Chapter 2

Numerical diagnostics of Lagrangian perturbation theory

Contents

"**Hector Barbossa**: The world used to be a bigger place.
Jack Sparrow: The world's still the same. There's just... less in it."
— Pirates of the Caribbean: At World's End (2007)

Abstract

This chapter is intended as a guide on the approximation error in using Lagrangian perturbation theory instead of fully non-linear gravity in large-scale structure analysis. We compare various properties of particle realizations produced by LPT and by N-body simulations. In doing so, we characterize differences and similarities, as a function of scale, resolution and redshift.

The goal of this chapter is to characterize the accuracy of Lagrangian perturbation theory in terms of a set of numerical diagnostics. It is organized as follows. In section 2.1, we look at the correlations functions of the density field, which usually are the final observable in cosmological surveys. As the displacement field plays a central role in LPT, we study its statistics in section 2.2. In particular, we illustrate that in some regimes, when the perturbative parameter is large, 2LPT performs worse than the ZA. We examine the decomposition of the displacement field in a scalar and rotational part and review various approximations based on its divergence. Finally, in section 2.3, we compare cosmic web elements (voids, sheets, filaments, and clusters) as predicted by LPT and by non-linear simulations of the LSS.[1]

Corresponding LPT and N-body simulations used in this chapter have been run from the same initial conditions, generated at redshift $z = 63$ using second-order Lagrangian perturbation theory. The N-body simulations have been run with the Gadget-2 cosmological code (Springel, Yoshida & White, 2001; Springel, 2005). Evolutions of the Zel'dovich approximation were performed with N-GenIC (Springel, 2005), and of second-order Lagrangian perturbation theory with 2LPTic (Crocce, Pueblas & Scoccimarro, 2006a). To ensure sufficient statistical significance, we used eight realizations of the same cosmology, changing the seed used to generate respective initial conditions. All computations are done after binning the dark matter particles with a Cloud-in-Cell (CiC) method (see section B.3). The simulations contain 512^3 particles in a 1024 Mpc/h cubic box with periodic boundary conditions. We checked that with this setup, the power spectrum agrees with

[1] In the following, we will often write "full gravity", even if, strictly speaking, N-body simulations also involve some degree of approximation.

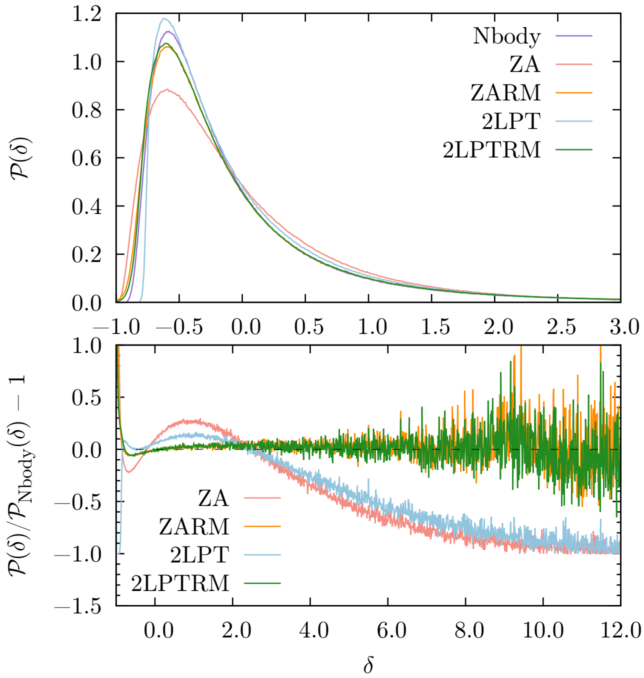

Figure 2.1: *Upper panel.* Redshift-zero probability distribution function for the density contrast δ, computed from eight 1024 Mpc/h-box simulations of 512^3 particles. The particle distribution is determined using: a full N-body simulation (purple curve), the Zel'dovich approximation, alone (ZA, light red curve) and after remapping (ZARM, orange curve), second-order Lagrangian perturbation theory, alone (2LPT, light blue curve) and after remapping (2LPTRM, green curve). *Lower panel.* Relative deviations of the same pdfs with reference to N-body simulation results. Note that, contrary to standard LPT approaches, remapped fields follow the one-point distribution of full N-body dynamics in an unbiased way, especially in the high density regime.

the non-linear power spectrum of simulations run with higher mass resolution, provided by Cosmic Emulator tools (Heitmann *et al.*, 2009, 2010; Lawrence *et al.*, 2010) (deviations are at most sub-percent level for $k \lesssim 1$ (Mpc/h)$^{-1}$). Therefore, at the scales of interest of this work, $k \leq 0.4$ (Mpc/h)$^{-1}$ (corresponding to the linear and mildly non-linear regime at redshift zero), the clustering of dark matter is correctly reproduced by our set of simulations.

The cosmological parameters used are WMAP-7 fiducial values (Komatsu *et al.*, 2011),

$$\Omega_\Lambda = 0.728, \Omega_\mathrm{m} = 0.2715, \Omega_\mathrm{b} = 0.0455, \sigma_8 = 0.810, h = 0.704, n_\mathrm{s} = 0.967. \tag{2.1}$$

Thus, each particle carries a mass of 6.03×10^{11} $\mathrm{M}_\odot/h$.

2.1 Correlation functions of the density field

This section draws from section III in Leclercq *et al.* (2013).

In this section, we analyze the correlation functions of the density contrast field, δ, in LPT and N-body fields.

Note. All plots presented in this section contain lines labeled as "ZARM" and "2LPTRM" which correspond to remapped fields based on the ZA and on 2LPT, respectively. They are ignored in this chapter, which focuses on diagnostics of LPT. For a description of the remapping procedure and for comments on these approximations in comparison to the ZA, 2LPT and N-body dynamics, the reader is referred to chapter 6.

2.1.1 One-point statistics

Figure 2.1 shows the pdf for the density contrast, $\mathcal{P}_\delta$, at redshift zero, for N-body simulations, and for ZA and 2LPT density fields. All pdfs are non-Gaussian with a substantial skewness, are tied down to 0 at $\delta = -1$

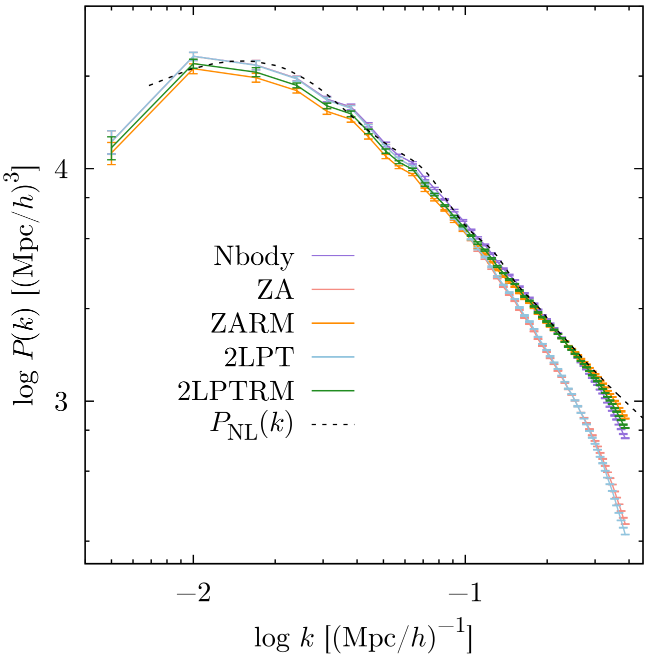

Figure 2.2: Redshift-zero dark matter power spectra in a 1024 Mpc/h simulation, with density fields computed with a mesh size of 8 Mpc/h. The particle distribution is determined using: a full N-body simulation (purple curve), the Zel'dovich approximation, alone (ZA, light red curve) and after remapping (ZARM, orange curve), second-order Lagrangian perturbation theory, alone (2LPT, light blue curve) and after remapping (2LPTRM, green curve). The dashed black curve represents $P_{\mathrm{NL}}(k)$, the theoretical power spectrum expected at $z = 0$. Both ZARM and 2LPTRM show increased power in the mildly non-linear regime compared to ZA and 2LPT (at scales corresponding to $k \gtrsim 0.1$ (Mpc/h)$^{-1}$ for this redshift), indicating an improvement of two-point statistics with the remapping procedure.

with a large tail in the high-density values. As discussed in section 1.2.3.4, the late-time pdf for density fields is approximately log-normal. However, already at the level of one-point statistics, the detailed behaviors of LPT and N-body simulations disagree: the peak of the pdf is shifted and the tails differ. In particular, LPT largely underpredicts the number of voxels in the high-density regime. This effect is more severe for the ZA than for 2LPT. This comes from the fact that 2LPT captures some of non-local effects involved in the formation of the densest halos.

The one-point pdf of the density is further analyzed in section 2.2.1, in comparison to the one-point pdf of the Lagrangian displacement field.

2.1.2 Two-point statistics

2.1.2.1 Power spectrum

We measured the power spectrum of dark matter density fields, as defined by equation (1.41). Dark matter particles have been displaced according to each prescription and assigned to cells with a CiC scheme, for different mesh sizes. Power spectra were measured from theses meshes, with a correction for aliasing effects (Jing, 2005). Redshift-zero results computed on a 8 Mpc/h mesh are presented in figure 2.2. There, the dashed line corresponds to the theoretical, non-linear power spectrum expected, computed with Cosmic Emulator tools (Heitmann *et al.*, 2009, 2010; Lawrence *et al.*, 2010). A deviation of full N-body simulations from this theoretical prediction can be observed at small scales. This discrepancy is a gridding artifact, completely due to the finite mesh size used for the analysis. As a rule of thumb, a maximum threshold in k for trusting the simulation data is set by a quarter of the Nyquist wavenumber, defined by $k_{\mathrm{N}} \equiv 2\pi/L \times N_{\mathrm{p}}^{1/3}/2$, where L is the size of the box and N_{p} is the number of cells in the Lagrangian grid on which particles are placed in the initial conditions; which makes for our analysis ($L = 1024$ Mpc/h, $N_{\mathrm{p}} = 512^3$), $k_{\mathrm{N}}/4 \approx 0.39$ (Mpc/h)$^{-1}$. At this scale, it has been observed that the power spectrum starts to deviate at the percent-level with respect to higher resolution simulations (Heitmann *et al.*, 2010). The relative deviations of various power spectra with reference

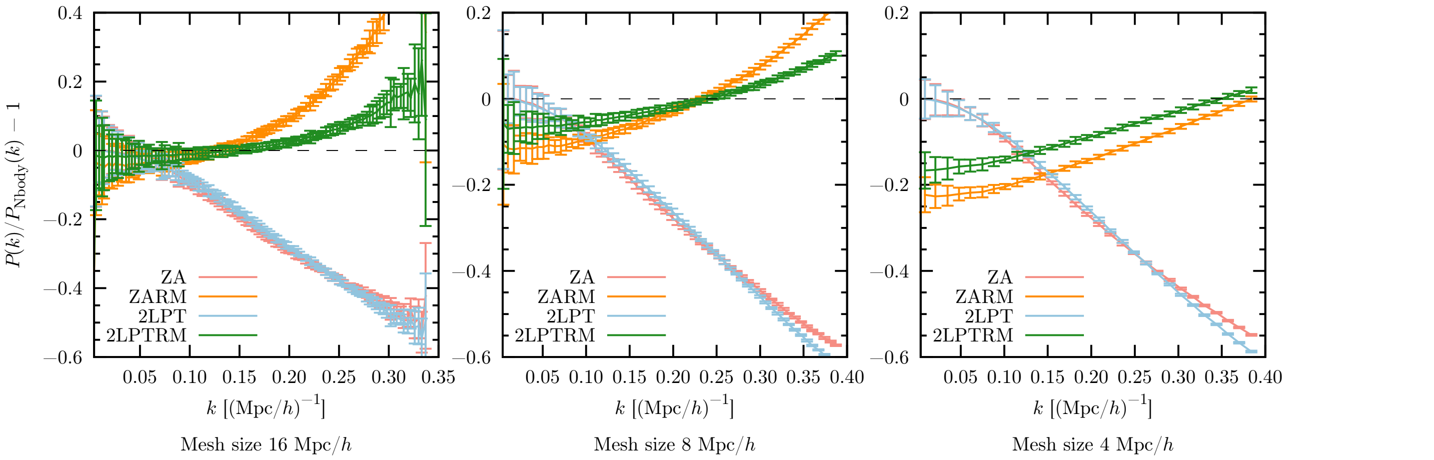

Figure 2.3: *Power spectrum: mesh size-dependence.* Relative deviations for the power spectra of various particle distributions, with reference to the density field computed with a full N-body simulation. The particle distribution is determined using: the Zel'dovich approximation, alone (ZA, light red curve) and after remapping (ZARM, orange curve), second-order Lagrangian perturbation theory, alone (2LPT, light blue curve) and after remapping (2LPTRM, green curve). The computation is done on different meshes: 16 Mpc/h (64^3-voxel grid, left panel), 8 Mpc/h (128^3-voxel grid, central panel) and 4 Mpc/h (256^3-voxel grid, right panel). All results are shown at redshift $z = 0$. LPT fields exhibit more small-scale correlations after remapping and their power spectra get closer to the shape of the full non-linear power spectrum.

to full gravity are presented in figures 2.3 and 2.4. In all the plots, the error bars represent the dispersion of the mean among eight independent realizations.

Generally, LPT correctly predicts the largest scales, when $k \to 0$ (the smallest wavelength mode accessible here is set by the box size: $k_{\mathrm{min}} = 2\pi/L$ with $L = 1024$ Mpc/h, giving $k_{\mathrm{min}} \approx 0.006$ (Mpc/h)$^{-1}$), as these are in the linear regime. These are affected by cosmic variance, but the effect is not visible in our plots, as corresponding LPT and N-body fields start from the same initial conditions. Differences arise in the mildly non-linear and non-linear regime, where LPT predicts too little power. Indeed, as LPT only captures part of the non-linearity of the Vlasov-Poisson system, presented in section 1.3.1, the clustering of dark matter particles is underestimated.

The discrepancy between LPT and N-body power spectra depends both on the target resolution (see figure 2.3) and on the desired redshift (see figure 2.4). For example, at a resolution of 8 Mpc/h and at a comoving wavelength of $k = 0.40$ (Mpc/h)$^{-1}$, 2LPT only lacks 5% power at $z = 3$ but more than 50% at $z = 0$. At fixed redshift, the lack of small scale power in LPT weakly depends on the mesh size.

2.1.2.2 Fourier-space cross-correlation coefficient

The Fourier space cross-correlation coefficient between two density fields δ and δ' is defined as the cross-power spectrum of δ and δ', normalized by the auto-power spectra of the same fields:

$$R(k) \equiv \frac{P_{\delta\times\delta'}(k)}{\sqrt{P_\delta(k)P_{\delta'}(k)}} \equiv \frac{\langle \delta^*(\mathbf{k})\delta'(\mathbf{k})\rangle}{\sqrt{\langle \delta^*(\mathbf{k})\delta(\mathbf{k})\rangle \, \langle \delta'^*(\mathbf{k})\delta'(\mathbf{k})\rangle}}. \tag{2.2}$$

It is a dimensionless coefficient, in modulus between 0 and 1, representing the agreement, at the level of two-point statistics, between the *phases* of δ and δ' (as the overall power has been divided out). Here we choose as a reference the density field predicted by N-body simulations, $\delta' = \delta_{\mathrm{Nbody}}$, and compare with approximate density fields generated from the same initial conditions with LPT. In this fashion, we characterize the phase accuracy of the ZA and 2LPT.

In figure 2.5 we present the Fourier-space cross-correlation coefficient between the redshift-zero density field in the N-body simulation and other density fields. At this point, it is useful to recall that an approximation well-correlated with the non-linear density field can be used in a variety of cosmological applications, such as the reconstruction of the non-linear power spectrum (Tassev & Zaldarriaga, 2012c). As pointed out by Neyrinck (2013), the cross-correlation between 2LPT and full gravitational dynamics is higher at small k than the cross-

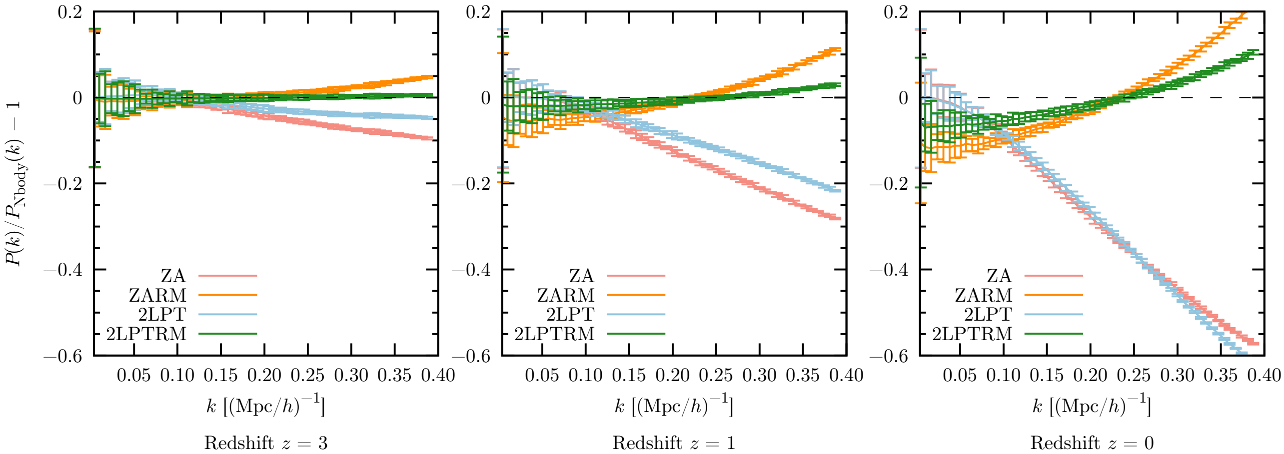

Figure 2.4: *Power spectrum: redshift-dependence.* Relative deviations for the power spectra of various particle distributions (see the caption of figure 2.3), with reference to the density field computed with a full N-body simulation. The computation is done on a 8 Mpc/h mesh (128^3-voxel grid). Results at different redshifts are shown: $z = 3$ (right panel), $z = 1$ (central panel) and $z = 0$ (left panel). The remapping procedure is increasingly successful with increasing redshift.

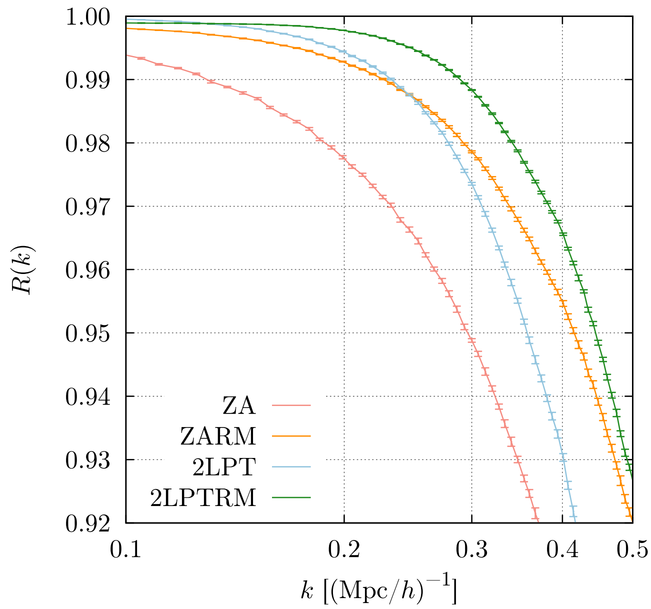

Figure 2.5: Fourier-space cross-correlation coefficient between various approximately-evolved density fields and the particle distribution as evolved with full N-body dynamics, all at redshift zero. The binning of density fields is done on a 8 Mpc/h mesh (128^3-voxel grid). At small scales, $k \geq 0.2$ (Mpc/h)$^{-1}$, the cross-correlations with respect to the N-body-evolved field are notably better after remapping than with LPT alone.

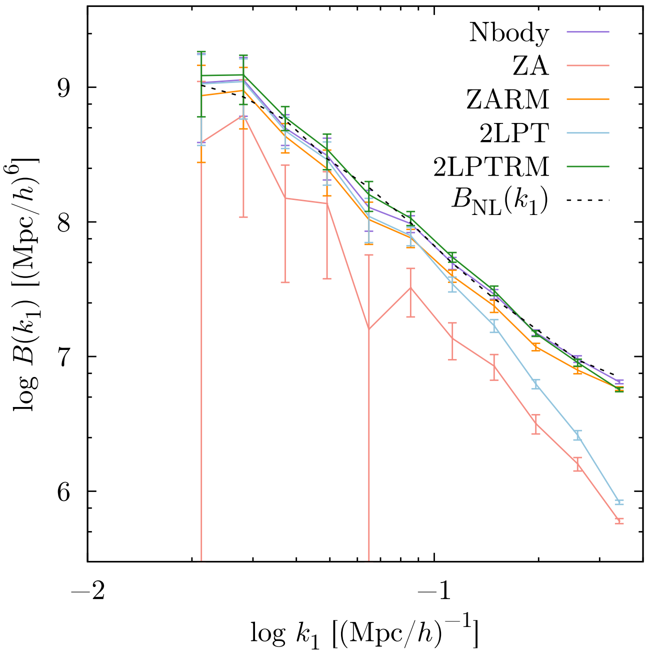

Figure 2.6: Redshift-zero dark matter bispectra for equilateral triangle shape, in 1024 Mpc/h simulations, with density fields computed on mesh of 8 Mpc/h size. The particle distribution is determined using: a full N-body simulation (purple curve), the Zel'dovich approximation, alone (ZA, light red curve) and after remapping (ZARM, orange curve), second-order Lagrangian perturbation theory, alone (2LPT, light blue curve) and after remapping (2LPTRM, green curve). The dashed line, $B_{\mathrm{NL}}(k)$, corresponds to theoretical predictions for the bispectrum, found using the fitting formula of (Gil-Marín *et al.*, 2012). Note that both ZARM and 2LPTRM show increased bispectrum in the mildly non-linear regime compared to ZA and 2LPT, indicating an improvement of three-point statistics with the remapping procedure.

correlation between the ZA and the full dynamics, meaning that the position of structures is more correct when additional physics (non-local tidal effects) is taken into account.

2.1.3 Three-point statistics

In this section, we analyze the accuracy of LPT beyond second-order statistics, by studying the three-point correlation function of the density field in Fourier space, i.e. the bispectrum, defined by equation (1.56). The importance of three-point statistics relies in their ability to test the shape of structures. Some of the natural applications are to test gravity (Shirata *et al.*, 2007; Gil-Marín *et al.*, 2011), to break degeneracies due to the galaxy bias (Matarrese, Verde & Heavens, 1997; Verde *et al.*, 1998; Scoccimarro *et al.*, 2001; Verde *et al.*, 2002) or to test the existence of primordial non-Gaussianities in the initial matter density field (Sefusatti & Komatsu, 2007; Jeong & Komatsu, 2009).

As for the power spectrum, we construct the dark matter density contrast field, by assigning particles to the grid using a CiC scheme. We then deconvolve the CiC kernel to correct for corresponding smoothing effects. The algorithm used to compute the bispectrum $B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ from this $\delta(\mathbf{k})$ field consists of randomly drawing k-vectors from a specified bin, namely Δk, and randomly orientating the $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ triangle in space. We chose the number of random triangles to depend on the number of fundamental triangle per bin, that scales as $k_1 k_2 k_3 \Delta k^3$ (Scoccimarro, 1997), where Δk is the chosen k-binning: given k_i we allow triangles whose i-side lies between $k_i - \Delta k/2$ and $k_i + \Delta k/2$. In this paper we always set $\Delta k = k_{\mathrm{min}} = 2\pi/L$, where L is the size of the box. For the equilateral case, at scales of $k \approx 0.1$ (Mpc/h)$^{-1}$ we generate $\sim 1.7 \times 10^6$ random triangles. We have verified that increasing the number of triangles beyond this value does not have any effect on the measurement. The rule of thumb presented in section 2.1.2.1 for the smallest scale to trust applies for the bispectrum as well. Also, as a lower limit in k, we have observed that for scales larger than $\sim 3\,k_{\mathrm{min}}$, effects of cosmic variance start to be important and considerable deviations with respect to linear theory can be observed. For this reason, we limit the largest scale for our bispectrum analysis to $3\,k_{\mathrm{min}} \approx 1.8 \times 10^{-2}$ (Mpc/h)$^{-1}$.

Error bars in bispectrum plots represent the dispersion of the mean among eight independent realizations, all

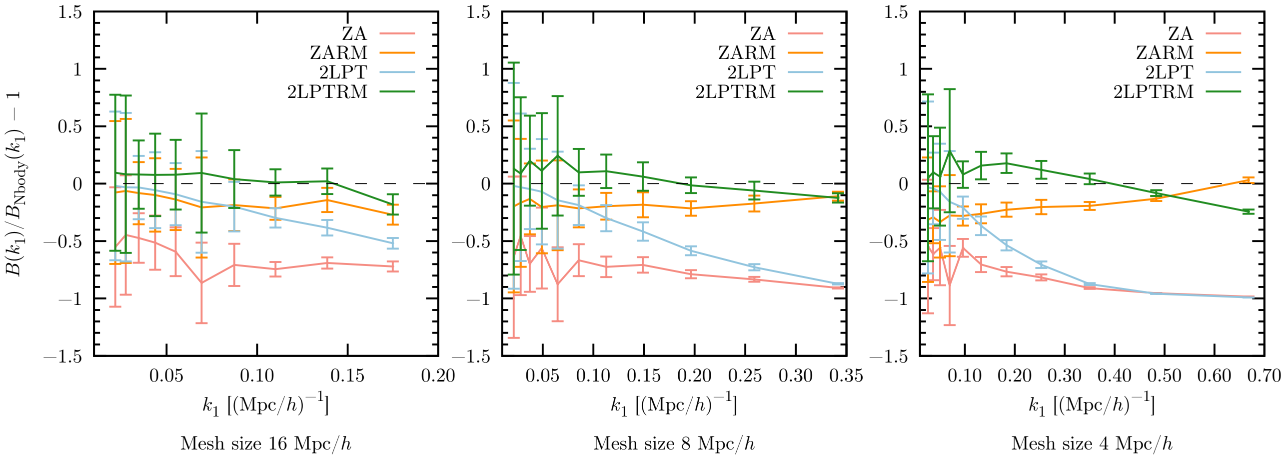

Figure 2.7: *Bispectrum: mesh size-dependence.* Relative deviations for the bispectra $B(k_1)$ of various particle distributions, with reference to the prediction from a full N-body simulation, $B_{\mathrm{Nbody}}(k_1)$. The particle distribution is determined using: the Zel'dovich approximation, alone (ZA, light red curve) and after remapping (ZARM, orange curve), second-order Lagrangian perturbation theory, alone (2LPT, light blue curve) and after remapping (2LPTRM, green curve). The computation of bispectra is done for equilateral triangles and on different meshes: 16 Mpc/h (64^3-voxel grid, left panel), 8 Mpc/h (128^3-voxel grid, central panel) and 4 Mpc/h (256^3-voxel grid, right panel). All results are shown at redshift $z = 0$. LPT fields exhibit more small-scale three-point correlations after remapping and their bispectra get closer to the shape of the full non-linear bispectrum.

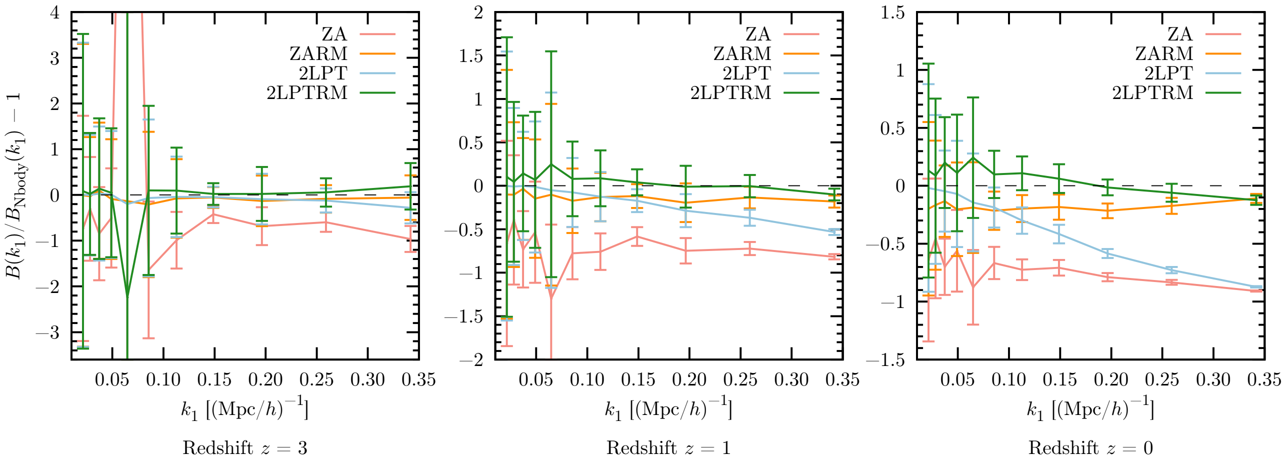

Figure 2.8: *Bispectrum: redshift-dependence.* Relative deviations for the bispectra $B(k_1)$ of various particle distributions (see the caption of figure 2.7), with reference to a full N-body simulation, $B_{\mathrm{Nbody}}(k_1)$. The computation of bispectra is done on a 8 Mpc/h mesh (128^3-voxel grid) and for equilateral triangles. Results at different redshifts are shown: $z = 3$ (right panel), $z = 1$ (central panel) and $z = 0$ (left panel).

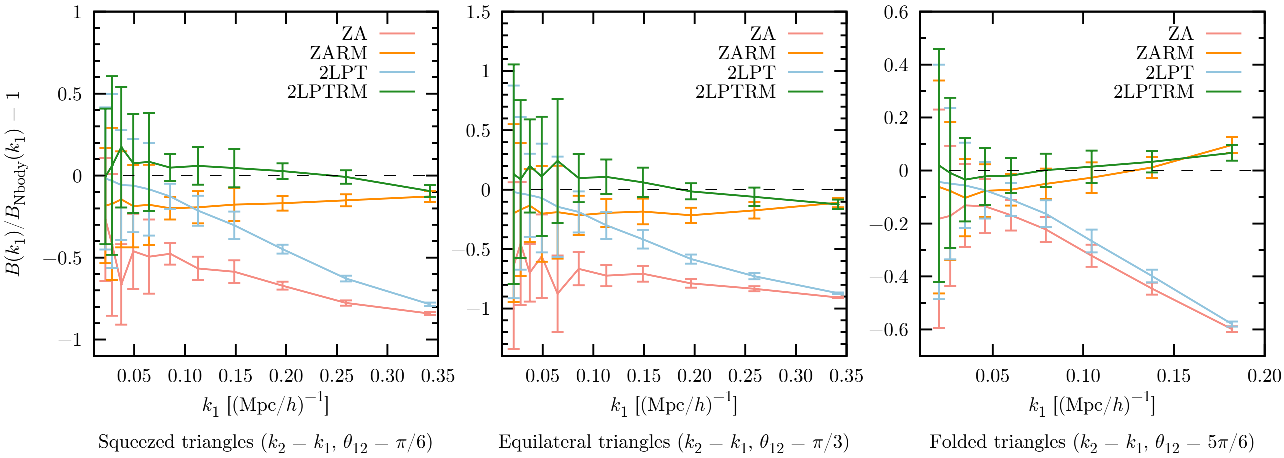

Figure 2.9: *Bispectrum: scale-dependence for different triangle shapes.* Relative deviations for the bispectra $B(k_1)$ of various particle distributions (see the caption of figure 2.7), with reference to a full N-body simulation, $B_{\mathrm{Nbody}}(k_1)$. The computation is done on a 8 Mpc/h mesh (128^3-voxel grid) and results are shown at redshift $z = 0$ for various triangle shapes as defined above.

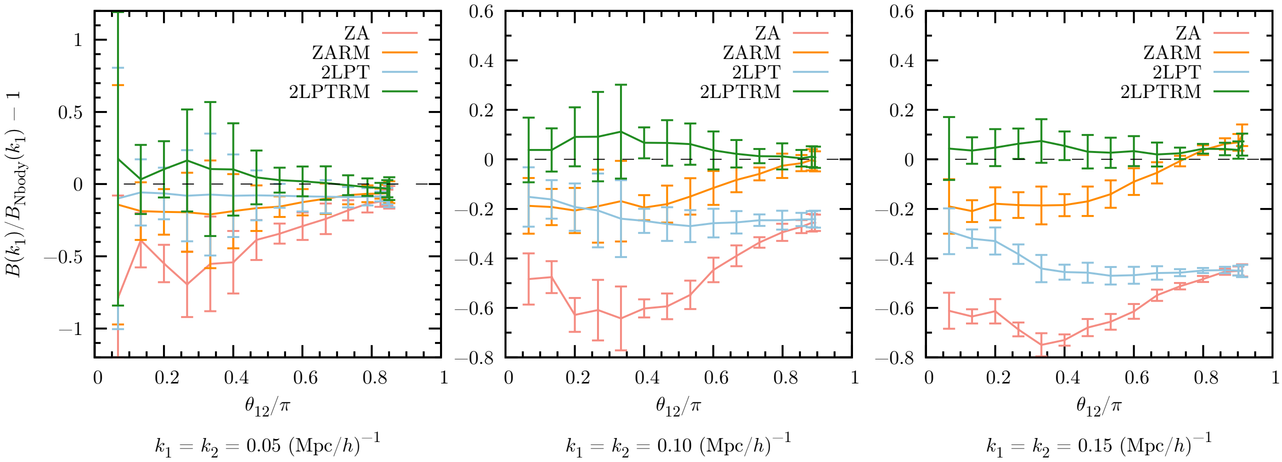

Figure 2.10: *Bispectrum: triangle shape-dependence.* Relative deviations for the bispectra $B(k_1)$ of various particle distributions (see the caption of figure 2.7), with reference to a full N-body simulation, $B_{\mathrm{Nbody}}(k_1)$. The computation is done on a 8 Mpc/h mesh (128^3-voxel grid) and results are shown at redshift $z = 0$. The dependence on the angle of the triangle $\theta_{12} = (\mathbf{k}_1, \mathbf{k}_2)$ is shown for different scales: $k_1 = k_2 = 0.05$ (Mpc/h)$^{-1}$ (corresponding to 125 Mpc/h), $k_1 = k_2 = 0.10$ (Mpc/h)$^{-1}$ (corresponding to 63 Mpc/h), $k_1 = k_2 = 0.15$ (Mpc/h)$^{-1}$ (corresponding to 42 Mpc/h).

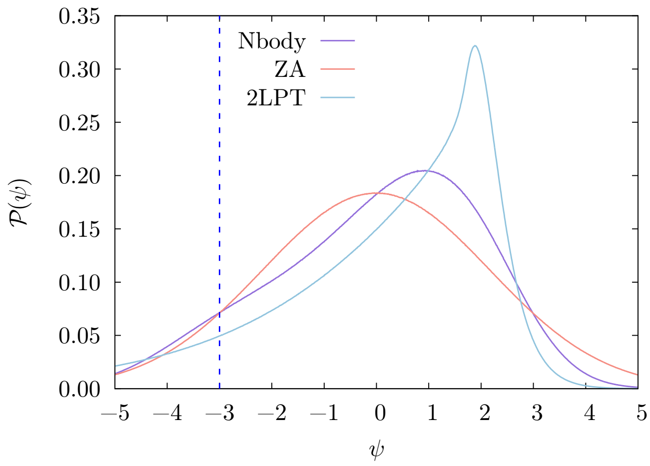

Figure 2.11: Redshift-zero probability distribution function for the divergence of the displacement field ψ, computed from eight 1024 Mpc/h-box simulations of 512^3 particles. A quantitative analysis of the deviation from Gaussianity of these pdfs is given in table 2.1. The particle distribution is determined using: a full N-body simulation (purple curve), the Zel'dovich approximation (ZA, light red curve) and second-order Lagrangian perturbation theory (2LPT, light blue curve). The vertical line at $\psi = -3$ represents the collapse barrier about which ψ values bob around after gravitational collapse. A bump at this value is visible with full gravity, but LPT is unable to reproduce this feature. This regime corresponds to virialized, overdense clusters.

of them with the same cosmological parameters. It has been tested (Gil-Marín *et al.*, 2012), that this estimator for the error is in good agreement with theoretical predictions based on the Gaussianity of initial conditions (Scoccimarro, 1998).

The subtracted shot noise is always assumed to be Poissonian:

$$B_{\mathrm{SN}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{1}{\bar{n}} \left[P(k_1) + P(k_2) + P(k_3)\right] + \frac{1}{\bar{n}^2}, \tag{2.3}$$

(see e.g. Peebles, 1980, and references therein), where $\bar{n}$ is the number density of particles in the box.

A triangle shape is defined by the relative length of vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ and the inner angle θ_{12}, in such a way that $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ and $\mathbf{k}_1 \cdot \mathbf{k}_2 = k_1 k_2 \cos(\pi - \theta_{12})$. In figure 2.6, we plot the redshift-zero bispectrum, computed on a 8 Mpc/h mesh, of the different density fields for equilateral triangles ($\theta_{12} = \pi/3$ and $k_2/k_1 = 1$). There, the dashed line corresponds to theoretical predictions for the non-linear bispectrum, found using the fitting formula of Gil-Marín *et al.* (2012). The relative deviations of various bispectra with reference to full N-body simulations are shown in figures 2.7, 2.8, 2.9 and 2.10.

The main result is that LPT predicts less three-point correlation than full gravity. This is true even at large scales for the ZA: as it is local, it generally fails to predict the shape of structures. 2LPT agrees with N-body simulations at large scales, with differences starting to appear only in the mildly non-linear regime, $k \gtrsim 0.1$ (Mpc/h)$^{-1}$ at $z = 0$.

2.2 Statistics of the Lagrangian displacement field

2.2.1 Lagrangian ψ versus Eulerian δ: one-point statistics

This section draws from Leclercq, Jasche & Wandelt (2015b), addendum to Leclercq *et al.* (2013).

As noted by previous authors (see in particular Neyrinck, 2013), in the Lagrangian representation of the LSS, it is natural to use the divergence of the displacement field ψ instead of the Eulerian density contrast δ. This section comments the one-point statistics of ψ in LPT and full gravity and comparatively analyzes key features of ψ and δ.

As seen in section 1.5, in the Lagrangian frame, the quantity of interest is not the position, but the displacement field $\mathbf{\Psi}(\mathbf{q})$ which maps the initial comoving particle position $\mathbf{q}$ to its final comoving Eulerian position $\mathbf{x}$

Model	$\mathcal{P}_\delta$	$\mathcal{P}_\psi$
	Skewness γ_1	
ZA	2.36 ± 0.01	-0.0067 ± 0.0001
2LPT	2.83 ± 0.01	-1.5750 ± 0.0002
N-body	5.14 ± 0.05	-0.4274 ± 0.0001
	Excess kurtosis γ_2	
ZA	9.95 ± 0.09	$-2.2154 \times 10^{-6} \pm 0.0003$
2LPT	13.91 ± 0.15	3.544 ± 0.0011
N-body	62.60 ± 2.75	-0.2778 ± 0.0004

Table 2.1: Non-Gaussianity parameters (the skewness γ_1 and the excess kurtosis γ_2) of the redshift-zero probability distribution functions $\mathcal{P}_\delta$ and $\mathcal{P}_\psi$ of the density contrast δ and the divergence of the displacement field ψ, respectively. The confidence intervals given correspond to the 1-σ standard deviations among eight realizations. In all cases, γ_1 and γ_2 are reduced when measured from ψ instead of δ.

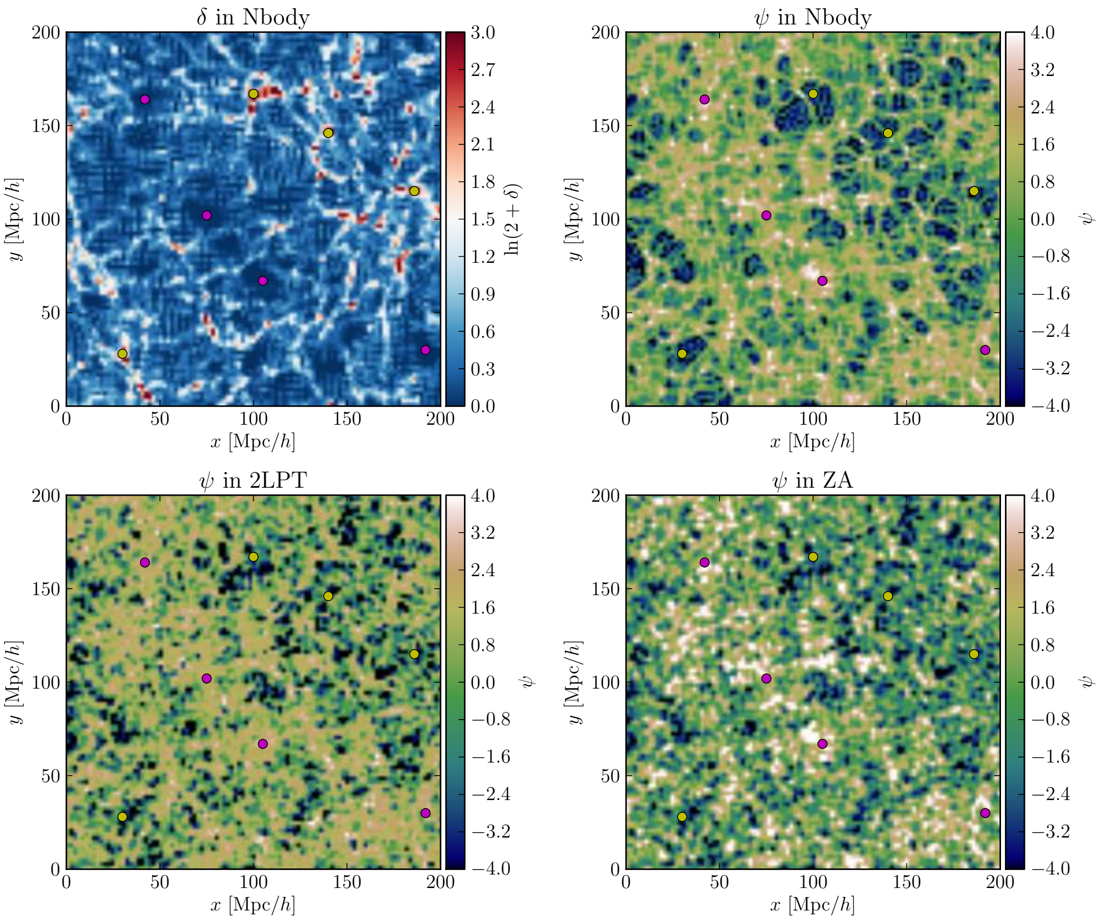

Figure 2.12: Slices of the divergence of the displacement field, ψ, on a Lagrangian sheet of 512^2 particles from a 512^3-particle simulation of box size 1024 Mpc/h, run to redshift zero. For clarity we show only a 200 Mpc/h region. Each pixel corresponds to a particle. The particle distribution is determined using respectively a full N-body simulation, the Zel'dovich approximation (ZA) and second-order Lagrangian perturbation theory (2LPT). In the upper left panel, the density contrast δ in the N-body simulation is shown, after binning on a 512^3-voxel grid. To guide the eye, some clusters and voids are identified by yellow and purple dots, respectively. The "lakes", Lagrangian regions that have collapsed to form halos, are only visible in the N-body simulation, while the "mountains", Lagrangian regions corresponding to cosmic voids, are well reproduced by LPT.

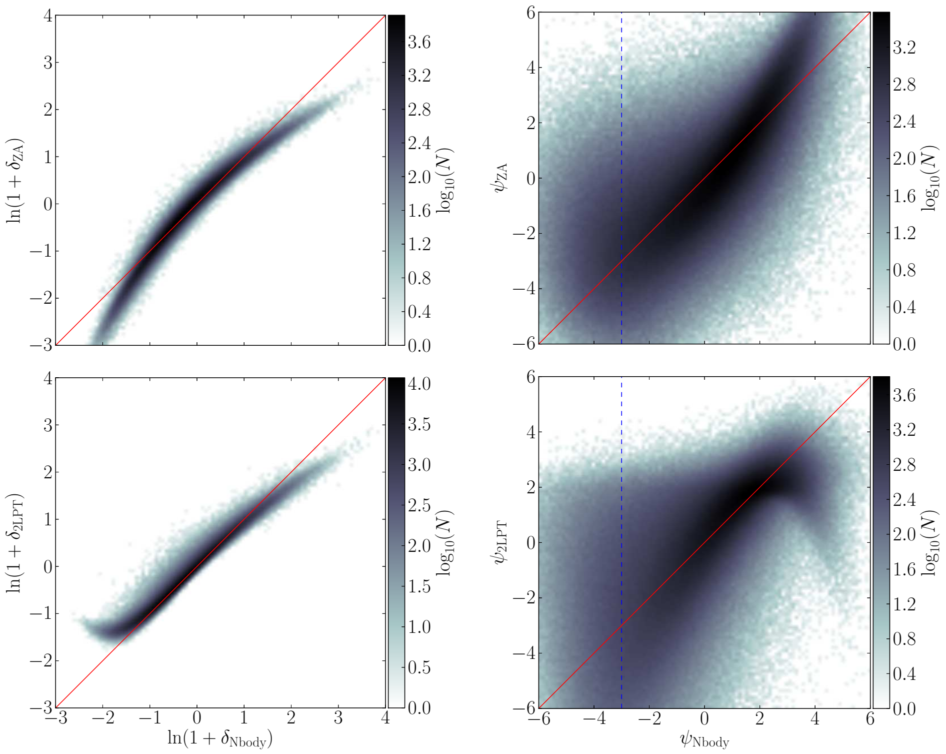

Figure 2.13: *Left panel.* Two-dimensional histograms comparing particle densities evolved with full N-body dynamics (the x-axis) to densities in the LPT-evolved particle distributions (the y-axis). The red lines show the ideal $y = x$ locus. A turn-up at low densities is visible with 2LPT, meaning that some overdense regions are predicted where there should be deep voids. *Right panel.* Same plot for the divergence of the displacement field ψ. Negative ψ corresponds to overdensities and positive ψ correspond to underdensities. The dotted blue line shows the collapse barrier at $\psi = -3$ where particle get clustered in full gravity. The scatter is bigger with ψ than with δ, in particular in overdensities, since with LPT, particles do not cluster. The turn-up at low densities with 2LPT, observed with the density contrast, is also visible with the divergence of the displacement field.

(see e.g. Bouchet *et al.*, 1995 or Bernardeau *et al.*, 2002 for overviews),

$$\mathbf{x} \equiv \mathbf{q} + \mathbf{\Psi}(\mathbf{q}). \tag{2.4}$$

It is important to note that, though $\mathbf{\Psi}(\mathbf{q})$ is *a priori* a full three-dimensional vector field, it is curl-free up to second order in LPT (appendix D in Bernardeau, 1994 or Bernardeau *et al.*, 2002 for a review). In this thesis, we do not consider perturbative contributions beyond 2LPT.

Let $\psi(\mathbf{q}) \equiv \nabla_\mathbf{q} \cdot \mathbf{\Psi}(\mathbf{q})$ denote the divergence of the displacement field, where $\nabla_\mathbf{q}$ is the divergence operator in Lagrangian coordinates. ψ quantifies the angle-averaged spatial-stretching of the Lagrangian dark matter "sheet" in comoving coordinates (Neyrinck, 2013). Let $\mathcal{P}_{\psi,\mathrm{LPT}}$ and $\mathcal{P}_{\psi,\mathrm{Nbody}}$ be the one-point probability distribution functions for the divergence of the displacement field in LPT and in full N-body fields, respectively. We denote by $\mathcal{P}_\delta$ the corresponding pdfs for the Eulerian density contrast.

In figure 2.11, we show the pdfs of ψ for the ZA, 2LPT and full N-body gravity. The most important feature of ψ is that, whatever the model for structure formation, the pdf exhibits reduced non-Gaussianity compared to the pdf for the density contrast δ (see the upper panel of figure 2.1 for comparison). The main reason is that $\mathcal{P}_\delta$, unlike $\mathcal{P}_\psi$, is tied down to zero at $\delta = -1$. It is highly non-Gaussian in the final conditions, both in N-body simulations and in approximations to the true dynamics. For a quantitative analysis, we looked at the first and second-order non-Gaussianity statistics: the skewness γ_1 and the excess kurtosis γ_2,

$$\gamma_1 \equiv \frac{\mu_3}{\sigma^3} \quad \text{and} \quad \gamma_2 \equiv \frac{\mu_4}{\sigma^4} - 3, \tag{2.5}$$

where μ_n is the n-th moment about the mean and σ is the standard deviation. We estimated γ_1 and γ_2 at redshift zero in our simulations, in the one-point statistics of the density contrast δ and of the divergence of the displacement field ψ. The results are shown in table 2.1. In all cases, we found that both γ_1 and γ_2 are much smaller when measured from $\mathcal{P}_\psi$ instead of $\mathcal{P}_\delta$.

At linear order in Lagrangian perturbation theory (the Zel'dovich approximation), the divergence of the displacement field is proportional to the density contrast in the initial conditions, $\delta(\mathbf{q})$, scaling with the negative growth factor, $-D_1(\tau)$:

$$\psi^{(1)}(\mathbf{q},\tau) = \nabla_\mathbf{q} \cdot \mathbf{\Psi}^{(1)}(\mathbf{q},\tau) = -D_1(\tau)\,\delta(\mathbf{q}). \tag{2.6}$$

Since we take Gaussian initial conditions, the pdf for ψ is Gaussian at any time with the ZA. In full gravity, non-linear evolution slightly breaks Gaussianity. $\mathcal{P}_{\psi,\mathrm{Nbody}}$ is slightly skewed towards negative values while its mode gets shifted around $\psi \approx 1$. Taking into account non-local effects, 2LPT tries to get closer to the shape observed in N-body simulations, but the correct skewness is overshot and the pdf is exceedingly peaked.

Figure 2.12 shows a slice of the divergence of the displacement field, measured at redshift zero for particles occupying a flat 512^2-pixel Lagrangian sheet from one of our simulations. For comparison, see also the figures in Mohayaee *et al.* (2006); Pueblas & Scoccimarro (2009) and Neyrinck (2013). We used the color scheme of the latter paper, suggesting a topographical analogy when working in Lagrangian coordinates. As structures take shape, ψ departs from its initial value; it takes positive values in underdensities and negative values in overdensities. The shape of voids (the "mountains") is found to be reasonably similar in LPT and in the N-body simulation. For this reason, the influence of late-time non-linear effects in voids is milder as compared to overdense structures, which makes them easier to relate to the initial conditions. However, in overdense regions where ψ decreases, it is not allowed to take arbitrary values: where gravitational collapse occurs, "lakes" form and ψ gets stuck around a collapse barrier, $\psi \approx -3$. As expected, these "lakes", corresponding to virialized clusters, can only be found in N-body simulations, since LPT fails to accurately describe the highly non-linear physics involved. A small bump at $\psi = -3$ is visible in $\mathcal{P}_{\psi,\mathrm{Nbody}}$ (see figure 2.11). We checked that this bump gets more visible in higher mass-resolution simulations (200 Mpc/h box for 256^3 particles), where matter is more clustered. This means that part of the information about gravitational clustering can be found in the one-point statistics of ψ. Of course, the complete description of halos requires to precisely account for the shape of the "lakes", which can only be done via higher-order correlation functions. More generally, it is possible to use Lagrangian information in order to classify structures of the cosmic web. In particular, DIVA (Lavaux & Wandelt, 2010) uses the shear of the displacement field and ORIGAMI (Falck, Neyrinck & Szalay, 2012) the number of phase-space folds. While these techniques cannot be straightforwardly used for the analysis of galaxy surveys, where we lack Lagrangian information, recently proposed techniques for physical inference of the initial conditions (chapters 4 and 5 Jasche & Wandelt, 2013a; Jasche, Leclercq & Wandelt, 2015) should allow their use with observational data.

Figure 2.13 shows two-dimensional histograms comparing N-body simulations to the LPT realizations for the density contrast δ and the divergence of the displacement field ψ. At this point, it is useful to note that a good mapping exists in the case where the relation shown is monotonic and the scatter is narrow. As pointed out by Sahni & Shandarin (1996) and Neyrinck (2013), matter in the substructure of 2LPT-voids has incorrect statistical properties: there are overdense particles in the low density region of the 2LPT δ-scatter plot. This degeneracy is also visible in the $\psi > 0$ region of the 2LPT ψ-scatter plot. On average, the scatter is bigger with ψ than with δ, in particular in overdensities ($\psi < 0$), since with LPT, particles do not cluster: ψ takes any value between 2 and -3 where it should remain around -3.

Summing up our discussions in this paragraph, we analyzed the relative merits of the Lagrangian divergence of the displacement field ψ, and the Eulerian density contrast δ at the level of one-point statistics. The important differences are the following:

1. $\mathbf{\Psi}$ being irrotational up to order two, its divergence ψ contains nearly all information on the displacement field in one dimension, instead of three. The collapse barrier at $\psi = -3$ is visible in $\mathcal{P}_\psi$ for N-body simulations but not for LPT. A part of the information about non-linear gravitational clustering is therefore encoded in the one-point statistics of ψ.

2. ψ exhibits much fewer gravitationally-induced non-Gaussian features than δ in the final conditions (figure 2.11 and table 2.1).

3. However, the values of ψ are more scattered than the values of δ with respect to the true dynamics (figure 2.13), meaning that an unambiguous mapping is more difficult.

2.2.2 Perturbative and non-perturbative prescriptions for ψ

Even if ψ does not contain all the information about the vector displacement field $\mathbf{\Psi}$, knowledge of its evolution allows for methods to produce approximate particle realizations at the desired redshift, for the variety of cosmological applications described in the introduction of this thesis. These methods include, but are not limited to, the ZA and 2LPT. On the contrary, 3LPT involves a non-zero rotational component and comes at the expense of significantly greater complexity, for an agreement with full gravity that does not improve substantially (Buchert, Melott & Weiß, 1994; Bouchet *et al.*, 1995; Sahni & Shandarin, 1996). Since we have adopted the approximation that the displacement field is potential, we stop our analysis of LPT at second order. However, we will describe various non-perturbative schemes.

Importantly, ψ-based methods are essentially as fast as producing initial conditions for N-body simulations. Their implementation can be decomposed in several steps:

1. Generation of a voxel-wise initial-density field δ. It is typically a grf, given a prescription for the linear power spectrum (see section B.6), but it can also include primordial non-Gaussianities.

2. Estimation of ψ from δ at the desired redshift.

3. Generation of the final vector displacement field $\mathbf{\Psi}$ from ψ with an inverse-divergence operator.

4. Application of $\mathbf{\Psi}$ to the particles of a regular Lagrangian lattice to get their final positions.

In practice, steps 1 and 3 are performed in Fourier space, using fast Fourier transforms to translate between configuration space and Fourier space when necessary. In the remainder of this paragraph, we review various prescriptions that have been proposed in the literature to estimate $\psi(\mathbf{q}, \tau)$ from $\delta(\mathbf{q})$ (step 2).

The Zel'dovich approximation. The first scheme, already studied in section 1.5.2, is the ZA (equation (1.128)),

$$\psi_{\mathrm{ZA}}(\mathbf{q}, \tau) = -D_1(\tau)\, \delta(\mathbf{q}) \equiv -\delta_{\mathrm{L}}(\mathbf{q}, \tau). \tag{2.7}$$

The ZA allows to separate prescriptions for ψ into two classes: *local* Lagrangian approximations, where ψ depends only on its linear value, $\psi_{\mathrm{L}}(\mathbf{q}, \tau) \equiv -\delta_{\mathrm{L}}(\mathbf{q}, \tau)$ and *non-local* ones (e.g. higher-order LPT) where ψ depends on derivatives of ψ_{L} as well (which means that the behavior of a Lagrangian particle depends on its neighbours).

Second-order Lagrangian perturbation theory. In 2LPT, the non-local prescription for ψ is (see equation (1.133))

$$\psi_{\mathrm{2LPT}}(\mathbf{q},\tau) = -D_1(\tau)\Delta_{\mathbf{q}}\phi^{(1)}(\mathbf{q}) + D_2(\tau)\Delta_{\mathbf{q}}\phi^{(2)}(\mathbf{q}), \tag{2.8}$$

where the Lagrangian potentials follow the Poisson-like equations (1.134) and (1.135). As pointed out by Neyrinck (2013), since 2LPT is a second-order scheme, ψ_{2LPT} is roughly parabolic in the local δ_{L}, which yields, using $D_2(\tau) \approx -\frac{3}{7}D_1^2(\tau)$ (Bouchet *et al.*, 1995),

$$\psi_{\mathrm{2LPT}}(\mathbf{q},\tau) \approx \psi_{\mathrm{2LPT,parab}}(\mathbf{q},\tau) \equiv -\delta_{\mathrm{L}}(\mathbf{q},\tau) + \frac{1}{7}\left(\delta_{\mathrm{L}}(\mathbf{q},\tau)\right)^2 . \tag{2.9}$$

The spherical collapse approximation. Bernardeau (1994) provides a simple formula for the time-evolution (collapse or expansion) of a spherical Lagrangian volume element, independent of cosmological parameters:

$$V(\mathbf{q},\tau) = V(\mathbf{q})\left(1 - \frac{2}{3}\delta_{\mathrm{L}}(\mathbf{q},\tau)\right)^{3/2} . \tag{2.10}$$

Building upon this result, Mohayaee *et al.* (2006); Lavaux (2008) and Neyrinck (2013) derived a prescription for the divergence of the displacement field. Considering the isotropic stretch of a Lagrangian mass element that occupies a cube of side length $1+\psi/3$ (giving $\nabla_{\mathbf{q}}\cdot\mathbf{\Psi} = \psi$), mass conservation imposes

$$\frac{V(\mathbf{q},\tau)}{V(\mathbf{q})} = \frac{1}{1+\delta} = \left(1+\frac{\psi}{3}\right)^3 . \tag{2.11}$$

Equations (2.10) and (2.11) yield

$$\psi = 3\left(\sqrt{1-\frac{2}{3}\delta_{\mathrm{L}}} - 1\right) . \tag{2.12}$$

However, there exists no solution for $\delta_{\mathrm{L}} > 3/2$. Neyrinck (2013) proposes to fix $\psi = -3$ in such volume elements. This corresponds to the ideal case of a Lagrangian patch contracting to a single point ($\nabla_{\mathbf{q}}\cdot\mathbf{x} = 0$). The final prescription for the spherical collapse (SC) approximation is then

$$\psi_{\mathrm{SC}}(\mathbf{q},\tau) = \begin{cases} 3\left(\sqrt{1-\frac{2}{3}\delta_{\mathrm{L}}(\mathbf{q},\tau)} - 1\right) & \text{if } \delta_{\mathrm{L}} < 3/2, \\ -3 & \text{if } \delta_{\mathrm{L}} \geq 3/2. \end{cases} \tag{2.13}$$

One possible concern with this formula is that, in full gravity, there are roughly as many particles with $\psi > -3$ as with $\psi < -3$ (see e.g. trajectories in ψ as a function of the scale factor a, figure 7 in Neyrinck, 2013). Yet, this remains more correct than what happens with LPT, where ψ can take any negative value, indicating severe unphysical over-crossing of particles in collapsed structures.

Compared to LPT, the SC approximation gives reduced stream-crossing, better small-scale flows and one-point pdf correspondence to the results of full gravity. However, a significant drawback is its incorrect treatment of large-scale flows, leading to a negative offset in the large-scale power spectrum (figure 14 in Neyrinck, 2013).[2] LPT realizations, on the other hand, give more accurate large-scale power spectra, as well as improved cross-correlation to the density field evolved with full gravity.

Local Lagrangian approximations. The SC approximation belongs to a more general family of "local Lagrangian" approximations investigated by Protogeros & Scherrer (1997), parameterized by $1 \leq \alpha \leq 3$, the effective number of axes along which the considered volume element undergoes gravitational collapse. The corresponding density is given by

$$\delta_\alpha(\psi) = \left(1+\frac{\psi}{\alpha}\right)^{-\alpha} - 1. \tag{2.14}$$

[2] An empirical correction may be added to the SC formula to fix this issue: multiplying δ_{L} in equation (2.13) by a factor such that the large-scale power spectrum of SC realizations agrees with that of LPT realizations (Neyrinck, 2013). See also the paragraph on MUSCLE.

Here, ψ is the actual non-linear displacement-divergence of a volume element, not necessarily related to the linearly evolved ψ_L. From equations (2.10) and (2.11), we get

$$\delta = \left(1 - \frac{2}{3}\delta_\mathrm{L}\right)^{-3/2} - 1 = \left(1 + \frac{2}{3}\psi_\mathrm{L}\right)^{-3/2} - 1, \tag{2.15}$$

therefore the spherical collapse approximation corresponds to the case $\alpha = 3/2$ for $\psi = \psi_\mathrm{L}$. The cubic mass-element approximation that would follow directly from using equation (2.11) without equation (2.10) corresponds to the case $\alpha = 3$ for the full ψ. Neyrinck (2013) shows that the δ–ψ relation closely follows $\delta_3(\psi)$ for $\psi < 0$, whereas for $\psi > 0$ the result is between $\delta_3(\psi)$ and $\delta_{3/2}(\psi)$, when accounting for the anisotropy of gravitational expansion.

Augmented Lagrangian Perturbation Theory. As discussed before, LPT correctly describes large scales and SC more accurately captures small, collapsed structures. Kitaura & Heß (2013) proposed a recipe to interpolate between the LPT displacement on large scales and the SC displacement on small scales, calling it Augmented Lagrangian Perturbation Theory (ALPT). It reads

$$\psi_\mathrm{ALPT}(\mathbf{q}, \tau) = (K_{R_\mathrm{s}} * \psi_\mathrm{2LPT})(\mathbf{q}, \tau) + [(1 - K_{R_\mathrm{s}}) * \psi_\mathrm{SC}](\mathbf{q}, \tau), \tag{2.16}$$

or, in Lagrangian Fourier space,[3]

$$\psi_\mathrm{ALPT}(\boldsymbol{\kappa}, \tau) = K_{R_\mathrm{s}}(\kappa)\, \psi_\mathrm{2LPT}(\boldsymbol{\kappa}, \tau) + [1 - K_{R_\mathrm{s}}(\kappa)]\, \psi_\mathrm{SC}(\boldsymbol{\kappa}, \tau). \tag{2.17}$$

This method introduces a free parameter, R_s, the width of the Gaussian kernel used in the above equations to filter between large and small displacements, $K_{R_\mathrm{s}}(k) \propto \exp(-k^2/2 \times (R_\mathrm{s}/2\pi)^2)$. In numerical experiments, Kitaura & Heß (2013) empirically found that the range $R_\mathrm{s} = 4-5$ Mpc/h yields the best density cross-correlation to full gravity.

Multi-scale spherical collapse evolution. Neyrinck (2016) argued that the major deficiency in the SC approximation is its treatment of the void-in-cloud process (in the terminology originally introduced by Sheth & van de Weygaert, 2004), i.e. of small underdensities in larger-scale overdensities. Such regions should eventually collapse, which is not accounted for in SC. To overcome this problem, he proposes to use the SC prescription as a function of the initial density contrast on multiple Gaussian-smoothed scales, thus including the void-in-cloud process. The resulting parameter-free scheme, MUSCLE (MUltiscale Spherical-CoLlapse Evolution), mathematically reads

$$\psi_\mathrm{MUSCLE}(\mathbf{q}, \tau) = \begin{cases} 3\left(\sqrt{1 - \frac{2}{3}\delta_\mathrm{L}(\mathbf{q}, \tau)} - 1\right) & \text{if } \delta_\mathrm{L} < 3/2 \text{ and } \forall R_\mathrm{s} \geq R_\mathrm{i}, K_{R_\mathrm{s}} * \delta_\mathrm{L} < 3/2, \\ -3 & \text{otherwise,} \end{cases} \tag{2.18}$$

where R_i is the resolution of the initial density field $\delta(\mathbf{q})$, and $K_{R_\mathrm{s}} * \delta_\mathrm{L}$ is the linearly extrapolated initial density field, smoothed using a Gaussian kernel of width R_s. In practice, a finite number of scales $r > R_\mathrm{i}$ have to be tried (for example $r = 2^n R_\mathrm{i}$ for integers $0 \leq n \leq n_\mathrm{max}$ such that $2^{n_\mathrm{max}} R_\mathrm{i} \leq L$ and $2^{n_\mathrm{max}+1} R_\mathrm{i} > L$).

Neyrinck (2016) checked that MUSCLE corrects the problems of SC at large scales and outperforms the ZA and 2LPT in terms of the density cross-correlation to full gravity.

2.2.3 Non-linear evolution of ψ and generation of a vector part

Beyond the approximations presented in the previous section, Chan (2014) analyzed the non-linear evolution of $\boldsymbol{\Psi}$ in full gravity, splitting it into its scalar and vector parts (the so-called "Helmholtz decomposition"):

$$\boldsymbol{\Psi}(\mathbf{q}) = \nabla_\mathbf{q}\phi(\mathbf{q}) + \nabla_\mathbf{q} \times \mathbf{A}(\mathbf{q}), \tag{2.19}$$

with

$$\Delta_\mathbf{q}\phi = \nabla_\mathbf{q} \cdot \boldsymbol{\Psi}(\mathbf{q}), \tag{2.20}$$

$$\Delta_\mathbf{q}\mathbf{A}(\mathbf{q}) = -\nabla_\mathbf{q} \times \boldsymbol{\Psi}(\mathbf{q}). \tag{2.21}$$

[3] We denote by $\boldsymbol{\kappa}$ a Fourier mode on the Lagrangian grid, κ its norm.

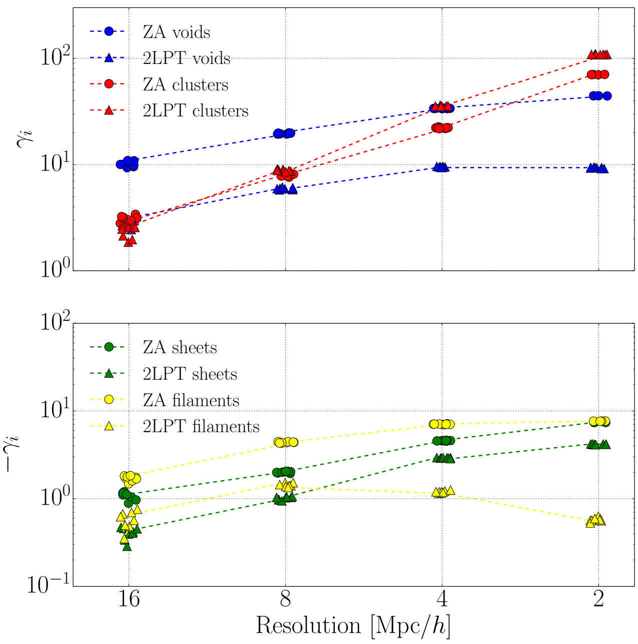

Figure 2.14: Relative volume fraction of voids, sheets, filaments and clusters predicted by LPT, compared to N-body simulations, as a function of the resolution used for the definition of the density fields. The points are sightly randomized on the x-axis for clarity. The estimators γ_i are defined by eq (2.22). Eight realizations of the ZA (circles) and 2LPT (triangles) are compared to the corresponding N-body realization, for various resolutions. The volume fraction of incorrectly predicted structures in LPT generally increases with increasing resolution.

Looking at two-point statistics of $\mathbf{\Psi}$, he found that shell-crossing leads to a suppression of small-scale power in the scalar part, and, subdominantly, to the generation of a vector contribution. Even at late-time and non-linear scales, the scalar part of the displacement field remains the dominant contribution. The rotational component is much smaller and does not have a coherent large-scale component. Therefore, the potential approximation is still good even when shell-crossing is non-negligible.

However, as pointed out by Neyrinck (2016), even if we neglect the rotational component, there is still a long way to go before we can perfectly predict ψ. Variants of LPT, such as ALPT (primarily motivated by the agreement in the scatter plot of final versus initial ψ – see figure 6 in Neyrinck, 2013 and figure 10 in Chan, 2014) or the inclusion of a suppression factor in the LPT displacement potential (Chan, 2014 – designed for fitting the non-linear power spectrum of $\mathbf{\Psi}$) extract information from simulations by taking the average of some statistics. Since shell-crossing is a highly non-linear process, it may not be surprising that such approaches yield limited success compared to standard LPT for some other statistics (such as the density power spectrum or phase accuracy). This suggests that a more detailed understanding and modeling of the small-scale physics beyond the simple phenomenological approach is required for improvement in ψ-based schemes, which would substantially increase the accuracy of particle realizations.

2.3 Comparison of structure types in LPT and N-body dynamics

This section draws from section II.B. in Leclercq *et al.* (2013).

In this section, we perform a study of differences in structure types in density fields predicted by LPT and N-body simulations. We employ the web-type classification algorithm proposed by Hahn *et al.* (2007a), which relies on estimating the eigenvalues of the Hessian of the gravitational potential (see section C.2). This algorithm dissects the voxels into four different web types (voids, sheets, filaments and clusters). Due to the different representations of the non-linear regime of structure formation, we expect differences in structure types in LPT and N-body simulations. In particular, overdense clusters are objects in the strongly non-linear regime, far beyond shell-crossing, where predictions of LPT fail, while underdense voids are believed to be better apprehended (e.g. Bernardeau *et al.*, 2002).

As an indicator of the mismatch between the volume occupied by different structure types in LPT and N-body dynamics, we use the quantities γ_i defined by

$$\gamma_i \equiv \frac{N_i^{\mathrm{LPT}} - N_i^{\mathrm{Nbody}}}{N_i^{\mathrm{Nbody}}}, \tag{2.22}$$

where i indexes one of the four structure types (T_0 = void, T_1 = sheet, T_2 = filament, T_3 = cluster), and N_i^{LPT} and N_i^{Nbody} are the numbers of voxels flagged as belonging to a structure of type T_i, in corresponding LPT and in N-body realizations, respectively. At fixed resolution, corresponding realizations have the same total number of voxels N_{tot}, so we also have

$$\gamma_i = \frac{\mathrm{VFF}_i^{\mathrm{LPT}}}{\mathrm{VFF}_i^{\mathrm{Nbody}}} - 1, \tag{2.23}$$

where the volume filling fraction of structure type T_i is defined by $\mathrm{VFF}_i \equiv N_i/N_{\mathrm{tot}}$.

In figure 2.14, we plot γ_i as a function of the voxel size used to define the density fields. γ_i is positive for clusters and voids, and negative for sheets and filaments, meaning that too large cluster and void regions are predicted in LPT, at the detriment of sheets and filaments. More specifically, LPT predicts fuzzier halos than N-body dynamics, and incorrectly predicts the surroundings of voids as part of them. This result indicates that even though LPT and N-body fields look visually similar, there are crucial differences in the representation of structure types. As demonstrated by figure 2.14, this mismatch increases with increasing resolution. This effect is of general interest when employing LPT in LSS data analysis.