Non-linear galaxy clustering in modified gravity cosmologies

CHENG-ZONG RUAN

Non-linear galaxy clustering in modified gravity cosmologies

Cheng-Zong Ruan

A thesis presented for the degree of

Doctor of Philosophy

Institute for Computational Cosmology

Department of Physics

The University of Durham

United Kingdom

July 2022

Non-linear galaxy clustering in modified gravity cosmologies

Cheng-Zong Ruan

Abstract

We present MG-GLAM, a code developed for the very fast production of full N-body cosmological simulations in modified gravity (MG) models. We describe the implementation, numerical tests, and first results of a large suite of cosmological simulations for a wide range of the viable MG models. The code is highly optimised, with a tremendous speedup of a factor of more than a hundred compared with earlier N-body codes, while still giving accurate predictions of the matter power spectrum and dark matter halo abundance. MG-GLAM is ideal for the generation of large numbers of MG simulations that can be used in the construction of mock galaxy catalogues and the production of accurate emulators for ongoing and future galaxy surveys.

The coming generation of galaxy surveys will provide measurements of galaxy clustering with unprecedented accuracy and data size, which will allow us to test cosmological models at much higher precision than previously achievable. This means that we must have more accurate theoretical predictions to compare with future observational data. As a first step towards more accurate modelling of the redshift space distortions (RSD) of small-scale galaxy clustering in modified gravity cosmologies, we investigate the validity of the so-called Skew-T (ST) probability distribution function (PDF) of halo pairwise peculiar velocities in these models. We show that, combined with the streaming model, the ST PDF substantially improves the small-scale predictions by incorporating skewness and kurtosis, for both Λ cold dark matter (ΛCDM) and two leading MG models: $f(R)$ gravity and the DGP braneworld model. The ST model reproduces the velocity PDF and redshift-space halo clustering measured from MG N-body simulations down to highly non-linear scales. By performing a simple Fisher analysis, we find a significant increase in constraining power to detect modifications of General Relativity by introducing small-scale information in the RSD analyses.

We introduce an emulator-based halo model approach for non-linear clustering of galaxies in modified gravity cosmologies. We construct accurate emulators, i.e. simulation-based theoretical templates, using neural networks for basic halo properties than can be calculated robustly from N-body simulations. The dark matter halo emulators can be combined with a halo-galaxy connection model to predict the galaxy clustering statistics down to non-linear scales through the halo model.

Supervisors: Prof. Baojiu Li, Prof. Carlton Baugh and Dr. Alexander Eggemeier

Acknowledgements

I would first like to thank my supervisors, Baojiu and Carlton. Thanks for your enduring support, patience, and insight over the past three years, as well as for providing me with the freedom to explore and develop my research interests. My inefficiency might have made you suspect the decision to hire me as your student. I could not have asked for a better supervisory team.

I would like to express my gratitude to Alex. Thank you for spending so much time answering my trivial questions, and saving me from the awkward silence when I have no idea in the group meeting.

I consider myself extremely lucky to have met Carol as a collaborator. Half of this thesis can be summarised as the application and imitation of your codes, which are the best programming textbooks and references for me.

I am grateful to Prof. Klypin and Prof. Prada for creating the great code GLAM and generously allowing us to implement modified gravity modules based on it.

Special thanks is given to my former MPhys supervisor, Prof. Hu, for inspiring me to work on the exciting fields of non-linear structure formation and modified gravity.

Thanks ICC for providing endless possibilities and opportunities in the research of the large-scale structure. Thanks the COSMA support team for maintaining the essential electronic infrastructure. For me, the periods of COSMA downtime are the biggest holiday than Christmas.

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

感谢爸爸妈妈对我深沉而无条件的溺爱。

Contents

Declaration

The work described in this thesis was undertaken between 2019 and 2022 while the author was a research student under the supervision of Prof. Baojiu Li, Prof. Carlton M. Baugh and Dr. Alexander Eggemeier at the Institute for Computational Cosmology in the Department of Physics at the University of Durham. No part of this thesis has been submitted for any other degree at the University of Durham or any other university. The contents of this work have appeared in the following papers:

- **Ruan, Cheng-Zong**; Hernández-Aguayo, César, et al., "Fast full N-body simulations of generic modified gravity: conformal coupling models." *Journal of Cosmology and Astroparticle Physics* 2022.05 (2022): 018. (**Chapter** 2 and 3)
- **Ruan, Cheng-Zong**; Cuesta-Lazaro, Carolina, et al., "Towards an accurate model of small-scale redshift-space distortions in modified gravity." *Monthly Notices of the Royal Astronomical Society* 514.1 (2022): 440-459. (**Chapter** 3)

Elements of Chapter 2 have appeared in:

- Hernández-Aguayo, César; **Ruan, Cheng-Zong**, et al., "Fast full N-body simulations of generic modified gravity: derivative coupling models." *Journal of Cosmology and Astroparticle Physics* 2022.01 (2022): 048.

The author has closely collaborated on this paper.

Chapter 4 is based on the following papers that will be submitted to *Monthly Notices of the Royal Astronomical Society*:

- **Ruan, Cheng-Zong**; Arnold, Christian, et al. "Towards an emulator-based halo model in modified gravity - I. Halo concentration-mass relations and density profiles."
- Hernández-Aguayo, César; **Ruan, Cheng-Zong**, et al. "Towards an emulator-based halo model in modified gravity - II. Halo mass functions"

All figures shown in this thesis were made by the author, with the exception of Fig. 1.1, which was taken from Fig. 2 of Kitaura et al. (2016) by permission of *Monthly Notices of the Royal Astronomical Society*, Oxford University Press.

List of Figures

List of Tables

Nomenclature

AMR adaptive mesh refinement

BDM bound density maximum

CDM cold dark matter

cHMF cumulative halo mass function

CIC cloud-in-cell

CLPT convolutional Lagrangian perturbation theory

CMB cosmic microwave background

DGP Dvali-Gabadadze-Porrati

dHMF differential halo mass function

FFT fast Fourier transformation

FLRW Friedmann-Lemaitre-Robertson-Walker

GELU Gaussian error linear unit

GR general relativity

GSM Gaussian streaming model

HAM halo abundance matching

HMF halo mass function

HOD halo occupation distribution

LPT Lagrangian perturbation theory

LSS large-scale structure

MCMC Markov chain Monte Carlo

MG modified gravity

NFW Navarro-Frenk-White

PDE partial differential equation

PDF probability distribution function

PM particle mesh

ReLU Rectified Linear Unit

RSD redshift space distortion

SM streaming model

ST skewed student-T

STSM skew-T streaming model

TPCF two-point correlation function

Chapter 1

Introduction

1.1 Our knowledge and understanding of cosmic large-scale structure

The past four decades since the 1980s have seen enormous advances in the study of the large-scale structure (LSS) of the Universe — the distribution of galaxies, quasars, clusters of galaxies and other tracers of matter — both in terms of our knowledge of the large-scale matter distribution from cosmological observations, and in the understanding of its origin from theories and simulations. Spectroscopic surveys, such as the 2dFGRS (Peacock et al., 2001), SDSS (Tegmark et al., 2004), WiggleZ (Blake et al., 2011), VIMOS (Pezzotta et al., 2017), BOSS (Alam et al., 2017), eBOSS (Alam et al., 2021a) and DESI (DESI Collaboration et al., 2016), have been used to measure the distribution of these large-scale tracers in the Universe with a level of detail and on spatial scales that were unthinkable in the 1990s. At the same time, observations of the cosmic microwave background (CMB) (e.g. Hinshaw et al., 2013; Planck Collaboration et al., 2016a,b), by showing us the Universe when it was only about $380,000$ years old, have supported the bold theoretical ideas conceived in the 1980s about the contents of the Universe (non-baryonic dark matter) and the mechanisms that initially created the seeds of structure formation shortly after the Big Bang (inflation: see Peebles (2017) for a historic review).

Cosmological N-body (Angulo & Hahn, 2022) and hydrodynamical (Vogelsberger et al., 2020) simulations provide the critical link between the early, near-uniform universe and the rich structure observed at late times. This work has taken advantage of the dramatic increase in computing power to create more realistic virtual universes, by running simulations of the growth of cosmic structures showing how gravity and astrophysical processes can produce galaxies and larger structures from the primordial particle "soup". Taken together, these advances have led to the emergence of a

"standard model" of cosmology. The Λ cold dark matter (ΛCDM) cosmological model, in which gravity is described by Einstein's theory of general relativity (GR), provides an excellent fit to many cosmological observations, such as the CMB anisotropies (e.g. Planck Collaboration et al., 2016a), weak and strong gravitational lensing (e.g. Kilbinger, 2015; Porredon et al., 2021; Suyu et al., 2017), light element abundances predicted by Big Bang Nucleosynthesis (e.g. Alpher et al., 1948; Turner et al., 2021), and the large-scale clustering of galaxies (e.g. Alam et al., 2017, 2021a).

Fig. 1.1 illustrates how well this standard cosmological model reproduces the observed large-scale distribution of galaxies. The top and left wedge plots show sub-regions of the SDSS-III Baryon Oscillation Spectroscopy Survey (BOSS) (Alam et al., 2017). These slices through the 3D galaxy distribution reveal extremely rich structures. Galaxies, groups and clusters are linked together in a pattern of sheets and filaments, commonly known as the "cosmic web" (Libeskind et al., 2018). The bottom and right panels present the mock galaxy catalogues constructed using the halo abundance matching (HAM) technique applied to the BigMultiDark N-body simulation (Rodríguez-Torres et al., 2016), selected with magnitude limits and survey geometries that match those used in the observations. The similarity of the structures between simulations and observations is striking and is supported by quantitative comparisons of clustering, such as galaxy two-point correlation functions and power spectra. So, what is the big picture of the structure formation in the standard ΛCDM paradigm?

Dark matter is believed to dominate the matter content of the Universe and the gravitational forces on large scales. It is thought to consist of a new weakly interacting elementary particle that is beyond the current standard model of particle physics. Because these particles are required to have small random velocities at early times, they were dubbed cold dark matter (CDM). Hot dark matter, such as massive neutrinos, has been ruled out by the fact that the galaxy distribution from hot dark matter dominated simulations can not reproduce observations (White et al., 1983). The need for dark matter stems from the realisation in the 1930s that the mass of the luminous component in galaxy clusters is too small for its gravity to hold clusters together (Zwicky, 1933; Smith, 1936). Zwicky (1933) found that the dynamical mass of the Coma cluster, inferred from the observed velocity dispersion of galaxies, was estimated to be ~ 400 times larger than the mass of the stellar components. Zwicky (1933) was based on the Hubble's linear relationship between redshift and distance, where the author used a Hubble constant of $H_0 = 558\,\mathrm{km\,s^{-1}Mpc^{-1}}$ from Hubble & Humason (1931). Adopting the modern value of $H_0 \approx 67\,\mathrm{km\,s^{-1}Mpc^{-1}}$ (Planck Collaboration et al., 2016b), we see that the mass-to-light ratio was overestimated by a factor of $\sim 558/67 \approx 8.3$, reducing the ratio from ~ 400 to ~ 5. The corrected value is consistent with the

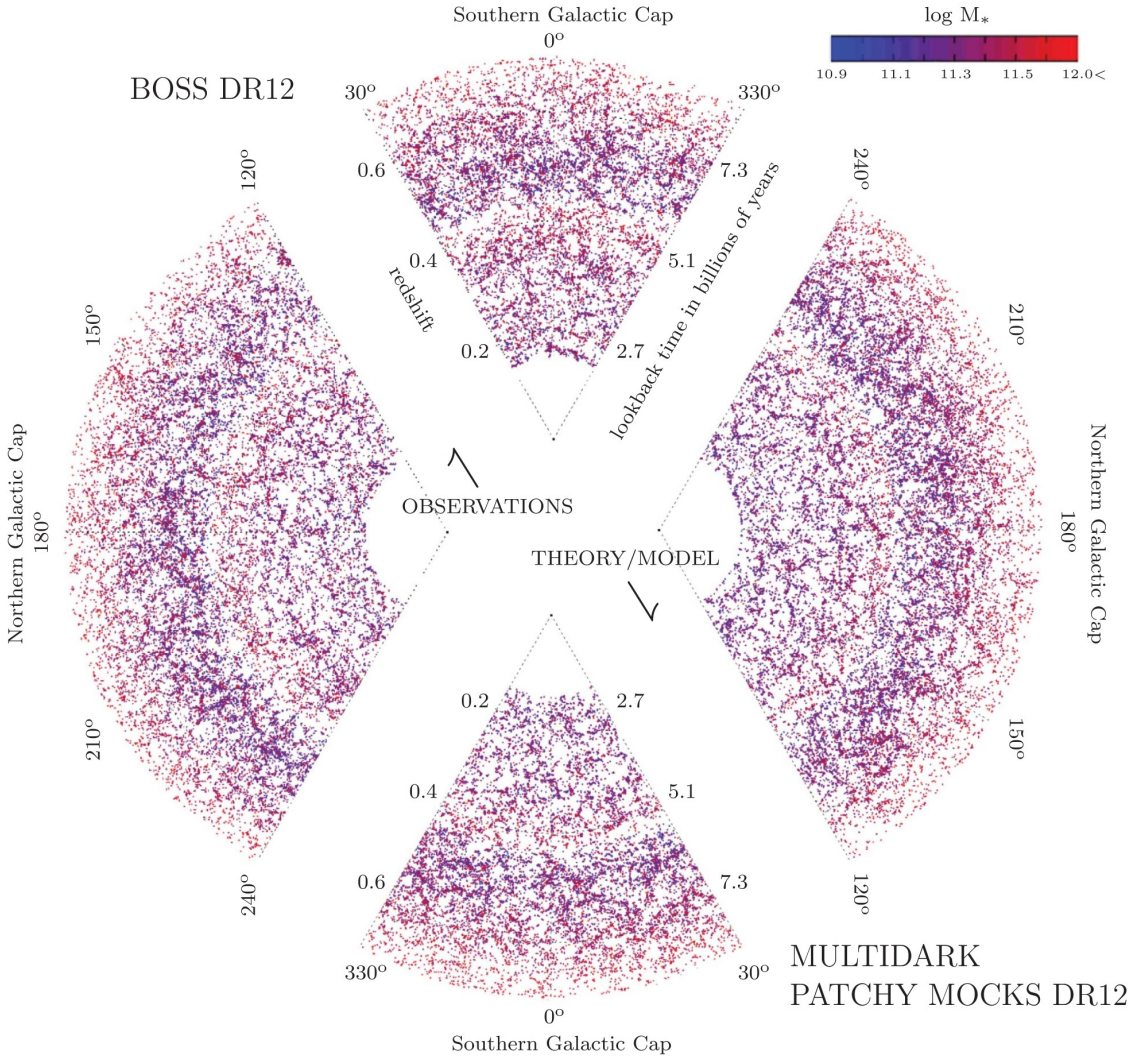

Figure 1.1: The large-scale galaxy distribution obtained from the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) data release 12 (upper and left region), and mock catalogues constructed from the MultiDark simulation (lower and right region), selected with the same selection criteria such as the survey mask and magnitude limits. Colours encode the (logarithmic) galaxy stellar mass in the unit of $M_{\odot}$. Image adapted from Fig. 2 of Kitaura et al. (2016) by permission of *Monthly Notices of the Royal Astronomical Society*, Oxford University Press.

ratio between the matter (including cold dark matter, baryon and massive neutrinos) and baryon density parameters inferred from cosmological observations. See Section III of Bertone & Hooper (2018) for a historical review of the discovery of dark matter.

Further evidence came in the 1970s when Rubin & Ford (1970) measured the rotation speeds of hydrogen gas in the outskirts of galaxies, which were approximately independent of radius, whereas the bulk of the light in the galaxies is concentrated towards the centres. This inference also received support from numerical studies (e.g. Ostriker & Peebles, 1973) which demonstrated that if most of the mass in a spiral galaxy were in a disk supported by circular motion, the disk would be unstable and prone to collapse, creating a chaotic flow. This instability can be avoided if the mass in the disk is subdominant to that in a spherical dark matter halo, in which the enclosed

mass increases roughly in proportion to the distance from the centre of the halo.

The CMB reflects the early-time Universe, when the density fluctuations of all matter species (such as photons, dark matter, baryons and neutrinos) were much smaller than today, $\delta \equiv (\rho - \bar{\rho})/\bar{\rho} \ll 1$, and so can be described accurately by linear perturbation theory. At the time the CMB was emitted there were none of the rich patterns and structures seen in Fig. 1.1. The structure formation in this linear regime is particularly simple, as all the non-linear terms in the dynamic equations can be ignored, and the matter perturbations on different scales evolve independently.

When the perturbations have grown sufficiently, the typical value of the density fluctuations could be substantially larger than unity. Perturbation theory approaches are no longer valid. The non-linear evolution can be fully solved only by numerical simulations, which will discuss in Section 1.3. Although no fully analytic solution exists, a number of approaches are possible in order to gain insight into the evolution of matter perturbations beyond linear order. Higher order perturbation theories, such as effective field theories of large-scale structure (e.g. Baumann et al., 2012), are valid in the quasi-linear regime, $\delta \lesssim 1$, and hence can be applied down to smaller scales than linear perturbation theory. A large fraction of the dark matter today resides in dark matter haloes, which are the building blocks of the LSS and provide the gravitational potential wells to accrete gas and form galaxies (White & Rees, 1978).

The *halo model* (Cooray & Sheth, 2002) of non-linear LSS assumes that each dark matter particle belongs to one and only one halo. Under this idealisation, the large-scale mass distribution can be studied in two steps: the distribution of mass within each halo, and the spatial distribution of the haloes themselves. As we will discuss in Section 1.4 and Chapter 4, we can use the basic dark matter halo properties — halo mass functions, halo density profiles and halo clustering statistics — as the model ingredients to explore the non-linear distribution of matter as well as being tracers of matter on large scales.

Let us now turn to 'Λ' of the ΛCDM model. At the close of the 20th century, observations of Type Ia supernovae in distant galaxies (Riess et al., 1998; Perlmutter et al., 1999) provided the first concrete evidence for the existence of dark energy — a mysterious component with negative pressure driving the expansion of the Universe to accelerate; in a universe dominated by matter, the expansion rate should always be slowing down. The cosmological constant, Λ, with a constant energy density ρ_Λ and equation of state $w = P/\rho = -1$, is favoured by many cosmological observations, over its extensions such as the models with time-varying equation-of-state parameters (e.g. Planck Collaboration et al., 2016b). Despite its simplicity, the concept of

dark energy is at least not as concrete and reliable as CDM, and Λ is responsible for, arguably, the most serious shortcomings of ΛCDM, as we will discuss in the next section.

1.2 Modified gravity as a possible alternative to dark energy

The cosmological constant (Λ), or more generally the dark energy, still lacks a reliable physical explanation, as the vacuum energy density predicted by the standard model of particle physics is many orders of magnitude larger than the value of Λ inferred from cosmological observations (Carroll, 2001). This suggests that a more fundamental theory is yet to be developed which can naturally explain the small value of Λ. The alternative theoretical models proposed so far can be roughly classified into two categories: those that involve some exotic new components beyond the standard model of particle physics, the so-called dark energy (Copeland et al., 2006), which usually has non-trivial dynamics; and the other which involve modifications to Einstein's GR on certain (usually cosmic) scales (Clifton et al., 2012; Joyce et al., 2015; Koyama, 2018), or introduces new fundamental forces between matter particles*. Leading examples include: quintessence (Ratra & Peebles, 1988; Wetterich, 1988; Zlatev et al., 1999; Steinhardt et al., 1999), k-essence (Armendariz-Picon et al., 2000, 2001), coupled quintessence (Amendola, 2000), $f(R)$ gravity (Sotiriou & Faraoni, 2010; De Felice & Tsujikawa, 2010) and chameleon model (Khoury & Weltman, 2004b,a; Mota & Shaw, 2007; Brax et al., 2008), symmetron model (Hinterbichler & Khoury, 2010; Hinterbichler et al., 2011), the Dvali-Gabadadze-Porrati (DGP) model (Davis et al., 2012), scalar (Nicolis et al., 2009; Deffayet et al., 2009) and vector (Heisenberg, 2014; Allys et al., 2016; Beltran Jimenez & Heisenberg, 2016) Galileons, K-mouflage (Babichev et al., 2009), and massive gravity (e.g., Hinterbichler, 2012). By considering different modified gravity (MG) models we can see how alternatives to GR might change measurable quantities, and therefore suggest which observables have the most potential to constrain deviations from GR.

Modifications to GR typically manifest themselves as changes to the cosmic expansion history and/or the evolution of structure, i.e., at the background and/or perturbation levels. Many viable MG models closely mimic the expansion history of ΛCDM, by construction, and therefore are hard to distinguish from GR using background cosmology alone. Even in the case of MG models where the expansion rate is modified, there can be degeneracies which cannot be fully broken using background observables. Hence, high hopes have been placed on the use of cosmological observations that involve perturbation dynamics to test gravity. The evolution of perturbations

*The two classes of models cannot always be clearly distinguished, and some of the modified gravity models studied here can also considered as coupled dark energy.

in linear theory in MG models has been well-studied both theoretically (e.g. Brax et al., 2011b; Barreira et al., 2012, 2015) and numerically (e.g. Hojjati et al., 2011; Hu et al., 2014; Bellini et al., 2018), and the model predictions have been confronted with observational data such as the CMB temperature fluctuations and the matter power spectrum (Hu et al., 2013; Dossett et al., 2014).

Nevertheless, on non-linear scales (e.g. $\lesssim 20\,h^{-1}\mathrm{Mpc}$ at the present day) where tremendous amounts of observational data are available, perturbation theory-based approaches break down and a fully non-linear treatment is needed in order to more accurately predict the model behaviour. An improved non-linear model is essential if, for example, one wishes to make the best use of the current and next generation cosmological surveys to test models. This point becomes even more acute in the context of MG cosmology, given that such models have intrinsically non-linear features, such as screening mechanisms and non-linear field equations for new degrees of freedom, which cannot be captured by linear theories (e.g. Li et al., 2013a).

For example, in typical galaxy redshift surveys, clustering measurements are most precise at the $\sim$ Mpc scale (Zhai et al., 2019), which is below the scales (20–25 $h^{-1}\mathrm{Mpc}$) on which state-of-the-art perturbation theories are valid (Chen et al., 2021). Such small scales offer critical information to improve model and parameter constraints, as demonstrated for standard ΛCDM in several recent works (e.g. Chapman et al., 2021); for the case of MG, we recently showed that including galaxy clustering down to 1–5 $h^{-1}\mathrm{Mpc}$ can also increase the precision of constraints by a factor of a few (Ruan et al., 2022). A fully non-linear treatment – N-body simulations – is essential to accurately describe the small-scale non-linear dynamics of cosmic structure formation. This will prove critical for a more complete exploitation of galaxy surveys to test gravity models. For this reason, here we focus on non-linear structure formation in MG cosmologies, with the objective of improving model tests by including data from deep into the non-linear regime of the LSS of the Universe.

The main hurdle to overcome when using N-body simulations is their expensive computational cost. A Monte Carlo Markov chain (MCMC) analysis, usually used to confront theoretical predictions with data, requires sampling at least 10^4–10^5 models in the cosmological and modified gravity parameter space. Such a large number of models is computationally prohibitive with simulations. The situation is even worse in MG scenarios, which usually involve highly non-linear partial differential equations governing the new physics, that must be solved if one aims to make the accurate theoretical predictions needed to test these models with incoming observational data. Current MG simulations can take between 2 to $\mathcal{O}(10)$ times longer than standard ΛCDM

simulations with the same specifications.

To overcome this hurdle and be prepared for exploring future observational data in gravity tests, this thesis presents an *emulator*-based halo model approach to modelling galaxy clustering statistics on non-linear scales, for a wide range of MG models of current interest, by combining the following numerical and statistical developments:

- a fast and full N-body simulation code **MG-GLAM** (Ruan et al., 2022; Hernández-Aguayo et al., 2022), which is able to reproduce the predictions of well-tested and high-fidelity codes such as MG-AREPO (Arnold et al., 2019a) and ECOSMOG (Li et al., 2012) with sufficiently high accuracy, whilst taking only a small fraction of the computation time.
- a parameter sampling algorithm in high-dimensional cosmological parameter space, *maximin distance sliced Latin hypercube designs* (**SLHD**, Ba et al., 2015; Bouhlel et al., 2019).
- a statistical technique to perform multidimensional regression — **neural networks** (Paszke et al., 2019) — to find approximate functions that react to the input parameters smoothly and predict quantities for new sets of parameters.

1.3 Dances with non-linearities: cosmological N-body simulations

1.3.1 Setting the stage: an expanding space-time with perturbations

On large scales, the universe is described by the homogeneous and isotropic Friedmann-Lemaitre-Robertson-Walker (FLRW) metric,

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right], \tag{1.1}$$

where we have employed natural units with $c = 1$ and the $(-,+,+,+)$ signature of the metric. The symmetries of this space-time reduce the 10 independent components of the metric tensor to a single function of time – the scale factor, $a(t)$, and a constant – the spatial curvature parameter k, which may take on the values -1 (for an open universe), $+1$ (closed), or 0 (flat). Throughout this thesis, we will only work in a spatially flat universe. Current cosmological observations strongly favour a flat universe, e.g., the combined *Planck* and baryon acoustic oscillation measurements (Planck Collaboration et al., 2020a).

Theoretically, it is more convenient to factorise the spatially flat FLRW metric into a static Minkowski metric $\eta_{\mu\nu}$ multiplied by a time-dependent factor, by introducing *conformal time*

defined as

$$\mathrm{d}\eta \equiv \frac{\mathrm{d}t}{a(t)}, \quad \eta(t) \equiv \int_0^t \frac{\mathrm{d}t'}{a(t')}, \tag{1.2}$$

and the metric becomes (in the Cartesian form)

$$g_{\mu\nu} = a^2(\eta) \begin{pmatrix} -1 & \\ & \delta_{ij} \end{pmatrix} = a^2(\eta)\, \eta_{\mu\nu}, \tag{1.3}$$

where δ_{ij} is the Kronecker delta symbol. Now we consider small perturbations around this flat FLRW space-time. The generic form of the perturbed metric is given by

$$g_{\mu\nu} = a^2(\eta) \begin{pmatrix} -1-2A & -B_i \\ -B_i & -2D\,\delta_{ij} + 2E_{ij} \end{pmatrix} \equiv a^2(\eta)(\eta_{\mu\nu} + h_{\mu\nu}), \tag{1.4}$$

where we have decomposed the spatial metric perturbation h_{ij} into a scalar part $D(\eta, x^i)$ and a trace-less part E_{ij}, which means $\delta^{ij}E_{ij} = 0$.

In GR perturbation theory, there are two coordinate transformations of interest: gauge transformations and background space-time transformations.

- In *gauge transformations*, the background coordinates remain the same, and the coordinates in the perturbed metric change. This means the correspondence between the points in the background and the perturbed space-time is changed. If we denote two different perturbed coordinated (two gauges) by $\hat{x}^\alpha$ and $\tilde{x}^\alpha$, then they are related by the so-called gauge transformation,

  $$\tilde{x}^\alpha = \hat{x}^\alpha + \xi^\alpha, \tag{1.5}$$

  where the gauge vector ξ^α and its derivatives $\xi^\alpha{}_{,\beta}$ are assumed to be first-order small quantities.

- In *background space-time transformations*, the gauge is fixed, but we do a coordinate transformation in the background space-time. This would consequently induce a coordinate transformation in the perturbed space-time. We are interested in the transformations which preserve the background symmetries, e.g., spatial rotation transformations, whose transformation matrices are given by

  $$x^{\mu'} = X^{\mu'}{}_{\rho} x^\rho, \quad X^{\mu'}{}_{\rho} = \begin{pmatrix} 1 & 0 \\ 0 & R^{i'}{}_{k} \end{pmatrix}, \tag{1.6}$$

  where $R^{i'}{}_{k}$ is a spatial rotation matrix, with the property of orthogonality, $R^{\mathrm{T}}R = I$, or $R^{i'}{}_{k}R^{i'}{}_{\ell} = \delta_{k\ell}$.

We can derive the transformation of the metric (Equation (1.4)) induced by the background transformation, and identify the transformations of the perturbation quantities,

$$\begin{cases} g_{\rho'\sigma'} = X^{\mu}{}_{\rho'} X^{\nu}{}_{\sigma'} g_{\mu\nu}, \\ g_{0'0'} = a^2(-1-2A), \\ g_{0'\ell'} = -a^2 R^{j}{}_{\ell'} B_j, \\ g_{k'\ell'} = a^2\Big(-2D\delta_{k\ell} + 2E_{ij} R^{i}{}_{k'} R^{j}{}_{\ell'}\Big). \end{cases} \Rightarrow \begin{cases} A' = A, \\ D' = D, \\ B_{\ell'} = R^{j}{}_{\ell'} B_j, \\ E_{k'\ell'} = R^{i}{}_{k'} R^{j}{}_{\ell'} E_{ij}. \end{cases} \tag{1.7}$$

Under the background spatial rotation transformation (keeping the gauge fixed), A and D transform as scalars; B_i transforms as 3D vectors and E_{ij} as 3D tensors. We can therefore treat them as scalar, vector and tensor fields in the *3D Euclidean background space*. We can further decompose B_i and E_{ij} into scalar (curl-free), pure-vector (divergence-free) and pure-tensor part. For B_i we have

$$B_i = -B_{,i} + B_i^{\mathrm{V}}, \text{ with } \delta^{ij} B^{\mathrm{V}}_{i,j} = 0. \tag{1.8}$$

Similarly, the symmetric trace-less tensor field E_{ij} is divided into three parts,

$$E_{ij} = \left(E_{,ij} - \frac{1}{3}\delta_{ij}\delta^{k\ell} E_{,k\ell}\right) - \frac{1}{2}(E_{i,j} + E_{j,i}) + E^{\mathrm{T}}_{ij}, \tag{1.9}$$

where $\delta^{ij} E_{i,j} = 0$, $\delta^{ik} E^{\mathrm{T}}_{ij,k} = 0$ and $\delta^{ij} E^{\mathrm{T}}_{ij} = 0$ (The tensor perturbation E^{T} is transverse and trace-less). After the full scalar-vector-tensor decomposition, the full perturbation metric is written as

$$g_{\mu\nu} = a^2(\eta_{\mu\nu} + h_{\mu\nu}), \tag{1.10}$$

$$h_{\mu\nu} = \begin{pmatrix} -2A & B_{,i} - B_i^{\mathrm{V}} \\ B_{,i} - B_i^{\mathrm{V}} & -2D\delta_{ij} + 2E_{,ij} - \frac{2}{3}\delta_{ij}\delta^{k\ell} E_{,k\ell} \\ & -(E_{i,j} + E_{j,i}) + 2E^{\mathrm{T}}_{ij} \end{pmatrix}. \tag{1.11}$$

The transformations of the perturbation quantities under rotations in background space are

$$\begin{aligned} A' = A, \quad B' = B, \quad D' = D, \quad E' = E, \\ B^{\mathrm{V}}_{\ell'} = R^{j}{}_{\ell'} B^{\mathrm{V}}_j, \quad E_{\ell'} = R^{j}{}_{\ell'} E_j, \\ E^{\mathrm{T}}_{k'\ell'} = R^{i}{}_{k'} R^{j}{}_{\ell'} E^{\mathrm{T}}_{ij}. \end{aligned} \tag{1.12}$$

We see that the 10 components of the metric perturbation $h_{\mu\nu}$ have been divided into

- 4 scalar degrees of freedom: A, B, D and E;
- 4 vector degrees of freedom: B_i^{V} and E_i along with the divergence-free constraints;

- 2 tensor degrees of freedom: E_{ij}^{T} with transverse and trace-less constraints.

We only consider scalar perturbations throughout this thesis. Firstly, the scalar, vector, and tensor perturbations do not couple to each other in first-order perturbation theory, as proved in Kodama & Sasaki (1984). The full evolution is a simple superposition of each components. Secondly, scalar perturbations couple to density and/or pressure perturbations of the matter fields, and exhibit gravitational instability. They are related to the structure formation. Vector perturbations couple to rotational velocity perturbations and tend to decay in an expanding universe, and therefore can be ignored. Tensor perturbations are nothing but gravitational waves propagating in an expanding background, which may have observable effects on very large scales, such as the anisotropies of the CMB.

So far, we have been working in a fixed gauge with scalar perturbations only. Bardeen (1980) found the following two combinations (called the *Bardeen potentials*) of scalar perturbations are invariant under gauge transformations,

$$\Phi \equiv A + \mathcal{H}\left(B - \frac{\partial E}{\partial \eta}\right) + \frac{\partial (B - \partial E/\partial \eta)}{\partial \eta}, \tag{1.13}$$

$$\Psi \equiv D + \frac{1}{3}\delta^{ij}E_{,ij} - \mathcal{H}\left(B - \frac{\partial E}{\partial \eta}\right), \tag{1.14}$$

where $\mathcal{H}(\eta) \equiv \dfrac{1}{a}\dfrac{\mathrm{d}a}{\mathrm{d}\eta}$ is the conformal Hubble parameter.

A gauge transformation $\tilde{x}^\mu = x^\mu + \xi^\mu$ can produce artificial vector perturbations. To avoid these and retain only scalar perturbations, we can again decompose the spatial components of the gauge vector $\xi^\mu = (\xi^0, \xi^i)$ as a scalar part and a pure-vector part,

$$\xi^i = \xi_{\mathrm{v}}^i - \delta^{ij}\xi_{,j}, \tag{1.15}$$

with $\vec{\nabla} \cdot \vec{\xi_{\mathrm{v}}} = 0$. ξ_{v}^i is responsible for the spurious vector perturbation. We will only consider these "scalar gauge transformations" that are fully specified by two scalar fields ξ^0 and ξ. The corresponding vector perturbation transformations are given by

$$\begin{aligned}
\tilde{A} &= A - \frac{\partial \xi^0}{\partial \eta} - \mathcal{H}\xi^0, \\
\tilde{B} &= B + \frac{\partial \xi}{\partial \eta} + \xi^0, \\
\tilde{D} &= D - \frac{1}{3}\delta^{ij}\xi_{,ij} + \mathcal{H}\xi^0, \\
\tilde{E} &= E + \xi.
\end{aligned} \tag{1.16}$$

There are still remaining degrees of gauge freedom. If we choose the gauge vector as

$$\xi = -E, \quad \xi^0 = -B + \frac{\partial E}{\partial \eta}, \tag{1.17}$$

then the scalar perturbations B and E are set to zero. Under this gauge transformation, we arrive at the commonly used *conformal Newtonian* gauge. In this gauge, we have

$$A = \Phi, \quad D = \Psi, \tag{1.18}$$

i.e., the Bardeen potentials are equal to the two nonzero metric perturbations in this gauge. From now on we shall work in the conformal-Newtonian gauge.

Let us end this subsection with a one-sentence summary. Starting from the generic form (Equation (1.4)) of the perturbed FLRW metric, assuming spatial flatness ($k = 0$), considering scalar perturbations only and using the conformal-Newtonian gauge, we finally arrive at the following space-time, which would serve as the stage for structure formation in our studies,

$$\mathrm{d}s^2 = a(\eta)^2\Big\{-[1+2\Psi(\eta,\boldsymbol{x})]\,\mathrm{d}\eta^2 + [1+2\Phi(\eta,\boldsymbol{x})]\delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\Big\}. \tag{1.19}$$

Note that there are a variety of notations for metric perturbations. Here we follow the convention of the standard cosmology textbook Dodelson & Schmidt (2020).

In Einstein's GR, the energy-momentum tensor is the source of the gravitational field, which determines the geometry of space-time described by the Einstein tensor $G_{\mu\nu}$. This concept is comprised in Einstein's field equation

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = T_{\mu\nu}, \tag{1.20}$$

where Λ is the cosmological constant and G is Newton's gravitational constant. Given a distribution function $f(\boldsymbol{x}, \boldsymbol{p}, t)$, the energy-momentum tensor in a curved space-time $g_{\mu\nu}$ is

$$T^{\mu}_{\ \nu}(\boldsymbol{x}, t) = \frac{g}{\sqrt{-\det[g_{\alpha\beta}]}} \int \frac{\mathrm{d}^3P^i}{(2\pi)^3} \frac{P^\mu P_\nu}{P^0} f(\boldsymbol{x}, \boldsymbol{p}, t), \tag{1.21}$$

where g is the degeneracy factor counting the different particle states as described by the distribution function; $P^\mu \equiv \dfrac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ is the comoving momentum; λ is the parameter of the particle's path; and $\boldsymbol{p} = (p^i)$ is the physical momentum, related to P^i by $p^i = aP^i$.

Putting the smooth part of the FLRW metric into the Einstein equation, we obtain the Friedmann equations describing the expansion of the background Universe,

$$H^2(t) = \frac{8\pi G}{3}\bar{\rho}(t) - \frac{kc^2}{a^2(t)}, \tag{1.22}$$

$$\frac{1}{a(t)}\frac{\mathrm{d}^2a(t)}{\mathrm{d}t^2} = -\frac{4\pi G}{3}[\bar{\rho}(t) + 3p(t)]. \tag{1.23}$$

where Hubble parameter is defined as $H(t) \equiv \frac{1}{a(t)}\frac{\mathrm{d}a(t)}{\mathrm{d}t}$; $\bar{\rho}(t)$ is the total density of all species in the cosmic inventory and the pressure of one component $p^{(i)}(t)$ is related with the density as $w^{(i)} \equiv p^{(i)}/\bar{\rho}^{(i)}$, where w is the equation-of-state parameter.

Plugging the perturbed FLRW metric Equation (1.19) into the Einstein equation and considering cold dark matter only (with the justification for this given in the next subsection), we arrive at the Poisson equation that describes relation between gravitational field and density configuration,

$$\nabla^2\Psi(\boldsymbol{x}, t) = 4\pi G a^2[\rho_{\mathrm{m}}(\boldsymbol{x}, t) - \bar{\rho}_{\mathrm{m}}(t)], \tag{1.24}$$

where $\Phi = -\Psi$ since the photons and neutrinos have been ignored.

1.3.2 N-body simulation technique

In principle, we can apply the similar procedure of constructing metric perturbations to the matter side of the Einstein equation, i.e., exploring the perturbations around a smooth energy-momentum tensor. However, in the non-linear regime of structure formation, such linear matter perturbation quantities are not as useful as their space-time counterparts. In the late universe with matter overdensity being highly non-linear, $\delta_{\mathrm{m}} \gg 1$, the metric is still close to FLRW and Equation (1.19) still works. That is why we adopt the statistical mechanics description using a phase space distribution function.

Collisionless cold dark matter particles dominate the non-linear clustering of the large-scale structure. In the ΛCDM model, the clustering components, more simply called matter, consist of CDM and baryons. According to the cosmological parameter constraints from the final *Planck* measurements of the CMB anisotropies (Planck Collaboration et al., 2020a), CDM makes up more than 80% of the total matter density. Moreover, the impact of electromagnetic interactions of baryons on structure formation is important only on very small scales. Baryons decoupled from the CMB photons well before the non-linear regime and cool rapidly. Thus, neglecting the non-gravity interactions of baryons, we will establish and solve the equations describing the evolution of collisionless cold dark matter particles under gravity, in the expanding space-time with perturbations as described in Equation (1.19).

Now we derive the equations governing the evolution of space-time and matter components. The former come from the Einstein equation with a perturbed FLRW metric, and the latter is obtained by taking moments of the Boltzmann equation. The collisionless Boltzmann equation, or the Vlasov equation, for CDM written in Cartesian coordinates is (cf. Equation (12.8) of Dodelson

& Schmidt (2020))

$$\frac{\mathrm{d}f_{\mathrm{m}}}{\mathrm{d}t} = \frac{\partial f_{\mathrm{m}}}{\partial t} + \frac{\partial f_{\mathrm{m}}}{\partial x^j}\frac{p^j}{ma} - \frac{\partial f_{\mathrm{m}}}{\partial p^j}\left(Hp^j + \frac{m}{a}\frac{\partial \Psi}{\partial x^j}\right) = 0, \tag{1.25}$$

where $f_{\mathrm{m}}(\boldsymbol{x}, \boldsymbol{p}, t)$ is the distribution function for matter, which counts the number of particles within an infinitesimal phase-space volume element $\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{p}$, m is the particle mass, p^i is the particle momentum.

To avoid solving for f_{m} in the 6D phase space $(\boldsymbol{x}, \boldsymbol{p})$, we can perform discrete Monte Carlo samplings of the distribution function and trace their evolution numerically. This is the N-body simulation technique. For any function $A(\boldsymbol{x}, \boldsymbol{p}, t)$ defined on the phase space, we can define the momentum average

$$\langle A \rangle_{f_{\mathrm{m}}}(\boldsymbol{x}, t) \equiv \int \frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3} A(\boldsymbol{x}, \boldsymbol{p}, t)\, f_{\mathrm{m}}(\boldsymbol{x}, \boldsymbol{p}, t). \tag{1.26}$$

For example, $\langle m \rangle_{f_{\mathrm{m}}}$ yields the sampled matter density field

$$\rho_{\mathrm{m}}(\boldsymbol{x}, t) = m \int \mathrm{d}^3\boldsymbol{p}\, f_{\mathrm{m}}(\boldsymbol{x}, \boldsymbol{p}, t). \tag{1.27}$$

We can find the dynamical equations of the matter field by taking the two lowest order moments of the the Vlasov equation. The zeroth moment is obtained by taking the momentum average of Equation (1.25) multiplied by m, and the first moment comes from multiplying Equation (1.25) with p^i and integrating over $\boldsymbol{p}$. The resulting particle equations of motions are given by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}a} = \frac{\boldsymbol{p}}{a^3 H}, \tag{1.28}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}a} = -\frac{\nabla\Psi}{aH}, \tag{1.29}$$

where we have switched the time variable from t to the scale factor a, and the Hubble parameter in the ΛCDM model is given by the Friedmann equation,

$$H(a) = H_0\sqrt{\Omega_{\mathrm{m}0}a^{-3} + \Omega_{\Lambda 0}}, \quad \text{for } \Omega_{\mathrm{m}0} + \Omega_{\Lambda 0} = 1. \tag{1.30}$$

In general, the N-body simulations have a simple logic: start from some initial conditions of the simulation particles $\left\{\boldsymbol{x}^{(i)}_{\mathrm{init}}, \boldsymbol{p}^{(i)}_{\mathrm{init}}\right\}_{i=1}^{N_{\mathrm{particles}}}$,

1. use the particle positions $\left\{\boldsymbol{x}^{(i)}\right\}$ to obtain the density $\rho(\boldsymbol{x})$ of the computational domain,
2. solve the Poisson equation (1.24) on the mesh, and
3. evolve the particle velocities and positions to the next moment of time using Equations (1.29) and (1.28), respectively.

Initial Conditions

The statistical properties of the density fluctuations $\delta(\boldsymbol{x})$ at early times are known to be very close to a statistically homogeneous and isotropic Gaussian random field (e.g. Planck Collaboration et al., 2020b). The Fourier modes of the overdensity field, $\delta_{\boldsymbol{k}} = |\delta_{\boldsymbol{k}}|e^{i\phi}$ are independent of each other and have uniformly distributed random phases ϕ and the following distribution for the amplitude (Bardeen et al., 1986),

$$p(|\delta_{\boldsymbol{k}}|)\,\mathrm{d}|\delta_{\boldsymbol{k}}| = \frac{|\delta_{\boldsymbol{k}}|}{P(k)} \exp\left[-\frac{|\delta_{\boldsymbol{k}}|^2}{2P(k)}\right] \mathrm{d}|\delta_{\boldsymbol{k}}|\,, \tag{1.31}$$

where $P(k)$ is the matter power spectrum. We can therefore draw values of ϕ from a uniform distribution in $[0, 2\pi]$, and $|\delta_{\boldsymbol{k}}|$ from Equation (1.31). To connect an initial matter overdensity field with the velocities of simulation particles, one can use the Zel'dovich approximation (ZA; Zel'Dovich 1970). Consider the displacement field $\boldsymbol{\chi}(\boldsymbol{x})$, which relates the initial (Lagrangian) coordinate $\boldsymbol{q}$ and late-time (Eulerian) coordinate $\boldsymbol{x}$ of a particle as

$$\boldsymbol{x}(t) = \boldsymbol{q} + D(z_{\rm init})\boldsymbol{\chi}(\boldsymbol{q}), \tag{1.32}$$

where $D(z)$ is the linear growth factor. The coordinate and velocity of a simulation particle at the initial time $z_{\rm init}$ are given by

$$\boldsymbol{x} = \boldsymbol{q} + D(z_{\rm init})\boldsymbol{\chi}, \quad \boldsymbol{v} = a_{\rm init}\left.\frac{\mathrm{d}D(z)}{\mathrm{d}t}\right|_{z_{\rm init}} \boldsymbol{\chi}. \tag{1.33}$$

In practice $\boldsymbol{q}$ can be taken as the grid points of a uniform Cartesian lattice (Efstathiou et al., 1985). In some implementations, the regular Cartesian lattice is replaced by a relaxed glass distribution of particles, which is obtained by evolving a random distribution of particles with the sign of gravity inverted, until an equilibrium state is reached; this can reduce the discreteness caused by placing particles initially on a (regularly spaced) grid (Baugh et al., 1995; White, 1996).

The Zel'dovich approximation used above to generate initial conditions can be considered as a first-order approximation, whose accuracy becomes worse when applied at lower initial redshifts, because at late times spurious high-order corrections become more important; with a late start these transients have less time to die away. As an example, for $z_{\rm init} \sim 50$ the ZA scheme can give few-percent inaccuracy in the initial particle distribution, which will propagate into late-time clustering statistics through the N-body simulation. One way to improve on this is to include second-order corrections following higher-order Lagrangian perturbation theories, such as the second-order (2LPT; Crocce et al. (2006b)) and third-order one (3LPT, Hahn et al. (2021); Michaux et al. (2021)); alternatively, the simulation can be started at higher $z_{\rm init} \sim 100$, where the ZA works better.

Gravitational Potential

With a density field $\rho(\boldsymbol{x})$ constructed from simulation particles, we can estimate the gravitational potential by solving the Poisson equation. We focus on the particle mesh (PM) technique as it is adopted by our base code GLAM. The computational domain in a PM code is a regular mesh with $N_{\rm g}$ grids in each direction and a box size L. We start by performing a 3D fast Fourier transformation (FFT) to the density field, which returns the Fourier components $\tilde{\rho}_{\boldsymbol{k}}$ on the same grid. In Fourier space, the right hand side of the Poisson equation can be schematically written as the multiplication between the density harmonics $\tilde{\rho}_{i,j,k}$ and the Green's functions $G(\boldsymbol{k})$,

$$\tilde{\Psi}_{i,j,k} = 4\pi G \tilde{\rho}_{i,j,k} G(\boldsymbol{k}). \tag{1.34}$$

The simplest method to derive the Green's function is to consider $\tilde{\Psi}_{i,j,k}$ and $\tilde{\rho}_{i,j,k}$ as amplitudes of the Fourier decomposition of the gravitational potential in the computational volume, and then to differentiate the Fourier harmonics analytically. This gives

$$G_0(\boldsymbol{k}) = -\frac{1}{k_x^2 + k_y^2 + k_z^2} = -\left(\frac{L}{2\pi}\right)^2 \frac{1}{i^2 + j^2 + k^2}. \tag{1.35}$$

where $G(\boldsymbol{k})$ is the Green's function. A better way of solving the Poisson equation that was used in the GLAM code (Klypin & Prada, 2018) is to start with the finite-difference approximation of the Laplacian, using a second-order Taylor expansion for the partial derivatives,

$$\begin{aligned} \nabla^2\Psi &= \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} && (1.36)\\ &\approx \big[\Psi_{i+1,j,k} - 2\Psi_{i,j,k} + \Psi_{i-1,j,k} \\ &\quad + \Psi_{i,j+1,k} - 2\Psi_{i,j,k} + \Psi_{i,j-1,k} \\ &\quad + \Psi_{i,j,k+1} - 2\Psi_{i,j,k} + \Psi_{i,j,k-1}\big]/(\Delta x)^2, && (1.37) \end{aligned}$$

where $\Delta x \equiv L/N_{\rm g}$. This approximation leads to a large system of linear algebraic equations $A\vec{\Psi} = 4\pi\vec{\rho}$, where A is the matrix of the coefficients. All of its diagonal components are equal to -6, and all six nearest off-diagonal components are unity. The solution of this matrix equation can be found by applying the Fourier Transformation, which provides another approximation for the Green's function:

$$G_1(\boldsymbol{k}) = \frac{(\Delta x)^2}{2}\left[\cos\left(\frac{2\pi i}{N_{\rm g}}\right) + \cos\left(\frac{2\pi j}{N_{\rm g}}\right) + \cos\left(\frac{2\pi k}{N_{\rm g}}\right) - 3\right]^{-1}. \tag{1.38}$$

For small indices (i, j, k), the above equation gives the same results as Equation (1.35). However, when (i, j, k) are close to $N_{\rm g}$, the finite-difference scheme Green's function G_1 provides less suppression for high-frequency harmonics and thus gives a stronger and more accurate force at distances closer to the grid spacing.

Time Integration

Once the initial condition is set up and the gravitational potential is solved, the remaining tasks are to evolve the particle velocities and positions. The commonly used scheme for the time integration of Equation (1.29) and (1.28) is the *leap-frog* method (Hockney & Eastwood 1988), which maps (x_n, v_n) to (x_{n+1}, v_{n+1}) as

$$v_{n+\frac{1}{2}} = v_n - \frac{\nabla\Psi_n}{aH}\frac{\Delta t}{2}, \tag{1.39}$$

$$x_{n+1} = x_n + v_{n+\frac{1}{2}}\Delta t, \tag{1.40}$$

$$v_{n+1} = v_{n+\frac{1}{2}} - \frac{\nabla\Psi_{n+1}}{aH}\frac{\Delta t}{2}. \tag{1.41}$$

The leap-frog is a structure-preserving symplectic integration method (e.g. Saha & Tremaine, 1992; Cohen et al., 2003) that observes important special properties of Hamiltonian systems. Such systems have first conserved integrals (such as the energy), they also exhibit phase-space conservation as described by the Liouville theorem.

1.4 Dances with non-linearities: the halo model approach

The idea of the halo model can be traced back to the work of Neyman & Scott (1952), where the authors built a theory of the spatial distribution of galaxies, assuming that galaxies only reside in clusters (equivalent to dark matter haloes in the modern framework). The problem of galaxy clustering is then split into the following questions:

- the distribution of cluster sizes, equivalent to the abundance of haloes, i.e. the halo mass function (HMF);
- the distribution of points around the cluster centre, equivalent to the density profile of haloes; and
- the clustering of clusters, equivalent to the clustering of haloes, e.g., the halo two-point correlation function (TPCF) in configuration space and the power spectrum in Fourier space.

On large scales, galaxies are simple tracers of the underlying matter field, as the statistical relation between the distribution of galaxies and matter, i.e. the galaxy *bias*, is relatively straight forward. The complicated physics of galaxy formation can be absorbed into a perturbative bias expansion with the cost of a finite set of bias parameters (see Desjacques et al. 2018 for a comprehensive review). On small scales ($\lesssim 20\,h^{-1}\mathrm{Mpc}$), however, the effect of galaxy formation can not

be bypassed using perturbative approaches anymore. We currently still lack a fully predictive theory for galaxy formation and evolution, in which various physical processes play an important role, such as gravitational instability, gas cooling, star formation and feedback processes. Moreover, the state-of-the-art modelling requires computationally expensive hydrodynamical simulations to make predictions, and are thus not suited for exploring and constraining large parameter spaces. Alternatively, semi-analytical models of galaxy formation can be used; whilst these require more approximations to be made than are required in hydrodynamical simulations, they can explore large parameter space (e.g. Elliott et al., 2021).

The halo model approaches divide the roles of cosmology and galaxy formation: cosmology dominates the halo formation and distribution, while galaxy formation determines how galaxies occupy haloes. Halo-based models assume that galaxies form and live inside dark matter haloes, which enables the quantitative modelling of galaxy clustering on small scales. These models are based on the fact that gravity is the dominant physical process in the formation of haloes. For example, the differences in the halo mass functions extracted from dark matter only N-body simulations and hydrodynamical simulations are at the few per cent level and almost independent of mass (Cui et al., 2012). Therefore, we can compute the statistical properties of haloes with N-body simulations accurately and robustly. Halo models then adopt a parametrisation to connect galaxies to haloes, thus bypassing the need to understand galaxy formation physics.

The connection between galaxy halo occupation and galaxy clustering was first made by semi-analytic models of galaxy formation (Baugh et al., 1999; Kauffmann et al., 1999), which provide the connection between galaxy clustering and halo occupation without requiring expensive hydrodynamical simulations (Somerville & Primack, 1999; Cole et al., 2000; Baugh, 2006; Benson, 2010). Basically, semi-analytic models populate N-body simulations with galaxies by solving a set of coupled differential equations which describe the cooling of gas in haloes, the formation of stars and black holes, feedback processes, metal enrichment, etc. It was realised that galaxy clustering, to a first approximation, is determined by halo occupation statistics as a function of halo mass, instead of all the details of galaxy formation (Benson et al., 2000). A series of works then sought to combine halo properties with occupation statistics to calculate the correlation function and power spectrum of galaxies on all scales (e.g. Peacock & Smith, 2000; Seljak, 2000; Scoccimarro et al., 2001). Berlind & Weinberg (2002) focused on the halo occupation distribution (HOD), which is the complete parametrisation of the galaxy-halo connection, and they studied how the HOD affects several galaxy clustering statistics and made a road-map to empirically constrain the HOD with galaxy clustering observations from spectroscopic surveys. The halo model has since been adopted

by a large number of studies to model galaxy clustering data in several galaxy redshift surveys, and constraint the relation between galaxy properties and the host haloes. Some leading examples include the TPCF of low-redshift galaxies and luminous red galaxies (Blake et al., 2008; Watson et al., 2010; White et al., 2011; Nikoloudakis et al., 2013; Reid et al., 2014).

The halo model ingredients — HMFs, density profiles and halo TPCFs — can in principle be derived from analytic methods. For example, the HMF can be predicted by the spherical collapse model of the linear matter density field (Press & Schechter, 1974) and the excursion set formalism (Bond et al., 1991; Sheth et al., 2001a). In these analytical approaches, all dependence of the HMF on redshift and cosmology is completely encoded in the root-mean-square fluctuations in the linear matter power spectrum. Jenkins et al. (2001) found that this universality of HMFs holds at an approximate level, given a certain halo mass definition. As simulation results were further refined, it was discovered that the redshift evolution of the mass function, even for CDM, deviates from the universal prediction (at the 5%-10% level), and several fitting formulas were proposed (e.g. Tinker et al., 2008; Courtin et al., 2011). Moreover, the performance of the universal HMF is further degraded in the extensions of ΛCDM (e.g. Bhattacharya et al., 2011, for wCDM) and modified gravity models (e.g. Lam & Li, 2012; Lombriser et al., 2013; Cataneo et al., 2016; Gupta et al., 2022).

In this thesis, we develop simulation-based theoretical templates, called *emulators*, with the aim to obtain accurate model predictions for halo properties as functions of halo mass in a range of cosmologies. To develop the emulator, we use a large number of MG N-body simulations for a set of cosmologies sampling a bounded parameter space. We then adopt the machine-learning-based interpolation technique, neural networks, to yield the desired predictions for halo properties for any set of parameters contained within the sampled region.

Following the philosophy of the Dark Quest project (Nishimichi et al., 2019a), we aim at constructing separate emulators for the halo model ingredients and assembling them in an analytical manner for different observables, instead of making an end-to-end mapping between the cosmological parameters to the final statistical quantities with the emulation process as in some previous work (e.g. Heitmann et al., 2006; Habib et al., 2007; Heitmann et al., 2010; Petri et al., 2015; Liu & Madhavacheril, 2019; DeRose et al., 2019; Zhai et al., 2019). Furthermore, this "numerical" version of the halo model allows us the flexibility of combining with any specific HOD prescription for different types of galaxies.

1.5 Outline of this thesis

In this thesis, we aim to develop a working pipeline to constrain cosmological parameters as well as the law of gravity with non-linear galaxy clustering. A key component of this pipeline is percent-level accuracy of the theoretical predictions of redshift-space galaxy clustering in a wide range of gravity models, such as $f(R)$ gravity, symmetron, K-mouflage and the DGP brane-world model. The pipeline is designed by providing the cosmological tools needed to explore the impact of modifications of gravity on the formation, evolution and clustering of galaxies using large and fast N-body cosmological simulations and small but high-resolution high-fidelity simulations.

In **Chapter 2**, we present MG-GLAM, a code for very fast production of full N-body simulations in a wide range of MG scenarios. As an application of the MG-GLAM products, **Chapter 3** shows how to use the improved streaming model to predict the redshift-space clustering of dark matter haloes down to small scales. In **Chapter 4**, we present the emulator-based halo model for non-linear galaxy-clustering, and demonstrate the accurate emulators for the halo properties that serve as the model ingredients. Finally, we summarise and conclude in **Chapter 5**, and also give a outlook of future projects that will complement the results presented in this thesis.

In the remaining chapters of this thesis, we adopt the usual conventions that Greek indices label all space-time coordinates ($\mu, \nu, \cdots = 0, 1, 2, 3$), while Latin indices label the space coordinates only ($i, j, k, \cdots = 1, 2, 3$). Our metric signature is $(-, +, +, +)$. We will strive to include the speed of light c explicitly in relevant equations, rather than setting it to 1, given that in numerical implementations c must be treated carefully. Unless otherwise stated, the symbol $\approx$ means 'approximately equal' or 'equal under certain approximations as detailed in the text', while the symbol $\simeq$ means that two quantities are of a similar order of magnitude. An over-dot denotes the derivative with respect to the cosmic time t, e.g., $\dot{a} \equiv \dfrac{\mathrm{d}a}{\mathrm{d}t}$ and the Hubble expansion rate $H(a)$ is defined as $H = \dfrac{\dot{a}}{a}$, while a prime ($'$) denotes the derivative with respect to the conformal time η, e.g., $a' = \dfrac{\mathrm{d}a}{\mathrm{d}\eta}, \mathcal{H}(a) \equiv \dfrac{a'}{a} = a\,H(a)$. Unless otherwise stated, we use a subscript 0 to denote the present-day value of a physical quantity, an over-bar for the background value of a quantity, and a tilde for quantities written in code units.

Chapter 2

From gravitational action to matter field: MG-GLAM I

This chapter covers both theoretical and numerical aspects of simulating cosmic structure formation in MG theories. The first two sections briefly describe the modified gravity models that were implemented in the MG-GLAM code (Ruan et al., 2022; Hernández-Aguayo et al., 2022). We start by introducing the generic Horndeski theory which has an additional scalar degree of freedom with second-order dynamical equations, in Section 2.1. We focus on one of the representative sub-classes in Section 2.2: scalar fields with conformal-coupling interactions with ordinary matter.

After setting the theoretical background, we turn to the numerical side, by firstly describing the N-body code GLAM (Klypin & Prada, 2018) for the standard ΛCDM model in Section 2.3. In Section 2.4, we present the numerical implementations to solve the MG scalar field equations additional to the standard gravity, including the code and data structure, the implementation of the multigrid relaxation method to solve the scalar field equations, and the tailored relaxation algorithms for each model. In Section 2.5, we show a series of code tests, which help us to verify the accuracy and reliability of the equation solvers.

In Appendix A, we present more detailed tests of MG-GLAM's power spectrum and HMF predictions at different force and mesh resolutions, and demonstrate that the increase of force resolution can lead to further improvement of the small-scale and low-mass predictions.

2.1 Theory: scalar field models for cosmic acceleration

Because of their simplicity, scalar fields are one of the most popular building blocks for models of dark energy (which acts as a new component species) and modified gravity (which mediates

additional forces between matter). This simplicity also makes it easier to design scalar field models that satisfy observational constraints; however, this does not mean that the scalar field models are trivial or have limited variation. Quite the contrary, the different ways in which scalar fields can interact with themselves and other species, or couple to the space-time curvature, give them a very rich phenomenology.

We consider the most general theory of gravity in a 4-dimension space-time, with field equations containing the metric tensor and a scalar field and leading to a second order equations of motion. The corresponding Lagrangian density, known as the Horndeski theory action, is given by (Horndeski, 1974)

$$\begin{aligned}\mathscr{S} = &\int \mathrm{d}^4x\sqrt{-g}\,\mathscr{L}_{\rm grav} + \int \mathrm{d}^4x\sqrt{-g}\,\mathscr{L}_{\rm scalar} \\ &+ \sum_n \int \mathrm{d}^4x\sqrt{-g}\,\mathscr{L}^{(n)}_{\rm matter}\Big[\psi^{(n)}, g_{\mu\nu}, \phi\Big],\end{aligned} \tag{2.1}$$

where g is the determinant of the metric $g_{\mu\nu}$, $\psi^{(n)}$ is the n-th components (such as normal and dark matter), ϕ is the scalar field. In this choice of metric, the coupling between the scalar field and matter species is explicit, and the matter Lagrangian depends on ϕ. The gravity Lagrangian density is the standard Einstein form:

$$\mathscr{L}_{\rm grav} = \frac{M_{\rm Pl}^2}{2} R[g_{\mu\nu}], \tag{2.2}$$

where $M_{\rm Pl} = (8\pi G)^{-1/2}$ is the reduced Planck mass and $R[g_{\mu\nu}]$ is the Ricci scalar constructed from the metric $g_{\mu\nu}$. Therefore, this frame is called the *Einstein frame*. In this case, a free-falling test particle does not follow the geodesics of $g_{\mu\nu}$ but feels an additional force due to the interaction with ϕ, which is different from the four known types of the fundamental forces and therefore is called the *fifth-force*.

It is possible to make the matter Lagrangian not explicitly depends on scalar fields, by re-scaling the metric $g_{\mu\nu}$ conformally ($B = 0$) or disformally ($B \neq 0$) as

$$\hat{g}_{\mu\nu} = A(\phi, \mathscr{H})g_{\mu\nu} + B(\phi, \mathscr{H})\nabla_\mu\phi\nabla_\nu\phi, \tag{2.3}$$

where ∇ is the covariant derivative compatible with $g_{\mu\nu}$ (i.e. $\nabla_\alpha g_{\mu\nu} = 0$), $\mathscr{H} \equiv \frac{1}{2}\nabla^\alpha\phi\nabla_\alpha\phi$, and A, B are some general functions of ϕ and $\mathscr{H}$.

In the frame of $\hat{g}_{\mu\nu}$, the geodesics of matter particles are determined by the metric alone, but the gravity Lagrangian and the resulting Einstein field equation are no longer the standard form. This frame is called the *Jordan frame* or the *matter frame*.

Despite the apparently different forms of (gravity and matter) equations in the Einstein and Jordan frames, it is widely accepted that a frame change as in Equation (2.3) does not change the physical contents of a model, so that physical observables are independent of the frame used in the analyses. In this thesis, we mainly introduce models in the Einstein frame.

We are not going to list the full expression of the scalar field Lagrangian $\mathscr{L}_{\rm scalar}$ of the Horndeski action. Instead, we will focus three classes of MG models which introduce scalar-type degrees of freedom that have conformal-coupling interactions: coupled quintessence (Amendola, 2000), chameleon $f(R)$ gravity (Khoury & Weltman, 2004b,a; Hu & Sawicki, 2007), and symmetron (Hinterbichler et al., 2011; Davis et al., 2012) models, which have been implemented in MG-GLAM (Ruan et al., 2022). These models generally introduce the fifth force between matter particles, and the latter two can be considered as special examples of the former, but differ in that they can both employ screening mechanisms to evade Solar System constraints on the fifth force.

These models are representative and have been widely studied in recent years. Further more, the implementation of them can lead to prototype MG N-body simulation codes that can be modified to work with minimal effort for other classes of interesting models. For example, the MG-GLAM twin paper (Hernández-Aguayo et al., 2022) investigated two classes of scalar field MG models with derivative-coupling interactions, the Vainshtein-type and the Kmouflage-type gravity models, which employ the Vainshtein and Kmouflage screening mechanism, respectively. As an example of Vainshtein-type models, we will investigate the normal branch of the DGP braneworld model in the next chapter, which serves as a prototype for other classes of models such as Galileons, vector Galileons, generalised Galileons and kinetic-gravity braiding models.

2.2 Theory: conformal coupling models

Consider a general model where a scalar field, ϕ, couples to matter, described by the following action

$$\mathscr{S} = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R - \frac{1}{2}\nabla^\mu\phi\nabla_\mu\phi - V(\phi)\right] + \sum_i \int \mathrm{d}^4x\sqrt{-\hat{g}}\,\mathscr{L}^m\left[\psi_i, \hat{g}_{\mu\nu}\right]. \tag{2.4}$$

Here the first term contains the gravitational and scalar field action, where $V(\phi)$ the potential energy of the scalar field ϕ. The second term is the matter action, which sums over all component species, such as normal and dark matter, labelled by i, with ψ being the matter field and $\hat{g}_{\mu\nu}$ the metric that couples to it. In principle, $\hat{g}_{\mu\nu}$ can be different for different matter species, but we only consider models with universal $\hat{g}_{\mu\nu}$.

For conformally coupled scalar field models, the Jordan-frame metric $\hat{g}_{\mu\nu}$ and the Einstein-frame metric $g_{\mu\nu}$ are related by the following conformal re-scaling

$$\hat{g}_{\mu\nu} = A^2(\phi) g_{\mu\nu}. \tag{2.5}$$

Note that the relation between $g_{\mu\nu}$ and $\hat{g}_{\mu\nu}$ can be more complicated, e.g., including a disformal term ($B \neq 0$ in Equation (2.3)), but these possibilities are beyond the scope of the present work.

The scalar field is a dynamical and physical degree of freedom in this model, which is governed by the following equation of motion from the stationary-action principle,

$$\nabla^\mu \nabla_\mu \phi = \frac{\mathrm{d}A(\phi)}{\mathrm{d}\phi} (\rho_\mathrm{m} - 3P_\mathrm{m}) + \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}, \tag{2.6}$$

where ρ_m and P_m are respectively the density and pressure of non-relativistic matter (radiation species do not contribute due to the conformal nature of Equation (2.5)). We also define the coupling strength $\beta(\phi)$ as a dimensionless function of ϕ:

$$\beta(\phi) \equiv M_\mathrm{Pl} \frac{\mathrm{d} \ln A(\phi)}{\mathrm{d}\phi}. \tag{2.7}$$

Note the M_Pl in this definition, which is because ϕ has dimensions of mass. For later convenience, we shall define a dimensionless scalar field as

$$\varphi \equiv \frac{\phi}{M_\mathrm{Pl}}. \tag{2.8}$$

We can see from Equation (2.6) that, in addition to the self-interaction of the scalar field ϕ, described by its potential energy, $V(\phi)$, the matter coupling means that the dynamics of ϕ is also affected by the presence of matter. We can therefore define an *effective total potential* of the scalar field, $V_\mathrm{eff}(\phi)$, as

$$V_\mathrm{eff}(\phi) \equiv A(\phi)\rho_\mathrm{m} + V(\phi), \tag{2.9}$$

where we have used $P_\mathrm{m} = 0$ for matter. With appropriate choices of $V(\phi)$ and $A(\phi)$, the effective potential $V_\mathrm{eff}(\phi)$ may have one or more minima, i.e., $\mathrm{d}V_\mathrm{eff}/\mathrm{d}\phi = 0$ at $\phi = \phi_\mathrm{min}$. Provided that the shape of $V_\mathrm{eff}(\phi)$ is sufficiently steep around ϕ_min, as in some classes of models to be studied below, the scalar field can oscillate around it, and we can define a scalar field mass, m, as

$$m^2 \equiv \frac{\mathrm{d}^2 V_\mathrm{eff}\left(\phi_\mathrm{min}\right)}{\mathrm{d}\phi^2}. \tag{2.10}$$

For non-relativistic matter particles, the interaction with the scalar field introduces new terms in their geodesic equations,

$$\dot{u}^\mu + \frac{\dot{\phi}}{M_\mathrm{Pl}} u^\mu = -c \frac{\beta(\phi)}{M_\mathrm{Pl}} \nabla^\mu \phi, \tag{2.11}$$

where $u^\mu \equiv \mathrm{d}x^\mu/\mathrm{d}\tau$ is the 4-velocity, and overdot denotes the time derivative.

In the weak-field limit where the metric $g_{\mu\nu}$ can be written through the following line element,

$$\mathrm{d}s^2 = -(1+2\Phi)c^2\mathrm{d}t^2 + (1-2\Phi)\mathrm{d}x^i\mathrm{d}x_i, \tag{2.12}$$

where Φ is the Newtonian potential, we can approximately write Equation (2.11) as

$$\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla}\Phi - c^2\frac{\beta(\phi)}{M_{\rm Pl}}\boldsymbol{\nabla}\phi - \frac{\beta(\phi)}{M_{\rm Pl}}\dot{\phi}\,\dot{\boldsymbol{r}}, \tag{2.13}$$

where $\boldsymbol{r}$ is the physical coordinate of the particle and $\boldsymbol{\nabla}$ is the gradient with respect to the physical coordinate.

The gravitational potential Φ and the perturbation to the MG scalar field have small values in Newtonian N-body simulations. Some relativistic cosmological simulation codes, such as GRAMSES (Barrera-Hinojosa & Li, 2020a,b), go beyond the weak-field approximation by including higher-order terms of the gravitational potentials, but find the effect on small scales is indeed small.

Equation (2.13) summarises three of the key effects that a coupled scalar field can have on cosmic structure formation: (1) a *fifth force*, as given by the gradient of ϕ, (2) a *frictional force* that is proportional to $\dot{\phi}$ and the particle's velocity $\dot{\boldsymbol{r}}$ — this is similar to the usual 'frictional' force caused by the Hubble expansion H, but because H can be modified by the coupled scalar field too, we have a third effect through a modified H, which is implicit in Equation (2.13).

In the same limit, the scalar field equation of motion, Equation (2.6), can be simplified as

$$c^2\boldsymbol{\nabla}^2\phi \approx V_\phi(\phi) - V_\phi(\bar{\phi}) + A_\phi(\phi)\rho_{\rm m} - A_\phi(\bar{\phi})\bar{\rho}_{\rm m}, \tag{2.14}$$

where an overbar denotes the background value of a quantity, and $V_\phi \equiv \mathrm{d}V(\phi)/\mathrm{d}\phi$, $A_\phi \equiv \mathrm{d}A(\phi)/\mathrm{d}\phi$. In deriving Equation (2.14) we have used the weak field approximation, as well as the quasi-static approximation which enables use to neglect the time derivative of the scalar field perturbation, $\delta\phi \equiv \phi - \bar{\phi}$, compared with its spatial gradient, i.e., $|\ddot{\delta\phi}| \simeq |H\dot{\delta\phi}| \ll |\boldsymbol{\nabla}^2\delta\phi| = |\boldsymbol{\nabla}^2\phi|$, where $H \equiv \dot{a}/a$ is the Hubble expansion rate. It is important to note that we do not assume that $\ddot{\bar{\phi}} \ll |\boldsymbol{\nabla}^2\phi|$, because $\ddot{\bar{\phi}}$ and $H\dot{\bar{\phi}}$ can be significant in certain models such as coupled quintessence, where $\bar{\phi}$ can evolve by a large amount throughout the cosmic history.

The quasi-static approximation has been tested for the modified gravity theories considered in this paper, such as $f(R)$ gravity (Oyaizu, 2008; Bose et al., 2015) and symmetron (Llinares & Mota, 2014). Oyaizu (2008) performed a consistency check of this approximation for Hu-Sawicki $f(R)$ gravity (Hu & Sawicki, 2007), where the simulations were run in the quasi-static limit but

it was checked that the time derivative of the scalar field perturbation is generally 5–6 orders of magnitude smaller than its spatial derivative in amplitude. Bose et al. (2015) directly examined this approximation by running full simulations including the time derivative terms, and found that the effects of the scalar field time derivative terms can be safely ignored in Hu-Sawicki $f(R)$ gravity. For the symmetron model, the quasi-static approximation has also been widely used in previous literature, e.g., Davis et al. (2012); Brax et al. (2012c). Llinares & Mota (2014) ran simulations with non-static terms and found very little difference in the matter power spectrum with the quasi-static simulations. However, the local power spectrum (defined as the $P(k)$ for the filtered matter field) shows deviations of the order of 1%. Therefore, it is expected that the quasi-static approximation is valid for usual cosmological probes such as power spectra which we are interested in, but other properties may be affected.

According to these researches, the quasi-static approximation is valid for our cosmological analyses. The effects of the scalar field time derivatives are small enough that can be safely ignored for the non-linear evolution of dark matter fields.

Finally, the Newtonian potential Φ is governed by the following Poisson equation, again written under the weak-field and quasi-static approximations,

$$\boldsymbol{\nabla}^2\Phi \approx 4\pi G A(\bar{\phi})\left(\rho_{\rm m} - \bar{\rho}_{\rm m}\right), \tag{2.15}$$

where we note the presence of $A(\bar{\phi})$ in front of $\rho_{\rm m}$, which is because the coupling to the scalar field ϕ can cause a time evolution of the particle masses of non-relativistic species, therefore affecting the depth of the resulting potential well Φ. This is the fourth key effect a coupled scalar field can have on cosmic structure formation. In the models considered in this paper, either the scalar field perturbation is small such that $A(\phi) \simeq A(\bar{\phi})$, or the scalar field has a small amplitude ($|\varphi| \ll 1$) in the entire cosmological regime so that $A(\phi) \simeq 1$ and $A(\bar{\phi}) \simeq 1$.

Equations (2.13, 2.14, 2.15) are the three key equations to be solved in our N-body simulations.

2.2.1 Coupled quintessence

The behaviour of the coupled scalar field, as well as its effect on the cosmological evolution, is fully specified with concrete choices of the coupling function $A(\phi)$ and scalar potential $V(\phi)$. Such models are known as *coupled quintessence* (Amendola, 2000), and have been studied extensively in the literature, including simulation analyses.

With some choices of $A(\phi)$ and $V(\phi)$, the scalar field dynamics can become highly non-linear, such as in the symmetron and chameleon models described below. These models are often display very little evolution of the background scalar field ($|\Delta\varphi| \ll 1$) throughout the cosmic history so that the background expansion rate closely mimics that of ΛCDM; the spatial perturbations of φ can reach $|\delta\varphi| \simeq |\bar{\varphi}| \ll 1$. In other, more general, cases, the scalar field can have a substantial dynamical evolution, $|\Delta\varphi| \sim \mathcal{O}(1)$ and $|\delta\varphi| \ll |\bar{\varphi}|$, which allows deviations from the ΛCDM expansion history, and the fifth force behaves in a less non-linear way. This latter case is the focus in this subsection.

We consider an exponential coupling function

$$A(\phi) = \exp\left(\beta\frac{\phi}{M_{\rm Pl}}\right) = \exp\left(\beta\varphi\right), \tag{2.16}$$

and an inverse power-law potential

$$V(\phi) = \frac{M_{\rm Pl}^{\alpha}\Lambda^4}{\phi^{\alpha}} = \frac{\Lambda^4}{\varphi^{\alpha}}, \tag{2.17}$$

where α, β are dimensionless model parameters, and Λ is a model parameter with mass dimension 1 which represents a new energy scale related to the cosmic acceleration. For convenience, we define a dimensionless order-unity parameter λ as

$$\frac{\Lambda^4}{M_{\rm Pl}^2} = H_0^2\lambda^2. \tag{2.18}$$

We consider parameters $\alpha > 0$, so that $V(\phi)$ is a runaway potential and the scalar field rolls down $V(\phi)$, and $\beta < 0$ so that the effective potential $V_{\rm eff}(\phi)$ has no minimum and the scalar field can keep rolling down $V_{\rm eff}(\phi)$ if not stopped by other effects. This means that we can have $|\bar{\varphi}| \sim \mathcal{O}(1)$ at late times (as mentioned in the previous paragraph) and kinetic energy makes up a substantial fraction of the scalar field's total energy (so that its equation of state w_ϕ can deviate substantially from -1).

While we specialise to Equations ((2.16), (2.17)) for the coupled quintessence models in this paper, the MG-GLAM code that we will illustrate below using this model can be applied to other choices of $A(\phi)$ and $V(\phi)$ with minor changes in a few places, to allow fast, inexpensive and accurate simulations for generic coupled quintessence models.

For completeness and convenience of later discussions, we also present here the linear growth equation for matter density contrast δ (or equivalently the linear growth factor D_+ itself) in the above coupled quintessence model:

$$\delta'' + \left[\frac{a'}{a} + \frac{\mathrm{d}\ln A(\bar{\varphi})}{\mathrm{d}\varphi}\bar{\varphi}'\right]\delta' - 4\pi G\bar{\rho}_{\rm m}(a)a^2A(\bar{\varphi})\left(1+2\beta^2\right)\delta = 0, \tag{2.19}$$

where $'$ denotes the derivative with respect to the conformal time τ. According to this equation, there are four effects that the coupled scalar field has on structure formation:

(*i*) a modified expansion history, a'/a;

(*ii*) a fifth force whose ratio with respect to the strength of the standard Newtonian force is given by $2\beta^2$;

(*iii*) a re-scaling of the matter density field by $A(\varphi) \neq 1$ in the Poisson equation, implying that the matter particle mass is effectively modified; and

(*iv*) a velocity-dependent force described by the term involving $\frac{\mathrm{d}\ln A}{\mathrm{d}\varphi}\bar{\varphi}'\delta'$.

The ratio between the fifth and Newtonian forces can be derived as follows: Equation (2.14) can be approximately rewritten as

$$\boldsymbol{\nabla}^2\left(c^2\delta\varphi\right) \approx 8\pi G\beta A(\bar{\varphi})\left(\rho_{\mathrm{m}} - \bar{\rho}_{\mathrm{m}}\right), \tag{2.20}$$

where we have used $A_\phi = \frac{\beta}{M_{\mathrm{Pl}}}\exp(\beta\varphi)$, $M_{\mathrm{Pl}}^{-2} = 8\pi G$, and neglected the contribution the scalar field potential $V(\phi)$ in the field perturbation $\delta\varphi$. Then, from Equations (2.15) and (2.13), it follows that the ratio of the two forces is $2\beta^2$, which means that the fifth force always boosts the total force experienced by matter particles in this model. In addition, since β is a constant, from Equation (2.19) we can conclude that the enhancement to linear matter growth, i.e., in the linear growth factor and matter power spectrum, will be scale-independent.

2.2.2 Symmetrons

The *symmetron* model (Hinterbichler & Khoury, 2010; Hinterbichler et al., 2011) features the following potential $V(\phi)$ and coupling function $A(\phi)$ for the scalar field:

$$V(\phi) = V_0 - \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\zeta\phi^4, \tag{2.21}$$

$$A(\phi) = 1 + \frac{1}{2}\frac{\phi^2}{M^2}, \tag{2.22}$$

where μ, M are model parameters of mass dimension 1, ζ is a dimensionless model parameter and V_0 is a constant parameter of mass dimension 4, which represents vacuum energy and acts to accelerate the Hubble expansion rate.

We can define

$$\phi_* \equiv \frac{\mu}{\sqrt{\zeta}}, \tag{2.23}$$

which represents the local minimum of the Mexican-hat-shaped symmetron potential $V(\phi)$. The total effective potential of the scalar field, however, is given in Equation (2.9). Because $A(\phi)$ is a quadratic function of ϕ, when $\rho_{\rm m}$ is large, the effective potential is dominated by $A(\phi)\rho_{\rm m}$, with single global minimum at $\phi = 0$; but when $\rho_{\rm m}$ is small, the effective potential is dominated by $V(\phi)$ and has two minima, $\pm\phi_{\rm min}$. Explicitly, it can be shown that $\phi_{\rm min} = 0$ when $\bar{\rho}_{\rm m} > \mu^2 M^2 \equiv \rho_*$ in background cosmology, while otherwise the symmetry in $V(\phi)$ is broken and the symmetron field solutions are given by

$$\pm\phi_{\rm min} = \sqrt{\frac{1}{\zeta M^2}\left(\rho_* - \bar{\rho}_{\rm m}\right)}, \tag{2.24}$$

from which we can confirm the above statement that as $\bar{\rho}_{\rm m} \to 0$ we have $\phi_{\rm min} \to \phi_*$. Because ρ_* has the dimension of density, it is more convenient to express it in terms of a characteristic scale factor a_* or redshift z_* corresponding to the time of symmetry breaking in $V_{\rm eff}(\phi)$:

$$\rho_* = \bar{\rho}_{\rm m0} a_*^{-3}, \tag{2.25}$$

where $\bar{\rho}_{\rm m0}$ is the background matter density today. According to Equation (2.24), as $\rho_{\rm m} \to 0$, $\phi_{\rm min} \to \phi_*$, i.e., $\phi_{\rm min}$ approaches the minimum of $V(\phi)$. Therefore, we must have $\phi_{\rm min} \in [0, \phi_*]$. For this reason we can define the following dimensionless variable

$$u \equiv \frac{\phi}{\phi_*} \in [0, 1), \tag{2.26}$$

Note that this is only true for background u, while in the perturbed case it is possible to have $u > 1$ in certain regions. Also, $u > 0$ is just a choice, because the symmetron field has two physically identical branches of solutions which differ by sign, and we choose the positive branch for simplicity*. In terms of the dimensionless scalar field φ, we have (Brax et al., 2012a)

$$\varphi_{\rm min}(a) = \varphi_* \sqrt{1 - \left(\frac{a_*}{a}\right)^3}, \tag{2.27}$$

with

$$\varphi_* \equiv \frac{\phi_*}{M_{\rm Pl}} = 6\Omega_{\rm m0}\beta_*\xi^2 a_*^{-3}, \tag{2.28}$$

where $\Omega_{\rm m0}$ is the matter density parameter today, $\xi \equiv H_0/m_*$ with m_* being the 'mass' of the scalar field at ϕ_*, given by

$$m_*^2 \equiv \frac{{\rm d}^2 V(\phi_*)}{{\rm d}^2\phi} = -\mu^2 + 3\zeta\phi_*^2 = 2\mu^2, \tag{2.29}$$

and β_* is a dimensionless parameter defined through

$$M_{\rm Pl}\frac{{\rm d}A}{{\rm d}\phi} = \frac{M_{\rm Pl}\phi}{M^2} \equiv \beta_*\frac{\phi}{\phi_*}, \tag{2.30}$$

*Indeed, it is possible that u can have different signs in different regions of the Universe, which are separated by domain walls, but we do not consider this more realistic possibility in this paper, as it does not have a big impact on the observables of interest to us.

which can be further expressed as

$$\beta_* \equiv \frac{M_{\rm Pl}}{M^2}\frac{\mu}{\sqrt{\zeta}} = \frac{M_{\rm Pl}m_*^2}{2\rho_*}\phi_*. \tag{2.31}$$

Therefore, the model can be fully specified by three dimensionless parameters — β_*, a_* (or z_*) and ξ — as opposed to the original, dimensional, parameters μ, ζ, M. We are interested in the regime of $\beta_*, a_* \sim \mathcal{O}(0.1)$ and $\xi \sim \mathcal{O}\left(10^{-3}\right)$. It is then evident from Equations (2.28) that $\varphi_* \ll 1$ and therefore $\varphi_{\rm min}(a) \ll 1$, confirming our claim above that in this model the scalar field has little evolution throughout the cosmic history. For simplicity we will assume that in the background the scalar field always follows $\varphi_{\rm min}$, namely $\bar{\varphi}(a) = \varphi_{\rm min}(a)$*. Further, because $\varphi_{\rm min} \simeq \varphi_* \ll 1$, we have

$$A(\phi) = 1 + \frac{1}{2}\beta_*\frac{\varphi}{\varphi_*}\varphi \simeq 1, \tag{2.32}$$

which implies that the time variation of particle mass is negligible in this model, and

$$\beta(\phi) = M_{\rm Pl}\frac{{\rm d}\ln A(\phi)}{{\rm d}\phi} \simeq \frac{{\rm dA}(\varphi)}{{\rm d}\varphi} = \beta_*\frac{\varphi}{\varphi_*} = \beta_* u, \tag{2.33}$$

so that β_* characterises the coupling strength between the scalar field and matter in this model.

With all the newly-defined variables, the scalar field equation of motion, Equation (2.6), in this model can be simplified as

$$c^2\nabla^2\frac{\varphi}{\varphi_*} = \frac{1}{2}\xi^{-2}H_0^2a^2\frac{\varphi}{\varphi_*}\left(\frac{\varphi^2}{\varphi_*^2} - 1\right) + \frac{1}{2}\xi^{-2}H_0^2a_*^3\frac{\varphi}{\varphi_*}\frac{\rho_m}{\bar{\rho}_{\rm m}}a^{-1}, \tag{2.34}$$

or equivalently

$$c^2\nabla^2 u = \frac{1}{2}\xi^{-2}H_0^2a^2u\left(u^2 - 1\right) + \frac{1}{2}\xi^{-2}H_0^2a_*^3u(1+\delta)a^{-1}, \tag{2.35}$$

where the density contrast is defined as

$$\delta \equiv \frac{\rho_{\rm m}}{\bar{\rho}_{\rm m}} - 1. \tag{2.36}$$

The symmetron model and its extensions have been studied with the help of numerical simulations in several works (Davis et al., 2012; Brax et al., 2012a), but the large computational cost has made it impossible to run large, high-resolution simulations for a very large number of parameter combinations, which is why we are implementing it in MG-GLAM. This model features the *symmetron screening mechanism* (Hinterbichler & Khoury, 2010), which helps to suppress the fifth force in dense environments by driving $\varphi \to 0$ so that the coupling strength $\beta(\phi) \to 0$,

*In practice, because $\varphi_{\rm min}(a)$ evolves with time, when trying to track it, $\bar{\varphi}$ can have oscillations around $\varphi_{\rm min}$ because $m_* \gg H(a) \simeq H_0$. Following most literature on the symmetron model, we will neglect these oscillations.

cf. Equation (2.33). This essentially decouples the scalar field from matter and therefore eliminates the fifth force in these environments, such that the model could evade stringent local and Solar System constraints. The *dilaton screening mechanism* (Brax et al., 2010) is another class of coupled scalar field models with a screening mechanism that works similarly, so in this paper we shall focus on the symmetron model only.

2.2.3 Chameleon $f(R)$ gravity

$f(R)$ gravity (Sotiriou & Faraoni, 2010; De Felice & Tsujikawa, 2010) is a very popular class of modified gravity models, which is described by the following gravitational action

$$S = \frac{M_{\rm Pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}\left[R + f(R)\right], \tag{2.37}$$

simply replacing the cosmological constant Λ with an algebraic function of the Ricci scalar, $f(R)$. It is well known that this theory can be equivalently rewritten as a scalar-tensor theory after a change of variable, and is therefore mathematically and physically equivalent to a coupled scalar field model in which the scalar field has a universal coupling to different matter species. Therefore, it belongs to the general models introduced in the beginning of this section. The model is fully specified by fixing the function $f(R)$, with different choices of $f(R)$ equivalent to coupled scalar field models with different forms of the scalar potential $V(\phi)$. Meanwhile, the coupling strength of the scalar field is a constant $\beta = 1/\sqrt{6}$ for all $f(R)$ models*, independent of $f(R)$. Despite this limitation, this model still has very rich phenomenology, and in this paper we will study it in the original form given by Equation (2.37), instead of studying its equivalent coupled scalar field model.

With certain choices of the function $f(R)$, the model can have the so-called *chameleon screening* mechanism (Khoury & Weltman, 2004b,a; Mota & Shaw, 2007; Brax et al., 2008), which can help the fifth force to hide from experimental detections in dense environments where $\rho_{\rm m}$ is high and the scalar field acquires a large mass m and therefore its strength decays exponentially and essentially vanishes beyond a typical distance of order m^{-1}. Of course, not all choices of $f(R)$ can lead to a viable chameleon screening mechanism, and in this paper we will focus only on those where the chameleon mechanism works, and we call the latter *chameleon $f(R)$ gravity*.

In $f(R)$ gravity, the Einstein equation is modified to

$$G_{\mu\nu} - X_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{2.38}$$

*This means that the ratio between the strengths of the fifth and the standard Newtonian forces is at most $1+2\beta^2 = 1/3$. For more details see below.

where $T_{\mu\nu}$ is the energy-momentum tensor, $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ is the Einstein tensor with $R_{\mu\nu}$ being the Ricci tensor, and $X_{\mu\nu}$ is defined as

$$X_{\mu\nu} \equiv -f_R R_{\mu\nu} + \frac{1}{2} \left[f(R) - \nabla^\alpha \nabla_\alpha f_R \right] g_{\mu\nu} + \nabla_\mu \nabla_\nu f_R, \tag{2.39}$$

where $f_R \equiv \mathrm{d}f(R)/\mathrm{d}R$ is a new dynamical scalar degree of freedom, with the following equation of motion

$$\nabla^\mu \nabla_\mu f_R = \frac{1}{3} \left[R - f_R R + 2f(R) - 8\pi G \rho_\mathrm{m} \right]. \tag{2.40}$$

One of the leading choices of the function $f(R)$ was the one proposed by Hu & Sawicki (2007). In this paper, instead of using the original function form provided in Hu & Sawicki (2007), we present it in an approximate form which will allow us to generalise it. Let's start with the following expression of $f_R(R)$,

$$f_R(R) = -\left| f_{R0} \right| \left(\frac{\bar{R}_0}{R} \right)^{n+1}, \tag{2.41}$$

where f_{R0} is the present-day value of the background f_R, $\bar{R}_0$ is the background Ricci scalar today, and $n \geq 0$ is an integer. For $n > 0$, the functional form $f(R)$ can be written as

$$f(R) \approx -6H_0^2 \Omega_{\Lambda 0} + \frac{1}{n} \left| f_{R0} \right| \left(\frac{\bar{R}_0}{R} \right)^{n+1} R, \tag{2.42}$$

where $\Omega_{\Lambda 0} = 1 - \Omega_{\mathrm{m}0}$ and the first term represents a cosmological constant that is responsible for the cosmic acceleration. For $n = 0$, we have

$$f(R) \approx -6H_0^2 \Omega_{\Lambda 0} + \left| f_{R0} \right| \bar{R}_0 \ln \left(\frac{\bar{R}_0}{R} \right). \tag{2.43}$$

Most of the simulation works to date have been performed for the case of $n = 1$, while the cases of $n = 2$ and $n = 0$ are not as well explored. In this paper we implement all three cases into MG-GLAM.

In general, for the model to have a viable chameleon screening, the parameter f_{R0} in Equation (2.41) should satisfy $|f_{R0}| \ll 1$. At late times, when $\bar{R}(a) \simeq \bar{R}_0$, we can see from Equations (2.42) and (2.43) that the relation $f(R) \simeq -6H_0^2 \Omega_{\Lambda 0}$ holds. On the other hand, from Equation (2.41) we have $|f_R| \ll 1$ throughout the cosmic history, i.e., it has a negligible evolution in time. This implies that all the terms in $X_{\mu\nu}$ in Equation (2.39) other than $f(R)$ can be neglected compared with the $f(R)$ term, and so the model behaves approximately like ΛCDM in the background expansion rate, with the background Ricci scalar given by

$$\bar{R}(a) = 3\mathbb{M}^2 \left(a^{-3} + 4 \frac{\Omega_{\Lambda 0}}{\Omega_{\mathrm{m}0}} \right), \tag{2.44}$$

and $\mathbb{M}^2 \equiv H_0^2\Omega_{\rm m0}$. This is compatible with what we mentioned above, i.e., in the coupled scalar field model that is equivalent to these $f(R)$ models, the scalar field ϕ has little time evolution and therefore has an equation of state which is very close to -1. It also implies that the weak-field approximation, where we can neglect the time evolution of the scalar degree of freedom f_R, is a good approximation, so that in an inhomogeneous Universe we have

$$\boldsymbol{\nabla}^2 f_R \approx \frac{1}{3c^2}\left(\delta R - 8\pi G\delta\rho_{\rm m}\right)a^2, \tag{2.45}$$

where $\boldsymbol{\nabla}$ is the gradient with respect to the comoving coordinate, as before,

$$\delta\rho_{\rm m} \equiv \rho_{\rm m} - \bar{\rho}_{\rm m} = \bar{\rho}_{\rm m}\delta, \tag{2.46}$$

and

$$\delta R = R - \bar{R}. \tag{2.47}$$

By realising that Equation (2.41) can be inverted to give

$$R = \bar{R}_0\left(\frac{f_{R0}}{f_R}\right)^{\frac{1}{n+1}}. \tag{2.48}$$

Then, Equation (2.45) becomes a non-linear dynamical equation for f_R with Equation (2.48).

Also under the quasi-static and weak-field approximations, the Poisson equation takes the following modified form

$$\begin{aligned}\boldsymbol{\nabla}^2\Phi &\approx \frac{16\pi G}{3}\delta\rho_{\rm m}a^2 - \frac{1}{6}\delta R a^2 \\ &= 4\pi G\bar{\rho}_{\rm m}a^2\delta - \frac{1}{2}c^2\boldsymbol{\nabla}^2 f_R,\end{aligned} \tag{2.49}$$

where in the second step we have used Equation (2.45).

One can have a quick peek into two opposite regimes of solutions for Equations (2.45) and (2.49). In the large field limit, when $|f_R|$ is relatively large (e.g., in the case of large $|f_{R0}|$), the perturbation δf_R is small compared to the background field $|\bar{f}_R|$, and $|\delta R| \ll 8\pi G\delta\rho_{\rm m}$, so that the Poisson equation (2.49) can be approximated as

$$\boldsymbol{\nabla}^2\Phi \approx \frac{16\pi G}{3}\delta\rho_{\rm m}a^2. \tag{2.50}$$

Comparing this with the standard Poisson equation in ΛCDM,

$$\boldsymbol{\nabla}^2\Phi \approx 4\pi G\delta\rho_{\rm m}a^2, \tag{2.51}$$

we confirm that the fifth force, i.e., the enhancement of gravity, is $1/3$ of the strength of the standard Newtonian force. In the opposite, small-field, limit where $|f_R|$ takes very small values, the left-hand side of Equation (2.45) is negligible and so we have $\delta R \approx 8\pi G\delta\rho_{\rm m}$, and plugging this into Equation (2.49) we recover Equation (2.51): this is the screened regime where the fifth force is strongly suppressed.

2.2.4 Summary and comments

In this section we have briefly summarised the essentials of the three classes of scalar field modified gravity models to be considered in this work. Among these, coupled quintessence is technically more trivial, because the fifth force is unscreened nearly everywhere, while $f(R)$ gravity and symmetrons are both representative thin-shell screening models Brax et al. (2012b) featuring two of the most important screening mechanisms respectively. Compared with previous simulation work, we will consider $f(R)$ models with more values of the parameter n: as discussed below, instead of the common choice of $n = 1$, we will also look at $n = 0$ and 2 to see how the phenomenology of the model varies.

We remark that, even with the additional modified gravity models implemented in the MG-GLAM papers (Ruan et al., 2022; Hernández-Aguayo et al., 2022), we are still far from covering all possible models. Changing the coupling function $A(\varphi)$ or the scalar field potential $V(\varphi)$, for example, will lead to new models. However, our objective is to have an efficient simulation code that covers different *types* of models, which serves as a 'prototype' that can be very easily modified for any other models belonging to the same type. This differs from the model-independent (Srinivasan et al., 2021) or parameterised modified gravity (Hassani & Lombriser, 2020) approaches adopted elsewhere, and we prefer this approach since there is a direct link to some fundamental Lagrangian here, and because, any parameterisation of models, one its parameters specified, also corresponds to a fixed model.

2.3 Simulation: the base code GLAM

The GLAM code is presented in Klypin & Prada (2018), and is a promising tool to quickly generate N-body simulations for ΛCDM models with reasonable speed and acceptable resolution, which are suitable for the massive production of galaxy survey mocks.

As a PM code, GLAM solves the Poisson equation for the gravitational potential in a periodic cube using FFT. The code uses a 3D mesh for density and potential estimates, and only one mesh is needed for the calculation: the density mesh is replaced with the potential. The gravity solver uses FFT to solve the discrete analogue of the Poisson equation, by applying it first in x- and then to y-direction, and finally transposing the matrix to improve data locality before applying FFT in the third (z-)direction. After multiplying this data matrix by the Green's function, an inverse FFT is applied, performing one matrix transposition and three FFTs, to compute the Newtonian

potential field on the mesh. The potential is then differentiated using a standard three-point finite difference scheme to obtain the x, y and z force components at the centres of the mesh cells. These force components are next interpolated to the locations of simulation particles, which are displaced using a leapfrog scheme. A standard cloud-in-cell (CIC) interpolation scheme is used for both the assignment of particles to calculate the density values in the mesh cells and the interpolation of the forces.

A combination of parameters that define the resolution and speed of the GLAM code are carefully selected. For example, it uses the FFT5 code (the FORTRAN 90 version of FFTPACK5.1) because it enables the option of real-to-real FFT to save memory. It typically uses 30%-50% of the number of particles (in 1D) as compared with the mesh size — given that the code is limited by available memory, this is a better combination than using the same number of particles and mesh points.

GLAM uses OPENMP directives to parallelise the solver. Overall, the code scales nearly perfectly, as has been demonstrated by tests run with different mesh sizes and on different processors (later in the paper we will present some actual scaling test of MG-GLAM as well, which again is nearly perfect). MPI parallelisation is used only to run many realisations on different supercomputer nodes with very little inter-node communications. Load balance is excellent since theoretically every realisation requires the same number of CPUs.

Initial conditions are generated on the spot by GLAM, using the standard Zel'dovich approximation (Zel'Dovich, 1970; Efstathiou et al., 1985) from a user-provided linear matter power spectrum $P(k)$ at $z = 0$. The code back-scales this $P(k)$ to the initial redshift $z_{\rm init}$ using the scale-independent linear growth factor for ΛCDM with the specified cosmological parameters. As the Zel'dovich approximation is less accurate at low redshifts (Crocce et al., 2006a), the simulation is started at an initial redshift $z_{\rm init} \geq 100$. Starting at a higher redshift such as $z_{\rm init} = 100$ also has the additional advantage that, for the MG models of interest here, the effect of the scalar field is smaller at earlier times, which means that it is an increasingly better approximation to use the same initial conditions in the MG models as in the ΛCDM model with the same cosmological parameters, as we practice throughout this work. If, as in the general scenarios, there is non-negligible MG effect prior to $z_{\rm init}$, such effect should be taken into account in the generation of initial conditions, e.g., Li & Barrow (2011b). We note that using ΛCDM initial conditions in the MG simulations means that we do not need to back-scale the linear $P(k)$ (e.g., at $z = 0$) of the corresponding MG models, which are usually scale-dependent — this latter approach has been checked for clustering

dark energy models in Hassani et al. (2019), where it is found to be unable to give the correct matter and gravitational potential power spectra at late times simultaneously (see Hassani et al. (2020) for a way to overcome this issue).

GLAM uses a fixed number of time steps, but this number is user-specified. The standard choice is about 150-200. Here, we have compared the model difference of the matter power spectra between modified gravity MG-GLAM and ΛCDM GLAM simulations and found that the result is converged with 160 time steps. Doubling the number of steps from 160 to 320 makes negligible difference.

The code generates the density field, including peculiar velocities, for a particular cosmological model. Non-linear matter power spectra and halo catalogues at user-specified output redshifts (snapshots) are measured on the fly. For the latter, GLAM employs the bound density maximum (BDM) algorithm (Klypin & Holtzman, 1997; Riebe et al., 2011) to get around the usual limitations placed on the completeness of low-mass haloes by the lack of force resolution in PM simulations. Here we briefly describe the idea behind the BDM halo finder, and further details can be found in Riebe et al. (2011); Knebe et al. (2011). The code starts by calculating a local density at the positions of individual particles, using a spherical top-hat filter containing a constant number $N_{\rm filter}$ (typically 20) of particles. It then gathers all the density maxima and, for each maximum, finds a sphere that contains a mass $M_\Delta = \frac{4}{3}\pi\Delta\rho_{\rm crit}(z)R_\Delta^3$, where $\rho_{\rm crit}(z)$ is the critical density at the halo redshift z, and Δ is the overdensity within the halo radius R_Δ. Throughout this work we will use the virial density definition for Δ given by (Bryan & Norman, 1998a)

$$\Delta_{\rm vir}(z) = 18\pi^2 + 82\left[\Omega_{\rm m}(z) - 1\right] - 39\left[\Omega_{\rm m}(z) - 1\right]^2, \tag{2.52}$$

where $\Omega_{\rm m}(z)$ is the matter density parameter at z. To find distinct haloes, the BDM halo finder still needs to deal with overlapping spheres. To this end, it treats the density maxima as halo centres and finds the one sphere, amongst a group of overlapping ones, with the deepest Newtonian potential. This is treated as a distinct, central, halo. The radii and masses of the haloes which correspond to the other (overlapping) spheres are then found by a procedure that guarantees a smooth transition of the properties of small haloes when they fall into the larger halo to become sub-haloes of the latter. The latter is done by defining the radius of the infalling halo as $\max(R_1, R_2)$, where R_1 is its distance to the surface of the larger, soon-to-be host, central halo, and R_2 is its distance to the nearest density maximum in the spherical shell $[\min(R_\Delta, R_1), \max(R_\Delta, R_1)]$ centred around it (if no density maximum exists in this shell, $R_2 = R_\Delta$). The BDM halo finder was compared against a range of other halo finders in Knebe et al. (2011), where good agreement was found.

MG-GLAM extends GLAM to a general class of modified gravity theories by adding extra modules for solving MG scalar field equations, which will be introduced in the following sections.

2.3.1 The GLAM code units

Like most other N-body codes, GLAM uses its own internal unit system. The code units are designed such that the physical equations can be cast in dimensionless form, which is more convenient for numerical solutions.

Let the box size of simulations be L and the number of grid points in one dimension be $N_{\rm g}$. We can introduce dimensionless coordinates $\tilde{\boldsymbol{x}}$, momenta $\tilde{\boldsymbol{p}}$ and potentials $\tilde{\Phi}$ using the following relations (Klypin & Prada, 2018)

$$\tilde{\boldsymbol{x}} = \left(\frac{N_{\rm g}}{L}\right)\boldsymbol{x}\,, \qquad \tilde{\boldsymbol{p}} = \left(\frac{N_{\rm g}}{H_0 L}\right)\boldsymbol{p}\,, \qquad \tilde{\Phi} = \left(\frac{N_{\rm g}}{H_0 L}\right)^2 \Phi. \tag{2.53}$$

Having the dimensionless momenta, we can find the peculiar velocity,

$$\boldsymbol{v}_{\rm pec} = 100\left[\frac{L/(h^{-1}\,{\rm Mpc})}{N_{\rm g}}\right]\left(\frac{\tilde{\boldsymbol{p}}}{a}\right)\ {\rm km\ s^{-1}}\,, \tag{2.54}$$

Using these notations, we write the particle equations of motion and the Poisson equation as

$$\frac{{\rm d}\tilde{\boldsymbol{p}}}{{\rm d}a} = -\left(\frac{H_0}{\dot{a}}\right)\tilde{\boldsymbol{\nabla}}\tilde{\Phi}, \tag{2.55}$$

$$\frac{{\rm d}\tilde{\boldsymbol{x}}}{{\rm d}a} = -\left(\frac{H_0}{\dot{a}}\right)\frac{\tilde{\boldsymbol{p}}}{a^2}, \tag{2.56}$$

$$\tilde{\nabla}^2\tilde{\Phi} = \frac{3}{2}\Omega_{\rm m0}a^{-1}\tilde{\delta}, \tag{2.57}$$

where $\tilde{\delta}$ is the code unit expression of the density contrast δ.

From Equations (2.53) we can derive the following units,

$$\tilde{\boldsymbol{\nabla}} = \left(\frac{L}{N_{\rm g}}\right)\boldsymbol{\nabla}\,, \quad {\rm d}\tilde{t} = H_0{\rm d}t\,, \quad \tilde{\rho}_{\rm m} = \left(\frac{a^3}{\rho_{\rm crit,0}\Omega_{\rm m0}}\right)\rho_{\rm m}\,, \quad \tilde{\delta} = \delta\,. \tag{2.58}$$

In what follows, we will also use the following definition

$$\tilde{c} = \left(\frac{N_{\rm g}}{H_0 L}\right)c \tag{2.59}$$

for the code-unit expression of the speed of light, c.

GLAM uses a regularly spaced three-dimensional mesh of size $N_{\rm g}^3$ that covers the cubic domain L^3 of a simulation box. The size of a cell, $\Delta x = L/N_{\rm g}$, and the mass of each particle, $m_{\rm p}$, define the force and mass resolution respectively:

$$m_{\rm p} = \Omega_{\rm m0}\,\rho_{\rm crit,0}\left(\frac{L}{N_{\rm p}}\right)^3 = 8.517\times10^{10}\left(\frac{\Omega_{\rm m0}}{0.30}\right)\left[\frac{L/(h^{-1}{\rm Gpc})}{N_{\rm p}/1000}\right]^3 h^{-1}M_\odot, \tag{2.60}$$

$$\Delta x = \left[\frac{L/(h^{-1}{\rm Gpc})}{N_{\rm g}/1000}\right]h^{-1}{\rm Mpc}, \tag{2.61}$$

where $N_{\rm p}^3$ is the number of particles and $\rho_{\rm crit,0}$ is the critical density of the Universe at present.

2.4 Simulation: solvers for the extra degrees of freedom

This section describes in detail how the different theoretical models of Section 2.2 can be incorporated in a numerical simulation code for the standard ΛCDM cosmology, so that the scalar degree of freedom can be solved at any given time with any given matter density field. In this way, the various effects of the scalar field on cosmic structure formation can be accurately predicted and implemented.

We have seen in Section 2.2 that in modified gravity models we usually need to solve a new, dynamical, degree of freedom, which is governed by a non-linear and elliptical type partial differential equation (PDE). Being a nonlinear PDE, unlike the linear Poisson equation in the standard gravity solved in base code GLAM, the equation can not be solved by a one-step FFT∗, but requires a *multigrid relaxation* scheme to obtain a solution.

For completeness, we will first give a concise summary of the relaxation method and its multigrid implementation (Section 2.4.1). Next, we will specify the practical side, discussing how to efficiently arrange the memory in the computer, to allow the same memory space to be used for different quantities at different stages of the calculation, therefore minimising the overall memory requirement (Section 2.4.2), and also saving the time for frequently array allocating and de-allocating operations. After that, in Section 2.4.3–2.4.5, we will respectively discuss how the nonlinear PDEs in general coupled quintessence, symmetron and $f(R)$ models can be solved most efficiently. Much effort will be devoted to replacing the common Newton-Gauss-Seidel relaxation method by a nonlinear Gauss-Seidel, which has been found to lead to substantial speedup of simulations (Bose et al., 2017) (but we will generalise this to more models than focused on in Bose et al. (2017)). For the coupled quintessence model, we will also briefly describe how the background evolution of the scalar field is numerically solved as an integral part of MG-GLAM, to further increase its flexibility.

∗This does not mean that FFT cannot be used under any circumstances. For example, Chan & Scoccimarro (2009) used a FFT-relaxation method to solve nonlinear PDEs iteratively. In each iteration, the equation is treated as if it were linear (by treating the nonlinear terms as a 'source') and solved using FFT, but the solution in the previous step is used to update the 'source', for the PDE to be solved again to get a more accurate solution, until some convergence is reached.

2.4.1 Multigrid Gauss-Seidel relaxation

Let the partial differential equation to be solved take the following form:

$$\mathcal{L}(u) = 0, \tag{2.62}$$

where u is the scalar field and $\mathcal{L}$ is the PDE operator. To solve this equation numerically, we use finite difference to get a discrete version of it on a mesh*. Since MG-GLAM is a PM code, it has a uniform mesh resolution and does not use adaptive mesh refinement (AMR). When discretised on a uniform mesh with cell size h, the above equation can be denoted as

$$\mathcal{L}^h(u^h) = f^h, \tag{2.63}$$

where we have added a non-zero right-hand side, f^h, for generality (while $f^h = 0$ on the mesh with cell size h, later when we discretise it on coarser meshes needed for the multigrid implementation, f is no longer necessarily zero). Both u^h and f^h are evaluated at the cell centres of the given mesh.

The solution we obtain numerically, $\hat{u}$, is unlikely to be the true solution u^h to the discrete equation, and applying the PDE operator on the former gives the following, slightly different, equation:

$$\mathcal{L}^h(\hat{u}^h) = \hat{f}^h. \tag{2.64}$$

Taking the difference between the above two equations, we get

$$\mathcal{L}^h(u^h) - \mathcal{L}^h(\hat{u}^h) = f^h - \hat{f}^h = -d^h, \tag{2.65}$$

where $d^h \equiv \hat{f}^h - f^h$ is the *local residual*, which characterises the inaccuracy of the solution $\hat{u}^h$ (this is because if $\hat{u}^h = u^h$, we would expect $\hat{f}^h = f^h$ and hence there is zero 'inaccuracy'). d^h is also evaluated at cell centres. Later, to check if a given set of numerical solution $\hat{u}^h$ is acceptable, we will use a *global residual*, ϵ^h, which is a single number for the given mesh of cell size h. In this work we choose to define ϵ^h as the root-mean-squared of d^h in all mesh cells (although this is by no means the only possible definition). We will call both d^h and ϵ^h 'residual' as the context will make it clear which one is referred to.

Relaxation solves Equation (2.63) by starting from some approximate trial solution to u^h, $\hat{u}^h_{\rm old}$, and check if it satisfies the PDE. If not, this trial solution can be updated using a method that is similar to the Newton-Ralphson iterative method to solve nonlinear algebraic equations

$$\hat{u}^h_{\rm new} = \hat{u}^h_{\rm old} - \frac{\mathcal{L}^h\left(\hat{u}^h_{\rm old}\right) - \hat{f}^h}{\partial\mathcal{L}^h\left(\hat{u}^h_{\rm old}\right)/\partial\hat{u}^h}. \tag{2.66}$$

*In this paper we consider the simplest case of cubic cells.

This process can be repeated iteratively, until the updated solution satisfies the PDE to an acceptable level, i.e., ϵ^h becomes small enough. In practice, because we are solving the PDE on a mesh, Equation (2.66) should be performed for all mesh cells, which raises the question of how to order this operation for the many cells. We will adopt the Gauss-Seidel 'black-red chessboard' approach, where the cells are split into two classes, 'black' and 'red', such that all the six direct neighbours* of a 'red' cell are black and vice versa. The relaxation operation, Equation (2.66), is performed in two sweeps, the first for 'black' cells (i.e., only updating $\hat{u}^h$ in 'black' cells while keeping their values in 'red' cells untouched), while the second for all the 'red' cells. This is a standard method to solve nonlinear elliptical PDEs by using relaxation, known as the *Newton-Gauss-Seidel method*. However, although this method is generic, it is not always efficient, and later we will describe a less generic alternative which is nevertheless more efficient.

Relaxation iterations are useful at reducing the Fourier modes of the error in the trial solution $\hat{u}^h$, whose wavelengths are comparable to that of the size of the mesh cell h. If we do relaxation on a fine mesh, this means that the short-wave modes of the error are quickly reduced, but the long-wave modes are generally much slower to decrease, which can lead to a slow convergence of the relaxation iterations. A useful approach to solve this problem is by using *multigrid*: after a few iterations on the fine level, we 'move' the equation to a coarser level where the cell size is larger and the longer-wave modes of the error in $\hat{u}^h$ can be more quickly decreased. The discretised PDE on the coarser level is given by

$$\mathcal{L}^H(u^H) = \mathcal{L}\left(\mathcal{R}\hat{u}^h\right) - \mathcal{R}d^h \equiv S^H, \tag{2.67}$$

where the superscript ^{H} denotes the coarse level where the cell size is H (in our case $H = 2h$), and $\mathcal{R}$ denotes the *restriction* operator which interpolates quantities from the fine level to the coarse level. In our numerical implementation, a coarse (cubic) cell contains 8 fine (cubic) cells of equal volume, and the restriction operation can be conveniently taken as the arithmetic average of the values of the quantity to be interpolated in the 8 fine cells.

Equation (2.67) can be solved using relaxation similarly to Equation (2.64), for which the numerical solution is denoted as $\hat{u}^H$. This can be used to 'correct' and 'improve' the approximate solution $\hat{u}^h$ on the fine level, as

$$\hat{u}^{h,\text{new}} = \hat{u}^{h,\text{old}} + \mathcal{P}\left(\hat{u}^H - \mathcal{R}\hat{u}^h\right), \tag{2.68}$$

*The direct neighbours of a given cell are the six neighbouring cells which share a common face with that cell.

where $\mathcal{P}$ is the *prolongation* operation which does the interpolation from the coarse to the fine levels. In this work we shall use the following definition of the prolongation operation: for a given fine cell,

1. find its parent cell, i.e., the coarser cell that contains the fine cell;

2. find the seven neighbours of the parent cell, i.e., the coarser cells which share a face (there are 3 of these), an edge (there are 3 of these) or a vertex (just 1) with the above parent coarser cell;

3. calculate the fine-cell value of the quantity to be interpolated from the coarse to the fine levels, as a weighted average of the corresponding values in the 8 coarse cells mentioned above: $27/64$ for the parent coarse cell, and $9/64$, $3/64$ and $1/64$ respectively for the coarse cells sharing a face, an edge and a vertex with the parent cell.

The above is a simple illustration of how multigrid works for two levels of mesh resolution, h and H. In principle, multigrid can be and is usually implemented using more than two levels. In this paper we will use a hierarchy of increasingly coarser meshes with the coarsest one having 4^3 cells.

There are flexibilities in how to arrange the relaxations at different levels. The most-commonly used arrangement is the so-called V-cycle, where one starts from the finest level, moves to the coarsest one performing relaxation iterations on each of the intermediate levels (cf. Equation (2.67)), and then moves straight back to the finest performing corrections using Equation (2.68) on each of the intermediate levels. Other arrangements, such as F-cycle and W-cycle (cf. Fig. 2.1), are sometimes more efficient in improving the convergence rate of $\hat{u}^h$ to u^u, and we have implemented them in MG-GLAM as well.

2.4.2 Memory usage

GLAM uses a single array to store mesh quantities, such as the matter density field and the Newtonian potential, because at any given time only one of these is needed. The Newtonian force at cell centres is calculated by finite-differencing the potential and then interpolated to the particle positions. To be memory efficient, GLAM also opts not to create a separate array to store the forces at the cell centres, but instead directly calculates them at the particle positions immediately before updating the particle velocities.

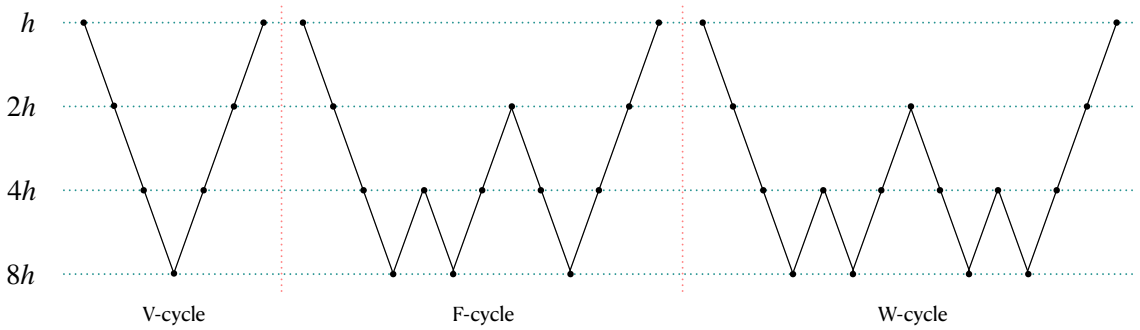

Figure 2.1: An illustration of the three different arrangements of multigrid relaxation method used in this paper: from left to right, V-cycle, F-cycle and W-cycle. The horizontal dotted lines depict 4 multigrid levels of mesh, with the finest mesh (denoted by its cell size h) on top, and the coarsest mesh (with cell size $8h$) at the bottom. The relaxation always starts on the finest level, and the solid lines show how the multigrid solver walks through the different levels, performing Gauss-Seidel relaxation iterations at each level (denoted by the circles), called smoothing. Only one single full cycle is shown for each case. The solver walks over the multigrid levels more times in W-cycle than in F-cycle and V-cycle, and thus it requires fewer cycles in the former case to arrive at a converged solution. However, it is also computationally more expensive.

With the new scalar field to be solved in modified gravity models, we need two additional arrays of size $N_{\rm g}^3$, where $N_{\rm g}^3$ is the number of cells of the PM grid (i.e., there are $N_{\rm g}$ cells in each direction of the cubic simulation box). This leads to three arrays. `Array 1` is the default array in GLAM, which is used to store the density field ρ and the Newtonian potential Φ (at different stages of the simulation). Note that the density field is also needed when solving the scalar field equation of motion during the relaxation iterations, and so we cannot use this array to also store the scalar field. On the other hand, we will solve the Newtonian potential after the scalar field, by when it is safe to overwrite this array with Φ. `Array2` is exclusively used to store the scalar field solution $\hat{u}^h$ on the PM grid, which will be used to calculate the fifth force. `Array3` is used to store the various intermediate quantities which are created for the implementation of the multigrid relaxation, such as d^h, $\hat{u}^H$, $\mathcal{R}\hat{u}^h$, $\mathcal{R}d^h$, S^H and ρ^H, the last of which is the density field on the coarser level ^{H}, which appears in the coarse-level discrete PDE operator $\mathcal{L}^H$.

To be concrete, we imagine the 3D array (`Array3`) as a cubic box with $N_{\rm g}^3$ cubic cells of equal size. An array element, denoted by (i, j, k), represents the i-th cell in the x direction, j-th cell in the y direction and k-th cell in the z direction, with $i, j, k = 1, \cdots, N_{\rm g}$. We divide this array into 8 sections, each of which can be considered to correspond to one of the 8 octants that equally divide the volume of the cubic box. The range of (i, j, k) of each section and the quantity stored in that section of `Array3` are summarised in Table. 2.1.

Let us explain this more explicitly. First of all, the whole `Array3`, of size $N_{\rm g}^3$, will be used to store the residual value d^h on the PM grid (which has $N_{\rm g}^3$ cells). From now on, we label this grid

Table 2.1: The index ranges and the stored quantities of each section of `Array3` in MG-GLAM.

Section	i range	j range	k range	Quantity
1	$1, \cdots, N_\mathrm{g}/2$	$1, \cdots, N_\mathrm{g}/2$	$1, \cdots, N_\mathrm{g}/2$	d^ℓ, $\mathcal{R}d^\ell$
2	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$1, \cdots, N_\mathrm{g}/2$	$1, \cdots, N_\mathrm{g}/2$	d^ℓ, $\rho^{\ell-1} = \mathcal{R}\rho^\ell$
3	$1, \cdots, N_\mathrm{g}/2$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$1, \cdots, N_\mathrm{g}/2$	d^ℓ, $\mathcal{R}\hat{u}^\ell$
4	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$1, \cdots, N_\mathrm{g}/2$	d^ℓ, $\hat{u}^{\ell-1}$
5	$1, \cdots, N_\mathrm{g}/2$	$1, \cdots, N_\mathrm{g}/2$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	d^ℓ, recursion
6	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$1, \cdots, N_\mathrm{g}/2$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	d^ℓ, $d^{\ell-1}$
7	$1, \cdots, N_\mathrm{g}/2$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	d^ℓ, $S^{\ell-1}$
8	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	$N_\mathrm{g}/2+1, \cdots, N_\mathrm{g}$	d^ℓ

by 'level-ℓ', and use 'level-$(\ell-m)$' to denote the grid that are m times coarser, i.e., if the cell size of the PM grid is h, then the cells in this coarse grid have a size of $2^m h$. In the table above we have used d^ℓ to denote the d^h on level-ℓ, and so on. Note that we always use $N_\mathrm{g} = 2^\ell$.

The local residual d^h on a fine grid is only needed for two purposes: (1) to calculate the global residual on that grid, ϵ^h, which is needed to decide convergence of the relaxation, and (2) to calculate the coarse-level PDE operator $\mathcal{L}^H$ that is needed for the multigrid acceleration, as per Equation (2.67). This suggests that d^h does not have to occupy `Array3` all the time, and so this array can be reused to store other intermediate quantities (see the last column of the above table) *after* we have obtained ϵ^h.

In our arrangement, Section 1 stores the residual $\mathcal{R}d^\ell$, Section 2 stores the restricted density field $\rho^{\ell-1} = \mathcal{R}\rho^\ell$, Sections 3 and 4 store, respectively, the restricted scalar field solution $\mathcal{R}\hat{u}^\ell$ and the coarse-grid scalar field solution $\hat{u}^{\ell-1}$ — the former is needed to calculate $S^{\ell-1}$ in Equation (2.67) and to correct the fine-grid solution using Equation (2.68), which is fixed after calculation, while the latter is updated during the coarse-grid relaxation sweeps$*$. Section 7 stores the coarse-grid source $S^{\ell-1}$ for the PDE operator $\mathcal{L}^{\ell-1}$ as defined in Equation (2.67), and finally Section 6 stores the residual on the coarse level, $d^{\ell-1}$. Note that all these quantities are for level-$(\ell-1)$, so that they can be stored in section of `Array3` of size $(N_\mathrm{g}/2)^3$. Section 8 is not used to store anything other than d^ℓ.

We have not touched Section 5 so far — this section is reserved to store the same quantities as above, but for level-$(\ell-2)$, which are needed if we want to use more than two levels of multigrid. It is further divided into 8 sections, each of which will play the same roles as detailed in the table

$*$We use $\mathcal{R}\hat{u}^\ell$ as the initial guess for $\hat{u}^{\ell-1}$ for the Gauss-Seidel relaxations on the coarse level.

above∗. In particular, the (sub)Section 5 of Section 5 is reserved for quantities on level-$(\ell-3)$, and so on. In this way, there is no need to create separate arrays of various sizes to store the intermediate quantities on different multigrid levels which therefore saves memory.

There is a small tricky issue here: as we mentioned above, the local residual d^ℓ on the PM grid is needed to calculate the coarse-grid source $S^{\ell-1}$ using Equation (2.67), thus we will be using the quantity d^ℓ stored in `Array3` to calculate $\mathcal{R}d^\ell$ and then write it to (part of) the same array, running the risk of overwriting some of the data while it is still needed. To avoid this problem, we refrain from using the d^ℓ data already stored in `Array3`, but instead recalculate it in the subroutine to calculate $\mathcal{R}d^\ell$ (this only needs to be done for level-ℓ). With a bit of extra computation, this enables use to avoid creating another array of similar size to `Array3`.

Since `Array3` stores different quantities in different parts, care must be excised when assessing these data. There is a simple rule for this: suppose that we need to read or write the quantities on the coarse grid of level-$(\ell-m)$ with $m\geq 1$. These are 3-dimensional quantities with the three directions labelled by I,J,K, which run over $1,\cdots,2^{\ell-m}$, and we have

$$
\begin{aligned}
\mathcal{R}(d^{\ell-m+1})\,[I,J,K] &\leftrightarrow \texttt{Array3}[i=I, && j=J && ,k=K+(2^m-2)\cdot 2^{\ell-m}],\\
\mathcal{R}(\rho^{\ell-m+1})\,[I,J,K] &\leftrightarrow \texttt{Array3}[i=I+2^{\ell-m}, && j=J && ,k=K+(2^m-2)\cdot 2^{\ell-m}],\\
\mathcal{R}(u^{\ell-m+1})\,[I,J,K] &\leftrightarrow \texttt{Array3}[i=I, && j=J+2^{\ell-m} && ,k=K+(2^m-2)\cdot 2^{\ell-m}],\\
\hat{u}^{\ell-m}\,[I,J,K] &\leftrightarrow \texttt{Array3}[i=I+2^{\ell-m}, && j=J+2^{\ell-m} && ,k=K+(2^m-2)\cdot 2^{\ell-m}],\\
d^{\ell-m}\,[I,J,K] &\leftrightarrow \texttt{Array3}[i=I+2^{\ell-m}, && j=J && ,k=K+(2^m-1)\cdot 2^{\ell-m}],\\
S^{\ell-m}[I,J,K] &\leftrightarrow \texttt{Array3}[i=I, && j=J+2^{\ell-m} && ,k=K+(2^m-1)\cdot 2^{\ell-m}],
\end{aligned}
$$

where $i,j,k=1,\cdots,N_{\rm g}$ run over the entire `Array3`.

We can estimate the required memory for MG-GLAM simulations as follows. As mentioned above, the code uses a 3D array of single precision to store both the density field and the Newtonian potential, and one set of arrays for particle positions and velocities. In addition, two arrays are added to store the scalar field solution (`Array2`) and various intermediate quantities in the multigrid relaxation solver (`Array3`). In the cosmological simulations described in this paper, we have used double precision for the two new arrays, and we have checked that using single precision slightly speeds up the simulation, while agreeing with the double-precision results within 0.001% and 0.5% respectively for the matter power spectrum and halo mass function. Given its fast speed and its shared-memory nature, memory is expected to be the main limiting factor for large MG-GLAM

∗The exception is that, as $d^{\ell-1}$ is already stored in Section 6, it does not have to be stored in Section 5 again.

jobs. For this reason, we assume that all arrays are set to be single precision for future runs, and this leads to the following estimate of the total required memory:

$$\begin{aligned} M_{\rm total} &= \left(12N_{\rm g}^3 + 24N_{\rm p}^3\right)\,\text{bytes}\,, \\ &= \left[89.41\left(\frac{N_{\rm g}}{2000}\right)^3 + 22.35\left(\frac{N_{\rm p}}{1000}\right)^3\right]\,\text{GB}\,, \\ &= 112\left(\frac{N_{\rm p}}{1000}\right)^3\,\text{GB}\,, \quad \text{for } N_{\rm g} = 2N_{\rm p}\,, \end{aligned} \tag{2.69}$$

where we have used $1\,\text{GB} = 1024^3\,\text{bytes}$. This is slightly more than twice the memory requirement of the default GLAM code, which is $52\,(N_{\rm p}/1000)^3$ GB (Klypin & Prada, 2018).

2.4.3 Implementation of coupled quintessence

Defining the code unit of the dimensionless scalar field perturbation, $\delta\varphi = \varphi - \bar{\varphi}$, as*

$$\tilde{\varphi} \equiv \frac{c^2 N_{\rm g}}{H_0^2 L^2}\delta\varphi = \tilde{c}^2\delta\varphi, \tag{2.70}$$

with $\delta\varphi$ being the perturbation to φ, we can rewrite its equation of motion as

$$\tilde{\boldsymbol{\nabla}}^2\tilde{\varphi} = 3\beta\Omega_{\rm m}a^{-1}e^{\beta\bar{\varphi}}\left[\exp\left(\beta\frac{\tilde{\varphi}}{\tilde{c}^2}\right)(1+\tilde{\delta}) - 1\right] - \alpha\lambda^2 a^2\left[\frac{1}{(\bar{\varphi}+\tilde{c}^{-2}\tilde{\varphi})^{1+\alpha}} - \frac{1}{\bar{\varphi}^{1+\alpha}}\right], \tag{2.71}$$

where $\bar{\varphi}$ is the background value of φ, and λ is defined in Equation (2.18). The Poisson equation becomes

$$\tilde{\boldsymbol{\nabla}}^2\tilde{\Phi}_{\rm N} = \frac{3}{2}\Omega_{\rm m}a^{-1}e^{\beta\bar{\varphi}}\left[\exp\left(\beta\frac{\tilde{\varphi}}{\tilde{c}^2}\right)(1+\tilde{\delta}) - 1\right] + \lambda^2 a^2\left[\frac{1}{(\bar{\varphi}+\tilde{c}^{-2}\tilde{\varphi})^{\alpha}} - \frac{1}{\bar{\varphi}^{\alpha}}\right]. \tag{2.72}$$

In practice, as we know that the scalar field density perturbation is small in the models of interest, the second term on the right-hand side of the Poisson equation can be dropped approximately. We have also chosen to neglect the term $\exp\left(\beta\tilde{c}^{-2}\tilde{\varphi}\right)$ in front of $(1+\tilde{\delta})$, to simplify the simulation — this is again justified because $|\delta\varphi| \ll |\bar{\varphi}| \simeq \mathcal{O}(1)$ at late times, although including this in the simulation is trivial.

The modified particle coordinate and velocity updates can be rewritten as

$$\frac{{\rm d}\tilde{\boldsymbol{x}}}{{\rm d}a} = \frac{H_0}{a^2\dot{a}}\tilde{\boldsymbol{p}}, \tag{2.73}$$

$$\frac{{\rm d}\tilde{\boldsymbol{p}}}{{\rm d}a} = -\frac{H_0}{\dot{a}}\left[\tilde{\boldsymbol{\nabla}}\tilde{\Phi}_{\rm N} + \beta\tilde{\boldsymbol{\nabla}}\tilde{\varphi}\right] - \beta\frac{{\rm d}\bar{\varphi}}{{\rm d}a}\tilde{\boldsymbol{p}}. \tag{2.74}$$

Here we can observe more explicitly the effect of a modified background expansion history in coupled quintessence models, encoded in the $H_0\dot{a}^{-1}$ terms.

*Note that, for brevity, we have slightly abused the notations, by using the same symbol φ with a tilde for the code-unit expression of $\delta\varphi$. Given that the code-unit quantity always comes with a tilde, this should not cause any confusion with, e.g., the background scalar field $\bar{\varphi}$, or the total dimensionless scalar field φ in physical units.

In MG-GLAM, Equation (2.71) is solved using the Newton-Gauss-Seidel method described in Section 2.4.1. Equation (2.72) is not directly solved, but instead we solve the (standard) Poisson equation not having $e^{\beta\bar{\varphi}}$: since this is a background quantity, we instead multiply it when calculating the Newtonian force from $\tilde{\Phi}_{\rm N}$. Equations (2.73) and (2.74) are then solved — the fifth force $\beta\tilde{\boldsymbol{\nabla}}\tilde{\varphi}$ is incorporated by first summing up the two potentials, $\tilde{\Phi}_{\rm N}+\beta\tilde{\varphi}$, and then doing the finite difference.

MG-GLAM background cosmology solver

As Equations (2.73) and (2.74) contain background quantities $\dot{a}$ and $\mathrm{d}\tilde{\varphi}/\mathrm{d}a$, for every given coupled quintessence model we need to solve its background evolution. This is governed by the following system of equations — the equation of motion for the background scalar field $\bar{\varphi}$:

$$\ddot{\bar{\varphi}}+3\frac{\dot{a}}{a}\dot{\bar{\varphi}}+\frac{\mathrm{d}V(\bar{\varphi})}{\mathrm{d}\varphi}+\frac{\mathrm{d}A(\bar{\varphi})}{\mathrm{d}\varphi}8\pi G\bar{\rho}_{\rm m}=0, \tag{2.75}$$

the Friedmann equation (with a flat Universe, $k=0$, being assumed)

$$H^2=\left(\frac{\dot{a}}{a}\right)^2=\frac{8\pi G}{3}\left[\bar{\rho}_{\rm r}(a)+A(\bar{\varphi})\bar{\rho}_{\rm m}(a)\right]+\frac{1}{6}\dot{\bar{\varphi}}^2+H_0^2\frac{\lambda^2}{\bar{\varphi}^\alpha}, \tag{2.76}$$

and the Raychaudhuri equation

$$3\left(\dot{H}+H^2\right)=-4\pi G\left[2\bar{\rho}_{\rm r}(a)+A(\bar{\varphi})\bar{\rho}_{\rm m}(a)\right]-\dot{\bar{\varphi}}^2+H_0^2\frac{\lambda^2}{\bar{\varphi}^\alpha}, \tag{2.77}$$

where $\bar{\rho}_{\rm r}$ denotes the background density of radiations (we assume that all three species of neutrinos are massless and thus counted as radiation). Note that in Equations (2.76) and (2.77), to see the dimensions of the different terms clearly, we have already explicitly substituted the inverse-powerlaw potential and used the definition of λ in Equation (2.18). In MG-GLAM the scalar field equation is solved by a fifth-sixth order continuous Runge-Kutta method*.

For numerical solutions in background cosmology, instead of directly working with Equations (2.75), (2.76) and (2.77)), it is convenient to use $N\equiv\ln(a)$ as the time variable, for which we have

$$\bar{\varphi}'=\mathcal{H}\frac{\mathrm{d}\bar{\varphi}}{\mathrm{d}N},\quad \bar{\varphi}''=\mathcal{H}^2\frac{\mathrm{d}^2\bar{\varphi}}{\mathrm{d}N^2}+\mathcal{H}'\frac{\mathrm{d}\bar{\varphi}}{\mathrm{d}N}, \tag{2.78}$$

where $'$ is the derivative with respect to the conformal time τ and $\mathcal{H}\equiv a'/a$. In this convention, the background quintessence field equation of motion, Equation (2.75), can be written as

$$\frac{\mathcal{H}^2}{H_0^2}\frac{\mathrm{d}^2\bar{\varphi}}{\mathrm{d}N^2}+\left[2\frac{\mathcal{H}^2}{H_0^2}+\frac{\mathcal{H}'}{H_0^2}\right]\frac{\mathrm{d}\bar{\varphi}}{\mathrm{d}N}-3\alpha\lambda\mathrm{e}^{2N}\bar{\varphi}^{-(1+\alpha)}+3\beta\mathrm{e}^{-N}\Omega_{\rm m}\exp(\beta\bar{\varphi})=0, \tag{2.79}$$

*For this numerical integrator we have used `subroutine dverk` from the CAMB code, originally developed in FORTRAN 66 by K. R. Jackson.

where the quantities $\mathcal{H}^2/H_0^2$ and $\mathcal{H}'/H_0^2$ can be obtained from Equations (2.76, 2.77) as

$$\frac{\mathcal{H}^2}{H_0^2} = \left[1 - \frac{1}{6}\left(\frac{\mathrm{d}\bar{\varphi}}{\mathrm{d}N}\right)^2\right]^{-1} \left[\Omega_\mathrm{r}\mathrm{e}^{-2N} + \exp(\beta\bar{\varphi})\Omega_\mathrm{m}\mathrm{e}^{-N} + \lambda\mathrm{e}^{2N}\bar{\varphi}^{-\alpha}\right], \tag{2.80}$$

$$\frac{\mathcal{H}'}{H_0^2} = -\frac{1}{3}\left(\frac{\mathrm{d}\bar{\varphi}}{\mathrm{d}N}\right)^2 \frac{\mathcal{H}^2}{H_0^2} + \lambda\mathrm{e}^{2N}\bar{\varphi}^{-\alpha} - \Omega_\mathrm{r}\mathrm{e}^{-2N} - \frac{1}{2}\Omega_\mathrm{m}\mathrm{e}^{-N}\exp(\beta\bar{\varphi}). \tag{2.81}$$

Here Ω_r denotes the present-day radiation density parameter, with 'radiation' including CMB photons with a present-day temperature of $2.7255\ \mathrm{K}$ and 3.046 flavours of massless neutrinos; we defer the implementation of massive neutrinos, both as a non-interacting particle species and in the context of coupling to scalar fields, to future works.

We note that λ is *not* a free parameter of the model. Rather, once the density parameters Ω_m, Ω_r and H_0 are specified (or equally once the present-day densities of matter and radiation are specified), λ, which quantifies the size of the potential energy of the scalar field, must take some certain value in order to ensure consistency. If λ is too large, the predicted $H(a=1)$ will be larger than the desired (input) value of H_0, and vice versa. In practice, MG-GLAM starts from a trial value of $\lambda = 1$, evolves Equations (2.75, 2.76, 2.77) from some initial redshift ($z_\mathrm{init} = 10^5$) to $z = 0$, and checks if the calculated value of $H(a=1)$ is equal to the desired value H_0 (within a small relative error of order $\mathcal{O}\left(10^{-6}\right)$); if the predicted $H(a=1)$ is larger than the desired H_0, λ is decreased, and vice versa. This process is repeated iteratively to obtain a good approximation to λ with a relative error smaller than 10^{-6}. The initial conditions of $\bar{\varphi}$ and $\dot{\bar{\varphi}}$ at $z_\mathrm{init} = 10^5$ are not important, as long as their values are small enough. Once the value of λ has been determined in this way, it is stored to be used in other parts of the code; also stored are a large array of the various background quantities such as $H, \dot{H}, \bar{\varphi}$ and $\dot{\bar{\varphi}}$ — if needed at any time by the scalar field solver of MG-GLAM, these quantities can be linearly interpolated in the scale factor a.

2.4.4 Implementation of symmetrons

The scalar field equation of motion in the symmetron model, Equation (2.35), can be written in code unit as

$$\tilde{c}^2\tilde{\nabla}^2 u = \frac{a^2}{2\xi^2}\left(\tilde{\rho}\frac{a_*^3}{a^3} - 1\right)u + \frac{a^2}{2\xi^2}u^3. \tag{2.82}$$

While this equation can be solved similarly to the case of coupled quintessence by using the standard Newton-Gauss-Seidel relaxation method we described in Section 2.4.3, the 'Newton' approximation of this method, Equation (2.66), is indeed unnecessary, as can be seen from the following derivation. Defining

$$L_{i,j,k}(u) \equiv u_{i+1,j,k} + u_{i-1,j,k} + u_{i,j+1,k} + u_{i,j-1,k} + u_{i,j,k+1} + u_{i,j,k-1}, \tag{2.83}$$

where a subscript $_{i,j,k}$ denotes the value of a quantity in a cell that is the i-th (j-th, k-th) in the x (y, z) direction, the discretised version of Equation (2.82), after some rearrangement, can be written as

$$u_{i,j,k}^3 + \left[\tilde{\rho}_{i,j,k}\frac{a_*^3}{a^3} - 1\right] u_{i,j,k} + \frac{12}{h^2}\frac{\tilde{c}^2\xi^2}{a^2}u_{i,j,k} - \frac{2}{h^2}\frac{\tilde{c}^2\xi^2}{a^2}L_{i,j,k} = 0. \tag{2.84}$$

We can define

$$p \equiv \tilde{\rho}_{i,j,k}\frac{a_*^3}{a^3} - 1 + \frac{12}{h^2}\frac{\tilde{c}^2\xi^2}{a^2}, \tag{2.85}$$

$$q \equiv -\frac{2}{h^2}\frac{\tilde{c}^2\xi^2}{a^2}L_{i,j,k}, \tag{2.86}$$

so that the above equation can be simplified as

$$u_{i,j,k}^3 + pu_{i,j,k} + q = 0. \tag{2.87}$$

This is similar to the discrete equation of motion in the Hu-Sawicki $f(R)$ gravity model with $n = 1$, as discussed in Bose et al. (2017), which can be treated as a cubic equation of $u_{i,j,k}$ that can be solved exactly (analytically). Therefore, given the (approximate) values of the field u in the six direct neighbouring cells of (i, j, k), we can calculate $u_{i,j,k}$ analytically, and there is no need to solve it using the Newton approximation as in Equation (2.66). The relaxation iterations are still needed, since the values of u in the six direct neighbours are *approximate* and therefore need to be updated iteratively, but the replacement of the Newton solver with an exact solution of $u_{i,j,k}$ (therefore the name *nonlinear Gauss-Seidel* as opposed to *Newton Gauss-Seidel*) has been found to significantly improve the convergence speed of the relaxation (Bose et al., 2017). This method for the symmetron model was briefly mentioned in an Appendix of Bose et al. (2017) but no numerical implementation was shown there.

The solution to Equation (2.87) can be found as

$$u_{i,j,k} = \begin{cases} -\frac{1}{3}\left(C + \frac{\Delta_0}{C}\right), & \Delta > 0, \\ \sqrt[3]{-q}, & \Delta = 0, \\ -\frac{2}{3}\sqrt{\Delta_0}\cos\left(\frac{\Theta}{3} + \frac{2\pi}{3}\right), & \Delta < 0, \end{cases} \tag{2.88}$$

where we have defined $\Delta_0 \equiv -3p$, $\Delta_1 \equiv 27q$, $\Delta \equiv \Delta_1^2 - 4\Delta_0^3$ and

$$C \equiv \sqrt[3]{\frac{1}{2}\left(\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}\right)}, \tag{2.89}$$

$$\Theta \equiv \arccos\left(\frac{\Delta_1}{2\sqrt{\Delta_0^3}}\right). \tag{2.90}$$

It can be shown that all the 3 branches of solutions in Equation (2.88) can be the physical solution in certain regimes, depending on model parameters, density values, mesh size, and so on. In our implementation in MG-GLAM, we have used Equation (2.88) instead of Equation (2.66) for the symmetron model.

The acceleration on particles, Equation (2.13), can be written as following in the symmetron model:

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}}{\mathrm{d}a} = \frac{H_0}{a^2\dot{a}}\tilde{\boldsymbol{p}}, \tag{2.91}$$

$$\frac{\mathrm{d}\tilde{\boldsymbol{p}}}{\mathrm{d}a} = \tilde{\boldsymbol{F}}_{\mathrm{N}} + \tilde{\boldsymbol{F}}_5 + \tilde{\boldsymbol{F}}_{\times}, \tag{2.92}$$

where $\tilde{\boldsymbol{F}}_{\mathrm{N}}$, $\tilde{\boldsymbol{F}}_5$ and $\tilde{\boldsymbol{F}}_{\times}$ denote, respectively, the standard Newtonian acceleration, the fifth force acceleration and the frictional force acceleration, in code units, given by

$$\tilde{\boldsymbol{F}}_{\mathrm{N}} = -\frac{H_0}{\dot{a}}\tilde{\boldsymbol{\nabla}}\tilde{\Phi}_{\mathrm{N}}, \tag{2.93}$$

$$\tilde{\boldsymbol{F}}_5 = -6\frac{H_0}{\dot{a}}\xi^2\Omega_{\mathrm{m}}\beta_*^2\tilde{c}^2a_*^{-3}u\tilde{\boldsymbol{\nabla}}u = -3\frac{H_0}{\dot{a}}\xi^2\Omega_{\mathrm{m}}\beta_*^2\tilde{c}^2a_*^{-3}\tilde{\boldsymbol{\nabla}}\left(u^2\right), \tag{2.94}$$

$$\tilde{\boldsymbol{F}}_{\times} = -9\Omega_{\mathrm{m}}\beta_*^2\xi^2\sqrt{1-\left(\frac{a_*}{a}\right)^3}\frac{1}{a^4}u\tilde{\boldsymbol{p}}. \tag{2.95}$$

In practice, as mentioned earlier, the frictional force is much weaker than the other two force components because of the very slow time evolution of the symmetron field. Likewise, any time variation of the matter particle mass due to the coupling with the symmetron field must be tiny and negligible. Therefore, for the Poisson equation, which governs Φ_{N} and thus $\boldsymbol{F}_{\mathrm{N}}$, we simply approximate it to be the same as in ΛCDM.

2.4.5 Implementation of $f(R)$ gravity

In Section 2.2.3 we have introduced a class of $f(R)$ models with an (inverse) power-law function f_R, Equation (2.41), and mentioned that we will focus on the cases of $n = 0,\ 1,\ 2$. In this subsection, we shall first derive equations that apply to general values of n, and then specialise to these three cases, for which we will develop case-specific algorithms of nonlinear Gauss-Seidel relaxation.

In code unit, the f_R equation of motion of this model, Equation (2.45), can be written as

$$\tilde{c}^2\tilde{\boldsymbol{\nabla}}^2\tilde{f}_R = -\Omega_{\mathrm{m}}a^{-1}\left(1+\tilde{\delta}\right) + \frac{1}{3}\tilde{\bar{R}}(a)a^2\left(\frac{\tilde{\bar{f}}_R}{\tilde{f}_R}\right)^{\frac{1}{n+1}} - 4\Omega_\Lambda a^2, \tag{2.96}$$

where $\tilde{f}_R \equiv f_R$ and $\tilde{\bar{f}}_R \equiv \bar{f}_R(a)$ is the background value of f_R. The Newtonian force is still given by Equation (2.93) with $\tilde{\Phi}_{\mathrm{N}}$ governed by Equation (2.57). On the other hand, the fifth force in

code unit can be written as

$$\tilde{\boldsymbol{F}}_5 = \frac{1}{2}\tilde{c}^2\tilde{\boldsymbol{\nabla}}\tilde{f}_R, \tag{2.97}$$

It is more convenient to define the following new, positive-definite, scalar field variable (Bose et al., 2017)

$$u \equiv (-f_R)^{1/(n+1)}, \tag{2.98}$$

where the minus sign is because $f_R < 0$. Equation (2.96) then becomes

$$-\tilde{c}^2\tilde{\boldsymbol{\nabla}}^2\left(u^{n+1}\right) + \frac{\Omega_{\rm m}}{a}\delta + \frac{1}{3}\tilde{\bar{R}}(a)a^2 - \frac{1}{3}\tilde{\bar{R}}(a)a^2\left[-\bar{f}_R(a)\right]^{1/(n+1)}\frac{1}{u} = 0, \tag{2.99}$$

where we have defined the following dimensionless background quantity:

$$\tilde{\bar{R}}(a) \equiv \frac{\bar{R}(a)}{H_0^2} = 3\Big(\Omega_{\rm m}a^{-3} + 4\Omega_\Lambda\Big), \tag{2.100}$$

with $\bar{R}(a)$ being the background value of the Ricci scalar at scale factor a. Equation (2.99) can be further simplified to

$$(u_{i,j,k})^{n+2} + p\,u_{i,j,k} + q = 0, \tag{2.101}$$

where

$$p \equiv \frac{h^2}{6\tilde{c}^2}\left[\frac{\Omega_{\rm m}}{a}\delta_{i,j,k} + \frac{1}{3}\tilde{\bar{R}}(a)a^2\right] - \frac{1}{6}L_{i,j,k}, \tag{2.102}$$

$$q \equiv -\frac{h^2}{6\tilde{c}^2}\frac{1}{3}\tilde{\bar{R}}(a)a^2\left[-\bar{f}_R(a)\right]^{1/(n+1)} \tag{2.103}$$

where $L_{i,j,k}$ was defined in Equation (2.83), and we have neglected the tilde in $\tilde{\bar{f}}_R(a)$ because $\tilde{\bar{f}}_R = \bar{f}_R$ anyway.

Equation (2.101) is a polynomial for $u_{i,j,k}$, which can be analytically solved for the cases of $n = 0$, 1 and 2. The case of $n = 1$ has been discussed in Bose et al. (2017), while cases of $n = 0, 2$ have not been studied before using nonlinear Gauss-Seidel schemes*. Here we discuss all three cases with equal details.

- **The case of $n = 2$**

  In this case, Equation (2.101) is a quadratic equation of $u_{i,j,k}$. Define

$$\Delta_0 \equiv 12q,$$
$$\Delta_1 \equiv 27p^2. \tag{2.104}$$

*The case of $n = 2$ has been studied using simulations based on Newton-Gauss-Seidel relaxation (e.g., Li & Hu, 2011).

We see that $q < 0$ and so $\Delta_0 < 0$ and $\Delta_1 > 0$. Equation (2.101) has 4 branches of analytical solutions:

$$u_{i,j,k} = -S \pm \frac{1}{2}\sqrt{-4S^2 + \frac{p}{S}}, \tag{2.105}$$

$$u_{i,j,k} = S \pm \frac{1}{2}\sqrt{-4S^2 - \frac{p}{S}}, \tag{2.106}$$

where we have defined

$$S \equiv \frac{1}{2}\sqrt{\frac{1}{3}\left(Q + \frac{\Delta_0}{Q}\right)},$$
$$Q \equiv \sqrt[3]{\frac{1}{2}\left(\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}\right)}. \tag{2.107}$$

We need to find the correct branch of solution. First, note that S is a square root, and so we can show that if the quantity under the square root is a positive number, then $S > 0$. This is straightforward, as.

$$12S^2 = Q + \frac{\Delta_0}{Q} = \sqrt[3]{\frac{1}{2}\left(\sqrt{\Delta_1^2 - 4\Delta_0^3} + \Delta_1\right)} - \sqrt[3]{\frac{1}{2}\left(\sqrt{\Delta_1^2 - 4\Delta_0^3} - \Delta_1\right)} > 0, \tag{2.108}$$

Consider first the limit $p \to 0$. From the above equation we have

$$12S^2 \approx \sqrt[3]{(-\Delta_0)^{3/2} + \frac{1}{2}\Delta_1} - \sqrt[3]{(-\Delta_0)^{3/2} - \frac{1}{2}\Delta_1} \approx -\frac{1}{3}\frac{\Delta_1}{\Delta_0} = -\frac{3}{4}\frac{p^2}{q}, \tag{2.109}$$

which means that $S \simeq |p| \to 0$ but $p/S \to \pm 4\sqrt{-q}$ depending on the sign of p. This leads to the solution $u_{i,j,k} = \sqrt[4]{-q}$.

Given that $S > 0$, if $p > 0$, Equation (2.106) cannot be the physical branch because $u_{i,j,k}$ in this branch is complex. The '$-$' branch of Equation (2.105) cannot be chosen either, because $u_{i,j,k} < 0$, inconsistent with the requirement that $u_{i,j,k} > 0$.

If $p < 0$, Equation (2.105) cannot be the physical branch because $u_{i,j,k}$ in this branch is complex. Out of the two branches of Equation (2.106), we should choose '$+$', because this guarantees that when $p \to 0^-$ we still have $u_{i,j,k} > 0$.

Therefore, the analytical solution can be summarised as

$$u_{i,j,k} = \begin{cases} -S + \frac{1}{2}\sqrt{-4S^2 + \frac{p}{S}}, & p > 0, \\ \sqrt[4]{-q}, & p = 0, \\ S + \frac{1}{2}\sqrt{-4S^2 - \frac{p}{S}}, & p < 0. \end{cases} \tag{2.110}$$

Note that it can be shown that $8S^3 < |p|$ because $\Delta_1 = 27p^2$ and $\Delta_0 = 12q < 0$. This fact guarantees that in Equation (2.110) the square roots are real; it also guarantees that in

the $p > 0$ branch the condition $u_{i,j,k} > 0$ is satisfied (in the $p < 0$ branch it is satisfied automatically).

The existence of analytical solutions Equation (2.110) indicates that, like in the symmetron model, in the $n = 2$ case of $f(R)$ gravity here, it is not necessary to use the Newton approximation within the Gauss-Seidel relaxation, but the solution $u_{i,j,k}$ of cell (i, j, k) can be solved given the density field in this cell and the approximate solutions of u in the neighbouring cells.

- **The case of $n = 1$**

In this case, Equation (2.101) is a cubic equation of $u_{i,j,k}$ (Bose et al., 2017). Define $\Delta_0 \equiv -3p$, $\Delta_1 \equiv 27q$ and the discriminant

$$\Delta \equiv \Delta_1^2 - 4\Delta_0^3. \tag{2.111}$$

We see that $q < 0$ and so $\Delta_1 < 0$. The solution is given by

$$u_{i,j,k} = \begin{cases} -\frac{1}{3}\left(C + \frac{\Delta_0}{C}\right), & \Delta > 0, \\ \sqrt[3]{-q}, & \Delta = 0, \\ -\frac{2}{3}\sqrt{\Delta_0}\cos\left(\frac{\Theta}{3} + \frac{2\pi}{3}\right), & \Delta < 0, \end{cases} \tag{2.112}$$

where

$$C \equiv \sqrt[3]{\frac{1}{2}\left[\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}\right]}, \tag{2.113}$$

$$\Theta \equiv \arccos\left(\frac{\Delta_1}{2\sqrt{\Delta_0^3}}\right). \tag{2.114}$$

Again, the exact analytical solutions given in Equation (2.112) eliminates the need for Newton-Gauss-Seidel relaxations, and this has led to a significant improvement in the speed and convergence properties of simulations of this model compared with previous simulations (Bose et al., 2017).

- **The case of $n = 0$**

In this case, Equation (2.101) is a quadratic equation of $u_{i,j,k}$. The solution in this case is simple and the physical branch is given by

$$u_{i,j,k} = \frac{1}{2}\left[-p + \sqrt{p^2 - 4q}\right], \tag{2.115}$$

which satisfies $u_{i,j,k} > 0$.

2.5 Simulation: MG-GLAM code tests

In this section, we present various code test results to demonstrate the reliability of the equations, algorithms and implementations described in the previous sections. We follow the code test framework of the ECOSMOG code papers (Li et al., 2012, 2013b). Apart from the background cosmology test, all the tests shown in this section were performed on a cubic box with size $L = 256\,h^{-1}\,\mathrm{Mpc}$ and $N_{\mathrm{g}} = 512$ grid cells in each direction, and all background quantities are calculated at $a = 1$.

2.5.1 Background cosmology tests

Of the models considered in this work, only the coupled quintessence model can substantially affect the background expansion history, while for (viable) $f(R)$ gravity and symmetron models the expansion rate is practically indistinguishable from that of ΛCDM. In MG-GLAM, the background cosmology in the coupled quintessence model is solved numerically, as described in Section 2.4.3.

To check the numerical implementation, we have compared the predictions of certain background quantities by MG-GLAM with the results produced by the modified CAMB code, for the same coupled quintessence model, described in Li & Barrow (2011a). The results are presented in Fig 2.2, where the left panel shows the ratio between the background expansion rates of three coupled quintessence models and that of a ΛCDM model with the same (non-MG) cosmological parameters, while the right panel shows the background evolution of the scalar field, $\bar{\varphi}(a)$, for the same three models. Lines are from the modified CAMB code and symbols are for MG-GLAM. We see that the background cosmology solver of MG-GLAM agrees with the CAMB code very well in all cases.

There are two additional interesting features displayed in Fig. 2.2. First, the results are much more sensitive to β than to α, as can be observed by comparing the closeness between the black vs red lines, and the large difference between the black vs blue lines. This shows that the coupling to matter has a stronger impact on the scalar field background evolution than the potential itself.

Second, as discussed in Section 2.2.1, the scalar field affects structure formation through a combination of the following four effects:

- *modified expansion rate*: in the models studied here, the expansion rate is slowed down, which can lead to enhancement of structure formation.

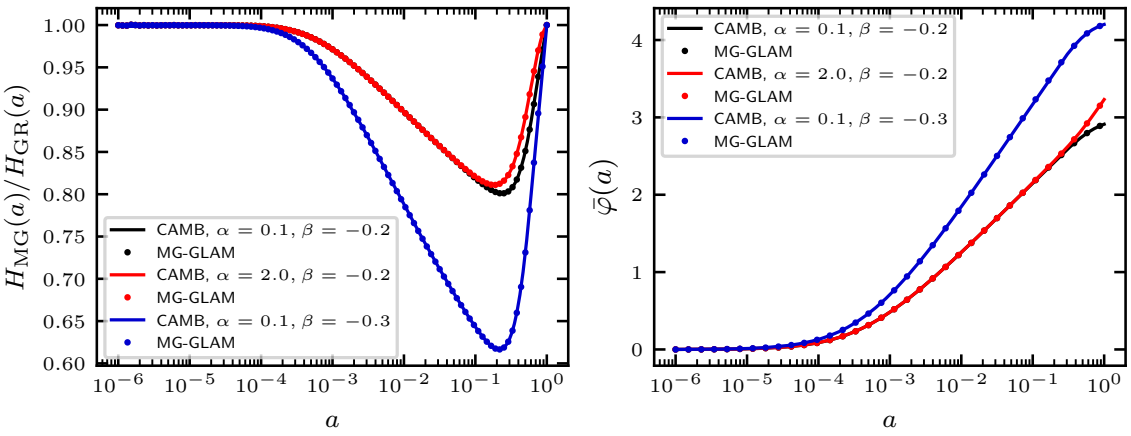

Figure 2.2: Cosmological background evolution tests. *Left panel:* The ratio of the Hubble parameters between the coupled quintessence and the ΛCDM models, from the modified CAMB (lines) and MG-GLAM (dots) codes for three kinds of model parameter values as labelled. *Right panel:* The evolution of the background scalar field in the coupled quintessence from CAMB (lines) and MG-GLAM codes.

- *fifth force*: the fifth-force-to-Newtonian-gravity ratio is a constant $2\beta^2$, and this boosts structure formation.

- *velocity-dependent force*: from the right panel of Fig. 2.2, we see that the scalar field is positive and grows over time such that, with $\beta < 0$, the term $(\mathrm{d}\ln A(\bar{\varphi})/\mathrm{d}\varphi)\,\bar{\varphi}' < 0$, which means that the velocity-dependent force is in the same direction as the particle velocity, i.e., it is essentially an 'anti-friction' force which tends to strengthen structure formation.

- *time variation of effective particle mass*: since the particle mass effectively depends on $\exp(\beta\bar{\varphi})$, with $\beta < 0$ and $\bar{\varphi} > 0$, at late times the effective mass decreases, which tends to weaken structure formation.

Therefore, the four effects work in different directions, and the net effect on structure formation — whether it is boosted or weakened — will need to be calculated numerically for specific models.

2.5.2 Density tests

This subsection is devoted to the tests of the multigrid solvers for the $f(R)$, symmetron and coupled quintessence models, using different density configurations for which the scalar field solution can be solved analytically or using a different numerical code.

2.5.2.1 Homogeneous density field tests

In a homogeneous density field the MG scalar field should also be homogeneous and exactly equal to its background value if the matter field is homogeneous, i.e.,

$$\tilde{\delta}(\tilde{x}) \equiv 0 \longrightarrow \begin{cases} f_R(\tilde{x})/\bar{f}_R \equiv 1, & f(R) \text{ gravity;} \\ \varphi(\tilde{x})/\varphi_* = \bar{\varphi}/\varphi_* \leq 1, & \text{symmetron;} \\ \tilde{c}^2\delta\varphi(\tilde{x}) \equiv 0, & \text{coupled quintessence.} \end{cases} \tag{2.116}$$

This offers a very simple test for the relaxation solvers described above, that is particularly useful for checking the implementation of multigrid.

We show the test results for a homogeneous density field in the left-hand panels of Fig. 2.3, where we display the scalar field values along the $\tilde{x}$ direction for fixed $\tilde{y}, \tilde{z}$ coordinates before (symbols) and after (lines) the multigrid relaxation, for two initial guesses (black and red). The three rows, from top to bottom, are respectively for the $f(R)$, symmetron and coupled quintessence models. For $f(R)$ gravity, the initial guesses are randomly generated from a uniform distribution within $\xi = f_R(\tilde{x})/\bar{f}_R \in [0, 2]$, and the model parameters used are $n = 1$ and $f_{R0} = -10^{-5}$; for the symmetron model, the random initial guesses are generated from a uniform distribution $\varphi(\tilde{x})/\varphi_* \in [0, 1]$ and the model parameters adopted are $a_* = 0.5, \xi = 10^{-3}, \beta_* = 0.1$; for coupled quintessence we consider the model parameters $\alpha = 0.1, \beta = -0.2$, and the initial guesses are from a uniform distribution $\delta\varphi(\tilde{x}) \in [-0.5, 0.5]$.

In all cases, we find that the solutions after relaxation agree very well to the analytical predictions given in Equation (2.116).

2.5.2.2 One-dimensional density tests

In the case of one spatial dimension, the scalar field satisfies ordinary differential equations. Therefore, we can construct a density field that has a known analytical solution of the scalar field, to check if the code returns the correct numerical solution, according to the scalar field equations of $f(R)$ gravity (Equation (2.99)), symmetron (Equation (2.82)) and coupled quintessence (Equation (2.71)) in code units. In practice this can be achieved by choosing a functional form of the scalar field in 1D, and applying the above equations to derive analytical expressions for $\delta(\tilde{x})$. For example, we can design density configurations in $f(R)$ gravity by manipulating Equation (2.99)

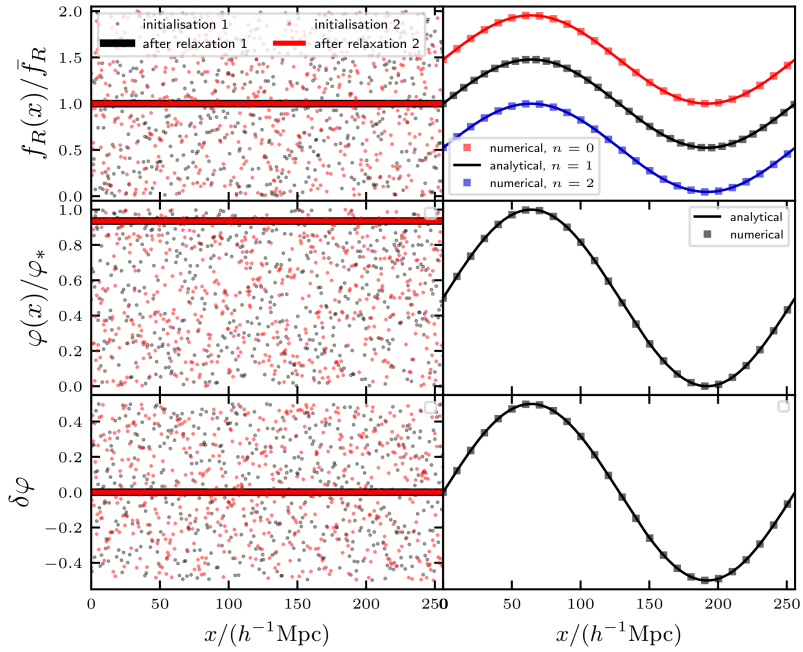

Figure 2.3: The uniform and one-dimensional code test results. The two columns show the cases with homogeneous (left) and sine (right) scalar fields, whilst the different rows represent the $f(R)$ gravity (upper), symmetron (middle) and coupled quintessence (bottom) models. *Left Panels:* Uniform density test, where the symbols represent the random initial guesses of the MG scalar field in the ranges of $[0, 2]$ ($f(R)$ gravity), $[0, 1]$ (symmetron) and $[-0.5, 0.5]$ (coupled quintessence), respectively. The solid lines show the numerical solutions after multigrid relaxation. Two random initialisations have been displayed in red and black. *Right Panels:* Sine field tests. The squares show the numerical results and the lines show the analytical solutions. The upper right panel shows the $f(R)$ gravity test results with $n = 0, 1$ and 2 as labelled.

in the 1D case as

$$\tilde{\delta}(\tilde{x}) = -\frac{a}{\Omega_{\rm m}}\left\{-\tilde{c}^2\tilde{\nabla}^2\left[u^{n+1}\right] - \frac{1}{3}\tilde{\bar{R}}(a)a^2\left[-\bar{f}_R(a)\right]^{1/(n+1)}\frac{1}{u} + \frac{1}{3}\tilde{\bar{R}}(a)a^2\right\}. \tag{2.117}$$

We have designed such tests where the scalar field solution is a sine function.

For $f(R)$ gravity, the scalar field takes the following sine-function form,

$$\frac{f_R(\tilde{x})}{\bar{f}_R} = 1 + A\sin\frac{2\pi\tilde{x}}{N_{\rm g}}, \tag{2.118}$$

if the density field is given by

$$\tilde{\delta}(\tilde{x}) = \frac{a}{\Omega_{\rm m}}\left[\tilde{c}^2\bar{f}_R\left(\frac{2\pi}{N_{\rm g}}\right)^2 A\sin\frac{2\pi\tilde{x}}{N_{\rm g}} + \frac{1}{3}\tilde{\bar{R}}(a)a^2\left(1 + A\sin\frac{2\pi\tilde{x}}{N_{\rm g}}\right)^{-\frac{1}{n+1}} - \frac{1}{3}\tilde{\bar{R}}(a)a^2\right], \tag{2.119}$$

where A is a constant and $|A| < 1$ as $f_R/\bar{f}_R$ should be positive. We have again adopted $f_{R0} = -10^{-5}$ and considered the three cases of $n = 0, 1, 2$ respectively.

For the symmetron model, we have taken the following form of the scalar field

$$u(\tilde{x}) = \frac{\varphi(\tilde{x})}{\varphi_*} = \frac{1}{2} + A\sin\frac{2\pi\tilde{x}}{N_{\rm g}}, \tag{2.120}$$

which corresponds to the following overdensity field,

$$\tilde{\delta}\left(\tilde{x}\right)=\frac{a^{3}}{a_{*}^{3}}\left[1-\left(\frac{1}{2}+A\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}\right)^{2}-\frac{2}{a^{2}}\xi^{2}\tilde{c}^{2}A\left(\frac{2\pi}{N_{\mathrm{g}}}\right)^{2}\frac{\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}}{\frac{1}{2}+A\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}}\right]-1. \tag{2.121}$$

The model parameter used here are the same as in the uniform density test above.

For the coupled quintessence model, we have taken the following form of the scalar field

$$\bar{\varphi}(\tilde{x})=\tilde{c}^{2}\delta\varphi(\tilde{x})=A\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}, \tag{2.122}$$

which corresponds to the following overdensity field,

$$\tilde{\delta}\left(\tilde{x}\right)=-\frac{a}{3\beta\Omega_{\mathrm{m}}}\exp(\beta\bar{\varphi})\left[A\left(\frac{2\pi}{N_{\mathrm{g}}}\right)^{2}\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}-\frac{\lambda^{2}a^{2}}{\left(\bar{\varphi}+\tilde{c}^{-2}A\sin\frac{2\pi\tilde{x}}{N_{\mathrm{g}}}\right)^{\alpha}}+\frac{\lambda^{2}a^{2}}{\bar{\varphi}^{\alpha}}\right]. \tag{2.123}$$

The model parameter used here are the same as in the uniform density test above.

The panels in the right column of Fig. 2.3 present the sine field test results for the three classes of models, in the same order as in the left column. The numerical solutions from MG-GLAM (squares) agree well with the analytical solutions of Equations (2.118, 2.120 and 2.122), shown by lines, indicating that the code works accurately to solve the scalar field equations. In all the tests shown here we have taken $A=0.5$, but we have checked other values of A, as well as sine functions with more than one oscillation period, and found similar agreements in all cases.

2.5.2.3 Three-dimensional density tests

As the final part of our tests of the multigrid relaxation solver, we consider slightly more complicated density configurations than the uniform and 1D density fields used previously. In order to get analytical and numerical solutions that can be compared with the predictions by MG-GLAM, we still would like to use density fields that have certain symmetries. To this end, we have done tests using a point mass (for $f(R)$ gravity) and spherical top-hat overdensity (for the symmetron and coupled quintessence models). These tests will see the scalar field values vary in x, y and z directions, and they are therefore proper 3D tests.

Point mass

For the first test in 3D space, we consider the solution of the scalar field around a point mass placed at the origin, for which we have approximated analytical solution that is valid in the regions far from the mass. This test has been widely performed in previous MG code papers such as Oyaizu

(2007); Brax et al. (2011a); Li et al. (2012); Arnold et al. (2019b). The matter overdensity array is constructed as

$$\tilde{\delta}_{i,j,k} = \begin{cases} 10^{-4}(N_{\rm g}^3 - 1), & i = j = k = 1; \\ -10^{-4}, & \text{otherwise.} \end{cases} \tag{2.124}$$

where $i, j, k = 1, ..., N_{\rm g}$ are the cell indices in x, y, z directions, respectively.

In $f(R)$ gravity, with this density configuration, the scalaron equation Equation (2.45) in regions far from the point mass simplifies to

$$\boldsymbol{\nabla}^2 \delta f_R \approx m_{\rm eff}^2 \delta f_R \,, \tag{2.125}$$

where $\delta f_R(\boldsymbol{x}) \equiv f_R(\boldsymbol{x}) - \bar{f}_R$, and the effective mass of the scalar field, $m_{\rm eff}$, is given by

$$m_{\rm eff}^2 \equiv -\frac{1}{3(n+1)} \frac{\bar{R}_0}{c^2 \bar{f}_{R0}} \left(\frac{\bar{R}}{\bar{R}_0}\right)^{n+2} = \frac{H_0^2 \Omega_{\rm m0}}{c^2 (n+1)(-\bar{f}_{R0})} \frac{\left(a^{-3} + 4\frac{\Omega_{\Lambda 0}}{\Omega_{\rm m0}}\right)^{n+2}}{\left(1 + 4\frac{\Omega_{\Lambda 0}}{\Omega_{\rm m0}}\right)^{n+1}} \,. \tag{2.126}$$

At $a = 1$, this only depends on the combination $(n+1)\bar{f}_{R0}$. For a spherically symmetric case such as the one considered here, the equation can be recast in the following form,

$$\frac{1}{r^2} \frac{\rm d}{{\rm d}r} \left(r^2 \frac{{\rm d}\delta f_R}{{\rm d}r} \right) = m_{\rm eff}^2 \, \delta f_R, \tag{2.127}$$

or equivalently

$$\frac{{\rm d}^2}{{\rm d}r^2} \left[r \delta f_R(r) \right] = m_{\rm eff}^2 \, r \, \delta f_R(r), \tag{2.128}$$

where r is the distance from the central point mass. This equation has the solution

$$r \, \delta f_R(r) = \alpha_1 \exp\left(-m_{\rm eff} r\right) + \alpha_2 \exp\left(m_{\rm eff} r\right), \tag{2.129}$$

where α_1, α_2 are constants of integral, and we must have $\alpha_2 = 0$ to prevent the solution from diverging at $r \to \infty$. This leads to the following solution

$$\delta f_R(r) \propto \frac{1}{r} \exp\left(-m_{\rm eff} r\right), \tag{2.130}$$

which in code unit can be rewritten as

$$\delta f_R\left(\tilde{r}\right) \propto \frac{1}{\tilde{r}} \exp\left(-\tilde{m}_{\rm eff} \tilde{r}\right), \tag{2.131}$$

where the $\tilde{m}_{\rm eff}$ is the scalar field mass $m_{\rm eff}$ in code unit, given by

$$\tilde{m}_{\rm eff}^2 \equiv \frac{\Omega_{\rm m0}}{\tilde{c}^2 (n+1)(-\bar{f}_{R0})} \frac{\left(a^{-3} + 4\frac{\Omega_{\Lambda}}{\Omega_{\rm m}}\right)^{n+2}}{\left(1 + 4\frac{\Omega_{\Lambda}}{\Omega_{\rm m}}\right)^{n+1}} \,. \tag{2.132}$$

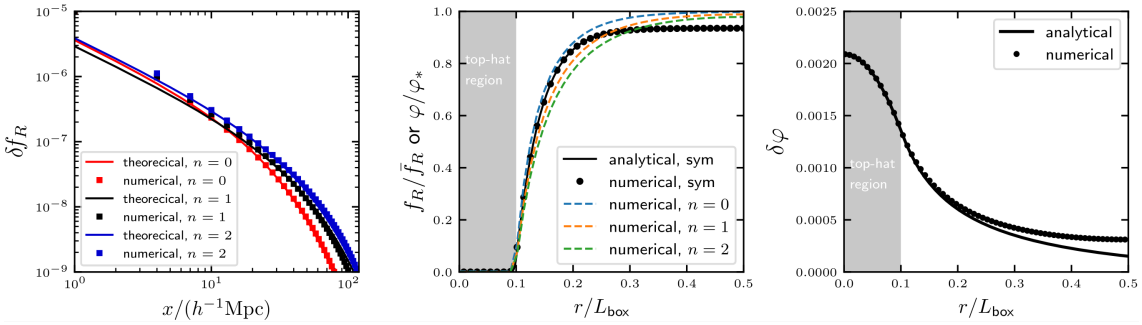

Figure 2.4: The three-dimensional code test results. *Left Panel:* The numerical (squares) and analytical (lines) solutions to $\delta f_R \equiv f_R - \bar{f}_R$ around a point mass located at $(x, y, z) = (0, 0, 0)$, for three $f(R)$ gravity models with $f_{R0} = -10^{-5}$ and $n = 0$ (red), 1 (black) and 2 (blue), respectively. The analytical approximations are only valid far from the point mass. Only the solutions along the x-axis are shown. *Middle Panel:* Top-hat overdensity test for the symmetron (black) and $f(R)$ gravity models (dashed lines). The lines correspond to the analytical solutions and the dots represent the numerical results. The quantities shown on the y-axis are $\varphi(r)/\varphi_*$ for the symmetron model, and $f_R(r)/\bar{f}_R$ for the $f(R)$ model. *Right Panel:* The same as the middle panel but for the coupled quintessence model.

Note that we have neglected the tilde for δf_R since in our code units $\tilde{f}_R = f_R$ and $\tilde{\bar{f}}_{R0} = \bar{f}_{R0} \equiv f_{R0}$.

In the left panel of Fig. 2.4, we show the numerical solutions from MG-GLAM and the analytical results given in Equation (2.131). Notice that the latter has an unknown coefficient, which we have tuned to match the amplitude of the MG-GLAM solution. Once that is done, the two agree very well for all three $f(R)$ gravity models with $\bar{f}_{R0} = -10^{-5}$ for $n = 0, 1$ and 2 respectively, except on scales smaller than $\simeq 5\,h^{-1}\,\mathrm{Mpc}$ since Equation (2.125) is not valid near the point mass, and far from the point mass where the MG-GLAM solution starts to see the effect of periodic boundary condition, which is absent in Equation (2.131).

Spherical top-hat overdensity

For the symmetron and coupled quintessence models, instead of a point mass test, we have considered a spherical top-hat overdensity with radius $\tilde{R}_{\mathrm{TH}}$ located at the centre of the simulation box $(\tilde{x}, \tilde{y}, \tilde{z}) = (N_{\mathrm{g}}/2, N_{\mathrm{g}}/2, N_{\mathrm{g}}/2)$. Note that code units are used here. The overdensity field is given by

$$\tilde{\delta}_{\mathrm{TH}}(\tilde{r}) = \begin{cases} \tilde{\delta}_{\mathrm{in}}, & \tilde{r} \leq \tilde{R}_{\mathrm{TH}} \\ \tilde{\delta}_{\mathrm{out}}, & \tilde{r} > \tilde{R}_{\mathrm{TH}} \end{cases}, \tag{2.133}$$

where $\tilde{r} \equiv \sqrt{(\tilde{x} - N_{\mathrm{g}}/2)^2 + (\tilde{y} - N_{\mathrm{g}}/2)^2 + (\tilde{z} - N_{\mathrm{g}}/2)^2}$ is the distance from the top-hat centre, and we have adopted $\tilde{R}_{\mathrm{TH}} = 0.1 N_{\mathrm{g}}$, $\tilde{\delta}_{\mathrm{in}} = 5000$ and $\tilde{\delta}_{\mathrm{out}} = 0$ in our tests. In spherical symmetry, the scalar field equations for the symmetron (Equation (2.82)) and coupled quintessence

(Equation (2.71)) models reduce to the following 1D ordinary differential equations,

$$\tilde{c}^2\frac{1}{\tilde{r}^2}\frac{\mathrm{d}}{\mathrm{d}\tilde{r}}\left(\tilde{r}^2\frac{\mathrm{d}u}{\mathrm{d}\tilde{r}}\right)=\frac{a^2}{2\xi^2}\left[\left(1+\tilde{\delta}_{\rm TH}(\tilde{r})\right)\frac{a_*^3}{a^3}-1\right]u+\frac{a^2}{2\xi^2}u^3 \tag{2.134}$$

and

$$\begin{aligned}\frac{1}{\tilde{r}^2}\frac{\mathrm{d}}{\mathrm{d}\tilde{r}}\left(\tilde{r}^2\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{r}}\right)=&\frac{3\beta\Omega_{\rm m}}{a}e^{\beta\bar{\varphi}}\left[\exp\left(\beta\frac{\tilde{\varphi}}{\tilde{c}^2}\right)\left(1+\tilde{\delta}_{\rm TH}(r)\right)-1\right]\\&-\alpha\lambda^2a^2\left[\frac{1}{(\bar{\varphi}+\tilde{c}^{-2}\tilde{\varphi})^{1+\alpha}}-\frac{1}{\bar{\varphi}^{1+\alpha}}\right],\end{aligned} \tag{2.135}$$

respectively.

These two 1D equations are numerically solved using the MAPLE software, with the following boundary conditions on the interval $r\in[0,N_{\rm g}/2]$,

$$u(\tilde{r}=N_{\rm g}/2)=\sqrt{1-\left(\frac{a_*}{a}\right)^3},\quad \frac{\mathrm{d}u}{\mathrm{d}\tilde{r}}(\tilde{r}=0)=0\,, \tag{2.136}$$

and

$$\tilde{\varphi}(\tilde{r}=N_{\rm g}/2)=0,\quad \frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{r}}(\tilde{r}=0)=0\,, \tag{2.137}$$

for the symmetron and coupled quintessence models respectively. Note that rigorously speaking the first boundary condition should really have been set at $\tilde{r}\to\infty$, but for numerical implementation this is impractical and we instead use $N_{\rm g}/2$ as an approximation to ∞.

We have obtained the numerical solutions of these ordinary differential equations for $u(\tilde{r})$ and $\tilde{\varphi}(\tilde{r})$, but still call them 'analytical' to distinguish from the numerical solutions directly solved from the original PDEs solved by MG-GLAM. The model parameters are the same as in the uniform and 1D density tests: for the symmetron model we have used $a_*=0.5,\xi=10^{-3},\beta_*=0.1$, while for coupled quintessence we have used $\alpha=0.1,\beta=-0.2$.

The analytical and numerical solutions for the symmetron and coupled quintessence models are displayed in the middle and right panels of Fig. 2.4, respectively as the black solid line and black symbols. They agree very well.

As a comparison, in the middle panel of Fig. 2.4 we have also shown, with coloured symbols, the MG-GLAM solutions for the $f(R)$ model with $\bar{f}_{R0}=-10^{-5}$ and $n=0$ (blue), 1 (orange) and 2 (green). This can serve as a quick comparison of the screening efficiencies in these four models. First, we note that in all three $f(R)$ models the solution, $f_R(\tilde{r})/\bar{f}_{R0}$, tends to 1 far from the spherical top-hat, which is expected because the scalar field approaches its background value far from the matter perturbation at the centre. Second, for all four models, the scalar field is strongly

suppressed inside the top-hat (grey shaded region), but increases sharply immediately outside $\tilde{R}_{\rm TH}$ such that within some small distance from the edge of the top-hat it already reaches $\gtrsim 50\%$ of the background value: this is what one would expect from the $f(R)$ and symmetron models — both of which are examples of the so-called *thin-shell screened* models (Brax et al., 2012b). Third, comparing the solutions of the three $f(R)$ models with the same f_{R0}, it seems that increasing the value of n increases the screening efficiency, implying that the $n = 2$ case has the strongest screening amongst them; we shall see the consequence of this in the cosmological simulations in the next section. Finally, comparing the tested symmetron model with the $f(R)$ ones, it seems that the solution of the former lies somewhere in between the $n = 0$ and $n = 1$ cases (at least near the top-hat); however, we caution that the fifth forces in the two models are obtained in different ways: in $f(R)$ gravity it is directly proportional to $\boldsymbol{\nabla} f_R$, while for symmetrons it is proportional to $\boldsymbol{\nabla}\left(u^2\right)$, cf. Equation (2.94), rather than $\boldsymbol{\nabla} u$.

2.5.3 Convergence tests

As mentioned in Sect. 2.4.1, we have implemented three different arrangements of the multigrid solver — V-, F- and W-cycles. To compare them we have run a series of small cosmological simulations for the $f(R)$ gravity model with $\bar{f}_{R0} = -10^{-5}$ and $n = 1$, the symmetron model with $a_* = 0.3, \xi = 10^{-3}$ and $\beta_* = 0.1$ and the coupled quintessence model with $\alpha = 0.1$ and $\beta = -0.2$. These runs all use $L = 256\,h^{-1}{\rm Mpc}$, $N_{\rm p}^3 = 512^3$ and $N_{\rm g}^3 = 1024^3$ for the smaller simulations. We consider 10 and 2 V-cycles (V10 and V2), 1 F-cycle (F1) and 1 W-cycle (W1) to test the convergence of the MG scalar field solutions. The V10 simulation results are used as the benchmark of our test. For F- and W-cycles we only consider one cycle because, as will be shown below, this already gives excellently converged results.

Figure 2.5 shows the relative differences of the matter power spectra at $z = 0$ between the test simulations described above and the benchmark case (V10), for $f(R)$ gravity (left), and the symmetron (middle) and coupled quintessence (right) models. We find that all the different schemes and different numbers of cycles used to solve the partial differential equations have good agreement on almost all scales probed by the simulations ($\lesssim 0.4\%$). Also in Fig. 2.6, we directly check the MG scalar field $f_R(\boldsymbol{x})$ obtained by different arrangements. The scalar fields solved by V2, F1 and W1 differ with that obtained from V10 by less than 1%. When more cycles are used, the run time gets longer, and the slowest simulations are those using V10. F-cycles and W-cycles are more effective in reducing residuals, both agreeing with V10 by $\lesssim 0.05\%$ after only one cycle,

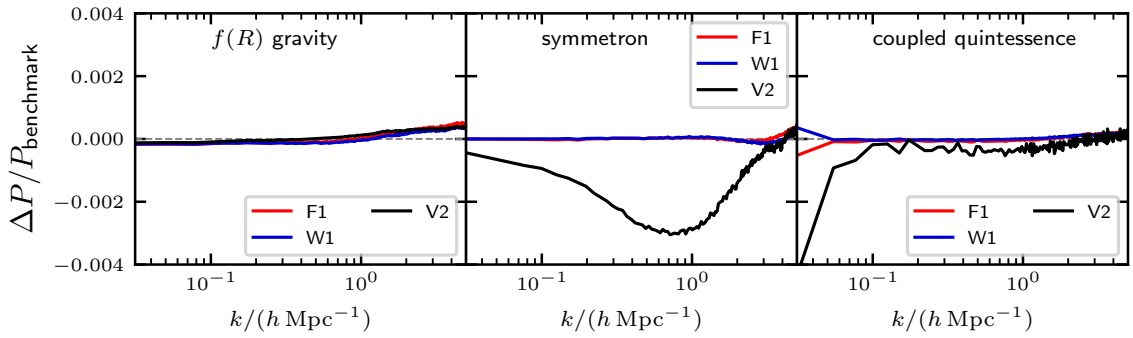

Figure 2.5: A comparison of the convergence of the Gauss-Seidel relaxation with different multigrid arrangements and numbers of cycles. The fractional differences in the matter power spectra are plotted at $z = 0$, obtained for different multigrid schemes (V2, V10, F1, W1) using V10 as the reference. The cases shown are for the $f(R)$ model with $f_{R0} = -10^{-5}$ and $n = 1$ (left panel), the symmetron model with $a_* = 0.3, \xi = 10^{-3}$ and $\beta_* = 0.1$ (middle panel), and the coupled quintessence model with $\alpha = 0.1, \beta = -0.2$ (right panel).

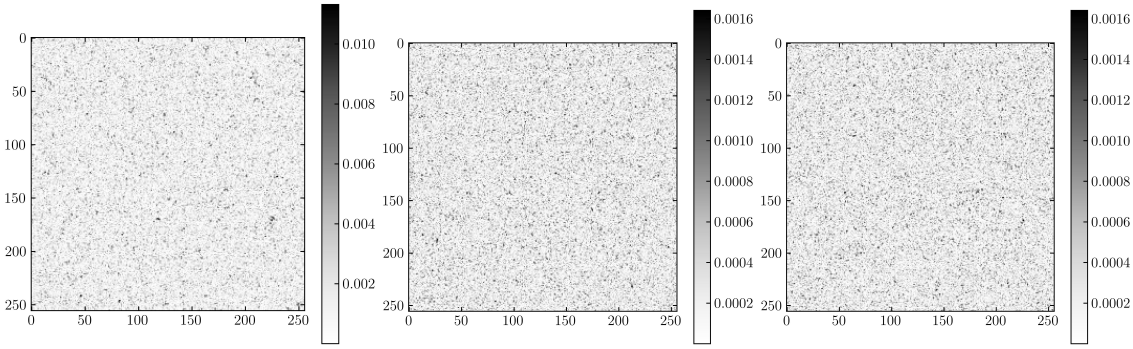

Figure 2.6: The fractional differences in the MG scalar field, $\Delta f_R(\boldsymbol{x})/f_R^{\rm benchmark}$, obtained for different multigrid schemes (V2, F1, W1, from left to right) using V10 as the reference. The model shown is for the $f(R)$ model with $f_{R0} = -10^{-5}$ and $n = 1$, at redshift $z = 0$.

which is not surprising since they walk more times across the fine and coarse multigrid levels. As a result, Both F1 and W1 are slower than V2. Therefore, as a compromise between accuracy and cost, we decide to always use V2 in our cosmological runs.

It is a great achievement for the multigrid solver to reach convergence after just 2 V-cycles (and 2 Gauss-Seidel passings of the entire mesh in each cycle), for non-linear equations in the $f(R)$ gravity and symmetron models.

2.5.4 Comparisons with previous simulations

As a final test of MG-GLAM, we compare its predictions from cosmological simulations with those by other modified gravity codes in the literature. We do this for the $f(R)$ and symmetron models only, since the coupled quintessence model is more trivial: the fifth force in this model is unscreened, and has a nearly constant ratio with the strength of Newtonian gravity in space (Li & Barrow, 2011a).

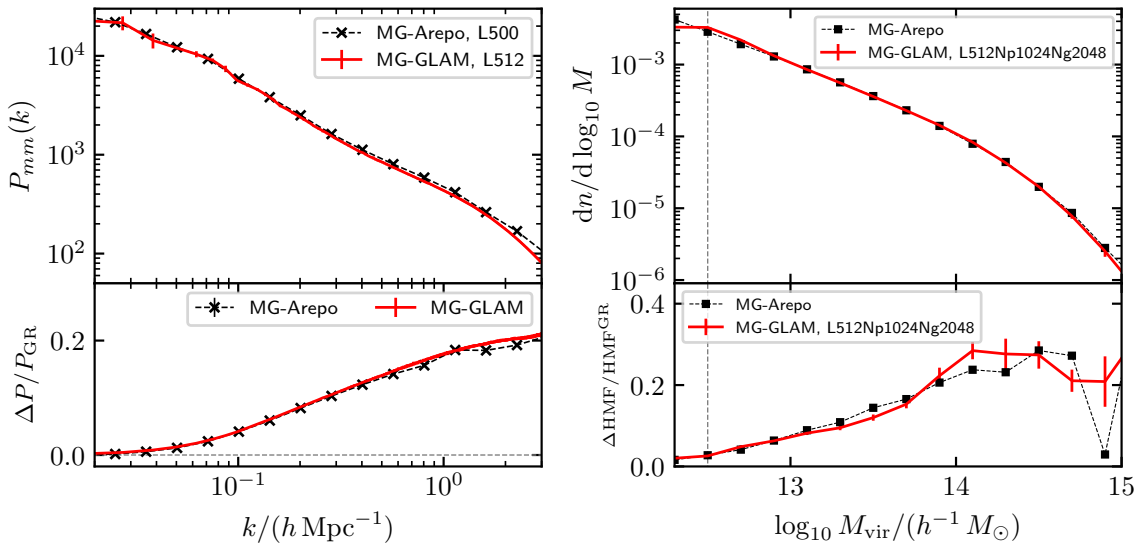

Figure 2.7: Comparison of matter power spectra (left panel) and halo mass functions (right panel) predicted by simulations with the same box size and particle number, using the MG-GLAM (black dashed lines with symbols) and MG-Arepo (red solid lines) codes for the same $f(R)$ model, $n = 1$ and $-f_{R0} = 10^{-5}$. The upper sub-panels show the absolute measurements from the simulations, while the lower sub-panels show the relative differences from the counterpart ΛCDM runs. The vertical dashed line in the right panel denotes $10^{12.5}h^{-1}M_\odot$. The two codes agree very well above this mass.

For $f(R)$ gravity, we have run two MG-GLAM simulations for the model $f_{R0} = -10^{-5}, n = 1$, using a box size $L = 512\,h^{-1}{\rm Mpc}$ with $N_{\rm p}^3 = 1024^3$ particles and $N_{\rm g}^3 = 2048^3$ mesh cells. These are compared with the predictions from a simulation using MG-Arepo, with $L = 500\,h^{-1}{\rm Mpc}$ and $N_{\rm p}^3 = 1024^3$. All these simulations have the same cosmological parameters, but they started from different realisations of initial conditions. Since MG-Arepo uses adaptive mesh refinement for the modified gravity force and trees for the Newtonian force with a softening length of $\approx 15\,h^{-1}{\rm kpc}$, it achieves better force resolution as compared with the MG-GLAM simulations that use a regular mesh with $N_{\rm g} = 2048$ giving a force resolution of $250\,h^{-1}{\rm kpc}$. Despite this, we will see that MG-GLAM can reproduce the MG-Arepo results on scales of interest.

In the left panel of Fig. 2.7 we compare the matter power spectra, $P_{\rm mm}(k)$, from the MG-GLAM (lines) and MG-Arepo (symbols) simulations. The upper sub-panel shows the absolute $P(k)$, where the two codes agree down to $k \approx 1\,h\,{\rm Mpc}^{-1}$. As shown in Klypin & Prada (2018), with a mesh resolution of $0.25\,h^{-1}{\rm Mpc}$ the GLAM code is capable of predicting $P_{\rm mm}(k)$ with percent-level accuracy down to $k \approx 1\,h\,{\rm Mpc}^{-1}$. The lower sub-panel shows the enhancements of the matter power spectrum due to $f(R)$ gravity. To obtain this, we have also run a counterpart ΛCDM simulation for each of the $f(R)$ simulations, using the same box size, grid number, particle number, cosmological parameters and initial conditions; we then take the relative difference between an $f(R)$ run and its counterpart ΛCDM run. We can see an excellent agreement between

the two codes, down to $k \approx 3\,h\,\mathrm{Mpc}^{-1}$ (even though the power spectra themselves agree only down to $k \approx 1\,h\,\mathrm{Mpc}^{-1}$).

The right panel of Fig. 2.7 extends the comparison to the differential halo mass function (dHMF). The dHMF is a description of the halo abundance; more accurately, it quantifies the number density of haloes, in a spatial volume, that falls into a given halo mass bin. In the upper sub-panel we present the dHMFs measured from the MG-GLAM and MG-AREPO simulations, while in the lower sub-panel we show the enhancements with respect to their counterpart ΛCDM runs. As we mentioned above, MG-GLAM uses the spherical overdensity halo mass definition with the virial halo overdensity, $M_{\rm vir}$. On the other hand, MG-AREPO by default uses the M_{200c} halo mass definition, which is defined by requiring the mean overdensity within the halo radius R_Δ, to be $\Delta = 200\rho_{\rm crit}(z)$. To be self-consistent, we have rerun MG-AREPO's halo finder, SUBFIND (Springel et al., 2001), using the $M_{\rm vir}$ definition. The upper sub-panel shows that, at this specific mesh resolution, the dHMF predicted by MG-GLAM is complete down to $10^{12.5}h^{-1}M_\odot$, and agrees with MG-AREPO for $M > 10^{12.5}h^{-1}M_\odot$. In addition, the dHMF enhancements due to $f(R)$ gravity predicted by these two codes also agree very well, despite being noisy at the high-mass end due to the small box sizes used here.

Overall, for the $f(R)$ version, we find very good agreement between MG-GLAM and MG-AREPO. We have also compared the MG-GLAM simulation results with predictions by ECOSMOG (although the results are not presented here), and obtained as good agreements as shown in Fig. 2.7 for both $P_{\rm mm}(k)$ and the HMF.

For the symmetron model, we have three MG-GLAM runs for the parameter values $a_* = 0.33$, $\beta_* = 1$ and $\xi = 3.34 \times 10^{-4}$, and we compare the measured matter power spectra with those presented in Brax et al. (2012c) and Winther et al. (2012) using the adaptive mesh refinements codes ECOSMOG and MLAPM, respectively. The ECOSMOG symmetron run followed the evolution of $N_{\rm p}^3 = 256^3$ particles in a box of size $L = 128\,h^{-1}\mathrm{Mpc}$, and the domain grid (defined as the finest uniform grid which covers the whole simulation box) has $N_{\rm g}^3 = 256^3$ cells. For the MLAPM simulation, the box size is $64\,h^{-1}\mathrm{Mpc}$, the particle number is 256^3 and the domain grid cell number is 128^3.

Fig. 2.8 shows the matter power spectrum enhancement, $\Delta P/P_{\rm GR}$, between a pair of ΛCDM and MG simulations starting from the same initial conditions, for the MG-GLAM runs (solid lines) and the data taken from Brax et al. (2012c) (red symbols) and Winther et al. (2012) (black symbols). For the power spectrum itself, the sample variance should be smaller on smaller scales, which have

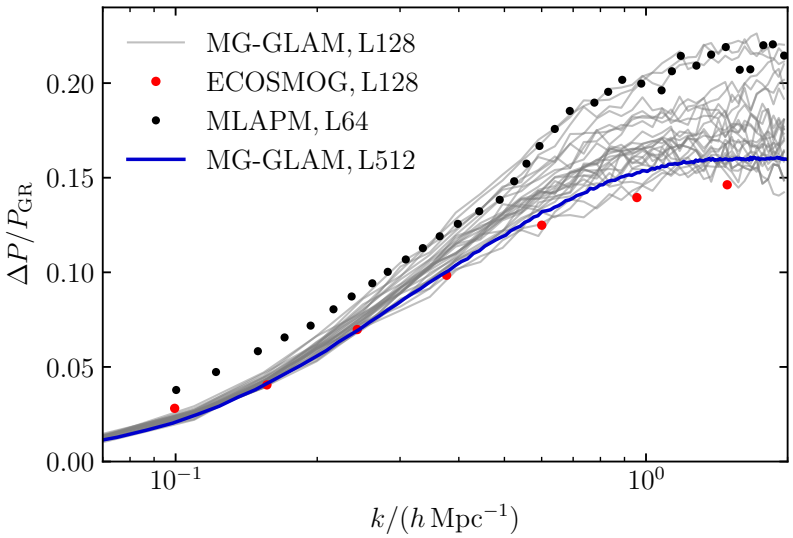

Figure 2.8: Comparison of the matter power spectrum enhancements with respect to ΛCDM, for the symmetron model with $a_* = 0.33$, $\beta_* = 1$ and $\xi = 3.34 \times 10^{-4}$ at $z = 0$, from two previous simulations run with ECOSMOG (Brax et al., 2012c) (red symbols) and MLAPM (Winther et al., 2012) (black symbols) respectively, and two groups of MG-GLAM runs with box sizes of 128 (thin grey lines) and $512\,h^{-1}{\rm Mpc}$ (thick blue line). The box sizes used in the MLAPM and ECOSMOG runs are respectively $64\,h^{-1}{\rm Mpc}$ and $128\,h^{-1}{\rm Mpc}$. The 25 small-box MG-GLAM realisations have substantial sample variance (as shown by the strong scatters) on small scales due to the box size, and these curves are also much noisier compared with the large-box result. The ECOSMOG and MLAPM results are close to the two limits of the scatters in the 25 MG-GLAM runs. The large scale behaviour of the MLAPM result is likely due to its very small box size.

more k modes than large scales. For $\Delta P/P_{\rm GR}$, however, we see the opposite behaviour: the sample variance is substantially suppressed on large scales ($k < 0.1\,h^{-1}{\rm Mpc}$) where the evolution is largely linear and different k modes uncoupled to each other. On small scales, different Fourier modes are coupled together, and two different gravity models that have different strengths of gravity would lead to different levels of such coupling. Hence, the difference between the power spectra in these two models at the same high k can be substantial, especially when the box size is small and therefore the result is more susceptible to rare large objects present in the box. An example to illustrate this point is the bottom panel of Fig. 5 of Zhao et al. (2011), which compares the $\Delta P/P_{\rm GR}$ from an $f(R)$ simulation and a 'linearised' (no-chameleon) $f(R)$ simulation which has its chameleon screening effects removed — the latter case corresponds to a much stronger gravity and a much larger power spectrum enhancement, and therefore much more significant scatters.

In Fig. 2.8, we find that the MG-GLAM simulations for the 25 independent realisations with a small box ($L = 128\,h^{-1}{\rm Mpc}$) have strong scatters in $\Delta P/P_{\rm GR}$ on small scales while not on large scales. The previous simulation results from ECOSMOG and MLAPM follow into this range of scatters. Note that the MLAPM simulation has a box size of $64\,h^{-1}{\rm Mpc}$ which can likely explain its behaviour on large scales. We also show the result from a single $512\,h^{-1}{\rm Mpc}$ box. Within the

uncertainties allowed by sample variance, all three codes seem to agree with each other.

2.5.5 Code test summary

To quickly sum up this section: we have done a number of tests of different aspects of the MG-GLAM code. These include the test of the background cosmology solver for the coupled quintessence model (cf. Section 2.5.1), tests of the multigrid relaxation solver of the scalar field equations for different density configurations (cf. Sections 2.5.2.1, 2.5.2.2 and 2.5.2.3), convergence property tests of the relaxation solvers with three different multigrid arrangements (V-cycle, F-cycle and W-cycle), and additionally comparisons of MG-GLAM cosmological simulations with runs using other codes. We see that MG-GLAM satisfactorily passes all these tests, and gives reasonable results.

Chapter 3

From matter to haloes: MG-GLAM II

This chapter shows the MG-GLAM simulation products for a large suite of MG models, as summarised in Section 3.1, which serve to showcase the potential power of this new code. In Section 3.2, we review the streaming model of redshift space distortion (RSD), with a particular focus on one of its ingredients — the pairwise velocity probability distribution function (PDF) of dark matter halos. In Section 3.3, we study the behaviour of the halo pairwise velocity PDFs in a wide range of MG models, show that the streaming model with the skew-T PDF accurately reproduces redshift-space two-point correlation functions, and perform a simple Fisher matrix analysis to assess the impact of including small-scale RSD on the model constraints. Throughout, our analysis is based on dark matter haloes, and we leave the extension of the RSD modelling to galaxy clustering for future work.

3.1 Cosmological runs

In this section, we will present some examples of cosmological runs using this code. In particular, we will run a very large suite of $f(R)$ simulations with different parameter values of n and $\bar{f}_{R0}$. These simulations only take a small fraction of time of a single high-resolution run of MG-Arepo or ECOSMOG for the box size and mass resolution.

The inventory of the cosmological runs we have performed is (see also Table 3.1 for a summary)

- Seventy-two ΛCDM simulations using GLAM.
- $f(R)$ gravity runs with $n = 0, 1$ and 2 and $\log_{10}(|\bar{f}_{R0}|)$ in 10 bins linearly spaced in the range $[-6.00, -4.50]$. One realisation for each model.
- Ten realisations of $f(R)$ gravity runs with $n = 0$ and 1 and $-\log_{10}(|\bar{f}_{R0}|) = 5.00$.

- Five symmetron models with fixed $a_* = 0.33$ and $\beta_* = 1$, with different values of ξ given by $c\xi/H_0 = 0.5, 1, 2, 2.5, 3$.
- Three variants of the coupled quintessence model, with (α, β) equal to $(0.1, -0.1)$, $(0.1, -0.2)$ and $(0.5, -0.2)$ respectively.
- For each MG simulation, we have a counterpart ΛCDM run with the same simulation specifications of cosmological parameters. We will label these runs as 'GR' runs, to contrast with 'MG' runs, even though none of our simulations is really general relativistic.

For all simulations, we followed the evolution of 1024^3 particles in a cubic box with size $512\,h^{-1}\mathrm{Mpc}$ using a grid with 2048^3 cells. The non-MG cosmological parameters are from the Planck 2015 (Planck Collaboration XIII, 2016) best-fitting ΛCDM parameters:

$$\{\Omega_{\rm b0}, \Omega_{\rm m0}, h, n_s, \sigma_8\} = \{0.0486, 0.3089, 0.6774, 0.9667, 0.8159\}.$$

The initial conditions of both the GR and MG runs are generated on the fly from the same ΛCDM linear perturbation theory power spectrum at $z_{\rm init} = 100$, which itself is generated using the CAMB code. We have used the same initial conditions for GR and MG simulations (for the same realisation), since the MG effect is very weak at $z > 100$, so that the linear matter power spectrum at $z_{\rm init} = 100$ is nearly identical to that of ΛCDM.

The dark matter halo catalogues are produced using the BDM spherical overdensity halo finder (Klypin et al., 2011). Only main haloes are studied in this work since the sub-haloes are not well resolved due to the limited force resolution. However, in the code papers we found that the main haloes of these simulations are complete down to a halo mass of $\simeq 10^{12.5}\,h^{-1}M_\odot$. This should allow us to construct mock galaxy catalogues based on the HOD model (see, e.g., Berlind et al., 2003; Zheng et al., 2005a), though in this paper we will focus on haloes for simplicity, and leave a more detailed analysis using realistic mock galaxy catalogues to future work. For the halo mass definition, the BDM halo finder adopts the virial mass $M_{\rm vir}$, which is the mass enclosed within a spherical overdensity of radius $R_{\rm vir}$, such that the mean overdensity within this sphere is $\Delta_{\rm vir} \approx 330$ times the mean matter density of the Universe. The virial overdensity is calculated according to Bryan & Norman (1998b). We saved halo catalogues at redshift $z = 0.0, 0.5$ and 1.0 for analysis, and show the results of $z = 0.5$ in the main text. The results obtained from the other snapshots are presented in Appendix D.

In Fig. 3.1, we show some of the basic cosmological quantities predicted by the (MG-)GLAM simulations. The left panel shows the cumulative halo mass function (cHMF) for the ΛCDM, F5n0,

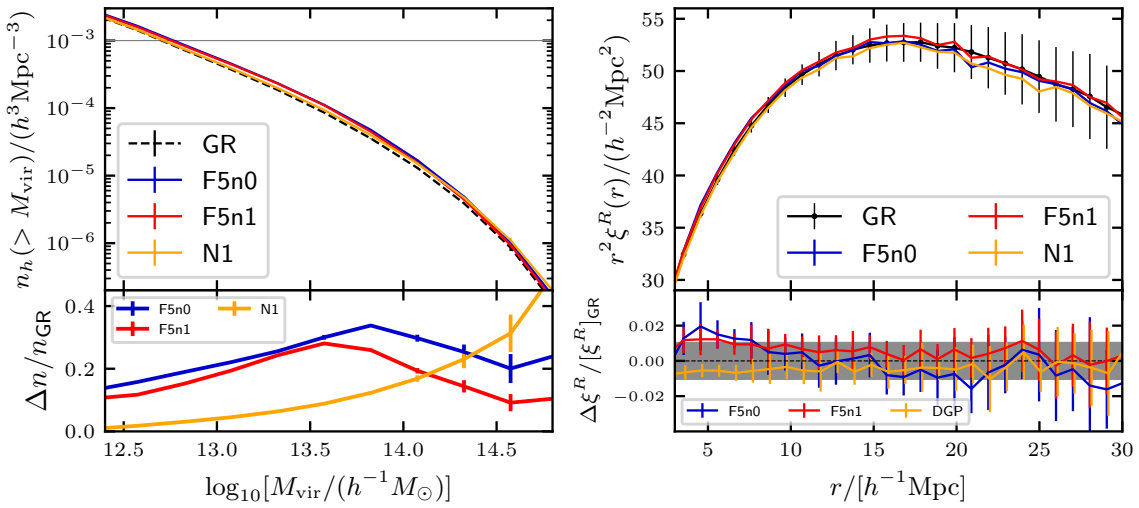

Figure 3.1: Cumulative halo mass functions (left panel) and halo real-space correlation functions (right panel) at $z = 0.5$, from (MG-)GLAM simulations of the $f(R)$ model with $f_{R0} = -10^{-5}$, $n = 0$ (F5n0, blue) and $n = 1$ (F5n1, red), the DGP model with $H_0 r_c = 1$ (N1, orange) and ΛCDM (black). The ΛCDM halo catalogues have fixed number density $n_h = 10^{-3}\,(h^{-1}{\rm Mpc})^{-3}$ (indicated by the grey horizontal line in the upper left panel) by selecting haloes more massive than a threshold value $M_{\rm min}$. For MG halo catalogues, the mass cuts are tuned to match the ΛCDM halo correlation functions over a range of scales. The lower sub-panels show the fractional difference between the MG and ΛCDM results, with the grey shaded region in the lower right panel indicating $\pm 1\%$. The error bars present the standard deviation over 10 realisations for each model (72 for ΛCDM results). Only the ΛCDM error bars are displayed in the upper left panel for clarity.

F5n1 and N1 models, each showing the mean of 10 independent realisations; the lower sub-panel shows the enhancements of the MG models with respect to GR, which agree very well with earlier simulation studies as shown in the code test section 2.5. The right panel of Fig. 3.1 compares the real-space halo TPCF of the same models; here we have tuned the number densities of the halo catalogues* such that the TPCFs in all models agree with each other within $\simeq 1\%$ between 5 and $30\,h^{-1}{\rm Mpc}$ (see the lower sub-panel, which shows the relative difference from GR after the tuning).

In Appendix C, we compare several halo pairwise velocity statistics predicted by these simulations with the predictions from a high-resolution simulation using a different code, and find that MG-GLAM gives reliable results down to small scales.

3.1.1 $f(R)$ gravity

We have measured the matter power spectra $P_{\rm mm}(k)$ and halo mass functions (HMF) at $z = 0$. The results are shown in Fig. 3.2 for the matter power spectra and Fig. 3.3 for the halo mass functions.

*Note that this tuning means that here we are not comparing halo catalogues in different models with exactly the same number density, but the latter is not our main interest anyway. On the other hand, as we shall see below, this tuning will make it easier when comparing other physical quantities.

Table 3.1: The summary of the specifications of the simulations used in this chapter.

Simulation	Code	Model (× #realisations)	Box size (h^{-1}Mpc)	$M_{\rm particle}$ ($h^{-1}M_\odot$)	$N_{\rm particle}$	Force resolution (h^{-1}kpc)
GLAM	GLAM	GR (×72)	512	1.07×10^{10}	1024^3	250.0
MG-GLAM	MG-GLAM	F5n0 (×10) F5n1 (×10) N1 (×10)	512	1.07×10^{10}	1024^3	250.0
MG-GLAM	MG-GLAM	28 $f(R)$ models (×1) 29 DGP models (×1)	512	1.07×10^{10}	1024^3	250.0
LIGHT-CONE	MG-GADGET	GR (×1) F5n1 (×1)	768	4.50×10^9	2048^3	10.0

The relative differences for $P_{\rm mm}(k)$ and HMF between $f(R)$ and GR are displayed in the lower sub-panels.

Fig. 3.2 shows that the matter clustering is boosted by 3-40% due to the fifth force, but the boost is scale-dependent and is weak on very large scales ($k \lesssim 0.03\,h\,{\rm Mpc}^{-1}$). The $P_{\rm mm}(k)$ enhancement, $\Delta P/P_{\rm GR}$, depends qualitatively on the value of $|\bar{f}_{R0}|$. When $|\bar{f}_{R0}|$ is small so that the MG effect is weak, $\Delta P/P_{\rm GR}$ increases monotonically with k. On the other hand, when the MG effect is strong, the fractional difference of matter power spectra no longer monotonically increases with k, but goes down at small scales after reaching some peak value at $k \sim 1\,h\,{\rm Mpc}^{-1}$ (although on even smaller scales the $P_{\rm mm}(k)$ enhancement increases again for some models, we only focus on the scales $k \lesssim 3\,h\,{\rm Mpc}^{-1}$ given the fixed simulation resolution, cf. Sec. 2.5.4). This behaviour can be explained in the context of the halo model (Cooray & Sheth, 2002), which assumes that on small scales the matter power spectrum is determined mainly by the matter distribution inside dark matter haloes (the one-halo term).

- In the regime of weak MG effect, haloes are well screened inside so that particles do not feel the fifth force during most of their evolution. When the haloes become unscreened at late times, the total gravitational potential rapidly becomes $1/3$ deeper, but the particle kinetic energy requires more time to respond, so that these particles tend to fall towards the halo centre, increasing the halo density profile and therefore the one-halo contribution to $P_{\rm mm}(k)$.

- When the MG effect is strong, particles have been accelerated for a long time (both well before and after they fall into haloes, as the latter are unscreened or less screened) due to the relatively strong fifth force. This means that the accelerations and velocities of particles can be boosted by a similar fraction as the enhancement in the depth of the gravitational

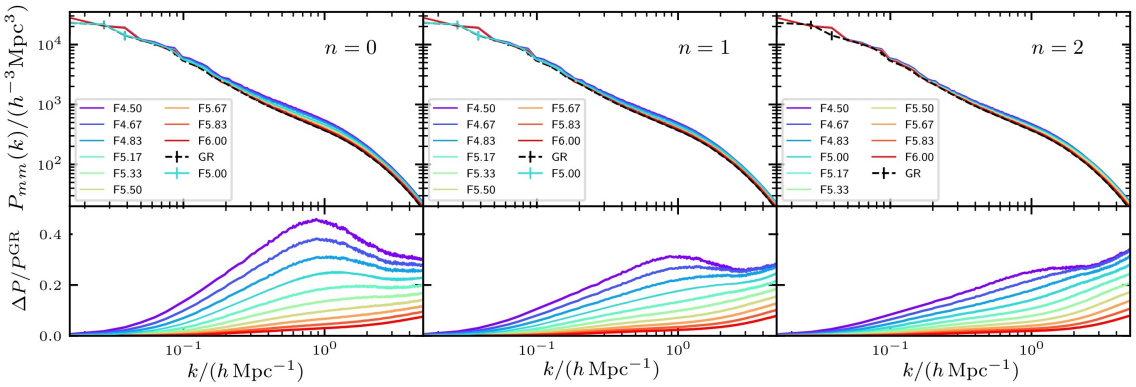

Figure 3.2: *Upper Panels:* The non-linear matter power spectra at redshift $z = 0$, from MG-GLAM simulations of the $f(R)$ model for $n = 0$ (left panel), $n = 1$ (middle) and $n = 2$ (right), each with 10 values of $|\bar{f}_{R0}|$ logarithmically spaced between 10^{-6} and $10^{-4.5}$, i.e., $-\log_{10}|\bar{f}_{R0}| = 4.50, 4.67, \ldots, 6.00$. These are indicated with different colours given in the legends. *Lower Panels:* The fractional difference, $\Delta P/P_{\mathrm{GR}}$, between the $f(R)$ and ΛCDM results, where $\Delta P \equiv P_{\mathrm{MG}} - P_{\mathrm{GR}}$. The $n = 0, 1$, $-\log_{10}|\bar{f}_{R0}| = 5.00$ and ΛCDM results are the mean of ten independent realisations, while other models only have one realisation.

potential, and hence the particle kinetic energy can be increased by a larger factor than the deepening of the potential, so that the particles are less likely to be trapped towards the centre of the potential. The small-scale structure can thus be erased out to a certain degree. This behaviour of $f(R)$ gravity has been discussed in previous works such as (Zhao et al., 2011; Li et al., 2013a; Mitchell et al., 2019). The explanation also works for other models in which screening has always been weak or absent, such as the coupled quintessence model (the left panel of Fig. 3.5) and the K-mouflage model (Hernández-Aguayo et al., 2022); in both cases we see a decay of $\Delta P/P_{\mathrm{GR}}$ at $k \gtrsim 1\,h\,\mathrm{Mpc}^{-1}$.

We note that the parameter n of the $f(R)$ model also has a considerable influence on structure formation. For fixed $\bar{f}_{R0}$, the larger the value of n, the more efficiently the fifth force is screened, as can be seen from Fig. 3.2, which shows that the matter clustering enhancement is strongest in the $n = 0$ while weakest in the $n = 2$ case. We have found similar behaviour when we checked the solution of scalar field around a top-hat overdensity in Sec. 2.5.2.3, see the middle panel Fig. 2.4: the $n = 2$ case has the strongest screening efficiency.

In MG theories, the dark matter halo populations are also affected. One of the elementary halo properties is their abundance, which we quantify using the dHMF, $\dfrac{\mathrm{d}n\,(M)}{\mathrm{d}\log_{10} M}$, which is defined as the halo number density per unit logarithmic halo mass. The dHMF result for the $f(R)$ gravity runs at $z = 0$ is shown in Fig. 3.3, where the lower sub-panels show the enhancements with respect to ΛCDM.

Firstly, we note that the abundance of haloes is enhanced due to the enhancement of total

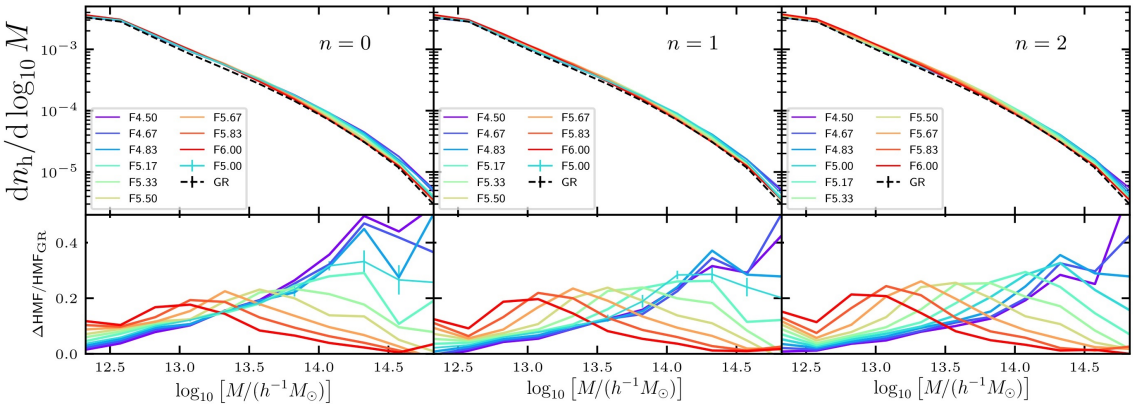

Figure 3.3: *Upper Panels:* Differential halo mass functions of $f(R)$ gravity for $n = 0$ (left panel), $n = 1$ (middle) and $n = 2$ (right), each with 10 values of $|\bar{f}_{R0}|$ logarithmically spaced between 10^{-6} and $10^{-4.5}$, i.e., $-\log_{10}|\bar{f}_{R0}| = 4.50, 4.67, \ldots, 6.00$, at redshift $z = 0$, from MG-GLAM cosmological runs. *Lower Panels:* The fractional difference $\Delta\mathrm{HMF}/\mathrm{HMF}_\mathrm{GR}$ between $f(R)$ and ΛCDM results, where $\Delta\mathrm{HMF} \equiv \mathrm{HMF}_\mathrm{MG} - \mathrm{HMF}_\mathrm{GR}$. The $n = 0, 1$, $-\log_{10}|\bar{f}_{R0}| = 5.00$ and ΛCDM results come from ten realisations (the standard deviation of which is shown as the error bars in the bottom left/central panels), while other models only have one realisation.

gravity. Secondly, for the weaker $f(R)$ models, the relative difference from ΛCDM is suppressed for massive haloes, where the fifth force is efficiently screened; going to smaller haloes the enhancement increases first, which is due to the less efficient screening and stronger MG effect for these objects; but for even smaller haloes the HMF enhancement decreases after reaching a peak, which is due to smaller haloes experiencing more mergers to form larger haloes. Apparently, this trend is not seen for the strong MG cases, such as F4.50 (purple) and F4.67 (dark blue), where the HMF enhancement seems to increase monotonically with halo mass. However, we speculate that the qualitative behaviour for the weaker $f(R)$ models should also hold even in these cases: note that our halo catalogues have been cut off for $M_\mathrm{vir} \gtrsim 10^{14.7}\, h^{-1} M_\odot$ due to the relatively small box size; should the simulations be run with larger box sizes (while keeping the same resolution), we expect the HMF enhancement to decay to zero for large enough haloes even in the strong MG cases. Finally, we note that the dHMFs are less sensitive to the model parameter n than to $\bar{f}_{R0}$, compared to the matter power spectra. The shapes and amplitude of dHMFs are similar for $n = 0, 1$ and 2, though we can still see that they are enhanced slightly more in the case of $n = 0$ than the cases of $n = 1$ and 2, for F4.50 and F4.67.

3.1.2 Symmetrons and coupled quintessence

We now present the measured matter power spectra and halo mass functions from our symmetron and coupled quintessence runs in Figs. 3.4 and 3.5, respectively.

Fig. 3.4 presents the symmetron model results with $a_* = 0.33$, $\beta_* = 1.0$ and five $c\xi/H_0$

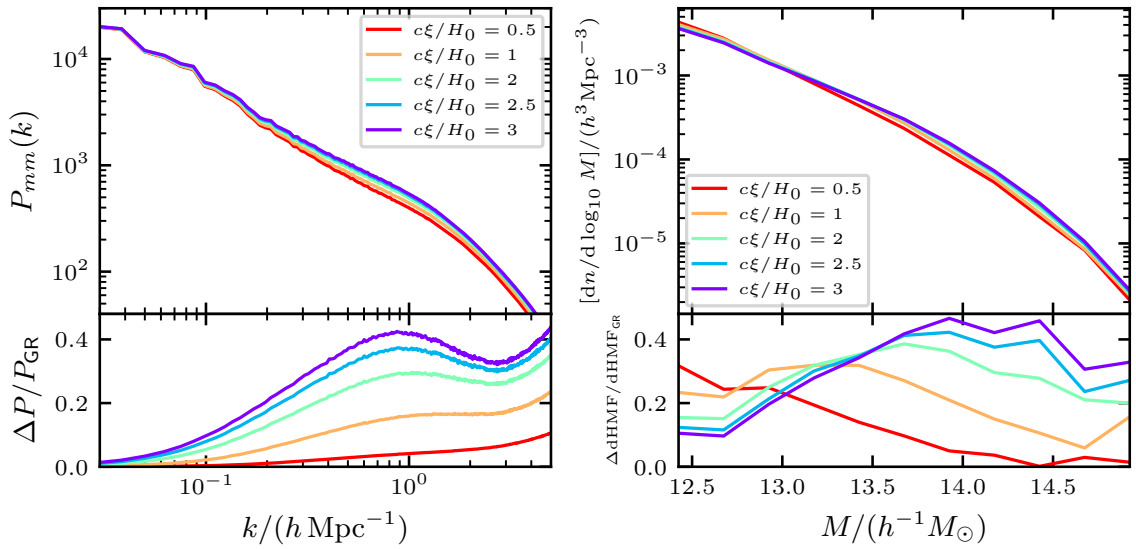

Figure 3.4: The matter power spectra (left panel) and differential halo mass functions (right) of the symmetron models at $z = 0$ for 5 different values of $c\xi/H_0$ as labelled. In all cases the remaining symmetron parameters are fixed as $a_* = 0.33$ and $\beta_* = 1.0$. As in Figs. 3.2 and 3.3, the upper sub-panels present the absolute measurements from simulations, while the lower sub-panels show the relative differences from ΛCDM.

values of $0.5, 1.0, 2.0, 2.5$ and 3.0. The behaviour of the symmetron model is qualitatively similar to that of the $f(R)$ model since both of them are thin-shell screened models (Brax et al., 2012b). This agrees with the middle panel of Fig. 2.4, which shows that these two models have qualitatively very similar scalar field profiles for a given spherical top-hat overdensity. A smaller value of $c\xi/H_0$ means m_*, the 'mass' of the symmetron field, is larger, which subsequently implies that the field can more easily settle to the potential minimum (which corresponds to $\varphi = 0$) in dense regions, and therefore be screened.

In Fig. 3.5 we show the $P_{\rm mm}(k)$ and dHMF from our three coupled quintessence models with $(\alpha, \beta) = (0.1, -0.1), (0.1, -0.2)$ and $(0.5, -0.2)$. The power spectrum enhancement remains approximately constant at $k \lesssim 0.1\,h\,{\rm Mpc}^{-1}$, which is the linear perturbation regime. This is different from the behaviour seen in the $f(R)$ and symmetron models above, where $\Delta P/P_{\rm GR}$ increases with k in this range, and the difference is because in coupled quintessence there is no screening, so that the fifth force is long ranged, with a ratio to the strength of Newtonian gravity that is almost constant in space. At small scales, $k \gtrsim 1\,h\,{\rm Mpc}^{-1}$, $\Delta P/P_{\rm GR}$ decays with k, as we found in the stronger $f(R)$ models in Fig. 3.2, and the physical explanation is the same as the case of the stronger $f(R)$ gravity model: different from the weaker $f(R)$ models, even inside dark matter haloes the particles still feel a strong fifth force, which is almost in constant proportion to the strength of Newtonian force, and on top of this the direction-dependent force can also speed up the particles; the result of the two forces is that the particles gain higher kinetic energy and tend to

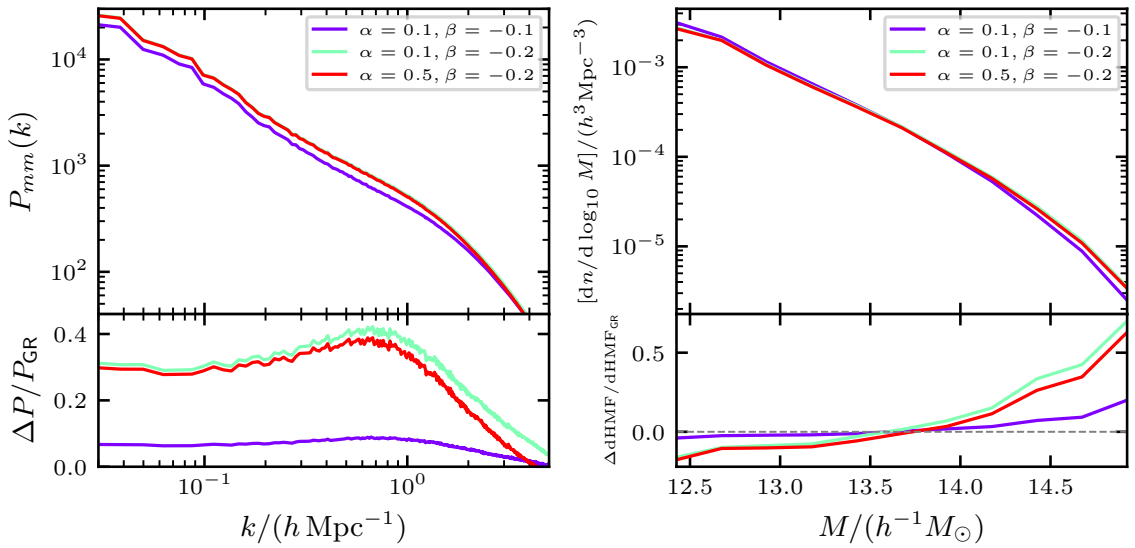

Figure 3.5: The matter power spectra (left panel) and the differential halo mass functions (right) of the coupled quintessence models at $z = 0$, for three different α and β values as labelled. As in Figs. 3.2, 3.3 and 3.4, the upper sub-panels present the absolute measurements from MG-GLAM simulations, while the lower sub-panels show the relative differences from ΛCDM.

move to and stay in the outer region of haloes, thereby reducing matter clustering on small scales compared to ΛCDM.

The right panel of Fig. 3.5 presents the dHMF results. We find that the coupled quintessence models studied here produce more high-mass haloes and fewer low-mass haloes than GR, which is the consequence of the competition between the four effects discussed in Section 2.5.1. Because these effects strongly entangle with each other through the complicated structure formation process, it is difficult to know quantitatively how they lead to the observed behaviour above, except by running simulations with different combinations of them switched on or off. Although this is obviously an interesting and important thing to do, it is beyond the scope of this paper, so we will leave it to future works.

3.1.3 Summary for the simulations

In this section we have had an initial taste of the MG-GLAM code, by running a large suite of simulations covering all three classes of models studied in this paper.

One particularly relevant aspect of the MG-GLAM code is its fast speed. The 40 $f(R)$ simulations described in this section have been run using 56 threads with OpenMP parallelisation, and we find that the run times vary randomly between $\simeq 17,000$ and $\simeq 33,000$ seconds, apparently depending on the real-time performance of the computer nodes used. The majority of them took $\sim 24,000$ seconds, or equivalently $\simeq 375$ core hours. This is roughly 100 times faster than MG-

Arepo, and 300 times faster than ECOSMOG, for the same simulation specifications. With such a high efficiency, we can easily ramp up the simulation programme to include many more models and parameter choices, and increase the size and/or resolution of the runs, e.g., using box size of at least $1\,h^{-1}\,\mathrm{Gpc}$. For the symmetron and coupled quintessence runs we have found similar speeds, though the run time for coupled quintessence models can perhaps be dramatically reduced if we do not explicitly solve the scalar field and the fifth force, by instead assume that the latter is proportional to the Newtonian force. We have also run a few even larger simulations for ΛCDM, F5n0 and F5n1 with $L = 512\,h^{-1}\mathrm{Mpc}$, $N_\mathrm{p} = 2048$ and $N_\mathrm{g} = 4096$ (for the same cosmology as above), and some of the results are presented in Appendix A — these runs took around $42,000$ seconds for ΛCDM, $80,000$ seconds for F5n0 and $125,000$ seconds (wall-clock time) for F5n1 with 128 threads using the SKUN6@IAA supercomputer at the IAA-CSIC in Spain, suggesting that a single run of specification L1000Np2048Ng4096, which would be useful for cosmological (e.g., galaxy clustering and galaxy clusters) analyses should take at most 1–1.5 days to complete and is therefore easily affordable with existing computing resources.

On the other hand, efficiency should not be achieved at the cost of a significant loss of accuracy. For the runs used here, we have used a mesh resolution of $0.25\,h^{-1}\,\mathrm{Mpc}$, which is sufficient to achieve percent-level accuracy of the matter power spectrum at $k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$ (Klypin & Prada, 2018), matter power spectrum enhancement at $k \lesssim 3\,h\,\mathrm{Mpc}^{-1}$, and (main) halo mass function down to $\sim\ 10^{12.5}\,h^{-1}M_\odot$. The particle number, N_p^3, in GLAM simulations is normally set according to $N_\mathrm{p} = N_\mathrm{g}/2$, so that in the simulations here we have used 1024^3 particles. However, we have checked that increasing the particle number to 2048^3 has little impact on the halo mass function. We notice that the completeness level of the HMFs here is similar to ECOSMOG runs with the same simulation specifications, suggesting that MG-GLAM is capable of striking an optimal balance between cost and accuracy. In Appendix A, we present more detailed tests of MG-GLAM's power spectrum and HMF predictions at different force and mesh resolutions, including our highest-resolution runs for ΛCDM and F5n1 with $L = 512\,h^{-1}\mathrm{Mpc}$, $N_\mathrm{p} = 2048$ and $N_\mathrm{g} = 4096$ (for the same cosmology as above). There we demonstrate that the increase of force resolution can lead to further improvement of the small-scale and low-mass predictions.

Before concluding this section, let us briefly describe some tests we have performed to understand how well the parallelisation of MG-GLAM works. This consists of a series of runs (taking $f(R)$ gravity F5n1 as a representative) to demonstrate the scaling of MG-GLAM, and these runs were all done on the SKUN6 facility managed by the IAA-CSIC in Spain. The strong scaling tests are presented in the left panel of Fig. 3.6. The test simulations employed a fixed resolution

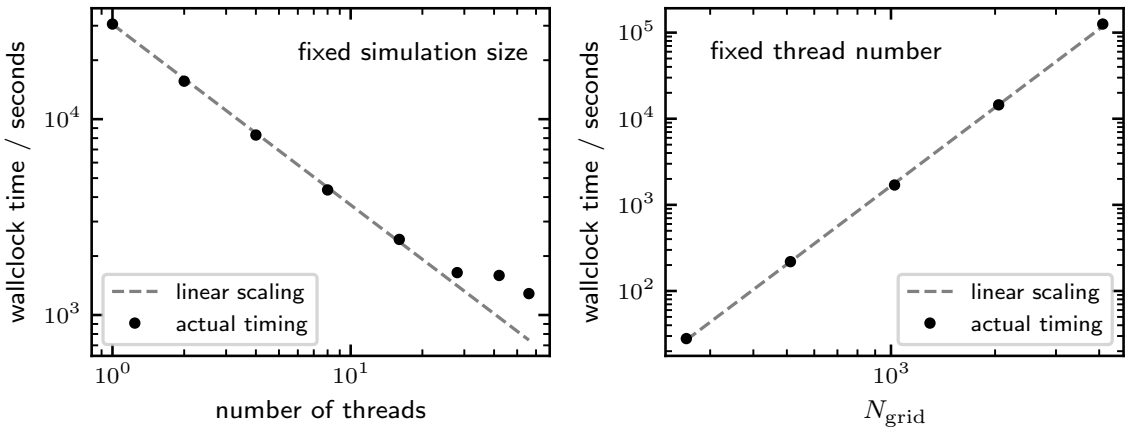

Figure 3.6: *Left panel:* The wall-clock time of the MG-GLAM test runs for the F5n1 model, with fixed simulation size/resolution ($L = 512h^{-1}$Mpc, $N_{\rm p} = 256$ and $N_{\rm g} = 512$), as a function of the number of threads used in OPENMP parallelisation. The scaling between run time and thread number is very close to be perfectly linear for number of threads up to ~ 30. *Right panel:* The wall-clock times of the MG-GLAM runs for F5n1 with varying simulation sizes and resolutions (from left to right: $N_{\rm g} = 256, 512, 1024, 2048, 4096$, and $N_{\rm p} = N_{\rm g}/2$), while the number of threads is chosen as 128. Again, the scaling is nearly perfectly linear. In both cases, the symbols denote the times taken by the test runs, and the lines denote the expected results with a 'perfect linear scaling'. We find similarly good scaling properties for ΛCDM and F5n0, but those are not shown here.

of 256^3 particles and 512^3 grids, with the same Planck 2015 cosmology as used in the main MG-GLAM runs of this chapter. This plot shows that, when the number of OpenMP threads ranges between 1 and ~ 30, the wall-clock time scales linearly with the thread number. The deviation from a perfect linear scaling (black dashed line) when the number of threads exceeds 30 is possibly due to the small size of the test run. In addition, we have also run a set of simulations of different sizes by varying the resolutions and keeping the number of threads fixed. The wall-clock time used is shown in the right panel of Fig. 3.6. We see that the time consumption again scales nearly perfectly linearly with the considered resolutions (up to $N_{\rm g} = 4096$ and $N_{\rm p} = N_{\rm g}/2$). These tests demonstrate that MG-GLAM is well scalable.

3.2 The streaming model of redshift space distortions

In this section, we focus on the mapping between real-space and redshift-space two-point statistics. The redshift-space halo correlation function depends on the real-space correlation function and the halo pairwise velocity PDF, as described by the so-called streaming model. We show that this PDF can be accurately described by an skewed student-T (ST) distribution in different MG scenarios. The ST distribution has four free parameters, which can be related to the first four moments of the pairwise velocity PDF. It can therefore match the skewness and kurtosis of the halo pairwise

velocity PDF predicted by N-body simulations. Finally, we will show that modelling skewness and kurtosis is relevant for constraining MG models through small-scale redshift space clustering measurements.

3.2.1 Redshift-space distortions

The real-space TPCF is defined as

$$\xi^{\mathrm{R}}(r) = \langle \delta(\boldsymbol{x})\, \delta(\boldsymbol{x}+\boldsymbol{r}) \rangle \,, \tag{3.1}$$

where $\delta(\boldsymbol{x})$ the number density contrast of the tracer field under investigation at position $\boldsymbol{x}$, and $\langle \cdots \rangle$ denotes ensemble averaging. This quantity only depends on the length, r, of the pair separation vector, $\boldsymbol{r}$, due to the assumed statistical isotropy and homogeneity of clustering in real space. $\xi^{\mathrm{R}}(r)$ describes the excess probability of finding a pair of tracers with separation r, compared with a random distribution of points.

In redshift space, the statistical isotropy is broken since there is a 'special' line-of-sight direction, the velocity component of which could induce additional redshifts or blueshifts, causing distortions to the tracer correlation function Equation (3.1). This is known as RSD. In this space (denoted by a superscript $^{\mathrm{S}}$), the correlation function, which is similarly defined as,

$$\xi^{\mathrm{S}}(s,\mu) = \langle \delta^{\mathrm{S}}(\boldsymbol{x})\, \delta^{\mathrm{S}}(\boldsymbol{x}+\boldsymbol{s}) \rangle \,, \tag{3.2}$$

depends not only on the pair separation $s = |\boldsymbol{s}|$, but also on the angle of $\boldsymbol{s}$ with respect to the line-of-sight direction $\hat{\boldsymbol{z}}$, characterised by the cosine $\mu \equiv \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{z}}$. This dependence can also be expressed by the separations perpendicular ($s_\perp$) and parallel ($s_\parallel$) to the line of sight, i.e., $\xi^{\mathrm{S}}(s_\perp, s_\parallel)$, where $s = \sqrt{s_\parallel^2 + s_\perp^2}$ and $s_\parallel = s\mu$, and we will use both notations. It is convenient to decompose the 2D anisotropic correlation function $\xi^{\mathrm{S}}(s,\mu)$ into multipole moments in a basis of Legendre polynomials, $L_\ell(\mu)$, as

$$\xi^{\mathrm{S}}(s,\mu) = \sum_\ell \xi_\ell(s)\, L_\ell(\mu) \,, \tag{3.3}$$

where ℓ is the order of the multipole. Odd ℓ moments vanish since $\xi^{\mathrm{S}}(s,\mu)$ is symmetric in μ. We will focus on the first three non-vanishing multipoles, i.e., the monopole ($\ell = 0$), the quadrupole ($\ell = 2$) and the hexadecapole ($\ell = 4$). We measure $\xi^{\mathrm{S}}(s,\mu)$ in the separation range $4 \leq s/(h^{-1}\mathrm{Mpc}) \leq 30$ for 26 linearly spaced bins of size $1\,h^{-1}\mathrm{Mpc}$, and in the angular cosine range $0 \leq \mu \leq 1$ for 240 equally spaced linear bins. We have explicitly checked that these choices lead to converged result at sub-percent level.

We use the publicly available PYTHON package HALOTOOLS (Hearin et al., 2017) to measure real- and redshift-space correlation functions of halo catalogues from the simulations. In cases where the simulation box is large enough, or where there are many independent realisations, we adopt the plane-parallel approximation, assuming that the line-of-sight direction $\hat{\boldsymbol{z}}$ is along one of the three axes of the Cartesian coordinate system for all haloes. The systematic deviations caused by this assumption have been shown to be small for the current surveys (Samushia et al., 2012; Yoo & Seljak, 2015). Under this approximation, the relation between the real ($\boldsymbol{r}$) and redshift ($\boldsymbol{s}$) space positions of a halo is given by

$$\boldsymbol{s} = \boldsymbol{r} + \frac{\boldsymbol{v}(\boldsymbol{r}) \cdot \hat{\boldsymbol{z}}}{a\,H(a)} \hat{\boldsymbol{z}} \,, \tag{3.4}$$

where $\boldsymbol{v}(\boldsymbol{r})$ is the peculiar velocity of the halo and $H(a)$ is the Hubble parameter at a given scale factor a.

3.2.2 The streaming model

The streaming model of RSD, introduced by Peebles (1980) and subsequently generalised by Fisher (1995), is a probabilistic approach to relate the clustering statistics of tracers in real and redshift spaces. The full complexities of modelling the redshift-space TPCF, $\xi^{\rm S}(s_\perp, s_\parallel)$, are encoded in the pairwise velocity PDF, which is the probability distribution of the relative velocities in a pair of tracers (haloes in our case) at a given halo separation; this is explicitly defined as $\boldsymbol{v}_{12} \equiv \boldsymbol{v}_2 - \boldsymbol{v}_1$, where $\boldsymbol{v}_1, \boldsymbol{v}_2$ are the velocities of the two haloes in the pair.

Because the number of tracers is conserved in real and redshift space, the fractional number overdensity in the two spaces is related by

$$\Big[1 + \delta^{\rm S}(\boldsymbol{s})\Big]\, {\rm d}^3\boldsymbol{s} = \Big[1 + \delta^{\rm R}(\boldsymbol{r})\Big]\, {\rm d}^3\boldsymbol{r} \ . \tag{3.5}$$

This equation can be further manipulated (Scoccimarro, 2004) to obtain the exact relationship between real- and redshift-space TPCFs,

$$1 + \xi^{\rm S}(s_\perp, s_\parallel) = \int_{-\infty}^{\infty} {\rm d}r_\parallel \Big[1 + \xi^{\rm R}(r)\Big] \mathcal{P}(s_\parallel - r_\parallel | \boldsymbol{r}) \,, \tag{3.6}$$

where $r_\perp \equiv s_\perp$, $\boldsymbol{r} \equiv (r_\perp, r_\parallel)$, $r \equiv \sqrt{r_\perp^2 + r_\parallel^2}$, and $\mathcal{P}(v_\parallel | \boldsymbol{r})$ is the PDF of line-of-sight relative velocities of halo pairs separated by $\boldsymbol{r}$.

The line-of-sight pairwise velocity PDF, $\mathcal{P}(v_\parallel | \boldsymbol{r})$ can be calculated from the full halo pairwise velocity distribution, $\mathcal{P}(v_r, v_t | r)$, where v_r and v_t are the pairwise velocity components parallel

and transverse to the pair separation vector $\boldsymbol{r}$, respectively. The line-of-sight projection of velocities is given by

$$v_\parallel = v_r \cos\theta + v_t \sin\theta \,, \tag{3.7}$$

where $\theta \equiv \arctan\Big(r_\perp / r_\parallel\Big)$ is the angle between the line of sight and the separation vector $\boldsymbol{r}$. Therefore,

$$\mathcal{P}(v_\parallel|\boldsymbol{r}) = \int \frac{\mathrm{d}v_r}{\sin\theta}\, \mathcal{P}\bigg(v_r, v_t = \frac{v_\parallel - v_r\cos\theta}{\sin\theta}\bigg| r\bigg) \,. \tag{3.8}$$

Note that the distribution $\mathcal{P}(v_r, v_t|r)$ only depends on the separation length r (instead of the vector $\boldsymbol{r}$) due to statistical homogeneity and isotropy in real space. It is an intrinsic property of N-body systems, which are determined by dynamical evolution under gravity.

Since we are discussing halo velocity fields, rather than the velocities of randomly chosen points in space, the moments m_{ij} and the central moments c_{ij} (where i, j are non-negative integers) of the pairwise velocity PDF, which are defined as

$$m_{ij}(r) \equiv \int \mathrm{d}v_r\, \mathrm{d}v_t\, (v_r)^i (v_t)^j \mathcal{P}(v_r, v_t|r) \,, \tag{3.9}$$

and

$$c_{ij}(r) \equiv \int \mathrm{d}v_r\, \mathrm{d}v_t\, [v_r - m_{10}(r)]^i [v_t - m_{01}(r)]^j\, \mathcal{P}(v_r, v_t|r), \tag{3.10}$$

should be weighted by halo mass when measured from simulations,

$$m_{ij}(r) = \frac{\Big\langle [1+\delta(\boldsymbol{x}_1)][1+\delta(\boldsymbol{x}_2)](v_r)^i (v_t)^j \Big\rangle}{\Big\langle [1+\delta(\boldsymbol{x}_1)][1+\delta(\boldsymbol{x}_2)] \Big\rangle} \,, \tag{3.11}$$

and

$$c_{ij}(r) = \frac{\Big\langle [1+\delta(\boldsymbol{x}_1)][1+\delta(\boldsymbol{x}_2)][v_r - m_{10}(r)]^i [v_t - m_{01}(r)]^j \Big\rangle}{\Big\langle [1+\delta(\boldsymbol{x}_1)][1+\delta(\boldsymbol{x}_2)] \Big\rangle}, \tag{3.12}$$

where $r \equiv |\boldsymbol{x}_2 - \boldsymbol{x}_1|$. Statistical isotropy in the transverse plane implies that only moments with even powers of the transverse component are non-zero. The four lowest order non-zero moments are

$$\begin{aligned} &m_{10}, \\ &c_{20}, c_{02}, \\ &c_{30}, c_{12}, \\ &c_{40}, c_{22}, c_{04} \,. \end{aligned} \tag{3.13}$$

Similarly, the line-of-sight velocity moments and central moments are defined as

$$m_n(\boldsymbol{r}) \equiv \int \mathrm{d}v_\parallel \, (v_\parallel)^n \, \mathcal{P}(v_\parallel|\boldsymbol{r}) \tag{3.14}$$

and

$$c_n(\boldsymbol{r}) \equiv \int \mathrm{d}v_\parallel \left[v_\parallel - m_1(\boldsymbol{r})\right]^n \mathcal{P}(v_\parallel|\boldsymbol{r}) \, . \tag{3.15}$$

According to Equation (3.7), the relations between the moments of the PDF $\mathcal{P}(v_\parallel|\boldsymbol{r})$ and $\mathcal{P}(v_r, v_t|r)$ are given by (Cuesta-Lazaro et al., 2020)

$$c_n(r_\perp, r_\parallel) = \sum_{k=0}^{n} \binom{n}{k} \mu^k (1-\mu^2)^{\frac{1}{2}(n-k)} c_{k,n-k}(r), \tag{3.16}$$

where $\mu \equiv r_\parallel / r$ is the direction cosine as before, c_n is the n-th central moment of the line-of-sight velocity distribution $\mathcal{P}(v_\parallel|r_\perp, r_\parallel)$, and $c_{k,n-k}$ denotes the moment of the k-th radial component, $(n-k)$-th transverse component of $\mathcal{P}(v_r, v_t|r)$. The n-th moment about the origin is denoted as m_n. In what follows, we will need the line-of-sight pairwise velocity moments $m_1(\boldsymbol{r}), c_{2\text{-}4}(\boldsymbol{r})$ for the streaming model predictions, and these will be obtained by first measuring the pairwise velocity moments $c_{ij}(r)$ from the simulated halo catalogues in real space and then performing the conversions using Equation (3.16), since the latter are an intrinsic property of halo catalogues while the former also depend on a specified line-of-sight direction.

3.2.2.1 The Gaussian streaming model

In its early applications, the streaming model was used to predict the galaxy clustering measured from the CfA survey (Davis & Peebles, 1983). In this case the best fit to the observational data was found using an exponential form for the pairwise velocity distribution. Fisher (1995) showed that the streaming model with a Gaussian velocity PDF and a scale-dependent velocity dispersion could reproduce the linear perturbation theory result for RSD on large scales. Scoccimarro (2004) demonstrated that the pairwise velocity PDF is not Gaussian, even for a Gaussian matter density field, but it can be approximated by a Gaussian near its peak. Based on a non-perturbative re-summation of the linearised limit (Fisher, 1995) of the streaming model equation (3.6), Reid & White (2011); Carlson et al. (2013) proposed that the line-of-sight pairwise velocity PDF can be approximated by a Gaussian function,

$$\mathcal{P}_{\mathrm{G}}(v_\parallel|\boldsymbol{r}) = \frac{1}{\sqrt{2\pi c_2(\boldsymbol{r})}} \exp\left[- \frac{(v_\parallel - m_1(\boldsymbol{r}))^2}{2c_2(\boldsymbol{r})} \right], \tag{3.17}$$

where we note that the Gaussian model parameters m_1 and c_2 are scale-dependent. The Gaussian streaming model (GSM) has become one of the most commonly used RSD models in galaxy surveys (e.g. Reid et al., 2012; Samushia et al., 2014; Satpathy et al., 2017; Tamone et al., 2020).

Considering the massive dark matter haloes, Kuruvilla & Porciani (2018) applied the convolutional Lagrangian perturbation theory (CLPT) formalism to compute the ingredients in the GSM, including the real-space TPCFs and the pairwise velocity moments up to the second order. They obtained predictions for the redshift-space correlation function monopole and quadrupole which are accurate to 2-4 per cent down to $\simeq 25\,h^{-1}\mathrm{Mpc}$, compared to statistics measured in N-body simulations. Bose & Koyama (2017) adopted the GSM combined with the regularised perturbation theory to compute the large-scale redshift-space halo power spectrum and TPCF, for Vainshtein screened and Chameleon screened MG models as well as GR. Also based on the GSM, Bose et al. (2020) presented a hybrid approach to predict the quasi non-linear redshift space matter power spectrum multipoles. Valogiannis et al. (2020) extended the GSM to calculate the redshift-space correlation functions for biased tracers in modified gravity models, by employing the Lagrangian perturbation theory (LPT) and CLPT resummation scheme to predict the ingredients of the GSM, including $\xi^R(r)$, $m_1(\boldsymbol{r})$ and $c_2(\boldsymbol{r})$. Their new approach qualitatively reproduces the redshift-space correlation function quadrupole in the MG simulations compared, down to at least $17\,h^{-1}\mathrm{Mpc}$, and traces the shape of the hexadecapole down to similar small scales.

Despite its simplicity and popularity, it is well known that the Gaussian model described by Equation (3.17) does not fully describe the pairwise velocity PDF, especially for pairs at small separations (Bianchi et al. (2015); Uhlemann et al. (2015); Bianchi et al. (2016); Kuruvilla & Porciani (2018), see also Fig. 3.7),because the true PDF as measured from simulations can have significant skewness and kurtosis, which are absent in a Gaussian PDF.

3.2.2.2 The Skewed Student-t (ST) distribution

Cuesta-Lazaro et al. (2020) proposed to use the so-called ST distribution (Azzalini & Capitanio, 2009) to model $\mathcal{P}(v_\parallel|\boldsymbol{r})$ (see also Zu & Weinberg, 2013, for an earlier application in a similar context). The ST distribution is constructed from the Student's t-distribution, whose PDF for a random variable x in one dimension is given by

$$t_1(x-x_c|w,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}w\Gamma(\nu/2)}\left[1+\frac{1}{\nu}\left(\frac{x-x_c}{w}\right)^2\right]^{-\frac{\nu+1}{2}}. \tag{3.18}$$

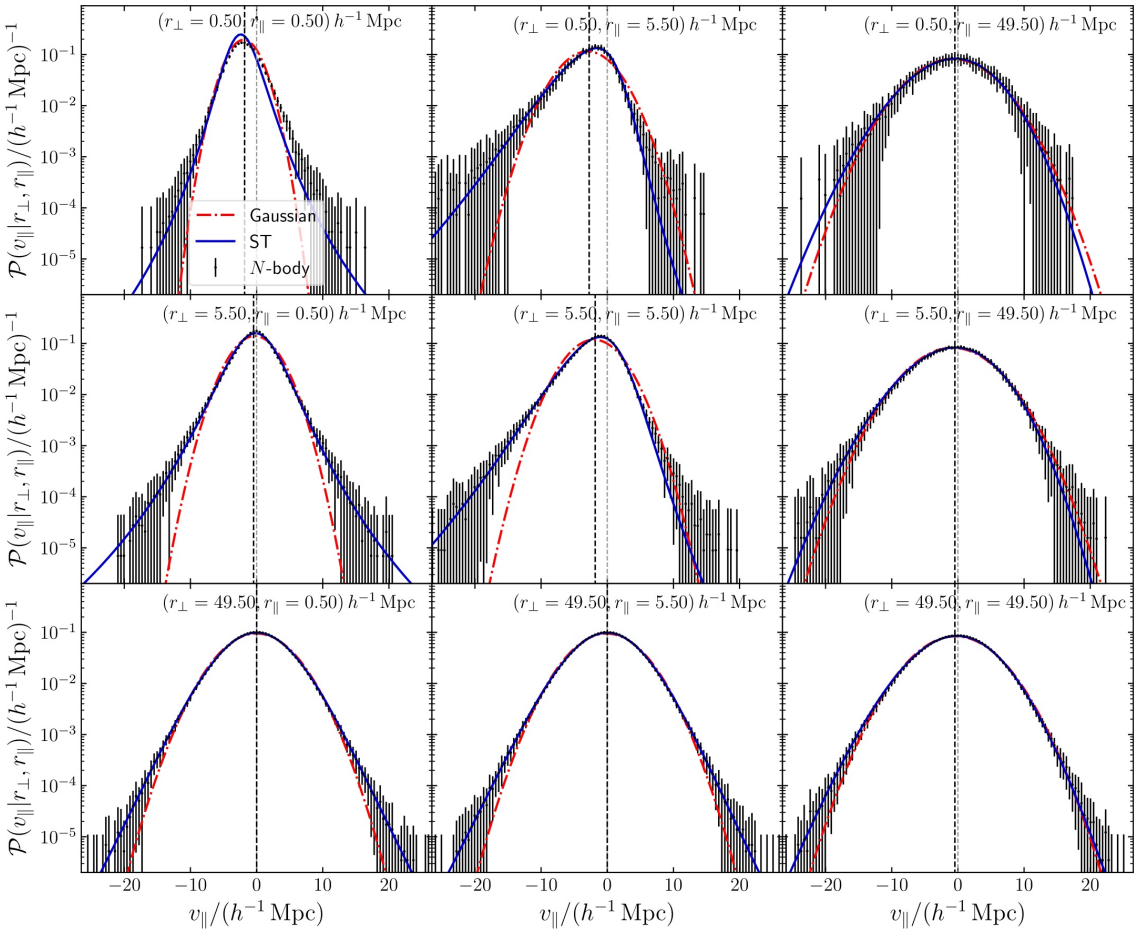

Figure 3.7: The pairwise line-of-sight velocity distribution $\mathcal{P}(v_\parallel|\boldsymbol{r})$ for dark matter haloes with number density $\bar{n}_\mathrm{h} = 10^{-3}(h^{-1}\mathrm{Mpc})^{-3}$ at $z = 0.5$, evaluated at different separations $\boldsymbol{r} = (r_\perp, r_\parallel)$, from ten MG-GLAM cosmological runs for the F5n1 model. The rows show increasing $r_\parallel$ separation, and the columns show increasing $r_\perp$. The black dots with error bars represent the mean and standard deviation from ten realisations. The red dash-dotted and blue solid lines show the Gaussian and ST models, respectively. The best-fitting parameters are obtained by converting the measured pairwise velocity moments, instead of directly fitting the black dots, as described in Section 3.2.2. The black dashed lines represent the line-of-sight pairwise velocity mean, $m_1(\boldsymbol{r})$, by integrating the measured velocity PDF, and the grey dashed lines show zero velocity value to aid visualisation. All velocities are re-scaled by $1/(aH)$ according to Equation (3.4) so that they have the unit of length.

This distribution is characterised by three parameters: the location of the centre x_c, the shape parameter w, and the number of degrees of freedom ν.

The expression for the ST distribution of line-of-sight pairwise velocities, which originates from the t-distribution (3.18), is given by

$$\mathcal{P}_\mathrm{ST}\Big(v_\parallel; v_c(\boldsymbol{r}), w(\boldsymbol{r}), \alpha(\boldsymbol{r}), \nu(\boldsymbol{r})|\boldsymbol{r}\Big) = \frac{2}{w} t_1(v_\parallel - v_c|1, \nu) \times T_1\left(\alpha \frac{v_\parallel - v_c}{w}\left[\frac{\nu+1}{\nu + ((v_\parallel - v_c)/w)^2}\right]^{1/2}; \nu+1\right), \tag{3.19}$$

where T_1 is the one-dimensional cumulative t-distribution with $\nu + 1$ degrees of freedom, and v_c, w, α and ν are the four free parameters, themselves functions of $\boldsymbol{r}$, which fully specify the ST

distribution.

Although Equation (3.19) looks quite lengthy, it has the advantage that the four parameters can be analytically related to its first four moments m_1 and $c_{2,3,4}$ (see Equations (A1)-(A6) in Appendix A of Cuesta-Lazaro et al., 2020, we have reproduced these relations in Appendix B of this paper for completeness). Furthermore, Cuesta-Lazaro et al. (2020) found that the ST distribution fits the $\mathcal{P}(v_\parallel|\boldsymbol{r})$ measured from simulations very well, in particular for close pairs, and consequently it leads to much more accurate predictions of the RSD multipoles $\xi^{\rm S}_\ell(s)$ at small scales. We will see that it also works very well for the modified gravity models described in Section 2.2.

In practice, the ST model parameters, $v_c(\boldsymbol{r}), w(\boldsymbol{r}), \alpha(\boldsymbol{r}), \nu(\boldsymbol{r})$, are determined in the following way: (1) measure the lowest four moments of the pairwise velocity distribution $\mathcal{P}(v_r, v_t|r)$ from the simulated halo catalogues, (2) convert these to the lowest four moments of the line-of-sight projected velocity PDF $\mathcal{P}(v_\parallel|\boldsymbol{r})$ using Equation (3.16), and (3) compute v_c, w, α, ν using the relations given in Appendix B. Schematically this can be illustrated as follows:

$$\begin{aligned} &\{m_{10};\ c_{20,02};\ c_{30,12};\ c_{40,22,04}\}\,(r) \\ \xrightarrow{\text{Equation (3.16)}}\ &\{m_1, c_2, c_3, c_4\}\,(\boldsymbol{r}) \\ \xrightarrow{\text{Appendix B}}\ &\{v_c, w, \alpha, \nu\}\,(\boldsymbol{r})\ . \end{aligned} \tag{3.20}$$

3.3 Results

Our aim is to demonstrate the importance of modelling higher order moments, beyond the mean and variance of the pairwise velocity distribution, to constrain modified gravity models using RSD. In this section, we show that using the ST distribution as a generic phenomenological model to convert the velocity moments into redshift-space clustering through the streaming model is also applicable to MG models, by showing that its prediction agree very well with the measurements from N-body simulations, down to scales around $5\,h^{-1}\mathrm{Mpc}$.

In Section 3.3.1 we analyse the ingredients of the streaming model (SM), in particular the halo pairwise line-of-sight velocity PDF and its moments, as measured from (MG-)GLAM simulations for a range of MG models. In Section 3.3.2 we show that the skew-T streaming model (STSM) can accurately predict both the redshift-space TPCF multipoles in different MG models, but also their relative enhancement with respect to GR, down to small scales. In Sect. 3.3.3 we perform a simple Fisher analysis to illustrate how, by including small-scale RSD information, the power of galaxy clustering analyses in constraining MG models can be significantly improved.

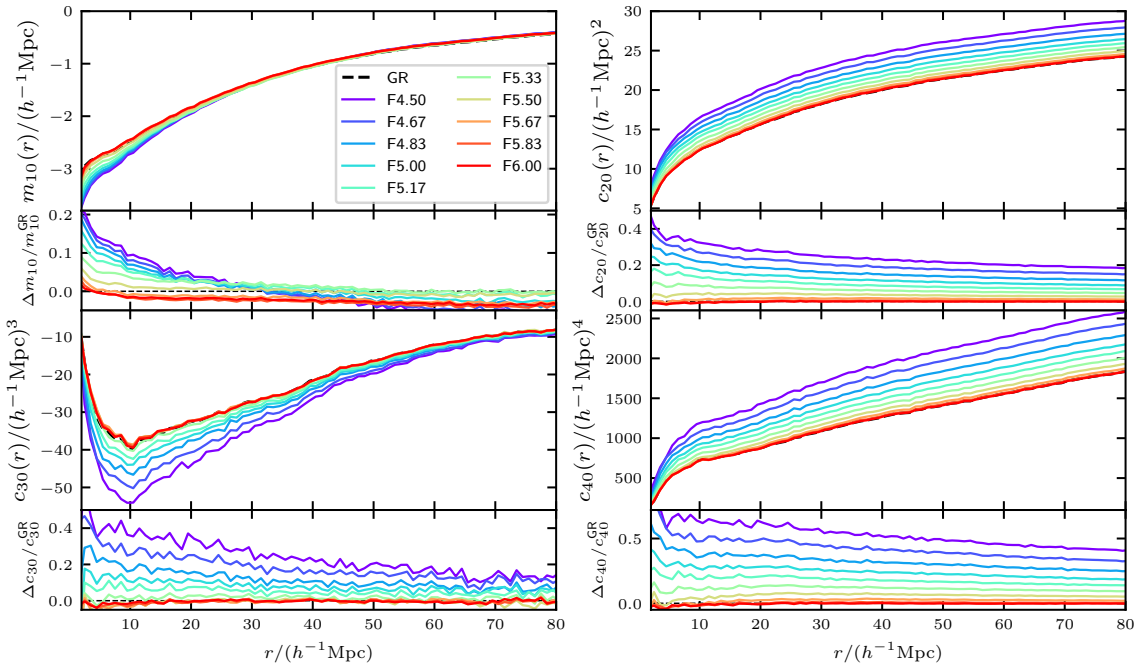

Figure 3.8: The four lowest-order moments of the radial and transverse halo pairwise velocity distribution for the $f(R)$ gravity model with $n = 1$ and 10 values of $\log_{10}|f_{R0}|$ equally spaced in the range between -6.0 and -4.5, as labelled, at $z = 0.5$, from the MG-GLAM simulations. The lower sub-panels show the relative differences between the $f(R)$ and GR models. The horizontal dashed line denotes 0. The halo catalogues have a fixed number density of $n_{\mathrm{h}} = 10^{-3}\ (h^{-1}\mathrm{Mpc})^{-3}$ for all models.

3.3.1 Streaming model ingredients

3.3.1.1 Halo line-of-sight pairwise velocity PDFs

In Fig. 3.7, we show the line-of-sight pairwise velocity PDF of dark matter haloes from the F5n1 simulations run using MG-GLAM, for nine selected combinations of $(r_\perp, r_\parallel)$ covering large, intermediate and small scales. The figure shows increasing $r_\perp$ values from top to bottom and increasing $r_\parallel$ values from left to right. The black dots represent the measured PDFs of dark matter haloes, and the lines show the Gaussian (red) and ST (blue) distributions. The best-fitting model parameters are obtained by converting the measured pairwise velocity moments, as described in Section 3.2.2 and specifically in Equation (3.20) for the ST model. Comparing this plot with Fig. 2 of Cuesta-Lazaro et al. (2020), which shows measurements from the standard gravity simulation suite Dark Quest (Nishimichi et al., 2019b), we see that the same conclusions can be reached regarding the relative performance of the Gaussian and ST models, even though several aspects of the analyses are different, such as: the gravity models ($f(R)$ gravity versus ΛCDM), the N-body codes used, the halo finders applied and the mass definitions.

In Section 3.2 of Cuesta-Lazaro et al. (2020) there is a comprehensive discussion on how the

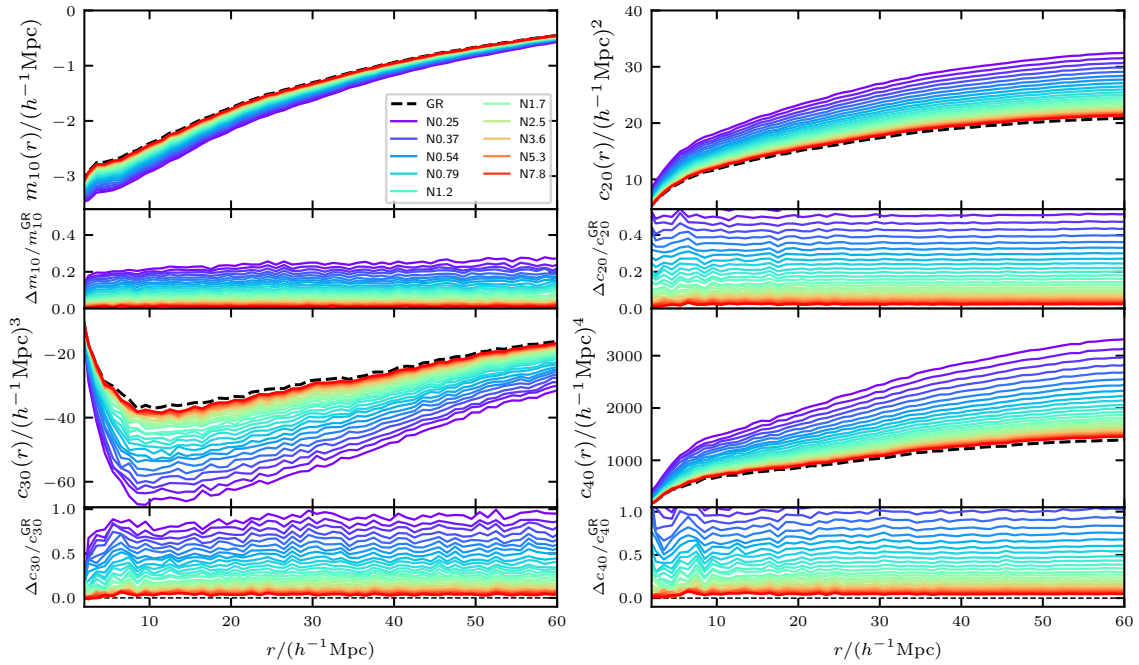

Figure 3.9: The same as Fig.3.8, but for the DGP model with 30 $H_0 r_c$ values in the range of $[0.25, 10]$, as labelled.

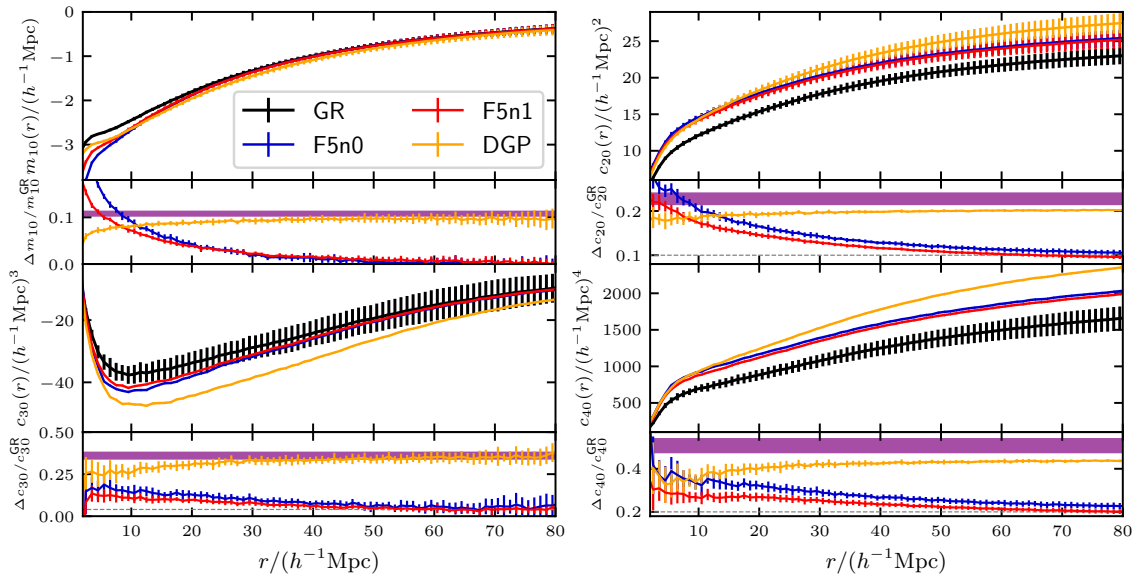

Figure 3.10: The four lowest order moments of the radial and transverse halo pairwise velocity distribution at $z = 0.5$, for the $f(R)$ model with $f_{R0} = -10^{-5}$, $n = 0$ (F5n0, blue) and 1 (F5n1, red), the DGP model with $H_0 r_c = 1$ (N1, orange) and the ΛCDM (black) model. The lower sub-panels show the relative differences between the MG and GR models. The error bars present the standard deviation of 10 realisations for each model. For the third and fourth order moments we only show the error bars of the ΛCDM results to avoid clutter. The purple bands show the theoretical prediction by Equation (3.26). This figure differs from Figs. 3.8 and 3.9 in that here the number densities of haloes have been tuned slightly so that the MG models all match the real-space halo TPCF of GR in the range $r \in [5, 30]\, h^{-1}\mathrm{Mpc}$.

behaviour of $P(v_\parallel|\boldsymbol{r})$ depends on pair separation. Here we make similar observations:

- The ST model is a better description of the simulation measurements than the Gaussian distribution on all scales we have looked at, at the expense of requiring two more parameters to quantify the skewness and kurtosis of the PDF. The improvement is significant for small separations ($r_\perp, r_\parallel \lesssim 25\,h^{-1}\mathrm{Mpc}$).
- The behaviour of $P(v_\parallel|\boldsymbol{r})$ in GR and $f(R)$ gravity is qualitatively similar, although the gravity is enhanced in the latter. The ST model achieves a similar level of agreement as in the GR case. Since the ST distribution describes the measured velocity PDF accurately, we can use the "best-fit"$*$ ST results as proxies to explore the differences in $\mathcal{P}(v_\parallel|\boldsymbol{r})$ between $f(R)$ gravity and ΛCDM.
- For small separations, such as $r_\parallel = 5.50\,h^{-1}\mathrm{Mpc}$, $r_\perp = 0.5$ or $5.50\,h^{-1}\mathrm{Mpc}$, the line-of-sight velocity distributions are strongly skewed towards negative pairwise velocities. This can be explained by the fact that such close halo pairs are more likely to be located in high-density regions where haloes infall towards each other ($v_\parallel < 0$). This skewness is less obvious when we go to large $r_\perp$ *or* $r_\parallel$ (e.g., $49.50\,h^{-1}\mathrm{Mpc}$), since for large separations the probabilities of finding infalling and receding halo pairs tend to differ less.
- The measured line-of-sight velocity PDFs are heavily tailed compared with their best-fit Gaussian ones.

3.3.1.2 Halo pairwise velocity moments

Fig. 3.8 shows the four lowest-order moments of the radial and transverse halo pairwise velocity PDFs from halo catalogues at $z = 0.5$ with fixed number density $\bar{n}_\mathrm{h} = 10^{-3}(h^{-1}\mathrm{Mpc})^{-3}$ for the $f(R)$ gravity model with $n = 1$ and ten $\log_{10}|\bar{f}_{R0}|$ values evenly spaced between -6.0 (the weakest modification) and -4.5 (the strongest), along with the relative differences with respect to the ΛCDM results displayed in the lower sub-panels. We have checked the results for $f(R)$ models with $n = 0$ and 2, and found similar results, but for clarity those are not shown here. Fig. 3.9 is the same as Fig. 3.8, but presents the DGP model with 30 $H_0 r_c$ values logarithmically spaced between 0.25 and 10. We only show one of each higher-order moment to make the plot easier to read.

The differences in the velocity moments between the MG and GR models are caused by the MG effects and the different halo populations. The latter occurs because, at fixed halo number

$*$Note that quotation marks are used here since strictly speaking this is not a fit. Instead, as described above, the ST distribution parameters have been calculated directly using the measured velocity moments.

density, the haloes from different models are likely not to have a one-to-one correspondence, even though the simulations start from the same initial conditions. For example, the contribution of the $f(R)$ gravity effect is suppressed on scales larger than the range of the fifth force. Therefore, we expect that the radial mean velocity relative enhancement, $\Delta m_{10}/m_{10}^{\rm GR}$, tends to be zero on large scales (e.g., $\gtrsim 40\,h^{-1}{\rm Mpc}$). However, due to the halo population difference, we see that this is not the case as shown in the upper left panel of Fig. 3.8: $\Delta m_{10}/m_{10}^{\rm GR}$ on large scales is scale-independent but $\bar{f}_{R0}$-dependent. We are mainly interested in the difference caused by MG effects; however, Figs. 3.8 and 3.9 only provide an incomplete picture of how the velocity moments depend on the MG model parameters.

We can isolate the MG effects, at least on large scales, on halo pairwise velocity moments and halo clustering by tuning the mass cut of MG halo catalogues, so that the real-space correlation functions $\xi^R(r)$ agree with the GR ones on large scales. Due to the small simulation box size, the correlation functions are noisy and the tuning is not reliable for a single realisation. We only tuned the MG models for which we have run ten or more realisations, i.e. GR, F5n0, F5n1 and N1. The resulting real-space correlation functions are presented in the right panel of Fig. 3.1. In the rest of this paper, we will always use these tuned MG halo catalogues for the halo clustering analysis unless otherwise stated.

Fig. 3.10 shows the same halo velocity moments measurements as in Figs. 3.8 and 3.9, but for the matched halo catalogues and models only. We can see that after matching the large scale halo correlation functions, the mean velocity relative difference of $f(R)$ gravity with respect to ΛCDM is consistent with zero on scales $\gtrsim 40\,h^{-1}{\rm Mpc}$, for both F5n1 and F5n0. The different values of n only affect the small-scale mean velocities, with the boost in $n = 0$ being larger as the fifth force is less screened in this case (cf. Sect. 5.1 of Ruan et al., 2022). For higher order moments, the relative differences on large scales converge toward non-zero constants, whose values are slightly different for $n = 0$ and 1. For the DGP model (N1), the behaviour of the velocity moments is qualitatively different from that of $f(R)$ gravity on all scales. We will interpret this result in the context of MG effects next.

The DGP and $f(R)$ gravity models feature different screening mechanisms (Vainshtein vs. thin-shell chameleon screening). In the DGP case, the fifth force is screened close to and inside massive bodies, but is unscreened and proportional to the Newtonian force when placed at a far distance. This means that structure formation is enhanced on large scales here, which is in contrast to $f(R)$ gravity, where the finite range of the fifth force means that structure formation is enhanced only

below the Compton wavelength of the scalaron field. As a result, unlike in $f(R)$ gravity, the large-scale value of the radial mean velocity enhancement in DGP is non-zero. We find that this scale-independent value agrees well with the linear perturbation theory prediction of the first pairwise velocity moment, m_{10}, which is related to the halo TPCF as (see e.g., Sheth et al., 2001b)

$$m_{10}(r) = -\frac{2}{3}\beta\frac{r\,\bar{\xi}^{\rm R}(r)}{1+\xi^{\rm R}(r)}\,, \tag{3.21}$$

where $\beta \equiv f/b_1$, $f(z)$ is the linear growth rate, b_1 is the linear halo bias and $\bar{\xi}^{\rm R}(r)$ is the volume-averaged halo correlation function

$$\bar{\xi}^{\rm R}(r) \equiv \frac{3}{4\pi r^3}\int_0^r 4\pi\xi^{\rm R}(r')r'^2\,{\rm d}r'\,. \tag{3.22}$$

Accordingly, the relative difference on large scales is approximately given by

$$\frac{m_{10}^{\rm N1}}{m_{10}^{\rm GR}} - 1 \approx \frac{\beta^{\rm N1}}{\beta^{\rm GR}} - 1\,. \tag{3.23}$$

Note that we have used the fact that the large-scale real-space halo TPCF of N1 has been tuned to match the ΛCDM one. The values of $f(z)$ at $z = 0.5$ and the linear bias b_1 are respectively calculated and measured* as

$$f^{\rm GR}(z=0.5) = 0.761, \quad b_1^{\rm GR} = 1.602 \pm 0.007\,, \tag{3.24}$$

$$f^{\rm N1}(z=0.5) = 0.804, \quad b_1^{\rm N1} = 1.527 \pm 0.006\,, \tag{3.25}$$

which give $m_{10}^{\rm N1}/m_{10}^{\rm GR} - 1 = 0.108$. This value (the purple shade region in the lower sub-panel of the upper left panel in Fig. 3.10) agrees well with what we find in the simulation data (orange line and data points).

In fact, we can qualitatively explain the behaviour of pairwise velocity moments enhancement for $\xi^{\rm R}$-tuned MG halo catalogues, for all the 4 velocity moments shown in Fig. 3.10, as follows.

The pairwise velocity moments can broadly be thought of as having two contributions: the bulk flow of haloes, which mainly contributes on large scales, and the random motion caused by small-scale shell crossing and virialisation inside dark matter haloes. The pairwise velocity moments can be approximated by the bulk flow and random motion terms, assuming that they are independent of each other so that the cross correlation between them can be ignored. We note that, while the random motions occur on small scales, their contribution still affects the even-order pairwise velocity moments for pairs of haloes at large separations — this is because random

*To find the linear halo bias b_1, we have measured the halo and matter auto-power spectra, taken their ratio and calculated the square root.

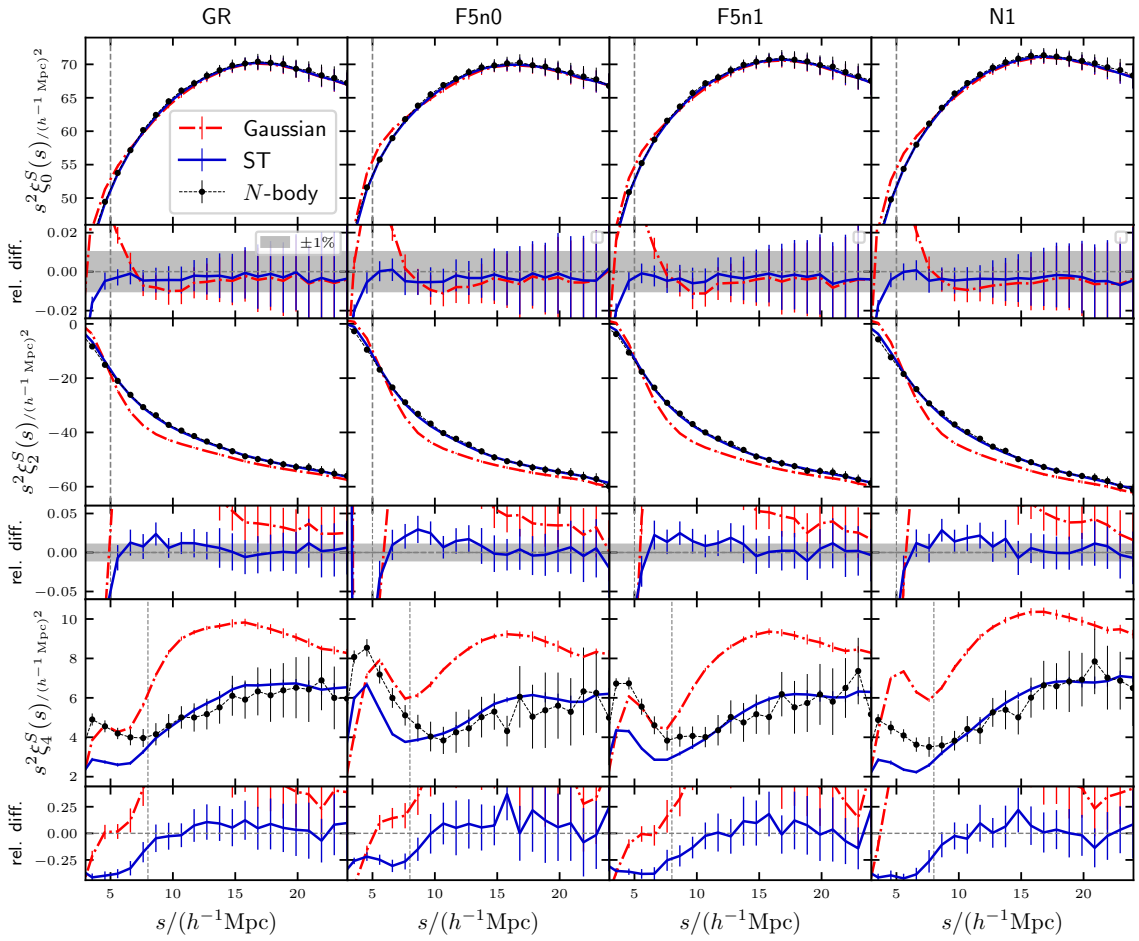

Figure 3.11: The monopole, quadrupole and hexadecapole of the redshift-space two-point correlation functions for GR (left column), the DGP model with $H_0 r_c = 1$ (N1; right column), and $f(R)$ gravity model with $\bar{f}_{R0} = -10^{-5}$ and $n = 0$ (F5n0; second column) and 1 (F5n1; third column) at $z = 0.5$, from the (MG-)GLAM simulations (black dots). Also shown are the Gaussian (dash-dotted red lines) and ST (solid blue lines) streaming model predictions, where the ingredients of the model are measured from simulations. In the lower sub-panels the relative differences between the SM predictions and the simulation measurements, $\xi^{\mathrm{model}}(s)/\xi^{\mathrm{sim}}(s) - 1$, are shown. The horizontal dashed lines in the lower sub-panels denote 0, the grey shaded regions shows $\pm$ 1% for the monopole and quadrupole. The vertical dashed lines indicate where the STSM predictions start to differ significantly from simulation measurements.

motions do contribute to the velocity difference of the pair, v in Equation (3.11), and when taking even powers of v there can be no cancellation and this contribution stays in the final m_{ij}.

As mentioned above, the effect of the fifth force in $f(R)$ gravity is suppressed on large scales which are well beyond the range of the force (the inverse of the scalaron Compton wavelength), whereas on small scales we would expect to observe some effect (except where chameleon screening works efficiently to suppress it). On the other hand, in DGP models, gravity is enhanced by a constant factor on large scales, but is very efficiently suppressed within a few times the typical halo virial radius (see, e.g., Hernández-Aguayo et al., 2021). This implies that in the two MG models the contributions from the bulk flow and the random motion will behave very differently.

On small scales where the random motions are strong, the velocity moment boost in $f(R)$ gravity can be considerable, since the fifth force is only fully screened in a few very massive haloes, and is unscreened for most objects (at least for the F5n1 and F5n0 models considered here). In DGP models, the Vainshtein screening is efficient on scales smaller than the Vainshtein radius, which causes the MG enhancement to be small toward these scales. This is qualitatively consistent with the small-scale behaviour of $\Delta m_{10}/m_{10}^{\rm GR}$ observed in the upper left sub-panel of Fig. 3.10.

Linear theory has explained the large-scale behaviour of the first order moment enhancement in both gravity models. Similarly, the leading (linear) term of higher order moment enhancements in perturbation theory, which describes the bulk flow, is given by (see Appendix E for a heuristic derivation):

$$\frac{c_n^{\rm MG}}{c_n^{\rm GR}} - 1 \approx \left(\frac{\beta^{\rm MG}}{\beta^{\rm GR}}\right)^n - 1, \text{ on large scales.} \tag{3.26}$$

The horizontal purple bands in Fig. 3.10 present the linear predictions for the N1 model. For odd-order pairwise velocity enhancements, the large-scale N-body measurements (orange lines) agree well with the linear theory, while for even-order moments, the measurements are systematically smaller. This behaviour can be explained by employing the bulk flow (bf) and random motion (rm) decomposition mentioned above. To be specific, we consider the second-order moment c_{20}, but the argument works for any other even-order moments. Consider the large-scale difference of c_{20} between N1 and ΛCDM and let us decompose the moments into bulk flow and random motion contributions,

$$\begin{aligned}
\Delta c_{20} &\equiv [c]_{\rm MG} - [c]_{\rm GR} && (3.27)\\
&= c_{\rm MG}^{\rm bf} + c_{\rm MG}^{\rm rm} - (c_{\rm GR}^{\rm bf} + c_{\rm GR}^{\rm rm}) && (3.28)\\
&\overset{\rm DGP}{\approx} c_{\rm MG}^{\rm bf} - c_{\rm GR}^{\rm bf}, && (3.29)
\end{aligned}$$

where in the last line we have made use of the fact that the random motion contributions in the DGP and GR models are approximately the same (i.e., $c_{\rm DGP}^{\rm rm} \approx c_{\rm GR}^{\rm rm}$) due to the screening of the fifth force. Taking the ratio with respect to the moment of ΛCDM, we have

$$\begin{aligned}
\frac{\Delta c}{c_{\rm GR}} &\approx \frac{c_{\rm MG}^{\rm bf} - c_{\rm GR}^{\rm bf}}{c_{\rm GR}^{\rm bf} + c_{\rm GR}^{\rm rm}}\\
&< \frac{c_{\rm MG}^{\rm bf} - c_{\rm GR}^{\rm bf}}{c_{\rm GR}^{\rm bf}} \approx \left(\frac{\beta_{\rm MG}}{\beta_{\rm GR}}\right)^2 - 1\,,
\end{aligned} \tag{3.30}$$

where for the inequality we have implicitly used the facts $c_{\rm DGP}^{\rm bf} > c_{\rm GR}^{\rm bf}$ and $c_{\rm GR}^{\rm rm} > 0$ (as mentioned above, for even moments, $c_{\rm GR}^{\rm rm} \neq 0$ even for halo pairs at large separations). This is the reason why the even-order moments of the simulation measurements are lower than those predicted by the linear theory, Equation (3.26).

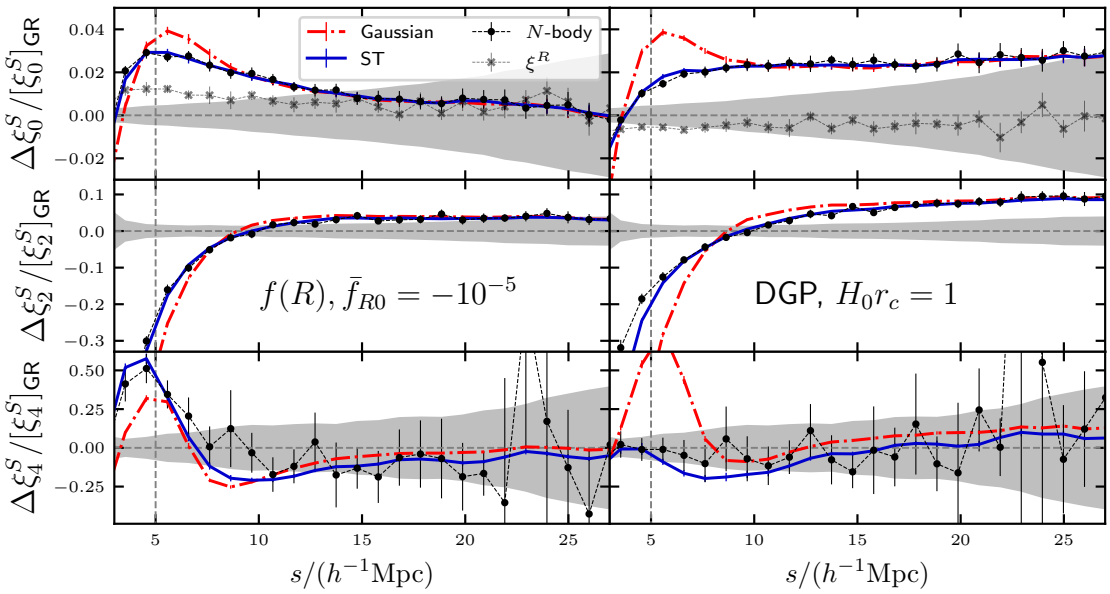

Figure 3.12: The relative difference of redshift-space two-point correlation function multipoles between F5n1 and GR (left column), and between N1 and GR (right column), from the (MG)-GLAM simulations. The black dots with error bars are the measurements from simulations, and the small crosses with error bars in the top row are the relative differences in the real-space halo TPCFs — these are measured from the tuned halo catalogues and are there just for comparison. The streaming model predictions are plotted as the red dash-dotted (GSM) and blue solid (STSM) lines. The gray-shaded areas correspond to the standard deviation for GR over 72×3 measurements obtained from 72 realisations of the GLAM code.

For the odd-order moments, Equations (3.27)-(3.29) are still valid. The difference arises in the inequality of Equation (3.30). Random motion contributions to odd-order velocity moments are significantly smaller than those of even-order moments, since even-order powers of v are always positive and therefore can have no cancellation in Equation (3.11), but odd-order powers of v in Equation (3.11) can so that $c_{\rm GR}^{\rm rm} \approx 0$. We therefore have, for odd-order moments,

$$\begin{aligned} \frac{\Delta c}{c_{\rm GR}} &\approx \frac{c_{\rm MG}^{\rm bf} - c_{\rm GR}^{\rm bf}}{c_{\rm GR}^{\rm bf} + c_{\rm GR}^{\rm rm}} \\ &\approx \frac{c_{\rm MG}^{\rm bf} - c_{\rm GR}^{\rm bf}}{c_{\rm GR}^{\rm bf}} \approx \left(\frac{\beta_{\rm MG}}{\beta_{\rm GR}}\right)^2 - 1 \,. \end{aligned} \tag{3.31}$$

3.3.1.3 Halo real space two-point correlation functions

As mentioned above, we have constructed the GR halo catalogues to have a fixed number density, and tuned the minimum mass cut $M_{\rm min}$ of the various MG halo catalogues so that the halo real-space TPCFs match their GR counterparts closely. As shown in the right panel of Fig. 3.1, the correlation functions of the tuned $f(R)$ gravity and DGP halo catalogues agree with the ΛCDM

counterparts on scales larger than $\sim 10\,h^{-1}\mathrm{Mpc}$ to within a relative difference of less than $1\simeq$ per cent.

The effect of MG on the TPCF of halo catalogues with fixed $n_{\rm h}$, or fixed minimum mass cut $M_{\rm min}$, is interesting in its own right. However, this has been investigated in various previous works, more recently by Alam et al. (2021b). On the other hand, as we have discussed above, the tuning of $n_{\rm h}$ to make it MG model dependent — in order to achieve a matching of the real-space halo TPCFs in different models — leads to catalogues where the effect of different halo populations can be more cleanly separated from that of the fifth force. In addition, since the real-space halo TPCFs are matched, any difference in the redshift-space clustering is necessarily caused by the difference in the pairwise velocities. This makes the interpretation of the underlying physics more straightforward. As a result, for the rest of this paper we will only use the tuned halo catalogues.

3.3.2 Halo redshift-space two-point correlation function multipoles

In this subsection, we will apply the ST velocity distribution to the streaming model, to predict halo redshift-space correlation function monopoles, quadrupoles and hexadecapoles. Since our goal is to show that the ST model is generic and applicable to both ΛCDM and MG models, we measure all the ingredients of the streaming model, including halo real-space correlation functions and pairwise velocity moments, from the simulations. We use the MG-GLAM simulation data for this investigation. The halo catalogues at $z=0.5$ with the halo number density around $10^{-3}\,(h^{-1}\mathrm{Mpc})^{-3}$ are used in this section. We measure the model ingredients for each realisation, compute the streaming model predictions and then present the average and standard deviation. The results with other number densities and in other redshifts are qualitatively similar, and some of these will be shown in Appendix D.

In Fig. 3.11, we show the multipoles of the redshift-space correlation function, $\xi^{\rm S}_{0,2,4}(s)$ (the different rows), for the ΛCDM, F5n0, F5n1 and N1 models (the different columns). In the lower sub-panels of each panel, the relative differences between the model predictions and the simulation measurements are displayed. The monopole predictions are quite accurate for both GSM and STSM, while it is apparent that STSM performs slightly better. If one targets at percent-level accuracy, then GSM fails at $s\lesssim 10\,h^{-1}\mathrm{Mpc}$ while STSM works well down to $s\simeq 5\,h^{-1}\mathrm{Mpc}$. The improvement made by using the ST velocity distribution is significant when we consider the quadrupole. The GSM is biased for scales $\lesssim 20\,h^{-1}\mathrm{Mpc}$, whilst the STSM only starts failing on scales smaller than $5\,h^{-1}\mathrm{Mpc}$. Similar behaviour is found for the hexadecapole. Although the

measurement of the hexadecapole is rather noisy, mainly due to the small simulation box size, the ST model agrees with simulation measurements within one standard deviation for scales larger than $\simeq 8\,h^{-1}\mathrm{Mpc}$, while the GSM is biased on all scales considered here.

The STSM matches the $\xi^{\mathrm{S}}_{0,2,4}(s)$ on scales larger than the maximum separation shown in Fig. 3.11, so to improve the readability of the figure, we opt not to show the behaviour on larger scales. We conclude that the ST pairwise velocity distribution with the streaming model is competent in predicting redshift-space correlation functions in $f(R)$ gravity and the DGP model, as well as in GR.

Note that this excellent performance of the ST model is under ideal conditions: all ingredients of the streaming model of RSD, e.g., the halo real-space two-point correlation function, $\xi^{\mathrm{R}}(r)$, and the four lowest order pairwise velocity moments, are all measured from simulations, instead of using theoretical models. We will briefly discuss our plan on constructing simulation-based emulators for $\xi^{\mathrm{R}}(r)$ and higher-order pairwise velocity moments in Section 3.4 to extend the unbiased predictions down to highly non-linear scales.

Fig. 3.12 compares the enhancements, with respect to ΛCDM, of the measured RSD monopole (upper panels), quadrupole (middle) and hexadecapoe (lower) from the simulated halo catalogues (symbols with error bars), against the predictions by the Gaussian (red dashed lines) and ST (blue solid) streaming models. The left column shows the results for F5n1 and the right panel for N1. We note that, again, for both MG models, STSM outperforms the GSM in matching the simulation data. For the monopole, GSM starts to fail at $\sim 10\,h^{-1}\mathrm{Mpc}$ while STSM works well down to $\sim 3\,h^{-1}\mathrm{Mpc}$. For the quadrupole, the GSM prediction deviates from simulation data at $s \lesssim 20\,h^{-1}\mathrm{Mpc}$, while STSM remains in good agreement with the latter down to $\sim 5\,h^{-1}\mathrm{Mpc}$. For hexadecapole, we can see an improvement in STSM as well, though here the simulation data is noisier.

3.3.3 Schematic demonstration of scale dependence of the constraint on MG parameters

We have seen that, compared with the traditional Gaussian model, the ST model has achieved greater success in predicting halo clustering on smaller scales (5-25 $h^{-1}\mathrm{Mpc}$). In order to quantitatively demonstrate the constraining power gained from small-scale RSD signals, we will forecast the constraints on the MG parameters using a highly simplified Fisher analysis, in which all parameters are fixed except the MG parameters such $\bar{f}_{R0}$ in $f(R)$ gravity and $H_0 r_c$ in DGP.

The Fisher matrix method provides a way to propagate the observable uncertainty to the constraints of cosmological parameters. Our calculation of the Fisher matrix is based on Tegmark (1997) and Seo & Eisenstein (2003), assuming a Gaussian likelihood function for our measurements of the correlation function multipoles. Additionally we ignore any parameter dependence of the covariance matrix, in which case the Fisher matrix of a redshift slice centred at z can be approximated as

$$F_{ij} = \sum_{\alpha\beta} \frac{\partial f_\alpha}{\partial p_i} \mathrm{Cov}^{-1}[f_\alpha, f_\beta] \frac{\partial f_\beta}{\partial p_j} \tag{3.32}$$

where Greek indices α, β label the spatial separation bins, e.g., s_α; $f_\alpha = \{\xi_0^{\rm S}(s_\alpha), \xi_2^{\rm S}(s_\alpha)\}$ are the redshift-space halo correlation function multipoles at redshift z; $\mathrm{Cov}[f_\alpha, f_\beta]$ is the corresponding covariance matrix and p_i, p_j are the model parameters (only one parameter in our simplified case here) being considered, which are $\{\bar{f}_{R0}\}$ in $f(R)$ gravity and $\{H_0 r_c\}$ in the DGP model. The covariance matrix of the redshift-space multipole moments are calculated from the halo catalogues of 72 GLAM ΛCDM runs. The 1σ error is given by $\sqrt{(F^{-1})_{11}}$. We fix the maximum separation $s_{\rm max} = 35\,h^{-1}{\rm Mpc}$, and vary the minimum scale $s_{\rm min}$ from 22 to $2.5\,h^{-1}{\rm Mpc}$ to explore the constraining power on MG parameters gained from small-scale information.

For the F5n1 and N1 models considered here, the derivatives in Equation (3.32) are approximated by (taking $\xi_0^{\rm S}(s)$ as an example)

$$\frac{\partial \xi_0^{\rm S}(s_\alpha; \bar{f}_{R0})}{\partial \bar{f}_{R0}} \approx \frac{\xi_0^{\rm S}(s_\alpha; -10^{-5}) - \xi_0^{\rm S}(s_\alpha; {\rm GR})}{-10^{-5}} \tag{3.33}$$

and

$$\frac{\partial \xi_0^{\rm S}(s_\alpha; \gamma)}{\partial (H_0 r_c)} \approx \frac{\xi_0^{\rm S}(s_\alpha; 1) - \xi_0^{\rm S}(s_\alpha; \gamma_{\rm GR} = 0)}{1}, \tag{3.34}$$

respectively, where $\gamma \equiv 1/(H_0 r_c)$.

The Fisher forecast results are presented in Fig. 3.13, in which we have considered three scenarios: using monopole data only (red), using quadrupole data only (blue) and using both monopole and quadrupole (black). In all cases it is clear that the constraining power on both $\bar{f}_{R0}$ and $H_0 r_c$ monotonically increases with decreasing $s_{\rm min}$. Compared with using the monopole data alone, the addition of the quadrupole data (which on its own does not produce very strong constraints) tightens the constraints by $\sim 20\%$. Including the hexadecapoles leads to little improvement, which is unsurprising given the rather noisy hexadecapole measurements. Most interestingly, we note that, for both $f(R)$ gravity and DGP, including scales of $s \lesssim 10\,h^{-1}{\rm Mpc}$ can markedly improve the constraints on the MG parameter. This confirms that small-scale RSD, if measured precisely

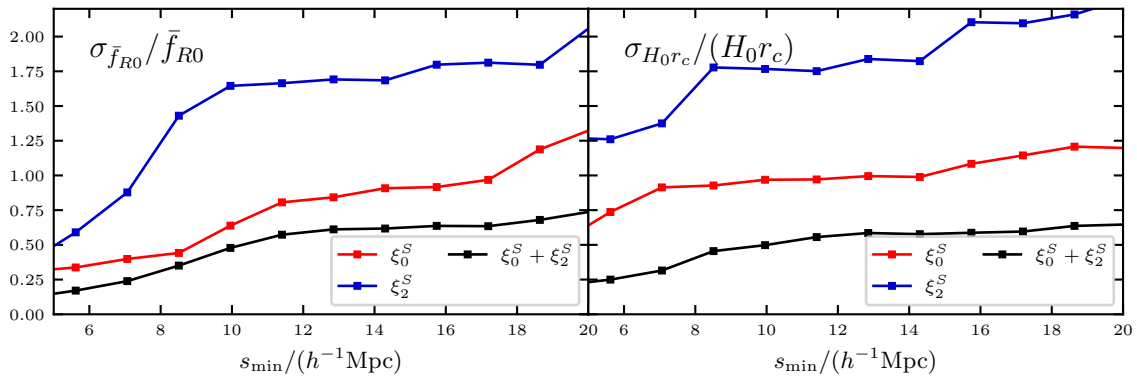

Figure 3.13: The dependence of the constraining power on $\bar{f}_{R0}$ (left) and $H_0 r_c$ (right) from halo redshift-space correlation function multipole measurements, on the minimum scale $s_{\rm min}$ included in the constraints, based on our simplified Fisher forecast with different minimum scales considered. The maximum scale is fixed to $s_{\rm max} = 35\,h^{-1}{\rm Mpc}$. Three scenarios are considered: monopole data alone (red), quadrupole data only (blue), and including both monopole and quadrupole (black).

and modelled accurately, can be a promising tool to help test gravity models using galaxy clustering data.

3.4 Conclusions and discussion

Percent-level accuracy in modelling the anisotropies of redshift-space galaxy clustering is required to accurately recover cosmological information from RSD signals in order to distinguish between dark energy and modified gravity scenarios. Within the framework of the streaming model of RSD, this requires that, compared with the current status, we must: (i) improve the mapping of real- to redshift-space clustering statistics, i.e., find a better description of the pairwise velocity distribution, (ii) increase the accuracy of the predictions of the streaming model ingredients, including the halo real-space correlation function $\xi^{\rm R}(r)$ and the pairwise velocity moments. We have investigated both aspects in the context of modified gravity cosmologies. For the first aspect, we have demonstrated that the ST probability distribution for the halo pairwise velocity, which was introduced by Cuesta-Lazaro et al. (2020) as an alternative to the traditional Gaussian model, is applicable to the measurements from the N-body simulations of all MG models considered here. For the second aspect, we have explored MG signals in both the individual ingredients and predictions of the streaming model. In this chapter, we have made full use of the cosmological simulations from MG-GLAM.

The traditional Gaussian distribution fails to fully capture some properties of the halo pairwise velocity PDF found in N-body simulations, such as the skewness and kurtosis. The ST distribution can be tuned to match the four lowest order velocity moments of $\mathcal{P}(v_\parallel|\boldsymbol{r})$ with four parameters (two

more than a Gaussian). Compared with the Gaussian form, the ST model extends the validity of the streaming model from $\simeq 7$ to $\lesssim 5\,h^{-1}\mathrm{Mpc}$ for the monopole, and $\simeq 15$ to $\simeq 5\,h^{-1}\mathrm{Mpc}$ for the quadrupole. For the hexadecapole, the ST model gives predictions that are correct down to about $8\,h^{-1}\mathrm{Mpc}$, while the GSM is biased on all scales shown. The performance of ST is equally good among all considered gravity models, including ΛCDM, $f(R)$ gravity with $\bar{f}_{R0} = -10^{-5}, n = 0$ (F5n0) and 1 (F5n1), and the normal branch of DGP with $H_0 r_c = 1$ (N1).

We have investigated the MG enhancements of halo pairwise velocity moments and redshift-space correlation functions with respect to the ΛCDM baseline. To remove the effect of different halo populations on large scales and make the physics in the results easier to interpret, we have tuned the mass cut of MG halo catalogues to match the real-space correlation functions to that of the ΛCDM halo catalogues on large scales. With this, the MG pairwise velocity moment enhancements on large scales can be explained by linear theory.

We have performed a simple Fisher forecast analysis to assess the impact of including small-scale information on the power of RSD in testing and distinguishing different gravity models. Fig. 3.13 demonstrates that both including the RSD quadrupole and including data from scales of $s \lesssim 15\,h^{-1}\mathrm{Mpc}$ can substantially decease the uncertainty in the constrained MG parameters. This highlights the potentially important role played by small-scale RSD in cosmological tests of gravity using data from upcoming galaxy surveys such as DESI and Euclid.

The analysis in this work has been largely theoretical, since we have focused on haloes and made use of direct measurements from simulations for a fixed number of theoretical models. To apply the ST model to real galaxy survey data and fully exploit its accuracy on small scales, we need to improve in a couple of aspects. First of all, we need accurate predictions of the two SM model ingredients — the pairwise velocity moments and real-space correlation functions — for arbitrary cosmological models. While on linear and quasi-linear scales, perturbation based approaches have proven very useful in this regard, since our focus here is on the small, non-linear scales, where the perturbative approach fails, alternatives need to be sought. We plan to build emulators for the pairwise velocity PDF moments and TPCFs, taking advantage of the large number of big simulations that will be enabled by the fast MG-GLAM code. Note that this is different from directly emulating the 2D redshift-space correlation functions or their multipoles, since the SM ingredients are quantities with clearer physical meanings.

Secondly, we need to extend our analysis to observable tracers of the large-scale structure, such as galaxies. The inclusion of satellite galaxies in clustering analysis will lead to a substantial

finger-of-God effect, which can pose new challenges to the accurate modelling of small-scale RSD, and this needs to be investigated. Also, since our simulations are dark matter only, a model of galaxy-halo connection, such as HOD (e.g., Berlind et al., 2003; Zheng et al., 2005a), abundance matching (e.g., Conroy et al., 2006; Moster et al., 2010; Reddick et al., 2013), and semi-analytic galaxy formation models (e.g., Kauffmann et al., 1993; Cole et al., 1994; Lacey et al., 2016), is needed, and this will likely introduce additional uncertainties in the predicted signal. For simulations with relatively low resolution, the HOD method is usually adopted to construct galaxy mock catalogues by populating the simulated haloes, where the HOD parameters can be calibrated to match the observed galaxy number density and certain properties of their large-scale correlation. In a forthcoming project, we will focus on the redshift-space TPCF multipoles for HOD galaxies, and use these to reassess the constraining power of small-scale RSD in gravity tests. Modelling the redshift-space galaxy clustering within the streaming model gives us more flexibility when combining with the HOD prescription.

Chapter 4

From haloes to galaxies: the emulator-based halo model approach

In the modern classic textbook *Modern Cosmology* (Dodelson & Schmidt, 2020), the halo model was introduced as a useful but less precise approach as

> Given its simplicity, the halo model does remarkably well, predicting the power spectrum to within 25% over a wide range of scales, with the most significant departure being an underprediction of the power spectrum on small scales (which can be improved by tuning the assumed concentration–mass relation of halos).
>
> For precision applications, this is not sufficient, but the wide range of scales that are described approximately does illustrate that the halo model can be a useful phenomenological framework.

This chapter tries to demonstrate that the numerical version halo model, in which all the model ingredients are calculated from simulation-based templates (the so-called *emulators*), could be accurate enough for modelling non-linear galaxy clustering statistics such as TPCFs.

Following the philosophy of the Dark Quest project (Nishimichi et al., 2019a; Kobayashi et al., 2020), we develop emulators for the basic properties of dark matter haloes (instead of galaxies), which can be estimated from N-body simulations robustly for fixed cosmologies. One can adopt a galaxy-halo connection model and combine it with the halo property emulators to compute the clustering statistics for a galaxy sample of interest, down to highly non-linear scales.

We begin with a discussion of the halo model formalism in Section 4.1, including a full derivation of the TPCFs of matter, halo mass, and galaxies. Section 4.2 describes in detail fully

connected neural networks — the machine learning scheme we adopt to construct emulators. Section 4.3 shows the results of the ideal emulation test, as an assess of the fitting ability of neural networks. We present the main result in Section 4.4: the "real" emulation results of the basic halo properties, including the HMF and halo density profile, in $f(R)$ gravity. In Section 4.5, we show how to combine the ingredients of the halo model with a HOD prescription to accurately reproduce the clustering of galaxies.

4.1 The halo model basics

4.1.1 The matter two-point correlation function

As mentioned in Section 1.1, the halo model assumes that the matter density field consists of a superposition of haloes at locations $\boldsymbol{x}_i$ with masses M_i, so that the matter field can be written as

$$\rho_{\rm m}(\boldsymbol{x}) = \sum_i M_i\, u\Big(|\boldsymbol{x}-\boldsymbol{x}_i|\Big|M_i\Big) \tag{4.1}$$

$$= \int \mathrm{d}^3\boldsymbol{x}' \int \mathrm{d}M \underbrace{\sum_i \delta^{(\rm D)}(M-M_i)\,\delta^{(\rm D)}(\boldsymbol{x}'-\boldsymbol{x}_i)}_{=\mathrm{d}n(\boldsymbol{x}';M)/\mathrm{d}M} M\, u\Big(|\boldsymbol{x}-\boldsymbol{x}'|\Big|M\Big), \tag{4.2}$$

$$= \int \mathrm{d}^3\boldsymbol{x}' \int \mathrm{d}M\, \frac{\mathrm{d}n(\boldsymbol{x}';M)}{\mathrm{d}M} M\, u\Big(|\boldsymbol{x}-\boldsymbol{x}'|\Big|M\Big), \tag{4.3}$$

where

- The summation is for all haloes; the same results can be derived by taking the summation over all microcells which are made to be so small that each cell contains no more than one halo centre (e.g. Peebles, 1980);
- $\delta^{(\rm D)}(\cdots)$ is the Dirac delta function;
- $\dfrac{\mathrm{d}n(\boldsymbol{x}';M)}{\mathrm{d}M} \equiv \sum_i \delta^{(\rm D)}(M-M_i)\,\delta^{(\rm D)}(\boldsymbol{x}'-\boldsymbol{x}_i)$ is the *local* HMF, whose integral over the halo mass gives the halo number density field at the field point $\boldsymbol{x}'$:

$$\int \mathrm{d}M\, \frac{\mathrm{d}n(\boldsymbol{x}';M)}{\mathrm{d}M} = n_{\rm h}(\boldsymbol{x}';M), \tag{4.4}$$

and its ensemble average gives the common HMF,

$$\left\langle \frac{\mathrm{d}n(\boldsymbol{x}';M)}{\mathrm{d}M} \right\rangle = \left\langle \sum_i \delta^{(\rm D)}(M-M_i)\,\delta^{(\rm D)}(\boldsymbol{x}'-\boldsymbol{x}_i) \right\rangle = \frac{\mathrm{d}n(M)}{\mathrm{d}M}; \tag{4.5}$$

- $u\Big(|\boldsymbol{x}-\boldsymbol{x}_i|\Big|M_i\Big) \equiv \dfrac{\rho_{\rm h}\Big(|\boldsymbol{x}-\boldsymbol{x}_i|\Big|M_i\Big)}{M_i}$ denotes the density profile of a halo centred at $\boldsymbol{x}_i$, which is also assumed to be spherically symmetric and depend only on its mass,

$$u(\boldsymbol{x}-\boldsymbol{x}_i|M_i) = \begin{cases} \rho_{\rm h}\Big(|\boldsymbol{x}-\boldsymbol{x}_i|\Big|M_i\Big)/M_i, & |\boldsymbol{x}-\boldsymbol{x}_i| \leq R_{\rm halo}; \\ 0, & |\boldsymbol{x}-\boldsymbol{x}_i| > R_{\rm halo}. \end{cases} \tag{4.6}$$

  u is normalised: $\int \mathrm{d}^3\boldsymbol{x}\, u(\boldsymbol{x}-\boldsymbol{x}_i|M_i) = 1; \quad \int \mathrm{d}^3\boldsymbol{x}'\, u(\boldsymbol{x}-\boldsymbol{x}'|M) = 1;$

- Throughout this thesis, we have ignored the halo assembly bias, which refers to the fact that halo clustering also depends on halo properties other than mass, amongst which the most notable is halo assembly history (e.g. Gao et al., 2005; Wechsler et al., 2006). These different assembly histories influence secondary properties of haloes, which, in turn, might also affect the formation of galaxies, and hence result in different galactic contents for haloes of the same mass. The full treatment will be left in future works. Schematically, if halo correlation functions also depend on a secondary halo property S, the halo-model integral would be modified as

$$\int \mathrm{d}M\, \frac{\mathrm{d}n(M)}{\mathrm{d}M} \xi_{\rm hh}(r|M)\cdots \quad \rightarrow \quad \int \mathrm{d}M\, \frac{\mathrm{d}n(M)}{\mathrm{d}M} \int \mathrm{d}S\, \xi_{\rm hh}(r|M,S)\cdots . \tag{4.7}$$

  The candidates of the secondary halo property include halo concentration, spin, half-mass scale, etc (e.g. Mao et al., 2018).

Now we derive the expression of matter TPCFs in the halo model, by substituting Equation (4.2) with the matter density as

$$\begin{aligned} \langle \rho_{\rm m}(\boldsymbol{x})\rho_{\rm m}(\boldsymbol{x}')\rangle = \Bigg\langle & \left[\sum_i \int \mathrm{d}M_1 \int \mathrm{d}^3\boldsymbol{x}_1\ \delta^{(\rm D)}(M_1-M_i)\,\delta^{(\rm D)}(\boldsymbol{x}_1-\boldsymbol{x}_i)\,M_1 u(\boldsymbol{x}-\boldsymbol{x}_1|M_1)\right]\times \\ & \left[\sum_j \int \mathrm{d}M_2 \int \mathrm{d}^3\boldsymbol{x}_2\ \delta^{(\rm D)}(M_2-M_j)\,\delta^{(\rm D)}(\boldsymbol{x}_2-\boldsymbol{x}_j)\,M_2 u(\boldsymbol{x}'-\boldsymbol{x}_2|M_2)\right]\Bigg\rangle \end{aligned} \tag{4.8}$$

$$\begin{aligned} = \int \mathrm{d}M_1 \int \mathrm{d}M_2 \int \mathrm{d}^3\boldsymbol{x}_1 \int \mathrm{d}^3\boldsymbol{x}_2\, M_1 M_2\, u(\boldsymbol{x}-\boldsymbol{x}_1|M_1)\, u(\boldsymbol{x}'-\boldsymbol{x}_2|M_2) \\ \left\langle \left[\sum_i \delta^{(\rm D)}(M_1-M_i)\,\delta^{(\rm D)}(\boldsymbol{x}_1-\boldsymbol{x}_i)\right]\left[\sum_j \delta^{(\rm D)}(M_2-M_j)\,\delta^{(\rm D)}(\boldsymbol{x}_2-\boldsymbol{x}_j)\right]\right\rangle \end{aligned} \tag{4.9}$$

The ensemble average term in red can be split into the so-called 1-halo and 2-halo terms, corres-

ponding to the summation of the parts $i = j$ and $i \neq j$, respectively.

1-halo term $(i = j)$

$$= \left\langle \sum_i \left[\delta^{(\mathrm{D})}(M_1 - M_i) \delta^{(\mathrm{D})}(M_2 - M_i)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_i) \delta^{(\mathrm{D})}(\boldsymbol{x}_2 - \boldsymbol{x}_i) \right] \right\rangle \tag{4.10}$$

$$= \left\langle \sum_i \left[\delta^{(\mathrm{D})}(M_1 - M_2) \delta^{(\mathrm{D})}(M_1 - M_i)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_i) \right] \right\rangle \tag{4.11}$$

$$= \delta^{(\mathrm{D})}(M_1 - M_2)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \left\langle \sum_i \left[\delta^{(\mathrm{D})}(M_1 - M_i)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_i) \right] \right\rangle \tag{4.12}$$

$$= \delta^{(\mathrm{D})}(M_1 - M_2)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\, \frac{\mathrm{d}n(M_1)}{\mathrm{d}M_1}. \tag{4.13}$$

2-halo term $(i \neq j)$

$$= \left\langle \sum_i \sum_{j;j\neq i} \left[\delta^{(\mathrm{D})}(M_1 - M_i) \delta^{(\mathrm{D})}(M_2 - M_j)\, \delta^{(\mathrm{D})}(\boldsymbol{x}_1 - \boldsymbol{x}_i) \delta^{(\mathrm{D})}(\boldsymbol{x}_2 - \boldsymbol{x}_j) \right] \right\rangle \tag{4.14}$$

$$= \frac{\mathrm{d}n(M_1)}{\mathrm{d}M_1} \frac{\mathrm{d}n(M_2)}{\mathrm{d}M_2} \left[1 + \xi_{\mathrm{hh}}\Big(|\boldsymbol{x}_2 - \boldsymbol{x}_1| \Big| M_1, M_2 \Big) \right]. \tag{4.15}$$

A side note on the halo auto-correlation function

For a mass threshold halo sample (consisting of haloes more massive than M), denote the fractional number density field as $\delta_{\mathrm{h}}(\boldsymbol{x}| > M)$ and the average number density as $\bar{n}_{\mathrm{h}}(> M)$ (i.e. the cHMF). Analogy to the definition of the matter TPCF (Peebles, 1980), the halo two-point correlation function is defined as the *excess probability of finding a pair of haloes at* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$:

$$\mathrm{d}\mathcal{P}_{12} = \left[\bar{n}_{\mathrm{h}}(> M) \right]^2 \left[1 + \xi_{\mathrm{hh}}\Big(|\boldsymbol{x}_2 - \boldsymbol{x}_1| \Big| > M \Big) \right] \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2\,. \tag{4.16}$$

For two halo samples in two narrow different mass intervals $[M_1 + \mathrm{d}M_1)$ and $[M_2 + \mathrm{d}M_2)$, the cross-correlation function is similarly defined as the *excess probability of finding a halo with mass* $[M_1, M_1 + \mathrm{d}M_1)$ *at* $\boldsymbol{x}_1$ *and a halo with mass* $[M_2, M_2 + \mathrm{d}M_2)$ *at* $\boldsymbol{x}_2$:

$$\mathrm{d}\mathcal{P}_{12} = \frac{\mathrm{d}n(M_1)}{\mathrm{d}M_1}\, \mathrm{d}M_1 \frac{\mathrm{d}n(M_2)}{\mathrm{d}M_2}\, \mathrm{d}M_2 \left[1 + \xi_{\mathrm{hh}}\Big(|\boldsymbol{x}_2 - \boldsymbol{x}_1| \Big| M_1, M_2 \Big) \right] \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2\,. \tag{4.17}$$

The matter TPCF is given by

$$\langle \rho_{\mathrm{m}}(\boldsymbol{x})\, \rho_{\mathrm{m}}(\boldsymbol{x}') \rangle = \bar{\rho}_{\mathrm{m}}^2 \left[1 + \langle \delta_{\mathrm{m}}(\boldsymbol{x})\, \delta_{\mathrm{m}}(\boldsymbol{x}') \rangle \right] \tag{4.18}$$

$$= \bar{\rho}_{\mathrm{m}}^2 + \bar{\rho}_{\mathrm{m}}^2 \xi_{\mathrm{mm}}(|\boldsymbol{x}' - \boldsymbol{x}|). \tag{4.19}$$

Using this equation, we obtain the expression of the halo model of $\xi_{\rm mm}$ as

$$\begin{aligned}\xi_{\rm mm}(r) &= \xi^{\rm 1h}_{\rm mm}(r) + \xi^{\rm 2h}_{\rm mm}(r) \\ &= \int {\rm d}M\, \frac{M^2}{\bar{\rho}^2_{\rm m}} \frac{{\rm d}n(M)}{{\rm d}M} \int {\rm d}^3\boldsymbol{y}\; u(y|M)\, u\Big(|\boldsymbol{x}-\boldsymbol{x}'+\boldsymbol{y}|\Big|M\Big) \\ &+ \int {\rm d}M_1\, \frac{M_1}{\bar{\rho}_{\rm m}} \frac{{\rm d}n(M_1)}{{\rm d}M_1} \int {\rm d}M_2\, \frac{M_2}{\bar{\rho}_{\rm m}} \frac{{\rm d}n(M_2)}{{\rm d}M_2} \int {\rm d}^3\boldsymbol{x}_1\, u\Big(|\boldsymbol{x}-\boldsymbol{x}_1|\Big|M_1\Big) \\ &\quad \int {\rm d}^3\boldsymbol{x}_2\, u\Big(|\boldsymbol{x}'-\boldsymbol{x}_2|\Big|M_2\Big)\, \xi_{\rm hh}\Big(|\boldsymbol{x}_2-\boldsymbol{x}_1|\Big|M_1,M_2\Big),\end{aligned} \tag{4.20}$$

where $r \equiv |\boldsymbol{x}'-\boldsymbol{x}|$.

4.1.2 Halo-mass cross-correlation as an estimation of the averaged halo density profile

For two different populations of objects with overdensity fields $\delta_a(\boldsymbol{x})$ and $\delta_b(\boldsymbol{x})$, the two-point cross-correlation function is defined as

$$\xi_{ab}(\boldsymbol{r}) \equiv \langle \delta_a(\boldsymbol{x})\, \delta_b(\boldsymbol{x}+\boldsymbol{r}) \rangle\,, \tag{4.21}$$

where $\langle \cdots \rangle$ denotes ensemble averaging. Assuming statistical isotropy reduces $\xi_{ab}(\boldsymbol{r})$ to a function of separation only, that is, $\xi_{ab}(r)$. In the case of the cross-correlation between halo centres and the matter field, the cross-correlation function is given by

$$\xi_{\rm hm}(r) \equiv \langle \delta_{\rm h}(\boldsymbol{x})\, \delta_{\rm m}(\boldsymbol{x}+\boldsymbol{r}) \rangle \tag{4.22}$$

$$= \frac{1}{\bar{n}_{\rm h}\bar{\rho}_{\rm m}} \langle n_{\rm h}(\boldsymbol{x})\, \rho_{\rm m}(\boldsymbol{x}') \rangle - 1, \tag{4.23}$$

where $\boldsymbol{x}' \equiv \boldsymbol{x}+\boldsymbol{r}$, and the halo number density fluctuation field is defined as

$$\delta_{\rm h}(\boldsymbol{x}) \equiv \frac{n_{\rm h}(\boldsymbol{x}) - \bar{n}_{\rm h}}{\bar{n}_{\rm h}}. \tag{4.24}$$

Now we derive the expressions of the halo and mass fields in Equation (4.23) under the halo model framework. In realistic N-body simulations, halo catalogues always have a finite halo mass range, at least limited by the simulation force/mass resolution. Define the halo selection function for a given mass range (MR) as

$$\phi(m|{\rm MR}) \equiv \begin{cases} 1, & \text{if } m \in {\rm MR}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.25}$$

In practise, MR can be a narrow mass interval, $[M, M+{\rm d}M) \approx [M, M+\Delta M)$, or a mass threshold interval, $[M, \infty)$. The corresponding halo number density field is given by

$$n_{\rm h}(\boldsymbol{x}) = \sum_i \delta^{\rm (D)}(\boldsymbol{x}-\boldsymbol{x}_i)\, \phi(m_i|{\rm MR}). \tag{4.26}$$

To obtain the expression of $\langle n_{\rm h}(\boldsymbol{x})\,\rho_{\rm m}(\boldsymbol{x}')\rangle$ in the halo model, we can plug the expressions for the halo and mass fields, Equations (4.2) and (4.26), into Equation (4.23). The full expression can be found in Equations (7) and (8) of García et al. (2021). We focus only on the internal structure of the halo and, therefore, on the 1-halo term, which reads

$$\langle n_{\rm h}(\boldsymbol{x})\,\rho_{\rm m}(\boldsymbol{x}')\rangle_{\rm 1h} = \int {\rm d}m\,\frac{{\rm d}n}{{\rm d}m} m\,\phi(m|{\rm MR})\,u(r|m) \tag{4.27}$$

$$= \begin{cases} {\rm d}n(M)\,M\,u(r|M), & {\rm MR} = [M, M+{\rm d}M), \\ \displaystyle\int_M^\infty {\rm d}m\,\frac{{\rm d}n}{{\rm d}m}\,m\,u(r|m), & {\rm MR} = [M, \infty), \end{cases} \tag{4.28}$$

where ${\rm d}n(M)$ is the density of the halo number in the mass range $[M, M+{\rm d}M)$. The 1-halo term of the halo-mass correlation function is

$$\xi^{\rm 1h}_{\rm hm}(r|M) = \frac{M\,u(r|M)}{\bar{\rho}_{\rm m}} - 1 = \frac{\rho_{\rm h}(r|M)}{\bar{\rho}_{\rm m}} - 1, \tag{4.29}$$

$$\xi^{\rm 1h}_{\rm hm}(r|>M) = \frac{1}{\bar{\rho}_{\rm m}\,\bar{n}_{\rm h}(>M)} \int_M^\infty {\rm d}m\,\frac{{\rm d}n}{{\rm d}m}\,m\,u(r|m) - 1. \tag{4.30}$$

We are interested in the average density profile of haloes in a narrow mass interval $[M, M+{\rm d}M)$. However, the noise level of the simulation measurements for such halo properties and statistics is high because of the low number density. We bypass this problem by measuring $\xi^{\rm 1h}_{\rm hm}(r|>M)$ (which involves more haloes and therefore smoother) and taking the partial derivative with respect to mass,

$$u(r|M) = -\frac{\bar{\rho}_{\rm m}}{M}\left[\frac{{\rm d}n(M)}{{\rm d}M}\right]^{-1} \frac{\partial}{\partial M}\Big\{\bar{n}_{\rm h}(>M)\left[1+\xi^{\rm 1h}_{\rm hm}(r|>M)\right]\Big\}. \tag{4.31}$$

4.1.3 Galaxy clustering in the halo occupation distribution prescription

In this subsection, we examine how the clustering properties of galaxies are modelled in the halo model framework. Since galaxies reside in dark matter haloes, it is natural to describe the clustering of galaxies in terms of the halo clustering (on large scales) and the distribution of galaxies in dark matter haloes (small scales). In the spirit of the halo model, we split the galaxy TPCF into a one-halo term and a two-halo term.

Before the central-satellite split

In the HOD prescription, we suppose that a halo of mass M contains N galaxies, where the occupation number N follows a random distribution $P(N)$ called HOD. In the absence of *halo assembly bias*, HODs depend only on halo masses, that is, $P(N) = P(N|M)$. The further

dependencies on the secondary halo properties, which usually correlate with the formation history of haloes (Gao & White, 2007; Croton et al., 2007; Li et al., 2008), can be directly incorporated into the halo model framework due to its flexibility. We will ignore the assembly bias temporarily in this section.

The average spatial distribution of galaxies inside haloes is given by a normalised profile (usually related to the halo density profile)

$$u_{\rm g}(\boldsymbol{x}-\boldsymbol{x}_0|M) \equiv \frac{n_{\rm g}(\boldsymbol{x}-\boldsymbol{x}_0|M)}{N} = u_{\rm g}\Big(|\boldsymbol{x}-\boldsymbol{x}_0|\Big|M\Big), \tag{4.32}$$

where $\boldsymbol{x}_0$ is the centre of the halo, and we have assumed that the (averaged) profile is spherically symmetric, $u_{\rm g}(\boldsymbol{r}|M) = u_{\rm g}(r|M)$.

Motivated by the idea that galaxies can be treated as Poisson sampling of a smooth underlying density field, TPCFs of galaxies can be estimated as (Mo et al., 2010)

$$\xi_{\rm gg}(r) = \frac{\mathrm{DD}(r)}{\mathrm{RR}(r)} - 1, \tag{4.33}$$

where $\mathrm{DD}(r)\,\Delta r$ is the number of galaxy-galaxy pairs with separations in the range $[r-\Delta r/2, r+\Delta r/2]$, and $\mathrm{RR}(r)\,\Delta r$ is the expected number of such pairs in a Poisson distributed random sample with the same number density and volume as the galaxy catalogues. We can model galaxy pair counting in the HOD framework as follows.

The 1-halo term: galaxy pairs in one halo

- In a halo with properties $\{M, \boldsymbol{x}_0, N\}$, the average number of galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$ is

$$\begin{aligned}&\mathrm{GG}^{\rm 1h}(\boldsymbol{r}|M,N,\boldsymbol{x}_0)\,\mathrm{d}^3\boldsymbol{x}_1\,\mathrm{d}^3\boldsymbol{x}_2\\ =&N u_{\rm g}\Big(|\boldsymbol{x}_1-\boldsymbol{x}_0|\Big|M\Big)\,(N-1)u_{\rm g}\Big(|\boldsymbol{x}_2-\boldsymbol{x}_0|\Big|M\Big)\,\mathrm{d}^3\boldsymbol{x}_1\,\mathrm{d}^3\boldsymbol{x}_2\ ,\end{aligned} \tag{4.34}$$

  where $\mathrm{d}^3\boldsymbol{x}_i$ is the volume element in $\boldsymbol{x}_i$.

- Integrating over the halo centre $\boldsymbol{x}_0$ gives the average number density of galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$ in the haloes with properties $\{M, N\}$,

$$\mathrm{GG}^{\rm 1h}(\boldsymbol{r}|M,N) = \int \mathrm{d}^3\boldsymbol{x}_0\,\mathrm{GG}^{\rm 1h}(\boldsymbol{r}|M,N,\boldsymbol{x}_0) \tag{4.35}$$

$$= \int \mathrm{d}^3\boldsymbol{x}_0\,N u_{\rm g}\Big(|\boldsymbol{x}_1-\boldsymbol{x}_0|\Big|M\Big)\,(N-1)u_{\rm g}\Big(|\boldsymbol{x}_2-\boldsymbol{x}_0|\Big|M\Big) \tag{4.36}$$

$$= \int \mathrm{d}^3\boldsymbol{x}\,N(N-1)\,u_{\rm g}(x|M)u_{\rm g}\Big(|\boldsymbol{x}+\boldsymbol{r}|\Big|M\Big), \tag{4.37}$$

  where we have transformed the variable as $\boldsymbol{x}_0 \to \boldsymbol{x} \equiv \boldsymbol{x}_1 - \boldsymbol{x}_0$ in the second line.

- Now, we take into account HOD: Different haloes of the same mass M can host different N with probability $P(N|M)$. Therefore, the average number density of galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$ in haloes with properties $\{M\}$ is

$$\mathrm{GG}^{1\mathrm{h}}(\boldsymbol{r}|M) = \sum_N P(N|M)\,\mathrm{GG}^{1\mathrm{h}}(\boldsymbol{r}|M,N) \tag{4.38}$$

$$= \int \mathrm{d}^3\boldsymbol{x} \underbrace{\left[\sum_N P(N|M)\,N(N-1)\right]}_{\equiv\langle N(N-1)|M\rangle} u_\mathrm{g}(x|M) u_\mathrm{g}\Big(|\boldsymbol{x}+\boldsymbol{r}|\Big|M\Big) \tag{4.39}$$

$$= \int \mathrm{d}^3\boldsymbol{x}\,\langle N(N-1)|M\rangle\, u_\mathrm{g}(x|M) u_\mathrm{g}\Big(|\boldsymbol{x}+\boldsymbol{r}|\Big|M\Big), \tag{4.40}$$

  where we have defined the halo occupation number as

$$\langle N^k|M\rangle \equiv \sum_N N^k P(N|M). \tag{4.41}$$

- Finally, we arrive at the total number density of pairs of galaxies separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$, by integrating over the halo mass M weighted by HMFs,

$$\mathrm{GG}^{1\mathrm{h}}(\boldsymbol{r}) = \int \mathrm{d}M \frac{\mathrm{d}n\,(M)}{\mathrm{d}M} \langle N(N-1)|M\rangle \int \mathrm{d}^3\boldsymbol{x}\, u_\mathrm{g}(x|M) u_\mathrm{g}\Big(|\boldsymbol{x}+\boldsymbol{r}|\Big|M\Big) \tag{4.42}$$

The 2-halo term: galaxy pairs in two haloes

Consider one halo with the properties $\{M_1, N_1, \boldsymbol{x}\}$ and another halo $\{M_2, N_2, \boldsymbol{x}'\}$.

- The number density of inter-halo galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$ contributed by these two haloes is given by

$$N_1 N_2\, u_\mathrm{g}\Big(|\boldsymbol{x}_1 - \boldsymbol{x}|\Big|M_1\Big)\, u_\mathrm{g}\Big(|\boldsymbol{x}_2 - \boldsymbol{x}'|\Big|M_2\Big). \tag{4.43}$$

  The joint probability of finding a halo $\{M_1, \boldsymbol{x}\}$ and another halo with $\{M_2, \boldsymbol{x}'\}$ is related to the halo TPCF as

$$\mathrm{Prob}\left\{\begin{array}{c}\text{finding a halo with mass } M_1 \text{ at } \boldsymbol{x} \\ \text{AND} \\ \text{finding a halo with mass } M_2 \text{ at } \boldsymbol{x}'\end{array}\right\}$$

$$= \frac{\mathrm{d}n\,(M_1)}{\mathrm{d}M_1}\frac{\mathrm{d}n\,(M_2)}{\mathrm{d}M_2}\,\mathrm{d}M_1\,\mathrm{d}M_2\left[1 + \xi_\mathrm{hh}\Big(|\boldsymbol{x}-\boldsymbol{x}'|\Big|M_1,M_2\Big)\right]\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{x}'\,. \tag{4.44}$$

- The average number density of the inter-halo galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$, with one galaxy at $\boldsymbol{x}_1$ living in a halo $\{M_1, N_1, \boldsymbol{x}\}$ and another galaxy at $\boldsymbol{x}_2$ living in a halo

$\{M_2, N_2, \boldsymbol{x}'\}$, is given by

$$\begin{aligned}
&\mathrm{GG}^{\mathrm{2h}}(\boldsymbol{r}|M_1, M_2; N_1, N_2; \boldsymbol{x}, \boldsymbol{x}') \\
&= N_1 N_2\, u(\boldsymbol{x}_1 - \boldsymbol{x}|M_1)\, u(\boldsymbol{x}_2 - \boldsymbol{x}'|M_2) \\
&\quad \times \frac{\mathrm{d}n\,(M_1)}{\mathrm{d}M_1} \frac{\mathrm{d}n\,(M_2)}{\mathrm{d}M_2}\, \mathrm{d}M_1\, \mathrm{d}M_2 \left[1 + \xi_{\mathrm{hh}}\Big(|\boldsymbol{x} - \boldsymbol{x}'|\Big|M_1, M_2\Big)\right] \mathrm{d}^3\boldsymbol{x}\, \mathrm{d}^3\boldsymbol{x}'
\end{aligned} \tag{4.45}$$

- Similarly to the procedure in the 1-halo term, we consider HOD and integrate on the halo masses M_1, M_2 and the centres $\boldsymbol{x}, \boldsymbol{x}'$. The average number density of the inter-halo galaxy pairs separated by $\boldsymbol{r} \equiv \boldsymbol{x}_2 - \boldsymbol{x}_1$ is

$$\begin{aligned}
\mathrm{GG}^{\mathrm{2h}}(\boldsymbol{r}) = &\int \mathrm{d}M_1 \frac{\mathrm{d}n\,(M_1)}{\mathrm{d}M_1} \langle N_1|M_1\rangle \int \mathrm{d}M_2 \frac{\mathrm{d}n\,(M_2)}{\mathrm{d}M_2} \langle N_2|M_2\rangle \\
&\int \mathrm{d}^3\boldsymbol{x} \int \mathrm{d}^3\boldsymbol{x}'\, u_{\mathrm{g}}\Big(|\boldsymbol{x}_1 - \boldsymbol{x}|\Big|M_1\Big)\, u_{\mathrm{g}}\Big(|\boldsymbol{x}_2 - \boldsymbol{x}'|\Big|M_2\Big) \left[1 + \xi_{\mathrm{hh}}\Big(|\boldsymbol{x} - \boldsymbol{x}'|\Big|M_1, M_2\Big)\right].
\end{aligned} \tag{4.46}$$

In the halo model formalism and the HOD prescription, the galaxy TPCF is given by

$$\xi_{\mathrm{gg}}(\boldsymbol{r}) = \frac{\left[\mathrm{GG}^{\mathrm{1h}}(\boldsymbol{r}) + \mathrm{GG}^{\mathrm{2h}}(\boldsymbol{r})\right] \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2}{\mathrm{RR}(\boldsymbol{r})\, \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2} - 1, \tag{4.47}$$

where $\mathrm{RR}(\boldsymbol{r})\, \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2 = \bar{n}_{\mathrm{g}}^2\, \mathrm{d}^3\boldsymbol{x}_1\, \mathrm{d}^3\boldsymbol{x}_2$ and $\bar{n}_{\mathrm{g}} \equiv \int \mathrm{d}M\, \dfrac{\mathrm{d}n\,(M)}{\mathrm{d}M} \langle N|M\rangle$ is the mean number density of galaxies. $\mathrm{RR}(\boldsymbol{r})$ is the expected number of galaxy pairs in the absence of clustering.

After the central-satellite split

The above derivation assumes that every galaxy in a halo samples the profile $u_{\mathrm{g}}(r|M)$ uniformly. This is not the case for all galaxies. According to simulation studies of galaxy formation (Zheng et al., 2005b), a central galaxy is assumed to always sit near the centre of the host halo, whereas the distribution of satellite galaxies may be described by a spherical profile. We should treat the central and satellite galaxies separately.

Suppose that the profile of the satellite distribution in a halo (with mass M and centre $\boldsymbol{x}_0$) is given by $u_{\mathrm{s}}\Big(|\boldsymbol{x} - \boldsymbol{x}_0|\Big|M\Big)$. The galaxy pairs can be separated by 1-halo and 2-halo terms as

$$\mathrm{GG}(r) = \mathrm{CS}^{\mathrm{1h}}(r) + \mathrm{SS}^{\mathrm{1h}}(r) + \mathrm{CC}^{\mathrm{2h}}(r) + \mathrm{CS}^{\mathrm{2h}}(r) + \mathrm{SS}^{\mathrm{2h}}(r). \tag{4.48}$$

The occupation of central galaxies is parameterised as a Bernoulli distribution, whereas that of satellites is a Poisson distribution (Zheng et al., 2005b). Both distributions are described by their mean occupation number,

$$\langle N_{\mathrm{g}}\rangle_M = \langle N_{\mathrm{c}}\rangle_M + \langle N_{\mathrm{s}}\rangle_M. \tag{4.49}$$

The galaxy number density $\bar{n}_{\rm g}$ can then be obtained by integrating the HMF weighted by the mean occupation,

$$\bar{n}_{\rm g} = \int {\rm d}M \, \frac{{\rm d}n(M)}{{\rm d}M} (\langle N_{\rm c} \rangle (M) \, + \langle N_{\rm s} \rangle (M)). \tag{4.50}$$

To be specific, we adopt the HOD model in Zheng et al. (2005b) by introducing the following HOD parameters,

$$\boldsymbol{\mathcal{G}} = \{M_{\rm min}, \sigma_{\log M}, M_1, \kappa, \alpha\}. \tag{4.51}$$

The mean HOD for central galaxies is given by

$$\langle N_{\rm c} \rangle (M|\boldsymbol{\mathcal{G}}) = \frac{1}{2} \left[1 + {\rm erf} \left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}} \right) \right]. \tag{4.52}$$

The mean central HOD, $\langle N_{\rm c} \rangle (M)$, can be interpreted as the probability that a halo with mass M hosts a central galaxy. The mean central HOD considered here has properties $\langle N_{\rm c} \rangle \to 0$ for haloes with $M \ll M_{\rm min}$, while $\langle N_{\rm c} \rangle \to 1$ for haloes with $M \gg M_{\rm min}$.

The mean satellite HOD is parameterised as

$$\langle N_{\rm s} \rangle (M|\boldsymbol{\mathcal{G}}) = \langle N_{\rm c} \rangle (M|\boldsymbol{\mathcal{G}}) \lambda_{\rm s}(M) \tag{4.53}$$

$$\lambda_{\rm s}(M) = \left(\frac{M - \kappa M_{\rm min}}{M_1} \right)^{\alpha}. \tag{4.54}$$

Following the commonly-used prescription, we assume that satellite galaxies reside only in a halo that already hosts a central galaxy. Hence, in the above equation, the satellite galaxy(ies) can only reside in haloes with $\langle N_{\rm c} \rangle = 1$. Then we assume that the number distribution of satellite galaxies in a given host halo follows the Poisson distribution with mean $\lambda_{\rm s}(M)$:

$$P(N_{\rm s}|N_{\rm c} = 1) = \frac{[\lambda_{\rm s}(M)]^{N_{\rm s}} {\rm e}^{-\lambda_{\rm s}(M)}}{N_{\rm s}!} \quad \text{and} \quad P(N_{\rm s}|N_{\rm c} = 0) = \delta^{\rm Kr}_{N_{\rm s},0}, \tag{4.55}$$

where $\delta^{\rm Kr}_{ij}$ stands for the Kronecker delta.

Given these assumptions, we can express galaxy TPCF as a function of dark matter halo properties in the halo model framework. First, we split the one- and two-halo terms into correlation of central and satellite galaxies,

$$\begin{aligned} \xi_{\rm gg}(r) = \, & \xi^{\rm 1h}_{\rm cs}(r) + \xi^{\rm 1h}_{\rm ss}(r) \\ & + \xi^{\rm 2h}_{\rm cc}(r) + \xi^{\rm 2h}_{\rm cs}(r) + \xi^{\rm 2h}_{\rm ss}(r). \end{aligned} \tag{4.56}$$

Note that some terms involving both centrals and satellites lead to convolution of the halo profiles and/or the halo TPCF, such as $\int {\rm d}^3\boldsymbol{x} \, u(x|M) \, u\Big(|\boldsymbol{x} + \boldsymbol{r}| \Big| M\Big)$. It is therefore more convenient

to compute these terms in Fourier space, where convolutions are simple products of the Fourier modes. Following the similar procedure in the last subsection, the expressions of the galaxy TPCF after the central-satellite split are given by

$$\xi(r) = \int_0^\infty \frac{\mathrm{d}k}{(2\pi)^3} 4\pi k^2 \frac{\sin(kr)}{kr} P(k), \tag{4.57}$$

$$P_{\rm cs}^{\rm 1h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \int \mathrm{d}M \frac{\mathrm{d}n(M)}{\mathrm{d}M} \langle N_{\rm c}\rangle(M)\, \lambda_{\rm s}(M)\, u(k|M), \tag{4.58}$$

$$P_{\rm ss}^{\rm 1h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \int \mathrm{d}M \frac{\mathrm{d}n(M)}{\mathrm{d}M} \langle N_{\rm c}\rangle(M)\, [\lambda_{\rm s}(M)]^2\, [u(k|M)]^2, \tag{4.59}$$

$$P_{\rm cs}^{\rm 2h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \int \mathrm{d}M \frac{\mathrm{d}n(M)}{\mathrm{d}M} \langle N_{\rm c}\rangle(M)\, u(k|M) \int \mathrm{d}M' \frac{\mathrm{d}n(M')}{\mathrm{d}M'} \langle N_{\rm c}\rangle(M')\, \lambda_{\rm s}(M')\, P_{\rm hh}(k|M, M'), \tag{4.60}$$

$$P_{\rm ss}^{\rm 2h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \int \mathrm{d}M \frac{\mathrm{d}n(M)}{\mathrm{d}M} \langle N_{\rm c}\rangle(M)\, \lambda_{\rm s}(M)\, u(k|M) \int \mathrm{d}M' \frac{\mathrm{d}n(M')}{\mathrm{d}M'} \langle N_{\rm c}\rangle(M')\, \lambda_{\rm s}(M')\, P_{\rm hh}(k|M, M')\, u(k|M'). \tag{4.61}$$

We can compute the central-central TPCF term without the Fourier transform as

$$\xi_{\rm cc}^{\rm 2h}(r) = \frac{1}{\bar{n}_{\rm g}^2} \int \mathrm{d}M \frac{\mathrm{d}n(M)}{\mathrm{d}M} \langle N_{\rm c}\rangle(M) \int \mathrm{d}M' \frac{\mathrm{d}n(M')}{\mathrm{d}M'} \langle N_{\rm c}\rangle(M')\, \xi_{\rm hh}(r|M, M'). \tag{4.62}$$

In the expressions above, we highlight the emulated quantities in blue, such as halo mass functions $\frac{\mathrm{d}n}{\mathrm{d}M}$ and halo auto-correlation function $\xi_{\rm hh}(r|M, M')$, following the convention of the Dark Quest paper (Miyatake et al., 2020). In the following section, we will introduce the machine learning algorithm — neural networks — to construct simulation-based templates, or *emulators*.

4.2 Interlude: neural networks for interpolation in high dimensions

Given a dataset comprised of inputs (also called *features*) and outputs (*labels*), we assume that there *exists* an unknown underlying function mapping the inputs to the outputs. We can use supervised learning algorithms to approximate this function. In the context of cosmology, for example, we can run N-body simulations given a set of cosmological parameters $\boldsymbol{\mathcal{C}}$, save the snapshot at a given redshift, and measure the HMFs,

$$\frac{\mathrm{d}n(M|\boldsymbol{\mathcal{C}}, z)}{\mathrm{d}M}, \quad \boldsymbol{\mathcal{C}} \xrightarrow{\text{for example}} \{\Omega_{\rm m0}, \Omega_{\rm b0}, H_0, \sigma_8, n_s, \cdots\}.$$

It is prohibitively expensive to perform $\gtrsim \mathcal{O}(10^4)$ cosmological simulations to explore the parameter space in a typical Markov chain Monte Carlo (MCMC) analysis. As we will demonstrate in the next sections, it is possible to construct *emulators* by running affordable numbers (e.g. 50-100)

of simulations and "interpolate" in high-dimensional parameter space. Emulators can make accurate predictions for new models without running simulations. Thanks to the development of algorithms and computing power, statistics and machine learning provide us with a wealth of tools to solve such regression problems.

To build emulators, we adopt fully connected neural networks (Alom et al., 2019), a type of artificial neural network in which all nodes (or neurons) in one layer are connected to the neurons in the next layer. The neural network algorithm has been widely applied to cosmology (e.g. Agarwal et al., 2012, 2014; Jennings et al., 2019; Kobayashi et al., 2020), and its performance was shown to be competitive or sometimes better than other existing methods. Neural networks have a strong fitting ability, as reflected in the *universal approximation theorem* (Cybenko, 1989; Hornik et al., 1989; Goodfellow et al., 2016):

> Let $\phi(\cdot)$ be a non-constant, bounded, monotonically increasing and continuous function, $\mathcal{J}^D$ is a D-dimensional hyper-cube $[0,1]^D$ and $C(\mathcal{J}^D)$ is the set of continuous functions defined on $\mathcal{J}^D$.
>
> For any function $F(\boldsymbol{x}) \in C(\mathcal{J}^D)$, there exist an integer M, a group of numbers $v_m, b_m \in \mathbb{R}$ and vectors $\boldsymbol{w}_m \in \mathbb{R}^D, m = 1, \ldots, M$, so that we can define the function,
>
> $$f(\boldsymbol{x}) \equiv \sum_{m=1}^{M} v_m \, \phi\Big(\boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{x} + b_m\Big), \tag{4.63}$$
>
> such that $f(\boldsymbol{x})$ is a function approximation of $F(\boldsymbol{x})$, that is,
>
> $$\big|f(\boldsymbol{x}) - F(\boldsymbol{x})\big| < \varepsilon, \forall \boldsymbol{x} \in \mathcal{J}^D, \tag{4.64}$$
>
> where $\varepsilon > 0$ is a small positive number.

However, this theorem does not provide a construction for the optimised neural networks, but merely guarantees the existence. Also, this strong fit ability also makes neural networks more prone to the overfitting problem compared to Gaussian processes. Therefore, we need to carefully check the emulation performance using the validation data set and tune the network architecture so that the generalisation error is successfully suppressed.

Many cosmological emulation projects, such as Dark Quest (Nishimichi et al., 2019a), the Coyote Universe (Heitmann et al., 2010), the Mira-Titan Universe (Bocquet et al., 2020), Aemulus (DeRose et al., 2019), AbacusSummit (Maksimova et al., 2021) and FORGE (Arnold et al., 2021), fitted matter, halo and/or galaxy properties from N-body simulations using a combination

of principal component analysis, to reduce the dimensionality of the data vector, and Gaussian processes (Williams & Rasmussen, 2006), to fit the dependency of the principal component coefficients as a function of cosmology. Here, we show how dimensionality reduction can be avoided through neural networks, which also improve the accuracy of the predictions.

Fully connected neural networks approximate a function f such that

$$\boldsymbol{y} = f(\boldsymbol{x}|\boldsymbol{\theta}), \tag{4.65}$$

where $\boldsymbol{x}$ represents the inputs or features, $\boldsymbol{y}$ the desired outputs or labels, and $\boldsymbol{\theta}$ the network free parameters, also called trainable. The optimal function f is defined by the set of values $\boldsymbol{\theta}_\star$ that minimises the loss function, which provides a measure of the performance of the model when evaluated in the dataset.

Rectified Linear Unit (ReLU) (Agarap, 2018) is the most commonly used activation function in current neural networks used to add non-linearities in the mapping between inputs and outputs. ReLU is defined as

$$\mathrm{ReLU}(x) = \max(0, x), \tag{4.66}$$

where x is the output of the previous layer of the neural network. Note that the activations of ReLU are not differentiable at $x = 0$. In this work, however, we are interested in functions that are differentiable with respect to its inputs and, in particular, with respect to the cosmological parameters. Therefore, throughout, we use Gaussian error linear unit (GELU) (Hendrycks & Gimpel, 2016) as the activation function instead,

$$\mathrm{GELU}(x) = 0.5x\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \tag{4.67}$$

To find the optimal parameters $\boldsymbol{\theta}_\star$ that reproduce the statistics measured in the N-body simulations, we minimise the loss function L1,

$$\mathcal{L} = \frac{1}{N}\sum_{i=0}^{N} |y^i_{\mathrm{true}} - y^i_{\mathrm{predicted}}| \tag{4.68}$$

using Adam optimiser (Kingma & Ba, 2014). Compared to the mean squared error, the loss of L1 reduces the importance given to outlier errors. To avoid fine-tuning the learning rate, we adopt a learning rate scheduler that reduces the learning rate by a factor of 10 every time the validation loss does not improve after 30 epochs. We also stop the training process when the validation loss does not improve after 100 epochs. This iterative reduction of the learning rate allows the model to quickly learn the broad characteristics of the data and later reduce the errors with a smaller learning rate. The initial learning rate is always set to 0.01.

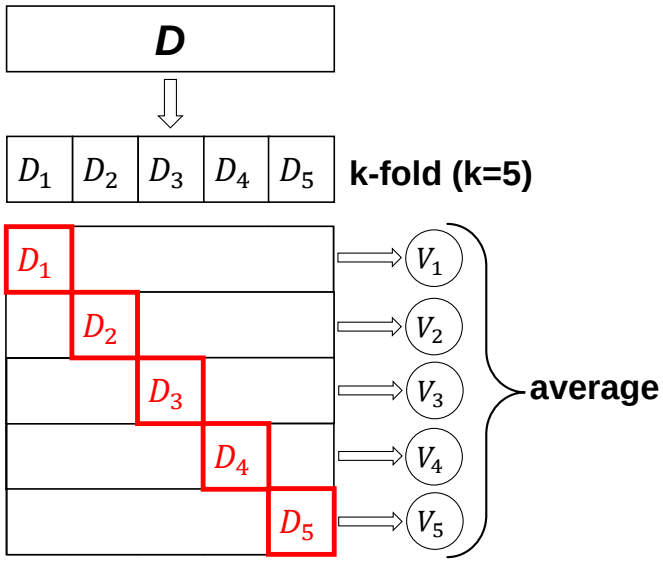

Figure 4.1: Illustration of the k-fold cross-validation. To avoid overfitting, the whole dataset D is divided into k (here $k = 5$) mutually exclusive subsets. For each subset, we take it as the test set (red boxes) and the remaining subsets as the training set. The final emulator output is the mean of the k versions' outputs.

Suppose that we have a dataset D comprising n instances of data:

$$D = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}.$$

For example, the feature $\boldsymbol{x}$ could be the cosmological parameters, and the label $\boldsymbol{y}$ could be the HMF for the given cosmology. If we simply use the entire data set D as a training set, the resulting neural networks are prone to the overfitting problem: the training error is small, but performance is poor for the new models outside of D. We use the k-fold cross-validation to overcome this problem, which is illustrated in Fig. 4.1. First, we randomly split the dataset D into k mutually exclusive subsets of similar size;

$$D = D_1 \cup D_2 \cdots \cup D_k. \tag{4.69}$$

For each subset, we take it as the test set and the remaining subsets as the training set, which returns k versions of the trained neural networks. The final output of the emulator is the mean of the k versions.

In the following sections, we demonstrate the precision of fully connected networks in some ideal emulation tests, as well as in reproducing the HMF and the halo-mass cross-correlation function measured from the FORGE simulations (Arnold et al., 2021).

4.3 Ideal emulation tests

To have a preliminary impression of the emulation process, and to guide the design of the emulators in the future, we construct emulators under ideal conditions. We generate halo properties for randomly selected cosmologies with the existing analytical methods or fitting formulas, which are noise-free mappings from cosmologies to halo properties, unlike simulation measurements. Then we use these data to train neural networks. To evaluate the emulator performance, we compare the "true" values with the emulator predictions on independent test datasets and calculate the fractional error.

The cosmologies of the training set cover 50 flat geometry w_0w_aCDM models (Linder, 2003), where the equation-of-state parameter for dark energy is parameterised as

$$w(a) = w_0 + w_a(1 - a). \tag{4.70}$$

A key aspect of building emulators is an efficient sampling scheme. As the training dataset, fifty cosmologies were sampled using optimal minimax distance sliced Latin hypercube designs (Ba et al., 2015), which enable efficient sampling from the seven-dimensional cosmological parameter space,

$$\boldsymbol{\mathcal{C}} = \left\{\Omega_{\rm m0}, \Omega_{\rm b0}, h, \sigma_8, n_s, w_0, w_a\right\}, \tag{4.71}$$

as shown in the grey dots in Fig. 4.2. The range of parameters is

$$\begin{aligned} &0.1 < \Omega_{\rm m0} < 0.7, \quad 0.02 < \Omega_{\rm b0} < 0.06, \quad 0.5 < h < 0.9, \\ &0.5 < \sigma_8 < 1.2, \quad 0.92 < n_s < 0.99, \\ &-1.3 < w_0 < -0.7, -0.1 < w_a < 0.1, \end{aligned} \tag{4.72}$$

while the upper and lower parameter limits deviate significantly from the current best-fit ΛCDM background cosmology from the *Planck* satellite (Planck Collaboration et al., 2020a). To evaluate the performance of the emulators, we also generate two test datasets that were not used in the training: both consist of 500 random cosmologies; one set covers the same parameter range as that of the training set, and the other test set covers the inner half-region (in terms of the length per dimension, instead of volume) of the parameter space. The cosmologies in the full- and half-range test datasets are presented as green and blue dots in Fig. 4.2, respectively.

We consider two basic properties: the halo concentration-mass relation, denoted as $c(M)$, and HMF. For given cosmologies, we generate the $c(M)$ relation calibrated in Prada et al. (2012),

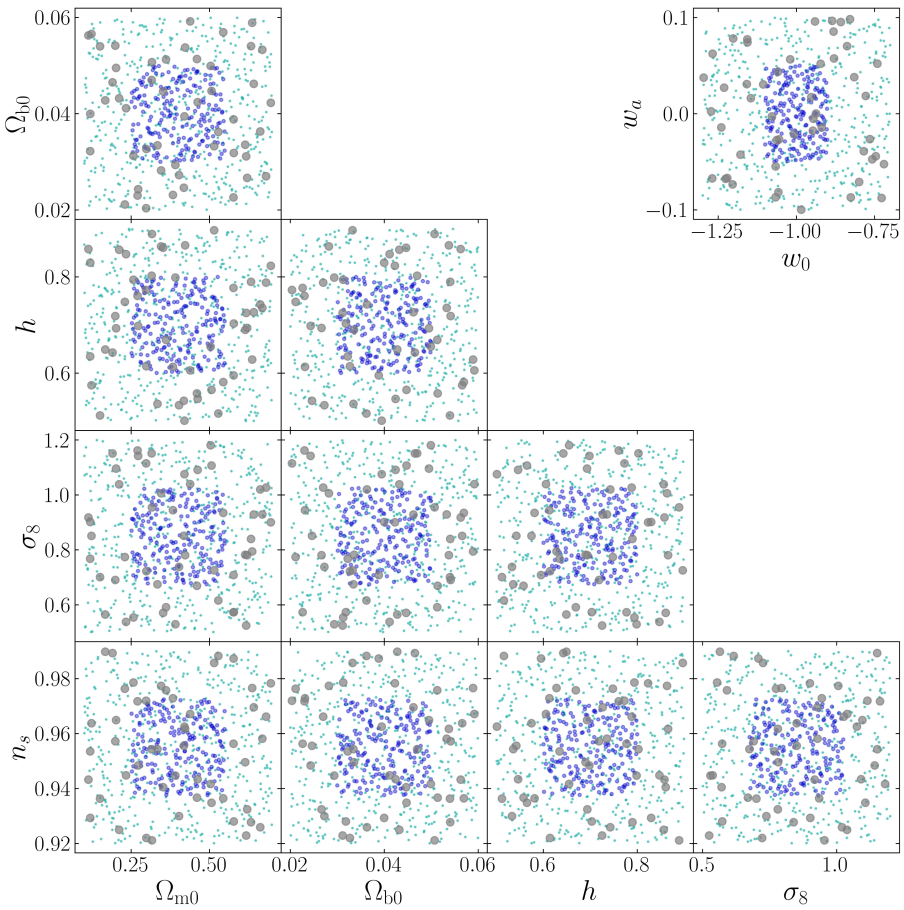

Figure 4.2: Seven-parameter w_0w_aCDM cosmologies of the training and test sets in the ideal emulation tests. Grey dots: **training** set including 50 nodes. Green dots: full-range **test** set consisting of 500 nodes covering the save range of parameters as that of the training set. Blue dots: half-range **test** set including 500 nodes in the inner region of the parameter space.

using the publicly available Python toolkit COLOSSUS (Diemer, 2018), and compute the Tinker et al. (2008) HMF with the Python application HMFcalc (Murray et al., 2013). The upper panels of Fig. 4.3 show the halo properties calculated with both the analytical methods and the emulators. The fractional error results are shown in the lower panels. The error is estimated as the mean fractional difference between the model and emulator outputs for all of the test cosmologies. The results show that the emulator reproduces the analytical halo $c(M)$ relation with a sub-percent error in the 7D parameter space with 50 training data. The performance of the HMF emulator is a bit worse, as the HMF data span ~ 20 orders of magnitude in our tests. However, the emulator still performs well: the typical emulator precision is better than 1% for $\lesssim 10^{13.5}\,h^{-1}M_\odot$ and $< 5\%$ for $M \lesssim 10^{15}\,h^{-1}M_\odot$.

We also note that the emulator precision is generally different at the edges and inside the parameter space, as revealed by the green and blue lines in the bottom panels of Fig. 4.3. This suggests that we should design the parameter space to be wider than the existing cosmological constraints. The parameter space in our ideal tests is designed to be wide enough that even half of

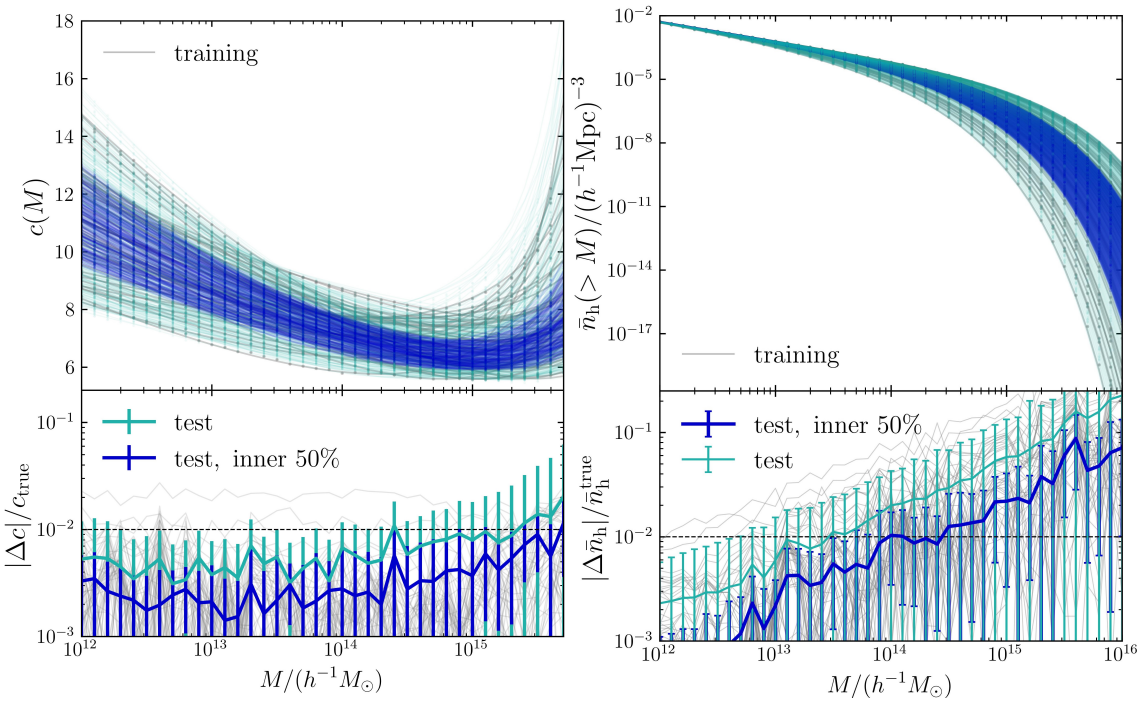

Figure 4.3: The ideal emulation test results. **Top**: The halo concentration-mass relation and halo mass function of the cosmologies in the training (grey), full-range test (green) and half-range test (blue) sets. **Bottom**: The absolute value of the relative difference between the outputs of the models and emulators. The thick green and blue lines and errorbars present the mean and standard deviation of the emulation fractional errors among 500 cosmologies of each test set.

it still covers the Planck Collaboration et al. (2020a) best fit ΛCDM model well.

In general, ideal emulation tests show that under noise-free conditions, the emulator can provide accurate interpolations in a high-dimensional parameter space, using 50 efficiently sampled models as the training set. The fractional error shows a power-law dependence on the halo mass, with larger errors for more massive haloes. In the next section, we will present the real emulation results: The data are measured from simulations and, therefore, are influenced by noise.

4.4 The emulator-based halo model

4.4.1 The FORGE simulation suite

To train emulators for dark matter halo properties in modified gravity, we use the FORGE suite of simulations (Arnold et al., 2021). The simulations were performed with 1024^3 dark matter particles in $512\,h^{-1}\,{\rm Mpc}$ (hereafter high-resolution runs, denoted HR) or $1500\,h^{-1}\,{\rm Mpc}$ (low-resolution runs, labelled LR) side-length boxes, using the MG-Arepo N-body code (Arnold et al., 2019b). The mass resolutions of the HR and LR runs are 9.1×10^9 and $1.5\times10^{12}(\Omega_{\rm m0}/0.3)\,h^{-1}\,M_\odot$, respectively. The initial conditions were generated using the second-order Lagrangian perturbation

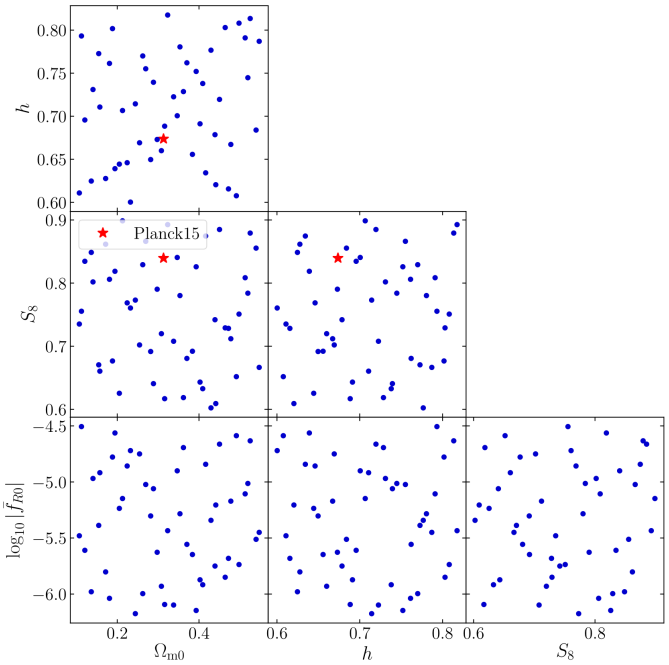

Figure 4.4: Visualisation of the cosmological parameters covered in the FORGE suite of simulations. The red star shows the fiducial *Planck* cosmology from Planck Collaboration et al. (2020a). ΛCDM models correspond to $\bar{f}_{R0} = 0$ and $\log_{10}|\bar{f}_{R0}| = -\infty$, therefore are not shown in the bottom subpanels. The ASCII version of the parameter table can be found at this link.

theory (2LPT, Crocce et al. (2006b)) and the redshift at $z_{\rm init} = 127$.

The cosmologies of the simulations cover 50 flat geometry $f(R)$ gravity models with a ΛCDM background, as shown in Fig. 4.4. The set of cosmological parameters is defined using optimal minimax distance sliced Latin hypercube designs (Ba et al., 2015), which enable efficient sampling from the four-dimensional parameter space,

$$\mathcal{C} = \left\{\Omega_{\rm m0}, h, S_8, \bar{f}_{R0}\right\}, \tag{4.73}$$

where $h \equiv H_0/(100\,{\rm km\,s^{-1}\,Mpc^{-1}})$. The other parameters are fixed at $n_s = 0.9652$ and $\Omega_{\rm b0} = 0.049199$. Since the FORGE simulations were partly designed to emulate weak lensing statistics, the authors sampled directly in the composite structure growth parameter

$$S_8 \equiv \sigma_8\sqrt{\frac{\Omega_{\rm m0}}{0.3}},$$

instead of the physical matter fluctuation amplitude parameter σ_8. The use of S_8 accounts better for the degeneracy between $\Omega_{\rm m0}$ and σ_8 in the cosmic shear analysis. In additional to the 50 models, there are two $f(R)$ gravity models with $\bar{f}_{R0} = -10^{-5}$ (F5) and -10^{-6} (F6) and the fiducial *Planck* cosmology (Planck Collaboration et al., 2020a),

$$\{\Omega_{\rm m0}, \Omega_{\Lambda 0}, \Omega_{\rm b0}, n_s, h, \sigma_8\} = \{0.31315, 0.68685, 0.049199, 0.9652, 0.6737, 0.82172\}. \tag{4.74}$$

Each of them has 8 realisations.

The dark matter halo catalogues were obtained with the SUBFIND (Springel et al., 2001) halo finder. The haloes were first identified using a fast parallel friends-of-friends (FOF) algorithm with link length set to $b = 0.168$. Spherical overdensity halo catalogues are then built out from the potential minimum of each FOF halo. The halo mass definition adopted is

$$M_{200\mathrm{c}} \equiv \frac{4\pi}{3}(R_{200\mathrm{c}})^3 \times 200\rho_{\mathrm{crit}},$$

where $\rho_{\mathrm{crit}}(z) \equiv 3H^2(z)/(8\pi G)$ is the critical density of the Universe, and $R_{200\mathrm{c}}$ is the spherical halo radius within which the spherically averaged mass density equals $200\rho_{\mathrm{crit}}$. Only main haloes with masses above $10^{12}\,h^{-1}M_\odot$ are studied in this work. The halo catalogues at $z = 0.0, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75$ and 2.00 are available. In this chapter, we present the performance of the emulators at redshift zero, since the measurement and emulation at other redshifts are performed in the same way.

4.4.2 The emulator for halo mass functions

For each halo catalogue, we bin the halo masses with a default bin width $\Delta \log_{10} M = 0.1$. For each cosmology, we remove the high-mass bins which contain less than 50 haloes to avoid the impact of shot noise. As a conservative choice, the resulting halo mass range is between 10^{12} and $10^{14}\,h^{-1}M_\odot$. We have checked that the measured mass functions and constructed emulators are robust to changes in the details of the binning scheme. In the upper panel of Fig. 4.5, we show the measured HMFs for the 50 FORGE cosmologies.

We built HMF emulators for each snapshot, instead of treating the redshift as another feature of the neural networks. As mentioned in Section 4.2, the data reduction procedures in the commonly used Gaussian process-based emulators, such as principal component analysis, are not needed with neural networks. We summarise the key steps in constructing the emulators as follows.

- For every input cosmology, we measure a halo mass function from the halo catalogue. To obtain a smoother function, we choose to emulate the *cumulative* halo mass function,

$$\bar{n}_{\mathrm{h}}(> M) = \int_{\log_{10} M}^{\infty} \mathrm{d}\log_{10} M' \, \frac{\mathrm{d}n(M')}{\mathrm{d}\log_{10} M}.$$

- Follow the k-fold cross-validation scheme described in Section 4.2, we randomly split the 50 cosmologies into 5 subsets. Then we take each subset as the test set and the remaining subsets as the training set. The neural network training procedure was implemented using

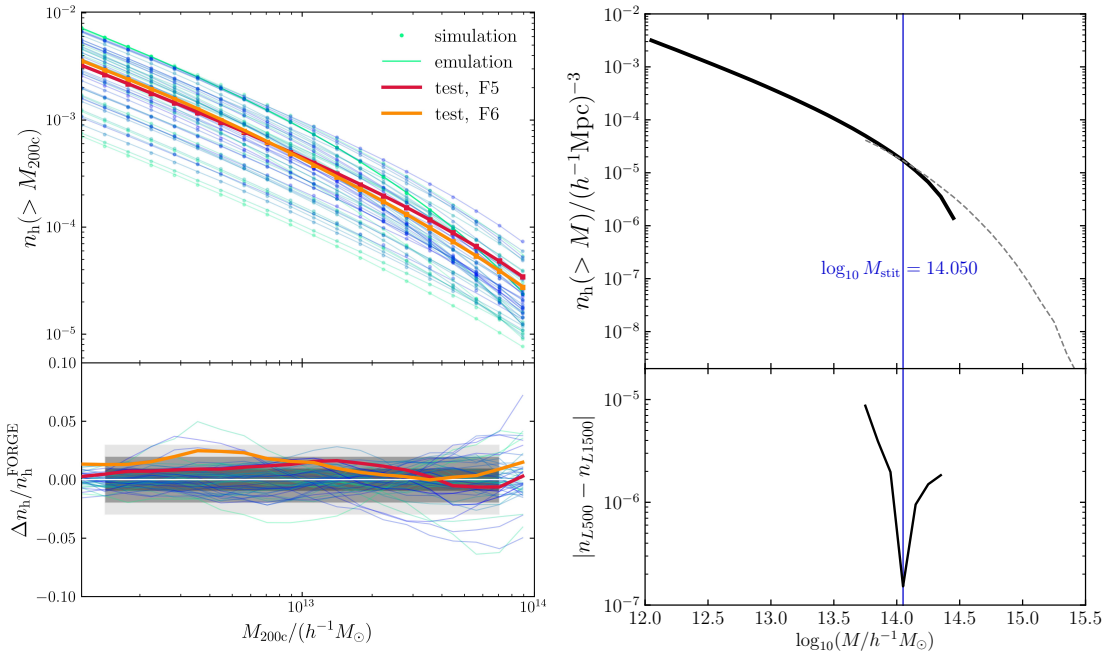

Figure 4.5: **Left** panel: The performance of the HMF emulator at $z = 0$ for $f(R)$ gravity. The upper subpanel shows the cumulative halo mass functions (cHMF) from simulation measurements (dots) and emulator outputs (lines), for 50 FORGE cosmologies (green) and two test cosmologies F5 (red) and F6 (orange) that were not used in the emulator construction. The fractional errors of the emulator are shown in the lower panel, where the light, medium and deep grey bands denote respectively ± 3, ± 2 and ± 1 % differences. **Right** panel: The demonstration of extending the mass range of the HMFs, by stitching the HMF measurements from a pair of small- (solid black line) and large-box size (dashed grey line) simulations. The upper subpanel presents the cumulative HMFs of the two simulations with the same specifications and different box sizes: 500 and 1500 h^{-1}Mpc. The lower subpanel shows their absolute differences. The pivot mass is identified as the mass where the difference is minimised, as shown in the solid blue line.

the open-source Python framework PyTorch (Paszke et al., 2019), and the adaptive stochastic optimisation algorithm Adam (Kingma & Ba, 2014).

- After the training, the emulators would learn the mapping between the input cosmological parameters and the output cHMFs,

$$\boldsymbol{\mathcal{C}}_{\rm in} \rightarrow \bar{n}_{\rm h}(>M|\boldsymbol{\mathcal{C}}_{\rm in}). \tag{4.75}$$

The performance of the HMF emulator is shown in the left panel of Fig. 4.5. In the lower subpanel, we present the fractional errors of the emulator,

$$\frac{\bar{n}_{\rm h}^{\rm emulator} - \bar{n}_{\rm h}^{\rm simulation}}{\bar{n}_{\rm h}^{\rm simulation}}, \tag{4.76}$$

estimated in the training set, as well as two test cosmologies (F5 and F6) that were independent with the emulator. The emulator achieves 2% accuracy for the two test cosmologies around the centre of the parameter space, and the errors for most of the 50 models are lower than 3%. Interestingly, we do not observe the mass-dependence of the emulator accuracy as in the ideal emulation test

(Fig. 4.3). The possible explanations are (1) the cosmology-dependence in the presence of modified gravity may be substantially different; and (2) the orders of magnitude in the real HMFs are much less than those in the ideal tests.

According to previous studies of emulating HMFs (e.g. Nishimichi et al., 2019a; McClintock et al., 2019; Bocquet et al., 2020), the most challenging part lies in the high-mass end $> 10^{14}\,h^{-1}M_{\odot}$, because of the large Poisson noise that affects the measurements caused by the small number of haloes in the simulations. Our conservative choice of the maximum mass is due to the relatively small box size ($500\,h^{-1}\mathrm{Mpc}$). We can extend to dynamic range of emulator predictions by combining simulations performed in different box sizes and stitching their measurements. This is demonstrated in the right panel of Fig. 4.5, where we show the stitching the cHMFs from a pair of HR and LR FORGE simulations. Their HMFs overlap at the absolute difference reaches at some pivot mass. We can then smoothly stitch the two measurements and extend to a maximum mass of $10^{15}\,h^{-1}M_{\odot}$. The emulation of the stitched HMFs is beyond the scope of the thesis, and we will explore it in future works.

4.4.3 The emulator for halo density profiles

4.4.3.1 The individual halo profile

One of the most remarkable discoveries from cosmological N-body simulations was that the dark matter haloes share a universal density profile (Navarro et al., 1996, 1997; Wang et al., 2020), from the mass of the host haloes of dwarf galaxies to those of massive galaxy clusters. Specifically, it was shown that the spherically averaged density profile of individual relaxed haloes can be described by the well-known Navarro-Frenk-White (NFW) profile. The NFW profile is described by two parameters, the characteristic density and scale radius of the halo, or equivalently the halo mass and concentration,

$$\rho_{\mathrm{NFW}}(r) = \frac{4\rho_{-2}}{\left(\frac{r}{r_{-2}}\right)\left(1+\frac{r}{r_{-2}}\right)^{2}}, \tag{4.77}$$

where r_{-2} is the characteristic radius (also denoted as r_{s}) of a halo at which the logarithmic slope of the density profile equal to -2,

$$\left.\frac{\mathrm{d}\log\left[\rho_{\mathrm{NFW}}(r)\right]}{\mathrm{d}\log r}\right|_{r=r_{-2}} = -2, \tag{4.78}$$

and $\rho_{-2} \equiv \rho_{\rm NFW}(r = r_{-2})$. The halo concentration c is defined as the ratio between the halo radius (which is adopted as $R_{\rm 200c}$ in this work) and r_{-2},

$$c \equiv \frac{R_{\rm 200c}}{r_{-2}}. \tag{4.79}$$

The other NFW parameter is related to the concentration as

$$\rho_{-2} = \frac{\rho_{\rm crit}(a)}{4} \frac{200}{3\Omega_{\rm m}(a)} \frac{c^3}{f(c)}, \tag{4.80}$$

where $\rho_{\rm crit}(a) \equiv 3H^2(a)/(8\pi G)$ is the critical density of the Universe at the redshift of the halo catalogues; $\Omega_{\rm m}(a) \equiv \bar{\rho}_{\rm m}(a)/\rho_{\rm crit}(a)$ is the matter density parameter; $f(c) \equiv \ln(1+c) - c/(1+c)$.

Halo profiles themselves are of considerable interest. Individual halo profiles can be detected by X-ray observations assuming hydrostatic equilibrium, or inferred indirectly for less massive subhaloes using statistical lensing probes. The (averaged) halo profile is a key ingredient of the halo model of the galaxy clustering, as described in Section 4.1.3. As the observations continue to improve, a corresponding sharpening in theoretical predictions for halo concentration-mass relations and density profiles of different cosmologies is required.

In the previous years, the researches were focused on measuring halo concentrations for one or a few cosmology(ies) around the best-fit *WMAP* or *Planck* cosmologies of that time (e.g. Prada et al., 2012; Diemer & Kravtsov, 2015). The calibrated fitting functions cannot be simply extended beyond the cosmological models they have been tested for. In order to overcome the many shortcomings of extrapolating fitting functions to a wider range of cosmologies, we try to build emulators for the halo concentration-mass relations and density profiles.

The FORGE simulations were carried out using the high-fidelity simulation code MG-Arepo (Arnold et al., 2019b) with a Tree Particle-Mesh algorithm to calculate the gravitational forces and a small force-resolution, $15\,h^{-1}{\rm kpc}$, which enable to reliably resolve the internal structure of haloes. We consider only the haloes containing more than $1,000$ particles, corresponding to $9.1 \times 10^{12}\,h^{-1}M_\odot$ for the fiducial *Planck* cosmology. We did not exclude unrelaxed haloes that contain a large amount of substructure as in some works (e.g. Neto et al., 2007; Prada et al., 2012; Klypin et al., 2016) for several reasons. First, we aim at predicting the halo profile as an ingredient of the halo model, instead of studying the formation of haloes theoretically. Galaxies are expected to reside in all haloes, instead of in relaxed ones only. Second, excluding unrelaxed haloes would bias the concentration high because haloes in the rapid mass accretion stage tend to have low concentrations. Third, such a cut removes typically 30-50% haloes, which would make the derived halo statistics less reliable, since the box size of the FORGE simulations and therefore the size of the halo samples are relatively small.

Now we describe the measurement of the halo concentration. As mentioned in Section 4.4.1, we use M_{200c} and R_{200c} as the definitions of the halo mass and radius, respectively. The halo centre is adopted as the gravitational potential minimum. The halo particles are split into 20 logarithmically radial bins, uniformly covering the range $[0.05, 1.00]R_{200c}$. We then fit NFW profile Equation (4.77) to the the density in the radial bins of each halo, by minimising the χ^2 defined as

$$\chi^2 = \sum_{i=1}^{N_{\rm bin}} [\log_{10}\rho_{\rm sim}(r_i) - \log_{10}\rho_{\rm NFW}(r_i)]^2, \tag{4.81}$$

where $N(r_i)$ is the number of particles in the i-th radial bin. Note that Equation (4.81) assigns equal weight to each bin. We have explicitly checked that the best-fitting concentration parameters are not sensitive to other plausible choices, such as Poisson weighting.

The mean value of the best-fit concentration depends on technical details such as the halo finders, the number and radial range of bins and/or the χ^2 function that is minimised. These details would have a large impact if the NFW profile is not a good fit to the halo profile (Meneghetti & Rasia, 2013; Dooley et al., 2014). It was argued that using the median instead of the mean concentration would relieve such non-convergence (Diemer & Kravtsov, 2015).

To perform test the convergence, we use three N-body simulations from Mitchell et al. (2021) for the fiducial *Planck* cosmology with fixed particle number 1024^3 and three box sizes, $200, 500$ and $1000\,h^{-1}{\rm Mpc}$. We then bin the halo particles, fit the NFW profile and calculate the mean and median value of the concentrations in each mass bin, keeping the maximum radius fixed at $1.00R_{200c}$ and checking with four $r_{\rm min}$ values: $(0.05, 0.07, 0.10$ and $0.13)R_{200c}$. The results are shown in Fig. 4.6. The median concentrations (as well as the mean) are still sensitive to the minimum radius, with a relative difference up to 10%.

The convergence test also confirms our choice of the particle-number cut of haloes. The concentrations from the lower resolution simulations start deviating from those of the higher resolution runs around the halo mass corresponding to ~ 800 times the particle mass. This pivot mass also depends weakly on the radial range, with smaller minimum scales having larger pivot masses.

The concentration measurements do not converge, indicating that it is not the optimal way to describe the halo density profiles. To complete this study, we still present the emulator for the $c(M)$ relation of the FORGE simulations, choosing $r_{\rm min} = 0.10R_{200c}$ as representative. The emulator was constructed following the same procedure as that of the HMF as described in the last subsection. The performance of the emulator for concentration-mass relations at $z = 0$ in

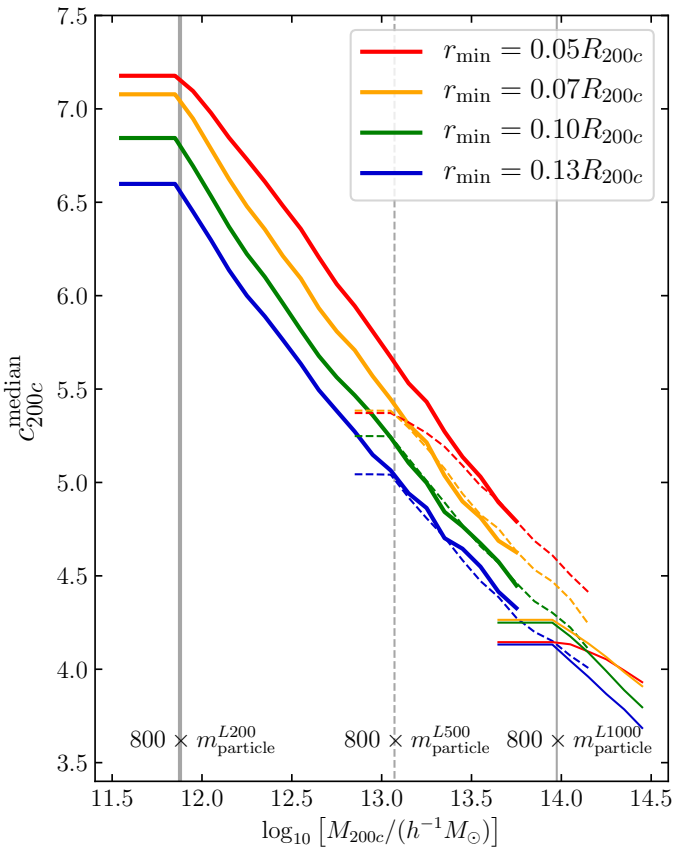

Figure 4.6: The convergence test of the halo concentration measurement. We fit the NFW profile and measure the halo concentration from three N-body simulations with the same particle number and different box sizes, $L200$ (the highest resolution run, thick solid lines), $L500$ (the medium run, thin dashed lines) and $L1000$ (the lowest one, thin solid lines). Colours denote the different minimum fitting range in the fitting, with the maximum radial range fixed at $1.00R_{200c}$. The vertical lines present the pivot mass scales smaller than which the concentration measurements are unreliable due to the limited resolution, which equal to ~ 800 times the simulation particle mass as long as $r_{\rm min} > 0.05R_{200c}$.

$f(R)$ gravity is shown in the left panel of Fig. 4.7. The fractional errors are sub-percent for most of the 50 cosmologies and smaller than the 2% level for all. This level of precision in the $c(M)$ predictions translates into $\lesssim 2\%$-level errors of the NFW profiles, as shown in the right panel of Fig. 4.7, where we compare the NFW profiles calculated using the $c(M)$ relations from the simulation measurements and emulator outputs, respectively.

4.4.3.2 The averaged halo profile estimated from the halo-mass correlation function

The (normalised) halo density profile $u(r|M)$ that appears in the halo model does have a direct counterpart in N-body simulations. As shown in Section 4.1.2, $u(r|M)$ can be estimated from the halo-mass cross-correlation function. The halo profile $u(r|M)$ is directly related to the matter density field cross-correlated with a halo sample with a narrow mass range $[M, M + \Delta M]$. However, such cross-correlations would be noisy because of the low halo number density. To feed

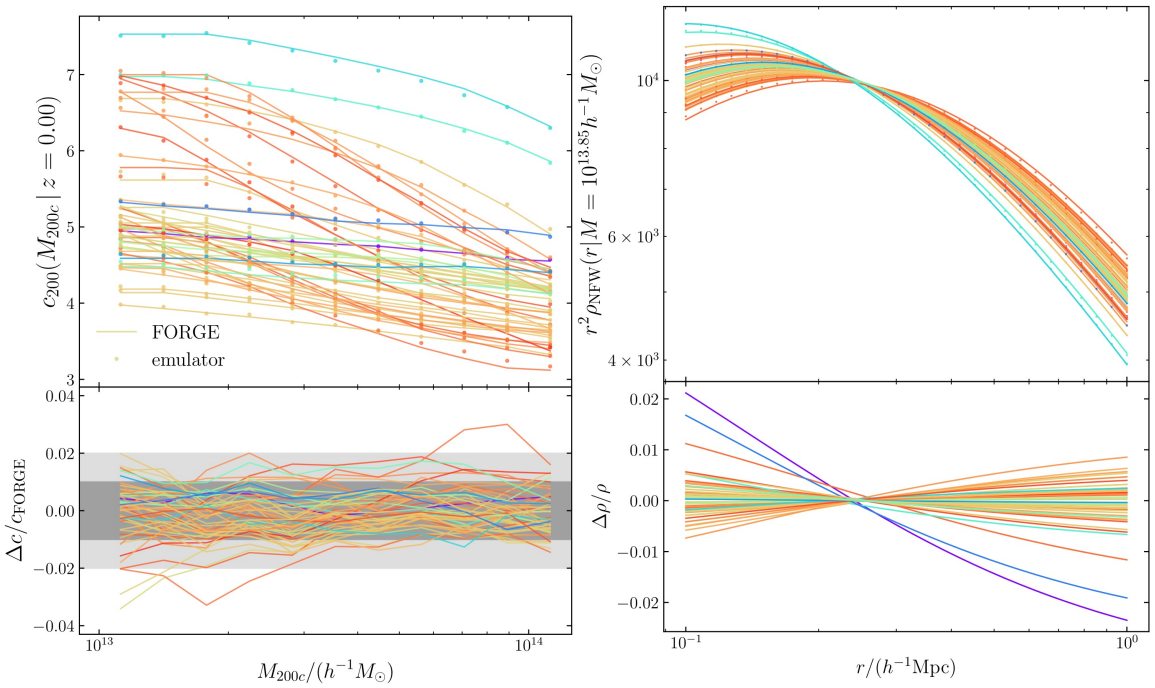

Figure 4.7: **Left** panel: The upper subpanel shows the halo concentration-mass emulator at $z = 0$ for 50 FORGE $f(R)$ gravity cosmologies. The fractional errors of the emulator outputs are presented in the lower subpanel, with the dark and light grey bands denoting $\pm1\%$-and $\pm2\%$-levels of difference. **Right** panel: The NFW profiles with mass $M = 10^{13.85}\,h^{-1}M_\odot$ for the 50 cosmologies, using the $c(M)$ relations measured from simulations (dots) and emulators (lines). The relative differences of the two groups of profiles are shown in the lower subpanel.

the neural networks with the smoother functions, we measure $\xi_{\rm hm}(r|\boldsymbol{\mathcal{C}}, \bar{n}_{\rm h})$ a function of the halo number density, $\bar{n}_{\rm h}$. We then use HMFs to translate $\xi_{\rm hm}(r|\boldsymbol{\mathcal{C}}, \bar{n}_{\rm h})$ as a function of number density into $\xi_{\rm hm}(r|\boldsymbol{\mathcal{C}}, M)$ as a function of differential mass using Equation (4.31). Furthermore, to reduce the dynamic range of the data vector, we opt to emulate $r^2\xi_{\rm hm}(r)$ instead of $\xi_{\rm hm}(r)$ itself. We construct emulators for with the fixed values of the following halo number densities,

$$\log_{10}\left[\frac{\bar{n}_{\rm h}}{(h^{-1}\,{\rm Mpc})^{-3}}\right] = -2.80, -2.90, -3.00, \ldots, -5.50.$$

The separation r is split into 30 logarithmic bins from 0.05 (three times the force resolution) to $3\,h^{-1}$Mpc bins. The upper left subpanel of Fig. 4.8 shows $\xi_{\rm hm}(r|\bar{n}_{\rm h} = 10^{-2.9}\,(h^{-1}\,{\rm Mpc})^{-3})$ at $z = 0$ for the 50 FORGE cosmologies.

The halo-mass correlation enters the transition between 1- and 2-halo terms as the scale increases. To estimate the range of the one-halo term, we only consider the scales below 1.5 times the halo virial radius, defined as

$$R_{\rm vir} = 0.163\,h^{-1}{\rm kpc}\left(\frac{M_{\rm vir}}{10^{12}\,h^{-1}M_\odot}\right)^{1/3}\left(\frac{\Delta_{\rm vir}}{200}\right)^{-1/3}(\Omega_{\rm m0})^{-1/3}(1+z)^{-1}, \tag{4.82}$$

where the virial overdensity is calculated from the non-linear spherical collapse model in the

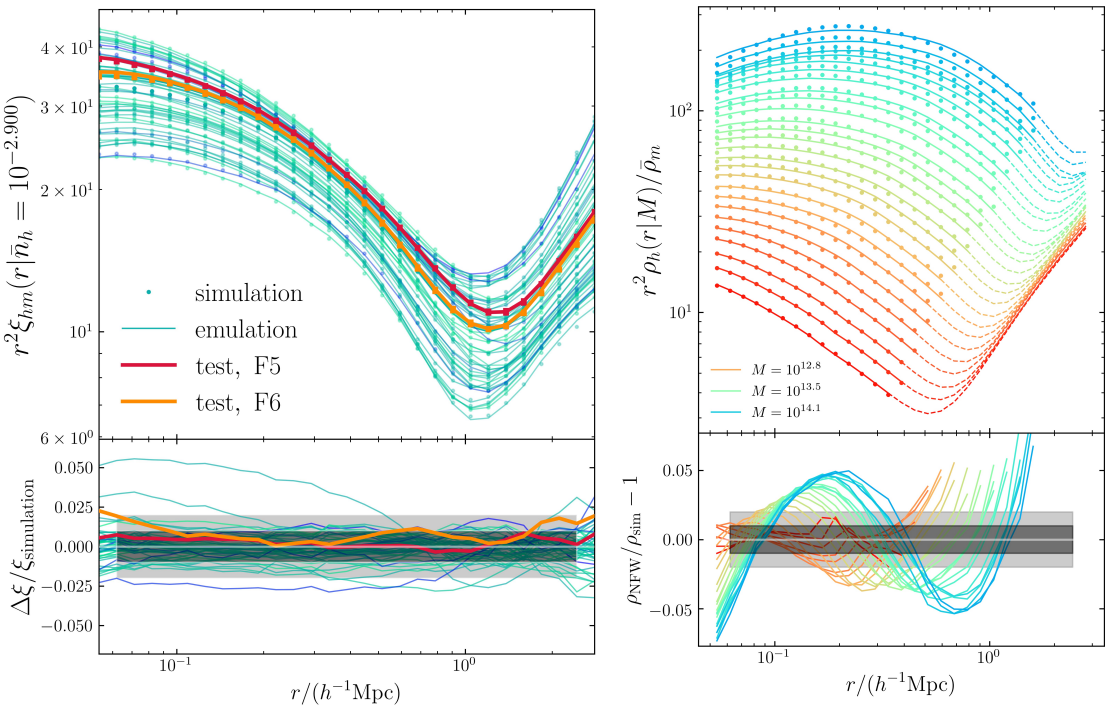

Figure 4.8: **Left** panel: The halo-mass cross-correlation functions from simulation measurements and emulator outputs at $z = 0$, for the 50 FORGE $f(R)$ gravity cosmologies (green) and two independent test models F5 (red) and F6 (orange). In the lower subpanel, we compare their relative differences. The dark and light grey bands denote $\pm 1\%$- and $\pm 2\%$-level errors, respectively. **Right** panel: The averaged halo density profiles derived from $\xi_{\rm hm}(r)$ for a random FORGE cosmology, using the simulation data (lines) and emulator predictions (dots). Colours represent halo masses as shown in the legend. Only the halo-mass correlation in the radial range lower than 1.5 times the halo virial radius (see the text for details) are identified as the halo profiles (shown in the solid lines). In the lower right panel, the accuracy of the NFW fitting is presented.

ΛCDM model as (Bryan & Norman, 1998a)

$$\Delta_{\rm vir}(z) = 18\pi^2 + 82[\Omega_{\rm m}(z) - 1] - 39[\Omega_{\rm m}(z) - 1]^2, \tag{4.83}$$

and we apply this estimation to $f(R)$ gravity models as well, since this is a simple and conservative approximation of the dominant scale range of the 1-halo term. The upper right panel of Fig. 4.8 presents the derived halo density profiles, where the solid lines are in the radial range smaller than $1.5R_{\rm vir}$ and treated as the profiles, and the dashed ones are not used.

We then build emulators for $\xi_{\rm hm}(r|\boldsymbol{\mathcal{C}}, \bar{n}_{\rm h}, z)$ for each halo number density and redshift, to reduce the number of features and minimise emulation errors. The lower left subpanel of Fig. 4.8 shows the fractional difference between the simulation measurements and the emulator predictions, for the FORGE simulations along with two test models F5 and F6. The emulator achieves sub-percent accuracy for the independent test models, which is not surprising since they are located near the centre of the parameter space. Only three FORGE cosmologies have errors larger than 2%, and we have checked that all of these models have at least one cosmological parameter near

the boundary of the parameter space.

We have also explored how well the NFW profile fits the averaged halo density profiles. The relative differences between the density profiles from the best-fit NFW and emulator predictions are shown in the lower right subpanel of Fig. 4.8 for a random FORGE $f(R)$ gravity cosmology (the results for the other cosmologies are similar). The NFW fits are accurate to better than 5% in the radial range studied.

4.5 *Tiro Finale (incompiuta)*: the assembly of the model ingredients

The relatively small box size of the FORGE simulations implies a larger sample-variance noise for halo statistical properties, such as halo auto-correlation functions and pairwise velocity moments. Neural networks are prone to overfitting problems especially when the training data have a low signal-to-noise ratio. Therefore, the emulators for the halo statistics in the MG models are still not available. However, we can demonstrate the feasibility of the emulator-based halo model by calculating the galaxy two-point correlation functions with the model ingredients provided by the publicly available Python module Dark Emulator (Nishimichi et al., 2019a). We keep the code structure highly modularised so that the model ingredients can be easily replaced to the modified gravity emulators when they are ready.

The Dark Emulators are based on the Dark Quest simulations (Nishimichi et al., 2019a), which were performed with 2048^3 dark matter particles in $1\,h^{-1}\mathrm{Gpc}$ or $2\,h^{-1}\mathrm{Gpc}$ side-length boxes, using the GADGET-2 N-body solver (Springel, 2005). The cosmologies used in the simulations cover 101 flat geometry wCDM models, where the equation of state for dark energy is assumed to be constant w. The set of cosmological parameters is defined using optimal maximin distance sliced Latin hypercube designs (Ba et al., 2015), which enable efficient sampling from the six-dimensional parameter space,

$$\mathcal{C} = \left\{\omega_{\rm b0}, \omega_{\rm c0}, \Omega_{\rm de}, \ln\left(10^{10} A_{\rm s}\right), n_{\rm s}, w\right\}, \tag{4.84}$$

where $\omega_{\rm b0} \equiv \Omega_{\rm b0} h^2$ and $\omega_{\rm c0} \equiv \Omega_{\rm c0} h^2$ are the physical density parameters of baryons and cold dark matter, respectively; $A_{\rm s}$ and $n_{\rm s}$ are the amplitude and slope of the primordial curvature power

spectrum normalised at $0.05\,\mathrm{Mpc}^{-1}$. The range of parameters explored is

$$\begin{aligned}
&0.0211375 < \omega_{\rm b0} < 0.0233625,\\
&0.10782 < \omega_{\rm c0} < 0.13178,\\
&0.54752 < \Omega_{\rm de} < 0.82128,\\
&2.4752 < \ln\left(10^{10} A_{\rm s}\right) < 3.7128,\\
&0.916275 < n_{\rm s} < 1.012725,\\
&-1.2 < w < -0.8,
\end{aligned} \tag{4.85}$$

which is centred on the fiducial best fitting ΛCDM model to the *Planck* 2015 data (Planck Collaboration et al., 2016b): $\omega_{\rm b0} = 0.02225, \omega_{\rm c0} = 0.1198, \Omega_{\rm de} = 0.6844, \ln\left(10^{10} A_{\rm s}\right) = 3.094, n_{\rm s} = 0.9645$ and $w = -1$.

We generated mock galaxy catalogues that mimic the LOWZ SDSS-like galaxies (Dawson et al., 2013) based on the fiducial *Planck* cosmology, following Kobayashi et al. (2020). We populate galaxies into the haloes in the 25 realisations of the fiducial Dark Quest simulations. We adopt the HOD prescription of Zheng et al. (2005b), as described in Section 4.1.3. The model specifications of the mock samples are given by

$$\begin{aligned}
\bar{z} = 0.251, \quad \bar{n}_{\rm g} = 2.174 \times 10^{-4}\,(h^{-1}\mathrm{Mpc})^{-3},\\
\log_{10}\left(\frac{M_{\rm min}}{h^{-1}M_\odot}\right) = 13.62, \quad \sigma_{\log M} = 0.6915,\\
\log_{10}\left(\frac{M_1}{h^{-1}M_\odot}\right) = 14.42, \quad \alpha = 0.9168, \quad \kappa = 0.51.
\end{aligned} \tag{4.86}$$

The right panel of Fig. 4.9 shows the mean occupation numbers of central and satellite galaxies.

In the left panel of Fig. 4.9, we compare the galaxy real-space TPCFs calculated using the emulator-based halo model with Dark Emulator, and the 1-loop Lagrangian perturbation theory (Chen et al., 2021) using the publicly available code velocileptors. Note that since we are only looking at the clustering in real space, we do not include the one-loop effective field theory counter-terms in the analysis. The figure clearly shows that the emulator-based approach fairly well reproduces the mock measurements over all scales within 1% accuracy, while the perturbation theory fails on scales smaller than $25\,h^{-1}\mathrm{Mpc}$. There are fluctuations around $2\,h^{-1}\mathrm{Mpc}$, which overlaps with the transition scale between the two-halo and one-halo term. This is a well-known defect of the halo model approach (e.g. Mead et al., 2021).

We aim at constructing emulators for halo statistical properties in modified gravity models. The high-fidelity MG simulation codes such as ECOSMOG and MG-Arepo are too expensive to

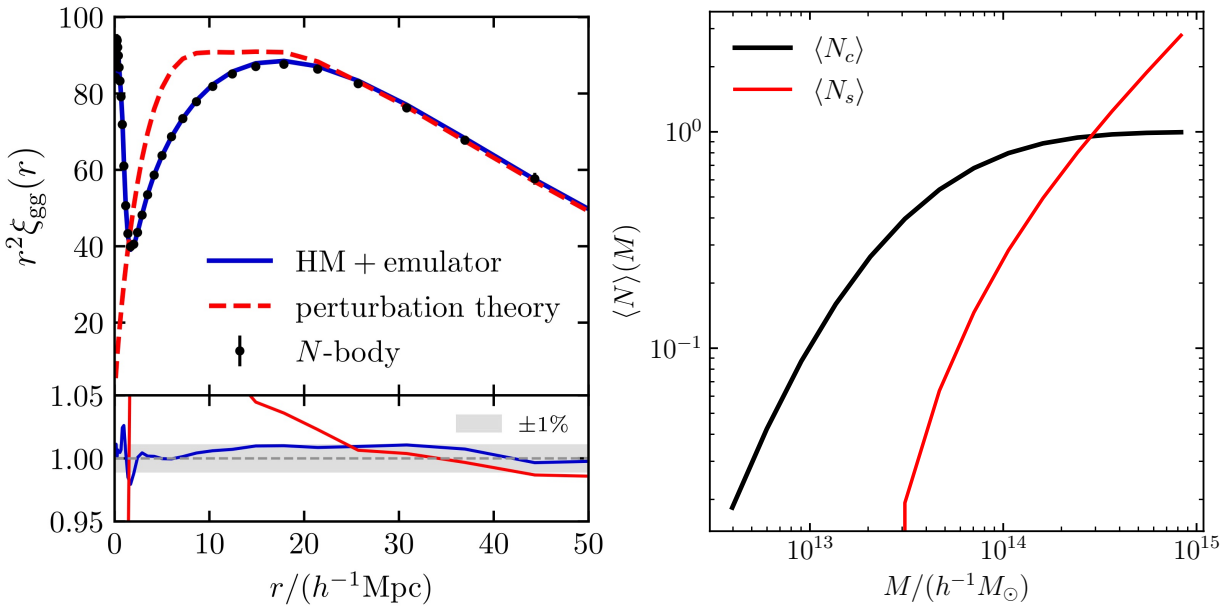

Figure 4.9: **Left** panel: The real-space two-point correlation function obtained by populating with mock galaxies the 25 realisations of the Dark Quest simulations for the fiducial *Planck* cosmology (black dots), compared with the predictions from the emulator-based halo model (solid blue line) and the Lagrangian perturbation theory (dashed red line). The fractional errors of the model predictions are shown in the lower subpanel, with the grey band denoting ±1% difference. **Right** panel: The mean occupation numbers of central and satellite galaxies as a function of halo masses, which mimics the LOWZ sample of the SDSS-III survey.

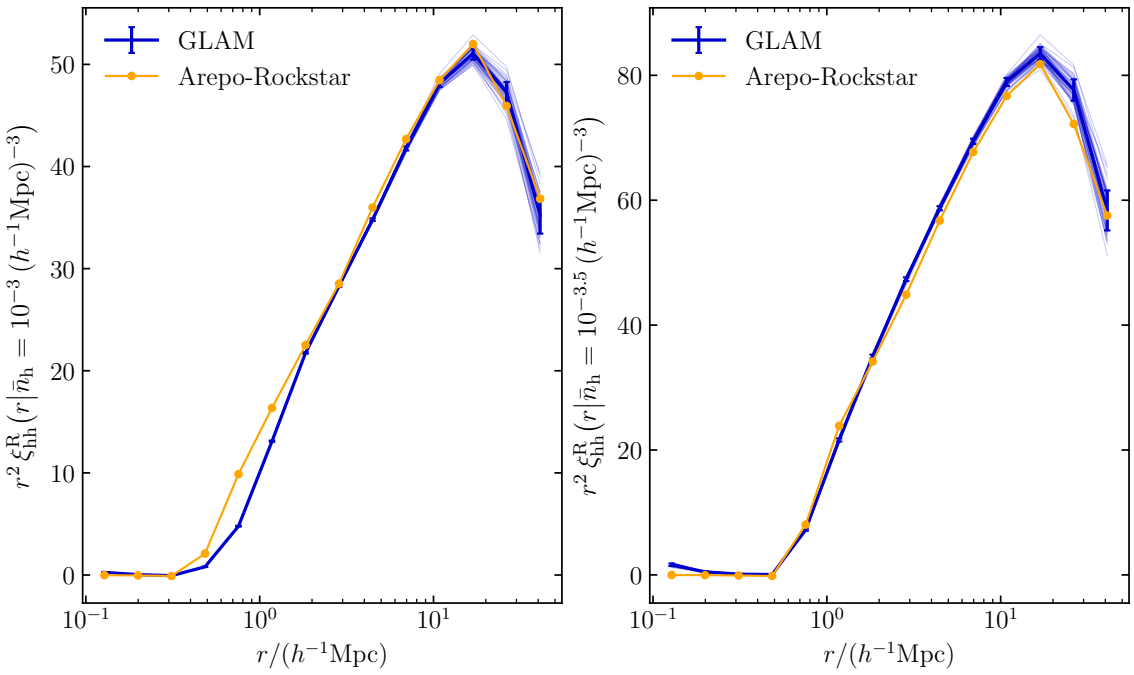

Figure 4.10: The comparison of the halo auto correlation functions from one MG-Arepo simulation (orange) and 50 MG-GLAM realisations of simulations (blue) at $z = 0.25$, for two halo number densities $\bar{n}_{\rm h} = 10^{-3.0}$ and $10^{-3.5}\,(h^{-1}{\rm Mpc})^{-3}$.

perform large-box simulations while keeping high mass resolutions. In Fig. 4.10, we compare the halo two-point correlation functions at $z = 0.25$ with two halo number densities, from an MG-Arepo simulation with the box size $500\,h^{-1}\mathrm{Mpc}$, and 50 realisations of MG-GLAM with $L_{\mathrm{box}} = 1024\,h^{-1}\mathrm{Mpc}$, for the same $f(R)$ gravity model with the *Planck* cosmological parameters and $\bar{f}_{R0} = -10^{-5}$. The halo catalogues were obtained with two different halo finders: BDM for MG-GLAM and Rockstar for MG-Arepo. But the adopted halo mass definitions are the same: the halo virial mass, where the spherical overdensity of haloes comes from the non-linear spherical collapse, instead of 200 times the critical density as in $M_{200\mathrm{c}}$. The two groups of halo TPCFs are in good agreement with each other. This enables us to combine the fast and high-fidelity simulation codes, as well as the different resolution and box size runs, to measure all the halo model ingredients accurately and economically.

Chapter 5

Conclusions and Outlook

5.1 Conclusions

This thesis presents an *emulator*-based halo model approach to model galaxy redshift-space clustering statistics down to non-linear scales, for a wide range of MG models of current interest. The key aspects of this approach include the following points, which are also summarized in the pipeline shown in Fig. 5.1.

- A fast and full N-body simulation code **MG-GLAM** (Ruan et al., 2022; Hernández-Aguayo et al., 2022). MG-GLAM is able to reproduce the predictions of well-tested and high-fidelity codes such as MG-Arepo (Arnold et al., 2019a) and ECOSMOG (Li et al., 2012) with sufficiently high accuracy, whilst taking only a small fraction of the computation time. (**Chapter** 2)

- A well-established *halo model* framework for non-linear galaxy clustering, combined with a *halo occupation distribution (HOD)* prescription as the galaxy-halo connection model, and a *streaming model of redshift-space distortions* accounting for the effects of peculiar velocities. The observational galaxy statistics, such as galaxy redshift-space two-point correlation functions and power spectra, are then expressed as combinations of the basic dark matter halo properties and a user-adopted HOD model. (**Chapters** 3 and 4)

- A parameter sampling algorithm in high-dimensional cosmological parameter space, **maximin distance sliced Latin hypercube designs** (Ba et al., 2015; Bouhlel et al., 2019) and a statistical technique to perform multidimensional regression — fully connected **neural networks** (Paszke et al., 2019). The combination of the efficient sampling scheme and the fitting algorithm enables *sub-percent accuracy* interpolation with only 50 models in a seven-

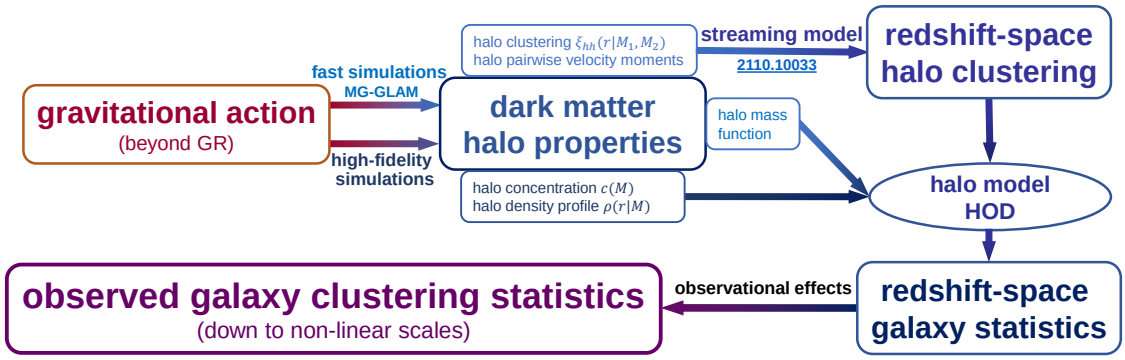

Figure 5.1: Schematic outline of the prediction of observed galaxy statistics (such as two-point correlation functions) in modified gravity theories. We start with a modified gravity of interest represented by its gravitational action, construct simulation-based theoretical templates or emulators for basic halo properties, and combine with a specific halo occupation distribution (HOD) prescription in the halo model framework.

dimensional parameter space, provided that the mappings between the inputs and outputs are noise-free, as shown in the ideal emulation tests presented in **Section 4.3**.

- **Accurate emulators** for the halo model ingredients, including the halo mass function (**Section** 4.4.2), the halo density profile (**Section** 4.4.3), the correlation function and peculiar velocity moments of haloes.

5.2 Outlook

We still lack high signal-to-noise ratio measurements of the halo statistics, including the halo auto-correlation function and halo pairwise velocity moments, from modified gravity simulations. To suppress the noise level, a large number of haloes and, therefore, a large simulation box size are required. The high-fidelity modified gravity N-body codes, such as MG-Arepo, are too expensive to perform such large box simulations while keeping the same resolution as in the currently available FORGE suite of simulations. The fast and cheap N-body code MG-GLAM could help, as shown in Fig. 4.10, where we compared the halo two-point correlation functions from the MG-GLAM and MG-Arepo simulations and found a good consistency. We plan to combine the fast and high-fidelity simulation codes, as well as the different resolution and box-size runs, to measure all the halo model ingredients accurately and economically.

In the future, we will apply this emulator-based pipeline to perform the cosmological and modified gravity parameter inference to constrain two of the most representative MG models, $f(R)$ gravity and the DGP braneworld model, by comparing the theoretical predictions with the measurements from the upcoming galaxy survey data, such as DESI and *Euclid*. In the mean time,

we plan to extend the analysis to other viable MG models such as symmetron, K-mouflage and coupled quintessence, which have been implemented in MG-GLAM. We expect this research to play an important role in constraining modified gravity models and testing the law of gravity on cosmic scales. We hope to reveal or exclude the modified gravity theories under consideration in the next few years.

Appendix A

Effects of mass and force resolution in MG-GLAM simulations

The original GLAM code has been well tested (cf. (Klypin & Prada, 2018)) by examining the effects of time-stepping and force resolution and comparing with the high-resolution MultiDark simulations (Klypin et al., 2016) which were performed using the GADGET code (Springel, 2005). In addition, Klypin et al. (2021) compared the GLAM results of halo mass functions and matter power spectra with those of the QUIJOTE simulations (Villaescusa-Navarro et al., 2020), and found good agreement. Denote $k_{1\%}$ as the wavenumber above which the GLAM matter power spectrum begins to deviate by more than 1% from those of the high-resolution simulations. Based on the comparison with the MultiDark simulations, the authors found that $k_{1\%}$ is related to the force resolution $\Delta x = L_{\rm box}/N_{\rm g}$ (cf. Eq. (2.61)) as

$$k_{1\%} = \frac{0.25 \pm 0.05}{(\Delta x)/(h^{-1}{\rm Mpc})}\, h\,{\rm Mpc}^{-1}\,. \tag{A.1}$$

If this relation also works for the MG-GLAM code, the power spectra of the modified gravity cosmological runs presented in the main text are reliable down to $k_{1\%} \sim 1\,h\,{\rm Mpc}^{-1}$. However, as mentioned in the main text above, the results of the power spectrum enhancement with respect to ΛCDM are empirically reliable down to larger k.

To test the effect of mass and force resolutions, we have performed four $f(R)$ gravity simulations for $f_{R0} = -10^{-5}, n = 1$ with fixed box size $512\,h^{-1}{\rm Mpc}$ and varying particle and grid numbers

$$(N_{\rm p}, N_{\rm g}) = \left\{ \begin{matrix} (1024, 2048), & (2048, 2048) \\ (1024, 4096), & (2048, 4096) \end{matrix} \right\}\,,$$

which correspond to mass and force resolutions of

$$\left(\frac{m_{\rm p}}{10^9\,h^{-1}M_\odot}, \frac{\Delta x}{h^{-1}{\rm Mpc}}\right) = \left\{\begin{matrix}(11.0, 0.25\ ), & (1.37, 0.25\ ), \\ (11.0, 0.125), & (1.37, 0.125)\end{matrix}\right\}\ .$$

Here the two runs in the same row (column) have the same force/mesh (mass) resolution. The adopted cosmological parameters are the same as the simulations used in the main text.

We compare the MG-GLAM matter power spectrum and halo mass function enhancement $\Delta P/P_{\rm GR}$ and $\Delta{\rm HMF}/{\rm HMF}_{\rm GR}$ with those of the MG-AREPO simulations. We focus on these quantities instead of comparing the absolute $P(k)$ and HMF from (MG-)GLAM and other codes, for the following reasons: (1) as mentioned above, the reliability of the ΛCDM results from GLAM has been carefully tested and established; (2) comparisons between different codes usually suffer from cosmic variance and different implementation details (such as the initial condition set up, force calculation, time stepping and halo finding), and as a result a large number of runs are needed to make reliable comparisons, after carefully calibrating simulation specifications of the different codes — such an effort is unnecessary and beyond the scope of this work given (1); (3) in MG simulations, people are often more interested in the enhancement with respect to ΛCDM, and this is indeed what has been tested in the code papers of the previous MG simulation codes. Also, we note that the MG-GLAM and MG-AREPO simulations presented in this work use slightly different cosmological parameters: we have checked explicitly (by running test simulations with MG-GLAM using identical cosmological parameters as the MG-AREPO runs) that the effect is small (few percent level), but this nevertheless still makes it difficult to justify directly comparing the absolute $P(k)$ or HMF from them; the enhancement, on the other hand, is known empirically to be less sensitive to cosmological parameter values and differences between simulation codes.

The left panel of Fig. A.1 presents the matter power spectrum enhancements at $z = 0$ from MG-GLAM and MG-AREPO, as well as two MG-GADGET simulations. We see that $\Delta P/P_{\rm GR}$ is relatively insensitive to the mass and force resolution variations considered here; this is consistent with previous experiences. However, increasing the mesh resolution from 0.25 to $0.125h^{-1}{\rm Mpc}$ does improve the agreement between MG-GLAM and MG-AREPO, by reducing $\Delta P/P_{\rm GR}$ (see, e.g., Li et al. (2013a) for a discussion of how a lower resolution simulation gives higher $\Delta P/P_{\rm GR}$). The highest resolution MG-GLAM run ($N_{\rm p} = 2048$ and $N_{\rm g} = 4096$) agrees with MG-AREPO nearly perfectly down to $k \sim 1\,h\,{\rm Mpc}^{-1}$, and the agreement is at the level of a couple percent down to $k \approx 5h{\rm Mpc}^{-1}$ (ignoring the dip in $\Delta P/P_{\rm GR}$ at $k \approx 4h{\rm Mpc}^{-1}$, which is apparently not physical). The slightly larger deviations at $k > 1\,h\,{\rm Mpc}^{-1}$ can be still due to the lower force resolutions in the GLAM simulations, but we note that the agreement between the MG-GADGET

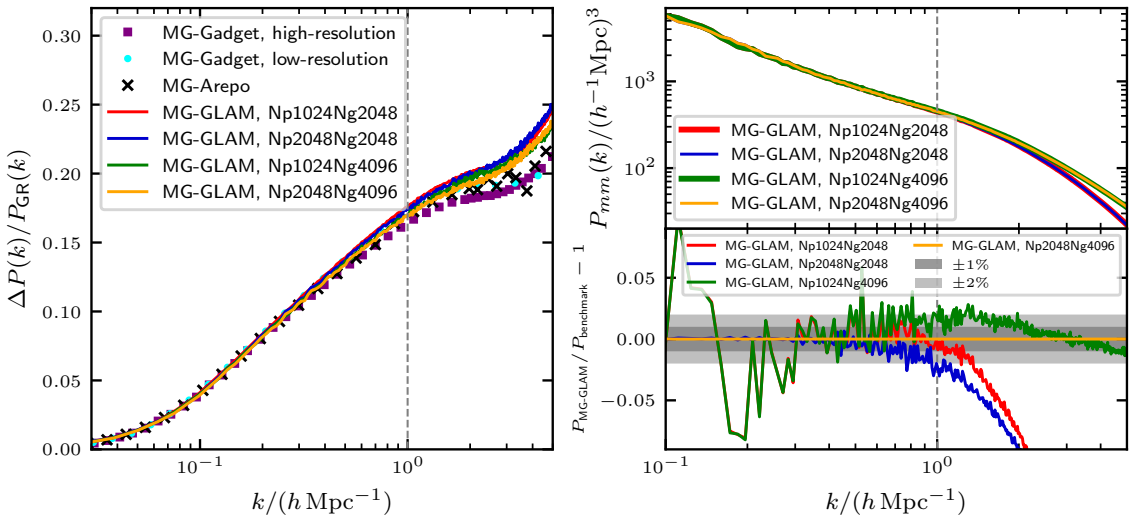

Figure A.1: Comparison of matter power spectra from MG-GLAM (lines), MG-GADGET (squares and circles) and MG-AREPO (crosses) simulations at $z = 0$. The left panel shows the matter spectrum enhancement, $\Delta P/P_{\rm GR}$, from the different codes and resolutions, as the legend labels. The MG-AREPO data are the same as in Fig. 2.7, while the MG-GADGET data are from the two LIGHT-CONE simulations, at higher ($L = 768h^{-1}$Mpc and $N_{\rm p}^3 = 2048^3$) and lower ($L = 1536h^{-1}$Mpc and $N_{\rm p}^3 = 2048^3$) resolutions, described in Arnold et al. (2019c). The upper right panel shows the absolute values of $P(k)$ from MG-GLAM simulations with the four combinations of mass and force resolutions. In the lower right panel, the ratios of the power spectrum in each simulation to that of the highest resolution run ($N_{\rm p} = 2048$ and $N_{\rm g} = 4096$) are displayed, where the dark and light grey shaded regions denote respectively $\pm1\%$ and $\pm2\%$ differences from the benchmark. The vertical lines represent $k = 1h$Mpc^{-1}.

and MG-AREPO runs (which have similar force resolutions) is at a comparable level, so the difference is likely also partly due to the different codes (or simulation realisations).

In the upper right panel of Fig. A.1, we present the absolute matter power spectra from the MG-GLAM simulations at different resolutions. As expected, increasing the mesh/force resolution leads to a $P(k)$ curve that decays much more slowly at small scales (orange and green lines), while increasing the mass resolution (blue) gives little improvement. The lower right panel of Fig. A.1 shows the ratio of the matter spectrum in each simulation to that from the highest resolution run. The figures indicate $\approx 1\%$ convergence for $k \lesssim 1\,h\,{\rm Mpc}^{-1}$, which is consistent with the convergence test of the original GLAM code presented in (Klypin & Prada, 2018).

The comparisons of halo mass functions are shown in Fig. A.2, where note that we used different halo finders for the MG-GLAM and MG-AREPO results, but the same halo mass definition, as described in Sect. 2.5.4. The HMFs of MG-GLAM simulations are accurate in the range of $M_{\rm vir} \gtrsim 10^{12.5}\,h^{-1}M_\odot$ for $\Delta x = 0.25\,h^{-1}$Mpc ($N_{\rm g} = 2048$), and $M_{\rm vir} \gtrsim 10^{12}\,h^{-1}M_\odot$ for $\Delta x = 0.125\,h^{-1}$Mpc ($N_{\rm g} = 4096$). There is excellent agreement between MG-GLAM's higher-resolution runs and MG-AREPO, in both the HMF and its enhancement, down to $10^{12}h^{-1}M_\odot$.

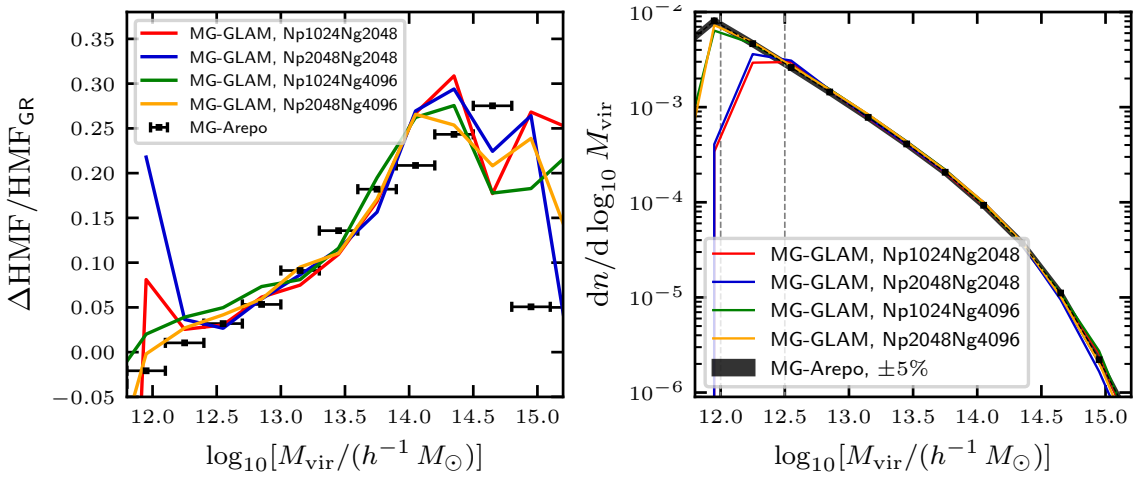

Figure A.2: Comparison of halo mass functions of MG-GLAM and MG-AREPO simulations at $z = 0$. The relative enhancements with respect to ΛCDM and the absolute values of the HMFs are shown in the left and right panels, respectively. The two vertical lines in the right panel denote respectively the masses 10^{12} and $10^{12.5}h^{-1}M_{\odot}$. There is excellent agreement between MG-GLAM's higher-resolution runs and MG-AREPO, in both the HMF and its enhancement, down to $10^{12}h^{-1}M_{\odot}$.

Appendix B

Method of moments for the ST distribution

The four parameters of the skew-T distribution, v_c, w, α, ν, can be analytically related to the first four moments. To simplify the relation between these moments and parameters, let us introduce

$$\begin{aligned} b_\nu &= \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma(\nu/2)}, \\ \delta &= \frac{\alpha}{\sqrt{(1+\alpha^2)}}, \\ \gamma_1 &= \frac{c_3}{c_2^{3/2}}, \\ \gamma_2 &= \frac{c_4}{c_2^2}. \end{aligned} \tag{B.1}$$

The moments can then be written as,

$$m_1 = v_c + w\delta b_\nu, \tag{B.2}$$

$$c_2 = w^2 \left[\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right], \tag{B.3}$$

$$\gamma_1 = \delta b_\nu \left[\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\delta^2 b_\nu^2\right] \left[\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right]^{-\frac{3}{2}}, \tag{B.4}$$

$$\begin{aligned} \gamma_2 = &\left[\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\delta^2 b_\nu^2 \nu(3-\delta^2)}{\nu-3} - \frac{6\delta^2 b_\nu^2 \nu}{\nu-2} - 3\delta^4 b_\nu^4\right] \\ &\left[\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right]^{-2}. \end{aligned} \tag{B.5}$$

These form a system of non-linear algebraic equations that can be solved numerically: parameters w, α and ν are obtained from the last three equations given the variance, skewness and kurtosis of

the distribution, and the remaining parameter, v_c, can then directly be obtained from the equation for the mean.

Appendix C

Convergence Tests

The GLAM and MG-GLAM simulations used in this work have a relatively small box size, $512\,h^{-1}\mathrm{Mpc}$. While their mass resolution is high compared to many other MG simulations to date, the particle-mesh nature of the GLAM-based codes means that the force resolution in these runs is poorer than what could be achieved using adaptive-mesh-refinement simulations with the same particle number and box size. In addition, MG-GLAM is a relatively new code and, while it has passed various tests as demonstrated in the code papers, those tests do not include velocity field statistics such as the ones considered in this work. For these reasons, in this Appendix we will carry out a test of the latter, by comparing the first four halo pairwise velocity moments measured from MG-GLAM simulations of F5n1 and GR, with the predictions by a higher-resolution Light-cone simulation for the same models.

The Light-cone simulation project (Arnold et al., 2019d) provides a set of high-resolution cosmological simulations of GR and F5n1, which employs the MG N-body code MG-GADGET (Puchwein et al., 2013), adopting the same Planck15 ΛCDM cosmology. Starting from identical initial conditions, the pair of GR and F5n1 simulation runs followed the dynamical evolution of 2048^3 dark matter particles in a box with $768\,h^{-1}\mathrm{Mpc}$ comoving length, reaching a mass resolution of $M_{\rm particle} = 4.50 \times 10^9\,h^{-1}M_\odot$. This high mass resolution make sure that the measured halo clustering signals are precise on small scales (down to $\sim 1\,h^{-1}\mathrm{Mpc}$). The halo catalogues are obtained with the Subfind (Springel et al., 2001) algorithm. The halo mass definition adopted is $M_{200\mathrm{c}} \equiv \frac{4\pi}{3}(r_{200\mathrm{c}})^3 200\rho_{\rm c}$, where $\rho_{\rm c} \equiv 3H^2/(8\pi G)$ is the critical density of the Universe, and $r_{200\mathrm{c}}$ is the spherical halo radius within which the spherically averaged mass density equals 200 times $\rho_{\rm c}$. The halo catalogues at redshifts $z = 1$ and 0 are available.

The results are shown in Fig. C.1, where the four panels from upper left to lower right are respectively for the first-, second-, third- and fourth-order moments. In each panel, the upper

sub-panel shows the magnitudes of the moments as a function of the halo separation r, while the lower sub-panel shows the relative difference between F5n1 and GR. The line styles and colours are indicated by legends. All results are at $z = 0$.

When reading Fig. C.1, let us bear in mind that the halo populations in the (MG)-GLAM and MG LIGHT-CONE simulations are necessarily different. This suggests that these two sets of simulations can have different halo biases, which would affect the amplitudes of the velocity moments, although the shapes are broadly the same, as can be readily seen. We have not made attempts to force an agreement by adjusting the halo number densities in the different simulations, but we have checked this using several ΛCDM simulations that use different simulation codes, resolutions and halo finders, and found similar levels of discrepancy among all of them.

More interestingly, the lower sub-panels show that the model difference between F5n1 and GR predicted by the two sets of simulations agree very well for all the analysed velocity moments, and all halo separations shown in Fig. C.1. In particular, the first moment, m_{10}, which the redshift-space halo correlation function quadrupole is most sensitive to, agrees between the two sets of simulations down to $\sim 5\,h^{-1}$Mpc. This shows that the MG-GLAM simulation results can be used to study RSD at such small scales.

Fig. C.1 also shows that the enhancements of the velocity moments are different on small scales between the MG-GLAM and LIGHT-CONE simulations. This difference is in most cases at a few percent at most above $\simeq 10\,h^{-1}$Mpc, and — in the case of m_{10} — around the percent level. This level of difference is expected given the many differences in the two sets of simulations, from codes to technical specifications, such as details of halo finding. The effect of these simulation/technical differences can also be seen by comparing the values (rather than the model differences) of the moments predicted by the two simulations (the upper sub-panels of each panel).

At $< 10\,h^{-1}$Mpc, we see a stronger discrepancy between the two simulations for c_{30} and c_{40}, at $\simeq 10\%$ (for c_{04} the agreement is much better — see the green/purple curves in the lower right panel of Fig. C.1). Again, given the more significant differences between the absolute curves of c_{30} and c_{40} for the same models, such as GR, as shown in Fig. C.1 (the green solid and purple solid curves in the upper sub-panel of the lower left panel), this is not surprising. Note in particular that MG-GLAM uses $M_{\rm vir}$ while LIGHT-CONE uses M_{200c} as the halo mass definition.

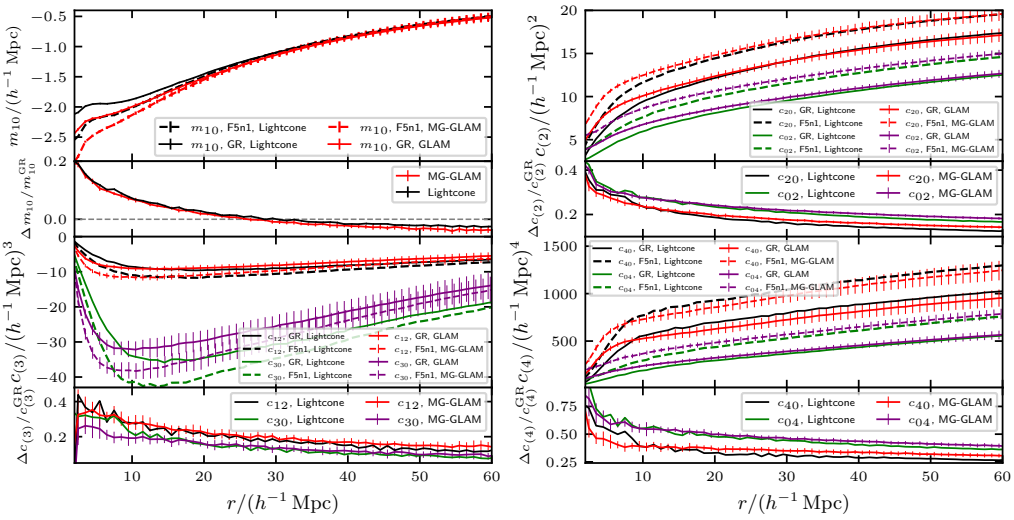

Figure C.1: The four lowest order moments of the radial and transverse halo pairwise velocity for the GR and F5n1 models at $z = 0$, from the LIGHT-CONE project and MG-GLAM simulations. The lower sub-panels show the relative difference between the velocity moments of the F5n1 and GR.

Appendix D

The performance of the skew-T model in more cases

In Fig. 3.11, we have demonstrated that the ST streaming model works very well in predicting the redshift-space correlation function multipoles $\xi^{\rm S}_{0,2,4}$ in not only the GR model, as found by Cuesta-Lazaro et al. (2020), but also for several modified gravity models. However, due to space limit, in that figure we have only presented the results at a single redshift ($z = 0.5$) and around a single halo number density $n_h = 10^{-3.0}\,(h^{-1}{\rm Mpc})^{-3}$.

We have also carried out similar checks for a range of other redshifts and halo number densities, and in all cases we found similarly good agreement between the ST streaming model and simulation predictions of RSD multipoles. A few selected examples are shown in Fig. D.1. The left, middle and right columns are respectively GR $n_h = 10^{-3.5}\,(h^{-1}{\rm Mpc})^{-3}$ at $z = 0$, F5n1 $n_h = 10^{-3.0}\,(h^{-1}{\rm Mpc})^{-3}$ at $z = 0.5$, and F5n1 $n_h = 10^{-3}\,(h^{-1}{\rm Mpc})^{-3}$ at $z = 1$. The three rows are for $\xi^{\rm S}_{0,2,4}$ respectively. In each panel, the upper sub-panel compares simulation measurement (symbols with error bars) with the predictions of the Gaussian (red) and ST (blue) streaming models, and the lower sub-panel shows the relative differences between the two streaming models with respect to the simulation measurement. In all the cases, the ST streaming model clearly gives more reliable predictions than the Gaussian one, indicating that the former can be applied to the modified gravity models studied in this work. There is no apparent reason why we should not expect it to work for other models as well. The performance of the GSM is better in the lower halo number density case, which is expected, since the pairwise velocity PDF becomes more Gaussian for more massive haloes.

Although not shown here, we have also checked the ST streaming model for a few other ΛCDM simulations which were run using different codes, at different resolutions and with different halo

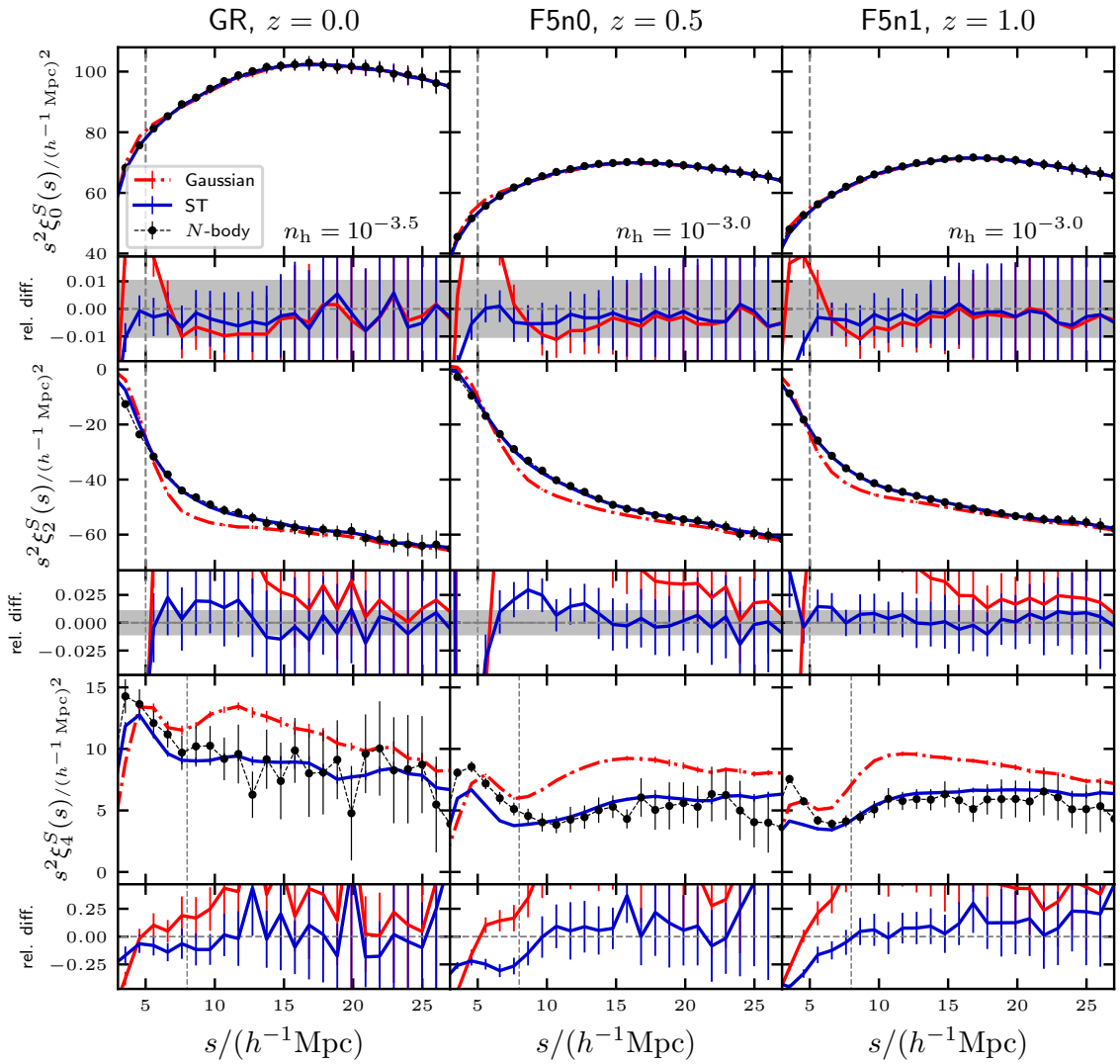

Figure D.1: The monopole, quadrupole and hexadecapole of the redshift-space two-point correlation functions for GR (left column), $f(R)$ gravity model with $\bar{f}_{R0} = -10^{-5}$ and $n = 0$ (F5n0; middle column) and 1 (F5n1; right column), at different redshifts as indicated by the subtitles, from the MG-GLAM simulations (black dots), along with the Gaussian (red lines) and ST (blue lines) streaming model predictions. In the lower sub-panels the relative differences between the SM predictions and the simulation measurements, $\xi^{\rm model}(s)/\xi^{\rm sim}(s) - 1$, are shown. The horizontal dashed lines in the lower sub-panels denote 0, the grey shaded regions denoting $\pm 1\%$ for the monopoles and quadrupoles. The vertical dashed lines indicate where the STSM predictions start to differ significantly from simulation measurements.

finders. In all cases the agreement with simulation prediction is equally good.

Appendix E

Linear perturbation predictions of halo pairwise velocity moments

In this appendix we aim to present a derivation of Equation (3.26), showing that in *linear* perturbation theory the n-th order halo pairwise velocity (central) moments, c_n, scale as β^n, where $\beta \equiv f/b_1$ was introduced in Equation (3.21).

We start from the line-of-sight peculiar velocity difference between points $\boldsymbol{x}$ and $\boldsymbol{x}'$, expressed in terms of the velocity divergence $\theta(\boldsymbol{x}) \equiv \boldsymbol{\nabla} \cdot \boldsymbol{v}(\boldsymbol{x})$,

$$\begin{aligned} v_{\parallel} &= \left[\boldsymbol{v}(\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x}')\right] \cdot \hat{z} = -i \int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \frac{\boldsymbol{k}\cdot\hat{z}}{k^2} \theta(\boldsymbol{k}) \\ &= -i\, aHf \int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \frac{\boldsymbol{k}\cdot\hat{z}}{k^2} \delta_L(\boldsymbol{k}) + \mathcal{O}\left[(\delta_L)^2\right] \\ &= \beta\,(-i\, aH) \int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \frac{\boldsymbol{k}\cdot\hat{z}}{k^2} \delta_h(\boldsymbol{k}) + \mathcal{O}\left[(\delta_L)^2\right] \,, \end{aligned} \tag{E.1}$$

where $\hat{z}$ stands for an arbitrary line of sight, and the integration symbol $\int_{\boldsymbol{k}}$ is a short-hand for $(2\pi)^{-3} \int \mathrm{d}^3\boldsymbol{k}$. In the second and third lines we have used that at linear order we can relate the velocity divergence to the linear matter perturbations as $\theta(\boldsymbol{k}) = aHf\,\delta_L(\boldsymbol{k})$, while the halo overdensity is given by $\delta_h(\boldsymbol{k}) = b_1\,\delta_L(\boldsymbol{k})$, showing that in linear perturbation theory $v_{\parallel}$ is determined by β times a quantity depending only on the halo density field.

Consequently, at leading order the n-th moment of the pairwise velocity PDF,

$$m_n(\boldsymbol{r}) = \frac{\left\langle v_{\parallel}^n \left[1+\delta_h(\boldsymbol{x})\right] \left[1+\delta_h(\boldsymbol{x}')\right] \right\rangle}{1+\xi_{\mathrm{hh}}(r)} \,, \tag{E.2}$$

where $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{x}'$, can be written as β^n multiplied by a term depending on the halo auto power spectrum or correlation function. Explicitly, making use of Equation (E.1) and keeping only the

leading order contributions, we obtain for the first moment:

$$\begin{aligned}[1+\xi_{\rm hh}(r)]\, m_1(\boldsymbol{r}) &\approx 2\left\langle v_\parallel\, \delta_h\right\rangle \\ &= 2\beta\, aH\, \frac{\boldsymbol{r}\cdot\hat{z}}{r} \int_{\boldsymbol{k}} j_1(kr)\, \frac{P_{\rm hh}(k)}{k}\,, \end{aligned} \tag{E.3}$$

whereas the second moment gives

$$\begin{aligned}&[1+\xi_{\rm hh}(r)]\, m_2(\boldsymbol{r}) \\ &\approx (1+\xi_{\rm hh})\, \left\langle v_\parallel^2\right\rangle + 2\left\langle v_\parallel\, \delta_h\right\rangle \left\langle v_\parallel\, \delta_h'\right\rangle \\ &= 2\beta^2\, (aH)^2 \left\{ \left(\frac{\boldsymbol{r}\cdot\hat{z}}{r}\right)^2 \left[(1+\xi_{\rm hh}) \int_{\boldsymbol{k}} j_2(kr) \frac{P_{\rm hh}(k)}{k^2} \right.\right. \\ &\qquad\qquad \left.\left. + \left(\int_{\boldsymbol{k}} j_1(kr) \frac{P_{\rm hh}(k)}{k} \right)^2 \right] \right. \\ &\quad \left. - \frac{1+\xi_{\rm hh}}{3} \int_{\boldsymbol{k}} \left[j_2(kr) + j_0(kr) - 1 \right] \frac{P_{\rm hh}(k)}{k^2} \right\}, \end{aligned} \tag{E.4}$$

and similar relations can be derived for the higher-order moments.

Crucially, because in this work the halo catalogues from different gravity (or, more generally, different cosmological) models have been tuned so that they have the same halo correlation function $\xi_{\rm hh}(r)$ and halo power spectrum $P_{\rm hh}(k)$, these expressions show that, when taking ratios of the pairwise velocity moments from different models all terms involving $\xi_{\rm hh}$ or $P_{\rm hh}$ cancel. This leaves only factors of β, and given two models, A and B, we therefore have

$$\frac{m_{n,A}}{m_{n,B}} = \left(\frac{\beta_A}{\beta_B}\right)^n . \tag{E.5}$$

It is important to stress that this only holds in linear theory and for that reason it is not guaranteed that Equation (E.5) is valid on sufficiently large scales, as it is well known that for instance the large-scale variance receives significant contributions from small-scale virialised motions (Scoccimarro, 2004). For more discussion on this point and how this alters the ratio in Equation (E.5) for even-order velocity moments, see Sec. 3.3.1.2.

Appendix F

Numerical details of the streaming model integration

This appendix presents the numerical details in the computation of the streaming model predictions from the ingredients measured from simulations. The streaming model for the redshift-space TPCF (Equation (3.6)) has two ingredients: the real-space TPCF and the line-of-sight pairwise velocity PDF, i.e. the position and velocity information of tracers.

As mentioned in Equation (3.20), we do not directly use $\mathcal{P}(v_\parallel|\boldsymbol{r})$ in our model predictions, but approximate it with the ST distribution

$$\mathcal{P}_{\rm ST}(v_\parallel|v_c(\boldsymbol{r}), w(\boldsymbol{r}), \alpha(\boldsymbol{r}), \nu(\boldsymbol{r})).$$

The four ST parameters for a given pair separation $\boldsymbol{r}$ can be fixed by the first four line-of-sight pairwise velocity moments m_1, $c_{2\text{-}4}(\boldsymbol{r})$, i.e., by solving the four non-linear algebraic equations, Equations (B.2-B.5). This is done by using the `fsolve` function of the standard open-source `scipy` (Virtanen et al., 2020) library.

The line-of-sight pairwise velocity moments can be obtained by:

- either directly measuring the line-of-sight pairwise distribution $\mathcal{P}(v_\parallel|\boldsymbol{r})$, or
- measuring the two-dimensional pairwise velocity distribution $\mathcal{P}(v_r, v_t|r)$ and projecting its moments along the line of sight according to Equation (3.16).

We prefer the second approach since $\mathcal{P}(v_r, v_t|r)$ takes advantage of symmetries and does not require fixing a particular line-of-sight when using the simulation data. We have checked that the projected line-of-sight moments from these two approaches are in good agreement. Fig. F.1

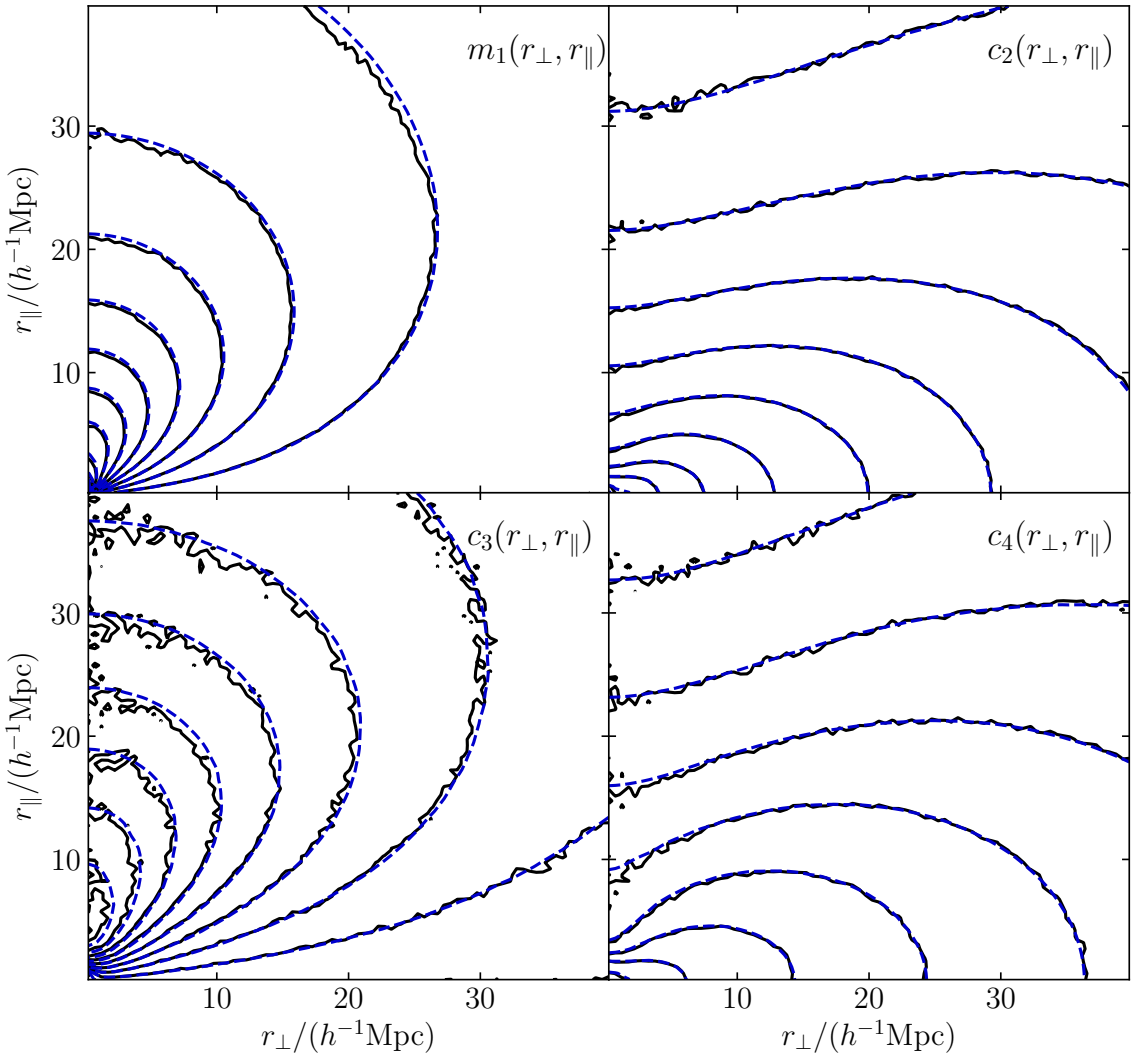

Figure F.1: The first four orders of the line-of-sight halo pairwise velocity moments, from direct measurements (black solid lines) and pairwise velocity moments projection (blue dashed lines). The halo catalogues are from the GLAM simulations with a fixed number density $n_h = 10^{-3.5}\,(h^{-1}\mathrm{Mpc})^{-3}$ at $z = 0.5$.

shows the case of the halo catalogues from GLAM simulations with the number density $n_h = 10^{-3.5}\,(h^{-1}\mathrm{Mpc})^{-3}$ at $z = 0.5$. The ST distribution with the model parameters obtained by this method has been compared with the measured line-of-sight velocity PDF in Fig. 3.7.

In practice, the streaming model numerical integrals can be sensitive to various factors, such as the choice of integration method, the binning scheme etc., and it is important to make sure that one's choices lead to converged results. We have created an example code for this, which can be found here. The example code calculated the redshift-space correlation function multipoles for the GLAM halo catalogues with the number density $10^{-3.0}\,(h^{-1}\mathrm{Mpc})^{-3}$ at $z = 0.25$. For the model ingredients $\xi^{\mathrm{R}}(r)$, $m_{10}(r)$ and $c_{(2)\text{-}(4)}(r)$, we measured them in the separation bins linearly spaced over $1 \leq r/(h^{-1}\mathrm{Mpc}) \leq 120$ with a bin width $\Delta r = 1\,h^{-1}\mathrm{Mpc}$. The optimal configurations depend on the tracers' type (e.g. haloes versus galaxies), number density and redshift, etc.

Bibliography

Agarap A. F., 2018, arXiv preprint arXiv:1803.08375

Agarwal S., Abdalla F. B., Feldman H. A., Lahav O., Thomas S. A., 2012, MNRAS, 424, 1409

Agarwal S., Abdalla F. B., Feldman H. A., Lahav O., Thomas S. A., 2014, MNRAS, 439, 2102

Alam S., et al., 2017, MNRAS, 470, 2617

Alam S., et al., 2021a, Phys. Rev. D, 103, 083533

Alam S., et al., 2021b, J. Cosmology Astropart. Phys., 2021, 050

Allys E., Peter P., Rodriguez Y., 2016, JCAP, 02, 004

Alom M. Z., et al., 2019, Electronics, 8

Alpher R. A., Bethe H., Gamow G., 1948, Physical Review, 73, 803

Amendola L., 2000, Phys. Rev. D, 62, 043511

Angulo R. E., Hahn O., 2022, Living Reviews in Computational Astrophysics, 8, 1

Armendariz-Picon C., Mukhanov V. F., Steinhardt P. J., 2000, Phys. Rev. Lett., 85, 4438

Armendariz-Picon C., Mukhanov V. F., Steinhardt P. J., 2001, Phys. Rev. D, 63, 103510

Arnold C., Leo M., Li B., 2019a, Nature Astronomy, 3, 945

Arnold C., Leo M., Li B., 2019b, Nature Astron., 3, 945

Arnold C., Fosalba P., Springel V., Puchwein E., Blot L., 2019d, MNRAS, 483, 790

Arnold C., Fosalba P., Springel V., Puchwein E., Blot L., 2019c, Mon. Not. Roy. Astron. Soc., 483, 790

Arnold C., Li B., Giblin B., Harnois-Déraps J., Cai Y.-C., 2021, arXiv e-prints, p. arXiv:2109.04984

Azzalini A., Capitanio A., 2009, arXiv e-prints, p. arXiv:0911.2342

Ba S., Myers W. R., Brenneman W. A., 2015, Technometrics, 57, 479

Babichev E., Deffayet C., Ziour R., 2009, Int. J. Mod. Phys. D, 18, 2147

Bardeen J. M., 1980, Phys. Rev. D, 22, 1882

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Barreira A., Li B., Baugh C. M., Pascoli S., 2012, Phys. Rev. D, 86, 124016

Barreira A., Brax P., Clesse S., Li B., Valageas P., 2015, Phys. Rev. D, 91, 063528

Barrera-Hinojosa C., Li B., 2020a, J. Cosmology Astropart. Phys., 2020, 007

Barrera-Hinojosa C., Li B., 2020b, J. Cosmology Astropart. Phys., 2020, 056

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Baugh C. M., Gaztanaga E., Efstathiou G., 1995, MNRAS, 274, 1049

Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999, MNRAS, 305, L21

Baumann D., Nicolis A., Senatore L., Zaldarriaga M., 2012, J. Cosmology Astropart. Phys., 2012, 051

Bellini E., et al., 2018, Phys. Rev. D, 97, 023520

Beltran Jimenez J., Heisenberg L., 2016, Phys. Lett. B, 757, 405

Benson A. J., 2010, Phys. Rep., 495, 33

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Berlind A. A., et al., 2003, ApJ, 593, 1

Bertone G., Hooper D., 2018, Reviews of Modern Physics, 90, 045002

Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, ApJ, 732, 122

Bianchi D., Chiesa M., Guzzo L., 2015, MNRAS, 446, 75

Bianchi D., Percival W. J., Bel J., 2016, MNRAS, 463, 3783

Blake C., Collister A., Lahav O., 2008, MNRAS, 385, 1257

Blake C., et al., 2011, MNRAS, 415, 2876

Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, ApJ, 901, 5

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bose B., Koyama K., 2017, J. Cosmology Astropart. Phys., 2017, 029

Bose S., Hellwing W. A., Li B., 2015, J. Cosmology Astropart. Phys., 2015, 034

Bose S., Li B., Barreira A., He J.-h., Hellwing W. A., Koyama K., Llinares C., Zhao G.-B., 2017, J. Cosmology Astropart. Phys., 2017, 050

Bose B., Winther H. A., Pourtsidou A., Casas S., Lombriser L., Xia Q., Cataneo M., 2020, J. Cosmology Astropart. Phys., 2020, 001

Bouhlel M. A., Hwang J. T., Bartoli N., Lafage R., Morlier J., Martins J. R. R. A., 2019, Advances in Engineering Software, p. 102662

Brax P., van de Bruck C., Davis A.-C., Shaw D. J., 2008, Phys. Rev. D, 78, 104021

Brax P., van de Bruck C., Davis A.-C., Shaw D., 2010, Phys. Rev. D, 82, 063519

Brax P., van de Bruck C., Davis A.-C., Li B., Shaw D. J., 2011a, Phys. Rev. D, 83, 104026

Brax P., van de Bruck C., Davis A.-C., Li B., Schmauch B., Shaw D. J., 2011b, Phys. Rev. D, 84, 123524

Brax P., Davis A.-C., Li B., Winther H. A., Zhao G.-B., 2012a, J. Cosmo. Astropart. Phys., 10, 02

Brax P., Davis A.-C., Li B., Winther H. A., 2012b, Phys. Rev. D, 86, 044015

Brax P., Davis A.-C., Li B., Winther H. A., Zhao G.-B., 2012c, J. Cosmology Astropart. Phys., 2012, 002

Bryan G. L., Norman M. L., 1998a, Astrophys. J., 495, 80

Bryan G. L., Norman M. L., 1998b, ApJ, 495, 80

Carlson J., Reid B., White M., 2013, MNRAS, 429, 1674

Carroll S. M., 2001, Living Reviews in Relativity, 4, 1

Cataneo M., Rapetti D., Lombriser L., Li B., 2016, J. Cosmology Astropart. Phys., 2016, 024

Chan K. C., Scoccimarro R., 2009, Phys. Rev. D, 80, 104005

Chapman M. J., et al., 2021, arXiv e-prints, p. arXiv:2106.14961

Chen S.-F., Vlah Z., Castorina E., White M., 2021, J. Cosmology Astropart. Phys., 2021, 100

Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Phys. Rept., 513, 1

Cohen D., Hairer E., Lubich C., 2003, Foundations of Computational Mathematics, 3, 327

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, Astrophys. J., 647, 201

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Copeland E. J., Sami M., Tsujikawa S., 2006, Int. J. Mod. Phys. D, 15, 1753

Courtin J., Rasera Y., Alimi J. M., Corasaniti P. S., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911

Crocce M., Pueblas S., Scoccimarro R., 2006a, Mon. Not. Roy. Astron. Soc., 373, 369

Crocce M., Pueblas S., Scoccimarro R., 2006b, MNRAS, 373, 369

Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303

Cuesta-Lazaro C., Li B., Eggemeier A., Zarrouk P., Baugh C. M., Nishimichi T., Takada M., 2020, MNRAS, 498, 1175

Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, MNRAS, 423, 2279

Cybenko G., 1989, Mathematics of control, signals and systems, 2, 303

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

Davis A.-C., Li B., Mota D. F., Winther H. A., 2012, Astrophys. J., 748, 61

Dawson K. S., et al., 2013, AJ, 145, 10

De Felice A., Tsujikawa S., 2010, Living Reviews in Relativity, 13, 3

DeRose J., et al., 2019, ApJ, 875, 69

Deffayet C., Esposito-Farese G., Vikman A., 2009, Phys. Rev. D, 79, 084003

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

Diemer B., 2018, ApJS, 239, 35

Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108

Dodelson S., Schmidt F., 2020, Modern cosmology. Academic Press

Dooley G. A., Griffen B. F., Zukin P., Ji A. P., Vogelsberger M., Hernquist L. E., Frebel A., 2014, ApJ, 786, 50

Dossett J., Hu B., Parkinson D., 2014, J. Cosmology Astropart. Phys., 2014, 046

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241

Elliott E. J., Baugh C. M., Lacey C. G., 2021, MNRAS, 506, 4011

Fisher K. B., 1995, ApJ, 448, 494

Gao L., White S. D. M., 2007, MNRAS, 377, L5

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

García R., Rozo E., Becker M. R., More S., 2021, MNRAS, 505, 1195

Goodfellow I., Bengio Y., Courville A., 2016, Deep learning. MIT press

Gupta S., Hellwing W. A., Bilicki M., García-Farieta J. E., 2022, Phys. Rev. D, 105, 043538

Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, Phys. Rev. D, 76, 083503

Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426

Hassani F., Lombriser L., 2020, Mon. Not. Roy. Astron. Soc., 497, 1885

Hassani F., Adamek J., Kunz M., Vernizzi F., 2019, J. Cosmology Astropart. Phys., 2019, 011

Hassani F., L'Huillier B., Shafieloo A., Kunz M., Adamek J., 2020, JCAP, 04, 039

Hearin A. P., et al., 2017, AJ, 154, 190

Heisenberg L., 2014, JCAP, 05, 015

Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, ApJ, 646, L1

Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104

Hendrycks D., Gimpel K., 2016, arXiv e-prints, p. arXiv:1606.08415

Hernández-Aguayo C., Arnold C., Li B., Baugh C. M., 2021, Mon. Not. Roy. Astron. Soc., 503, 3867

Hernández-Aguayo C., Ruan C.-Z., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, J. Cosmology Astropart. Phys., 2022, 048

Hinshaw G., et al., 2013, ApJS, 208, 19

Hinterbichler K., 2012, Rev. Mod. Phys., 84, 671

Hinterbichler K., Khoury J., 2010, Phys. Rev. Lett., 104, 231301

Hinterbichler K., Khoury J., Levy A., Matas A., 2011, Phys. Rev. D, 84, 103521

Hojjati A., Pogosian L., Zhao G.-B., 2011, J. Cosmology Astropart. Phys., 2011, 005

Horndeski G. W., 1974, International Journal of Theoretical Physics, 10, 363

Hornik K., Stinchcombe M., White H., 1989, Neural networks, 2, 359

Hu W., Sawicki I., 2007, Phys. Rev. D, 76, 064004

Hu B., Liguori M., Bartolo N., Matarrese S., 2013, Phys. Rev. D, 88, 123514

Hu B., Raveri M., Frusciante N., Silvestri A., 2014, Phys. Rev. D, 89, 103530

Hubble E., Humason M. L., 1931, ApJ, 74, 43

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jennings W. D., Watkinson C. A., Abdalla F. B., McEwen J. D., 2019, MNRAS, 483, 2907

Joyce A., Jain B., Khoury J., Trodden M., 2015, Phys. Rep., 568, 1

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Khoury J., Weltman A., 2004a, Phys. Rev. D, 69, 044026

Khoury J., Weltman A., 2004b, Phys. Rev. Lett., 93, 171104

Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901

Kingma D. P., Ba J., 2014, arXiv e-prints, p. arXiv:1412.6980

Kitaura F.-S., et al., 2016, MNRAS, 456, 4156

Klypin A., Holtzman J., 1997, arXiv e-prints, pp astro–ph/9712217

Klypin A., Prada F., 2018, MNRAS, 478, 4602

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Klypin A., et al., 2021, Mon. Not. Roy. Astron. Soc., 504, 769

Knebe A., et al., 2011, MNRAS, 415, 2293

Kobayashi Y., Nishimichi T., Takada M., Takahashi R., Osato K., 2020, Phys. Rev. D, 102, 063504

Kodama H., Sasaki M., 1984, Progress of Theoretical Physics Supplement, 78, 1

Koyama K., 2018, International Journal of Modern Physics D, 27, 1848001

Kuruvilla J., Porciani C., 2018, MNRAS, 479, 2256

Lacey C. G., et al., 2016, MNRAS, 462, 3854

Lam T. Y., Li B., 2012, MNRAS, 426, 3260

Li B., Barrow J. D., 2011a, Phys. Rev. D, 83, 024007

Li B., Barrow J. D., 2011b, Phys. Rev. D, 83, 024007

Li Y., Hu W., 2011, Phys. Rev. D, 84, 084033

Li Y., Mo H. J., Gao L., 2008, MNRAS, 389, 1419

Li B., Zhao G.-B., Teyssier R., Koyama K., 2012, J. Cosmology Astropart. Phys., 2012, 051

Li B., Hellwing W. A., Koyama K., Zhao G.-B., Jennings E., Baugh C. M., 2013a, MNRAS, 428, 743

Li B., Zhao G.-B., Koyama K., 2013b, J. Cosmology Astropart. Phys., 2013, 023

Libeskind N. I., et al., 2018, MNRAS, 473, 1195

Linder E. V., 2003, Phys. Rev. Lett., 90, 091301

Liu J., Madhavacheril M. S., 2019, Phys. Rev. D, 99, 083508

Llinares C., Mota D. F., 2014, Phys. Rev. D, 89, 084023

Lombriser L., Li B., Koyama K., Zhao G.-B., 2013, Phys. Rev. D, 87, 123511

Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, MNRAS, 508, 4017

Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, MNRAS, 474, 5143

McClintock T., et al., 2019, ApJ, 872, 53

Mead A. J., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401

Meneghetti M., Rasia E., 2013, arXiv e-prints, p. arXiv:1303.6158

Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, MNRAS, 500, 663

Mitchell M. A., Arnold C., He J.-h., Li B., 2019, Mon. Not. Roy. Astron. Soc., 487, 1410

Mitchell M. A., Hernández-Aguayo C., Arnold C., Li B., 2021, MNRAS, 508, 4140

Miyatake H., et al., 2020, arXiv e-prints, p. arXiv:2101.00113

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, The Astrophys. J., 710, 903

Mota D. F., Shaw D. J., 2007, Phys. Rev. D, 75, 063501

Murray S. G., Power C., Robotham A. S. G., 2013, Astronomy and Computing, 3, 23

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Neto A. F., et al., 2007, MNRAS, 381, 1450

Neyman J., Scott E. L., 1952, ApJ, 116, 144

Nicolis A., Rattazzi R., Trincherini E., 2009, Phys. Rev. D, 79, 064036

Nikoloudakis N., Shanks T., Sawangwit U., 2013, MNRAS, 429, 2032

Nishimichi T., et al., 2019a, ApJ, 884, 29

Nishimichi T., et al., 2019b, ApJ, 884, 29

Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

Oyaizu H., 2007, Phys. Rev. D, 78, 123523

Oyaizu H., 2008, Phys. Rev. D, 78, 123523

Paszke A., et al., 2019, in , Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp 8024–8035

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peacock J. A., et al., 2001, Nature, 410, 169

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., 2017, Nature Astronomy, 1, 0057

Perlmutter S., et al., 1999, Astrophys. J., 517, 565

Petri A., Liu J., Haiman Z., May M., Hui L., Kratochvil J. M., 2015, Phys. Rev. D, 91, 103511

Pezzotta A., et al., 2017, A&A, 604, A33

Planck Collaboration XIII 2016, Astron. Astrophys., 594, A13

Planck Collaboration et al., 2016a, A&A, 594, A1

Planck Collaboration et al., 2016b, A&A, 594, A13

Planck Collaboration et al., 2020a, A&A, 641, A6

Planck Collaboration et al., 2020b, A&A, 641, A9

Porredon A., et al., 2021, arXiv e-prints, p. arXiv:2105.13546

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Press W. H., Schechter P., 1974, ApJ, 187, 425

Puchwein E., Baldi M., Springel V., 2013, MNRAS, 436, 348

Ratra B., Peebles P. J. E., 1988, Phys. Rev. D, 37, 3406

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, The Astrophys. J., 771, 30

Reid B. A., White M., 2011, MNRAS, 417, 1913

Reid B. A., et al., 2012, MNRAS, 426, 2719

Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, MNRAS, 444, 476

Riebe K., et al., 2011, arXiv e-prints, p. arXiv:1109.0003

Riess A. G., et al., 1998, Astron. J., 116, 1009

Rodríguez-Torres S. A., et al., 2016, MNRAS, 460, 1173

Ruan C.-Z., Hernández-Aguayo C., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, J. Cosmology Astropart. Phys., 2022, 018

Rubin V. C., Ford W. Kent J., 1970, ApJ, 159, 379

Saha P., Tremaine S., 1992, AJ, 104, 1633

Samushia L., Percival W. J., Raccanelli A., 2012, MNRAS, 420, 2102

Samushia L., et al., 2014, MNRAS, 439, 3504

Satpathy S., et al., 2017, MNRAS, 469, 1369

Scoccimarro R., 2004, Phys. Rev. D, 70, 083007

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20

Seljak U., 2000, MNRAS, 318, 203

Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720

Sheth R. K., Mo H. J., Tormen G., 2001a, MNRAS, 323, 1

Sheth R. K., Hui L., Diaferio A., Scoccimarro R., 2001b, MNRAS, 325, 1288

Smith S., 1936, ApJ, 83, 23

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Sotiriou T. P., Faraoni V., 2010, Rev. Mod. Phys., 82, 451

Springel V., 2005, Mon. Not. Roy. Astron. Soc., 364, 1105

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Srinivasan S., Thomas D. B., Pace F., Battye R., 2021, JCAP, 06, 016

Steinhardt P. J., Wang L.-M., Zlatev I., 1999, Phys. Rev. D, 59, 123504

Suyu S. H., et al., 2017, MNRAS, 468, 2590

Tamone A., et al., 2020, MNRAS, 499, 5527

Tegmark M., 1997, Phys. Rev. Lett., 79, 3806

Tegmark M., et al., 2004, ApJ, 606, 702

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

Turner M. S., KICP/UChicago Kavli Foundation T., 2021, arXiv e-prints, p. arXiv:2111.14254

Uhlemann C., Kopp M., Haugg T., 2015, Phys. Rev. D, 92, 063004

Valogiannis G., Bean R., Aviles A., 2020, J. Cosmology Astropart. Phys., 2020, 055

Villaescusa-Navarro F., et al., 2020, ApJS, 250, 2

Virtanen P., et al., 2020, Nature Methods, 17, 261

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Wang J., Bose S., Frenk C. S., Gao L., Jenkins A., Springel V., White S. D. M., 2020, Nature, 585, 39

Watson D. F., Berlind A. A., McBride C. K., Masjedi M., 2010, ApJ, 709, 115

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

Wetterich C., 1988, Nuclear Physics B, 302, 668

White S. D. M., 1996, in Schaeffer R., Silk J., Spiro M., Zinn-Justin J., eds, Cosmology and Large Scale Structure. p. 349

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

White S. D. M., Frenk C. S., Davis M., 1983, ApJ, 274, L1

White M., et al., 2011, ApJ, 728, 126

Williams C. K., Rasmussen C. E., 2006, Gaussian processes for machine learning. MIT press Cambridge, MA

Winther H. A., Mota D. F., Li B., 2012, ApJ, 756, 166

Yoo J., Seljak U., 2015, MNRAS, 447, 1789

Zel'Dovich Y. B., 1970, A&A, 500, 13

Zhai Z., et al., 2019, ApJ, 874, 95

Zhao G.-B., Li B., Koyama K., 2011, Phys. Rev. D, 83, 044007

Zheng Z., et al., 2005a, Astrophys. J., 633, 791

Zheng Z., et al., 2005b, ApJ, 633, 791

Zlatev I., Wang L.-M., Steinhardt P. J., 1999, Phys. Rev. Lett., 82, 896

Zu Y., Weinberg D. H., 2013, MNRAS, 431, 3319

Zwicky F., 1933, Helvetica Physica Acta, 6, 110