsymbol{x}}}\left[\sqrt{\dot{\boldsymbol{x}}^2}\right] = n\frac{\dot{\boldsymbol{x}}}{|\dot{\boldsymbol{x}}|}. \tag{3.10}$$

Since $\dot{\boldsymbol{x}}$ is the vector tangent to the light's trajectory, we can normalize this vector by a suitable choice of λ and then define the normalized tangent vector $\boldsymbol{e} \equiv \dot{\boldsymbol{x}}$. Thus, plugging eqs. 3.9 and 3.10 into 3.8, we get

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(n\boldsymbol{e}) - \boldsymbol{\nabla} n = 0. \tag{3.11}$$

The second term in the derivative $d/d\lambda(n\boldsymbol{e})$ can be written as

$$\frac{d}{d\lambda} n = \boldsymbol{e}(\boldsymbol{\nabla} n \cdot \boldsymbol{e}), \tag{3.12}$$

by noticing that $dn/d\lambda = (dn[\boldsymbol{x}]/d\boldsymbol{x})\dot{\boldsymbol{x}}$.

Finally we get

$$n\dot{\boldsymbol{e}} = \boldsymbol{\nabla} n - \boldsymbol{e}(\boldsymbol{\nabla} n \cdot \boldsymbol{e}), \tag{3.13}$$

where the second term in the right-hand side is the component of the gradient perpendicular to the light path :

$$\dot{\boldsymbol{e}} = \frac{\boldsymbol{\nabla}_\perp n}{n} = \boldsymbol{\nabla}_\perp \ln(n). \tag{3.14}$$

Now using the result we derived earlier and that $\Phi/c^2 \ll 1$, we get

$$\dot{\boldsymbol{e}} = -\frac{2}{c^2}\boldsymbol{\nabla}\Phi(\vec{x}(\lambda)). \tag{3.15}$$

Finally, the total absolute deflection of the light path is

$$\hat{\boldsymbol{\alpha}} = \frac{2}{c^2}\int_{-\infty}^{\infty} \boldsymbol{\nabla}_\perp \Phi(\vec{x}(\lambda))d\lambda. \tag{3.16}$$

This result shows that the photon going through the geodesics will have its observed position altered by a cumulative effect, that given our assumptions, is simply given by integrating the gradient of the potential along its path.

Let's consider a point mass. Then the gradient Φ of its potential is

$$\boldsymbol{\nabla}_\perp \Phi = \partial_x \Phi \hat{\boldsymbol{x}} + \partial_y \Phi \hat{\boldsymbol{y}} = GM/r^3(x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}). \tag{3.17}$$

Then, by defining the impact parameter $b = \sqrt{x^2 + y^2}$ and integrating the above

gradient in the interval $[-\infty, \infty]$ along the z direction :

$$\hat{\boldsymbol{\alpha}} = \frac{2GM}{c^2} \int_{-\infty}^{+\infty} \frac{dz}{(b^2 + z^2)^{3/2}} = \frac{4GM}{c^2 b} \left(\cos(\phi)\hat{\boldsymbol{x}} + \sin(\phi)\hat{\boldsymbol{y}}\right). \tag{3.18}$$

There are two important things to notice about this result : (i) the deviation angle $\hat{\boldsymbol{\alpha}}$ is linear in M, which means that in the case we have a set of N lenses in a plane, we can find the total deviation angle as $\alpha(M_1 + ... + M_N) = \alpha(M_1) + ... + \alpha(M_N)$. (ii) In the Newtonian case, the solution for the equation of motion of a photon under the influence of a point mass potential, $\ddot{\boldsymbol{r}} = GM/r^3\boldsymbol{r}$, is $2GM/c^2 b$ and differs from our solution in the context of GR by a factor of two. This factor is directly coming from the spatial curvature in the metric 3.2, i.e., GR says that the effect of the scalar spatial curvature (small perturbations from Minkowski) is to double the effect of the gravitational potential in the photon's momentum.

Given that the matter distribution acting as a lens has negligible line-of-sight dimension compared to the other distances involved in the problem, we can express the deviation of a light bundle $\hat{\boldsymbol{\alpha}}(\boldsymbol{r}_\perp)$ (which crosses the lens plane at $\boldsymbol{r}_\perp$) in terms of the surface density

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{r}_\perp) = \frac{4GM}{c^2} \int_{\mathbb{R}^2} \frac{d^2 r'_\perp}{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2} (\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp) \Sigma(\boldsymbol{r}'_\perp), \tag{3.19}$$

where the surface density is defined as

$$\Sigma(\boldsymbol{r}_\perp) = \int_{-\infty}^{\infty} d\chi \rho(\boldsymbol{r}_\perp, \chi), \tag{3.20}$$

with $\rho(\boldsymbol{r}_\perp, \chi)$ being the density profile of the lens, e.g. the number density of DM particles around a void multiplied by the mass of each DM particle.

Therefore, given our assumptions, the deviation angle will be completely determined by the density profile of the lensing structure.

3.2.1 The Lens equation

Now that we have the angle that a light bundle will suffer due to some structure, we can relate the latter to the observed and original angular positions of light-bundles.

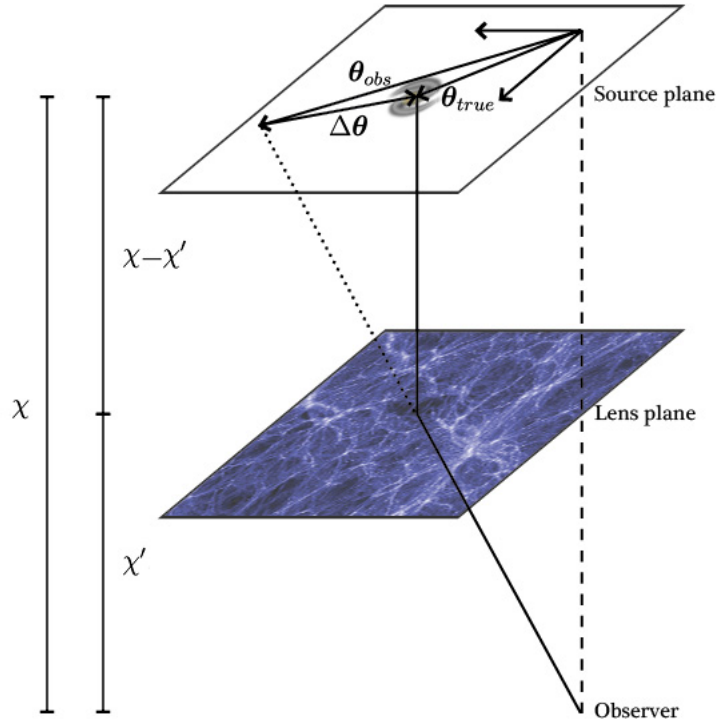

FIGURE 3.3 : Illustration of a generic lensing setup. The origin of the coordinate system we use through out this section is defined at the intersection between the lens and source planes and the line connecting the observer and the source.

Figure 3.3 shows a generic observer-lens-source system. The light is emitted with an impact parameter $\boldsymbol{\theta}_{true}$ and comoving distance $\chi \equiv \chi(z_S)$ and is deviated by a lens at comoving distance χ', resulting in an observed angular position $\boldsymbol{\theta}_{obs}$. Then, by simply relating the observed and true positions[2], the relation between the emitted and observed angular positions is

[2]notice that we are assuming a flat Universe

$$\boldsymbol{\theta}_{obs} = \Delta\boldsymbol{\theta} + \boldsymbol{\theta}_{true}, \tag{3.21}$$

where, using the result from the previous section,

$$\Delta\boldsymbol{\theta} = \int_0^{\chi} d\chi' \frac{(\chi - \chi')}{\chi} d\hat{\boldsymbol{\alpha}} = \frac{2}{c^2} \int_0^{\chi} d\chi' \frac{(\chi - \chi')}{\chi} \boldsymbol{\nabla}_{\perp} \Phi. \tag{3.22}$$

The above equation can be interpreted as each lens at comoving distance χ' is adding a small deviation $d\boldsymbol{\theta} = (\chi - \chi')/\chi \, d\hat{\boldsymbol{\alpha}}$ to the true source angular position. Then, the total shift in the true position is given by the integral along the whole path from the observer to the source.

Equation 3.21 is called the "lens equation" and is the starting point for the fundamental results of weak-lensing theory. Indeed, this equation encapsulates the GR effect in light rays through $\hat{\boldsymbol{\alpha}}$ and the geometry of the system. We derived this equation from a very simple geometric argument and under the small angle approximation, which is almost always the case in cosmology, but it is simply the geodesic equation of a photon going from the source plane to the observer.

In WL we are interested in quantifying the distortion and change of an image shape of a source galaxy due to the foreground gravitational potential. Of course, the image has to have a non-negligible apparent size. The treatment is largely facilitated in the case in which the apparent image size of source galaxies are small compared to the changes in the density profile of the lensing structure. In other words, we can expand the mapping between positions in the lens and source planes and use only the first order term. The distortion in the source image is then encoded in the Jacobian

$$\frac{\partial \theta_{true,i}}{\partial x_j} = \delta_{ij} - \frac{\partial \Delta\theta_i}{\partial x_j}, \tag{3.23}$$

where the derivatives are taken with respect to a coordinate system defined in the lens plane.

It is convenient to define the lensing potential as

$$\Delta\theta_i = \frac{\partial}{\partial x_i}\psi, \tag{3.24}$$

$$\psi = \frac{2}{c^2} \int_0^{\chi} d\chi' \frac{(\chi - \chi')}{\chi\chi'} \Phi(\boldsymbol{\theta}\chi', \chi'). \tag{3.25}$$

In the above definition we used that $\partial\boldsymbol{\theta}\chi'/\partial x_i = \chi'$.

We can define then the *distortion tensor* A_{ij} as :

$$A_{ij} \equiv \delta_{ij} - \frac{\partial^2\psi}{\partial x_i \partial x_j}, \tag{3.26}$$

which can be parameterized as

$$\boldsymbol{A} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{3.27}$$

In this parameterization, κ is the *convergence* and it is associated with an overall increase or decrease of the image size, whereas $\gamma_{1,2}$ are the *shear* components and are associated with elongations along the axes x_1 and x_2. Therefore, in the simplest case in which the source image is a circle, the convergence will be associated with a change in the circle size and the shear components to the eccentricity of an ellipse.

Notice that $\boldsymbol{A}$ has an identity component, which is the limit of no gravitational potential.

By definition, the shear components and the convergence are given by $\gamma_1 = 1/2(\partial_1\partial_1 - \partial_2\partial_2)\psi$, $\gamma_2 = \partial_1\partial_2\psi$ and $\kappa = 1/2(\partial_1^2 + \partial_2^2)\psi$. The shear components are often expressed as a complex number :

$$\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2\phi i}. \tag{3.28}$$

Thus, γ can be understood as a spin-2 field, i.e. a rotation of π will turn it into itself.

The convergence is simply the projection of the Poisson equation along the line-of-sight, weighted by a geometric factor :

$$\kappa(\boldsymbol{\theta}) = \frac{1}{c^2}\int_0^\chi d\chi' \frac{(\chi-\chi')\chi'}{\chi}\boldsymbol{\nabla}^2\Phi(\boldsymbol{\theta}\chi', \chi'). \tag{3.29}$$

The geometrical weight can be absorbed into the definition of the critical surface density :

$$\Sigma_c = \frac{c^2}{4\pi G}\frac{\chi}{\chi'(\chi-\chi')}. \tag{3.30}$$

Notice that the inverse of the critical surface density is weighing the Poisson equation, and Σ_c^{-1} is a parabola in χ, with maximum at $\chi/2$, i.e. the lens positions which maximise the distortions in source galaxy shapes, are the ones about half-way

between the observer and the source.

By introducing this definition and taking the right hand side of the Poisson equation, we can write the convergence as

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi} d\chi' \Sigma_c^{-1} a^2 \bar{\rho}_m(a) \delta(\boldsymbol{\theta}\chi', \chi') = \frac{3H_0^2\Omega_m^{(0)}}{8\pi G} \int_0^{\chi} d\chi' \frac{\delta(\boldsymbol{\theta}\chi', \chi')}{a(\chi')\Sigma_c}. \tag{3.31}$$

By defining the surface density contrast as

$$\Sigma(\boldsymbol{\theta}) = \frac{3H_0^2\Omega_m^{(0)}}{8\pi G} \int_0^{\chi} d\chi' \frac{\delta(\boldsymbol{\theta}\chi', \chi')}{a(\chi')} \tag{3.32}$$

and assuming that Σ_c is approximately a constant over the values of χ where $\delta(\boldsymbol{\theta}\chi)$ is non-zero, we can express the convergence as

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_c}. \tag{3.33}$$

The assumption that we can take Σ_c outside the integral, allows us to relate a quantity which is only dependent on the local density field, Σ, with a weak-lensing observable, κ.

In Fig. 3.4 an elliptical galaxy is lensing source galaxies in the background, drastically distorting their shapes. A weak version of this effect will happen in weak-lensing in cosmology.

One of the usefulness of weak-lensing in cosmology is to infer the density profile of clusters and voids. To make this kind of inference we always work with averaged profiles, which are isotropic. Therefore, the effect on source galaxy shapes will be a stretch either on the direction perpendicular or radial to the line connecting the cluster and the source galaxy positions.

In these cases, γ_1 and γ_2 are not necessarily aligned with the tangential or radial direction. For instance, in the case that the line connecting the cluster to the source galaxy position in the lens place is parallel to the θ_x-axis, then non-zero γ_1 means ellipticity aligned either with the θ_x-axis (negative γ_1) or with the θ_y-axis (positive γ_1), whereas ellipticity aligned with the direction rotated by 45° w.r.t. θ_x-axis means non-zero γ_2. However, for a generic source galaxy position, the ellipticity will be described by a combination of γ_1 and γ_2, even if the actual ellipticity is aligned with the tangential direction.

Therefore, it is more useful to express the shear in a basis which has one of

the axis always aligned with the tangential direction around a certain structure. This problem is exactly analogous to the polarization of the CMB, in which the polarization tensor is decomposed into the E and B modes.

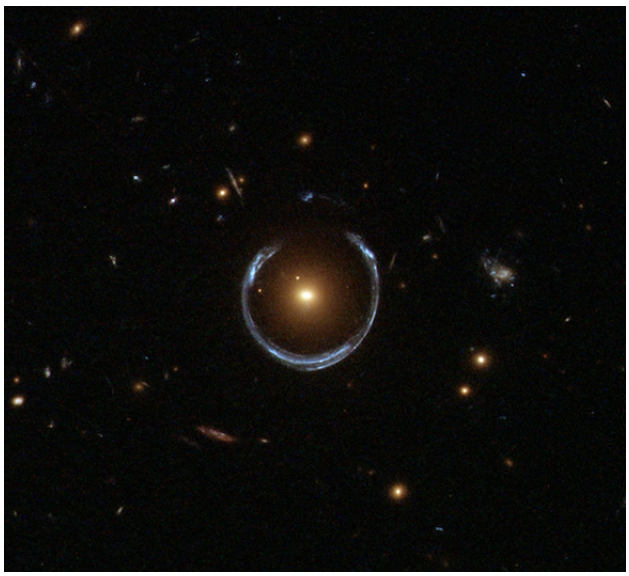

FIGURE 3.4 : Strong lensing around a galaxy from the Hubble telescope

Let's take the traceless part of the distortion tensor :

$$A_{ij}^T = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}. \tag{3.34}$$

This is a tensor which has two independent components. We can then decompose it in terms of a scalar component plus a transverse-traceless tensor, as we did for the metric perturbations. Thus, A_{ij}^T is written as

$$A_{ij}^T = 2\left(\frac{\theta_i\theta_j}{\theta^2} - \frac{1}{2}\delta_{ij}\right)\gamma_t + A_{ij}^{TT}, \tag{3.35}$$

where we introduced the scalar "perturbation" (in analogy to scalar metric perturbations) γ_t and A_{ij}^{TT} is a transverse-traceless tensor. It is easy to see that the scalar component is extracted as

$$\gamma_t = \frac{\theta_i\theta_j}{\theta^2} A_{ij}^T. \tag{3.36}$$

By writing a position in the $\theta_x - \theta_y$ plane in polar coordinates $\boldsymbol{\theta} = |\boldsymbol{\theta}|(\cos(\phi), \sin(\phi))$,

$$\gamma_t = (\cos^2(\phi) - \sin^2(\phi))\gamma_1 + 2\sin(\phi)\cos(\phi)\gamma_2 = \cos(2\phi)\gamma_1 + \sin(2\phi)\gamma_2, \quad (3.37)$$

which can be written in short-hand as

$$\gamma_t = \mathbb{R}[\gamma e^{-2i\phi}]. \quad (3.38)$$

The transverse-traceless components are :

$$\begin{aligned} A_{12}^{\rm TT} &= A_{12}^{\rm T} - 2\frac{\theta_1\theta_2}{\theta^2}\gamma_t \\ &= \gamma_2 - \sin(2\phi)\left(\cos(2\phi)\gamma_1\right) + \sin(2\phi)\gamma_2) \\ &= \left(1 - \sin^2(2\phi)\right)\gamma_2 - \sin(2\phi)\cos(2\phi)\gamma_1 \\ &= \cos(2\phi)\gamma_\times, \end{aligned} \quad (3.39)$$

where we defined

$$\gamma_\times = \cos(2\phi)\gamma_2 - \sin(2\phi)\gamma_1, \quad (3.40)$$

which can be written as :

$$\gamma_\times = \mathbb{I}[\gamma e^{-2i\phi}.] \quad (3.41)$$

It is straightforward to show that the remaining components of A^{TT} are also written in terms of $\gamma_\times$.

Finally, we can write the transverse part of the distortion matrix in terms of γ_t and $\gamma_\times$ as :

$$A_{ij}^T = \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}(-\gamma_t) + \begin{pmatrix} -\sin 2\phi & \cos 2\phi \\ \cos 2\phi & \sin 2\phi \end{pmatrix}(-\gamma_\times) \quad (3.42)$$

and for convenience we have redefined $\gamma_{t,\times} \to -\gamma_{t,\times}$. For instance, if $\phi = \pi/2.$, then

$$A_{ij}^T = \begin{pmatrix} \gamma_t & \gamma_\times \\ \gamma_\times & -\gamma_t \end{pmatrix}, \quad (3.43)$$

which means that $\gamma_t > 0$ is associated with distortions along the θ_x direction, whereas $\gamma_t < 0$ is associated with distortions along the θ_y direction. For any ϕ, $\gamma_t > 0$ will always be associated with distortions along the direction tangential to the line connecting the cluster and the source, whereas $\gamma_t < 0$ is associated with distortions along the radial direction. The reason why this decomposition is so useful, is because isotropic structures only induce non-zero *tangential shear*, γ_t, and never induce non-zero *cross-shear*, $\gamma_\times$. These components are often called E-mode (γ_t) and B-mode ($\gamma_\times$), in analogy to the E-B decomposition of the polarization tensor.

In a real observation, we choose a basis and measure the $\gamma_{1,2}$ components with respect to this basis. We can then convert these measured numbers in $\gamma_{t,\times}$ through equations 3.38 and 3.41.

On the other hand, we would like to have a prediction for the averaged tangential shear around a structure in terms of the expected profile of this structure. This prediction requires a relation between the tangential shear and the convergence, which can be achieved as follows.

Let's apply the Gauss theorem in 2D to the gradient of the gravitational potential over a circle :

$$\oint_{\partial S} d\boldsymbol{l} \cdot \boldsymbol{\nabla}\psi(\boldsymbol{\theta}, \phi) = \oint_S dS \boldsymbol{\nabla}^2\psi(\boldsymbol{\theta}, \phi). \tag{3.44}$$

By definition, $\boldsymbol{\nabla}^2\psi = 2\kappa$, so we can rewrite the above equation as

$$\theta \int_0^{2\pi} d\phi \frac{\partial\psi}{\partial\theta} = \int_0^{2\pi} d\phi \int_0^{\theta} d\theta' \theta' 2\kappa, \tag{3.45}$$

where we used that $d\boldsymbol{l} = \hat{\boldsymbol{n}}|\boldsymbol{\theta}|$, with $\hat{\boldsymbol{n}} = (\cos(\phi), \sin(\phi))$ the unitary vector pointing in the radial direction. Notice that the dot product between the line element and the gradient of ψ selects the component of the gradient parallel do the radial direction.

By applying the derivative w.r.t. θ in both sides :

$$\frac{\partial\langle\psi\rangle(\theta)}{\partial\theta} + \theta\frac{\partial^2\langle\psi\rangle(\theta)}{\partial\theta^2} = 2\theta\langle\kappa\rangle(\theta), \tag{3.46}$$

where we introduced the brackets indicating average along the circle

$$\langle f\rangle(\theta) = \frac{1}{2\pi}\int_0^{2\pi} d\phi f(\phi, \theta), \tag{3.47}$$

and used the Leibniz rule to integrate the right hand side.

Now we have to use a bit of intuition to deal with the second term in the left hand side. The second derivative of $\langle\psi\rangle$ is the Laplacian along the radial direction of a quantity which is averaged over ϕ, i.e., this result must be independent of our coordinate choice. This observation motivates the *Ansatz* that this term must be written in terms of γ_t. Indeed the distortion tensor is the identity plus the second derivative of ψ in the directions i, j (Eq. 4.4). If we take the definition of γ_t and test the cases, $\phi = n\pi/2$, we will always find the distortion matrix if we take

$$\frac{\partial^2\langle\psi\rangle}{\partial\theta^2} = \langle\kappa\rangle - \langle\gamma_t\rangle. \tag{3.48}$$

Thus, by using the above *Ansatz* and replacing the first term equation 3.47 by equation 3.45 :

$$\frac{2}{\theta}\int_0^\theta d\theta'\theta'\langle\kappa(\theta')\rangle + \theta(\langle\kappa\rangle(\theta) - \langle\gamma_t\rangle(\theta)) = 2\theta\langle\kappa\rangle(\theta)) \tag{3.49}$$

and we finally get to the relation between the averaged tangential shear and convergence around a structure :

$$\langle\gamma_t\rangle(\theta) = \overline{\langle\kappa\rangle}(<\theta) - \langle\kappa\rangle(\theta), \tag{3.50}$$

where we defined the averaged convergence inside a circle of radius θ :

$$\overline{\langle\kappa\rangle}(<\theta) = \frac{2}{\theta^2}\int_0^\theta d\theta'\theta'\langle\kappa(\theta')\rangle. \tag{3.51}$$

This result is a relation between a single quantity that parameterizes the net effect of a structure in the shapes of source galaxies and the projection of the total matter density field along the line-of-sight. In one hand we can take the observed $\gamma_{1,2}$ from observed photometric galaxies, convert them in γ_t and them compare it to the total projection of the density field, i.e., it allows to "see" the density profile of structures directly, without the need for converting between the observed density field through luminous tracers and the one we can calculate from Effective field theory of large-scale structure, N-body simulations, or another analytical/numerical method.

Of course that life is not too simple and we have some immediate issues with this inference. The most obvious one is that we are not really inferring the projected

density field, but rather the latter weighted by Σ_c^{-1}, which depends on the source distribution. Furthermore, there are observational issues inherent to weak-lensing, such as errors in the estimation of photometric galaxy shapes.

Regarding the first problem, there is way of overcoming it. Let's remind that the convergence can be approximated by Σ/Σ_c is the case that Σ_c varies very slowly over the region along the line-of-sight where the (averaged) density field is non-zero. Using this approximation, we can rewrite the tangential shear as

$$\langle\gamma_t\rangle(\theta) = \Sigma_c^{-1}(\overline{\Sigma}(<\theta) - \Sigma(\theta)) \tag{3.52}$$

and we can define the Excess Surface Mass Density (ESMD) :

$$\Delta\Sigma(\theta) = \overline{\Sigma}(<\theta) - \Sigma(\theta) = \Sigma_c\langle\gamma_t\rangle(\theta). \tag{3.53}$$

In other words, by observing the distortion caused by a targeted structure on the shapes of source galaxies, we can infer the projected density field around this structure.

Excess Surface Mass Density Estimator

As we have to find the optimal weighing to estimate the power spectrum, as was done by FKP, we also need to find the optimal weights to estimate $\Delta\Sigma$ from the given shear estimates. Here we reproduce the estimator first introduced in Sheldon et al. 2004 and that will be latter used to estimate $\Delta\Sigma$ from simulated shear estimates.

In the weak-lensing regime, the galaxy ellipticity can be written as :

$$\epsilon = \mathcal{R}2\gamma_t + \epsilon^{int}, \tag{3.54}$$

where $\mathcal{R}$ is the responsivity function and ϵ^{int} is the intrinsic shape. The responsivity function describes how susceptible a galaxy is to an induced shear by foreground structures. It is easy to imagine why this might vary from galaxy to galaxy : the induced shear is merely a consequence of light-rays from different points of the galaxy image having different impact parameters and therefore their photons path's being subjected to different potential gradient. Having this intuition in mind, clearly the responsivity might depend on the intrinsic shape and size of the source galaxy.

The error in the estimated shear must be then

$$4\sigma(\gamma_t) = \sigma(\epsilon) + \sigma(\epsilon^{int}), \tag{3.55}$$

where $\sigma(\epsilon)$ is the error in the estimation of galaxy shapes from photometric images and $\sigma(\epsilon^{int}) = \langle(\epsilon^{int})^2\rangle$ is the variance in the intrinsic shape of galaxies.

We can write the logarithm likelihood for $\Delta\Sigma$ as

$$\log \mathcal{L}(\Delta\Sigma) = \sum_{j,i} \left(-\frac{1}{2} \left[\frac{\gamma_t^{ij} - \Sigma_{crit,ji}^{-1} \Delta\Sigma}{\sigma\left(\gamma_t^{ij}\right)} \right]^2 \right), \tag{3.56}$$

where j, i runs, respectively, over the lenses and sources and γ_t^{ij} is the shear induced by the lens j into the source i.

The maximum likelihood is given by

$$\Delta\Sigma = \frac{\sum_{i=0}^{N_S} \sum_{j=0}^{N_L} w_{ij} \Sigma_{crit,ij} \gamma_t^{ij}}{\sum_{i=0}^{N_S} \sum_{j=0}^{N_L} w_{ij}}, \tag{3.57}$$

where $w_{ij} = \sigma_{ij}^{-2}$ and $\sigma_{ij} = \sigma(\gamma_t^{ij})\Sigma_{crit,ij}$.

Therefore, this estimator is basically taking the induced shear for each pair and attributing an effective weight for it, given by $\Sigma_{crit,ij}^{-1} = 4\pi G/c^2(\chi_L^2/\chi - \chi_L)$, a parabola with maximum at $\chi/2$.

3.2.2 Ellipticities and shear

When averaging millions or billions of source galaxy shapes, we expect that the resulting shape will be circular, so we can say that any deviation from a circle, or ellipticity, will be due to gravitational lensing (neglecting intrinsic alignments). However, we have not defined at exactly we mean by ellipticity and how it relates to the shear.

Let's take, as before, in the lens plane and center the image at the origin. Then we can calculate the quadrupole moment of the image as :

$$Q_{ij} \equiv \langle \theta_i \theta_j \rangle_{I_{obs}} \equiv \frac{1}{F} \int d^2\theta I_{obs}(\theta) \theta_i \theta_j, \tag{3.58}$$

where $I_{obs}(\theta)$ is the energy incident on a detector per solid angle, per unit of area

and time, per unit of frequency, that is

$$dE = Id\Omega dA_{\perp} dt d\nu. \tag{3.59}$$

The backet notation makes clear the meaning of the quadrupole : it is an average over the image weighted by the intensity. The normalisation constant F is

$$F = \int d^2\theta I_{\rm obs}(\boldsymbol{\theta}), \tag{3.60}$$

or the total flux of the image. Notice that the dimensions of $I(\theta)$ guarantees that the quadrupole moment is properly normalised.

Since Q_{ij} is a 2×2 symmetric tensor and, therefore, it has three independent components, we can write it as :

$$Q_{ij} = \frac{1}{2} Q \begin{pmatrix} 1+\epsilon_1 & \epsilon_2 \\ \epsilon_2 & 1-\epsilon_1 \end{pmatrix}, \tag{3.61}$$

where Q is the trace $Q = \mathrm{Tr}[Q_{ij}]$ and $\epsilon_{1,2}$ are the ellipticities. Since Q is multiplying all the components, it is associated with increase or decrease of the image size, whereas the ellipticities are distortions in the θ_x, θ_y (ϵ_1) or $\theta_x = \theta_y$ (ϵ_2) directions, i.e., a circular image implies $\epsilon_1 = \epsilon_2 = 0$.

Now we would like to link these ellipticities, which can be directly measured, with the shear, for which we know how to calculate through the gravitational potential.

Since we know the Jacobian of the transformation between positions in the source plane and the lens plane, which depends on the shear and convergence, we just have to write down $\epsilon_{1,2}$ in terms of the integral 3.58 and do the calculation. Therefore, following our parameterization in 3.61, we can write the ellipticities as

$$\epsilon_1 = \frac{Q_{xx} - Q_{yy}}{Q_{xx} + Q_{yy}} \tag{3.62}$$

and

$$\epsilon_2 = \frac{2Q_{xy}}{Q_{xx} + Q_{yy}}, \tag{3.63}$$

where we have used a simplified notation $x \equiv \theta_x$, $y \equiv \theta_y$, but keep in mind that these are the positions in the lens plane.

Plugging the integral 3.58 into 3.62 :

$$\epsilon_1 = \frac{\int d^2\theta I_{\rm obs}(\theta)\left[\theta_x\theta_x - \theta_y\theta_y\right]}{\int d^2\theta I_{\rm obs}(\theta)\left[\theta_x\theta_x + \theta_y\theta_y\right]}. \tag{3.64}$$

Now, by energy conservation, i.e., because lensing does not create photons, but simply change their path[3], $I_{true}(\boldsymbol{\theta}_{true}) = I_{obs}(\boldsymbol{\theta}_{obs})$, where $\boldsymbol{\theta}_{true}$ is the position in the source plane and we reused the subscript *obs* for quantities in the lens plane. Using that $\theta_i = A^{-1}_{ij}\theta^S_j$ and $d^2\theta = |\boldsymbol{A}^{-1}|d^2\theta^S$, equation 3.64 is rewritten as

$$\epsilon_1 = \frac{\int d^2\theta^S \left|A^{-1}\right|\left[\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xj} - \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yj}\right] I_{true}\left(\theta^S\right)\theta^S_i\theta^S_j}{\int d^2\theta^S \left|A^{-1}\right|\left[\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xj} + \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yj}\right] I_{true}\left(\theta^S\right)\theta^S_i\theta^S_j}. \tag{3.65}$$

We can pull all the matrices A outside the integral, since it depends only on variables in the lens plane. To perform the integrals, we use that the galaxy shape in the source plane is a circle. Since in all our calculations we are always using averaged shapes over many galaxies, this is a reasonable assumption. As a consequence of circular shape, the integrals when $i \neq j$ will be zero and we will be left with δ_{ij}. After contracting the indices we get :

$$\epsilon_1 = \frac{\left[\left(A^{-1}_{xx}\right)^2 - \left(A^{-1}_{yy}\right)^2\right]}{\left[\left(A^{-1}_{xx}\right)^2 + 2\left(A^{-1}_{xy}\right)^2 + \left(A^{-1}_{yy}\right)^2\right]}. \tag{3.66}$$

Finally, taking the inverse of $\boldsymbol{A}$,

$$\begin{aligned}\epsilon_1 &= \frac{(1-\kappa+\gamma_1)^2 - (1-\kappa-\gamma_1)^2}{(1-\kappa+\gamma_1)^2 + 2\gamma_2^2 + (1-\kappa-\gamma_1)^2} \\ &= \frac{4\gamma_1(1-\kappa)}{2(1-\kappa)^2 + 2\gamma_1^2 + 2\gamma_2^2}\end{aligned}. \tag{3.67}$$

Since we are in the weak-lensing regime ($\kappa, \gamma_{1,2} \ll 1$), all second order quantities can be dropped, yielding

$$\epsilon_1 \simeq 2\gamma_1 \tag{3.68}$$

[3]Photons might loose energy through gravitational redshift, which we neglect. The point here is that photons are not created or destroyed and all of the ones which left the source will reach the telescope.

and an analogous calculation will lead to

$$\epsilon_2 \simeq 2\gamma_2. \tag{3.69}$$

This result shows that the observed ellipticities are direct measurements of the projected density field of total matter.

In this derivation, we have used the Jacobian of the transformation of areas between the lens and source planes,

$$\mu \equiv |\boldsymbol{A}^{-1}| = \frac{1}{(1-\kappa)^2 - \gamma_1^2 - \gamma_2^2}, \tag{3.70}$$

which is the *magnification* and describes the change in the flux, which is equivalent to the change in the image size, since the number of photons is conserved. By dropping second order quantities,

$$\mu \simeq \frac{1}{1-2\kappa} \simeq 1 + 2\kappa. \tag{3.71}$$

By looking to equation 3.31, we see that if $\delta > 0$, the images sizes are magnified. However, if $\delta < 0$, which usually occurs in voids, the image size suffers a decrease in size, or *demagnification*.

3.2.3 Theoretical challenges

Despite being a promising observable and having increasing attention in literature due to its potential, weak-lensing science also has clear limitations and/or challenges to be overcome over the next years.

Projection

The most obvious unavoidable limitation is the fact that weak-lensing is only capable of recovering 2D information. We can partially overcome this limitation by taking tomographic bins, which is informally known as to recovering "2.5D" information. It has been shown that a suitable tomographic binning is capable of enhancing the precision on (Ω_m, σ_8) measurements by a factor of 5 Yuan et al. 2019 in comparison to the 2D peak analysis.

Intrinsic alignments

The next "leading order" limitation is related to one of the most important assumptions in all the results we have derived : the assumption that galaxies might have intrinsic shapes, but that these shapes are randomly oriented. It is well known that this assumption is not necessarily true. Galaxy shapes are subjected to tidal forces and tend to align in preferentially in the direction of tidal field. Near to massive halos, galaxies tend to have shapes aligned with the radial direction, whereas near to filaments, galaxy shapes tend to align in the direction of the filament. Intuitively one would think that this effect is important only on small scales, but it has been shown that the correlations between intrinsic shapes of galaxies can span up to $\simeq 100h^{-1}$Mpc Delgado et al. 2023. It is well established that ignoring the intrinsic-intrinsic correlations can lead to significant errors in weak-lensing analysis Kirk et al. 2015.

Up to date, and up to the knowledge of the present author, there are not studies on the impact of intrinsic alignments on void-lensing. Intuitively, it is expected to have less impact on void-lensing analysis since we are correlating larger scales compared to halos. However, the void-lensing signal is also fainter and hence any systematic effect will have more importance.

Baryonic effects

The weak-lensing signal is sensitive to the total matter field, since photons couples to dark matter and baryons. However, when making analytical or numerical predictions, we only account for the dark-matter field. It is known that the halo profile is highly affected by baryons Duffy et al. 2010, specially on small scales. Recent weak-lensing analysis in real data have avoided this problem by applying a cutoff at some scale. However, if we aim to extract the whole potential from weak-lensing, we have to model baryonic effects.

For void-lensing we expect that this effect will be minor for two reasons : we are probling larger scales and baryons have a marginal effect on void profiles.

Non-linearities

Weak-lensing cosmological analysis use to apply a scale cut to avoid the non-linear modelling Abbott et al. 2022, losing then constraining power from these scales, what is expect to be non-negligible due to less cosmic variance. The account of

non-linearities on weak-lensing is arguably more challenging that 3D tracer-tracer clustering since all scales will be important in the projection of the power spectrum. Therefore, effective field theoy of large scale structure shall not be enough, since it has a well defined scale up to which the theory works. The most promising strategy in the near future is to use emulators to interpolate numerical predictions accounting for the baryonic effects.

In the case of void-lensing, we also expect the non-linearities to be less important. However we don't have any analytical prediction for internal void profiles and it might no be possible to do it analytically. Emulators are highly promising for voids since it potentially does not require the inclusion of baryonic effects, i.e., a dark-matter only simulation should be sufficient.

4 Void-Lensing

4.1 The Void-Lensing measurement and numerical interpretation

In the following section I reproduce the published version of this work. For the sake of being self-contained, we keep the sections which are summaries of the theory developed in the previous sections.

4.1.1 Introduction

Given the abundance of cosmological data coming from Large-Scale Structure (LSS) surveys such as Euclid Laureijs et al. 2011 and DESI Aghamousa et al. 2016, one of the central challenges of this era is to know how to interpret and optimize the extraction of relevant physical information. This task passes through the identification of discrete tracers of LSS, each one tracing the underlying dark matter field in a particular way and therefore carrying information about the details of the underlying Dark Matter (DM) structures.

One particular type of tracer of LSS is cosmic voids, which consists, generically speaking, of large underdensities located in-between clusters, filaments and walls, dominating the volume of the Universe. These voids typically range from a few h^{-1}Mpc up to $\simeq 100h^{-1}$Mpc in radius. Hence, they populate the LSS in a very distinct manner compared to galaxies or halos, as evidenced, for instance, by their negative linear bias Chan, Li et al. 2020. Thus, these structures carry complementary information to luminous tracer statistics. In fact, it has been shown that the density profiles of voids, as well as their abundance, are particularly sensitive to dark energy Pisani et al. 2015, massive neutrinos Massara, Villaescusa-Navarro et al. 2015 ; Schuster et al. 2019, primordial non-gaussianities Kamionkowski et al. 2009 and modifications to general relativity Voivodic, Lima, Llinares et

David F. Mota 2017b ; Perico et al. 2019. Arguably, since voids are underdense in DM, their evolution should be more sensitive to dark energy, modifications to gravity and neutrinos.

Despite their potential as a cosmological probe, using voids in cosmological analysis is challenging for a couple of reasons, amongst them are the high shot-noise in the case of void auto-correlation, the presence of cosmic variance and the mismatch between voids found in the sparse galaxy field and the true voids in the underlying DM field, as well as the lack of theoretical knowledge regarding the distribution of matter in the vicinity of the void center (see Massara et Sheth 2018 for a recent development in this direction).

One particular observable involving voids is their imprint on the shape of background galaxies, the so-called void lensing (VL). Unlike overdensities, around which the shapes of background galaxies are strengthened around the target structure, voids will leave the opposite imprint : galaxy shape ellipticities tend to be radially oriented around the void, as a result of an "anti-lensing", or negative tangential shear. It has received increasing attention in the last two decades, since the first proposition (up to our knowledge) by Amendola et al. 1999 in 1998. Since then, a few measurements were made Carles Sánchez et al. 2016 ; Fang et al. 2019 ; Melchior et al. 2014, as well as analytical and numerical investigations have shown the sensitivity of this observable to modifications of gravity Barreira et al. 2015 ; Baker et al. 2018 ; Davies, Cautun et al. 2019.

Weak-lensing combined with voids can mitigate the limitation imposed by the sparsity of luminous tracers, since it is sensitive to the total matter and therefore we are able to "see" voids in this field. Furthermore, weak-lensing in general is useful to test gravity because it is sensitive to the sum of the Newtonian potentials $\Psi_L = (\Phi + \Psi)/2$ which in general are not equal in modified gravity scenarios. Basically, by measuring the distortion on the shapes of distant galaxies when their light passes through voids, we are directly (the total matter) probing the cleanest (less baryonic complications), simplest (less non-linearities) and more sensitive environment to modified gravity.

In the literature, the detection of weak lensing around voids is basically treated in two distinct ways : (i) by measuring the tangential shear γ_t, or the convergence κ, which parameterize the distortion on the shapes of background galaxies caused by the underdense structures spanning from the observer up to the source Shimasue et al. 2023 ; Gruen et al. 2016 ; Higuchi et al. 2013 ; Davies, Paillas et al. 2021

and (ii) by measuring the Excess Surface Mass Density (ESMD), $\Delta\Sigma$, which is basically the projection of the total matter contained in a thin lens (in this case, voids) localized somewhere in-between the observer and the source. Therefore, the approach (i) is measuring the projected profile of ultra-large troughs (with sizes of hundreds of h^{-1}Mpc), whereas the approach (ii) is measuring the projected profiles of voids with radii $\leq 50h^{-1}$Mpc, which is the limiting radius for which the thin lens approximation is still valid, as we will show in this work. The aim of the approach (ii) is to extract dynamical and morphological information about the usual generic definition of voids, i.e., underdense structures located in-between overdense structures (halos, sheets and filaments). The inference of void profiles and mass function through lensing can potentially give cleaner information about voids in the total matter and be a complementary to void analysis by, for instance, being able to measure the void bias Fang et al. 2019, whereas approach (i) measures the weak-lensing signature of ultra-large structures along the line-of-sight, the so-called troughs, and the relation of these objects to a generic definition of voids (with predictable void abundance) is not clear. Furthermore, approach (i) has less 3D information, which approach (ii) partially recovers. Although approach (i) could be interesting on its own, here we make a distinction between them and call by "void lensing" approaches such as (ii) and by "troughs" the approach (i). This work investigates aspects of VL.

Previous works on void lensing used two definitions of voids Fang et al. 2019; Carles Sánchez et al. 2016. In short, the first definition finds voids in the 3D galaxy field and the second finds voids in projections of the galaxy field with width of $100h^{-1}$Mpc. The first approach has the advantage of having trivial interpretation, i.e., $\Delta\Sigma$ is basically the projection of the void profile, but has the disadvantage of signal contamination by overdensities surrounding the voids. The second approach has no clear interpretation in terms of void definitions in the 3D field, but has the potential of obtaining higher S/N per void, since these voids fill the whole projected slices along the line-of-sight, minimizing the contamination by overdensities.

In this work we aim to compare VL measurements for two types of void-finders : one which is widely used in literature and also in previous measurements of VL in real data, based on the `ZOBOV` algorithm Fang et al. 2019; Carles Sánchez et al. 2016, and the one that we introduce in this work. We show that it is worth exploring the different sensitivity of these different approaches to cosmology and modified gravity, due to the very distinct profiles of voids that they yield. We also

investigate the consistency between the VL signal as measured through shear, and the same signal as estimated directly using the DM field. If we aim to do precision cosmology with this observable, we have to, first of all, be sure that we really measure the projected void profile through the shear in a controlled environment. As we will show, when working with voids to measure the ESMD, problems related to the size of voids might arise. We believe this work paves the way for a numerical interpretation of void lensing and therefore for precision cosmological analysis using the data from up-coming surveys with large sky-fraction coverage.

This work is organized as follows. We begin by recapping the weak-lensing basics that we are going to use in the following sections. In section 4.1.2, we describe the void-finding algorithm introduced in this work, as well as the resulting profile and abundance of voids measured in a simulation. In section 4.1.3, we compare the performance of our void-finding algorithm with the widely used `ZOBOV` void finder by measuring the ESMD in a galaxy mock. In section 4.1.7, we compare the ESMD as measured through the tangential shear γ_t and the one measured directly from the DM field of the same realisation, giving then a numerical interpretation of the observational ESMD and addressing the issue of the thin-lens approximation in the context of void lensing. Finally, we conclude with some observational considerations and avenues for future works.

Weak-Lensing review

In this section, we briefly review a few results of the weak-lensing theory that constitute the basis of this work. We refer to Schneider 2005 for an extensive review.

As the light of background galaxies propagate through LSS, its path is perturbed by the gravitational field in the foreground. The difference between the unperturbed and observed positions, respectively $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, is given by the lens equation :

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}\,. \tag{4.1}$$

Assuming small scalar perturbations to the Minkowsky metric, $\boldsymbol{\alpha}$ is given by the projection of the gradient of the gravitational potential along the line of sight :

$$\boldsymbol{\alpha} = \frac{2}{c^2}\int_0^{\chi_S} d\chi' \frac{(\chi_S - \chi')\chi'}{\chi_S} \boldsymbol{\nabla}_\perp \Phi(\boldsymbol{\theta}\chi', \chi'). \tag{4.2}$$

The integral is accounting for all lenses at positions χ' up to the source position χ_S. The factor $(\chi_S - \chi)\chi/\chi_S$ is a geometrical weight of the projection. Therefore, it is expected that the observed shapes of background galaxies will be different than the original shapes. The distortion in those shapes is expressed by the distortion matrix

$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \boldsymbol{\alpha}}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}, \tag{4.3}$$

which can be written as

$$\boldsymbol{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{4.4}$$

In eq. 4.3 we have defined the lensing potential

$$\psi = \frac{2}{c^2} \int_0^{\chi_S} d\chi' \frac{(\chi_S - \chi')}{\chi_S \chi'} \Phi(\boldsymbol{\theta}\chi', \chi'). \tag{4.5}$$

The convergence, κ, is related to the increase or decrease of the overall image size, whereas $\gamma_{1,2}$ parameterize the deviation of the image from a circle and are given by $\gamma_1 = 1/2(\partial_1\partial_1 - \partial_2\partial_2)\psi$, $\gamma_2 = \partial_1\partial_2\psi$. Using the previous results, the convergence ($\kappa = 1/2(\partial_1^2 + \partial_2^2)\psi$) can be written as

$$\kappa = \frac{1}{4\pi G} \int_0^{\chi_s} \frac{\nabla^2 \Phi}{\Sigma_c} \mathrm{d}\chi, \tag{4.6}$$

where $\Sigma_c = c^2\chi_S/(4\pi G\chi(\chi_S - \chi))$ is the critical surface mass density. The tangential (E mode) and cross (B mode) components of the shear are defined respectively as $\gamma_t = -\mathbb{R}[\gamma e^{-2i\phi}]$ and $\gamma_\times = -\mathbb{I}[\gamma e^{-2i\phi}]$, where $\gamma = \gamma_1 + i\gamma_2$. Notice that by factorizing out the convergence in Eq. 4.4, the deviation from identity becomes the reduced shear $g = \gamma/(1 - \kappa)$, which is the actual observable. In the weak field regime (our case), $g \simeq \gamma$ is a reasonable approximation. The tangential shear γ_t is positive in the case in which the foreground lenses are overdensities and negative in the case of underdensities, whereas the cross-component, $\gamma_\times$, is related to curl, which is not produced by weak-lensing and therefore must vanish.

In the case of axially symmetric lenses, we can write the tangential shear in terms of the convergence as

$$\gamma_t(r_\perp) = \bar{\kappa}(< r_\perp) - \kappa(r_\perp), \tag{4.7}$$

where

$$\bar{\kappa}(< r_\perp) = \frac{2}{r_\perp^2} \int_0^{r_\perp} r'_\perp \kappa(r'_\perp) dr_\perp. \tag{4.8}$$

In this work, we are interested in measuring the projected profile of voids in the thin lens approximation. Therefore, we take Σ_c out of the integral in Eq. 4.6, and integrate only in the radial bin that we regard as acting as a thin lens. By inserting the Poisson equation $\boldsymbol{\nabla}^2\Phi(\boldsymbol{\theta}\chi, \chi, a) = 4\pi G a^2 \bar{\rho}_m(a)\delta(\boldsymbol{\theta}\chi, \chi)$, the convergence then becomes

$$\kappa = \frac{3H_0^2\Omega_m^{(0)}}{8\pi G\Sigma_c} \int_{\chi_l-\Delta\chi/2}^{\chi_l+\Delta\chi/2} \frac{\delta(\boldsymbol{\theta}\chi, \chi)}{a(\chi)} \mathrm{d}\chi, \tag{4.9}$$

where $\Delta\chi$ is the bin width in comoving distance, which we consider as a thin lens. Based on this approximation, we define the ESMD in terms of the tangential shear, which is the quantity that we can measure in photometric surveys :

$$\Delta\Sigma(r_\perp) = \Sigma_c \gamma_t(r_\perp). \tag{4.10}$$

On the other hand, we can directly calculate the ESMD as

$$\Delta\Sigma(r_\perp) = \bar{\Sigma}(< r_\perp) - \Sigma(r_\perp), \tag{4.11}$$

where

$$\Sigma(r_\perp) = \frac{3H_0^2\Omega_m^{(0)}}{8\pi G} \int_{\chi_l-\Delta\chi/2}^{\chi_l+\Delta\chi/2} \frac{\delta_v(\boldsymbol{\theta}\chi, \chi)}{a(\chi)} \mathrm{d}\chi \tag{4.12}$$

and

$$\bar{\Sigma}(< r_\perp) = \frac{2}{r_\perp^2} \int_0^{r_\perp} dr'_\perp r'_\perp \Sigma(r'_\perp). \tag{4.13}$$

In the context of this work we use the void density contrast, δ_v, to calculate $\Delta\Sigma$.

4.1.2 Optimum Centering Void Finder (OCVF)

In this section we present the void finder algorithm we developed for this work. First we give the intuition and the recipe. Then we show two void statistics produced by this algorithm applied to a N-body simulation : the void density profile and the

void abundance.

4.1.2.1 Intuition and recipe

In order to have an intuition for how a void finder algorithm should be to deliver an ideal VL profile [1], we explore the possibilities of VL signals produced by voids described by an hyperbolic tangent-like profile :

$$\delta_v(r|r_v) \equiv \frac{\rho_v\left(r \mid r_v\right)}{\bar{\rho}_m} - 1 = |\delta_c| \left\{ \frac{1}{2} \left[1 + \tanh\left(\frac{y - y_0}{s\left(r_v\right)} \right) \right] - 1 \right\}, \tag{4.14}$$

where $y = \ln(r/r_v)$, $y_0 = \ln(r_0/r_v)$, $r_0 = 0.37s^2 + 0.25s + 0.89$ which is calibrated to describe voids with average density which is $\bar{\rho}_v(< r_v)/\bar{\rho}_m = 0.2$ of the average density of the Universe, where

$$\bar{\rho}_v(< r_v) = \frac{3}{r_v^3} \int_0^{r_v} dr r^3 \bar{\rho}_m(\delta_v(r) + 1), \tag{4.15}$$

s parameterises the gradient of the profile and δ_c the density contrast at the void center. This profile was first introduced in Voivodic, Rubira et al. 2020.

We choose to use this profile instead of the widely used HSW profile Nico Hamaus, Sutter et al. 2014 because the latter has more free parameters and therefore would make this exercise more complicated. However, it is important to notice that the HSW is more general than the profile 4.14, which is particularly suitable in our case.

This exercise will give us an intuition for what kind of voids we have to pursue in order to maximize certain desirable properties, namely the signal amplitude and emptiness, since emptier regions are more sensitive to the kind of physics we are interested in when working with voids, namely modifications to gravity, neutrino masses, or complementary information from underdensities which will increase the constraining power in a multi-tracer analysis.

Comparisons between different types of void finders was made in the context of N-body simulations by Cautun et al. 2018, where they found that voids found in the projected DM field have more power of distinguishing between modified gravity and GR. Also, Davies, Paillas et al. 2021 shows that the same type of voids provide the highest S/N. However, the void finder (called Tunnels) which

[1] henceforth, we use VL profile and ESMD ($\Delta\Sigma$) interchangeably

presents the desirable features in both works have overlapping between the voids. Our algorithm is intended to have similar voids but without overlapping, which is desirable in order to be able to predict the theoretical void abundance Sheth et Van De Weygaert 2004.

Figures 4.1 and 4.2 show, respectively, the void density profiles (left) and the corresponding VL profile (right) using Eq. 4.11, when one varies the parameter s (with fixed δ_c) and the density at the void center δ_c (with fixed s) in the hyperbolic tangent profile (Eq. 4.14). By Fig. 4.1 it is clear that voids with profiles that go faster from their minimum density to the average density $\bar{\rho}_m$ (smaller s) produce deeper VL profiles, with minimum at the void radius, which means that these voids are emptier inside. Figure 4.2 shows that voids less dense in their centers also produce deeper profiles.

Therefore the void finder must satisfy two criteria for defining each void in the catalogue : (i) the voids must have the smallest possible value of density at the center, δ_c (ii) their central position must be as far as possible from overdensities. The criteria (i) can be satisfied by the usage of Delaunay triangulation to define void center candidates.

The Delaunay triangulation is a set of d-symplexes (triangles for d = 2 or thetrahedrons for d = 3) $\mathbb{D}(\boldsymbol{P})$ defined in a set of discrete points $\boldsymbol{P}$, such that no point in $\boldsymbol{P}$ is inside any circum-hypersphere of $\mathbb{D}(\boldsymbol{P})$. In the case d = 2, the Delaunay triangles are those which are circumscribed by circles devoided of any point in $\boldsymbol{P}$ in their interior. In our context, the discrete set $\boldsymbol{P}$ can be any discrete tracer of LSS. By defining void positions as the centers of circum-hyperspheres, or circles circumscribed in Delaunay triangles in d = 2, we guarantee that criteria (i) is being satisfied.

The criteria (ii) means that it is not enough to define void positions in empty regions, but also that amongst all the void position candidates we must choose the one which is further away from overdensities. This idea is based on the intuition that voids are empty regions surrounded by walls, filaments and clusters, all of which are overdense structures. Therefore, there is a point for each underdense region which must be further away from these structures. Having all the candidates from the Delaunay triangulation, the point we are looking for is the one from which we can grow the largest possible circle (for d = 2) until it reaches a certain fraction of the average density of the Universe, $\bar{\rho}_v(< r_v)$, which is a free parameter of the void finder. Since all the candidates will have the same average density defined by

$\bar{\rho}_v(< r_v)$, the largest is the one which satisfies criteria (ii).

Hence the algorithm can be roughly expressed in three steps :

- Perform the Delaunay triangulation to obtain the set of points which are candidates as void positions
- Grow circles (or spheres) around them until the average density of these circles reaches a certain density threshold, specified by $\bar{\rho}_v(< r_v)$.
- The largest circle will be the first void in the catalogue and all the other voids which intercept it will be discarded. The same process will be repeated for the second largest remaining void. This process will be repeated until a void which has radius smaller than the cutting radius R_c is included in the catalogue.

The value of the cutting radius R_c is arbitrary and can be regarded as a free-parameter.

This algorithm also captures the "essence" of the excursion set theory Sheth et Van De Weygaert 2004, since it is a way of finding the voids which correspond to the trajectories which first cross the threshold for void formation.

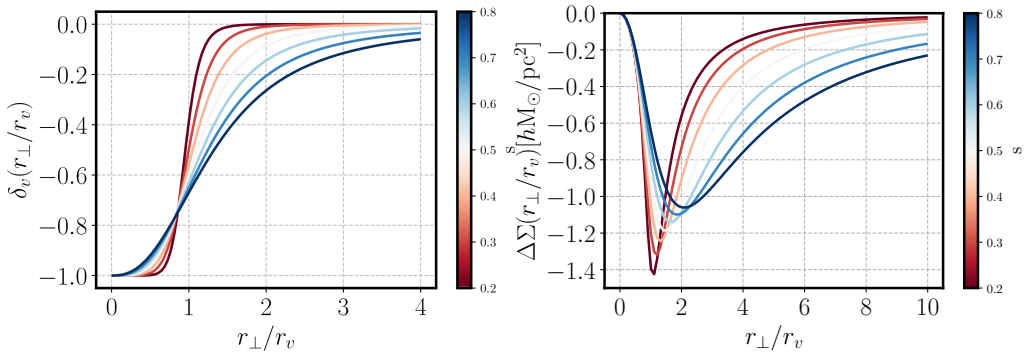

Figure 4.1 : Left : Generic void profiles of type eq. 4.14 for different choices of s, which controls the profile gradient. Right : the resulting differential surface densities, $\Delta\Sigma$ obtained through eq. 4.11. This exercise shows that deeper $\Delta\Sigma$ profiles are produced by voids with density profiles which have drastic transitions between δ_c and the average density, $\bar{\rho}_m$.

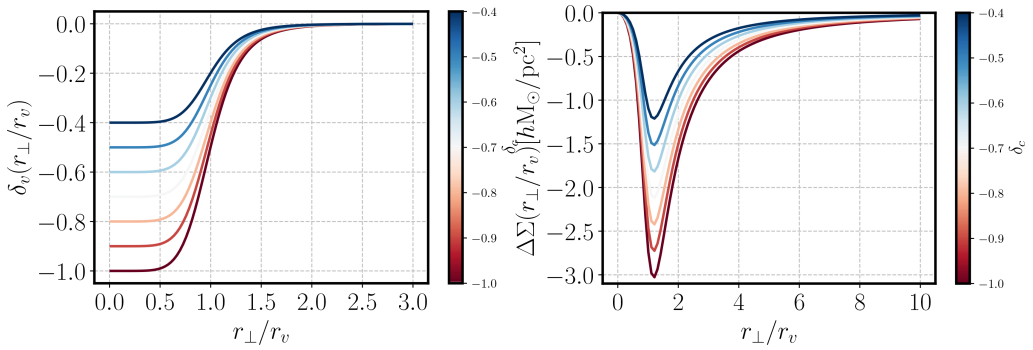

FIGURE 4.2 : Left : Generic void profiles of type eq. 4.14 for different choices of δ_c, which controls the density at the void center position. Right : the resulting differential surface densities, $\Delta\Sigma$ obtained through eq. 4.11. This exercise shows that voids with smaller values of density at its center, δ_c produces deeper differential surface densities, $\Delta\Sigma$.

Void profile and abundance

In this section, we present the resulting void profiles and abundance after applying the algorithm described in the last section to an N-body simulation. The simulation we use in this section is a $1h^{-1}$Gpc DM only box from the MultiDark suite PRADA et al. 2012. The simulation has 3840^3 DM particles which we sub sampled to have a density of 1 $h^3\mathrm{Mpc}^{-3}$. The cosmology in this simulation is $(h, \Omega_\Lambda, \Omega_m, \Omega_b, n, \sigma_8) = (0.67, 0.69, 0.307, 0.048, 0.96, 0.82)$.

Void Profile

We apply the algorithm to the simulation in its 3D version. For the purpose of visualisation, we show in Fig. 4.3 the voids found in a 2D slice of 50 h^{-1}Mpc. It is clear that voids are well fitted in underdense regions surrounded by filaments and walls, specially the largest voids. Although we don't show the same feature in the 3D field, we expect the same result, since the algorithm is exactly the same.

We can understand how voids are on average by looking at their stacked profiles, i.e $\delta_v(r') = \rho_v(r')/\bar{\rho}_m - 1$, where $r' \equiv r/R_v$ and ρ_v is the density averaged over a spherical shell at reduced distance r' to the void center. Figure 4.4 shows the measured void profiles (black dots), compared to the fit profile given by Eq. 4.14 (dashed-blue). We fit the free parameter s in each bin of radius $R_v = \{[3,4],[4,5],[5,6],[6,7],[7,8],[8,9],[9,10],[10,12]\}[h^{-1}\mathrm{Mpc}]$ obtaining

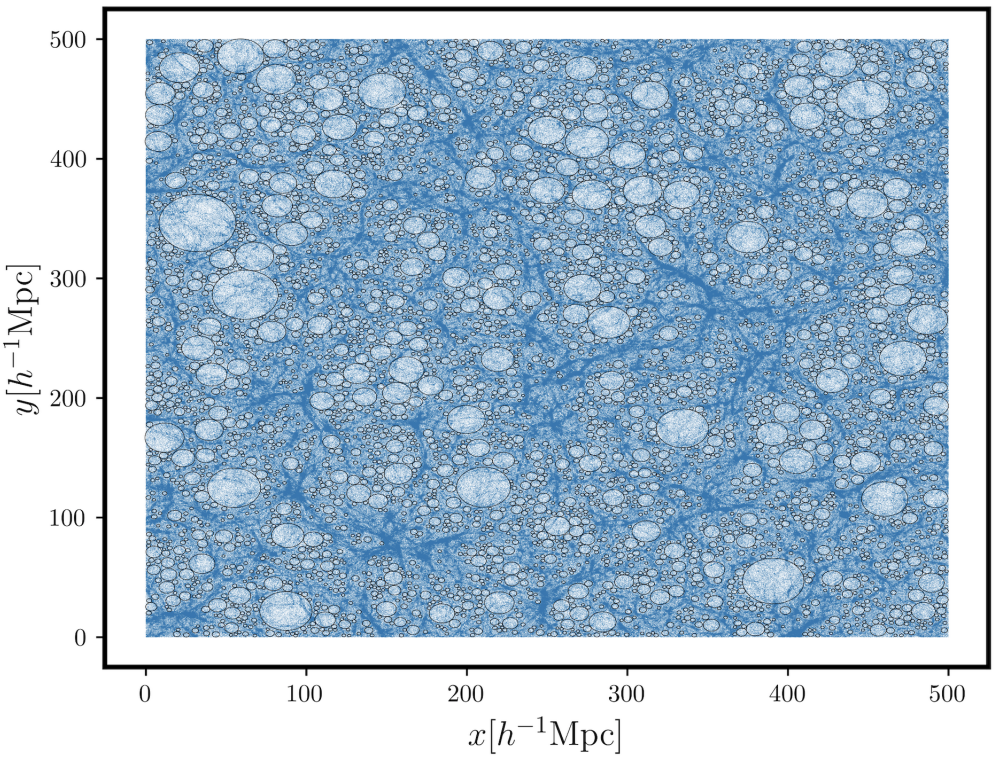

FIGURE 4.3 : Visualisation of voids found in the projection of a slice of 50 h^{-1}Mpc of DM particles.

$s = \{0.38, 0.4, 0.46, 0.48, 0.51, 0.52, 0.53, 0.54\}$, meaning that the smallest voids goes from the minimum density to $\delta_c = -1$ slightly faster than large voids. The reason for this is that smaller voids are mainly found inside overdense regions, being "voids-in-clouds" SHETH et VAN DE WEYGAERT 2004, they don't present internal structure but rather they are simply almost empty places in the process of being collapsed by the overdense surroundings. In contrast, the largest voids present more substructures (this can be clearly seen in Fig. 4.3), smoothing out their profiles.

The density profiles do not present a compensation wall such as the `ZOBOV` voids NEYRINCK 2008. This is directly related to the usage of the Delaunay approach that defines void center positions in empty places, and the `OCVF` post-processing, where only the largest void in every region is kept and all the intercepting ones are discarded. This approach selects voids better placed into the "holes" in LSS. Consequently, the averaged density profile will not present the compensation wall - a signature of overdensities nearby void center positions. It is important to notice

that these voids will not be more useful than the ZOBOV ones, but will simply be different tracers of LSS.

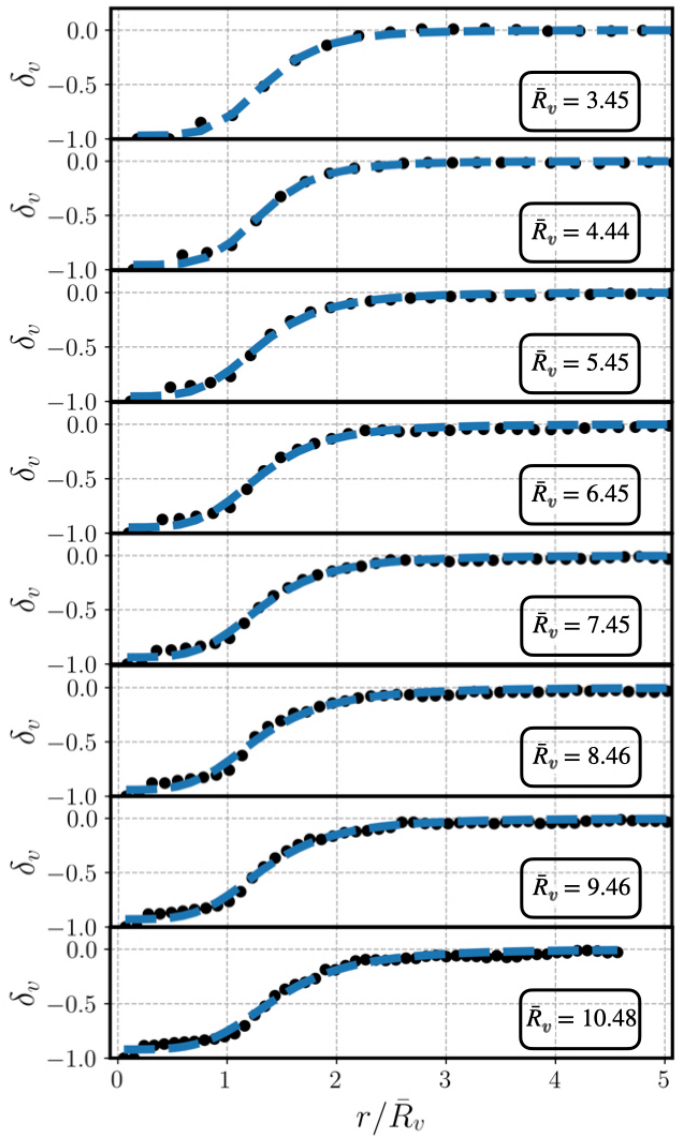

FIGURE 4.4 : The measured 3D void profiles, $\delta_v = \rho_v/\bar{\rho}_m - 1$ for each bin of radius compared to the profile given by Eq. 4.14 VOIVODIC, RUBIRA et al. 2020. The radii are given in units of h^{-1}Mpc.

Void Abundance

The void abundance (void mass function, or void radius function) is the void analog of mass function for halos, i.e., the counts of objects per bin of mass.

Several works have demonstrated its usefulness as a cosmological probe. For instance, Perico et al. 2019 and Cai et al. 2015 show the sensitivity of the abundance to modifications of gravity, Kreisch et al. 2019 and Massara, Villaescusa-Navarro et al. 2015 to massive neutrinos and Verza et al. 2019 to alternative dark energy scenarios. Recently, the first cosmological constraints using the void abundance was obtained by Contarini, Pisani et al. 2023.

The theory involved in the prediction of the void abundance is analogous to the theory developed for the halo mass function, which goes back to the pioneering work of Press-Schechter Press et al. 1974 and its further developments (e.g. Bond et al. 1991 ; Peacock et al. 1990 ; Maggiore et al. 2010). In the following, we briefly review the theory behind the prediction of the halo mass function. For a complete review, see reference Desjacques et al. 2018.

The so-called excursion set theory is based on the spherical collapse (expansion for voids), in which an isolated overdensity (underdensity) embedded into an Einstein de-Sitter background evolves and eventually collapses and virializes for halos, or experiences shell-crossing between internal shells which expand faster than outer shells for voids. Then, the linearly extrapolated value (from the initial conditions) of the density contrast for which the virialization (shell crossing) occurs is used as a threshold to define a halo (void). These values are $\delta_c = 1.686$ and $\delta_v = -2.7$, for halos and voids.

In the excursion set, the Lagrangian density field is smoothed at some scale R as

$$S(R) \equiv \sigma^2(R) = \left\langle \delta^2(x, R) \right\rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_L(k) W_R^2(k), \tag{4.16}$$

where $P_L(k)$ is the linear power spectrum and $W_R(k)$ is a smoothing function. The smoothed field $\delta(S)$ is then taken to perform a random walk, starting from $S = 0$ ($R \to \infty$). The excursion set predicts the fraction of walks which will cross the threshold δ_c for the first time in the bin of mass $[M, M + dM]$, $f(M)$. This fraction is then converted into the number density of objects per bin of mass as

$$\frac{dn}{d\ln M} \equiv \frac{d^2N}{dV\, d\ln M} = \bar{\rho}_m f(M). \tag{4.17}$$

The exact form of the function $f(M)$ depends on assumptions regarding the smoothing function $W_R(k)$. The most popular choice, which simplifies the calculations, is a sharp-k function Bond et al. 1991. For this choice, the random walk performed by $\delta(S)$ is Markovian and then the probability that a walk first crosses

the barrier δ_c is easily obtained as a solution of the Fokker-Planck equation with an appropriate boundary condition MAGGIORE et al. 2010. In this case, the halo mass function is given by

$$\frac{dn}{d\ln M} = \frac{\bar{\rho}_m}{M} f_h(\nu) \frac{d\ln \sigma^{-1}}{d\ln M}, \tag{4.18}$$

where $\nu = \delta_c/\sigma$, and the *multiplicity function* for halos f_h is defined as (in the Press-Schechter formalism)

$$f_h(\nu) = \sqrt{\frac{2}{\pi}} \nu e^{-\nu^2/2}. \tag{4.19}$$

The prediction for the void abundance is analogous and was first proposed by SHETH et VAN DE WEYGAERT 2004. The main difference from the halo case is the need for two density barriers, δ_c and δ_v. So the problem reduces to finding the fraction of walkers which first cross δ_v, without having never crossed δ_c for smaller S (larger R).

The solution obtained in ibid. is

$$\frac{dn_v}{d\ln R_L} = \frac{f_v^{2SB}(\sigma)}{V(R_L)} \frac{d\ln \sigma^{-1}}{d\ln R_L}, \tag{4.20}$$

with the void multiplicity function given by

$$f_v^{2SB}(\sigma) = 2\sum_{n=1}^{\infty} \frac{n\pi}{\delta_T^2} S \sin\left(\frac{n\pi\delta_v}{\delta_T}\right) \exp\left(-\frac{n^2\pi^2}{2\delta_T^2} S\right). \tag{4.21}$$

In the above equations $\delta_T = |\delta_v| + \delta_c$, the subscript L and the superscript 2SB stands for, respectively, the linear radius and two static barriers. This radius is the linear comoving radius of the encompassing region in the Lagrangian space, which will expand until the shell crossing (non-linear radius). At the shell crossing, the void has expanded by a factor of $\simeq 1.7$ (see BLUMENTHAL et al. 1992). The two static barriers refer to the density thresholds used in the excursion set, which are constant lines $\delta(S) = \delta_c, \delta_v$.

In reference JENNINGS et al. 2013, it was noticed that the cumulative fraction of the number of voids exceeds unity. This was interpreted as a brake in the void number conservation. The solution given by Jennings et al. ibid. was to impose the

conservation of the volume fraction :

$$V(r)dn = V(r_L)dn_L|_{r_L(r)}. \tag{4.22}$$

where the linear and non-linear radii are related as $r \simeq 1.7 r_L$. Then the void abundance becomes

$$\frac{dn_v}{d\ln R} = \frac{f_v^{2SB}(\sigma)}{V(R)} \frac{d\ln \sigma^{-1}}{d\ln R_L}. \tag{4.23}$$

In the second paper of the series Maggiore et al. 2010, the static barrier for halo formation was generalised to a stochastic barrier. The stochastic barrier captures the arbitrariness involved in the halo/void finders, as well as complications which arise from environmental conditions. These might act against or in favour of halo/void formation and hence the critical density δ_c/δ_v varies depending on the local variance.

In Sheth et Van De Weygaert 2004, an extended excursion set was first presented, i.e. the prediction for the abundance using two linear diffusing barriers :

$$\begin{aligned} f_v^{\rm 2LDB}(\sigma) = 2\left(1+D_v\right) \exp\left[-\frac{\beta_v^2\sigma^2}{2\left(1+D_v\right)} - \frac{\beta_v\delta_c}{1+D_v}\right] \\ \times \sum_n \frac{n\pi}{\delta_T^2}\sigma^2 \sin\left(\frac{n\pi\delta_c}{\delta_T}\right) \exp\left[-\frac{n^2\pi^2\left(1+D_v\right)}{2\delta_T^2}\sigma^2\right]. \end{aligned} \tag{4.24}$$

The superscript in 2LDB stands for two linear diffuse barriers and (β_v, D_v) are free parameters which describe, respectively, the slope and the diffusive coefficient of the barriers (see Voivodic, Lima, Llinares et David F. Mota 2017b for more details). Notice that we use the same slope and diffusive coefficient for the two barriers. The values of the linearly extrapolated thresholds for halo and void formation are kept fixed in ($\delta_c = 1.686, \delta_v = -2.7$).

Figure 4.5 shows the agreement between the abundance measured in the DM only simulation with size $L = 500h^{-1}$Mpc and the prediction given by the 2LDB prediction, with best fit free-parameters ($\beta_v = 0.02, D_v = 0.2$). The error bars are estimated by splitting the 1Gpc in 8 subboxes, estimating the abundance in each subbox and performing a jackknife in these 8 samples. The measured abundance is compatible with the theoretical expectation with only 2 free-parameters.

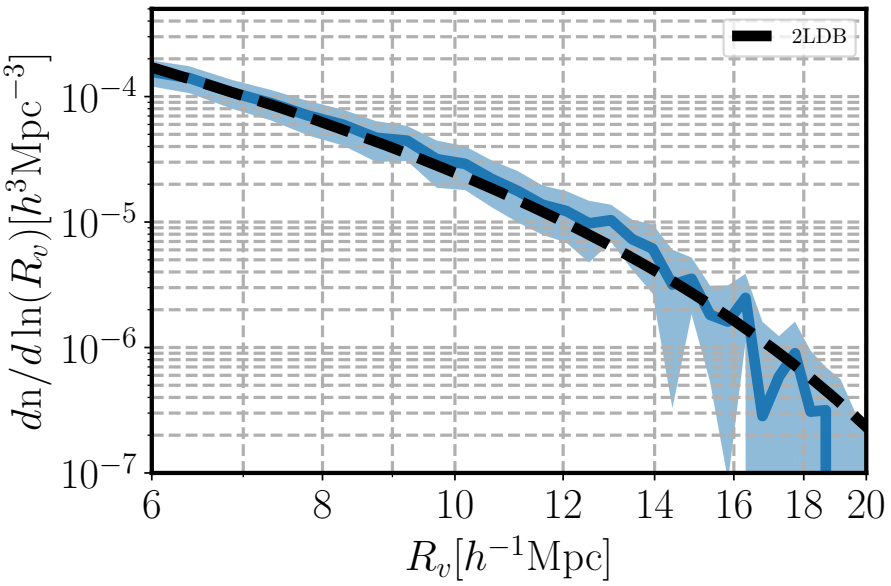

FIGURE 4.5 : The measured abundance in a DM only N-body simulation of size $L = 500h^{-1}$Mpc compared to the 2LDB model prediction using two free-parameters (β_v, D_v).

4.1.3 Void lensing on galaxies

In this section we compare our algorithm `OCVF` to `ZOBOV` (ZOnes Bordering On Voidness algorithm) NEYRINCK 2008 in the context of void lensing. This algorithm is applied through the wrapper `Revolver` Seshadri NADATHUR et al. 2019.

We apply both algorithms to the `Buzzard` mock DEROSE, Risa H WECHSLER et al. 2019 ; DEROSE, R. WECHSLER et al. 2022 ; Risa H WECHSLER, DEROSE et al. 2022 ; DEROSE, Matthew R BECKER et al. 2022, which models the observed spectroscopic redshifts of galaxies matching a DESI-like survey (DeRose et al. in prep). The simulated light-cone covers $10,313.25\ deg^2$, and contains 5,434,414 (BGS) galaxies brighter than $r = 20.2$ distributed over the redshift range $z \in [0.1, 0.3]$. These galaxies are used as lenses, i.e., we find the voids using them as tracers. The source galaxy catalogue is modelled to match the photometric redshifts of a DES-like survey, occupying the same surface area, with density of 4.4 galaxies/arcmin^2 in the redshift range $z \in [0.5, 1.5]$.

The main difference between the `ZOBOV` and `OCVF` is that the latter is Delaunay based, whereas the former is Voronoi based. Moreover, here we use the 2D version of `OCVF` and `ZOBOV` finds voids in 3D.

We chose to show the comparison between the 2D `OCVF` because former works

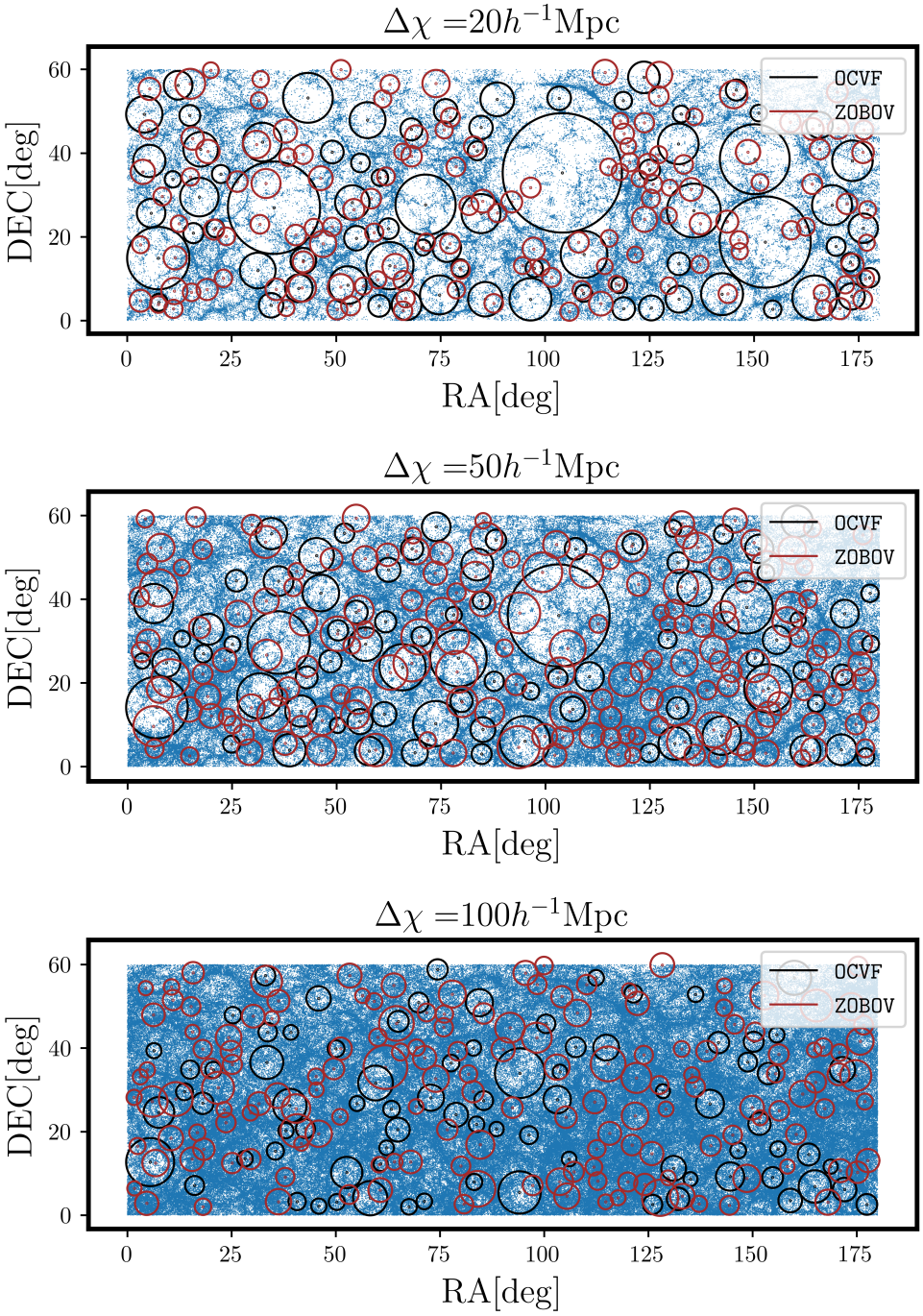

FIGURE 4.6 : Visualisation of OCVF and ZOBOV voids plotted over BGS galaxies, in slices of widths $\Delta\chi = 20$, 50 and 100$[h^{-1}\text{Mpc}]$. The ZOBOV voids were chosen to have their central positions well within the slices.

have indicated that voids found in projected fields are more efficient at measuring the weak-lensing signal by voids Davies, Paillas et al. 2021 and more sensitive to modifications to gravity Cautun et al. 2018. We confirmed that by using the same algorithm to find voids in the 2D and 3D fields, the former provides a signal with larger amplitude. The reason for this is that underdensities in the projected field correspond to anisotropic underdensities in the 3D field, with major-axis aligned with the line-of-sight, as shown in 4.1.7.1. As a consequence, the photon's path is guaranteed to be deviated by underdensities over a greater distance. In contrast, 3D voids are, in general, surrounded by overdensities, which partially or completely erase the signature of overdensities in the photon's path.

This comparison aims to highlight the qualitative differences between these two approaches, which evidences that unlike other statistics, the choice of void finder and how it is applied to data (in the 3D or 2D field, for instance) can drastically change the estimated observable ($\Delta\Sigma$ in this case).

4.1.4 Void catalogues

The `OCVF` is applied to the light-cone by splitting it in slices of equal widths along the line of sight. We use the widths $\Delta\chi = 20, 50, 100h^{-1}$Mpc. Therefore, voids are initially detected as underdensities in projected slices, which in fact correspond to voids in the 3D field as we will discuss in section 4.1.7.1. The `ZOBOV` algorithm is applied to the 3D galaxy field.

The void radius distribution in each version of `OCVF` (slice width), as well as in `ZOBOV` are shown in Figure 4.7. Table 4.1 shows the total number of voids for each algorithm. We choose to apply a radius cut in $R_v = 10h^{-1}$Mpc in all void catalogues. This choice is related to the minimum resolution we have in the galaxy field, but is rather conservative, especially for the `OCVF` sample, which has the vast majority of voids with radii smaller than $10h^{-1}$Mpc. However, we believe that these smaller voids are predominantly spurious, i.e. does not correspond to real underdensities in the DM field or are mostly voids-in-clouds.

The `OCVF` voids found in larger slices are significantly less numerous than those found in smaller slices, however, if we normalize the curves to be integrated to unity, all histograms have the same shape.

Figure 4.6 shows the visualisation of `ZOBOV` and `OCVF` voids plotted over the BGS galaxy distribution. The chosen `ZOBOV` voids were those with central positions as

close as possible from the bin center we used to find the OCVF voids, in a way that these voids must correspond to the underdensities in this projected slice. By visual inspection it is possible to notice that the OCVF voids are better fitted into the regions with less galaxies in the projected field, whereas ZOBOV voids sometimes correspond to those regions but frequently are placed into overdense regions. This is expected, since OCVF finds voids in the projected slices and ZOBOV is applied to the 3D field. It shows that three-dimensional voids are sometimes erased in the projected field and, as a consequence, won't be detectable through lensing, or will simply add noise to the estimated VL profile.

4.1.5 The $\Delta\Sigma$ Estimator and the TL approximation

The ESMD estimator is given by

$$\widehat{\Delta\Sigma}(r_\perp/\mathrm{R_v}) = \frac{\sum_{ij} w_{ij}\Sigma_{c,ij}\gamma_{t,ij}(r_\perp/R_v)}{\sum_{ij} w_{ij}}, \tag{4.25}$$

where $\gamma_{t,ij}$ is the tangential shear in the source galaxy i due to the void j. For each pair, the weights that minimize the variance of the signal are $w_{ij} = \Sigma_{c,ij}^{-2}$ Sheldon et al. 2004, and the critical mass density is defined as

$$\Sigma_{c,ij} = \frac{c^2}{4\pi G}\frac{\chi(z_i)}{\chi(z_j)(\chi(z_i)-\chi(z_j))}. \tag{4.26}$$

The tangential shear is calculated from the shear components $\gamma_{1,2}$ (obtained in DeRose, Risa H Wechsler et al. 2019 through ray-tracing) and the angle ϕ between the void and source positions.

The estimated ESMD is expressed in terms of the reduced perpendicular distance to the line-of-sight $r_\perp/R_v$ because voids with different sizes present very similar profiles in reduced coordinates. This is what we mean by stacked void profile.

The estimator 4.25 is based under the assumption that the shear $\gamma_{t,ij}$ for *all* voids can be obtained by assuming the thin lens approximation, in which the net effect on the photons path will depend only on the target structure which can be regarded as being contained in a single plane, i.e. ignoring the line-of-sight direction. In other words, the remaining structures when subtracting the void will have zero net effect on the photons path and the only effect can be computed by projecting the density field of the voids, as in eq. 4.9. In this case the equality in 4.25 is true.

However, it is possible that individual voids extend to distances larger than the maximum distance which does not allow us to write that $\Delta\Sigma = \Sigma_c\gamma_t$. In this case, a systematic error will be introduced in the estimator 4.25.

4.1.6 Void finder comparison

In Figure 4.8, we show the measurement of the ESMD performed using voids found in the 3D-galaxy field with the `ZOBOV` algorithm, as well as voids found in projected fields with the `OCVF` algorithm. The `ZOBOV` voids are split into two sub-samples, $R_v \in [10, 15]$ and $R_v \in [10, 25]$ (h^{-1}Mpc). The reason why we use these two sub-samples and $25h^{-1}$Mpc as maximum radius is justified in section 4.1.7. The `OCVF` voids are presented in three samples, each of which corresponds to the algorithm applied to slices of widths $\Delta\chi = 20, 50, 100h^{-1}$Mpc. The right and left plots show, respectively, the tangential ($\Delta\Sigma_t$) and cross ($\Delta\Sigma_\times$) components.

The tangential component presents several differences. Firstly, the `ZOBOV` voids present a compensation wall, whereas the `OCVF` does not. The compensation wall is a direct consequence of the same feature in the void's 3D density profile, which means that these voids are closer to overdensities, whereas `OCVF` voids are far enough from overdensities to not have correlation with them. Another notorious difference is that `ZOBOV` voids are shallower. There are two reasons for this. The first is the existence of a compensation wall, which produces positive tangential shear, γ_t. The second reason can be intuitively seen in Figure 4.6 : not all `ZOBOV` voids correspond to underdensities in the projected field, whereas the voids found in projected fields are guaranteed to not be correlated with overdensities along the line-of-sight. Moreover, larger projected slices tend to select underdensities aligned with the line-of-sight. The `OCVF` voids are presented in three samples, each of which corresponds to the algorithm applied to slices of widths $\Delta\chi = 20, 50, 100h^{-1}$Mpc. These voids correspond to 3D underdense structures which have their major axis aligned with the line-of-sight (see Figure 4.10). In the Appendix **B** we briefly discuss that not only the voids in the projected field present an anisotropic 3D density profile, but also the subsample voids found using the 3D version of the same algorithm (`OCVF`), therefore following the theoretical abundance (Figure 4.16), also present anisotropic 3D density profiles. This subsample is chosen using the distances from the 3D voids to the projected voids which present correlation between them (Figure 4.14). This result indicates that what we really find by applying the void

finder in the projected field are combinations of 3D voids which present intrinsic alignment between them.

The cross-components must be consistent with zero. It is almost always satisfied, except in some cases for small $r_\perp/R_v$, where it is slightly below zero. The same feature also appears in the tangential component (it must go to zero as $r_\perp/R_v \to 0$). This indicates a systematic error for small $r_\perp/R_v$ that we do not comprehend.

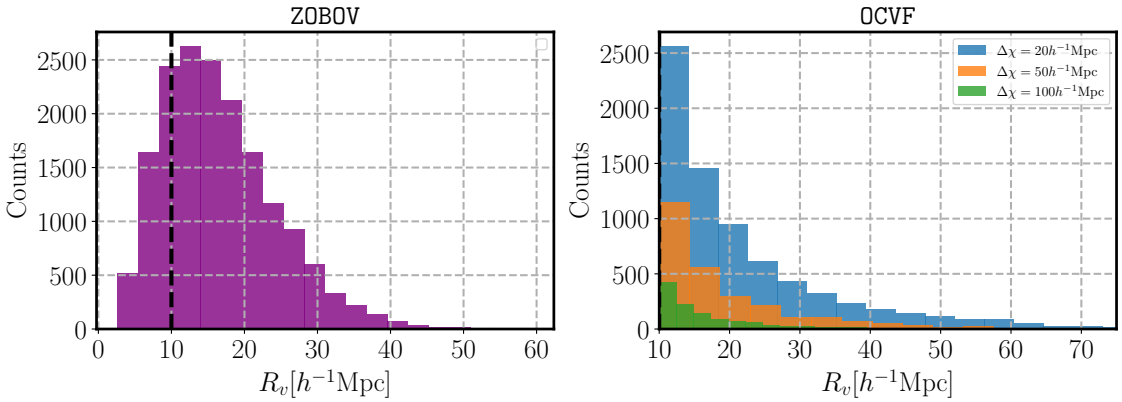

FIGURE 4.7 : Left : Radius distribution of ZOBOV voids. Right : Radius distribution of OCVF voids

In Figure 4.9 we compare the cumulative signal-to-noise (S/N) of the tangential component, defined as

$$(S/N)^2(< r_{\perp,k}) = \sum_{i\leq k, j\leq k} \Delta\Sigma_{t,i} C_{ij}^{-1} \Delta\Sigma_{t,j}, \tag{4.27}$$

where C_{ij} is the covariance matrix estimated through a void-by-void jack-knife.

The left plot shows the cumulative S/N for each sample, whereas the right plot shows the same quantity normalised by the number of voids in the corresponding sample (Table 4.1). The cumulative S/N is quite similar for all samples, except for the ZOBOV sample with maximum radius of $15h^{-1}$Mpc. The normalised S/N is significantly higher for the OCVF samples corresponding to the slices widths of $\Delta\chi = 50, 100h^{-1}$Mpc. This result is expected : the larger is the slice width, the larger is the photon's path which corresponds to underdensities, i.e., less contamination from overdensities.

It should be stressed that this comparison does not assess which sample is more useful to constrain cosmology. Indeed, larger S/N does not necessarily mean more sensitive to cosmological parameters or modifications to gravity. However, we can

say that measuring the lensing signal by voids using projected slices and the OCVF will produce a signal which tells more about underdensities than the one obtained using ZOBOV voids. By thinking of voids as being the most underdense tracers of LSS and assuming that tracers with different biases carry complementary information (see e.g. Mergulhão et al. 2022 ; Abramo et al. 2013), then the OCVF profiles in Figure 4.8 have biases more negative than the ZOBOV profiles.

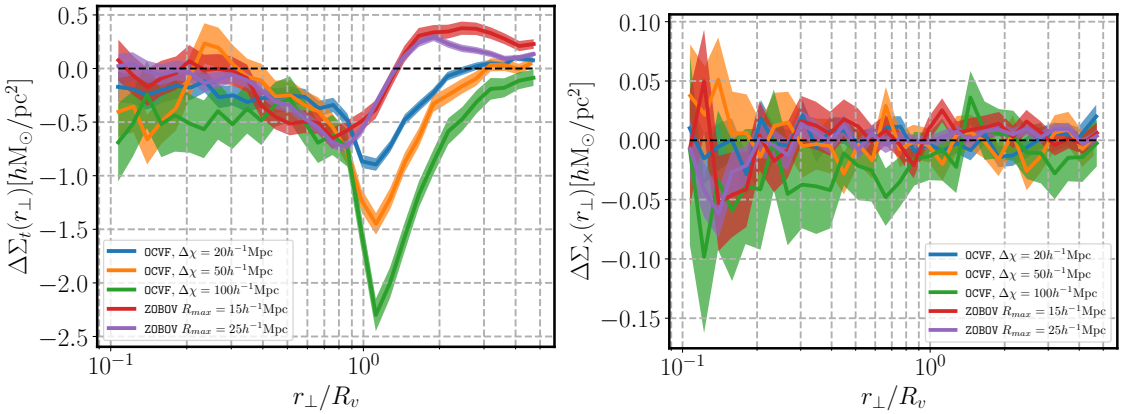

Figure 4.8 : Left : Comparison between the $\Delta\Sigma_t$ measurements performed using the OCVF in slices of width $\Delta\chi = 20,\ 50,\ 100\,h^{-1}$ Mpc (blue, orange, and green) and ZOBOV in the bins of radius $[10, 15]h^{-1}$Mpc (red) and $[10, 25]h^{-1}$Mpc (purple). Right : The same comparison for the cross component $\Delta\Sigma_\times$.

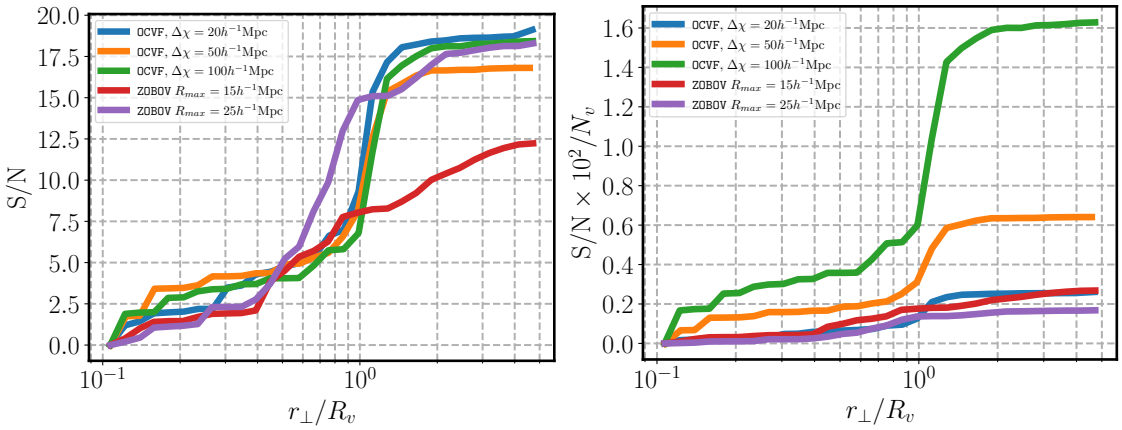

Figure 4.9 : Left : Comparison between the cumulative S/N on $\Delta\Sigma_t$ obtained using the OCVF on slices of width $\Delta\chi = 20,\ 50,\ 100\,h^{-1}$ Mpc (blue, orange, and green) and ZOBOV (red) in the redshift bin $0.1 < z < 0.3$. Right : S/N normalized by the number of voids

Void Finder	$N_v(0.1 < z < 0.3)$
`OCVF` ($\Delta\chi = 20\ h^{-1}$Mpc)	7321
`OCVF` ($\Delta\chi = 50\ h^{-1}$Mpc)	2623
`OCVF` ($\Delta\chi = 100\ h^{-1}$Mpc)	1131
`ZOBOV`	13438

TABLEAU 4.1 : The total number of voids for each void finder in the redshift bin $z \in [0.1, 0.3]$.

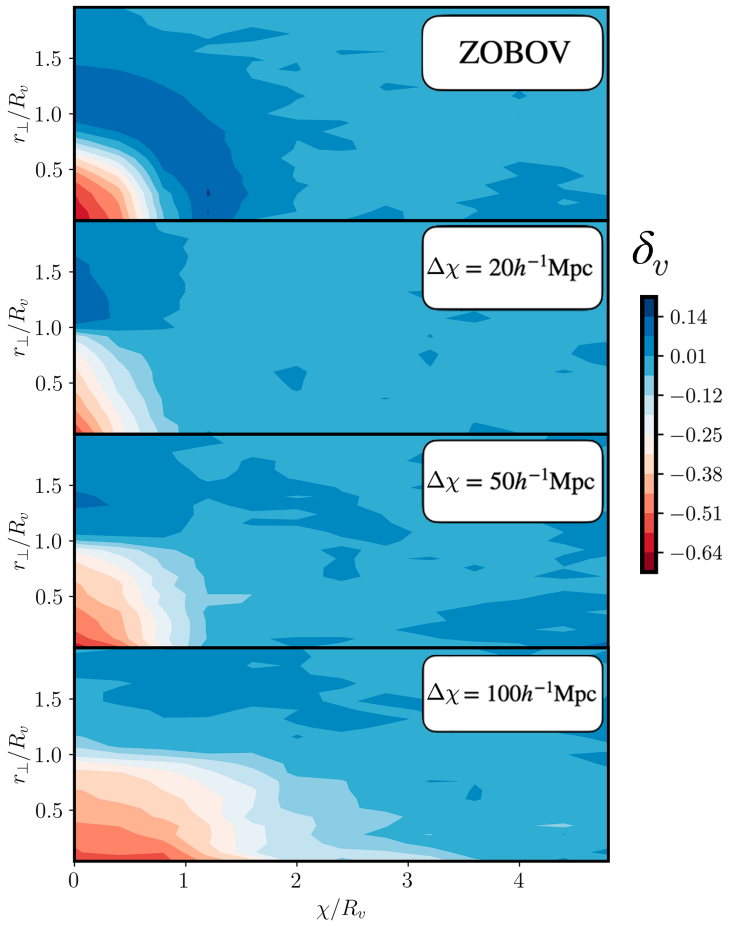

FIGURE 4.10 : Void DM profiles as a function of radial comoving distance, χ/R_v (centered at the void position) and the perpendicular distance to the line-of-sight, $r_\perp/R_v$. The top panel shows profiles estimated with `ZOBOV` voids found in the 3D galaxy distribution, and the profiles labeled by $\Delta\chi = 20, 50, 100h^{-1}$Mpc are `OCVF` voids found in slices of labeled sizes.

4.1.7 Numerical interpretation

Voids are peculiar tracers of LSS and it is not clear whether an analytic approach, based on perturbation theory, or effective field theory of large scale structure would be able to predict their density profiles, i.e., the cross-correlation void-tracer. The main reason for this impossibility is that voids are not uniquely defined, imposing a puzzle on how the condition related to the void center definition will be incorporated in the growth equation, in the case of linear perturbation theory. Therefore, a promising approach to extract this cosmological information is to take their profiles directly from N-body simulations and apply some emulation technique to predict their profiles for a set of cosmologies. In the context of VL, we have the advantage that we can directly use a DM only box to perform the emulation, since this observable is probing directly the total matter field[2].

In this section we pave the way for this kind of approach. We use the DM particles of the same realisation we used in the previous section and estimate the DM profiles of the same voids we found in the galaxy field. Then we use these void DM profiles to check the consistency between the $\Delta\Sigma$ profiles we estimated in the previous section (through the shear of background galaxies), and the same quantity, but estimated directly from the DM field. This consistency test will show whether we are really having access to the DM profiles of these voids through weak-lensing. This consistency test is crucial if we aim to use the VL profile to do precision cosmology in the near future. As we will show, this consistency is not trivially given in the context of voids, as it is in the context of halos.

4.1.7.1 VL and Void Dark Matter profiles

Figure 4.10 shows the stacked void profiles as a function of parallel and perpendicular distances to the line-of-sight, estimated using the DM particles of the Buzzard mock, i.e., the cross-correlation between void centers and DM particles for voids found in the BGS galaxy field :

$$\delta_v(r'_\perp, \chi') = \frac{1}{N_v} \sum_i^{N_v} \frac{n_p^i(\Delta r'_\perp, \Delta\chi')}{\langle n_p \rangle} - 1, \tag{4.28}$$

[2]Assuming that we can calibrate the density threshold in the void finder to find the voids in the tracer field which correspond to voids in the DM field found using another density threshold. That is, the map between the thresholds used in the DM and tracer fields.

where $r'_\perp \equiv r_\perp/R_v$, $\chi' \equiv \chi/R_v$ and $\Delta r'_\perp(\chi)$ denotes a bin in $(r'_\perp, \chi')$. Since we are estimating these profiles in configuration space, we expect that they will be isotropic. However, the profiles for increasingly larger slices become increasingly anisotropic. Since only voids aligned with the line-of-sight are detected as voids in the projected field, the resulting stacked profile will be anisotropic and might contain some information encoded in the void ellipticities of voids defined in the 3D distribution (see Appendix 4.1.8). This can be directly accessed with void lensing measurements. The relation between anisotropic and isotropic profiles is the topic of an ongoing work.

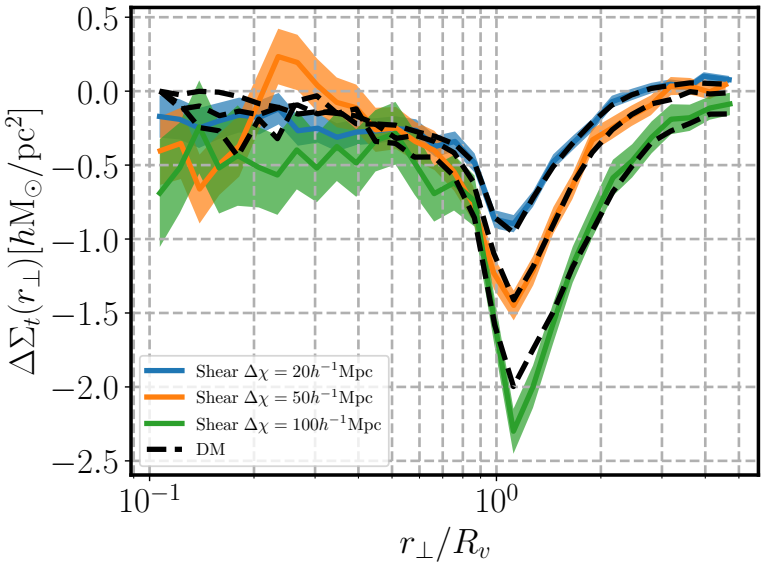

FIGURE 4.11 : Comparison between $\Delta\Sigma_t$ as measured through the shear (blue, orange, and green) and directly using the DM particles (dashed black).

4.1.7.2 Consistency between shear and Dark matter density profile around voids

In this section, we check the consistency between the $\Delta\Sigma_t$ profile of voids measured from the shear, i.e., $\Delta\Sigma_t = \Sigma_c\gamma_t$ and the one directly calculated using the DM density profiles of voids presented in figure 4.10. We take the same realisation of the density field, as traced by BGS galaxies. Therefore, we compare the observable we have access through observations against the same observable but computed using the DM field, which we do not have access to in real observations. In the case

where this consistency test is successful, we know exactly what we would observe in real data, given that the assumption we are using in simulations are well calibrated w.r.t real data, namely, the galaxy bias and possible systematic effects such as intrinsic alignment and survey mask. Before proceeding, it is important to mention the reasons why the two quantities might differ :

- The impact of the source distribution : In the estimator 4.25, the weights and the γ_t depend on Σ_c which contains information about the source distribution. By predicting $\Delta\Sigma_t$ only in terms of the void DM profile, without saying anything about the source distribution, it is not clear whether we can recover the same quantity than the one estimated.

- The impact of the thin lens approximation : To quote the results in terms of $\Delta\Sigma_t$ we are assuming that the void profile, or everything that will have an impact on the average shape of background galaxies is contained in a thin lens (see eq. 4.9). Since voids can have radius as large as $\simeq 100h^{-1}$Mpc and, furthermore, might have correlations with overdensities beyond the void radius, it is not clear whether the thin lens assumption still holds.

In order to calculate the $\Delta\Sigma_t$ profiles from the 3D void profiles in figure 4.10, we will exclude small scales to avoid resolution effects by using the annular differential surface density (ADSD) :

$$\begin{aligned}\Upsilon\left(R|R_0\right) \equiv & \Delta\Sigma_t(R) - \frac{R_0^2}{R^2}\Delta\Sigma_t\left(R_0\right) \\ = & \frac{2}{R^2}\int_{R_0}^{R} \mathrm{d}R' R'\Sigma\left(R'\right) - \frac{1}{R^2}\left[R^2\Sigma(R) - R_0^2\Sigma\left(R_0\right)\right],\end{aligned} \tag{4.29}$$

where R_0 is the cutting scale. For sufficiently small R_0, Υ reduces to $\Delta\Sigma_t$. We checked that the estimated profile does not depend on the particular choice of R_0, for $R_0 \leq 0.5$ in reduced coordinate $(r_\perp/R_v)$.

Since we are working with stacked void profiles, $\Delta\Sigma$ is proportional to the void radius, which comes from the Jacobian when transforming the integral from $r'_\perp$ to $r_\perp/R_v$ coordinate. Thus,

$$\Sigma(r'_\perp) = R_v\Sigma(r_\perp), \tag{4.30}$$

where $\Sigma(r_\perp)$ is given by eq. 4.12 and, consequently,

$$\Delta\Sigma_t(r'_\perp) = R_v \Delta\Sigma_t(r_\perp). \tag{4.31}$$

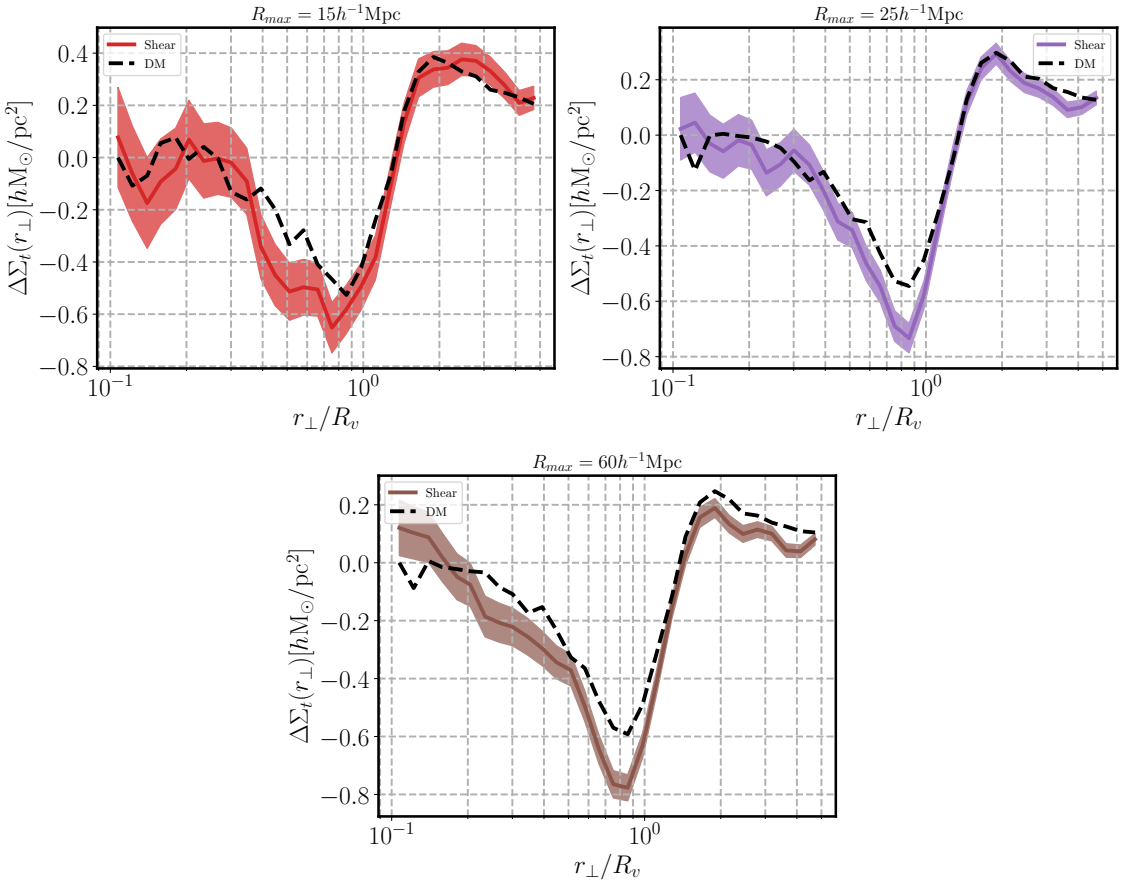

FIGURE 4.12 : The consistency check between the $\Delta\Sigma_t$ estimated through the shear of background galaxies and the same quantity estimated directly from the DM profiles of the same voids. Respectively, the top-left, top-right and bottom plots show the consistency test for ZOBOV voids with radii in the ranges $[10, 15]$, $[10, 25]$ and $[10, 60]$ in units of h^{-1}Mpc.

Figure 4.11 shows the results of the same $\Delta\Sigma_t$ profiles as shown in figure 4.8, through the estimator 4.25, against the $\Delta\Sigma_t$ estimated using the DM particles around the voids. In this estimation we project the void profiles using a bin width $\Delta\chi'$:

$$\Sigma(r_\perp) = R_v \int_{\chi'_l - \Delta\chi'/2.}^{\chi'_l + \Delta\chi'/2.} \frac{d\chi'}{a(\chi')} \delta_v(r'_\perp, \chi'). \tag{4.32}$$

and then compute the ADSD signal using eq. 4.29.

We obtain consistency between shear and DM profiles for the OCVF samples computed in slices of widths $\Delta\chi = 20, 50h^{-1}$Mpc. However for the sample with $L = 100h^{-1}$Mpc, the two profiles become inconsistent at reduced radius $r'_\perp \sim 1$. We obtain the same trend for the ZOBOV voids (Figure 4.12). The subsample in the radius bin [10, 15] (h^{-1}Mpc) is quite consistent, whereas the subsamples including larger voids, namely, [10, 25](h^{-1}Mpc) and [10, 60](h^{-1}Mpc), present increasing inconsistencies.

The trend with both void finders is, the larger is the stacked void profile, the larger is the inconsistency between shear and DM profiles. Furthermore, the void size in the direction perpendicular to the line-of-sight seems to be unimportant, since the normalised radius distribution of OCVF voids is the same amongst the different bins widths. We conclude that what really matters for this consistency is the void size along the line-of-sight, which indicates that the thin lens approximation might be broken in the cases in which the correlations between the void and its surroundings extend beyond a certain limit.

In the Appendix 4.1.8 we show an example of how a break of the thin lens approximation affects the VL profiles. We use generic analytical void profiles and calculate $\Delta\Sigma$ with and without the thin lens approximation. The wrong assumption of the thin lens approximation tends to underestimate the VL profile, specially around $r_\perp/R_v \simeq 1$, which is exactly what we see in the DM profiles of Figures 4.11 and 4.12. The source of the inconsistency might be, then, the wrong assumption of the thin lens approximation in the estimator 4.25.

4.1.8 Conclusions

In this work we have studied the VL profile in the context of galaxy mock. First, we proposed a new void finder algorithm which is particularly designed to deliver voids with deep VL profiles. Then, we apply this algorithm to the galaxy mock and contrast it with the widely used algorithm in literature, ZOBOV. The latter has been used in previous measurements of the VL profile in real data and provided a higher S/N compared to an algorithm similar to ours Fang et al. 2019.

We show that, compared to ZOBOV, voids found in projected slices using our algorithm can provide a higher S/N per void, with a deeper VL profile. These voids correspond to combinations of voids in the 3D DM field which present intrinsic

alignment between them, suggesting that the VL profile might be sensitive to the void intrinsic alignment. This opens some questions, namely : (i) is the void intrinsic alignment connected to tidal fields in LSS and, therefore, to cosmology ? (ii) How sensitive is the VL profile to the intrinsic alignment ?

Then we checked the consistency between the VL profile as estimated through the shear of background galaxies and the same quantity as estimated directly from the DM density profiles of voids. This consistency test has never been made (up to the knowledge of the present authors) and it is crucial for future cosmological analysis using VL profile as an observable. Unlike halos, voids have density profiles that extend over hundreds of h^{-1}Mpc, which might break the assumption of voids being contained in a thin slice, which is a basic assumption when estimating the VL profile from the shear of background galaxies. Furthermore, voids are much less numerous than halos and therefore residual contributions from structures along the line-of-sight might not be averaged out, as it easily is in the context of halos.

This consistency test shows that voids with larger sizes along the line-of-sight present inconsistencies between the shear and DM VL lensing profiles, suggesting that the thin lens approximation assumed in the estimator is not appropriate in the case of voids in general. For `ZOBOV` voids smaller than $15h^{-1}$Mpc and `OCVF` found in projected slices smaller than $50h^{-1}$Mpc the shear and DM VL profiles are consistent.

In future works, we plan to further understand the relation between the VL profile around voids in the projected field and the intrinsic alignments of voids in the 3D field. Also, we plan to model the VL profile with dependency on the cosmological parameters. This work paves the way for a trivial numerical prediction : since we have shown that for a suitable choice of void definition, the shear and DM VL profile are consistent, emulation techniques which are becoming widely used in cosmology can be used to interpret real data observations. Is it also important to have an analytical prediction of the VL observable, which is the subject of an ongoing work by the present author.

A How the thin lens approximation affects the VL profiles

In this appendix we explore the limits of the thin lens approximation, which is relevant in the case of voids, given their large size.

As pointed out in section 4.1.7, the matching between measured and predicted

ESMD depends on whether we can consider voids as thin lenses in between the source and the observer or not, i.e., whether we can write the convergence as eq. 4.9. To test this assumption we use a void profile with similar shape to the stacked profile produced by `OCVF` to compute $\Delta\Sigma$ with and without the thin lens approximation. We use a "worst case scenario" in which the void position is too close to the source with redshifts $z_l = 0.48$ and $z_s = 0.5$, respectively for the void and the source, and vary the void radius $R_v = (50, 100, 150)h^{-1}$Mpc. The combination of void radius and distance between the source and the void are the variables that control whether the thin lens is a good approximation or not.

Figure 4.13 shows the predictions for the ESMD for different void radius (left) and the relative difference between the same quantity calculated with and without the thin lens approximation (right). This exercise shows that for voids with radius of $50h^{-1}$Mpc the difference is below $\simeq 5\%$.

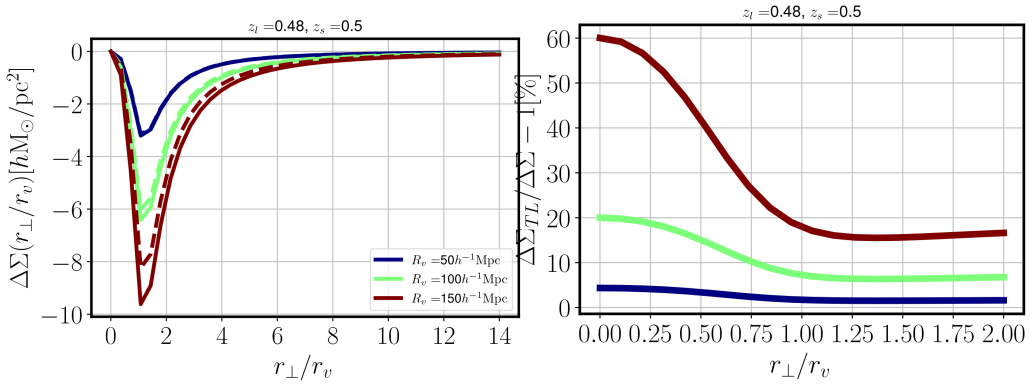

FIGURE 4.13 : Left : The ESMD profiles for different void radius with (solid) and without (dashed) the thin lens approximation. Right : The relative difference between the ESMD with and without thin lens approximation.

B The void intrinsic alignment around projected voids

In this appendix we show some preliminary results about the shapes of 3D voids around the voids found in the projected slices that we use in this work. Both void samples were created using the `OCVF` algorithm.

We use the same simulation we used in section 4.1.2 to find both void samples. The left plot in Figure 4.14 shows the correlation between a sample of voids found in projected fields of width $\Delta\pi = 50h^{-1}$Mpc with radii in the range $8 < R_v^b <$

$15h^{-1}$Mpc and the 3D voids in different radius bins. The right plot shows the stacked profile of the 3D voids as a function of the perpendicular, σ, and parallel, π, distances to the line of sight.

The correlations shows that voids in 3D and 2D with similar sizes present stronger correlations. The stacked profiles of the 3D voids are clearly anisotropic. This shows that 3D voids around voids found in the projected slices present intrinsic alignment between them, or, equivalently, that the voids found in projected slices are actually the combination of 3D voids, which are aligned between them.

A more detailed study of the correlations and the statistical relation between voids found in projected slices and 3D voids is the subject of ongoing work.

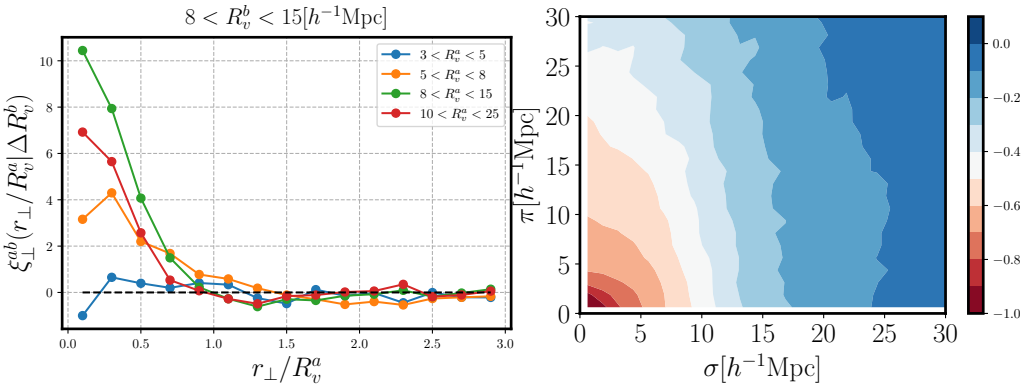

FIGURE 4.14 : Left : Correlation between voids found in the projected field, labelled by "b" and voids found in the 3D field, labelled by "a". Right : the stacked profile of 3D voids which are at a distance from the projected voids which presents non-zero correlation, as a function of the parallel, π, and perpendicular, σ, distance to the line-of-sight.

4.2 The Void-Lensing model

4.2.1 Introduction

So far, we have made a consistency check between the excess surface mass density (ESMD) that we infer through the shear of background galaxies around voids and the real quantity that we measure directly from the void profile. Have found that the two quantities are consistent under some conditions. The next natural step is to learn how to extract cosmological information from this signal, i.e., to have an interpretation of this signal

We've seen that 2D voids are more suitable for measuring the ESMD. Therefore, it would be useful to have a model for them. Since we don't have so far an analytical prediction for the 3D void profile in the literature, it is not possible to have an analytical interpretation for their ESMD. The 2D voids provide an opportunity in this regard. The opportunity comes from the fact that the 2D voids must be a combination of 3D voids projected along the line-of-sight. Since the 3D voids are not equally distributed across the Universe, it is expected that this quantity depends on the 3D void abundance, for which we do have an analytical prediction. Unfortunately, this prediction will, evidently, also depend on the 3D profiles themselves.

This section is dedicated to the model which combines the 3D void profiles and their abundance to produce the projected profile we measure. This is a work in its infancy and more effort is required to transform it in an useful tool for extracting cosmological information from the ESMD.

4.2.2 The relation between 2D and 3D voids

The most general relation between the density profiles of 2D and 3D must be a functional of the 3D profiles, i.e., the 2D voids must be the combination of 3D voids, since the only way of having underdensities in the projected field, is to have them in the 3D field. Therefore, the most general statement about this relation is

$$\delta_{2D}(r_\perp|R_{2D},\Delta_{2D},\Delta_{3D}) = \mathcal{F}[\delta_{3D}(R_{3D},\Delta_{3D})|R_{2D},\Delta_{2D}], \tag{4.33}$$

where Δ_{3D} and Δ_{2D} are, respectively, the density thresholds used to find the 3D and 2D voids.

To find the form of this functional is the goal of this work.

The trivial idea for the form of the above functional, is that the definition of a 2D void must be a point in a region dominated by 3D voids. Therefore, we expect that 3D void positions must be distributed around the 2D center. Moreover, the contributions from different void radii must be weighted by the abundance. Based on these ideas, I suggest the following form for the functional :

$$\delta_{2D}(r_\perp|R_{2D},\Delta_{2D},\Delta_{3D}) = \frac{1}{N}\int dR_{3D}\,\frac{dn_v}{dR_{3D}}(R_{3D}|\Delta_{3D}) \times \int dx_\perp d\phi\, P(x_\perp|R_{3D},R_{2D},\Delta_{3D},\Delta_{2D}) \times \int dr_\parallel \delta_{3D}(|\boldsymbol{r}_\perp-\boldsymbol{x}_\perp|,|r_\parallel-x_\parallel||R_{3D},\Delta_{3D}) \tag{4.34}$$

where $\boldsymbol{r}_\perp$ and $\boldsymbol{x}_\perp$ are, respectively, the distance to the 2D void center and the position of the 3D center in the plane perpendicular to the line-of-sight. Since the stacked profile is anisotropic, we take $\boldsymbol{r}_\perp$ aligned with the $x-$axis and the coordinate system centered at the 2D void center. Therefore, the distance between the 3D void center and the position to which the 3D void contributes is

$$|\boldsymbol{r}_\perp-\boldsymbol{x}_\perp| = \sqrt{(r_\perp-|\boldsymbol{x}_\perp|\cos(\phi))^2+|\boldsymbol{x}_\perp|^2\sin^2(\phi)}. \tag{4.35}$$

Figure 4.15 shows how the model works. At each $r_\perp$ a void at distance $|\boldsymbol{r}_\perp-\boldsymbol{x}_\perp|$ w.r.t. $r_\perp$ contributes with a quantity which is given by the weighted integral of the projected 3D profile

$$\delta_\Sigma(|\boldsymbol{r}_\perp-\boldsymbol{x}_\perp|) \equiv \int dr_\parallel \delta_{3D}(|\boldsymbol{r}_\perp-\boldsymbol{x}_\perp|,|r_\parallel-x_\parallel||R_{3D},\Delta_{3D}). \tag{4.36}$$

The weights are

- $\frac{dn_v}{dR_{3D}}$ is the usual void abundance, which depends on the linear power spectrum. Therefore, this is the main ingredient of the model, since it carries cosmological information.
- $P(x_\perp,x_\parallel|R_{3D},R_{2D},\Delta_{3D},\Delta_{2D})$ is the weight function that a 3D void is found at a distance $\sqrt{x_\parallel^2+|\boldsymbol{x}_\perp|^2}$ from the 2D void position.

To have an *Ansatz* for the weighting function P, let's consider the following reasoning. The 2D void profile δ_{2D} can be understood as being "traced" by the 3D void centers. In the case in which the void centers are point-like particles, then the projected profile would be simply given by the density profile of these particles around the 2D center, or the cross-correlation between 3D and 2D centers. This reasoning makes clear that the 2D profile at the distance $\boldsymbol{r}_\perp$ from the 2D void

center due to a 3D void at distance $\boldsymbol{x}_\perp$ from the 2D center must be proportional to this cross-correlation, i.e.

$$\delta^{2D}(\boldsymbol{r}_\perp, \boldsymbol{x}_\perp) \propto \xi_{2D,3D}(\boldsymbol{x}_\perp). \tag{4.37}$$

This *Ansatz* satisfies the limit in which there is no correlation between 2D and 3D centers, in which case the 2D profile must be flat, corresponding to taking random points in the projected field and using the model 4.34 to predict the flat profile.

However, these "tracers" have their own density profiles and, furthermore, they come in different "flavours", labelled by the 3D void radius. For a moment, let's think in the simplest possible case, in which there is only one 3D voids radius, i.e., only one type of 3D profile which is tracing a 2D center. In this case, the 2D profile must be given by equation 4.37 with only one contribution (only one type of 3D void profile). The factor which connects both quantities is simply given by the equation 4.36, i.e., the correlation $\xi_{2D,3D}$ acts as a weight, or a bias, for how strong the contribution from a projected 3D profile contributes :

$$\delta^{2D}\left(\boldsymbol{r}_\perp, \boldsymbol{x}_\perp\right) = \delta_\Sigma(|\boldsymbol{r}_\perp - \boldsymbol{x}_\perp|)\xi_{2D,3D}(\boldsymbol{x}_\perp) \tag{4.38}$$

and the isotropic 2D profile is then given by the contribution of all 3D voids distributed around the 2D center

$$\delta^{2D}(r_\perp) = \int dx_\perp d\phi \delta_\Sigma(\sqrt{(r_\perp - x_\perp \cos(\phi))^2 + x_\perp^2 \cos(\phi)})^2 \xi_{2D,3D}(x_\perp). \tag{4.39}$$

In the case where there are more than one 3D void radius, this quantity must be weighted by the 3D void abundance and then we arrive at the equation 4.34. Therefore, the weighting function P must be given by the cross-correlation between 2D and 3D centers.

It is important to notice the most dramatic approximation we are making in all this reasoning : that the Universe is only composed by voids. This is obviously a crude approximation, but we expect that it should work at some extent in regions dominated by large voids.

In the following sections we present measurements of the model's ingredients in a N-body dark matter only simulation box. Namely, we measure the abundance,

the 2D, 3D correlations and the density profiles of 3D voids. Then we use these ingredients to test the model given by equation 4.34.

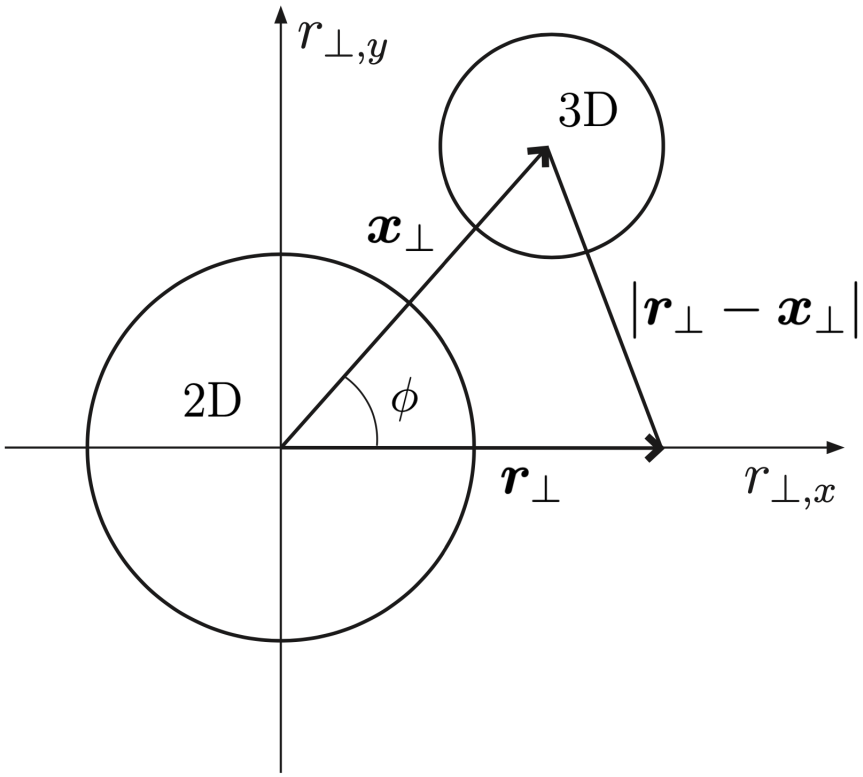

FIGURE 4.15 : Schematical representation of how the model works. A void at position $\boldsymbol{x}_\perp$ contributes to the profile at $r_\perp$ as $\delta^{3D}(|\boldsymbol{r}_\perp - \boldsymbol{x}_\perp|)$.

4.2.3 The void abundance

The void abundance is measured from the void found in the simulation using the density threshold $\Delta_{3D} = 0.2$. We then fit the measured abundance with the model

$$\frac{dn_v}{d\ln R} = \frac{f_v^{2LDB}(\sigma)}{V(R)} \frac{d\ln \sigma^{-1}}{d\ln R_L}, \tag{4.40}$$

where the multiplicity function f^{2SB} from the excursion-set with two linear diffusing barriers, is given by

$$\begin{aligned} f_v^{\mathrm{2LDB}}(\sigma) = {} & 2\,(1+D_v) \exp\left[-\frac{\beta_v^2\sigma^2}{2\,(1+D_v)} - \frac{\beta_v\delta_c}{1+D_v}\right] \\ & \times \sum_n \frac{n\pi}{\delta_T^2}\sigma^2 \sin\left(\frac{n\pi\delta_c}{\delta_T}\right) \exp\left[-\frac{n^2\pi^2\,(1+D_v)}{2\delta_T^2}\sigma^2\right]. \end{aligned} \tag{4.41}$$

Figure 4.16 shows the measured abundance and the fit. The best fit for the free-parameters are $(\beta_v = 0.02, D_v = 0.2)$. The abundance is the most important ingredient in the void-lensing model, since it is the only ingredient for which there is a theoretical prediction, which depends on the linear power-spectrum.

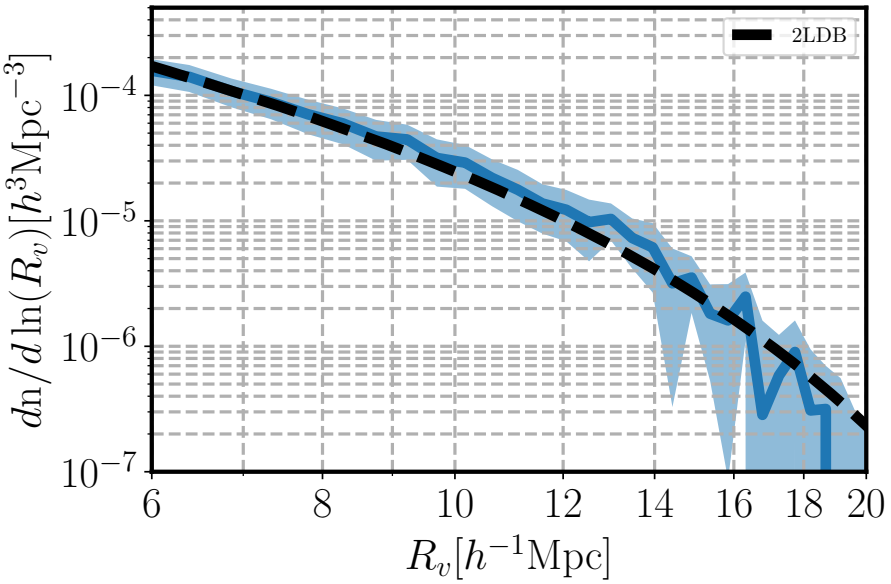

FIGURE 4.16 : The measured abundance in a DM only N-body simulation of size $L = 500h^{-1}$Mpc compared to the 2LDB model prediction using two free-parameters $(\beta_v,\ D_v)$.

4.2.4 The cross-correlation between 2D and 3D voids.

Figure 4.18 shows the correlation between 2D and 3D void positions as a function of the their separation $x_\perp$. We take three bins of 2D void radius R_v^{2D}, $[3, 5]h^{-1}$Mpc (top-left), $[5, 15]h^{-1}$Mpc (top-right) and $[10, 25]h^{-1}$Mpc (bottom). The 2D voids were found with a 2D density threshold of $\Delta_{2D} = 0.4$ and in a slice $\Delta L_z = 50h^{-1}$Mpc, whereas the 3D sample was found with $\Delta_{3D} = 0.2$.

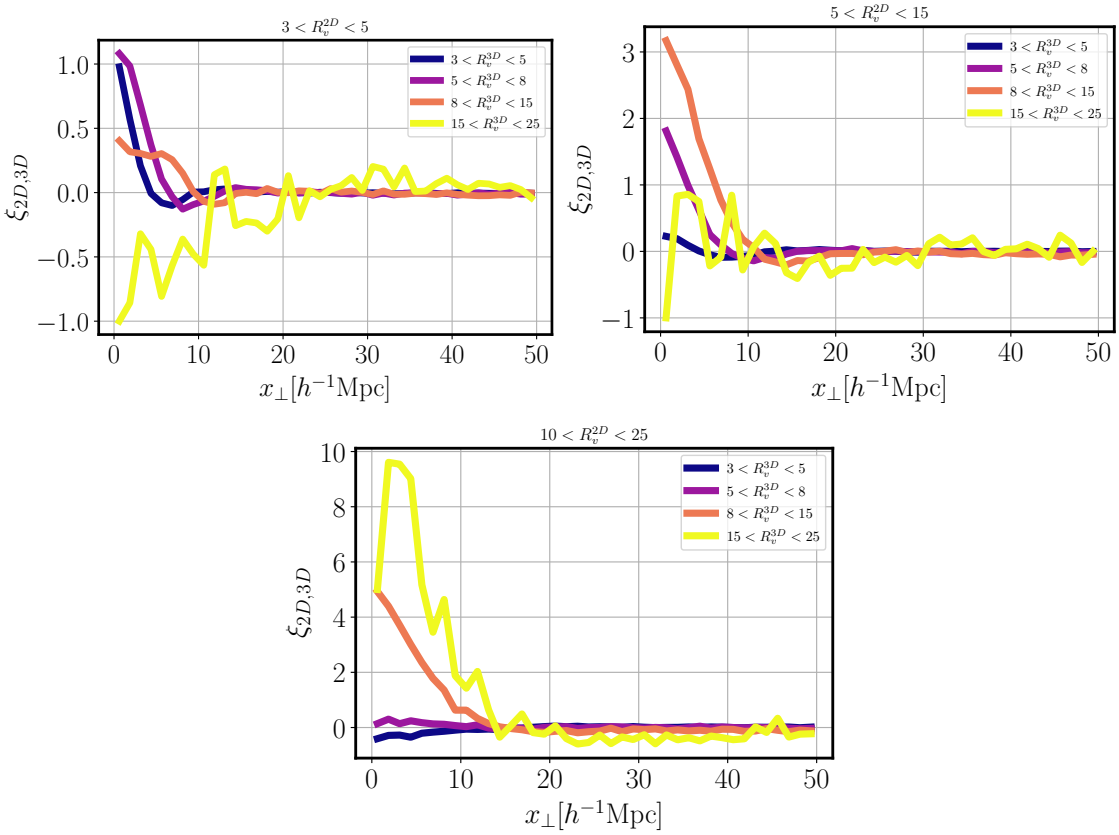

FIGURE 4.17 : Correlation between three bins of 2D radius and four bins of 3D radius. The correlations show that 2D voids are correlated with 3D voids with similar size.

The correlation $\xi_{2D,3D}$ is calculated as $n_{3D}(x_\perp)/n_r(x_\perp) - 1$, where $n_{3D}(x_\perp)$ is the number density of $3D$ centers at a distance $x_\perp$ from the 2D center and $n_r(x_\perp)$ is the same quantity for random points.

We can drawn some observations from these correlations :

- 2D voids of a given size tend to correlate with 3D voids of the similar size.
- The smallest sample 2D ($[3, 5]h^{-1}$Mpc) is anti-correlated with the largest 3D sample.
- The largest 2D sample present no-correlation with smallest 3D sample.
- The correlation between the largest 2D and 3D samples is stronger than the correlation between the smallest ones.

These observations suggest that the larger 2D voids are the result of combinations of 3D voids of similar size or larger, whereas the smallest 3D voids are randomly distributed around the 2D centers. This might be a direct consequence of the 3D algorithm, which tends to place small voids around large voids.

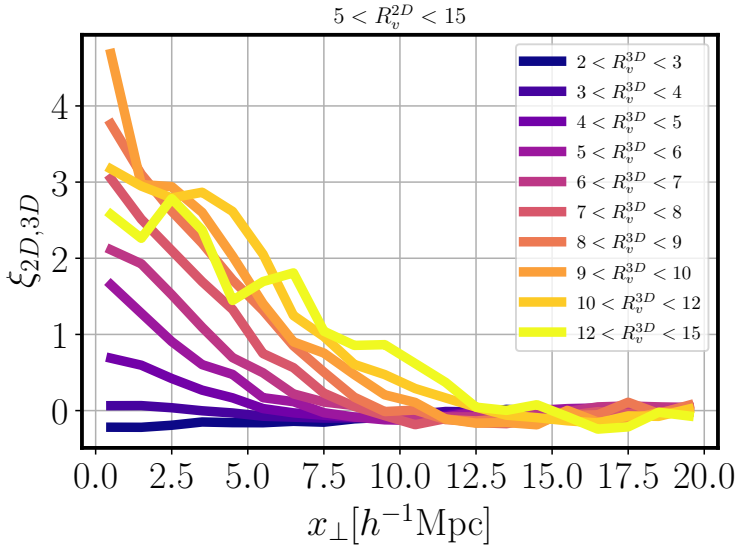

FIGURE 4.18 : Correlations between the bins of 2D and 3D that we interpolate to test the model.

We can say that the 2D voids are then tracing the same kind of structures than the 3D ones, i.e. large (small) 2D voids carry projected information from large (small) 3D voids.

4.2.5 Anisotropic void profiles

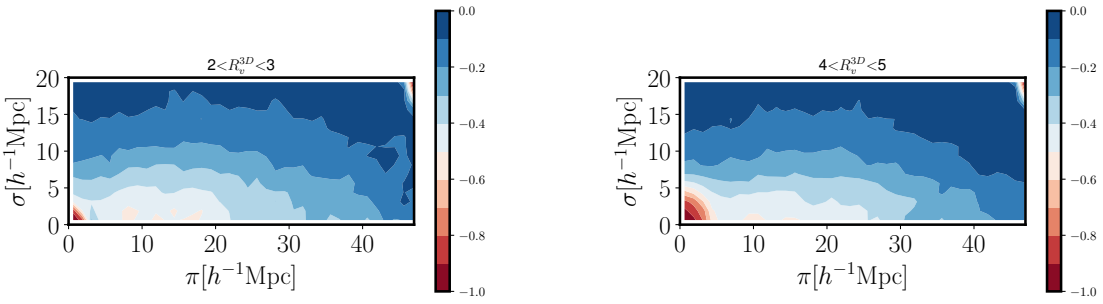

Figure 4.19 shows the density profiles of 3D voids which are at a distance from the 2D centers with non-zero correlation. In another words, it is the subsample of

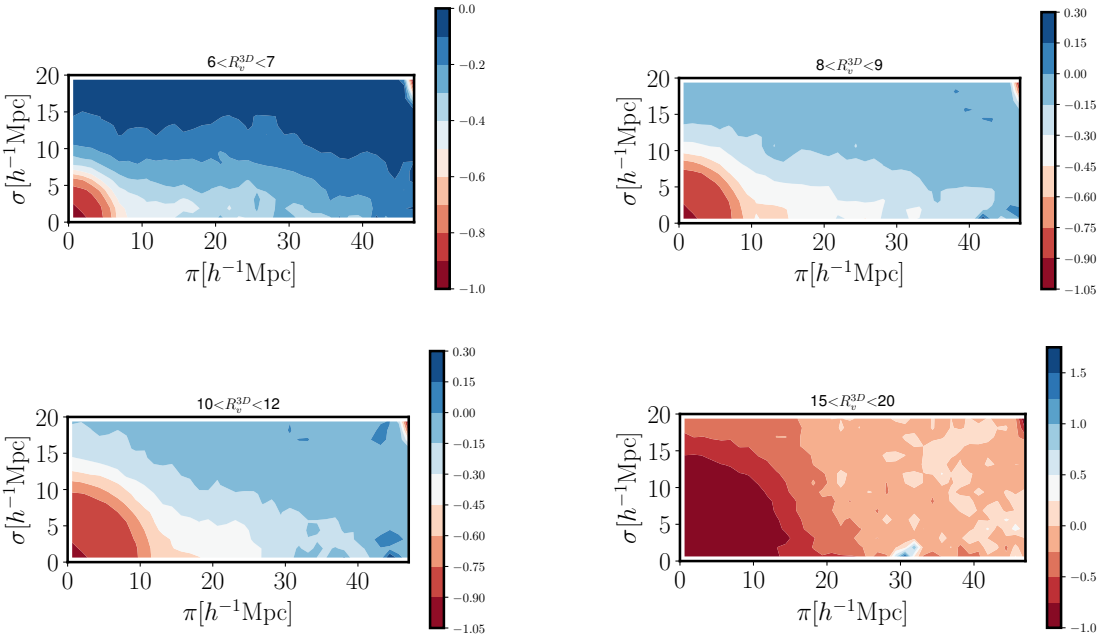

FIGURE 4.19 : Density profiles of 3D voids which are correlated with 2D voids, i.e., 3D voids at a distance from 2D centers which present non-zero correlation.

3D centers which are around the 2D centers and, therefore, it is the subsample of 3D voids which combine to form 2D voids through the model 4.34.

Clearly all the profiles present a certain degree of anisotropy. This anisotropy suggests that the 3D voids which combine to form the structures identified as voids in the projected field, are found in regions of tidal forces with direction aligned with the line-of-sight. In another words, the voids which combine to form 2D voids present an intrinsic alignment between them.

The smaller voids present profiles with higher degrees of anisotropy than the larger voids. This is expected, since the larger voids are better approximated by the spherical expansion and are found in the centers of large underdensities, whereas smaller voids are mainly voids-in-clouds.

4.2.6 Testing the model

In this section we test the model 4.34 and access the impact of each ingredient.

We use the ingredients presented in the previous sections and then numerically perform the integral three integrals 4.34 with interpolated 2D-3D correlations

(Figure 4.17), 3D density profiles (Figure 4.19) and abundance (Figure 4.16). We use the interpolated values of measured quantities from the simulation to avoid any difference arising from a poor modelling of any of these ingredients, as well as to highlight the usefulness and limitations of equation 4.34.

To show the impact of each ingredient in the prediction of the 2D density profile, we perform the model in four versions, namely, the integral 4.34, which we call the "full model" and three versions taking one ingredient at a time, i.e., one without weighting by the abundance, but only by the correlations, one weighting without the correlations, but only with the abundance and ,finally, one which is simply the integral 4.34 without any weight.

The upper plot in Figure 4.20 shows four versions compared to the 2D profile measured in the simulation. As expected, the version without weights (green) is the worst one compared to the measured profile (dashed-black). The incorporation of the correlations (red) corrects the profile specially for small $r_\perp$, but also slightly shifts the curve closer to the measured one for large $r_\perp$. The weighting by the abundance almost completely corrects the profile for large $r_\perp$, whereas it worsens the discrepancy for small $r_\perp$. The latter effect is easier to understand : the abundance privileges smaller voids, which should be more relevant for large $r_\perp$, since small voids tend to be found around the large underdense regions, but not close to the center of those regions. That is why the over weighting of small voids (as in the case of weighting only by the abundance) produces the large discrepancy at small $r_\perp$. The inclusion of the 2D-3D correlations corrects this by attributing more weight to larger voids. The lower plot in Figure 4.20 shows the same result as a function of the reduced radius $r_\perp/\bar{R}_v^{2D}$, where $\bar{R}_v^{2D} = 7.36h^{-1}$Mpc is the average radius in the 2D void sample. This result in reduced radius is useful to see in which regime, amongst one-void and two-void terms, the model works.

The model perfectly matches the measurement for $r_\perp \geq 7.5h^{-1}$Mpc ($r_\perp/\bar{R}_v^{2D} \geq 1$) and increasingly deviates for smaller $r_\perp$ ($r_\perp/\bar{R}_v^{2D}$). Given the intuition above, this discrepancy might be due to an over-weighting of small voids close to the void center, or an under-estimation of cross-correlation between 2D and larger 3D voids for small $x_\perp$.

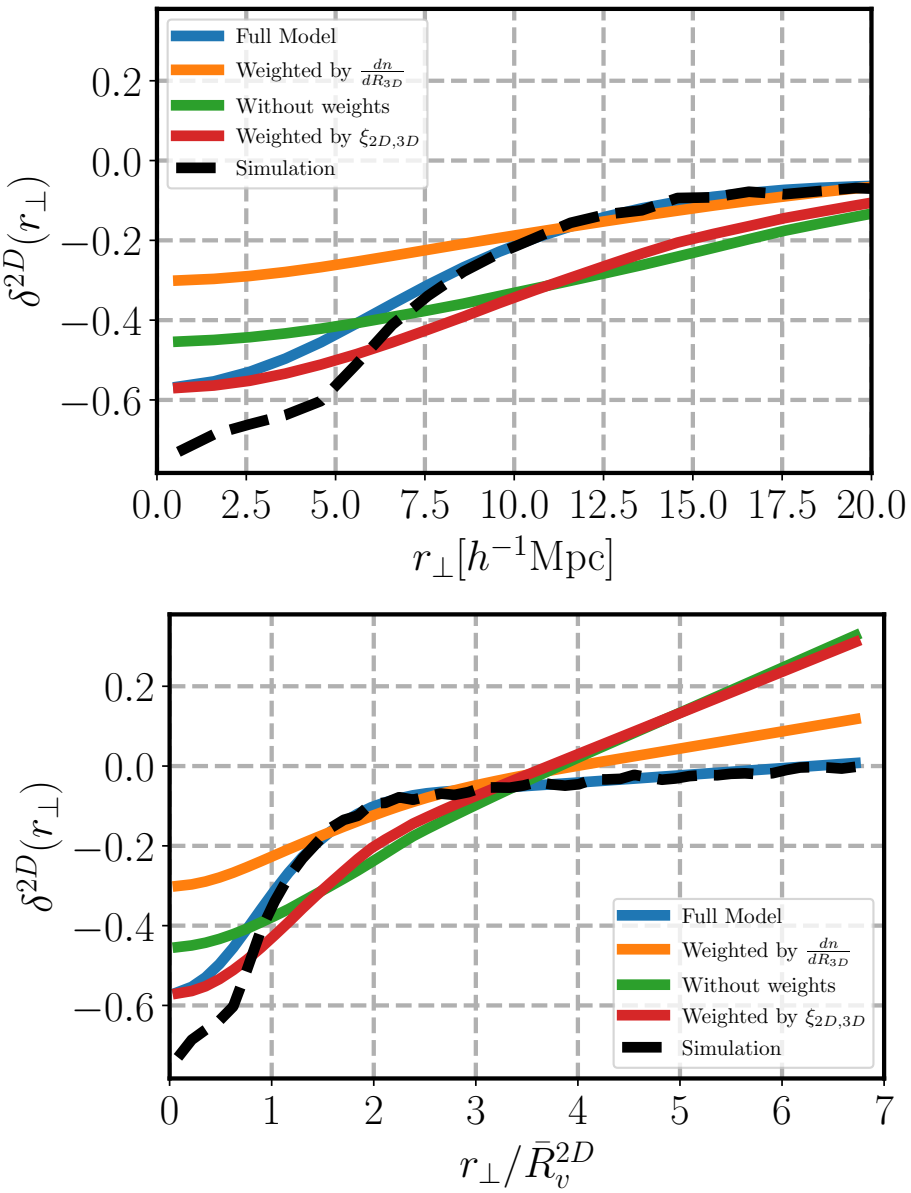

FIGURE 4.20 : Upper plot : The full model (blue, equation 4.34), the model only weighted by the abundance (orange), the model only weighted by the 2D-3D cross-correlations (red) and the model without weights (green). Lower plot : the same result presented as a function of reduced radius.

Conclusion

Cosmology is in an era in which we'll have an abundance of data. Basically, the next generation of galaxy surveys will map all the available sky area, up to high redshifts. All these data is a great opportunity to test our current knowledge about how structures form, which depends on how the laws of physics operate on large-scales.

The most important physical interaction on large scales is gravity. Since the cause for cosmic acceleration is not yet clear, it is possible that gravity is modified on large-scales.

Cosmic voids are the ideal environment to test deviations from general relativity on large-scales. If there is any deviation from general relativity, it will manifest, in its simplest form, as an additional scalar field, which is screened on high density environments, but can have its effects revealed by the most underdense and largest structures in the Universe.

The void science finds itself in an early stage of development, especially when it comes to theoretical predictions. However, recent works have shown its potential as a complementary probe, even beyond as a test of gravity. Arguably, the evolution of voids is sensitive to dynamical dark energy and massive neutrinos, since their relative density is higher inside voids.

Amongst the limitations of void science, there is the fact that we use the sparse galaxy field to run void finder algorithms. As a consequence, the galaxy voids does not correspond to the dark matter voids in general. One way of circumvent this limitation and gain information in void analysis, is to make usage of the fact that voids leave an imprint in the shapes of background galaxies. Therefore, we can infer the density profiles of voids in the dark matter field by measuring the tangential shear of background galaxies around voids.

The definition of what is a void is not unique, leaving freedom of choice when defining what is a void. The first contribution of this work is to show that this freedom of choice plays a significant role in the resulting $\Delta\Sigma$ profile around voids. We propose a new algorithm which provides a deep $\Delta\Sigma$ profile. We show that the

depth of these profiles is related to voids which are better defined, i.e. voids which are centered well within the underdensities in large-scale structures. We compare the $\Delta\Sigma$ as measured by our algorithm to the same observable measured by the `ZOBOV` algorithm and find significantly deeper $\Delta\Sigma$ profiles for our voids. This is consequence of the fact that we apply our algorithm on projected slices, rather than the 3D field, as well as due to the way in which we define voids.

Then, we make a consistency check between the $\Delta\Sigma$ as measured through the shear of background galaxies, as done in real observations, and the same quantity as measured directly from the dark matter density profiles of the same voids, the quantity we don't have access in real observations. This consistency check has never been done before for voids. We have two reasons to suspect that voids will present inconsistencies between these two measurements : (i) this measurement is based on the assumption that voids act as thin lenses. However, voids can extend over hundreds of h^{-1}Mpc and might not respect the thin-lens approximation, as it easily is in the case of halos. (ii) The contribution from the source distribution is cancelled out in the case of halos (or galaxies) due to the high number of them. Voids occupy large volumes and therefore are less numerous. The extent to which the $\Delta\Sigma$ does not depend on the source distribution is not clear.

The consistency test shows that the $\Delta\Sigma$ around voids defined in slices larger than $\simeq 100^{-1}$Mpc is not consistent between shear and dark-matter. The fact that these voids present elongations along-the-line of sight, and that the break of the thin-lens approximation has the same effect on shear profiles as we observe in this test, suggests that the inconsistency is coming from the large size of voids along-the-line of sight. For smaller slices, the shear and dark-matter $\Delta\Sigma$ profiles are consistent. We obtain a similar result for `ZOBOV` voids, i.e., larger voids present larger discrepancies between shear and dark matter $\Delta\Sigma$ profiles.

Since voids found in the projected field present deeper $\Delta\Sigma$ profiles, it is desirable to have an analytical model for their profiles. We propose the idea that the profiles we measure for 2D voids (and infer through $\Delta\Sigma$) is related to the profiles of 3D voids as an integral weighted by the correlation between 2D and 3D void positions and the abundance of 3D voids. We show that this model is capable of providing the exact same profile as the one directly measured from an N-body simulation in the two-void regime, whereas the one-void regime is increasingly inconsistent as the radial distance to the void center decreases. The reason of this discrepancy is the subject of an ongoing work.

The science concerning the cross-correlation between cosmic voids and shear (Void-Lensing) is in its infancy. In this work we have shown that this observable can be of great interest under a suitable choice of void-finder algorithm. We show that the $\Delta\Sigma$ profile that we are going to be able to measure in the next generation of galaxy surveys is exactly due to the projection of the dark matter density profiles of voids. Finally, we provide an analytical model to connect the measured $\Delta\Sigma$ from void in the projected field to voids defined in the 3D field, for which we know how to predict the abundance. We believe that these advancements are crucial to future analysis involving void-lensing.

Bibliographie

[Abb+22] TMC Abbott, M Aguena, A Alarcon et al. « Dark Energy Survey Year 3 results : Cosmological constraints from galaxy clustering and weak lensing ». In : *Physical Review D* 105.2 (2022), p. 023520 (cf. p. 96).

[AL13] L Raul Abramo et Katie E Leonard. « Why multitracer surveys beat cosmic variance ». In : *Monthly Notices of the Royal Astronomical Society* 432.1 (2013), p. 318-326 (cf. p. 26, 119).

[Agh+16] Amir Aghamousa, Jessica Aguilar, Steve Ahlen et al. « The DESI experiment part I : science, targeting, and survey design ». In : *arXiv preprint arXiv :1611.00036* (2016) (cf. p. 98).

[Agh+18] N Aghanim, Y Akrami, M Ashdown et al. « Planck 2018 results. VI. Cosmological parameters ». In : *arXiv preprint arXiv :1807.06209* (2018) (cf. p. 22).

[Agh+20] Nabila Aghanim, Yashar Akrami, Mark Ashdown et al. « Planck 2018 results-VI. Cosmological parameters ». In : *Astronomy & Astrophysics* 641 (2020), A6 (cf. p. 9, 28, 29).

[AFW99] Luca Amendola, Joshua A Frieman et Ioav Waga. « Weak gravitational lensing by voids ». In : *Monthly Notices of the Royal Astronomical Society* 309.2 (1999), p. 465-473 (cf. p. 27, 99).

[Bak+18] Tessa Baker, Joseph Clampitt, Bhuvnesh Jain et al. « Void lensing as a test of gravity ». In : *Physical Review D* 98.2 (2018), p. 023511 (cf. p. 27, 54, 99).

[Bar+15] Alexandre Barreira, Marius Cautun, Baojiu Li et al. « Weak lensing by voids in modified lensing potentials ». In : *Journal of Cosmology and Astroparticle Physics* 2015.08 (2015), p. 028 (cf. p. 27, 99).

[Bau09] Daniel Baumann. « TASI lectures on inflation ». In : *arXiv preprint arXiv :0907.5424* (2009) (cf. p. 36).

[BW02] Andreas A Berlind et David H Weinberg. « The halo occupation distribution : Toward an empirical determination of the relation between galaxies and mass ». In : *The Astrophysical Journal* 575.2 (2002), p. 587 (cf. p. 57).

[Blu+92] GR Blumenthal, L Nicolaci Da Costa, DS Goldwirth et al. « The largest possible voids ». In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 388, April 1, 1992, p. 234-241. Research supported by USIBSF and Smithsonian Institution.* 388 (1992), p. 234-241 (cf. p. 75, 111).

[Bon+91] JR Bond, Shaun Cole, George Efstathiou et al. « Excursion set mass functions for hierarchical Gaussian fluctuations ». In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 379, Oct. 1, 1991, p. 440-460. Research supported by NSERC, NASA, and University of California.* 379 (1991), p. 440-460 (cf. p. 57, 61, 63, 66, 68, 110).

[CPL15] Yan-Chuan Cai, Nelson Padilla et Baojiu Li. « Testing gravity using cosmic voids ». In : *Monthly Notices of the Royal Astronomical Society* 451.1 (2015), p. 1036-1055 (cf. p. 110).

[Car19] Sean M Carroll. « Beyond falsifiability : Normal science in a multiverse ». In : *Why trust a theory* (2019), p. 300-314 (cf. p. 22).

[Cau+18] Marius Cautun, Enrique Paillas, Yan-Chuan Cai et al. « The Santiago–Harvard–Edinburgh–Durham void comparison–I. SHEDding light on chameleon gravity tests ». In : *Monthly Notices of the Royal Astronomical Society* 476.3 (2018), p. 3195-3217 (cf. p. 27, 104, 115).

[CHD14] Kwan Chuen Chan, Nico Hamaus et Vincent Desjacques. « Large-scale clustering of cosmic voids ». In : *Physical Review D* 90.10 (2014), p. 103521 (cf. p. 56).

[Cha+20] Kwan Chuen Chan, Yin Li, Matteo Biagetti et al. « Measurement of Void Bias Using Separate Universe Simulations ». In : *The Astrophysical Journal* 889.2 (2020), p. 89 (cf. p. 98).

[Con+21] Sofia Contarini, Federico Marulli, Lauro Moscardini et al. « Cosmic voids in modified gravity models with massive neutrinos ». In : *Monthly Notices of the Royal Astronomical Society* 504.4 (2021), p. 5021-5038 (cf. p. 26, 54).

[Con+23] Sofia Contarini, Alice Pisani, Nico Hamaus et al. « Cosmological constraints from the BOSS DR12 void size function ». In : *The Astrophysical Journal* 953.1 (2023), p. 46 (cf. p. 26, 110).

[CS02] Asantha Cooray et Ravi Sheth. « Halo models of large scale structure ». In : *Physics reports* 372.1 (2002), p. 1-129 (cf. p. 56).

[DCL19] Christopher T Davies, Marius Cautun et Baojiu Li. « Cosmological test of gravity using weak lensing voids ». In : *Monthly Notices of the Royal Astronomical Society* 490.4 (2019), p. 4907-4917 (cf. p. 27, 99).

[Dav+21] Christopher T Davies, Enrique Paillas, Marius Cautun et al. « Optimal void finders in weak lensing maps ». In : *Monthly Notices of the Royal Astronomical Society* 500.2 (2021), p. 2417-2439 (cf. p. 99, 104, 115).

[Del+23] Ana Maria Delgado, Boryana Hadzhiyska, Sownak Bose et al. « The MillenniumTNG Project : Intrinsic alignments of galaxies and halos ». In : *Monthly Notices of the Royal Astronomical Society* (2023), stad1781 (cf. p. 96).

[DBW22] Joseph DeRose, Matthew R Becker et Risa H Wechsler. « Modeling Redshift-Space Clustering with Abundance Matching ». In : *The Astrophysical Journal* 940.1 (2022), p. 13 (cf. p. 113).

[DeR+22] Joseph DeRose, RH Wechsler, MR Becker et al. « Dark Energy Survey Year 3 results : Cosmology from combined galaxy clustering and lensing validation on cosmological simulations ». In : *Physical Review D* 105.12 (2022), p. 123520 (cf. p. 113).

[DeR+19] Joseph DeRose, Risa H Wechsler, Matthew R Becker et al. « The buzzard flock : Dark energy survey synthetic sky catalogs ». In : *arXiv preprint arXiv :1901.02401* (2019) (cf. p. 113, 116).

[DJS18] Vincent Desjacques, Donghui Jeong et Fabian Schmidt. « Large-scale galaxy bias ». In : *Physics reports* 733 (2018), p. 1-193 (cf. p. 25, 55, 59, 64, 65, 79, 110).

[Dod03] Scott Dodelson. *Modern cosmology*. Elsevier, 2003 (cf. p. 40, 41, 45, 46).

[DS20] Scott Dodelson et Fabian Schmidt. *Modern cosmology*. Academic press, 2020 (cf. p. 32).

[Dub+93] John Dubinski, L Nicolaci da Costa, Dalia S Goldwirth et al. « Void evolution and the large-scale structure ». In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 410, no. 2, p. 458-468.* 410 (1993), p. 458-468 (cf. p. 70).

[Duf+10] Alan R Duffy, Joop Schaye, Scott T Kay et al. « Impact of baryon physics on dark matter structures : a detailed simulation study of halo density profiles ». In : *Monthly Notices of the Royal Astronomical Society* 405.4 (2010), p. 2161-2178 (cf. p. 96).

[Fan+19] Yen Fang, Nico Hamaus, B Jain et al. « Dark Energy Survey year 1 results : the relationship between mass and light around cosmic voids ». In : *Monthly Notices of the Royal Astronomical Society* 490.3 (2019), p. 3573-3587 (cf. p. 27, 99, 100, 125).

[GT78] Stephen A Gregory et Laird A Thompson. « The Coma/A1367 supercluster and its environs ». In : (1978) (cf. p. 54).

[Gru+16] D Gruen, O Friedrich, A Amara et al. « Weak lensing by galaxy troughs in DES Science Verification data ». In : *Monthly Notices of the Royal Astronomical Society* 455.3 (2016), p. 3367-3380 (cf. p. 99).

[Ham+17] Nico Hamaus, Marie-Claude Cousinou, Alice Pisani et al. « Multipole analysis of redshift-space distortions around cosmic voids ». In : *Journal of Cosmology and Astroparticle Physics* 2017.07 (2017), p. 014 (cf. p. 54).

[HSW14] Nico Hamaus, PM Sutter et Benjamin D Wandelt. « Universal density profile for cosmic voids ». In : *Physical review letters* 112.25 (2014), p. 251302 (cf. p. 104).

[HOH13] Yuichi Higuchi, Masamune Oguri et Takashi Hamana. « Measuring the mass distribution of voids with stacked weak lensing ». In : *Monthly Notices of the Royal Astronomical Society* 432.2 (2013), p. 1021-1031 (cf. p. 99).

[HS96] Wayne Hu et Naoshi Sugiyama. « Small-scale cosmological perturbations : an analytic approach ». In : *The Astrophysical Journal* 471.2 (1996), p. 542 (cf. p. 46).

[Ick84] Vincent Icke. « Voids and filaments ». In : *Monthly Notices of the Royal Astronomical Society* 206.1 (1984), 1P-3P (cf. p. 69).

[Jab+23] Mariana Jaber, Marius Peper, Wojciech A Hellwing et al. « Hierarchical structure of the cosmic web and galaxy properties ». In : *arXiv preprint arXiv :2304.14387* (2023) (cf. p. 54).

[JLH13] Elise Jennings, Yin Li et Wayne Hu. « The abundance of voids and the excursion set formalism ». In : *Monthly Notices of the Royal Astronomical Society* 434.3 (2013), p. 2167-2181 (cf. p. 75, 111).

[KVJ09] Marc Kamionkowski, Licia Verde et Raul Jimenez. « The void abundance with non-Gaussian primordial perturbations ». In : *Journal of Cosmology and Astroparticle Physics* 2009.01 (2009), p. 010 (cf. p. 98).

[Kir+15] Donnacha Kirk, Michael L Brown, Henk Hoekstra et al. « Galaxy alignments : Observations and impact on cosmology ». In : *Space Science Reviews* 193 (2015), p. 139-211 (cf. p. 96).

[Kir+81] Robert P Kirshner, Augustus Oemler Jr, Paul L Schechter et al. « A million cubic megaparsec void in Bootes ». In : *Astrophysical Journal, Part 2-Letters to the Editor, vol. 248, Sept. 1, 1981, p. L57-L60.* 248 (1981), p. L57-L60 (cf. p. 54).

[KP11] Jurjen F. Koksma et Tomislav Prokopec. *The Cosmological Constant and Lorentz Invariance of the Vacuum State*. 2011. arXiv : 1105.6296 [gr-qc] (cf. p. 52).

[Kre+19] Christina D Kreisch, Alice Pisani, Carmelita Carbone et al. « Massive neutrinos leave fingerprints on cosmic voids ». In : *Monthly Notices of the Royal Astronomical Society* 488.3 (2019), p. 4413-4426 (cf. p. 110).

[Lau+11] R. Laureijs, J. Amiaux, S. Arduini et al. *Euclid Definition Study Report*. 2011. url : https://arxiv.org/abs/1110.3193 (cf. p. 98).

[MB95] Chung-Pei Ma et Edmund Bertschinger. « Cosmological perturbation theory in the synchronous and conformal Newtonian gauges ». In : *arXiv preprint astro-ph/9506072* (1995) (cf. p. 34).

[MR10] Michele Maggiore et Antonio Riotto. « The Halo Mass Function from Excursion Set Theory. I. Gaussian fluctuations with non-Markovian dependence on the smoothing scale ». In : *The Astrophysical Journal* 711.2 (2010), p. 907 (cf. p. 68, 69, 110-112).

[Mar12] Jerome Martin. « Everything you always wanted to know about the cosmological constant problem (but were afraid to ask) ». In : *Comptes Rendus Physique* 13.6-7 (2012), p. 566-665 (cf. p. 51).

[MS18] Elena Massara et Ravi K Sheth. « Density and velocity profiles around cosmic voids ». In : *arXiv preprint arXiv :1811.03132* (2018) (cf. p. 55, 99).

[Mas+15] Elena Massara, Francisco Villaescusa-Navarro, Matteo Viel et al. « Voids in massive neutrino cosmologies ». In : *Journal of Cosmology and Astroparticle Physics* 2015.11 (2015), p. 018 (cf. p. 26, 54, 98, 110).

[Mel+14] Peter Melchior, PM Sutter, Erin S Sheldon et al. « First measurement of gravitational lensing by cosmic voids in SDSS ». In : *Monthly Notices of the Royal Astronomical Society* 440.4 (2014), p. 2922-2927 (cf. p. 27, 99).

[Mer+22] Thiago Mergulhão, Henrique Rubira, Rodrigo Voivodic et al. « The effective field theory of large-scale structure and multi-tracer ». In : *Journal of Cosmology and Astroparticle Physics* 2022.04 (2022), p. 021 (cf. p. 119).

[MW96] HJ Mo et Simon DM White. « An analytic model for the spatial clustering of dark matter haloes ». In : *Monthly Notices of the Royal Astronomical Society* 282.2 (1996), p. 347-361 (cf. p. 25).

[Muk05] Viatcheslav Mukhanov. *Physical foundations of cosmology*. Cambridge university press, 2005 (cf. p. 33, 35).

[Nad+19] Seshadri Nadathur, Paul M Carter, Will J Percival et al. « REVOLVER : REal-space VOid Locations from suVEy Reconstruction ». In : *Astrophysics Source Code Library* (2019), ascl-1907 (cf. p. 113).

[Ney08] Mark C Neyrinck. « ZOBOV : a parameter-free void-finding algorithm ». In : *Monthly notices of the royal astronomical society* 386.4 (2008), p. 2101-2109 (cf. p. 108, 113).

[Nte16] Pierros Ntelis. « The Homogeneity Scale of the universe ». In : *arXiv preprint arXiv :1607.03418* (2016) (cf. p. 28).

[Pai+17] Enrique Paillas, Claudia DP Lagos, Nelson Padilla et al. « Baryon effects on void statistics in the EAGLE simulation ». In : *Monthly Notices of the Royal Astronomical Society* 470.4 (2017), p. 4434-4452 (cf. p. 54).

[PH90] JA Peacock et AF Heavens. « Alternatives to the Press–Schechter cosmological mass function ». In : *Monthly Notices of the Royal Astronomical Society* 243.1 (1990), p. 133-143 (cf. p. 62, 110).

[Per+19] Eder LD Perico, Rodrigo Voivodic, Marcos Lima et al. « Cosmic voids in modified gravity scenarios ». In : *Astronomy & Astrophysics* 632 (2019), A52 (cf. p. 26, 54, 99, 110).

[Pis+15] Alice Pisani, PM Sutter, Nico Hamaus et al. « Counting voids to probe dark energy ». In : *Physical Review D* 92.8 (2015), p. 083531 (cf. p. 26, 98).

[Pra+12] Francisco Prada, Anatoly A Klypin, Antonio J Cuesta et al. « Halo concentrations in the standard Λ cold dark matter cosmology ». In : *Monthly Notices of the Royal Astronomical Society* 423.4 (2012), p. 3018-3030 (cf. p. 107).

[PS74] William H Press et Paul Schechter. « Formation of galaxies and clusters of galaxies by self-similar gravitational condensation ». In : *Astrophysical Journal, Vol. 187, pp. 425-438 (1974)* 187 (1974), p. 425-438 (cf. p. 57, 60, 110).

[Rie+98] Adam G Riess, Alexei V Filippenko, Peter Challis et al. « Observational evidence from supernovae for an accelerating universe and a cosmological constant ». In : *The astronomical journal* 116.3 (1998), p. 1009 (cf. p. 28, 49).

[Sán+16] Carles Sánchez, Joseph Clampitt, András Kovacs et al. « Cosmic voids and void lensing in the Dark Energy Survey Science Verification data ». In : *Monthly Notices of the Royal Astronomical Society* (2016), stw2745 (cf. p. 27, 99, 100).

[Sar+09] Prakash Sarkar, Jaswant Yadav, Biswajit Pandey et al. « The scale of homogeneity of the galaxy distribution in SDSS DR6 ». In : *Monthly Notices of the Royal Astronomical Society : Letters* 399.1 (2009), p. L128-L131 (cf. p. 28).

[Sch05] Peter Schneider. « Gravitational lenses ». In : *Gravitational Lenses : Proceedings of a Conference Held in Hamburg, Germany 9–13 September 1991*. Springer. 2005, p. 196-208 (cf. p. 101).

[Sch+19] Nico Schuster, Nico Hamaus, Alice Pisani et al. « The bias of cosmic voids in the presence of massive neutrinos ». In : *Journal of Cosmology and Astroparticle Physics* 2019.12 (2019), p. 055 (cf. p. 98).

[She+04] Erin S Sheldon, David E Johnston, Joshua A Frieman et al. « The galaxy-mass correlation function measured from weak lensing in the Sloan Digital Sky Survey ». In : *The Astronomical Journal* 127.5 (2004), p. 2544 (cf. p. 91, 116).

[ST99] Ravi K Sheth et Giuseppe Tormen. « Large-scale bias and the peak background split ». In : *Monthly Notices of the Royal Astronomical Society* 308.1 (1999), p. 119-126 (cf. p. 25).

[SV04] Ravi K Sheth et Rien Van De Weygaert. « A hierarchy of voids : much ado about nothing ». In : *Monthly Notices of the Royal Astronomical Society* 350.2 (2004), p. 517-538 (cf. p. 56, 70-72, 75, 105, 106, 108, 111, 112).

[Shi+23] Takumi Shimasue, Ken Osato, Masamune Oguri et al. « Line-of-sight structure of troughs identified in Subaru Hyper Suprime-Cam Year 3 weak lensing mass maps ». In : *arXiv preprint arXiv :2307.11407* (2023) (cf. p. 99).

[SFW06] Volker Springel, Carlos S Frenk et Simon DM White. « The large-scale structure of the Universe ». In : *nature* 440.7088 (2006), p. 1137-1144 (cf. p. 24).

[Sus03] Leonard Susskind. *The Anthropic Landscape of String Theory*. 2003. arXiv : hep-th/0302219 [hep-th] (cf. p. 50).

[SSS84] Yasushi Suto, Katsuhiko Sato et Humitaka Sato. « Expansion of voids in a matter-dominated universe ». In : *Progress of theoretical physics* 71.5 (1984), p. 938-945 (cf. p. 71).

[Sut+12] PM Sutter, Guilhem Lavaux, Benjamin D Wandelt et al. « A first application of the Alcock–Paczynski test to stacked cosmic voids ». In : *The Astrophysical Journal* 761.2 (2012), p. 187 (cf. p. 54).

[Teg07] Max Tegmark. « 7 The multiverse hierarchy Max Tegmark Department of Physics, Massachusetts Institute of Technology ». In : *Universe or multiverse ?* (2007), p. 99 (cf. p. 22).

[Tel+16] Matteo Tellarini, Ashley J Ross, Gianmassimo Tasinato et al. « Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions ». In : *Journal of Cosmology and Astroparticle Physics* 2016.06 (2016), p. 014 (cf. p. 37).

[Tin+10] Jeremy L Tinker, Brant E Robertson, Andrey V Kravtsov et al. « The large-scale bias of dark matter halos : numerical calibration and model tests ». In : *The Astrophysical Journal* 724.2 (2010), p. 878 (cf. p. 57).

[TVW90] J Anthony Tyson, F Valdes et RA Wenk. « Detection of systematic gravitational lens galaxy image alignments-Mapping dark matter in galaxy clusters ». In : *The Astrophysical Journal* 349 (1990), p. L1-L4 (cf. p. 77).

[Ver+19] Giovanni Verza, Alice Pisani, Carmelita Carbone et al. « The void size function in dynamical dark energy cosmologies ». In : *Journal of Cosmology and Astroparticle Physics* 2019.12 (2019), p. 040 (cf. p. 54, 110).

[Voi+17a] Rodrigo Voivodic, Marcos Lima, Claudio Llinares et David F Mota. « Modeling void abundance in modified gravity ». In : *Physical Review D* 95.2 (2017), p. 024018 (cf. p. 54, 72, 74).

[Voi+17b] Rodrigo Voivodic, Marcos Lima, Claudio Llinares et David F. Mota. « Modeling void abundance in modified gravity ». In : *Physical Review D* 95.2 (jan. 2017). doi : 10.1103/physrevd.95.024018. url : https://doi.org/10.1103%2Fphysrevd.95.024018 (cf. p. 98, 112).

[VRL20] Rodrigo Voivodic, Henrique Rubira et Marcos Lima. « The Halo Void (Dust) Model of large scale structure ». In : *Journal of Cosmology and Astroparticle Physics* 2020.10 (2020), p. 033 (cf. p. 104, 109).

[Wec+22] Risa H Wechsler, Joseph DeRose, Michael T Busha et al. « ADDGALS : Simulated Sky Catalogs for Wide Field Galaxy Surveys ». In : *The Astrophysical Journal* 931.2 (2022), p. 145 (cf. p. 113).

[WT18] Risa H Wechsler et Jeremy L Tinker. « The connection between galaxies and their dark matter halos ». In : *Annual Review of Astronomy and Astrophysics* 56 (2018), p. 435-487 (cf. p. 56).

[Wey14] Rien van de Weygaert. « Voids and the Cosmic Web : cosmic depression & spatial complexity ». In : *Proceedings of the International Astronomical Union* 11.S308 (2014), p. 493-523 (cf. p. 54).

[WK93] Rien van de Weygaert et Eelco van Kampen. « Voids in gravitational instability scenarios–i. global density and velocity fields in an einstein–de sitter universe ». In : *Monthly Notices of the Royal Astronomical Society* 263.2 (1993), p. 481-526 (cf. p. 70).

[Yua+19] Shuo Yuan, Chuzhong Pan, Xiangkun Liu et al. « Cosmological studies from tomographic weak lensing peak abundances and impacts of photo-z errors ». In : *The Astrophysical Journal* 884.2 (2019), p. 164 (cf. p. 95).

[Zha+22] Cheng Zhao, Andrei Variu, Mengfan He et al. « The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey : cosmological implications from multitracer BAO analysis with galaxies and voids ». In : *Monthly Notices of the Royal Astronomical Society* 511.4 (2022), p. 5492-5524 (cf. p. 26).