Part III

The non-linear regime of structure formation

Remapping Lagrangian perturbation theory

Contents

"Everyone thinks of changing the world, but no one thinks of changing himself."
— Leo Tolstoy
Quoted in Bryan, Cameron & Allen (1999), *The Artist's Way at Work: Riding the Dragon*

Abstract

On the smallest scales, three-dimensional large-scale structure surveys contain a wealth of cosmological information which cannot be trivially extracted due to the non-linear dynamical evolution of the density field. Lagrangian perturbation theory is widely applied to the generation of mock halo catalogs and data analysis. In this chapter, we propose a method designed to improve the correspondence between these density fields, in the mildly non-linear regime. We develop a computationally fast and flexible tool for a variety of cosmological applications. Our method is based on a remapping of the approximately-evolved density field, using information extracted from N-body simulations. The remapping procedure consists of replacing the one-point distribution of the density contrast by one which better accounts for the full gravitational dynamics. As a result, we obtain a physically more pertinent density field on a point-by-point basis, while also improving higher-order statistics predicted by LPT. We quantify the approximation error in the power spectrum and in the bispectrum as a function of scale and redshift. Our remapping procedures improves one-, two- and three-point statistics at scales down to 8 Mpc/h.

This chapter is adapted from its corresponding publication, Leclercq *et al.* (2013).

6.1 Introduction

At present, observations of the three-dimensional LSS are major sources of information on the origin and evolution of the Universe. According to the current paradigm of cosmological structure formation, the presently observed structures formed via gravitational clustering of cold dark matter particles and condensation of baryonic matter in gravitational potential wells. Consequently, the large-scale matter distribution retains a memory of

its formation history, enabling us to study the homogeneous as well as the inhomogeneous evolution of our Universe.

Due to non-linearities involved in the formation process, at present there exists just limited analytic understanding of structure formation in terms of perturbative expansions in Eulerian or Lagrangian representations. Both of these approaches rely on a truncated sequence of momentum moments of the Vlasov-Poisson system, completed by fluid dynamic assumptions (see chapter 1 or e.g. Bernardeau *et al.*, 2002, and references therein). For this reason, the validity of these approaches ceases, once the evolution of the LSS enters the multi-stream regime (see e.g. Pueblas & Scoccimarro, 2009).

Nevertheless, Eulerian and Lagrangian approximations have been successfully applied to the analysis of three-dimensional density fields in regimes where they are still applicable, either at large scales or in the early Universe. Particularly, LPT captures significant mode-coupling information that is encoded beyond linear theory, such as large-scale flows and free-streaming, yielding three-dimensional matter distributions approximating those of full scale numerical simulations with reasonable accuracy (Moutarde *et al.*, 1991; Buchert, Melott & Weiß, 1994; Bouchet *et al.*, 1995; Scoccimarro, 1998, 2000; Scoccimarro & Sheth, 2002; Yoshisato *et al.*, 2006). Especially, second-order Lagrangian perturbation theory has been widely applied in data analysis and for fast generation of galaxy mock catalogs (e.g. PTHALOS: Scoccimarro & Sheth, 2002; Manera *et al.*, 2013; PINOC-CHIO: Monaco *et al.*, 2002; Monaco, Theuns & Taffoni, 2002; Taffoni, Monaco & Theuns, 2002; Heisenberg, Schäfer & Bartelmann, 2011; Monaco *et al.*, 2013) that can be useful to estimate error bounds when analyzing observations.

Modern cosmological data analysis has an increasing demand for analytic and computationally inexpensive models providing accurate representations of the mildly non-linear regime of structure formation. Over the years, various non-linear approximations and attempts to extend the validity of LPT have been proposed (see section 1.6). These include the spherical collapse model (Gunn & Gott, 1972; Bernardeau, 1994), the truncated Zel'dovich approximation (Melott, Pellman & Shandarin, 1994) and models with various forms for the velocity potential (Coles, Melott & Shandarin, 1993; Munshi & Starobinsky, 1994) or the addition of a viscosity term in the Euler equation (the adhesion model, Gurbatov, Saichev & Shandarin, 1989). Analytical techniques to improve the convergence and behavior of standard perturbation theory, successfully employed in quantum field theory and statistical physics, have also been applied in the context of gravitational clustering. These include renormalized perturbation theory (Crocce & Scoccimarro, 2006), the path integral formalism (Valageas, 2007), and the renormalization group flow (Matarrese & Pietroni, 2007). More recently, Tassev & Zaldarriaga (2012a,c) constructed a physical picture of the matter distribution in the mildly non-linear regime, and developed a method yielding improvements over LPT (Tassev, Zaldarriaga & Eisenstein, 2013), in particular at the scales relevant for baryon acoustic peak reconstruction (Tassev & Zaldarriaga, 2012b).

In this chapter, we propose a numerically efficient method designed to improve the correspondence between approximate models and full numerical simulations of gravitational large-scale structure formation. Generally, it can be applied to any approximate model of gravitational instability, but it is especially targeted to improving Lagrangian methods. We illustrate both these methods on fields evolved with LPT: at order one, the Zel'dovich approximation (Zel'dovich, 1970; Shandarin & Zel'dovich, 1989) and second-order Lagrangian perturbation theory.

LPT and *N*-body density fields are visually similar, which suggests that the properties of LPT could be improved by one-to-one mapping in voxel space, following a similar line of thoughts as the "Gaussianization" idea originally proposed by Weinberg (1992) and inspired existing techniques, widely used in cosmology (Melott, 1993; Croft *et al.*, 1998; Narayanan & Weinberg, 1998; Croft *et al.*, 1999; Feng & Fang, 2000; Neyrinck, Szapudi & Szalay, 2011; Yu *et al.*, 2011, 2012; Neyrinck & Yang, 2013). The method described in this chapter is based on a *remapping* of the approximately evolved particle distribution using information extracted from *N*-body simulations. It basically consists of replacing the one-point distribution of the approximately evolved distribution by one which better accounts for the full gravitational system. In this fashion, we adjust the one-point distribution to construct a physically more reasonable representation of the three-dimensional matter distribution, while retaining or improving higher order statistics, described already reasonably well by the ZA (Zel'dovich, 1970; Doroshkevich, 1970a; Shandarin & Zel'dovich, 1989; Buchert, 1989; Moutarde *et al.*, 1991; Yoshisato *et al.*, 2006) and by 2LPT (Moutarde *et al.*, 1991; Buchert, Melott & Weiß, 1994; Bouchet *et al.*, 1995; Scoccimarro, 1998, 2000; Scoccimarro & Sheth, 2002).

Major problems with naive approaches to remapping LPT density fields arise from minor deviations in structure types represented by LPT models and full gravity. For this reason, in chapter 2, we discussed the different representations of clusters, voids, sheets, and filaments, predicted by LPT and *N*-body simulations.

Besides being of general interest for LSS data analysis, the insights gained from this comparison will allow us to improve the remapping procedure.

Implementing and testing the accuracy and the regime of validity of our method is essential, and is subject of the present chapter. Our study quantifies the approximation error as a function of scale in terms of a set of statistical diagnostics. From cosmographic measurements, σ_8 is known to be of order unity, which means that gravity becomes highly non-linear at some scale around 8 Mpc/h. Our method is expected to break down due to shell-crossing in LPT, at some scale larger than 8 Mpc/h. Achieving a resolution of 16 Mpc/h would already constitute substantial improvement with respect to existing methods, since non-linearities begin to affect even large-scale cosmographic measurements such as the determination of the baryon acoustic oscillations scale from galaxy surveys (about 125 Mpc/h, e.g. Eisenstein *et al.*, 2005). However, we explore the validity of the improvement at 8 Mpc/h down to 4 Mpc/h, to see to what extent we can push the limit for possible data analysis applications into the non-linear regime. Recall that in three-dimensional LSS surveys, the number of modes usable for cosmological exploitation scales as the cube of the largest wavenumber, k^3, meaning that even minor improvements in the mildly non-linear regime would give access to much more cosmological information from existing and upcoming observations.

As will be demonstrated, this method can be used to generate realizations of density fields much faster than N-body simulations. Even though approximate, these fast realizations of mock density fields may be sufficient to model the salient features of the non-linear density field for certain applications.

This chapter is structured as follows. In section 6.2, we describe the remapping procedure for the density contrast of present-day density fields, analyze the obstacles to straightforward application and present an improved method. In section 6.3, we apply the procedure to cosmological models using data from numerical simulations, we study the statistics of remapped fields and quantify the approximation error. We discuss our results and give our conclusions in section 6.4.

The setup of LPT and N-body simulations used in this chapter are described at the beginning of chapter 2.

6.2 Method

In this section, we discuss the remapping procedure and apply it to cosmological density fields evolved with LPT. Naively following the approach of Weinberg (1992) for present-day density fields yields the procedure described in section 6.2.1. This approach is not entirely satisfactory and we analyze the reasons for its shortcomings in section 6.2.2. In consequence, we propose a improvement of the remapping procedure in section 6.2.3. The properties of the remapping function are examined in section 6.2.4.

6.2.1 Remapping procedure

In this section, we describe the remapping algorithm used to go from a low-redshift realization of a density field evolved with LPT to one evolved with full N-body gravitational dynamics. Note that both fields obey the same initial conditions but are evolved by different physical models.

Density fields are defined on Cartesian grids of cubic voxels. Linear gravitational evolution exactly maintains the relative amplitude of fluctuations in different voxels. Due to mode coupling, positive and negative fluctuations grow at different rates in the non-linear regime, but even non-linear evolution tends to preserve the *rank order* of the voxels, sorted by density.

The one-point probability distribution functions and the cumulative distribution functions (cdf) of the final density fields, evolved with either LPT or full N-body gravitational dynamics, exhibit similar, but not identical shapes. This result suggests a way to improve the approximation with information extracted from the N-body simulation: maintain the rank order of the voxels, but reassign densities so that the two cdfs match. The method therefore resembles the "Gaussianization" procedure proposed by Weinberg (1992), an attempt to reconstruct the initial conditions of a density field from its final cdf.

Let $\mathcal{P}_{\mathrm{LPT}}$ and $\mathcal{P}_{\mathrm{Nbody}}$ denote the probability distribution functions for the density contrast in the LPT and in the full N-body density fields, respectively. Let $\mathcal{C}_{\mathrm{LPT}}$ and $\mathcal{C}_{\mathrm{Nbody}}$ be their integrals, the cumulative distribution functions. $\mathcal{C}_{\mathrm{LPT}}(\delta_{\mathrm{LPT}})$ is the *fractional rank* for δ_{LPT} i.e. the probability that the density contrast at a given voxel is smaller than δ_{LPT}, $\mathcal{P}_{\mathrm{LPT}}(\delta \leq \delta_{\mathrm{LPT}})$, and the analogous for the N-body field. The remapping procedure works as follows. A voxel with rank order δ_{LPT} is assigned a new density δ_{Nbody} such that

$$\mathcal{C}_{\mathrm{LPT}}(\delta_{\mathrm{LPT}}) = \mathcal{C}_{\mathrm{Nbody}}(\delta_{\mathrm{Nbody}}) \tag{6.1}$$
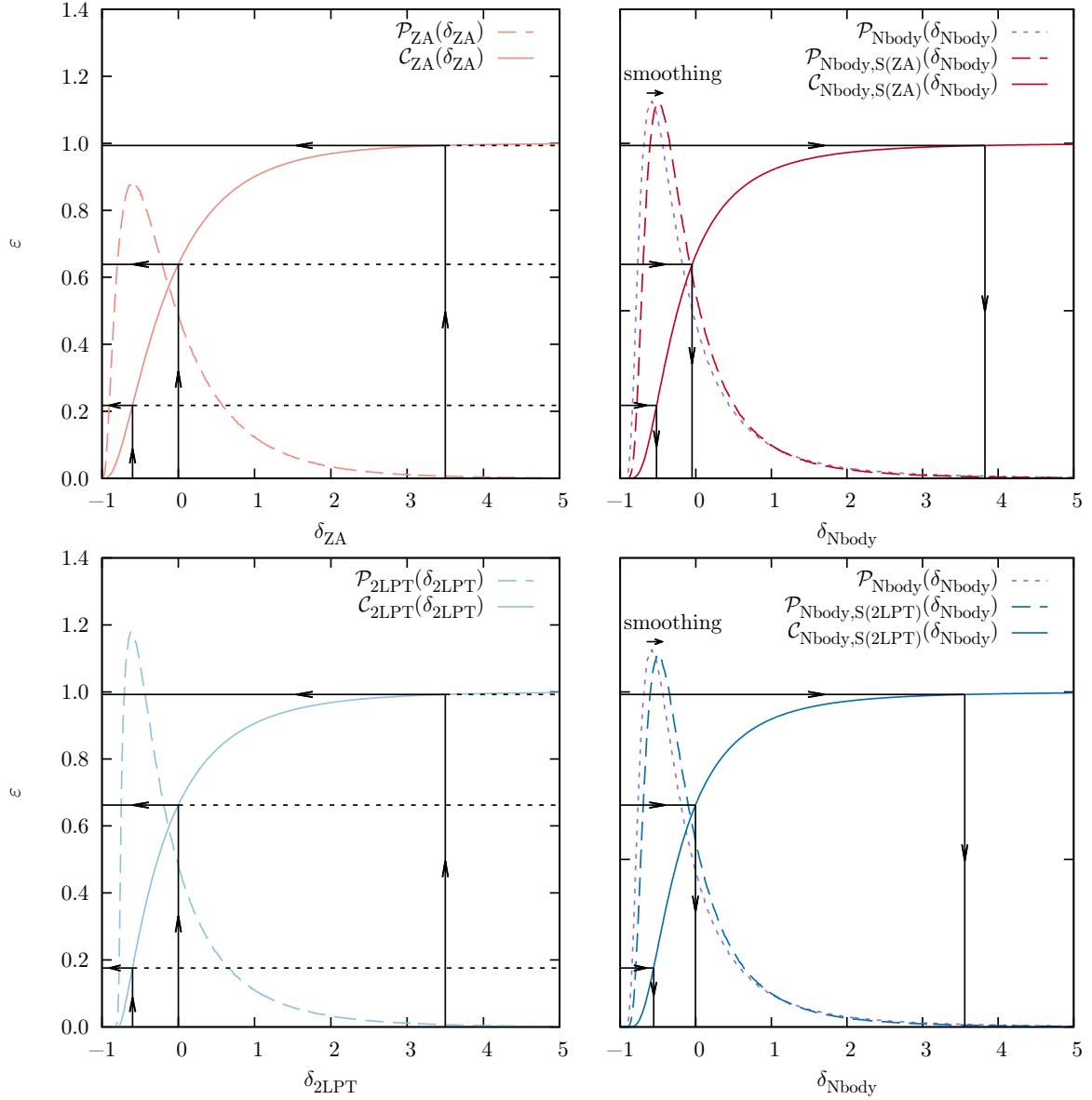
Figure 6.1: A graphical illustration of the improved remapping procedure at redshift zero, for the ZA (upper panel) and 2LPT (lower panel). On the right panels, the dotted purple curves are the probability distribution function for the density contrast in full N-body dynamics. The first step of the procedure is to smooth the N-body density field using the transfer function given by equation (6.2), which yields slightly different pdfs (dashed dark red and dark blue curves on the right panels). On the left panels, the dashed curves are the pdfs for the density contrast in the final density field, evolved from the same initial conditions with LPT (ZA: red line and 2LPT: blue line). The pdfs are computed on a 8 Mpc/h mesh (128^3-voxel grid) using the full statistics from eight realizations of 512^3-particles in a 1024 Mpc/h box with periodic boundary conditions. The solid curves are their respective integrals, the cumulative distribution functions. The second step is remapping, which assigns a voxel with density contrast δ_{LPT} and fractional rank ε the value of δ_{Nbody} that would have the same fractional rank in the smoothed N-body distribution (equation (6.3)). This remapping is illustrated for 2LPT with three sample points: $\delta_{\mathrm{2LPT}} = -0.60$ maps to $\delta_{\mathrm{Nbody}} = -0.56$, $\delta_{\mathrm{2LPT}} = 0.00$ maps to $\delta_{\mathrm{Nbody}} = -0.01$, and $\delta_{\mathrm{2LPT}} = 3.50$ maps to $\delta_{\mathrm{Nbody}} = 3.56$. The remapping procedure imposes the one-point distribution of the smoothed N-body field while maintaining the rank order of the LPT-evolved density fields. The last step is an increase of small-scale power in the remapped distribution using the reciprocal transfer function, equation (6.4).
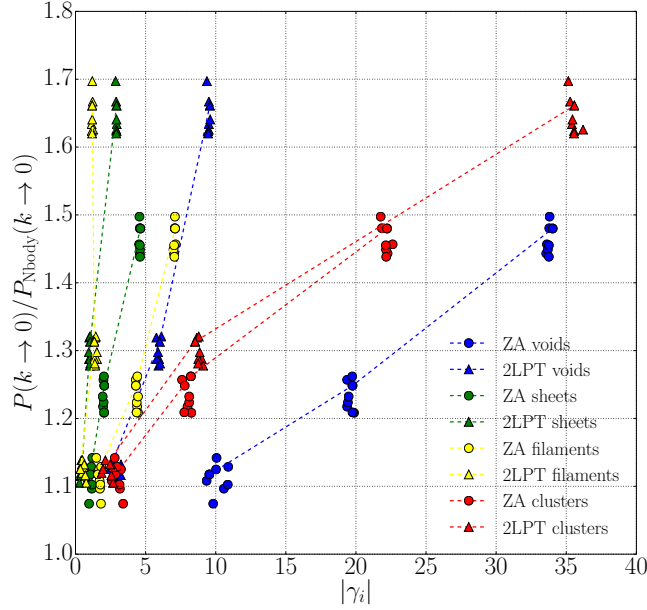
Figure 6.2: The large-scale bias in the power spectrum of density fields, remapped with the procedure described in section 6.2.1, as a function of the mismatch between structure types. The estimators γ_i are defined by equation (2.22). Eight realizations of the ZA (circles) and 2LPT (triangles) are compared to the corresponding N-body realization, for various resolutions (from bottom to top: 16 Mpc/h, 8 Mpc/h, 4 Mpc/h). The large-scale bias in the power spectra of remapped fields is strongly correlated to the volume fraction of structures incorrectly predicted by LPT.

(also see figure 6.1 for a schematic outline of this method). The left panels of figure 6.1 show $\mathcal{P}_{\mathrm{LPT}}$ (dashed curves) and the corresponding cumulative distributions, $\mathcal{C}_{\mathrm{LPT}}$ (solid curves). On the right panels, the dotted curves represent the pdf of the corresponding N-body realization, $\mathcal{P}_{\mathrm{Nbody}}$. Remapping assigns to a voxel with density contrast δ_{LPT} and fractional rank $\varepsilon = \mathcal{C}_{\mathrm{LPT}}(\delta_{\mathrm{LPT}})$ the value of δ_{Nbody} that would have the same fractional rank in the N-body distribution.

Since $\mathcal{C}_{\mathrm{Nbody}}$ contains exactly the same information as $\mathcal{P}_{\mathrm{Nbody}}$, the remapping procedure imposes the one-point distribution taken from the N-body-evolved density field while maintaining the rank order of the LPT-evolved density field. In other words, only the weight of underdensities and overdensities is modified, while their locations remain unchanged. In this fashion, we seek to adjust the density field while maintaining higher-order statistics provided by LPT with reasonable accuracy. We checked numerically that mass is always conserved in this procedure.

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

We implemented the remapping procedure described in the previous section and checked that we are able to modify LPT density fields so as to correctly reproduce the one-point distribution of a full N-body simulation. However, we experience a large-scale bias in the power spectrum, namely the amplitude of the spectrum of the remapped fields is slightly too high. Generally, a non-linear transformation in voxel space can change the variance of a field. This is consistent with the findings of Weinberg (1992), who found a similar effect in his reconstructions of the initial conditions, and who addressed the issue by rescaling the density field in Fourier space. However, such an approach will generally destroy the remapped one-point distribution, and may even further lead to Gibbs ringing effects which will make the remapped field unphysical.

The bias indicates a stronger clustering in the remapped density field. Since remapping is a local operation in voxel space, this large-scale bias means that erroneous remapping of small-scale structures affects the clustering at larger scales. To identify the cause of this bias, we performed a study of differences in structure types (voids, sheets, filaments, and clusters) in density fields predicted by LPT and N-body simulations. With this analysis, we wanted to understand the effect of remapping in different density and dynamical regimes of the LSS. The results are presented in section 2.3. We identified, in particular, a mismatch between the volume occupied by different structure types in LPT and N-body dynamics, quantified by the parameters γ_i defined by equation

(2.22).

In figure 6.2, we plot the large-scale bias observed in remapped fields obtained with the procedure of section 6.2.1 as a function of γ_i, for various resolutions. A strong correlation is observed between the bias and the mismatch in the volume occupied by different structure types. The difference in the prediction of the volume of extended objects is the cause of the bias: in clusters and in voids, remapping enhances a larger volume than should be enhanced, which yields on average a stronger clustering in the box.

6.2.3   Improvement of the remapping procedure

In the previous section, we noted that because too much volume of LPT density fields is mapped to the tails of the N-body one-point distribution (dense clusters and deep voids), the average two-point correlations are larger after remapping. Given this result, minor modifications of the procedure described in section 6.2.1 can solve the problem. We now propose to improve the remapping procedure of Weinberg (1992) for present-day density fields, in a similar fashion as in Narayanan & Weinberg (1998), combining it with transfer function techniques to deal with the mildly non-linear modes as in Tassev & Zaldarriaga (2012a,c). Our method works as follows.

1. We degrade the N-body density field to the same degree of smoothness as the LPT density field, by multiplying the Fourier modes of the density field by the transfer function

$$T(k) \equiv \sqrt{\frac{P_{\mathrm{LPT}}(k)}{P_{\mathrm{Nbody}}(k)}}. \tag{6.2}$$

   This steps yields a new density field, noted Nbody,S(LPT), whose power spectrum matches that of the LPT density field.

2. We remap the LPT density field in the fashion described in section 6.2.1, but using as a reference the cdf of the smoothed density field, $\mathcal{C}_{\mathrm{Nbody,S(LPT)}}$, instead of the full N-body density field (see figure 6.1). The remapping condition, equation (6.1), now reads

$$\mathcal{C}_{\mathrm{LPT}}(\delta_{\mathrm{LPT}}) = \mathcal{C}_{\mathrm{Nbody,S(LPT)}}(\delta_{\mathrm{Nbody}}). \tag{6.3}$$

3. We increase the power of small scales modes in the remapped distribution to the same value as in a full N-body simulation, using the reciprocal of the transfer function (6.2), namely

$$T^{-1}(k) = \sqrt{\frac{P_{\mathrm{Nbody}}(k)}{P_{\mathrm{LPT}}(k)}}. \tag{6.4}$$

This procedure cures the large-scale bias issue experienced with the simple implementation of the remapping described in section 6.2.1, without requiring any prior knowledge on the corresponding N-body simulation. As we will demonstrate in section 6.3, it yields improvement of one-, two- and three-point statistics of LPT.

6.2.4   Remapping function and transfer function

Since $\mathcal{C}_{\mathrm{LPT}}$ and $\mathcal{C}_{\mathrm{Nbody,S(LPT)}}$ are monotonically increasing functions, there is no ambiguity in the choice of δ_{Nbody}, and this procedure defines a *remapping function f* such that

$$\delta_{\mathrm{LPT}} \mapsto \delta_{\mathrm{Nbody}} = \mathcal{C}_{\mathrm{Nbody,S(LPT)}}^{-1}(\mathcal{C}_{\mathrm{LPT}}(\delta_{\mathrm{LPT}})) \equiv f(\delta_{\mathrm{LPT}}). \tag{6.5}$$

Establishing a remapping function f requires knowledge of both LPT and N-body density field statistics. Ideally, several realizations with different initial conditions should be combined in order to compute a precise remapping function. Indeed, a limited amount of available N-body simulations results in a lack of statistics and hence uncertainties for the remapping procedure in the high density regime. However, this effect is irrelevant from a practical point of view, since these high density events are very unlikely and affect only a negligible number of voxels. As a consequence this uncertainty will only affect to sub-percent level the usual statistical summaries of the density field. Note that in any case, if desired, the accuracy of the remapping function in the high density regime can be trivially enhanced by enlarging the size or number of N-body simulations used for

its construction. For the analysis presented in this chapter, the remapping functions have been computed using the full statistics from eight realizations of 512^3 particles in a 1024 Mpc/h box.

Note that once the relationship between the statistical behavior of the LPT fields and the full non-linear field is known, this procedure can be used on LPT realizations without the need of evolving corresponding N-body simulations. More specifically, the remapping function f (equation (6.5)) and the transfer function T (equation (6.2)) can be tabulated and stored, then used for the fast construction of a large number of large-scale structure density fields. Since producing LPT realizations is computationally faster than running full gravitational simulations by a factor of several hundreds, our method can be used to produce a large set of N-body-like realizations in a short time.

Some remapping functions are presented in figure 6.3. In each panel, the solid curves represent the remapping function f_z at redshift z, computed with the LPT and N-body simulations. The dashed black line shows the identity function. We checked that the remapping function converges to the identity function with increasing redshift, as expected. Critical values where the remapping function crosses the identity function are identified. Between these critical values, remapping either increases or decreases the local density.

The pdfs for the density contrast are evaluated on a grid after a CiC assignment of particles. This means that the remapping function *a priori* depends on the size of voxels. The problem of choosing a voxel size for the computation of the remapping function is linked to the more general problem of choosing a mesh for the CiC evaluation of density. Choosing too coarse a binning will result in an underestimation of the clustering of particles, whereas choosing too fine a binning will also result in artifacts in overdensities (some voxels may be empty due to their too small size). The right choice of voxel size for the evaluation of the remapping function is the one giving the best evaluation of the density contrast. This choice has to be made depending on the desired application of the remapped data.

The remapping function describes how the pdf for the density contrast is affected by non-linear structure formation. For this reason, it depends on the nature of the gravitational interaction, as described by LPT and by full N-body dynamics, but weakly depends on the detail of the cosmological parameters. We checked the cosmology-dependence of the remapping function in simulations with the dark matter and baryon density in the Universe, Ω_m and Ω_b, varying the WMAP-7 fiducial values (equation (2.1)) by $\pm\,3\sigma$ (still assuming a flat Universe):

$$\Omega_\Lambda = 0.750, \Omega_\mathrm{m} = 0.2494, \Omega_\mathrm{b} = 0.0428, \sigma_8 = 0.810, h = 0.704, n_\mathrm{s} = 0.967; \tag{6.6}$$

$$\Omega_\Lambda = 0.700, \Omega_\mathrm{m} = 0.2992, \Omega_\mathrm{b} = 0.0488, \sigma_8 = 0.810, h = 0.704, n_\mathrm{s} = 0.967. \tag{6.7}$$

Even for these models notably far from the fiducial values, we found that the remapping function almost perfectly overlaps that of our main analysis, for the density range $\delta \in [-1; 5]$, containing typically 98 to 99% of the voxels. We found a difference of less than 5% for $\delta = 5$ (see the left panel of figure 6.4).

The transfer function used in steps 1 and 3 of the improved procedure also exhibits very weak redshift-dependence, with deviations limited to a few percents at the smallest scales of interest of this work ($k \approx 0.4$ (Mpc/h)$^{-1}$, see the right panel of figure 6.4).

6.3   Statistics of remapped fields

In this section, we discuss the validity of the improved remapping procedure described in section 6.2.3, by studying the correlators of the remapped field in comparison to the input LPT and N-body fields. The remapping procedure based on the Eulerian density contrast essentially replaces the LPT one-point function by that of the smoothed N-body-evolved field. Since the position and shape of structures is left unchanged, we expect the higher-order correlators of the density field to be respected by the remapping procedure. Of particular interest is to check how remapping affects higher-order statistics and if possible improvements could be exploited in data analysis or artificial galaxy survey applications.

We implemented a numerical algorithm that computes and analyzes a remapped density field. The procedure can be divided in three steps:

1. We take as input two cosmological density fields, evolved from the same initial conditions with LPT (ZA or 2LPT) and with full N-body dynamics, and estimate the one-point statistics (pdf and cdf for δ) and the transfer function for this particular realization. We repeat this step for the eight realizations used in our analysis.
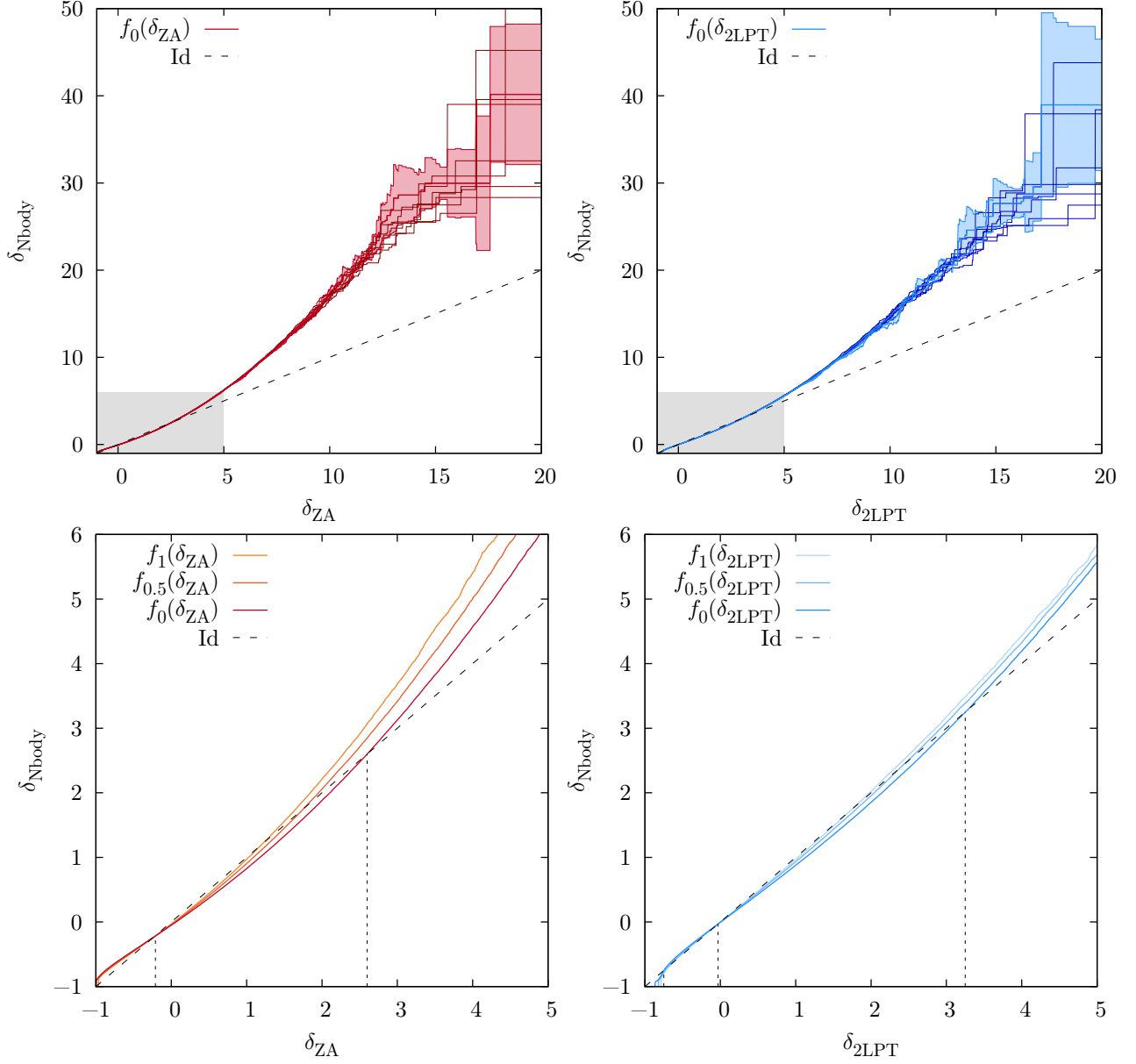
Figure 6.3: The remapping functions from LPT to smoothed N-body density fields, for the ZA (left panel) and 2LPT (right panel), all computed on a 8 Mpc/h mesh. The precise redshift-zero remapping functions f_0 (red and blue solid curves) have been computed using the full statistics from eight realizations (darker red and blue solid curves). The error bars shown are the 1-σ dispersion among the eight runs with reference to the full remapping function. The lower plots show the detail of the shaded area, in a density range containing most of the voxels. The redshift-dependence of the remapping function f_z is shown for $z = 1$, $z = 0.5$ and $z = 0$. The dashed line shows the identity function. Critical values of δ_{LPT} for which remapping does not change the local density are identified.
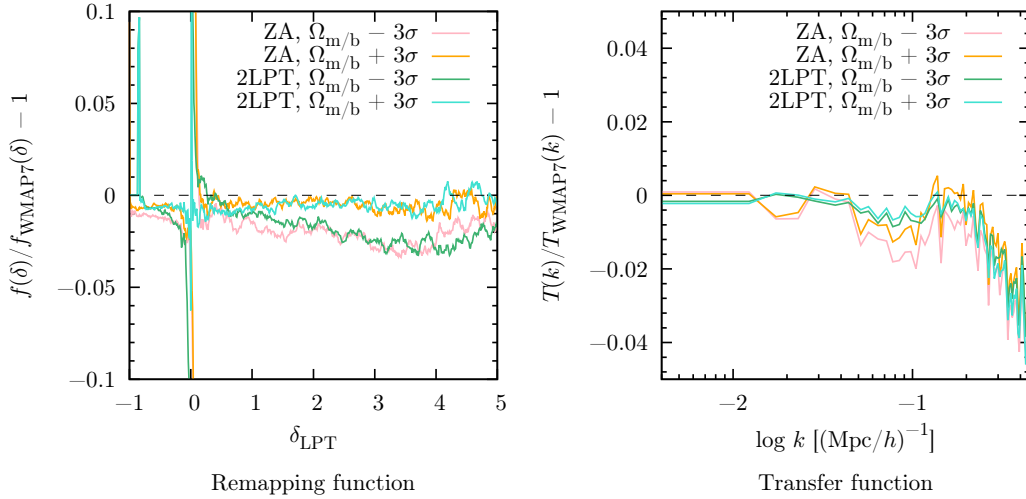
Figure 6.4: Relative deviations of the remapping function (left panel) and of the transfer function (right panel) for varying cosmological parameters (equations (6.6) and (6.7)), with respect to their behaviour in a fiducial cosmology (equation (2.1)).

2. We take as input all the one-point statistics computed with individual realizations, and we compute a precise remapping function using the full statistics of all available realizations, as described in section 6.2.1. The transfer function used as a reference is the mean of all available realizations. At this point, the remapping function and the transfer function can be tabulated and stored for later use, and N-body simulations are no longer required.

3. For each realization, we remap the density field using the improved three-step procedure described in section 6.2.3 and we analyze its correlation functions.

Our algorithm provides the one-point (section 6.3.1) and two-point (section 6.3.2) statistics. We used the code described in Gil-Marín *et al.* (2011, 2012) to study the three-point statistics (section 6.3.3). The results are presented below.

6.3.1   One-point statistics

The remapping procedure, described in section 6.2, is essentially a replacement of the cumulative distribution function of the density contrast δ of the input LPT-evolved field, $\mathcal{C}_{\mathrm{LPT}}$, by that of the reference N-body-evolved field after smoothing, $\mathcal{C}_{\mathrm{Nbody,S(LPT)}}$. After having applied the remapping procedure, we recomputed the pdf of the remapped field and verified that it matches that of the fiducial field as a sanity check.

Remapping and rescaling the density modes alters local density values but positions of structures remain unchanged. It is therefore important to check that remapping visually alters the LPT-evolved distribution in such a way that structures resemble more their N-body evolved counterparts. Figure 6.5 shows a slice of the density contrast δ, measured at redshift zero, on a 128^2-pixel sheet of a 512^3-particles realization in a 1024 Mpc/h box. The corresponding mesh size is 8 Mpc/h. Visually, remapped fields (ZARM and 2LPTRM) are closer to the full N-body result than their originals (ZA and 2LPT), with plausible particle distribution.

Since the improved remapping procedure involves a rescaling of the density modes in Fourier space (step 3), the pdf for the density contrast of the remapped fields is not guaranteed to be correct by construction, as would be the case with a naive remapping (section 6.2.1). Therefore, the one-point distribution has to be carefully checked at this point. In figure 2.1, we plot the pdf for the density contrast at redshift zero for N-body simulations and the approximately evolved fields: with the ZA and 2LPT alone, and after remapping (ZARM and 2LPTRM). It can be observed that the peaks of the pdfs get closer to the reference set by N-body dynamics and that the pdf of remapped fields accurately follows that of full gravitational dynamics for $\delta > 0$. The procedure being successful on average for one-point statistics and accurate for the most common events in overdensities, we expect the number count of objects such as clusters predicted by LPT to be made more robust by our procedure.
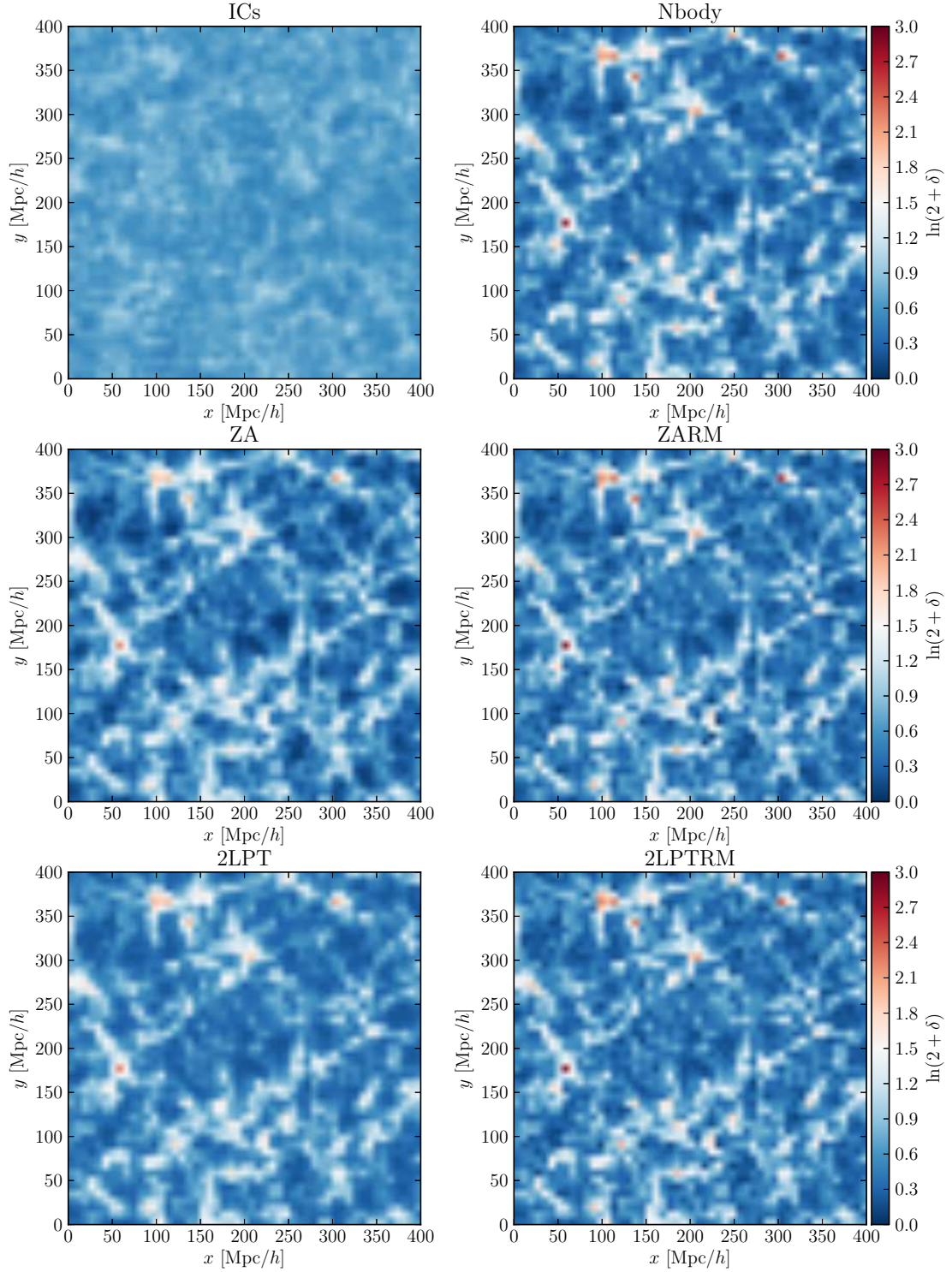
Figure 6.5: Redshift-zero density contrast on a 128^2-pixel slice of a 512^3-particles realization in a 1024 Mpc/h box with periodic boundary conditions. For clarity, the slice is limited to a square of 400 Mpc/h side, and the quantity shown is the log-density field, $\ln(2 + \delta)$. For comparison with the initial conditions, the density field at high redshift ($z = 5$) is shown in the top left corner. The redshift-zero density fields are determined using, from top to bottom and from left to right: a full N-body simulation, the Zel'dovich approximation, alone (ZA) and after remapping (ZARM), second-order Lagrangian perturbation theory, alone (2LPT) and after remapping (2LPTRM). The remapping and the transfer function operations are performed on a 128^3-voxel grid, corresponding to a mesh size of 8 Mpc/h.

6.3.2 Two-point statistics

6.3.2.1 Power spectrum

In figure 2.2, we plot the redshift-zero power spectrum of the different density fields.[1] The relative deviations of power spectra with reference to the density field computed with a full N-body simulation are presented in figures 2.3 and 2.4.

At high redshift ($z > 1$), we found no notable difference between the power spectrum of matter evolved with full N-body dynamics and that of LPT-remapped distributions. This indicates that our remapping procedure is increasingly successful as we go backwards in time towards the initial conditions, where LPT gives accurate results.

At low redshift, the power spectrum shape of remapped LPT fields is closer to the shape of the full non-linear power spectrum, turning down at smaller scales than the LPT power spectra. In particular, LPT fields exhibit more small-scale correlations after remapping, illustrating the success of our procedure in the mildly non-linear regime of large-scale structure formation.

Contrary to the density fields obtained via a naive remapping approach, whose power spectra exhibit a positive artificial offset at large scales as discussed in section 6.2.2, the fields obtained with the improved procedure have correct two-point statistics at all scales for coarse grids (down to 8 Mpc/h). For finer grids, a negative large-scale bias appears in the power spectrum, meaning that we have suppressed too much small-scale power in the N-body field in the first step of our procedure, which propagates to large scales with remapping. Comparing the panels of figures 2.3 and 2.4, it can be observed that this effect is suppressed at higher redshift and for coarser binning. We found that a voxel size of 8 Mpc/h is the best compromise, with a large-scale power spectrum compatible with the error bars and clear improvement at small scales, as can be seen in figure 2.2. This mesh size corresponds to the target resolution for analyses in the mildly non-linear regime, as discussed in the introduction of this chapter.

6.3.2.2 Fourier-space cross-correlation coefficient

In figure 2.5, we present the Fourier-space cross-correlation coefficient $R \equiv P_{\delta \times \delta'}/\sqrt{P_\delta P_{\delta'}}$ between the redshift-zero density field in the N-body simulation and several other density fields. At redshift zero and at small scales, the agreement is better with remapped LPT fields than with LPT alone, confirming the success of the remapping procedure to explore the mildly non-linear regime. In particular, the remapping of 2LPT predicts more than 96% level accuracy at $k = 0.4$ (Mpc/h)$^{-1}$ (corresponding to scales of 16 Mpc/h), where 2LPT gives only 93%. The cross-correlation coefficient indicates better agreements for the remapping of 2LPT than for the remapping of the ZA, which is consistent with the better performance of 2LPT in predicting the phases of the full N-body field (see section 2.1.2.2).

6.3.3 Three-point statistics

We analyzed the accuracy of our method beyond second-order statistics by studying the bispectrum, using the code described in Gil-Marín et al. (2011, 2012).[2] Figure 2.6 shows the redshift-zero bispectrum for equilateral triangles. The overall result is a clear improvement of the bispectrum of LPT-evolved fields with the remapping procedure, especially on the small scales shown, probing the mildly non-linear regime, $k \gtrsim 0.1$ (Mpc/h)$^{-1}$ corresponding to scales $\lesssim 62$ Mpc/h, where LPT predicts less three-point correlation than full gravity. At large scales, the approximation error remains $\lesssim 1\sigma$ of the estimated statistical uncertainty, even for a resolution of 8 Mpc/h and at late times ($z = 0$).

The relative deviations of approximate bispectra with reference to full N-body simulations are shown in figures 2.7, 2.8, 2.9 and 2.10. As expected, the success of our remapping procedure in exploring small scales ($k \gtrsim 0.1$ (Mpc/h)$^{-1}$) is increased for more coarsely-binned density fields (see figure 2.7) and at higher redshift (see figure 2.8). In figure 2.9 we examine the scale-dependence of the bispectrum for various triangle shapes. The precise dependence on the triangle shape at different scales is shown in figure 2.10.

---

[1] The reader is referred to section 2.1.2.1 for practical details on the computation of these power spectra.
[2] Technical considerations concerning the computation of bispectra are presented in section 2.1.3.

6.4   Discussion and conclusion

The main subject of this chapter is the development of a method designed to improve the correspondence between approximate models for gravitational dynamics and full numerical simulation of large-scale structure formation. Our methodology relies on a remapping of the one-point distribution of the density contrast of the approximately evolved particle distribution using information extracted from N-body simulations.

Due to the differences in the precise structure of the density fields in Lagrangian perturbation theory and in full gravity, the naive implementation of this procedure, inspired by Weinberg (1992), gives a large-scale bias in the power spectrum. This is not solved by a simple rescaling of Fourier modes, which leads to Gibbs ringing artifacts and an overall unphysical representation of the density fields. Smoothing LPT and N-body density fields with the same kernel is also unsuccessful, as smoothed N-body fields will always keep a sharper structure than smoothed LPT fields.

We figured out that the cause of this large-scale bias is not the different density predicted locally by LPT and N-body dynamics on a point-by-point basis, but a problem of mismatch between the volume of extended objects. Our findings question the reliability of LPT for LSS data analysis and generation of mock catalogs at low redshift and high mass resolution. They are also a likely explanation for the discrepancy between the power spectrum of initial conditions reconstructed via Gaussianization and linear theory expectations, encountered by Weinberg (1992).

Considering these results, we improved Weinberg's remapping procedure for present-day density fields by the use of a transfer function. In this fashion, we obtain a physically more accurate representation of the three-dimensional matter distribution in the mildly non-linear regime, while improving higher-order statistics. Since LPT captures well the cosmological dependence and remapping operates on small-scale aspects of the density field, we found that our procedure that is nearly independent of cosmological parameters.

The aim of this method is to develop a fast, flexible and efficient way to generate realizations of LSS density fields, accurately representing the mildly non-linear regime. Our procedure, therefore, responds to the increasing demand for numerically inexpensive models of three-dimensional LSS, for applications to modern cosmological data analysis. At the level of statistical error in our numerical experiments, the approach provides a good method for producing mock halo catalogs and low-redshift initial conditions for simulations, if desired. The resulting information can also be used in a variety of cosmological analyses of present and upcoming observations.

We showed that our approach allows fast generation of cosmological density fields that correlate with N-body simulations at better than 96% down to scales of $k \approx 0.4 \ (\mathrm{Mpc}/h)^{-1}$ at redshift zero and are substantially better than standard LPT results at higher redshifts on the same comoving scales. Remapping improves the fast LPT bispectrum predictions on small scales while the large scale bispectrum remains accurate to within about 1σ of the measurement in our N-body simulations. Since real observations will have larger statistical errors for the foreseeable future, our method provides an adequate fast model of the non-linear density field on scales down to $\sim 8 \ \mathrm{Mpc}/h$. These results constitute a substantial improvement with respect to existing techniques, since non-linearities begin to affect even large-scale measurements in galaxy surveys. Since the number of modes usable for cosmological exploitation scale as k^3, even minor improvements in the smallest scale k allow access to much more knowledge from existing and upcoming observations. This work is a step further in the non-linear regime, which contains a wealth of yet unexploited cosmological information. For possible applications, we provided a cosmographic and statistical characterization of approximation errors.

Our remapping procedure predicts the two-point correlation function at around 95% level accuracy and three-point correlation function at around 80% level accuracy at redshift 3, for k between 0.1 and 0.4 $(\mathrm{Mpc}/h)^{-1}$, illustrating the increasing success of our methods as we go backwards in time towards the initial conditions, when LPT is an accurate description of early structure formation. This is of particular interest in several areas of high-redshift cosmology, such as forecasting 21 cm surveys (Lidz *et al.*, 2007), analyzing the properties of the intergalactic medium via the Lyman-α forest (Kitaura, Gallerani & Ferrara, 2012) or probing the reionization epoch (Mesinger & Furlanetto, 2007). This work might also add to methods of data analysis for the ongoing and upcoming high-redshift galaxy surveys mentioned in the introduction.

However, the realization of density fields with these procedures stays approximate, since the full non-linear gravitational physics involves information contained in the shape of structures, which cannot be captured from a one-point modification of LPT, especially after shell-crossing. We studied the performance of one-point remapping of LPT and presented a statistical characterization of the errors, but additional refinements, such as a non-linear, density-dependent smoothing of the N-body field, could further improve on these approximations, for an increased computational cost. This is, however, beyond the scope and intent of this work. Generally,

the complications at large scales that we encounter when applying a local remapping seem difficult to solve in a Eulerian density field approach and would favor a Lagrangian, particle-based perspective.

As mentioned in section 6.1, fast and accurate methods to model the non-linearly evolved mass distribution in the Universe have the potential of profound influence on modern cosmological data analysis. Full Bayesian large-scale structure inference methods such as the BORG algorithm, which extract information on the matter distribution in the Universe from galaxy redshift surveys, rely on LPT (see chapter 4). The technique proposed in this chapter can be envisioned as a numerically efficient and flexible extension of these methods, permitting us to push dynamic analyses of the large scale structure further into the non-linear regime.

Non-linear filtering of large-scale structure samples

"While o'er him fast, through sail and shroud,
The wreathing fires made way.
They wrapt the ship in splendour wild,
They caught the flag on high,
And streamed above the gallant child,
Like banners in the sky."
— Felicia Hemans (1826), *Casabianca*

Abstract

Due to the approximate 2LPT model implemented in the BORG algorithm, inferred large-scale structure samples are only correct in the linear and mildly non-linear regime of structure formation. This chapter describes subsequent improvement of such samples at non-linear scales, via an operation that we refer to as "non-linear filtering". This process does not replace fully non-linear large-scale structure inference, but rather fills small scales with physically reasonable information. Several approaches to non-linear filtering are considered and discussed.

This chapter discusses the generation of non-linear, constrained realizations of the late-time large-scale structure via an operation that we call "filtering" of BORG samples. It is structured as follows. We give motivation for non-linear filtering and describe two different approaches (direct improvement of final conditions, and constrained simulations) in section 7.1. For later use in chapter 8, we describe a set of samples optimally filtered with GADGET in section 7.2. In section 7.3, we describe the efficient COLA scheme for fast production of non-linear large-scale structure realizations, and apply it to generate a large ensemble of samples, used in chapter 9.

7.1 Introduction

7.1.1 Motivation for non-linear filtering of large-scale structure samples

As noted in section 4.2.1.2, the likelihood for Bayesian large-scale structure inference involves a structure formation model to translate from the initial to the final density field:

$$\delta^{\mathrm{i}} \mapsto \delta^{\mathrm{f}} = \mathcal{G}(\delta^{\mathrm{i}}, a). \tag{7.1}$$

Ideally, this step should involve a numerical model that fully accounts for the non-linearities of the Vlasov-Poisson system, which describes structure formation (see chapter 1). Unfortunately, this is not currently computationally tractable. For this reason, BORG uses 2LPT as a proxy for gravitational dynamics.[1]

Nevertheless, the description of particular patterns of the cosmic web (as presented in part IV of this thesis) requires description of the LSS not only correct at the scales correctly described by 2LPT ($k \lesssim 0.1$ Mpc/h, see chapter 2), but also physically reasonable at smaller scales, up to $k \sim 1$ Mpc/h. At this point, it is also useful to recall that the number of Fourier modes usable for cosmology scales as the cube of the smallest accessible mode, k^3.

For these reasons, data-constrained, non-linear realizations of the LSS have a large variety of applications. As noted before, constraining small, non-linear scales within the inference framework is not yet possible; therefore, such realizations will rely on fusing data-constrained large scales and unconstrained small scales that only reflect our theoretical understanding of structure formation. Throughout this thesis, we refer to the production of data-constrained, non-linear realizations, on the basis of BORG large-scale structure samples, as *non-linear filtering*.

7.1.2   Filtering in the final conditions

One possible way to perform non-linear filtering is to directly improve the final conditions produced as BORG outputs. The technique of remapping Lagrangian perturbation theory can be useful in this context: as demonstrated in chapter 6, it cheaply yields improvements of density fields in the mildly non-linear regime. A particular advantage of remapping is its very low computational cost, which allows to process a large number of samples.[2] As seen in chapters 4 and 5, this is crucial for adequate uncertainty quantification.

7.1.3   Filtering via constrained simulations

Another idea is to capitalize on the inference of the initial conditions by BORG. Starting from inferred density fields, which contain the data constraints (see in particular section 5.3.3 for a discussion of information transport), it is possible to go forward in time using an alternative structure formation model, noted $\mathcal{G}_{\mathrm{NL}}$, that improves upon $\mathcal{G}$ for the description of small scales structures:

$$\delta^{\mathrm{i}} \mapsto \delta^{\mathrm{f}}_{\mathrm{NL}} = \mathcal{G}_{\mathrm{NL}}(\delta^{\mathrm{i}}, a). \tag{7.2}$$

This process is known in the literature as running *constrained simulations*. Final density fields $\delta^{\mathrm{f}}_{\mathrm{NL}}$ constructed in this way agree with corresponding BORG final conditions δ^{f} at large scales, but are also physically reasonable at smaller scales, up to the validity limit of $\mathcal{G}_{\mathrm{NL}}$.

In this picture, interesting questions are the determination of the smallest scale influenced by the data and the characterization of the reliability of structures extrapolated in unobserved regions, at high redshift or near survey boundaries. An upcoming publication will investigate the validity of constrained simulations, in particular the strength of data constraints in domains or at scales that have not been considered in the inference scheme.

In the following, we examine two particular cases for $\mathcal{G}_{\mathrm{NL}}$, corresponding to the GADGET-2 cosmological code (section 7.2) and to the fast COLA scheme (section 7.3).

7.2   Fully non-linear filtering with Gadget

This section draws from section II.B. in Leclercq *et al.* (2015).

Optimal non-linear filtering of BORG results is achieved when $\mathcal{G}_{\mathrm{NL}}$ fully accounts for non-linear gravitational dynamics. This is the case when a cosmological simulation code is used. For the purpose of this thesis, we consider that non-linear filtering of BORG results with the GADGET-2 cosmological code (Springel, Yoshida & White, 2001; Springel, 2005) is optimal.

For a variety of later uses, in particular for inference of dark matter voids in the Sloan volume (chapter 8), we generate a set of such optimally filtered, data-constrained realizations of the present large-scale structure.

---

[1] For the record, a BORG run, using 2LPT, takes of the order of a year (wall-clock time).

[2] The computational cost for remapping all the outputs of a BORG run, about $10,000$ samples, would be comparable to a few full-gravity dark matter simulations using GADGET-2.
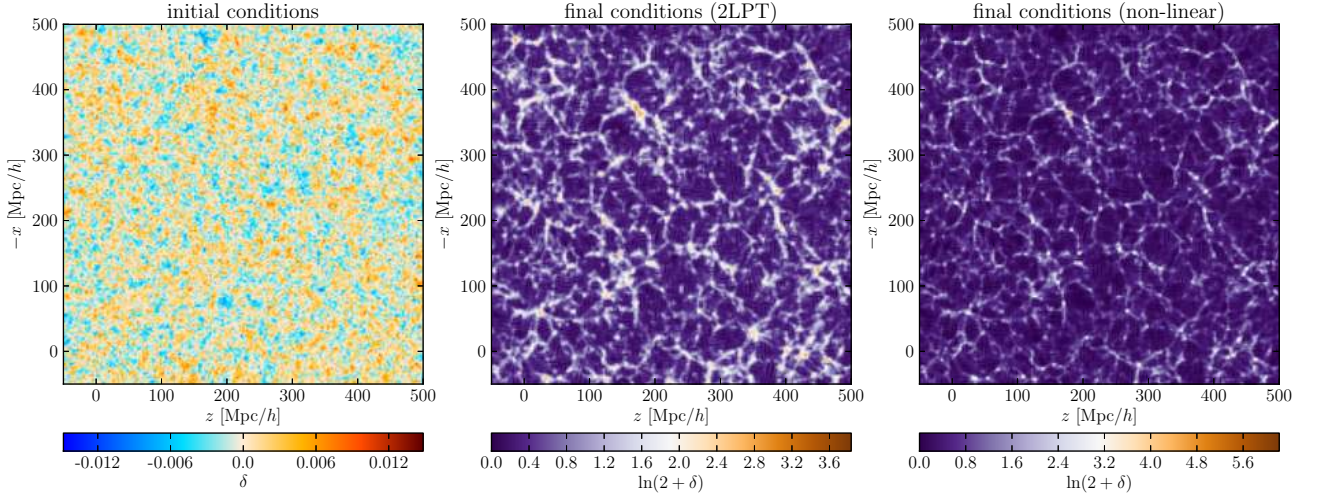
Figure 7.1: Non-linear filtering of BORG results. Slices through one sample of initial (left panel) and final density fields (middle panel) inferred by BORG. The final density field (middle panel) is a prediction of the 2LPT model used by BORG. On the right panel, a slice through the data-constrained realization obtained with the same sample via non-linear filtering (fully non-linear gravitational structure formation starting from the same initial conditions) is shown.

To do so, we rely on a subset of statistically independent initial conditions realizations, provided by Jasche, Leclercq & Wandelt (2015) (see chapter 5). The initial density field, defined on a cubic equidistant grid with side length of 750 Mpc/h and 256^3 voxels, is populated by 512^3 dark matter particles placed on a regular Lagrangian grid. The particles are evolved with 2LPT to the redshift of $z = 69$, followed by a propagation with GADGET-2 from $z = 69$ to $z = 0$. In this fashion, we generate fully non-linear, data-constrained reconstructions of the present-day large-scale dark matter distribution.

As discussed in section 7.1, final conditions inferred by BORG are accurate only at linear and mildly non-linear scales. Application of fully non-linear dynamics to the corresponding initial conditions acts as an additional filtering step, extrapolating predictions to unconstrained non-linear regimes. In a Bayesian approach, this new information can then be tested with complementary observations in the actual sky for updating our knowledge on the Universe.

An illustration of the non-linear filtering procedure is presented in figure 7.1.[3] By comparing initial and final density fields, one can see correspondences between structures in the present Universe and their origins. Comparing final density fields before and after filtering (middle and left panels), one can check the conformity of the linear and mildly non-linear structures at large and intermediate scales, correctly predicted by 2LPT. Small-scale structures, corresponding to the deeply non-linear regime, are much better represented after non-linear filtering (resulting particularly in sharper filaments and clusters). N-body dynamics also resolves much more finely the substructure of voids – known to suffer from spurious artifacts in 2LPT, namely the presence of peaky, overdense spots where there should be deep voids (Sahni & Shandarin, 1996; Neyrinck, 2013; Leclercq et al., 2013; see also chapter 2) – which is of particular relevance for the purpose of inferring dark matter voids (see chapter 8).

The improvement introduced by non-linear filtering at the level of two-point statistics is presented in figure 7.2, where we plot the power spectra of dark matter density fields at $z = 0$. The agreement between unconstrained and constrained realizations at all scales can be checked. The plot also shows that our set of constrained reconstructions contains the additional power expected in the non-linear regime[4], up to $k \approx 0.4$ (Mpc/h)$^{-1}$.

---

[3] In figure 7.1 and in all slice plots of the rest of this thesis, we keep the coordinate system of Jasche, Leclercq & Wandelt (2015), also used in chapter 5.

[4] Note that the lack of small scale power in GADGET and COLA with respect to theoretical predictions, for $k \gtrsim 0.5$ (Mpc/h)$^{-1}$, is a gridding artifact due to the finite mesh size used for the analysis. This value corresponds to around one quarter of the Nyquist wavenumber.
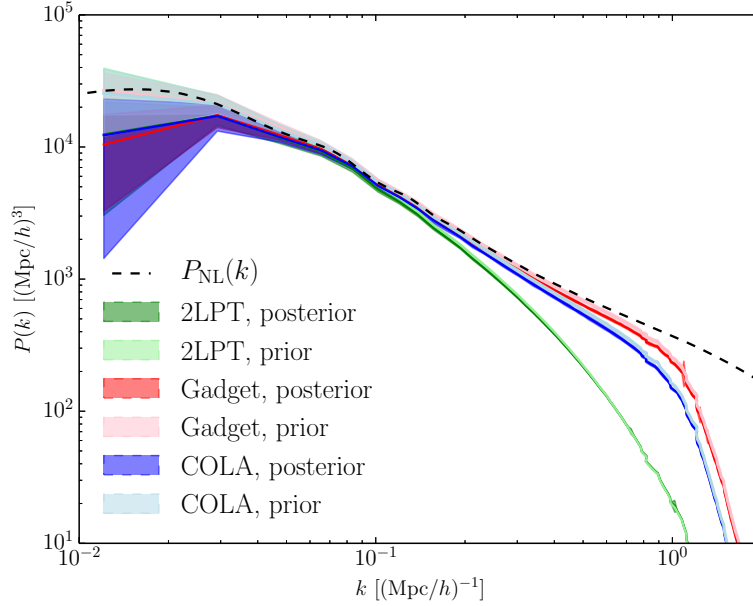
Figure 7.2: Power spectra of dark matter density fields at redshift zero, computed with a mesh size of 3 Mpc/h. The particle distributions are determined using: $1,000$ unconstrained 2LPT realizations ("2LPT, prior"), $4,473$ constrained 2LPT samples inferred by BORG ("2LPT, posterior"), 11 unconstrained GADGET-2 realizations ("Gadget, prior"), 11 constrained samples inferred by BORG and filtered with GADGET-2 ("Gadget, posterior"), $1,000$ unconstrained COLA realizations ("COLA, prior"), $1,097$ constrained samples inferred by BORG and filtered with COLA ("COLA, posterior"). The solid lines correspond to the mean among all realizations used in this work, and the shaded regions correspond to the 2-σ credible interval estimated from the standard error of the mean. The dashed black curve represents $P_{\mathrm{NL}}(k)$, the theoretical power spectrum expected at $z = 0$ from high-resolution N-body simulations.

7.3   Fast non-linear filtering with COLA

For means of uncertainty quantification within large-scale structure inference, it is necessary to process a large number of samples. Unfortunately, optimal non-linear filtering with GADGET-2 is too expensive for the $\sim 10,000$ samples of a single BORG run. However, an approximate model for non-linear structure formation, correct up to scales of a few Mpc/h, is enough for our purposes, as long as the approximation error is controlled and quantified.

7.3.1   The COLA method

The COLA (COmoving Lagrangian Acceleration, Tassev, Zaldarriaga & Eisenstein, 2013; Tassev *et al.*, 2015) technique offers a cheap way to perform non-linear filtering of a large number of BORG samples. A particular advantage (in opposition to standard particle-mesh codes) is its flexibility in trading accuracy at small scales for computational speed, without sacrificing accuracy at the largest scales.

The general idea of COLA is to use our analytic understanding of structure formation at large scales via LPT, and to solve numerically only for a subdominant contribution describing small scales. Specifically, Tassev & Zaldarriaga (2012c) propose to expand the Lagrangian displacement of particles as

$$\mathbf{\Psi}(\mathbf{x}, \tau) = \mathbf{\Psi}_{\mathrm{LPT}}(\mathbf{x}, \tau) + \mathbf{\Psi}_{\mathrm{MC}}(\mathbf{x}, \tau) \tag{7.3}$$

where $\mathbf{\Psi}_{\mathrm{LPT}}(\mathbf{x}, \tau)$ is the analytic displacement prescribed by LPT[5] (the ZA or 2LPT, see chapter 2) and $\mathbf{\Psi}_{\mathrm{MC}}(\mathbf{x}, \tau) \equiv \mathbf{\Psi}(\mathbf{x}, \tau) - \mathbf{\Psi}_{\mathrm{LPT}}(\mathbf{x}, \tau)$ is the "mode-coupling residual". Using this Ansatz, the Eulerian position is $\mathbf{x} = \mathbf{q} + \mathbf{\Psi}_{\mathrm{LPT}} + \mathbf{\Psi}_{\mathrm{MC}}$, and the equation of motion, which reads schematically (omitting constants and Hubble expansion; see equation (1.74))

$$\frac{\mathrm{d}^2 \mathbf{x}}{\mathrm{d}\tau^2} = -\nabla_{\mathbf{x}} \Phi, \tag{7.4}$$

---

[5] Following Tassev & Zaldarriaga (2012c), this first term can be written more generally in Fourier space as $\mathbf{\Psi}_{\star}(\mathbf{k}, \tau) = R_{\mathrm{LPT}}(k, \tau) \, \mathbf{\Psi}_{\mathrm{LPT}}(\mathbf{k}, \tau)$, where $R_{\mathrm{LPT}}(k, \tau)$ is a transfer function that we ignore here for simplicity.
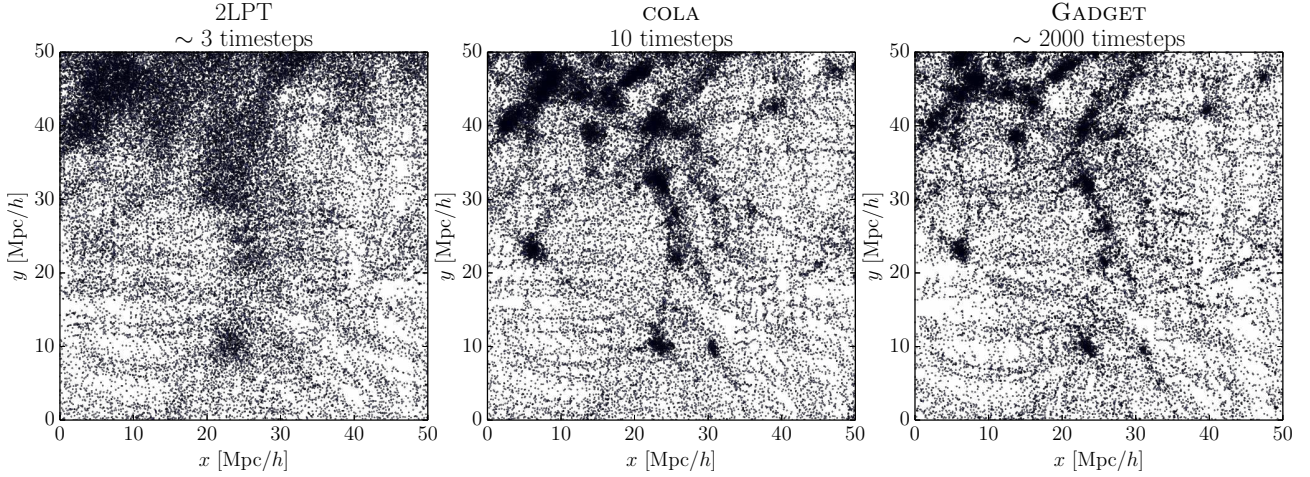
Figure 7.3: Slices through three particle realizations evolved from the same initial conditions up to $z = 0$. The particles are shown as black points. Each slice is 20 Mpc/h thick and 50 Mpc/h on the side. The left panel shows the 2LPT approximation, of computational cost roughly equivalent to 3 timesteps of a N-body code. The right panel shows the reference result obtained from GADGET-2 after ~ 2000 timesteps, starting from 2LPT initial conditions at $z = 69$. The middle panel shows the result obtained with COLA with 10 timesteps, starting from 2LPT initial conditions at $z = 9$.

can be rewritten in a frame comoving with "LPT observers", whose trajectories are given by $\mathbf{\Psi}_{\mathrm{LPT}}$, as

$$\frac{\mathrm{d}^2 \mathbf{\Psi}_{\mathrm{MC}}}{\mathrm{d}\tau^2} = -\nabla_{\mathbf{x}}\Phi - \frac{\mathrm{d}^2 \mathbf{\Psi}_{\mathrm{LPT}}}{\mathrm{d}\tau^2}. \tag{7.5}$$

In analogy with classical mechanics, $\mathrm{d}^2\mathbf{\Psi}_{\mathrm{LPT}}/\mathrm{d}\tau^2$ can be thought of as a fictitious force acting on particles, coming from the fact that we are working in a non-inertial frame of reference.

The standard approach in PM codes (see appendix B) is to discretize the second-derivative time operator in equation (7.4). At large scales, this is nothing more than solving for the linear growth factor. Therefore, if few timesteps are used in PM codes, the large-scale structure will be miscalculated only because of a faulty estimation of the growth factor, the exact value of which being well-known.

In contrast, the COLA method uses a numerical discretization of the operator $\mathrm{d}^2/\mathrm{d}\tau^2$ only on the left-hand side of equation (7.5) and exploits the exact analytic expression for the fictitious force, $\mathrm{d}^2\mathbf{\Psi}_{\mathrm{LPT}}/\mathrm{d}\tau^2$. The equation solved by COLA, equation (7.5), is obviously equivalent to (7.4). However, as demonstrated by Tassev, Zaldarriaga & Eisenstein (2013), using this framework requires a smaller number of timesteps to recover accurate particle realizations. In particular, they show that as few as 10 timesteps from $z = 9$ to $z = 0$ are sufficient to obtain an accurate description of halo statistics up to halos of mass 10^{11} M$_\odot/h$, without resolving the internal dynamics of halos. Concerning the description of the large-scale matter density field, 10 COLA timesteps achieve better than 95% cross-correlation with the true result up $k \sim 2$ Mpc/h.

As an illustration of the performance of COLA, we show slices through corresponding 2LPT, COLA and GADGET particle realizations in figure 7.3. The simulations contain 512^3 particles in a 750 Mpc/h cubic box with periodic boundary conditions. Forces are calculated on a PM grid of 512^3 cells. The initial conditions are generated with 2LPT at a redshift of $z = 69$ for GADGET and $z = 9$ for COLA.

7.3.2 Non-linear BORG-COLA realizations

This section draws from section II.B. in Leclercq, Jasche & Wandelt (2015c).

In chapter 9, we use an ensemble of $1,097$ large-scale structure realizations produced via non-linear filtering of BORG samples with COLA. The initial density field, defined on a cubic equidistant grid with side length of 750 Mpc/h and 256^3 voxels, is populated by 512^3 particles placed on a regular Lagrangian lattice. The particles are evolved with 2LPT to the redshift of $z = 69$ and with COLA from $z = 69$ to $z = 0$. The final density field is constructed by binning the particles with a CiC method on a 256^3-voxel grid. This choice corresponds to a resolution of around 3 Mpc/h for all the maps described in chapter 9. In this fashion, we generate a
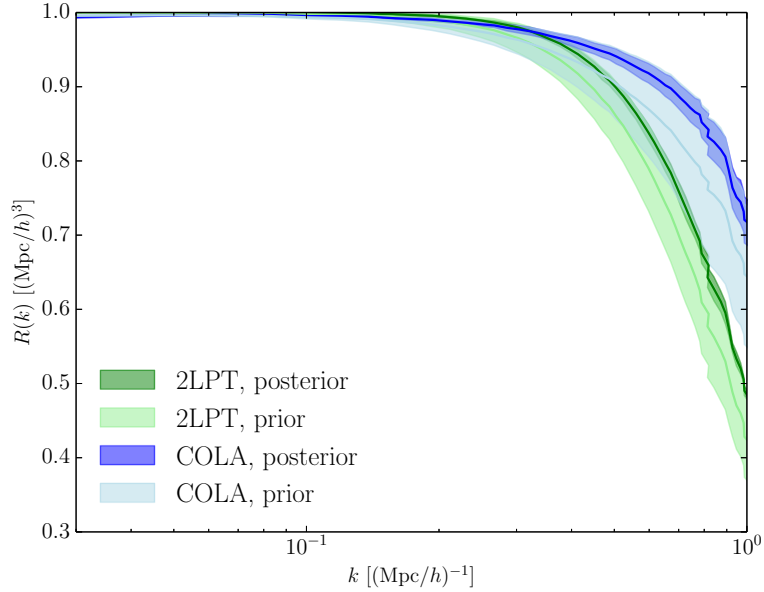
Figure 7.4: Cross-correlations between density fields at redshift zero, computed with a mesh size of 3 Mpc/h. The reference fields are the result of GADGET-2. The lines correspond to the cross-correlation between unconstrained 2LPT realizations and corresponding simulations ("2LPT, prior"), constrained 2LPT samples inferred by BORG and corresponding optimal filtering ("2LPT, posterior"), unconstrained COLA realizations and corresponding simulations ("COLA, prior"), constrained BORG-COLA samples and corresponding optimal filtering ("COLA, posterior"). In each case, we use 11 constrained or unconstrained realizations. The solid lines correspond to the mean among all realizations used in this work, and the shaded regions correspond to the 2-σ credible interval estimated from the standard error of the mean.

large set of data-constrained reconstructions of the present-day dark matter distribution (see also Lavaux, 2010; Kitaura, 2013; Heß, Kitaura & Gottlöber, 2013; Nuza *et al.*, 2014). To ensure sufficient accuracy, 30 timesteps logarithmically-spaced in the scale factor are used for the evolution with COLA.

COLA enables us to cheaply generate non-linear density fields at the required accuracy, as we now show. The power spectrum of non-linear BORG-COLA realizations is shown in figure 7.2 in comparison to that of unconstrained realizations and to samples optimally filtered with GADGET-2. In figure 7.4, we plot the cross-correlation between approximate density fields (predicted by 2LPT or by COLA) and the result of GADGET-2, for both unconstrained and constrained realizations. On these plots, it can be checked that our constrained samples, inferred by BORG and filtered with COLA, contain the additional power expected in the non-linear regime and cross-correlate at better that 95% accuracy with the corresponding fully non-linear realizations, up to $k \approx 0.4$ Mpc/h. Therefore, as for unconstrained simulations, our setup yields vanishing difference between the representation of constrained density fields with COLA and with GADGET-2, at the scales of interest of this work.