Title Here

Andrei Cuceu

Supervisor: Dr. Andrew Pontzen

University College London

March 25, 2018

Abstract

Eisenstein et al. 2007

Contents

Chapter 1

Introduction

1.1 Cosmological Context

Cosmology has seen many important advances in the last few decades. Piece by piece we are uncovering the history of the universe and it's components. We now know that our galaxy is just one in trillions (Conselice et al. 2016), and part of a rapidly expanding universe.

Our two main observational tools today are the Cosmic Microwave Background (CMB) and Galaxy Surveys. From the CMB we learn about the primordial universe, and Galaxy Surveys uncover the nature and structure of the Universe at late times. The latest surveys like Planck (Planck Collaboration et al. 2016), SDSS or DES lend strong support to the ΛCDM paradigm. Within this Cosmological Standard Model we now understand most of the important events that shaped the history of our Universe and dictate its future.

ref
ref

The two most important components of this model are arguably Dark Matter and Dark Energy. The two account for a total of about 95% of the matter-energy budget of the universe (e.g. Planck Collaboration et al. 2016). We have been able to constrain the properties of the two mysterious components quite well. Most observational probes are consistent with a Cosmological Constant model of dark energy , but they don't yet exclude other models like quintessence . On the other hand, the Cold Dark Matter model of a non-interacting (or very weakly interacting) particle that only has gravitational impact is leading on the dark matter side . However, despite this wealth of knowledge, the nature of the two most important components of the Universe still eludes us.

ref
ref
ref

1.2 Observational Probes

1.2.1 The Cosmic Microwave Background

The Cosmic Microwave Background is our main observational tool for studying the early universe. It is the first light emitted in the Universe after recombination, and it encodes plenty of useful information. The CMB is the most perfect black body ever observed (White 1999). Its existence is already a strong proof in support of the Big Bang model. On the other hand, the anisotropies found in the CMB strongly support both the ΛCDM paradigm and Inflation.

CMB surveys have measured these anisotropies very accurately (Figure 1.1). They are key in understanding the matter distribution in the universe. Before recombination Baryons and Radiation were coupled, so the matter distribution at recombination was imprinted in the CMB distribution. The matter distribution then continued to evolve on its own into the large scale structure we see today. Among the key anisotropies detected in the CMB are Baryon Acoustic Oscillations. This feature is a result of oscillations in the primordial Baryon-Photon plasma. Radiation opposed the collapse of baryonic matter into the potential wells created by collapsing dark matter (which does not interact and is free to collapse). This produced acoustic waves that were imprinted in both the matter distribution (we detect it today in the galaxy distribution) and the radiation distribution (we detect it in the CMB). This probe lends strong support both to the existence of dark matter which plays a key role in their creation, and to that of dark energy (as a cosmological constant) which dictates their evolution at late times.

The CMB also has plenty of data in support of Inflation. It shows a uniform, very flat universe that contains mostly gaussian anisotropies. All of these are outcomes of inflation and are hard to explain without it. Of particular interest for us are the gaussian anisotropies. There are many models of inflation , and at the moment we do not have enough precision to distinguish between them. Most models predict some primordial non-gaussianity, however constraining this is key to differentiating between them . refs!! ref

After the CMB was emitted we enter a period called the Dark Ages. Until the first stars and galaxies formed, the CMB was the only radiation in the Universe. This means we have little to no information about this important era. In it lie the secrets to the formation of the structure we see today in the universe. This shows the main problem of the CMB, it is primarily a 2 dimensional probe showing us the surface of last scattering. While CMB photons do interact with matter on their way to us producing secondary anisotropies, it is far from a complete picture of

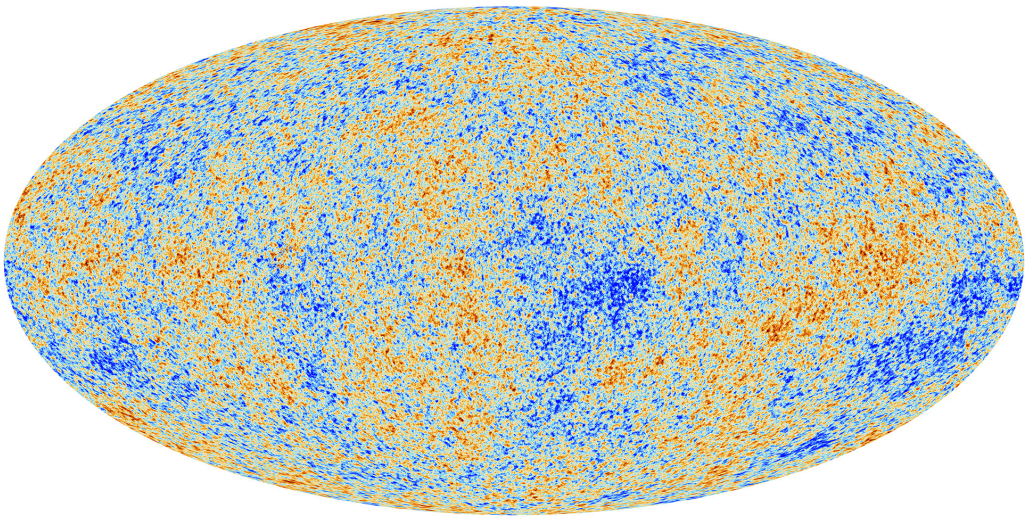

Figure 1.1: Map of the Cosmic Microwave Background acquired by the Planck Space Telescope (Planck Collaboration et al. 2016). After subtracting the effect of our own motion relative to the CMB, the anisotropies only start appearing at $10^{-5}K$. This gives us a glimpse into the primordial matter distribution.

the evolution of structure.

1.2.2 Large Scale Structure and Galaxy Surveys

Beginning with the first galaxy surveys in the 1980s (Davis et al. 1982), we started to piece together a picture of the large scale structure of the universe. Galaxies are clustered together and form massive filaments that are separated by huge voids. This web like structure is the result of gravitational collapse of initial perturbations. On the other hand, when looking on the largest scales, the Universe is very uniform.

ref

Recent galaxy surveys like 2dF, SDSS and DES have brought us a wealth knowledge about the Universe today. They are great tool for studying the universe because, unlike the CMB, we now have a 3 dimensional distribution. Future surveys such as LSST (LSST Science Collaboration et al. 2009) and DESI (Levi et al. 2013) are set to map larger and larger areas of the Universe with ever increasing sensitivity. One of their main objectives is to measure the imprint left by Baryon Acoustic Oscillations in the matter distribution.

This feature is even more important at these late times, because it was locked on a certain scale after recombination. This means we can use it as a characteristic length scale that evolves with Universe. However, there are some problems that also appear. Firstly, galaxies are biased tracers of the underlying matter distribution (e.g. Desjacques, Jeong, and Schmidt 2016). Secondly, non-linear gravitational collapse smears out this feature, reducing the precision of the length scale.

As this is a large scale feature, it is however possible to reconstruct it. Eisenstein et al. 2007 argued that most of the contributions to broadening are from super cluster formation and bulk flows. Therefore, Perturbation Theory can be used to perform this reconstruction. Many methods have been proposed to recover this feature. For example, Weinberg 1992 proposed a simple "Gaussianization", while Nusser and Dekel 1992 and Gramann 1993 introduced a technique based on the success of the Zel'dovich Approximation. More complex methods have also been proposed by Monaco and Efstathiou 1999, Croft and Gaztanaga 1997 and Brenier et al. 2003.

These methods have been instrumental in using survey data to its fullest. Weinberg et al. 2013 estimate that this increase in precision is equivalent to a factor $2-4$ increase in survey size.

1.3 The Growth of Structure

In this section we will take a closer look at some of the tools we use to study the growth of structure in the universe. The evolution of primordial perturbations into the structure we detect today is studied analytically using Perturbation Theory (PT). This field can be divided, according to the frame used to study the Universe, into Eulerian PT and Lagrangian PT (LPT). In the Eulerian frame the evolution of the spatial distribution of particles is studied (Bernardeau et al. 2002). On the other hand, in the Lagrangian frame, particles are tracked and their evolution with time is studied.

The Eulerian frame has been so widely used to study the growth of perturbations, that we refer to it as Standard Perturbation Theory (e.g. Vishniac 1983, Peebles 1980). The most popular approach is to consider an irrotational fluid characterized by its overdensity and peculiar velocity distributions (Carlson, White, and N. Padmanabhan 2009). However, this method suffers from divergences at large wavenumbers (on small scales). This lead to a number of extensions meant to bring it under control (Crocce and Scoccimarro 2006, Crocce and Scoccimarro 2008).

Lagrangian Perturbation Theory has been well developed in the 1990s (Buchert 1992, Buchert and Ehlers 1993, Buchert 1994). However, it has received less attention partly because the method breaks down after shell-crossing (Carlson, White, and N. Padmanabhan 2009). This event will be discussed further in Section 2.2. Recently it has been demonstrated that LPT correctly reproduces the SPT power spectrum, but also even it's linear first order approximation correctly predicts the decay of the correlation between the final (non-linear) and the initial fields (Matsubara 2008b, Matsubara 2008a). As this correlation is the main tool used in this

project, our main focus will be on the Lagrangian Frame.

Once we get into the deeply non-linear regime, these analytic methods break down. Instead we rely on large N-body simulations to continue our study of the evolution of structure. Generally, these simulations are used to help constrain Cosmological theories of the primordial universe or to test perturbation theories. A random initial field with properties based on the CMB and our general understanding of the early universe, is evolved through time until the present. The output of the simulation is then compared to observational data. Simulations like Millennium XXL (Angulo et al. 2012) or Illustris (Vogelsberger et al. 2014) gave us an unprecedented insight into structure formation and evolution.

On the other hand, simulations are also used as a laboratory for testing reconstruction methods they give us access to both the initial and the final density fields. We can attempt a reconstruction on the final field and then compare it to the initial field and test how well it worked.

1.4 Recovering Lost Information

Chapter 2

The Perfect Reconstruction

Our first step is to try to understand how much information from the primordial universe is preserved. We will look at the theoretical limit encountered when trying to reconstruct the final non-linear density field.

As a field is defined at every point in space, any attempt at representing it with data is inherently imperfect. We would have to measure the density field at every point in the Universe in order to obtain all the information it contains. This fact already implies that no data driven reconstruction will ever succeed at perfectly recovering the primordial density field (unless we manage to make an infinity of measurements).

To show this unavoidable loss of information we performed a 'perfect' reconstruction. We call this reconstruction 'perfect', as it uses data about the initial positions of all particles (which obviously is not available to observers). However, as we will show in this chapter, not even this perfect reconstruction succeeds at perfectly recovering the primordial matter distribution.

2.1 Methods

The first step in studying reconstruction is to find a way to test its effectiveness. To do this, we use cosmological N-body simulations. These simulations give us important insights into how structure evolves in the Universe. More importantly for this project, it allows us to compare any reconstructed density field to the real starting density field. We use data from three simulations available to us. The first one is Simulation A presented in Pontzen et al. 2016a (also referred to as Simulation A in this work). The other two simulations are variations of the same initial setup, with smaller size and smaller resolution respectively. The details of all three simulations are presented in Table 1.

The simulations only contain dark matter and the cosmological parameters

Label	Size	Number of Particles	Particle Mass (Solar Masses)
Sim A	$(200Mpc)^3$	512^3	6.59×10^9
Sim B	$(200Mpc)^3$	256^3	5.27×10^{10}
Sim C	$(100Mpc)^3$	256^3	6.59×10^9

Table 2.1: The sizes and number of particles of the three simulations used in this project.

used were the ones recommended by the WMAP 7-year observations (Komatsu et al. 2011). These are no longer our best estimates (Planck Collaboration et al. 2016), however for our present purposes they are sufficient as we do not compare to observational data. The simulations were evolved from $z = 99$ to $z = 0$ using the P-GADGET-3 code (Springel 2005, Springel et al. 2008).

The idea behind a perfect reconstruction is to use data about the initial state of the simulation to perform the reconstruction. We have access to multiple snapshots at various redshifts in our simulations, including the initial positions of all particles (at $z = 99$). Therefore, we used this information to reconstruct the density field. We first measured the density field of various snapshots in the redshift interval $z = 0 - 9$. The field was measured at the particle positions instead of being measured on a regular grid. This is because the density field will be moved along with the particles as described in the next section. After that, the particle positions were replaced with their initial positions taken from the initial snapshot at $z = 99$.

2.2 From Images to Statistics

2.2.1 The Reconstructed Density Field

As outlined above, the first step is to measure the density field in a snapshot. Each snapshot contains an indexed list with the positions, velocities and masses of all particles in the simulation. In this chapter, only the positions and masses are needed to perform a perfect reconstruction. To perform the first part of this analysis, we used the *pynbody*[1] package (Pontzen, Roškar, et al. 2013).

To have a visual understanding of the reconstruction, we first make some images of the density field. We use *pynbody* to import the initial snapshot and the snapshot at $z = 0$. The density field at the particle locations in the final snapshot is calculated and assigned as the density field of the initial snapshot. Density slices through this reconstructed field are compared to the initial and final fields in Figure .

put image in and ref

[1] https://github.com/pynbody/pynbody

We can already see from this comparison that the reconstruction has not recovered all the information, as it is not identical to the initial filed. However, the reconstruction spreads out the matter from the collapsed filaments onto a more uniform field. This is exactly the opposite of initial perturbations on a mostly uniform field collapsing into the filaments. In a way we are trying to run gravity in reverse. Also notice the large difference in the values of the density field. The reconstructed field has density values about 3 orders of magnitude larger than the initial field.

also talk about the density distribution

This large difference is an interesting side effect of our method. At late times, most particles tend to be clumped together. Therefore, when measuring the density field at the particle positions, we will mostly get very high values. These values do not change when moving the particles, so the final field will also have very high values, but this time distributed on an almost uniform grid. This results in an apparent increase in the total mass of the simulation. As this increase is just a result of the way we represent the density field, it needs to be accounted for when analysing the results. To conserve the mass of the simulation, we just divide the density field by the ratio of the total final mass to the total initial mass.

2.2.2 Correlation with the Initial Field

TALK ABOUT TAKING THE LOG OF THE DENSITY FIELD

In order to get a better understanding of how well this reconstruction worked, we turn to statistics. A good way to represent the reconstruction is to look at the normalized Cross-Spectrum between the initial and the reconstructed field:

$$\frac{P_{IX}(k)}{\sqrt{P_I(k) * P_X(k)}}$$

where I represents the initial field, and X the reconstructed field.

The power-spectrum will be introduced in chapter 2

We used the GENPK code[2] (Bird 2017) to measure auto and cross power-spectra of GADGET outputs. The original normalized cross-spectra between the initial and the final fields (from Sim A) can be seen in Figure 2.1. For small wavenumbers k (corresponding to large scales), the correlation is very good (converges to 1: perfect correlation). On the other hand, for large wavenumbers (corresponding to small scales), the two fields are completely decorrelated.

Talk about the binning??

The small k convergence towards perfect correlation indicates that information is preserved on these scales. Because of this, both the initial and the final

[2] https://github.com/sbird/GenPK.git

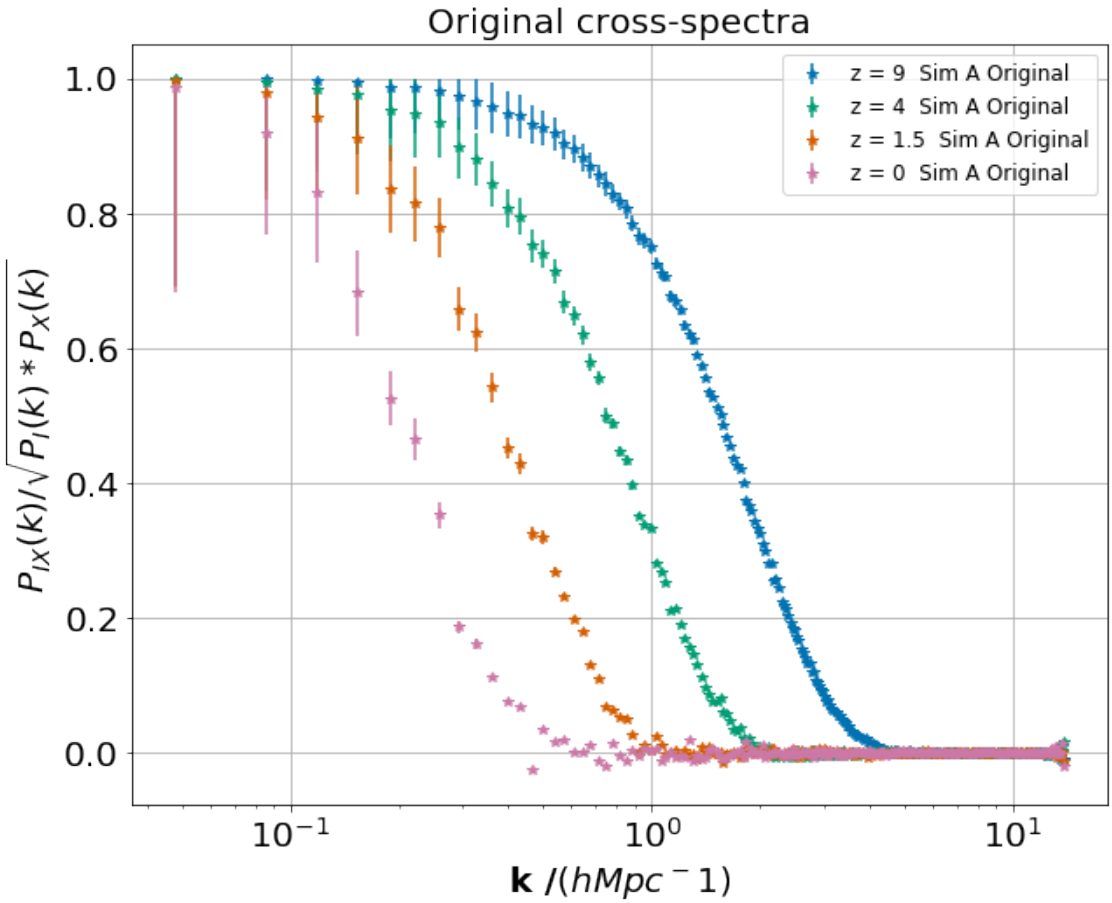

Figure 2.1: Normalized cross-spectra between the initial and final density fields as a function of scale. At small wavenumbers k (large scales) the two fields are perfectly correlated because the universe is not affected by gravitational collapse over such scales (they are both uniform). On the other hand, at large k (small scales) they are decorrelated because the initial field is very uniform, while the final field is very non-uniform on such scales (it contains large empty voids, and small and massive halos). The decorrelation scale moves to smaller k as time goes on due to the progressive collapse of larger and larger overdensities.

density fields tend to be very uniform, which preserves the correlation. However, over small scales information is lost. The results is a large discrepancy between massive collapsed regions and mostly empty voids. This is in stark contrast to the relative uniformity of the initial field, leading to a breakdown in correlation.

Figure 2.1 also shows the evolution of this correlation with redshift. The wavenumber at which the two fields decorrelate indicates the progress of gravitational collapse at that redshift. This results in the decorrelation scale moving to smaller wavenumbers with the progress of gravitational collapse. The objective of reconstruction methods is to bring this decorrelation scale to larger k (in order to recover information about the initial field).

Talk about errors here

In order to measure the power-spectra of our reconstructed fields, we modified GENPK to read the fields generated by *pynbody*. The results of the perfect reconstruction can be seen in Figure 2.2, where we compare it with the original correlation at different redshifts.

2.3 Analysis

The cross-spectra presented in Figure 2.2 show a large improvement in the correlation with the initial field. There is also an increase in the amount of information recovered for lower redshifts. This means redshift does not play a role as large in the perfect reconstruction as it originally did.

However, in order to understand this perfect reconstruction, we need to look at the key role played by the resolution of the simulation. Figure 2.3 shows a comparison of the cross-spectra across the three simulations. For the original correlations, the size of the simulation plays a larger role than the resolution. Simulation C (smaller size) shows a smaller scale of decorrelation. Simulations A and B (same size) are very close, with a slight edge for simulation B (lower resolution).

better explination here

A completely different structure can be seen once we perform the perfect reconstruction. Simulation A and C (same resolution) show identical reconstructed correlation, while the reconstruction in Simulation B (lower resolution) does not perform as well. This indicates that resolution plays a decisive role in the perfect reconstruction. The other limiting factor is the starting redshift. As we have seen, more and more information is lost as we go to lower redshift. The right panel in Figure 2.3 shows the lowest scale that can be reconstructed from $z = 9$ depending on the resolution of our simulation.

The perfect reconstructions in Figure 2.2 demonstrate that a perfect correlation

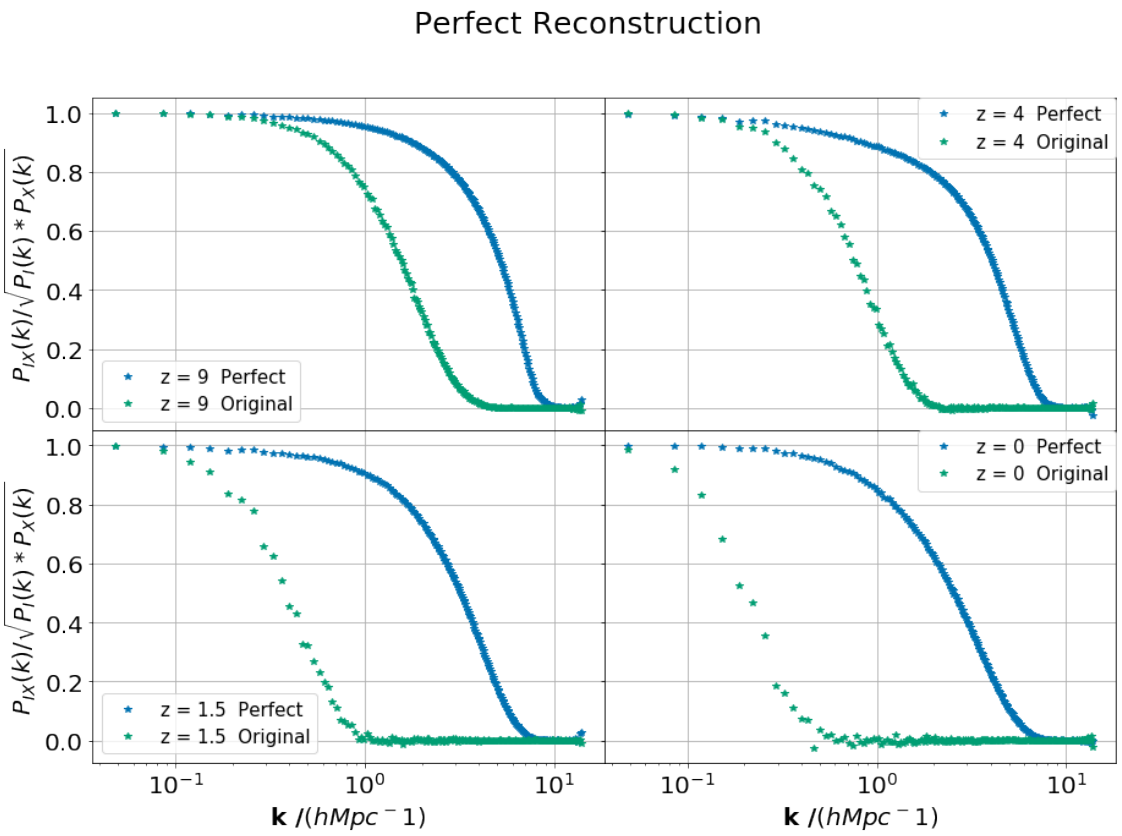

Figure 2.2: Normalized cross-spectra between the initial and the reconstructed fields compared to the original correlation. A large improvement in the correlation was achieved with the perfect reconstruction. This shows up as a shift of the decorrelation scale towards larger k (smaller scales). However, the perfect reconstruction does not lead to a perfect correlation due to the limiting resolution of our density field measurements. When comparing the reconstruction applied to fields at different redshift, we see a trend towards less information at lower redshift. This shows the progress of information loss between $z = 9$ and $z = 0$.

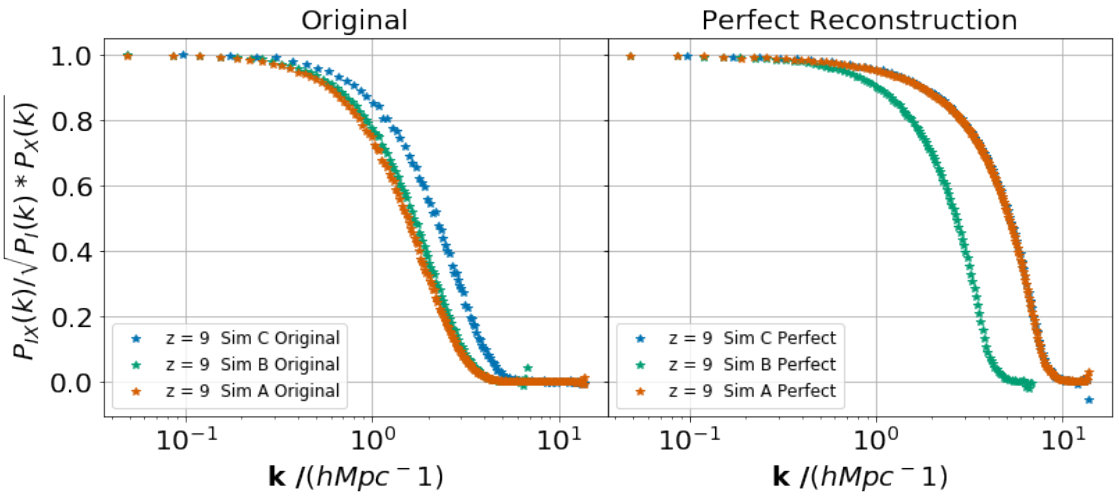

Figure 2.3: Normalized cross-spectra across the three simulations. The left plot shows the original correlation between the initial and the final fields, and the right panel shows the correlation with the reconstructed field. For the original correlations, the size of the simulation plays an important role, with the smaller simulation decorrelating on smaller scales. However, after the reconstruction, the size of the simulation does not seem to have any impact (with Sim A being almost identical to Sim C). In this case, the resolution of the simulation is the only factor that matters, with the larger resolution simulations showing a better correlation.

cannot be achieved even in the ideal case of perfect knowledge of all the starting particle positions. Therefore, we have shown the limits of reconstruction techniques. However, a large improvement in the correlation can be seen, with the decorrelation scale moving to very small scales (of the order $1Mpc$). This ideal reconstruction using perfect knowledge of the particle positions serves as a theoretical upper limit to reconstruction techniques. The perfect reconstruction, along with the original correlation, will always be present in the next chapter when we look at realistic reconstructions. This can give us a better understanding to how well our techniques work.

Chapter 3

Towards Realistic Reconstructions

With the tools developed in the previous chapter, and the perfect reconstruction serving as an upper limit for reconstruction, we are ready to dive into realistic reconstruction methods. In Section 1.2.2 we outlined different reconstruction methods used in practice, based on both Standard Perturbation Theory and Lagrangian Perturbation Theory. In this project, we base our reconstructions on the first order approximation to LPT (the Zel'dovich Approximation).

3.1 The Zel'dovich Approximation

3.1.1 Background

The central object of Lagrangian Perturbation Theory is the displacement field $\Psi(\mathbf{q})$ which maps the initial particle positions expressed by the Lagrangian coordinate $\mathbf{q}$ to the final Eulerian particle positions $\mathbf{x}$ (Bernardeau et al. 2002):

$$\mathbf{x}(t) = \mathbf{q} + \Psi(\mathbf{q}, t) \tag{3.1}$$

The first order approximation to this equation leads to a separation of variables $\mathbf{q}$ and t (T. Padmanabhan 1993). After also adding the expanding background we obtain:

$$\mathbf{r}(t) = a(t)\mathbf{x}(t) = a(t)[\mathbf{q} + b(t)\mathbf{p}(\mathbf{q})] \tag{3.2}$$

Where $\mathbf{x}(t)$ is now the comoving Eulerian coordinate.

This approximation was first proposed by Zel'dovich 1970 and it now carries his name, as the Zel'dovich Approximation (ZA). It only works for scales much smaller than the horizon, where Newtonian analysis is possible. However, as our purpose is to study the growth of structure on scales smaller than the BAO, this

approximation is a very good starting point.

An important detail is that the ZA (up to a point) works even in the non-linear regime. This regime is defined in terms of the overdensity δ. When $\delta \sim 1$ we enter the non-linear regime. Standard Perturbation Theory generally breaks down at this point. However, by switching to a description in terms of the particle trajectories, we can reach non-linear density contrasts without the particles being perturbed too much (T. Padmanabhan 1993).

On a final note, the ZA is very good at predicting the loss of correlation between the initial and the final density fields (e.g. Pontzen et al. 2016b). With all these considerations in mind, in this project we will investigate reconstruction methods within the Zel'dovich Approximation.

3.1.2 Tracing particles back

The key ingredient that we will use to perform realistic reconstructions is the density field. The idea behind the Zel'dovich approximation is to calculate a linear displacement field (we refer this field as Zel'dovich offset) based on the current peculiar velocities while taking into account the Hubble flow.

The peculiar velocity predicted by the ZA (T. Padmanabhan 1993) is given by:

$$\mathbf{V}(t) \equiv a(t)\frac{d\mathbf{x}}{dt} \tag{3.3}$$

Using 3.2, we arrive at a peculiar velocity given by:

$$\mathbf{V} = a(t)\dot{b}\mathbf{p}(\mathbf{q}) \tag{3.4}$$

Finally, bringing the displacement field back ($\Psi(\mathbf{q}, t) = b(t)\mathbf{p}(\mathbf{q})$), and switching variables to redshift we obtain:

$$\Psi_z(\mathbf{q}) = \frac{\mathbf{V}(\mathbf{q})}{a(z)} \times \frac{b(z)}{f(z)} \tag{3.5}$$

Where $b(z)$ is called the linear growth factor and $f(z) = \dot{b}(z)$ is the rate of linear growth. To calculate these two, we use the integration methods implemented in *Pynbody* (Pontzen, Roškar, et al. 2013).

As we are interested in looking at the correlation between the reconstructed field and the initial fields in our simulations (which are at $z = 99$), we need to calculate the Zel'dovich offset up to $z = 99$ only. In order to achieve this, we first used equation 3.5 to calculate the offset starting from the redshift z of the snapshot (Ψ_z). After that, we used the same equation to approximate this offset from $z = 99$ (Ψ_{99}) using the same velocity field. The displacement field we are

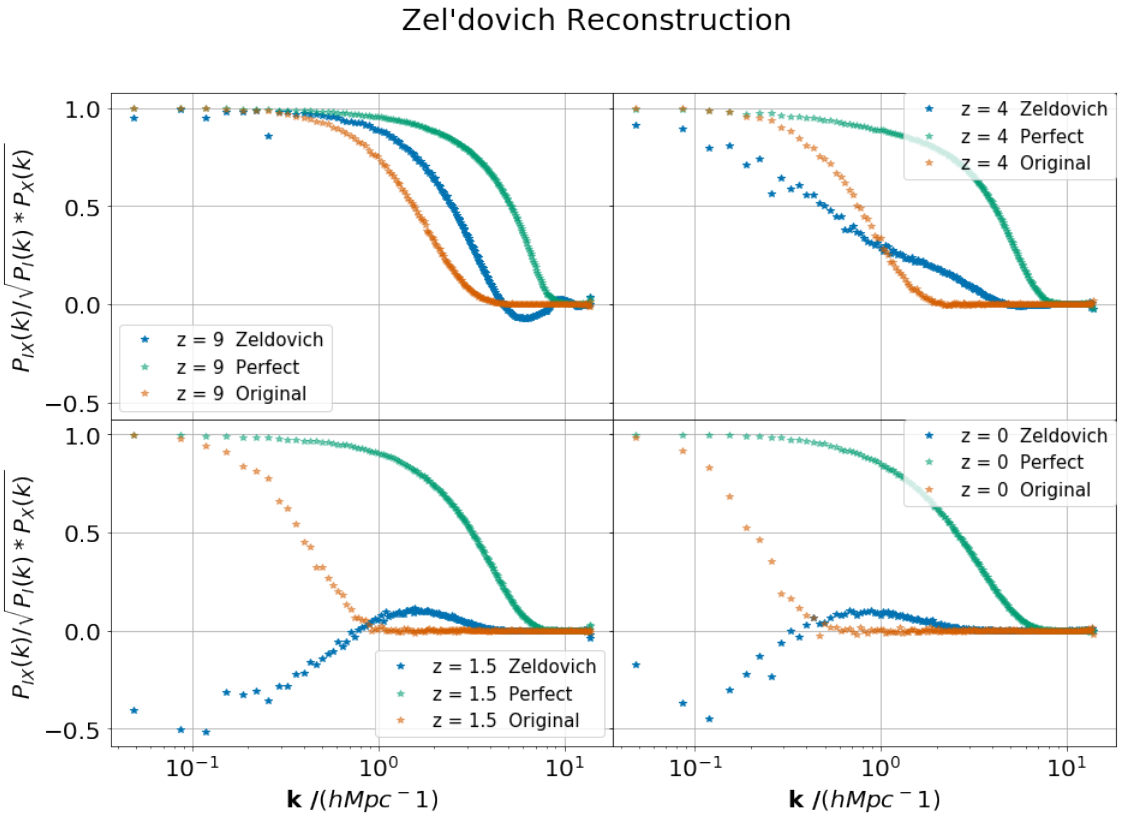

Figure 3.1: Normalized cross-spectra between the Zel'dovich reconstruction and the initial field. This reconstruction was performed by linearly moving the particles back in time (using the Zel'dovich approximation). As we apply the linear approximation directly to the particle velocities, this gives a good indicator of the regime we are in at that redshift. We see the reconstruction work very well when starting from $z = 9$, which indicates we are still in the quasi-linear regime there. However, as time progresses (lower redshift) the correlation breaks down even on the largest scales, indicating we are mostly in the non-linear regime.

after is then given by:

$$\Psi(\mathbf{q}) = \Psi_z(\mathbf{q}) - \Psi_{99}(\mathbf{q}) \tag{3.6}$$

3.1.3 The Reconstruction

We first start out in our investigation by performing a reconstruction using the Zel'dovich offset calculated directly from the particle velocities in each snapshot. The methodology of the reconstruction resembles the Perfect Reconstruction. We first calculate the density field at the particle positions in a snapshot, and then we apply the Displacement field $\Psi(\mathbf{q})$ to move the particles. The density field is carried along. After that, GENPK is used to measure the cross power-spectra of the reconstructed field with the initial ($z = 99$) field.

The results of this reconstruction (we call it the Zel'dovich reconstruction) can

be seen in 3.1. We again look at the normalized cross-spectra between this reconstruction and the initial density field. To give us a better understanding of how well this method works, the original correlation and the perfect reconstruction are also present. The figure presents the reconstruction starting from four different snapshots in the redshift interval $z = 9$ to $z = 0$.

The reconstruction starting at $z = 9$ gives very good results, bringing the decorrelation scale to an intermediate step between the original and the perfect reconstruction. At this redshift most particles are still in the quasi-linear regime, so this result was expected. An interesting feature is the small anti-correlation obtained at large k. This effect is most likely due to particles in non-linear regimes which are past shell-crossing. To understand what gives rise to this anti-correlation, consider two fronts of matter collapsing towards each other. After shell crossing there will be a turn-around as the two evolve into a single filament. If we linearly track these velocities back, we are effectively going the wrong way. This will lead to an anti-correlation over the affected scales. For the $z = 9$ reconstruction, this effect is very small, indicating that shell crossing has only occurred on the smallest scales, and that most particle motions can be well approximated with the linear regime.

this should be explained in chapter 2.

However, starting from lower redshifts produces results much worse than even the original. This was expected, as at these redshifts most particles are now in the non-linear regime. By still treating their motions as linear we are breaking even the correlation that was there to begin with. Figure 3.1 shows the largest scales decorrelating as we move to reconstructions from lower redshifts, and even leading to anti-correlation. For the lowest redshifts, we see a small improvement in the correlation on intermediate scales, but anti-correlation on large scales. This result is much harder to understand. A possible explanation is an extension to the reasons presented above for the small scales in the $z = 9$ reconstruction. The effect of anti-correlation due to particles that are past shell crossing moving the wrong way within the Zel'dovich approximation is expected to increase with decreasing redshift. This might lead to the large scales also becoming anti-correlated. This effect should be studied further, however, we leave this for future works, as our aim in this project is to achieve a good reconstruction.

The Zel'dovich approximation is then not a good reconstruction method at low redshift when most particles have non-linear velocities. For these regimes we need higher orders of LPT to perform the reconstruction. However, at this point it is hard to justify this pursuit from an observational stand point. In this section we have used the peculiar velocities of particles in our simulations. As the ultimate goal of any reconstruction technique is to be used in practice on real data, we need to consider the feasibility of our method. Observers usually detect

find ref

a few galaxies over an $8 - 10$ Mpc scale, and any peculiar velocity measurement inevitably come with errors. This means that the Zel'dovich reconstruction we just performed is very unrealistic in practice. The goal of the rest of this chapter is to modify the Zel'dovich reconstruction to make it more realistic, and also to improve its performance at low redshift.

3.2 Getting back to the linear regime

In order to make the Zel'dovich approximation work for our reconstruction, we must somehow get back into the linear regime. As discussed in Chapter 2, matter tends to be collapsed into filaments at late times. This means individual particles have non linear velocities, but ensembles of particles might still be in the linear regime. Our solution to the two problems outlined in the previous section is to use bulk velocities to calculate the Zel'dovich offset, instead of individual particle velocities.

We smooth particle velocities over 1 Mpc and 10 Mpc scales respectively before calculating the Zel'dovich offset. This means we are now considering bulk motions instead of particle motions. These bulk motions will hopefully provide a better start point when we calculate the Zel'dovich offset. This smoothing also improves the realism of our method. With the technology we currently have,

ref

observers can maybe detect a few galaxies in a 10 Mpc bin, so by smoothing our velocity field over that scale, we simulate a more realistic scenario. The reason for attempting a separate reconstruction using velocities smoothed over 1 Mpc scales is twofold. Firstly, we want to understand the effect of the velocity smoothing scale on the reconstruction. Secondly, we use the 1 Mpc case as a test for what could be achieved with improving technology and a better handling of systematics which could be useful for the next generation of Galaxy Surveys.

To perform these reconstructions, we first split a simulation into bins of a given size: $(1Mpc)^3$ or $(10Mpc)^3$. We then use the positions of the particles to identify the bin they are in. After that, an average velocity over the particles in each bin is calculated. This average velocity is assigned to the centre of the bin. In this manner, we construct a three-dimensional grid which contains a measure of the average velocity field. Finally, we use this average velocity field to linearly interpolate the value of the velocities at the particle positions. In this manner, velocities are smoothed over the scales of interest. Using these new velocities, the Zel'dovich reconstruction is performed as outlined in the previous section.

To further improve the practicality of our method, we also attempt a second type of velocity smoothing. We perform all the steps outlined above to create an average velocity field, but this time we use smaller bins: $(0.5Mpc)^3$ in size. We

then use a Gaussian Filter to smooth this field over the scales of interest (1 Mpc and 10 Mpc respectively). From here, the procedure carries on as outlined above. By first measuring the average velocity field in $(0.5Mpc)^3$ bins, we are closer to what an observer would detect. Galaxies are at most a few hundred thousand Kpc across . Even though we may be able to measure velocities on these small scales, we only detect a few galaxies on 10 Mpc scales.Also, on these small scales, velocities are most likely still in the non linear regime. This leads us to perform the Gaussian smoothing over larger scales.

find ref

ref

Before we move on to the results, an interesting side effect that should be mentioned showed up during the reconstruction. The nature of our method implies that we are creating coherent movements of particles. This coherent movement leaves large gaps in our reconstructed density field (regions where the density field is equal to 0). These gaps become a problem when we want to take the logarithm of the density field as discussed in Chapter 3. When calculating the density field, *Pynbody* uses a smoothing kernel which normally fills in these gaps. However, this method uses N nearest neighbours to calculate the smoothing scale in a region. Normally, there still are a few particles even in the largest voids. These particles will have very far away neighbours, imposing a large smoothing scale. On the other hand, by creating coherent movements, completely empty regions arise. Particles on the edges of these empty regions can easily find nearby neighbours and establish a relatively small smoothing scale. Our solution is to manually find these empty regions and assign a very small value to the density field.

3.3 Results

The first step in our investigation is to understand the effect of the velocity smoothing scale on the reconstruction. In section 4.1 we found that the Zel'dovich reconstruction works very well when starting from $z = 9$. We expect our new methods to have a similar performance when starting from this quasi-linear regime.

Figure 3.2 shows the impact of the velocity smoothing on the $z = 9$ Zel'dovich reconstruction. In this case we only present the first method of calculating an average velocity over 1 Mpc and 10 Mpc scales. For brevity we refer to them as 1 Mpc reconstruction and 10 Mpc reconstruction. As expected, when starting at $z = 9$, both methods work in reconstructing the density field. However, the 1 Mpc reconstruction performs better than the 10 Mpc one. This demonstrates a further loss of information when smoothing the velocity field. The 1 Mpc reconstruction produces a correlation which is very close to the original Zel'dovich reconstruction.

Once we know the impact of the smoothing scale and that reconstruction still

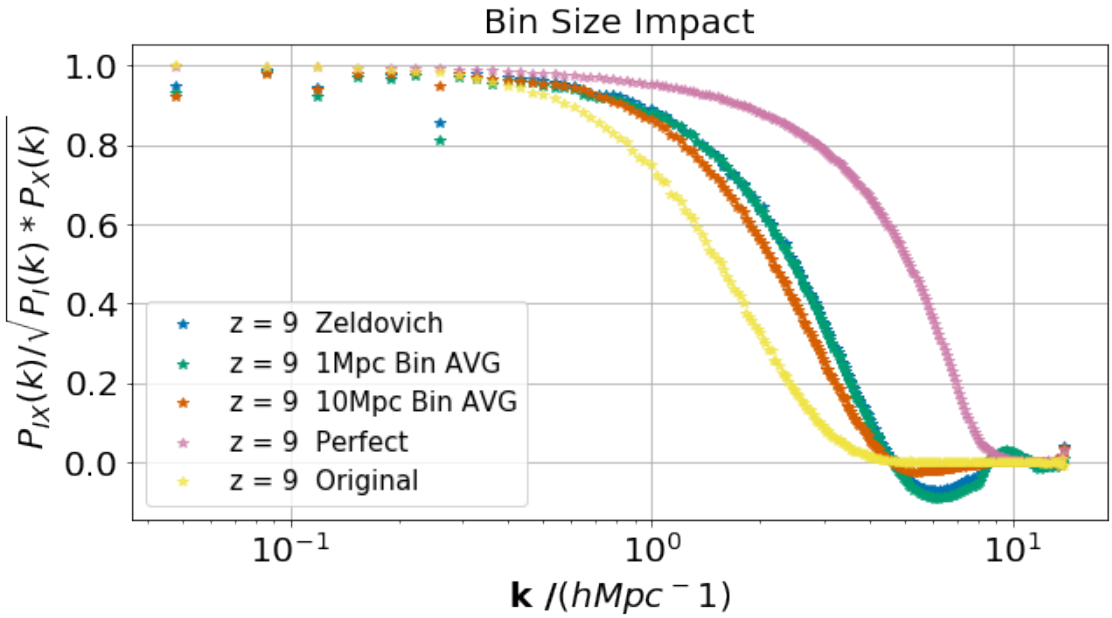

Figure 3.2: Normalized cross-spectra of realistic reconstructions with the initial field. This figure shows the impact of smoothing velocities over certain scales. Because we lose information by smoothing, the 10 Mpc reconstruction recovers less information compared to the 1 Mpc reconstruction. The Zel'dovich reconstruction is also present for comparison. As we are looking at reconstructions from $z = 9$, they all work quite well because most particles are still in the quasi-linear regime.

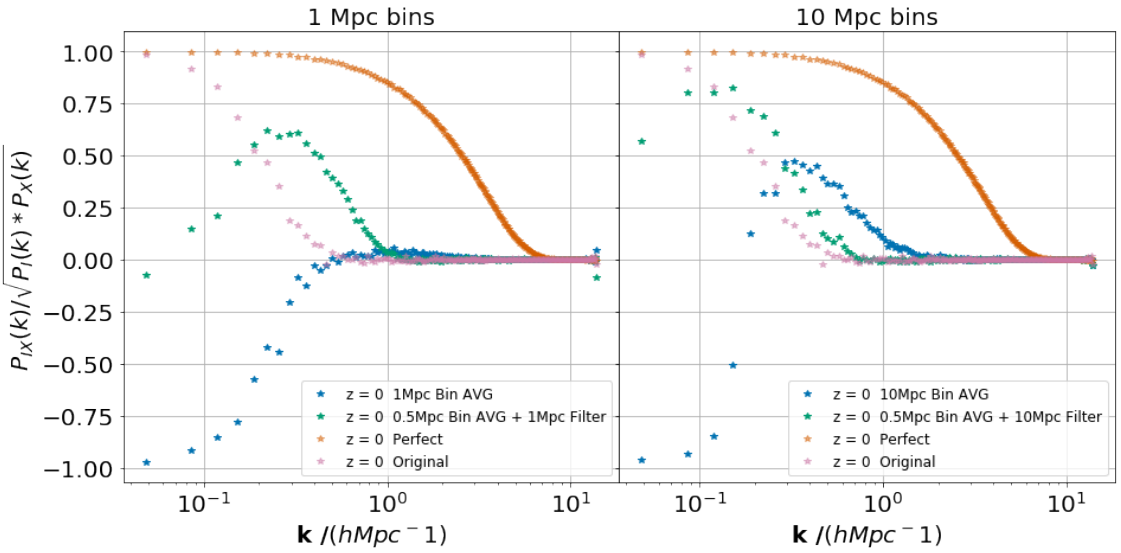

Figure 3.3: Normalized cross-spectra of fields reconstructed from $z = 0$ with the initial density field. We compare the two methods of smoothing velocities outlined in section 4.2 for the two scales of interest. The left panel shows the comparison between the performance of the two methods when velocities are smoothed over 1 Mpc scales, while the right panel shows 10 Mpc scales. If we just average velocities over those scales, we still obtain an anti-correlation over the largest scales. However, when averaging over 0.5 Mpc scales and then applying a Gaussian Filter over the required scales, we see a large improvement in the reconstruction. An interesting effect is that for the 10 Mpc reconstructions, the first method slightly outperforms the second on intermediate scales.

performs very well when starting from $z = 9$, we now want to compare their performance when starting from $z = 0$. Figure 3.3 shows the correlation between the $z = 0$ field, reconstructed with the two methods, and the initial field.

A normal velocity average over the interest scales produces results similar to what we found in section 4.1 with the Zel'dovich reconstruction. We recover some information on intermediate scales, but the large scales become anti-correlated. Here this seems to be taken to an extreme, in the sense that we have almost perfect anti-correlation on the largest scales. This result is very interesting and should be studied further in the future.

On the other hand, when we use smaller averaging bins and a Gaussian filter, we get a much better correlation. We recover quite a bit of information over intermediate to large scales. However, the largest scales still seem to decorrelate. We will investigate this further in the next section where we compare results from different simulations.

We found that using a Gaussian filter to smooth velocities works much better than normal averaging within the Zel'dovich reconstruction. This indicates that just averaging velocities in a 1 Mpc or even a 10 Mpc bin is not enough to bring

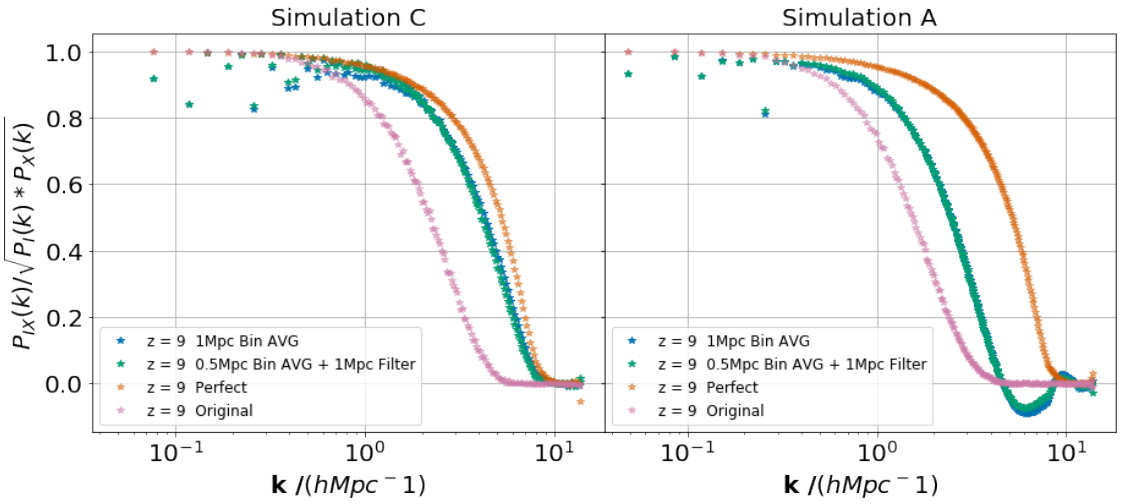

Figure 3.4: Normalized cross-spectra of reconstructed $z = 9$ fields with the initial field. We compare the same reconstruction methods (1 Mpc smoothing only) applied to two different simulations. On the left is Simulation C with a size of $(100Mpc)^3$ and on the right is simulation A with a size of $(200Mpc)^3$. The reconstruction in the smaller simulation recovers a lot more information, bringing the correlation very close to the perfect reconstruction.

us back into the quasi-linear regime. We do see an improvement in the reconstruction of intermediate scales when going from 1 Mpc to 10 Mpc averaging, but it still leads to an anti-correlation of the large scales.

An interesting outcome of these procedures is also the fact that for a small region, on intermediate scales (for 10 Mpc reconstructions), we do obtain a better correlation using normal averaging. This is probably an indication that we are losing more information when we smooth the field with a Gaussian filter due to the tail of the function. However, this tail is most likely also the reason this method works up to much larger scales. By taking into account velocities of particles further away than the scale of interest (even though they have a small weight) we are getting closer to the quasi-linear regime.

3.4 Analysis

We now turn our attention to how these methods work when we apply them to different simulations. We have so far only looked at results from Simulation A as it's the largest one and it has the best resolution. Figure 3.4 presents the correlation of the $z = 9$ reconstructed field with the initial density field in Simulation C and Simulation A. It shows that we are recovering information up to a smaller scale in the smaller simulation (C). This is not entirely surprising, as the smaller simulation is better correlated to begin with (as discussed in Chapter 3). What is

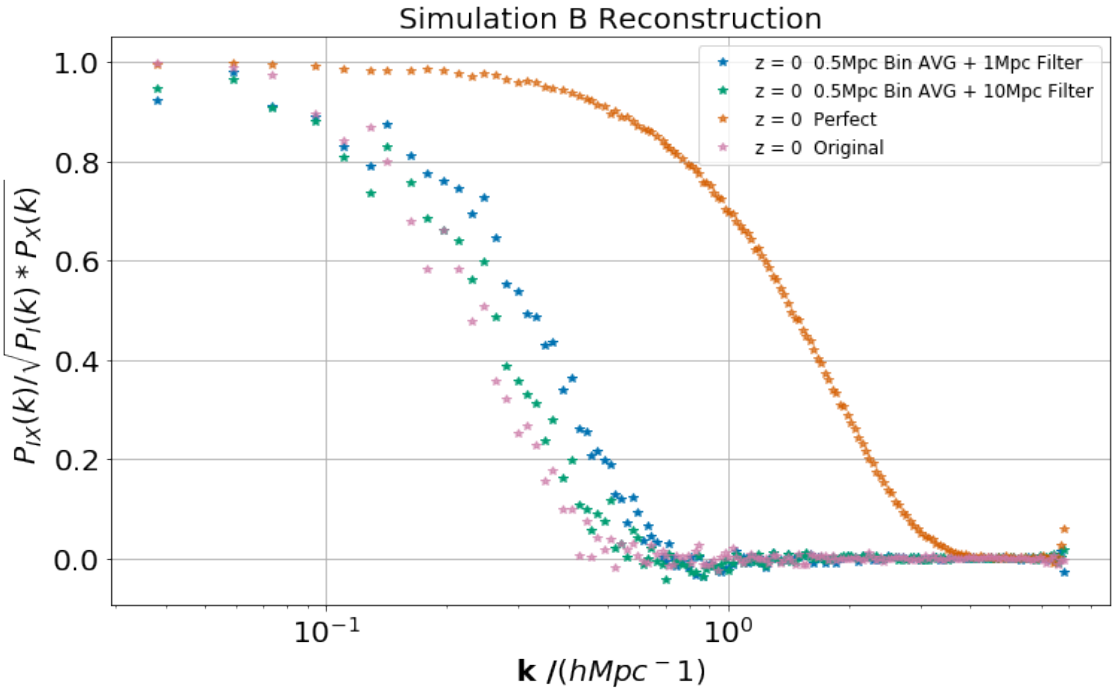

Figure 3.5: Normalized cross-spectra of the $z = 0$ reconstructed fields with the initial density field within Simulation B. Our reconstruction methods work very well in this low resolution simulation in the sense that we don't destroy the correlation that was there to begin with (the largest scales). We also recover some information on intermediate scales, however, it is very far away from the perfect reconstruction. Also, we again encounter the effect found in section 4.3: we recover less information when we smooth velocities over 10 Mpc compared to 1 Mpc.

surprising is how close to the perfect reconstruction the correlation gets.

The best overall correlation when reconstructing the $z = 0$ density field was achieved in Simulation B. This result is shown in Figure 3.5. Both the 1 Mpc and the 10 Mpc reconstructions achieve a better correlation than the original on most scales. This simulation also gives us the best solution from a practical standpoint. When we calculate the average velocity field using $(0.5Mpc)^3$) bins we have, on average, about 0.25 particles per bin. This means that most of our bins are actually empty, which for such small bins is exactly the case for observers as well. This is in contrast with simulation A, which has about 2 particles per bin.

The best results for Simulation A are presented in Figure 3.6. Compared to Simulation B, we recover more information on intermediate scales. However, the largest scales decorrelate. There is also a larger difference between using a 1 Mpc and a 10 Mpc smoothing range. When using a 10 Mpc filter, we recover less intermediate scale information, but the large scales are much better correlated. This is a very good showcase for a trend we have been seeing in this chapter. When we choose to smooth the velocity field we are giving up some information in

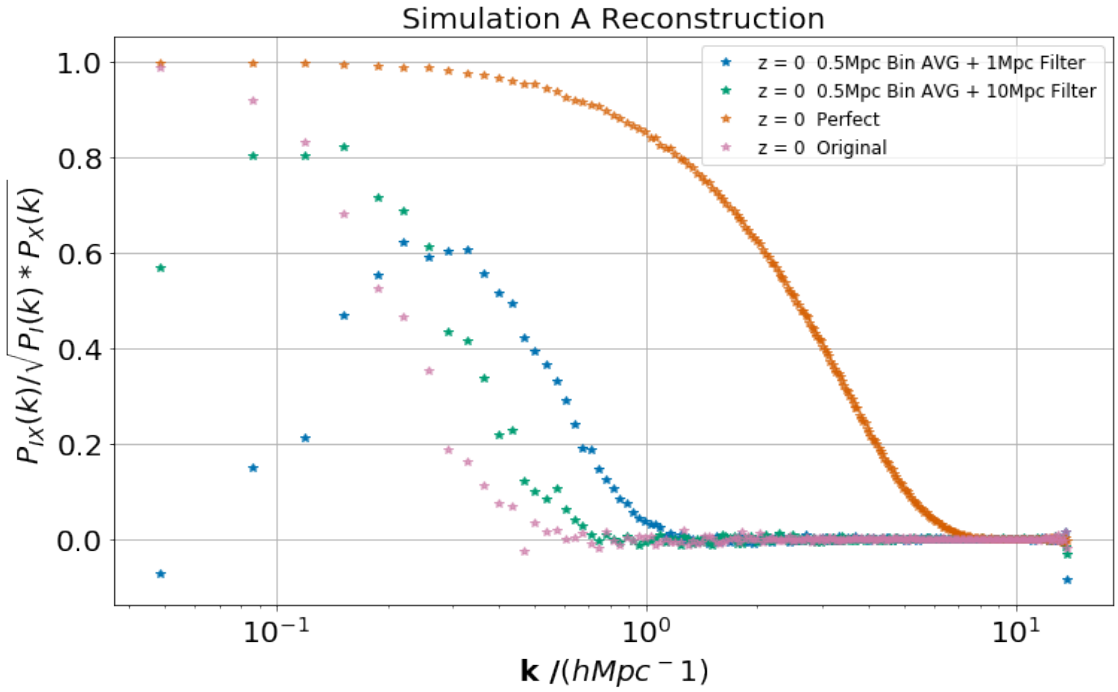

Figure 3.6: Normalized cross-spectra of the $z = 0$ reconstructed field with the initial density field within Simulation A. By smoothing velocities over 10 Mpc scales, we achieve a better correlation of the large scales, however we have a worse correlation over intermediate scales. The result for the intermediate scales was discussed in section 4.3. It is an outcome of the information lost when smoothing velocities. The interesting result is that within the Zel'dovich approximation, we manage to preserve the large scale correlation only by using less velocity information.

the hope of recovering the linear regime. The more velocity information we use, the more we destroy the large scale correlation by using the Zel'dovich approximation. However, using more velocity information generally leads to a better reconstruction on intermediate scales. This is an interesting information trade-off. The more intermediate scale information we recover, the more large-scale information we lose.

On the other hand, based on these results, a case could be made that by using less velocity information, we are just pushing the problem to larger scales. We still see the largest scales start to decorrelate even when using the Gaussian filter. If we had a much larger simulation with the same resolution we might still see the largest scales become anti-correlated. However, our initial goal was to study reconstruction on scales smaller than the BAO. In that respect, our methods do succeed, as we see an improvement in correlation on intermediate scales.

These results are still quite far away from the correlation we obtain with the perfect reconstruction. However, we only studied reconstructions within the Zel'dovich approximation. Higher orders of Lagrangian Perturbation Theory (for

example 2LPT) might be able to get us closer the perfect reconstruction, and also solve the problem with the information trade-off.

Chapter 4

Conclusions

4.1 Information loss

Talk about the inevitable information loss and the big discrepancy between the perfect and realistic reconstructions.

4.2 Future Work

Talk about the problems encountered and Future Work.

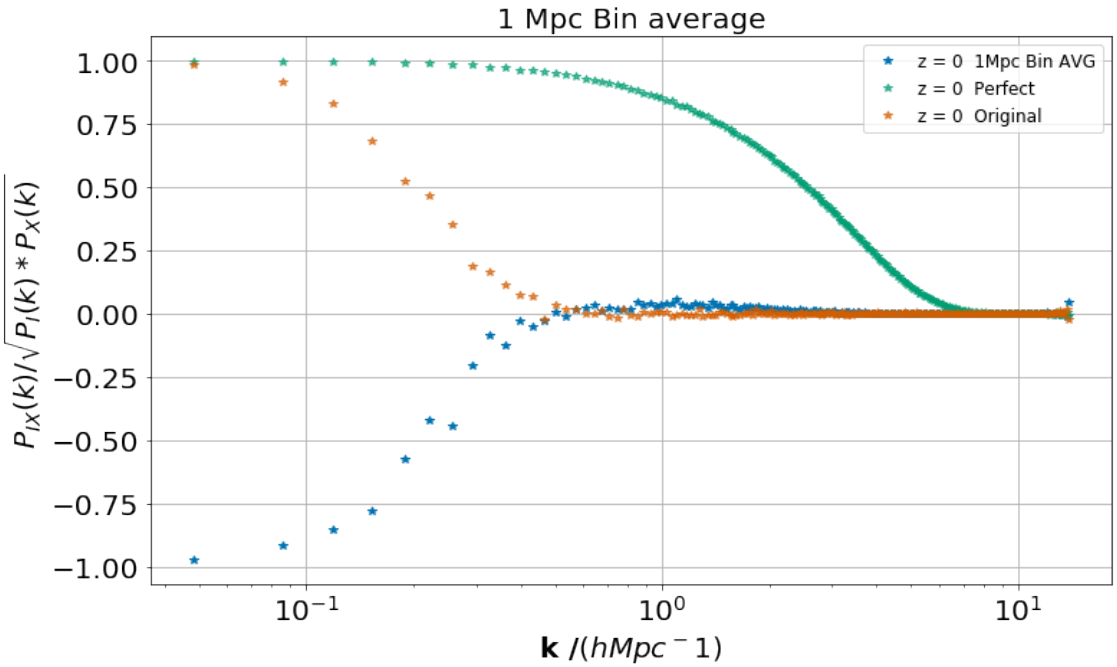

Figure 4.1: Simulation A reconstruction

Bibliography

Angulo, R. E. et al. (2012). "Scaling relations for galaxy clusters in the Millennium-XXL simulation". In: MNRAS 426, pp. 2046–2062. DOI: 10.1111/j.1365-2966.2012.21830.x. arXiv: 1203.3216.

Bernardeau, F. et al. (2002). "Large-scale structure of the Universe and cosmological perturbation theory". In: Phys. Rep. 367, pp. 1–248. DOI: 10.1016/S0370-1573(02)00135-7. eprint: astro-ph/0112551.

Bird, S. (2017). *GenPK: Power spectrum generator*. Astrophysics Source Code Library. ascl: 1706.006.

Brenier, Y. et al. (2003). "Reconstruction of the early Universe as a convex optimization problem". In: MNRAS 346, pp. 501–524. DOI: 10.1046/j.1365-2966.2003.07106.x. eprint: astro-ph/0304214.

Buchert, T. (1992). "Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies and the 'Zel'dovich approximation'". In: MNRAS 254, pp. 729–737. DOI: 10.1093/mnras/254.4.729.

— (1994). "Lagrangian Theory of Gravitational Instability of Friedman-Lemaitre Cosmologies - a Generic Third-Order Model for Nonlinear Clustering". In: MNRAS 267, p. 811. DOI: 10.1093/mnras/267.4.811. eprint: astro-ph/9309055.

Buchert, T. and J. Ehlers (1993). "Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies – second-order approach: an improved model for non-linear clustering". In: MNRAS 264. DOI: 10.1093/mnras/264.2.375.

Carlson, J., M. White, and N. Padmanabhan (2009). "Critical look at cosmological perturbation theory techniques". In: Phys. Rev. D 80.4, 043531, p. 043531. DOI: 10.1103/PhysRevD.80.043531. arXiv: 0905.0479 [astro-ph.CO].

Conselice, C. J. et al. (2016). "The Evolution of Galaxy Number Density at z = 8 and Its Implications". In: ApJ 830, 83, p. 83. DOI: 10.3847/0004-637X/830/2/83. arXiv: 1607.03909.

Crocce, M. and R. Scoccimarro (2006). "Renormalized cosmological perturbation theory". In: Phys. Rev. D 73.6, 063519, p. 063519. DOI: 10.1103/PhysRevD.73.063519. eprint: astro-ph/0509418.

— (2008). "Nonlinear evolution of baryon acoustic oscillations". In: Phys. Rev. D 77.2, 023533, p. 023533. DOI: 10.1103/PhysRevD.77.023533. arXiv: 0704.2783.

Croft, R. A. C. and E. Gaztanaga (1997). "Reconstruction of cosmological density and velocity fields in the Lagrangian Zel'dovich approximation". In: MNRAS 285, pp. 793–805. DOI: 10.1093/mnras/285.4.793. eprint: astro-ph/9602100.

Davis, M. et al. (1982). "A survey of galaxy redshifts. II - The large scale space distribution". In: ApJ 253, pp. 423–445. DOI: 10.1086/159646.

Desjacques, V., D. Jeong, and F. Schmidt (2016). "Large-Scale Galaxy Bias". In: *ArXiv e-prints*. arXiv: 1611.09787.

Eisenstein, D. J. et al. (2007). "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak". In: ApJ 664, pp. 675–679. DOI: 10.1086/518712. eprint: astro-ph/0604362.

Gramann, M. (1993). "An improved reconstruction method for cosmological density fields". In: ApJ 405, pp. 449–458. DOI: 10.1086/172377.

Komatsu, E. et al. (2011). "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation". In: ApJS 192, 18, p. 18. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538 [astro-ph.CO].

Levi, M. et al. (2013). "The DESI Experiment, a whitepaper for Snowmass 2013". In: *ArXiv e-prints*. arXiv: 1308.0847 [astro-ph.CO].

LSST Science Collaboration et al. (2009). "LSST Science Book, Version 2.0". In: *ArXiv e-prints*. arXiv: 0912.0201 [astro-ph.IM].

Matsubara, T. (2008a). "Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture". In: Phys. Rev. D 78.8, 083519, p. 083519. DOI: 10.1103/PhysRevD.78.083519. arXiv: 0807.1733.

— (2008b). "Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space". In: Phys. Rev. D 77.6, 063530, p. 063530. DOI: 10.1103/PhysRevD.77.063530. arXiv: 0711.2521.

Monaco, P. and G. Efstathiou (1999). "Reconstruction of cosmological initial conditions from galaxy redshift catalogues". In: MNRAS 308, pp. 763–779. DOI: 10.1046/j.1365-8711.1999.02747.x. eprint: astro-ph/9902119.

Nusser, A. and A. Dekel (1992). "Tracing large-scale fluctuations back in time". In: ApJ 391, pp. 443–452. DOI: 10.1086/171360.

Padmanabhan, T. (1993). *Structure Formation in the Universe*, p. 499.

Peebles, Phillip James Edwin (1980). *The large-scale structure of the universe*. Princeton university press.

Planck Collaboration et al. (2016). "Planck 2015 results. XIII. Cosmological parameters". In: A&A 594, A13, A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.

Pontzen, A., R. Roškar, et al. (2013). *pynbody: N-Body/SPH analysis for python*. Astrophysics Source Code Library. ascl: 1305.002.

Pontzen, A. et al. (2016a). "Inverted initial conditions: Exploring the growth of cosmic structure and voids". In: Phys. Rev. D 93.10, 103519, p. 103519. DOI: 10.1103/PhysRevD.93.103519. arXiv: 1511.04090.

— (2016b). "Inverted initial conditions: Exploring the growth of cosmic structure and voids". In: Phys. Rev. D 93.10, 103519, p. 103519. DOI: 10.1103/PhysRevD.93.103519. arXiv: 1511.04090.

Springel, V. (2005). "The cosmological simulation code GADGET-2". In: MNRAS 364, pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. eprint: astro-ph/0505010.

Springel, V. et al. (2008). "The Aquarius Project: the subhaloes of galactic haloes". In: MNRAS 391, pp. 1685–1711. DOI: 10.1111/j.1365-2966.2008.14066.x. arXiv: 0809.0898.

Vishniac, E. T. (1983). "Why weakly non-linear effects are small in a zero-pressure cosmology". In: MNRAS 203, pp. 345–349. DOI: 10.1093/mnras/203.2.345.

Vogelsberger, M. et al. (2014). "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe". In: MNRAS 444, pp. 1518–1547. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921.

Weinberg, D. H. (1992). "Reconstructing primordial density fluctuations. I - Method". In: MNRAS 254, pp. 315–342. DOI: 10.1093/mnras/254.2.315.

Weinberg, D. H. et al. (2013). "Observational probes of cosmic acceleration". In: Phys. Rep. 530, pp. 87–255. DOI: 10.1016/j.physrep.2013.05.001. arXiv: 1201.2434.

White, M., ed. (1999). *Anisotropies in the CMB*. eprint: astro-ph/9903232.

Zel'dovich, Y. B. (1970). "Gravitational instability: An approximate theory for large density perturbations." In: A&A 5, pp. 84–89.